生成式人工智能的幻觉特性如何反向塑造传媒内容的公信力边界

曹逸群
中国传媒大学，北京 100020
DOI:10.61369/HASS.2025090011

摘　　要： 随着ChatGPT等生成式人工智能在新闻写作、图像生成和舆情分析等传媒环节的广泛应用，媒体生产效率显著提升，但其固有的“幻觉”特性，即在缺乏事实依据时生成似真而假的信息，对内容真实性和媒体公信力构成挑战。本文首先分析生成式AI的幻觉机制及其在传媒内容生产中的表现，其次探讨幻觉特性对新闻真实性、机构依赖和公众信任的影响，最后提出技术防护、机构自律及公众媒介素养提升的综合策略。研究表明，在“人机共创”的传播语境下，重构公信力边界需实现技术、制度与社会认知的协同，保障新闻真实性与受众信任。

关 键 词： 生成式人工智能；幻觉特性；传媒公信力；内容真实性；人机共创

How the Hallucination Feature of Generative Artificial Intelligence Reshapes the Boundaries of Media Credibility

Cao Yiqun
Communication University of China, Beijing 100020

Abstract: With the widespread adoption of generative artificial intelligence (GAI) such as ChatGPT in news writing, image generation, and public opinion analysis, media production efficiency has significantly improved. However, the inherent “hallucination” characteristic of GAI—generating plausible but factually incorrect information—poses a challenge to content accuracy and media credibility. This study first examines the hallucination mechanisms of GAI and their manifestations in media content production. It then analyzes the impact of hallucinations on news authenticity, institutional reliance, and public trust, and finally proposes integrated strategies including technological safeguards, institutional self-regulation, and public media literacy enhancement. The findings indicate that in a “human-AI co-creation” context, reconstructing media credibility boundaries requires the coordinated development of technology, institutional frameworks, and audience cognition to ensure news accuracy and sustain public trust.

Keywords: generative artificial intelligence; hallucination; media credibility; content authenticity; human-AI co-creation

引言

近年来，以ChatGPT等为代表的生成式人工智能迅速渗透至传媒内容生产领域。新闻撰写、图像生成、舆情分析等环节的智能化，使传媒行业的效率与表达能力显著提升。然而，生成式AI在生成内容时存在固有的“幻觉”特性，即在缺乏事实依据的情况下生成似真而假的信息。这种“技术幻觉”源于算法建模与语料偏差，不仅挑战了内容真实性的边界，也引发了媒体公信力的结构性危机。在传统传播体系中，公信力依托于专业把关与事实核验机制，而AI内容的出现打破了“人控生产”的逻辑，使新闻源头更加模糊、责任主体趋于分散。幻觉信息一旦进入传播链条，经由社交媒体与算法推荐被不断复制放大，极易导致“事实似真、真伪难辨”的传播困境。由此，传媒领域亟需重新思考在“人机共创”语境下公信力的构建与边界重塑问题。

一、生成式人工智能与幻觉特性概述

生成式人工智能是当前人工智能发展的重要方向，其核心在于通过深度学习模型对大规模语料进行模式学习和概率预测，从而自动生成具有语义连贯性与创造性的文本、图像或音视频内容。以ChatGPT为代表的语言生成模型，已广泛应用于新闻写

作者简介：曹逸群（1996—），女，汉族，河南南阳人，硕士研究生，研究方向：新闻传播。

作、内容创作与传播分析等领域，显著提升了传媒行业的生产效率与信息处理能力。

然而，生成式AI在带来便利与创新的同时，也存在难以避免的“幻觉”问题，即模型在缺乏事实依据时生成似真而假的内容。这种现象源于其以“语言概率最优”替代“事实真实”的生成逻辑。由于模型本身不具备事实核查能力，其输出常在语言上连贯、形式上可信，却在事实层面存在虚构或错误[1]。

从传播学角度看，AI幻觉带来的风险主要体现在两方面：一是真实性的弱化，生成内容模糊了事实与虚构的界限，削弱了新闻报道的信源可靠性；二是信任的转移，受众在面对流畅自然的AI文本时，容易产生“认知真实感”，从而误将虚构信息视为权威内容。当传媒机构依赖AI生成信息而缺乏人工核查，幻觉内容便可能在算法推荐与社交传播中被放大，引发公信力危机[2]。

因此，生成式人工智能的幻觉特性不仅是技术层面的偏差，更是传媒真实性与信任边界面临重构的关键变量。深入理解其生成机制与传播效应，是探讨AI时代传媒公信力变迁的理论基础[3]。

二、生成式人工智能的幻觉特性与传媒内容的公信力

（一）幻觉特性对传媒内容真实性的挑战

生成式人工智能的“幻觉”现象已成为传媒真实性体系面临的核心挑战。AI模型在生成过程中依赖统计概率与语义预测，而非事实验证与逻辑推理。其生成机制决定了内容的“合理性”往往先于“真实性”，语言上虽连贯流畅，却可能在事实层面出现虚构、混淆甚至捏造。尤其在新闻写作、摘要生成或图像合成等环节中，这种幻觉信息极易被媒体工作者误认为真实材料而被采纳，形成“伪事实”的内容传播。

一旦此类内容以新闻或权威口吻发布，其欺骗性更强、传播性更广，不仅扰乱了信息生态，也冲击了新闻真实性原则这一传媒公信力的核心基石[4]。更深层的问题在于，AI幻觉并非偶然的技术缺陷，而是其算法逻辑与训练语料偏差的必然产物。当生成模型基于概率最优选择而非事实验证进行表达时，它实质上是在“构造一种似真之假”。因此，AI幻觉不仅是算法误差，更是媒体内容生产链中的潜在“信息污染源”，对新闻业长期以来以事实为中心的价值体系构成系统性侵蚀。

（二）传媒机构对人工智能生成内容的依赖与公信力危机

在新闻传播的数字化与自动化转型过程中，传媒机构对生成式AI的依赖程度不断提高。AI被广泛应用于选题策划、文本撰写、舆情监测与数据可视化等环节，大幅降低了人力成本，提升了信息生产的速度与规模。然而，这种依赖正在重塑传媒的权威机制与专业伦理[5]。

首先，AI生成内容普遍缺乏来源标识与事实溯源能力，模糊了“算法生成”与“人工采编”的边界。当媒体未经核实便将AI生成的信息直接发布，幻觉内容极易混入公共传播体系，削弱新闻的真实性与可验证性。其次，AI的“拟真”特征使虚构内容在语言和形式上更具迷惑性，从而降低了编辑审核机制的敏感度。久而久之，新闻机构可能在“效率优先”的逻辑下弱化人工把关，形成以产出速度取代真实性审查的结构性倾向。

更为严峻的是，一旦AI幻觉内容被揭露，媒体将同时遭遇“技术失误”与“信任崩塌”的双重危机。公众的不信任不仅针对特定新闻事件，更会波及整个机构的信誉体系，使媒体长期积累的公信资本遭受侵蚀。由此可见，传媒公信力的危机正是在“自动化效率”与“真实性原则”的张力中不断加剧。

（三）公众对AI生成内容的认知与信任问题

公众对AI生成内容的认知水平与信任态度，直接影响传媒公信力的社会基础。一方面，AI生成文本在语义结构和逻辑连贯性上的优越表现，使部分受众误以为“机器更理性、更客观”，从而形成对AI输出的过度信任[6]。这种“技术客观幻觉”使受众在潜意识中将流畅表达等同于真实信息，削弱了对内容真伪的主动辨识。另一方面，随着幻觉事件频繁曝光，公众对AI的信任又迅速转向怀疑甚至排斥，进而延伸为对媒体的不信任。这种情绪反转使传媒机构陷入“双重信任困境”——既要应对AI幻觉带来的内容风险，又要面对公众认知动荡带来的信任流失。

在“信息过载”与“算法推送”并存的媒介环境中，受众的注意力与判断力被进一步分散，形成认知疲劳与审查麻木。真假信息的交织使公众在事实判断上趋于依赖系统默认，导致“信任自动化”现象的出现[7]。结果是，传媒公信力不仅受制于AI生成内容本身的真实性，更受到公众信任逻辑的变化所影响。如何在AI赋能与信任维护之间重建平衡，成为传媒业必须面对的现实课题。这一过程需要技术透明、机构自律与公众教育的共同作用，方能在“人机共创”的传播格局中维系媒体的社会信任与价值边界。

三、应对生成式人工智能幻觉特性塑造公信力边界的策略

（一）技术层面的应对措施

从技术层面来看，减少生成式人工智能幻觉现象、保障传媒内容真实性，是重塑公信力的首要环节。首先，应完善生成模型的事实校验与信息验证机制。通过引入知识图谱、检索增强生成（RAG）等技术，实现“生成—验证”的双重路径，可以在内容生产过程中同步校验信息的可靠性，从源头降低虚假内容的出现概率。其次，建立内容溯源与可追踪体系同样不可或缺。借助数字水印、区块链存证等技术，为生成内容添加可识别标识，不仅提升信息透明度，也确保在传播链条中责任可追溯，从而增强公众对内容来源的信任[8]。此外，媒体机构应与技术研发团队密切合作，推动幻觉检测算法的标准化、工具化及行业化，使AI生成内容在生产环节即可实现自动识别、筛查与预警。值得强调的是，技术手段虽能有效减少幻觉风险，但其作用有限，必须与制度规范和受众认知能力相结合，才能形成系统化的防护体系，实现对传媒公信力的全面保障。

（二）传媒机构的自律与监管

在AI深度介入新闻生产的背景下，传媒机构自律成为维护

公信力的核心环节。媒体应建立严格的生成内容管理制度，明确AI内容使用范围、标识要求与责任归属，防止“无标识传播”带来的事实模糊与责任模糊问题。同时，应强化人工审核机制，确保编辑在真实性核查、伦理判断与信息筛选中保持主导地位，防止AI“拟真”特性削弱人工把关。除了机构自律，行业监管部门也需完善政策与制度保障，例如落实《生成式人工智能服务管理办法》，推动建立新闻机构AI应用备案、内容可追溯及责任追责机制。通过技术手段与制度约束的双重作用，可形成覆盖生成、发布与传播全过程的防护网络，既遏制幻觉内容的扩散，又保护媒体的专业声誉和社会公信力。此外，行业标准的统一化和透明化，也有助于增强公众对AI生成内容的信任，使监管成为公信力维护的有力支撑[9]。

（三）公众教育与媒体素养的提升

公众的媒介认知能力是传媒公信力持续维系的社会基础。面对生成式AI内容的快速扩散，提升公众媒体素养与AI素养成为不可或缺的策略。一方面，教育机构和传播平台应通过课程、讲座、专题报道等形式，使公众理解生成式AI的运作机制、信息生成逻辑及潜在局限，培养辨识虚假或误导信息的能力，增强理性判断与批判性思维。另一方面，媒体自身也应承担社会教育责任，通过透明披露AI参与内容生产的环节、开展“去幻觉”报道与案例分析，引导公众形成审慎、理性且可追溯的信任态度。在此过程中，公众不仅是信息的接收者，也应成为监督者，参与对AI生成信息的评价与反馈。只有当技术透明、机构自律与公众认知能力三者协同作用时，传媒公信力才能在“人机共创”的信息环境中获得新的社会支撑，实现真实性、效率与信任的平衡[10]。

通过技术防护、制度规范和公众教育的三位一体协作，可以有效遏制生成式AI幻觉对新闻真实性的侵蚀，同时为媒体在数字化、智能化时代构建稳固的公信力边界提供可行路径。

四、结语

生成式人工智能的幻觉特性正在深刻重塑传媒内容的公信力边界，其对新闻真实性、机构依赖与公众信任的多重影响凸显了传统公信力体系面临的结构性挑战。本文从技术机制、机构行为与公众认知三个维度分析了幻觉现象的生成原因及传播效应，指出其可能导致内容真实性弱化、编辑审核机制弱化以及公众信任波动。针对这一问题，研究提出了三方面的综合策略：一是通过知识图谱、检索增强生成及内容溯源技术降低幻觉风险；二是强化媒体机构自律和行业监管，建立责任可追溯的管理体系；三是提升公众媒体素养与AI素养，增强受众辨识与监督能力。综合来看，技术防护、制度规范与公众教育三者的协同作用，是在“人机共创”环境下重构公信力边界、保障新闻真实性与社会信任的关键路径。未来，媒体在数字化和智能化转型过程中，应持续关注AI幻觉带来的挑战，以实现效率、真实与信任的平衡，构建可持续的新闻生态。

参考文献

[1]齐硕 . 跨媒体智能对传媒产业的影响与变革 [J]. 科技视界 ,2024,14(22):1-4.

[2]宋光茂 . 人工智能条件下的“媒体再融合”[J]. 新闻战线 ,2024,(11):12-14.

[3]宁菁 . 生成式人工智能对传媒新生态的影响——以 ChatGPT 的运用为例 [J]. 上海广播电视研究 ,2024,(04):109-114.

[4]彭兰 . 生成式人工智能技术驱动传媒业再变革 [J]. 南方传媒研究 ,2024,(03):5-13.

[5]杨琴 . 以 ChatGPT 看生成式人工智能对传媒业的影响 [J]. 卫星电视与宽带多媒体 ,2024,21(02):65-67.

[6]余欣 , 余海霞 . 大模型时代传媒行业向新提智的实践与探索 [J]. 南方传媒研究 ,2024,(06):33-39.

[7]朱向阳 . 传媒新质生产力的培育路径研究 [J]. 中国地市报人 ,2024,(12):44-46.

[8]韩佳镔 . 新闻行业与人工智能融合发展探究 [J]. 中国报业 ,2024,(22):36-37.

[9]王强春 . 人工智能时代融媒体新闻发布路径与数字伦理风险探析 [J]. 北方传媒研究 ,2024,(05):67-71.

[10]李子青 . 基于人工智能的智慧传媒创新场景研究 [J]. 中国报业 ,2024,(19):92-93.

