

负责任创新视角下深度合成技术的传播伦理风险识别 与善治机制

孔昊瑜

广东外语外贸大学新闻与传播学院 广东 广州 510006

【摘要】：深度合成技术是中性的技术形态，往左走是“深度伪造”，往右走是“深度创造”。新兴的深度合成技术广泛应用于传媒领域，其带来的传播伦理风险日益复杂。既有研究多集中于“深度伪造”的负面影响，而忽视了其外延更广的“深度合成”在传播全链路中的伦理风险。为解决这一问题，本文以负责任创新理论为出发点，基于拉斯韦尔“5w”传播模式，从传者、内容、渠道、受众与效果五个维度系统探索深度合成技术涉及的伦理风险，包括传者价值嵌入风险、内容伪造风险、渠道责任风险、受众权益风险与意识形态传导风险。在此基础上，提出由技术治理、技术人文、责任压实和国际协同组成的负责任创新导向下的善治机制，以期为构建负责任、有底线的深度合成技术传播创新体系提供理论支撑。

【关键词】：深度合成；深度伪造；伦理风险

DOI:10.12417/3041-0630.26.02.060

1 引言：深度合成技术的传播伦理挑战

人工智能是引领未来发展的战略性技术，也是驱动经济发展、增强国家竞争力的关键力量。深度合成（Deep Synthesis）作为人工智能的重要应用，是指基于深度学习等智能化方式生成或编辑加工音频、视频、图像、文本等音视频类内容，其合成内容可高度逼真、难以分辨^[1]。在传播实践中，深度合成技术不仅扩展了内容形态，提高了传播效率和智能化水平，也广泛应用于内容生产各领域。

深度合成技术具有“双刃剑”特性，可以被用于内容创新和传播创新，也可以被用来生成大量具有迷惑性的“深度伪造”内容。既有研究多集中于后者，特别是涉政类伪造的负面影响，一定程度上忽视了“深度合成”这一概念所涵盖的多维伦理风险。“深度伪造”一词本身就带有贬义，“深度合成”是中性表达，更适于全面探讨其技术伦理影响。在工具理性与价值理性的拉力结构下，系统识别并治理深度合成技术在传播全流程中的伦理问题，是学术与实践面临的重大课题。深度合成技术或消解或再造新闻真实，其发展必须坚持人的价值导向，应置于工具理性之外，融入价值理性考量。

2 理论基础：传播视域下作为技术伦理的负责任创新

面对日新月异的技术迭代及其带来的颠覆性影响，如何构建技术发展与伦理规范之间的良性互动关系，成为一项关键课题。《华盛顿邮报》在抨击 OpenAI 以牺牲安全为代价追求模型快速迭代时指出，应对这一挑战的有效路径是推行“负责任创新”^[2]。

负责任创新（Responsible Innovation），又称负责任的研究和创新（Responsible Research and Innovation, RRI），源于学界对责任伦理的讨论。强调在技术研发中，技术主体、创新活动与反馈机制应在道德框架下形成良性互动。技术引领者需前瞻性地评估创新可能带来的后果与影响，坚持以道德价值为导向，从而引导技术朝着符合社会期待的方向发展^[3]。冯·尚伯格（René Von Schomberg）对负责任研究与创新也进行了相关界定，强调关注创新的过程及其伦理的可持续性、可接受性和社会可取性，认为负责任研究与创新是一个透明互动的过程，在这一过程中，科技进步与社会生活实现协调发展^[4]。

从负责任创新的视角看，技术发展和伦理建设并非互相制约，也不是一面强则另一面弱的零和游戏。一方面，技术的深度融入可能导致新闻真实受损、数字遗忘困境、信息过载等风险。另一方面，禁止或停止技术发展并非长远之计。历史经验表明，人类不会放弃对科技的追求，伦理不应成为创新的枷锁，而应成为创新过程中的助力。

因此，本文以负责任创新为出发点，将伦理融入传播的各个阶段，旨在通过创新过程符合道德期望和社会目的，使结果同时满足用户需求、道德原则和社会期望。从而实现既符合用户又符合道德原则和社会期望的技术结果，尤其适用于深度合成技术在传播领域的引导性应用和风险治理。

3 潜在伦理风险识别：基于“5W”模型的传播全链条剖析

深度合成技术作为一把“双刃剑”，在拓展传媒可能性的

作者简介：孔昊瑜（2005—），女，汉族，广东省清远市，本科在读，广东外语外贸大学新闻与传播学院，研究方向：技术伦理、国际传播。

同时,也伴生着不容忽视的伦理风险。为系统识别其潜在问题,本研究以拉斯韦尔(Harold Lasswell)的“5W”传播模式为分析框架,从传者(Who)、内容(Says What)、渠道(In Which Channel)、受众(To Whom)、效果(With What Effect)五个维度,对深度合成技术在传播链条中各环节可能引发的伦理风险进行系统性审视。

3.1 传者视角: 技术设计中的价值渗透与主体伦理缺位

技术本身不是价值中立的,其应用体现为两种形态:技术深度合成或技术深度伪造,由此导致的价值反差极化也构成了智能化时代主流意识形态传播的新挑战和新机遇。马克思也曾从技术的资本主义应用角度否认了技术中立主义的观点^[5]。在深度合成技术的运行逻辑中,从架构到训练到规则,每一个环节都蕴含技术人员、机构甚至行业领域关于价值伦理的判断和考量。当技术嵌入商业化传播语境,一些主体为了吸引流量和变现,可能会有意识地将迎合性、煽动性、偏见性等要素输入到合成过程中,使技术成为价值引导和意识形态传播的隐形载体,这种有偏见的嵌入式价值既挑战了新闻从业人员的职业伦理,也威胁着代表主流价值的意识形态传播秩序,因此必须被纳入传播伦理的批判性考察范畴。

3.2 内容视角: 深度伪造对新闻真实性与信息生态的冲击

当生成式的人工智能技术日益成熟,“深度伪造”手法已能轻松制造以假乱真的假新闻。互联网本就是谣言的集散地,人工智能将虚假信息生产带上了“真假难辨”之路^[6]。学术界将利用人工智能生成假事实的技术称为深度伪造(Deepfake)。新闻的真实性是新闻业的立身之本,也是吸引公众信任的基础。深度伪造技术生成逼真的假影像、假声音和假文字,不仅否定具体内容的真实性,更对社会共同依赖的“真实”产生质疑,污染媒介生态、使传播秩序失范,是对媒介生态的毒害。这种技术滥用已经对国家信息安全、个体新闻专业主义和新闻专业规范产生挑战^[7]。更大的风险在于,当大量的、难以辨别的虚假内容存在,公众将逐渐陷入持续的怀疑之中,甚至产生“一切都可以被伪造”的怀疑论,最终导致整个信息环境的信任崩塌和公共对话机制的失效。

3.3 渠道视角: 责任归属模糊与平台把关机制的双重困境

深度合成产品的传播链条牵扯上下游多主体、多环节的协同合作,而人工智能本身的“黑箱”特性,也加大了相关主体责任认定的难度。从创作生成、编辑加工、平台分发到用户二次传播,记者、算法工程师、内容创作者、平台企业、数据供给服务提供者等多方主体共同协作,很难在现行法律、伦理责任框架下厘清责任边界。与此同时,面对呈海量、实时、生成式爆发的深度合成内容,传统编辑把关机制面临巨大挑战。不同的传播渠道对内容把关要求的标准不同、人工把关能力有

限,自动化的把关技术尚未成熟,内容把关能力存在失灵和滞后,责任主体的失焦与把关机制的失能,是深度合成治理面临的结构性困境,也是虚假、有害信息得以流通的重大缺口。

3.4 受众视角: 数字人格权侵害与身份盗用风险

深度合成技术的广泛使用,让普通群众越来越成为深度合成的受害者。滥用最突出的表现之一,是未经许可使用他人肖像、声音等生物特征信息从事内容合成,如换脸、换表情、换声音,乃至在娱乐目的掩盖下侵犯他人肖像权、名誉权等数字人格权益。更有甚者用盗用的他人身份形象从事营销带货、诈骗等“数字身份盗窃”活动^[8]。这种侵犯不只侵犯个体的尊严和合法权益,也在更广泛水平上损坏数字社会中的身份信任,影响网络互信和相处伦理。

3.5 效果视角: 意识形态传导失衡与社会信任机制危机

从社会传播效果来看,深度合成的无序应用,可能会导致社会深层次的信任危机和意识形态安全风险。如果公众长期暴露在难以辨别的合成信息中,其判断的基础将不复存在,进而引发公众焦虑、意见极化和共识难以形成。从长期来看,这种真实性危机将逐渐蚕食媒体公信力、削弱主流传播力,甚至影响社会凝聚力和治理效能。在国际传播领域,深度合成也可能被用来构建信息泡沫、干扰舆论走向,甚至攻击国家形象和意识形态安全。政府、平台和社会如何从社会效果层面预判和管理这类技术带来的认知和信任风险,是一场亟待破解的治理考验。

4 善治机制探讨: 构建负责任创新的实践路径

深度合成技术是“潘多拉魔盒”还是“通往未来之门”,取决于人类如何使用技术,取决于人类能否趋利避害,共同推动科技向善。英国经济学家欧文(Richard Owen)指出,负责任创新注重技术创新活动与社会责任的有机结合,构建了由预测性、包容性、反身性以及反馈性组成的四维度体系^[9]。本研究基于前述伦理风险,结合“负责任创新”理论框架,从技术发展、责任落实与国际协同三个层面,提出如下治理路径。

4.1 预测性维度: 布局正向技术生态与风险预警

深度合成是人工智能领域最尖端、最核心的技术之一,在技术浪潮的推动下把握主动、服务大局,更好肩负起人工智能发展的责任,是践行负责任创新的题中之意。负责任创新的预测维度主张对技术的未来发展进行前瞻规划与主动引导,更好服务国家战略需求和公共利益。我国应当充分发挥新型举国体制优势,系统布局关键算法、检测工具、治理技术等方面的研发攻关,努力掌握技术标准和产业主导权^[10],把深度合成技术未来可能出现的潜在影响和技术伦理问题尽早融入研发过程,在研发设计阶段尽可能规避其潜在风险。新闻媒体应当发掘深度合成在内容创新、传播优化等方面的有益应用模式,例如创

新融媒产品,提升传播效能;打造适配不同终端和场景的个性化产品,更好满足用户需求,从而在造福社会大众的同时,积累有益于公共利益的技术实践和治理经验。

4.2 包容性维度:增强技术理性下的人文关怀

包容性维度主张在数字技术创新过程中纳入更多视角,特别是最容易被技术进程边缘化的主体和价值。在深度合成的传播实践中,对工具理性的追求,如对效率和流量的追求往往忽视了对人的尊严、文化多样性和社会信任的关照。在技术设计层面,应努力识别和降低算法偏见与歧视的风险;在应用层面,应尊重用户的文化背景和主体权利,避免生成排他性、刻板化的表达。媒体和平台应承担社会责任,提供简单易用的辨识工具、风险提示和反馈渠道。数字化时代使得“人机共生”的社会逐渐成为现实,从负责任创新的包容性维度出发^[11],我们需要超越技术发明的单向思考,提升人机协同功效,构建一个既敢于创新又充满人文温度、既高效又有公平的空间。

4.3 反身性维度:压实主体责任

反身性维度要求对创新过程及其后果进行反思,并建立相应的责任制度。深度合成治理的症结在于责任“是谁”和“何时缺席”不明。因此,治理的起点是厘清和落实责任:新闻记者和媒体应增强专业伦理意识和把关能力;技术开发者应秉持伦理设计,在算法中嵌入可解释性和可控性机制;平台对内容治理有重要节点责任,应承担审核责任。此外,还可以建立强制标识和风险分级制度。借鉴欧盟、德国等地的治理经验,结合我国实际构建一个分类精准、响应迅速的治理体系,从被动

处置转向主动调控。

4.4 反馈性维度:深化国际合作与国际协同治理

反馈性维度要构建动态响应、持续改进的善治生态。深度合成及其风险具有跨界特点,因此,单凭一国之力难以善治^[12]。因此,需要跨国家的凝聚共识与协同行动。一方面,可以倡导“网络空间命运共同体”,推动形成技术治理国际通用规则,反对安全与伦理双重标准;另一方面,积极主动参与并推动构建跨国技术检测网络、数据共享平台和联合治理机制,以提升识别和处置诸如深度伪造类型的恶意行为的能力。还可以探索建立“公众参与式”事实核查机制,鼓励媒体、专业机构和公众开展合作,基于群体智慧提高社会识谣辟谣能力。真正可持续、负责任的深度合成技术治理生态,需要多层次、跨主体的国际协同。

5 结语:迈向负责任创新:平衡技术理性与传播伦理

深度合成技术作为智能传播的典型应用,在带来新的媒介可能性的同时,也对传播伦理的实践语境和风险形态进行了重构。社会放大效应表明,即使客观风险极小,公众的风险感知也会造成巨大的社会疑虑并引发广泛的治理反应^[13]。因此,必须在技术发展与伦理规范之间寻找平衡,推动形成“负责任”的协同治理体系。

深度合成技术的健康发展最终取决于能否在工具理性与价值理性之间找到合适的平衡点,实现对包容创新与规范治理的兼容,真正服务好信息传播、文化传承和社会进步。

参考文献:

- [1] 梁瑞刚,吕培卓等.视听觉深度伪造检测技术研究综述[J].信息安全学报,2020(2):1-17.
- [2] WALKER D,TANEJA H.AI Changes Everything,We Need New Guardrails to Survive it And soon[EB/OL].<https://www.washingtonpost.com/opinions/2023/03/28/ai-machine-learning-safeguards/>.
- [3] 范毅强,李侠.新兴技术伦理信任的路径研究——以区块链技术的发展为例[J].自然辩证法研究,2020,36(04):37-43.
- [4] SCHOMBERG R V.Prospects for technology assessment in a framework of responsible research and innovation[C]//DUSSELDORP,M,BEECROFT R.Technikfolgen Abschtzen Lehren:Bildungspoten ziale Transdisziplin rer Methoden.Wiesbaden:VS Verlag für Sozialwissenschaften,2012:39-61.
- [5] 马克思.资本论(纪念版):第一卷[M].中共中央马克思恩格斯列宁斯大林著作编译局,译.北京:人民出版社,2018.
- [6] 全会.冲击·融合·协同:ChatGPT对传媒业的影响刍议[J].中国广播电视学刊,2023,(09):17-21.
- [7] 陈昌凤,徐芳依.智能时代的“深度伪造”信息及其治理方式[J].新闻与写作,2020,(04):66-71.
- [8] 张超,陈莎.深度合成的外生性风险及平台治理原则[J].中国新闻传播研究,2022,(05):39-49.
- [9] STILLGOE J,OWEN R,MACNAGHTEN P.Developing a framework for responsible innovation[J].Research policy,2013,42(9):1568-1580.
- [10] 姬德强.深度造假:人工智能时代的视觉政治[J].新闻大学,2020(7):1-16,121.
- [11] 吴越.负责任创新视域下高校生态文明教育体系构建研究[J].社会与公益,2025,(17):240-243.
- [12] 高泽阳.深度伪造对传播业态的机与危[J].中国传媒科技,2024,(10):39-42.
- [13] 杨博文,孙永军.理性赋能与向善赋权:科技伦理风险预警与敏捷治理体系的建构进阶[J].科学技术哲学研究,2023,40(04):122-128.