AI 幻觉挑战下循证科学传播的路径建设研究

程　曦[1,2,3]　苏　月[1,2,3]　吴丽婷[1,2,3]　王浩然[4]

（苏州大学传媒学院，苏州 215000）[1]
（苏州大学科技传播研究中心，苏州 215000）[2]
（苏州大学科技传播研究国际联合实验室，苏州 215000）[3]
（加拿大西安大略大学电子与计算机工程学院，加拿大安大略省 N6A3K7）[4]

[摘　要] 循证科学传播作为以客观事实和系统性研究证据为基础的传播方式，在应对人工智能（AI）幻觉方面具有潜在优势。本研究以循证科学传播为切入点，分析 AI 与循证传播的双重关系。（1）AI 对循证科学传播的赋能：AI 在科学传播的数据汇聚、证据甄别和效果反馈等环节具有赋能潜力，可提升循证传播的效率与覆盖面；（2）AI 幻觉对科学传播的挑战：AI 幻觉对科学传播的挑战不仅体现在内容的真实性，更涉及知识生产过程的合法性，在内容根基、知识生产、知识稳定性与共识达成效率等维度均带来挑战，导致失实信息从个案失真扩展为结构性错误。基于此，本研究提出以证据性、可追溯性、适应性与公共性为原则的循证科学传播人机协同路径，明确每项原则下 AI 端干预机制与人机协同的关键节点，构建贯穿“内容生产—流程管控—节奏调控—效果反馈”的循证科学传播体系。

[关键词] AI 幻觉　循证科学传播　知识生产　人工智能
[中图分类号] G206　[文献标识码] A　[DOI] 10.19293/j.cnki.1673-8357.2026.02.011

随着人工智能（Artificial Intelligence，AI），特别是生成式人工智能（Generative AI，GenAI）的迅速发展，传播型人工智能应运而生，标志着科学传播范式的转变[1]。然而，AI 幻觉现象的存在，使科学传播在信息权威性与可靠性方面面临前所未有的挑战。关于 AI 幻觉的定义目前尚无统一定论。例如，胡泳等将 AI 幻觉界定为“经由多元主体价值博弈后，由大模型技术所衍生、创造出的负荷着多元价值取向的内容文本”[2]；季子崴等则认为 AI 幻觉指“模型生成与事实不符或缺乏逻辑连贯性的内容且往往以逼真的形式呈现，从而模糊了真实与虚假的界限”[3]。ChatGPT 将 AI 幻觉定义为“机器（例如聊天机器人）生成看似逼真，却与任何现实世界输入毫无关联的感知体验的现象”[4]。科学的不确定性与前沿性特征，为 AI 幻觉的出现提供了温床。当信息模糊或不完整时，AI 可能基于碎片化或未经验证的知识进行推断，生成不准确甚至荒谬的结论[2]；而在面对争议性议题

收稿日期：2025-09-15
基金项目：国家社科基金青年项目“基于政策与科学共演过程的循证决策机制研究”（21CSH074）。
作者简介：程曦，苏州大学传媒学院副教授、苏州大学科技传播研究中心研究员、苏州大学科技传播国际传播联合实验室研究员，研究方向：科技政策、科技传播，E-mail：fxchxi@suda.edu.cn。

时，AI 过度简化或遗漏关键信息的倾向，则可能加剧公众的认知偏差 [5]。人工智能生成内容（Artificial Intelligence Generated Content，AIGC）在交互方式上具有便捷性，其非人格化形象强化了用户对其客观性的信任，促使信息消费从主动选择转向被动接受，在一定程度上放大了幻觉信息对公众认知的影响。在此情境下，AI 幻觉对科学传播的冲击不仅是个案失真，更可能演变为系统性错误：可能误导公众对科学本质的理解，削弱科学传播的准确性，进一步影响公众科学素质的提升与相应的行为决策。

循证科学传播（Evidence-Based Science Communication，EBSC）理念为应对这一挑战提供了可借鉴思路。循证科学传播理念源于循证医学思想，强调信息、结论或决策的产生应以客观的事实和科学的检验为基础 [6-7]。同时，循证科学传播也是一种决策模式，通过系统整合经验判断、专业知识以及基于系统研究获得“最佳”证据 [8]。循证科学传播强调在科学传播中以系统研究所提供的最佳证据为依据，并结合传播者的专业经验，提升传播的真实性、有效性和公共价值 [9-11]。在循证科学传播框架中，证据不仅支撑事实判断，也承担确保信息可信性与维护科学权威的功能 [12]。除此之外，循证科学传播还可以拓宽传播覆盖面、加快传播速度，促进公众参与，进而促成基于证据的讨论，这也是科学传播所强调的。近年来，在风险社会语境下对科学不确定性问题的关注不断增多，公众需要依托更便捷可靠的科学证据作出知情决策 [13]，循证科学传播的重要性随之提高。鉴于当前 AI 发展带来的风险与不确定性尚未完全厘清，且 AI 幻觉增加了失实信息的产生与传播概率，重申并落实“以可验证的证据为依据”的规范具有重要意义。

AI 与循证科学传播之间的关系并非单向的对立，而是呈现出张力与互补并存的双重特征。AI 在大规模数据处理、证据挖掘与模式识别等方面所展现的优势为循证科学传播的实践提供了新的技术支撑。例如，AI 能够辅助完成系统性文献回顾与多维度证据整合，从而在效率与广度上超越传统人工筛选 [14]；在设定明确规则与审查流程的前提下，基于规则的去人格化处理有助于在一定程度上弱化人为偏差。但同时，AI 幻觉可能通过生成似真非真的虚假信息，对事实证据的权威性与传播的可靠性构成威胁，从而凸显了循证科学传播的重要性与迫切性。

AI 于科学传播而言，既是潜在赋能者，也是风险制造者。针对 AI 幻觉所引发的挑战，循证科学传播的建设不仅关乎信息真实性与公众信任的维护，更直接关系到循证传播体系能否在新的技术语境下实现稳健运行。基于此，本研究聚焦以下 3 个问题，探讨如何在 AI 幻觉背景下构建有效且可持续的循证科学传播体系：（1）AI 如何赋能循证科学传播；（2）AI 幻觉给科学传播领域带来了哪些挑战；（3）面对 AI 幻觉，循证科学传播建设路径如何建构。

1 AI 赋能循证科学传播

循证科学传播以可核查的证据链为基础，组织传播信息的收集、提炼与传递，其主要流程包括原始数据的收集、证据的筛选与提炼、效果的评估与监测等 [15]。本研究基于上述 3 个重要流程，通过与传统循证实践对比，探讨 AI 如何赋能循证科学传播。

1.1 原始数据的收集

原始状态下的证据表现为信息与数据 [6]，因此数据收集是循证过程的起点。循证所需信息不仅来自自然科学，还覆盖政治、经济与文化等多维领域，其形态包括专家知识、

研究论文、统计数据、政策评估与利益相关方意见等。在传统路径中，研究者通常先设定检索词与数据库，再以人工方式逐条判读筛选。此法便于把关，但在合理成本下难以覆盖跨学科、多来源且快速演化的证据池，也难以及时捕捉舆论与政策场域的新生证据。

引入 AI 后，数据收集从一次性、静态汇编转向持续化、动态汇聚。基于主动学习的工具（如文献初筛）与风险偏倚自动判读，可借助 AI 在大规模语料中优先筛取相关材料，这已在健康政策证据的收集中得到应用[16]。开放学术知识图谱可支持信息的追踪检索，使原始证据的一次性收集转为可持续的更新追踪[17]。任务导向型 AI 智能体（AI Agent）的部署可进一步扩展证据采集能力。AI 智能体根据预先设定的证据需求与纳入标准自动理解和响应[18]，并自动规划数据源清单，执行周期性或事件触发的抓取，完成去重、格式统一与元数据补全等工作。与传统静态汇编式数据收集相比，AI 赋能后的动态汇聚式数据收集不仅扩大了证据外延，也为后续的筛选与提炼环节提供了可持续更新、可追踪演化的原始材料库。

1.2 证据的筛选与提炼

循证科学传播的关键在于确立科学、透明且可操作的证据筛选标准。证据在进入分析前通常处于原始与非结构化状态，需经严格甄别才能保证其可信度。传统模式下，筛选标准通常由研究者或政策制定者制定，难以避免主观性和目标导向倾向[19]。在认知与资源受限的情况下，研究者可能基于既定结论反向筛选证据，这削弱了循证传播的科学性与公信力。AI 赋能并非取代专家，而是在标准构建的过程中提供数据驱动的校准支持，例如：（1）借助历史系统综述与评估语料，训练监督、无监督模型对候选数据源进行一致性、相关性、可靠性与时效性评估；（2）基于舆情监测和互动平台数据，动态识别议题与证据缺口，为标准制定提供实践参考；（3）借助人类反馈强化学习（Reinforcement Learning from Human Feedback，RLHF），在专家监督下实现模型持续迭代，使自动化筛选规则与专家判断保持一致。在证据提炼环节，在标准明确的前提下，可利用自然语言处理与跨模态语义分析，从复杂数据中识别关键变量、去除冗余信息，并结合因果推断与模式识别、揭示变量间的关系结构。由此可将零散信息转化为结构化、可检验的证据单元，提升后续综合与解释的可操作性。

1.3 效果的评估与监测

循证科学传播需要对证据的实际应用效果进行持续监测与评估。传统模式多依赖小规模试点以及线下专家和公众反馈，虽可获得一线信息，但这类评估常以阶段性方式开展，覆盖面较为有限，且受权威压力与表达门槛影响，反馈效度受限。与传统的以阶段性评估为主的做法相比，AI 赋能可对证据在实践前、实践时和实践后的效果进行全流程的监测与评估。实践前，可借助生成式智能体等模拟人类行为的智能体构建应用环境，复现不同叙事框架与发布节奏下的受众响应与舆论演化，作为风险前置识别与方案预调优的依据[20]。实践时，可依托平台化的在线随机对照试验（例如 A/B 测试）收集用户意见，还可借助自适应实验设计提升传播信息与用户的匹配程度。实践后，结合监督学习与社会网络分析开展社交平台与新闻评论的扩散监测，对误解与谣言的传播链路进行早期识别并实施靶向澄清。

2 AI 幻觉给科学传播领域带来的挑战

AI 赋能科学传播的同时，也带来了 AI 幻

觉问题。在一般传播研究中，AI 幻觉多被理解为大模型输出不准确、编造或自相矛盾的内容，其关注点多集中于信息质量与平台扩散机制的治理。然而，科学传播不仅应关注知识的准确性，还应重视知识生产过程的规范性与可验证性。当大语言模型在生成过程中产生幻觉时，这种幻觉往往会超越单纯的事实失真，改变科学知识的产生过程与逻辑，削弱科学知识的稳定性，阻碍学术共识的达成。

2.1 动摇科学传播内容的证据根基

首先，AI 幻觉动摇了以证据为核心的科学知识基础。科学知识的构建离不开坚实的科学证据，而 AI 幻觉通过拼接、虚构甚至篡改传播证据破坏科学传播的严谨性 [21]。例如，GenAI 可能虚构并不存在的学术文献，编造专家发言，或将部分真实数据与伪造信息混合，从而制造出看似真实、严谨的科学内容 [4]。沃尔特斯（William H. Walters）等的实证研究发现，在 ChatGPT 3.5 生成的 222 条引文中，55% 的文献为捏造，学术论文类型捏造比高达 73%；切换到 4.0 版本，414 条引文中，18% 的文献为捏造，学术论文捏造比为 18%[22]。更为严重的是，AI 普遍存在自信度过高但提供内容有误的典型问题，即使缺乏事实依据，也会以高肯定性语气表达 [23]。其次，AI 幻觉利用并加剧了科学知识固有的不确定性。科学的不确定性既源于知识边界内的固有局限（如对未来的不可预见性），也源于知识需要不断适应外部客观世界而产生的动态变化 [24]。大模型基于既有数据和模式进行训练，这意味着尚未被观测或理解的内容可能成为幻觉生成的来源 [2]。同时，大模型常常缺乏最新数据 [25]，这可能导致 AI 输出的正确信息已然过时，难以适应外部变化。

这种真伪难辨的证据可能导致科学传播的过程从“知识普及”异化为“真伪验证”。即便是具备较高科学素质的学术共同体，也可能需要额外检索与比对才能确认信息真伪；而对于普通公众而言，验证成本更高。如果用户选择不加甄别地接受信息，虚假的证据便可能获得临时可信度。这不仅扰乱了学术共同体对证据有效性的判断，也可能重塑公众对证据的理解，将可言说性误认为可验证性，从而动摇科学知识合法性的根基。

2.2 改变科学知识生产过程与结构

刘海龙指出，人机协作的知识生产方式并非传统意义上的“思”，其可能引发人类思维逻辑机器化、知识权威和知识标准混乱等潜在后果 [26]。传统科学知识的生产范式包括经验归纳、理论建模、计算模拟、数据密集型科学，科学知识的生产依赖于观察与实验、假设与推理、严格统计等方法所产生的实证结果，最后须经同行评议才能确立其重要性 [27]。这一过程的核心在于强调研究过程的可重复性与可验证性 [28]。然而，GenAI 的逻辑建立在对已有知识的相关性学习与通过上下文预测语义概率之上 [29]，这也是 AI 幻觉产生的原因之一，因此 AI 幻觉只可缓解不可消除。在这一逻辑下，方法的可检验性逐渐让位于结果的可接受性或语言的可置信度，知识生产的过程与结构发生了改变。

用户在接受 AI 生成的科学内容时，通常不会追溯其知识生产过程，而该过程往往也是不可检验的。当用户选择相信 AI 生成的结果时，就意味着知识生产的结果取代了方法的地位。如果将此类无严格生产过程的知识传播出去，将为后续的科学研究带来错误示范。例如，当 AI 生成一篇关于气候变化的科普文章时，即使其结论看似正确，但若文章引用的研究数据来自虚构的论文，实验方法无法被还原验证，这种知识生产过程的缺陷将从根本上动摇科学传播的可信度基

础。AI 的知识生产不仅削弱了研究方法论在科学知识生产中的核心地位，也可能改变知识生产的基本结构，使失实的幻觉内容依靠语言的说服力和与原有知识的接近性而获得信任。

2.3 破坏科学知识的稳定性

科学知识体系建立在确定性与不确定性的辩证统一之上。科学知识一方面以规范性、逻辑性和严谨性为基石，要求能够提供稳定而可靠的结论[1]；另一方面又必须承认知识的动态性与可错性，以回应客观世界与科学实践不断演化的现实。科学知识的双重性要求科学传播需在“可检验的稳定性”与“可修正的开放性”之间取得平衡。

然而，不同 AI 平台针对相同科学话题常出现答案不一致或信息过时的情况，这也成为科学传播中 AI 幻觉的来源之一。例如，有家长发现同一小学奥数题 4 个不同 AI 平台给出的答案不同；如果改变身份提问同一问题，同一平台上也可能给出不同答案，因此社交媒体上常出现“当 AI 答案打架，我们该相信谁”“同一问题不同 AI 给出不同答案，哪个才是正确的”等疑问。不一致的答案破坏了科学知识所需的稳定性和可靠性，即使 AI 提供了正确答案，多版本回答也会使用户产生认知偏差，从不同甚至矛盾的视角理解同一科学问题，进而造成用户将 AI 回答的版本差异误认为是科学不确定性的体现。值得注意的是，AI 生成的答案差异与科学的不确定性存在本质区别。科学的不确定性是从确定性过渡而来[30]，需要经历一定时期的稳定积累。而因 AI 版本不一致导致的内容差异，既缺乏科学规范性和逻辑严谨性的支撑，也不遵循科学认知演进的时间逻辑，其快速变化的生成结果本质上属于技术问题，而非科学探索的必然特征。

2.4 增加科学共识达成难度

科学传播的目标不仅是知识的传递，更在于呈现多元观点、促进理性讨论并最终形成社会共识。然而，AI 为迎合用户的个性化输出机制使事实的准确性让位于用户的满意度，AI 可能选择性地呈现、甚至编造信息以迎合用户的偏好。这种取悦用户的机制不仅增强了错误信息的说服力，还会进一步强化用户的确认偏误。有研究表明，AI 生成的科学报道内容相比传统科学新闻更容易放大原有的情感倾向[31]。例如，在具有争议的气候变化议题中，AI 生成的相关描述相较于专家意见更为夸大，更易造成公众态度和意见的分歧[32]。这种分歧使用户只能接触到经过筛选的片面观点，从而加剧群体意见的极化。

所谓极化，是指群体成员意见逐渐向更加极端化的方向聚集[33]。在科学传播语境中，极化的加剧不仅削弱了科学的客观性，还可能使理性讨论异化为立场先行的情感对立，加剧社会撕裂，阻碍科学共识的形成。科学传播的社会价值最终体现在能否有效指导社会实践、增进人类福祉，而这离不开科学共识的达成以及社会成员的广泛认同与协作。因此，共识的难以达成不仅削弱了科学传播的社会功能，也可能导致政策执行受阻和社会协作效率下降。

3 面向 AI 幻觉的循证科学传播人机协同路径

AI 幻觉对科学传播的影响正在由个案失真演变为结构性错误。有研究指出，幻觉是大语言模型内在的局限性[33]，应对路径必须建立在幻觉不可消除的前提之上。基于此，本研究提出以证据性、可追溯性、适应性与公共性为原则的循证科学传播人机协同路径，并在每项原则下明确由 AI 工具承担的自动化执行任务与人工简单操作的关键节点，以 AI

工具承担可标准化的重复劳动，将人工投入压缩至一次点击或一次审定，避免增加科学传播者的日常运营负担。

3.1 以证据性为核心的内容生产路径

证据性原则旨在确保科学传播内容的真实性与可验证性。为重塑以证据为根基的科学传播生态，建议在内容生产的 AI 端构建如下干预机制。（1）基于检索增强生成（Retrieval–Augmented Generation，RAG）范式的引文链接自动嵌入。强制模型输出必须锚定已检索的权威文献片段，系统在正文中自动生成并插入原文链接或 DOI，从而将生成式幻觉严格约束于“检索—生成”框架之内。（2）基于古斯塔夫森（Gustafson Abel）与赖斯（Rice Ronald E）不确定性分类框架 [34–35] 的模型卡式自动声明。由系统为每条输出结果自动标注不确定性类型及适用边界，避免过度概括或以确定性修辞掩盖对科学事实的认知限度。（3）争议性议题对立证据的自动聚类。利用 AI 工具在预设的权威数据库内定向检索竞争性假说及其证据基础，并在前端界面并列呈现，为传播者与受众提供多元事实参照。（4）突发情境下最小可用证据集的智能推荐。在科学事实尚未完全确立时，系统可提供附带“初步显示 / 尚需验证”等过渡性限定语的证据候选集。

在人机协同层面，科学传播者需在发文系统中把关最终输出，对 AI 提取的证据来源及不确定性声明进行专业复核与事实背书。

3.2 以可追溯性为基准的流程管控路径

可追溯性原则主要回应 GenAI 在科学传播中“重结果、轻过程”的黑盒倾向，强调内容生产全周期的可还原与可核查。（1）以全生命周期版本号与日志的自动化记录为首要机制。内容的生成、修改与发布均由系统自动施加版本时间戳，并沿“数据—证据—内容”的反映链条，保存信息采纳、加工和删改的操作轨迹与算法依据。（2）人机协同决策数据的自动化捕获。系统需精确记录 AI 工具或人工在何时依据何种标准进行信息筛选、裁决证据冲突以及调整叙事策略，进而生成结构化的可查询决策链。（3）勘误时间与公示页面的自动生成。当传播内容发生实质性修订时，系统自动触发勘误机制，并在特定公示页输出修订前后的对照矩阵与事实依据，以降低人工维护成本。

在人机协同层面，科学传播者需在关键节点行使发布许可权，学术共同体需要对涉敏议题的价值导向与科学底线进行最终裁量。

3.3 以适应性为导向的节奏调控路径

适应性原则主要聚焦于科学传播在动态语境下稳定性与灵活性的平衡，旨在化解 AIGC 可能对科学知识稳定性造成的冲击。科学传播的适应性高度依赖对传播节奏的精准把控，即在科学共同体知识迭代的过程中，统筹信息发布的时效性与内容成熟度。此类动态调节高度契合 AI 的技术优势，具体应用包括 3 个方面。（1）同主题新证据的持续监测与触发式预警。当权威学术期刊或官方政策平台发布与既往发布内容相关的新研究成果、勘误声明或撤稿声明时，系统主动向传播者推送预警提示与复核建议。（2）勘误文稿的自动起草。由 AI 工具基于新增证据链自动生成勘误初稿（包含修订节点、逻辑说明和新证据溯源），经传播者与领域专家审阅后发布。（3）群体极化与误传路径的动态监测。利用 AI 自动识别潜在的极化议题或衍生误导信息，向运营方与专业共同体进行同步预警，为危机干预与前置响应提供决策支撑。

在人机协同层面，学术共同体需掌握勘误程序的最终启动权，主导科学共识的边界修订与责任界定，科学传播者需在文末规范标注“待复核”或“已更新”等状态标签。

3.4 以公共性为目标的效果反馈路径

公共性原则致力于在 AIGC 语境下重构公众参与和科学共识的凝聚机制。鉴于 AI 幻觉可能加剧事实替代与舆论极化，进而侵蚀基于事实的公共理性审议空间，科学传播必须由单向的知识灌输向共证协作的制度化参与模式转型。（1）同源证据的多版本文本生成。AI 工具可依据目标受众的认知负荷，将同一证据自动转化为专家版、公众版或青少年版，并对专业术语密度与逻辑推演步距进行动态控制。同时自动生成图文、短视频、问答等适配不同媒介载体的格式，并强制将不确定性类型及证据出处嵌入固定位置。（2）公众纠错线索的智能聚合与分级校验。对公众提交的原始数据、现场观察或质疑建议，AI 按信度与可验证性标准进行初步筛选与分层，再推送至对应处理环节。（3）高频疑问的自动聚类与标准化回应，以规避信息过载导致的线索遗漏。

在人类主导的社会反馈闭环中，公众可通过“提交线索—查询进度—接收回执”的交互路径参与共证协作。科学传播者可将日常高频问题交由 AI 调取标准化模板予以响应，而复杂或高度敏感的质询则流转至人工复核池。对于受限于当前科学认知尚无法定论的问题，需告知暂不回应的理由与后续观测计划，避免因信息真空引发公众误解，即沉默效应风险。

为避免证据性、可追溯性、适应性与公共性 4 项原则各行其是，可设立相应的核心指标作为循证科学传播建构的依据以实现 4 项原则协同，如证据可靠度（证据等级匹配、对立证据呈现、证据跟踪兑现情况）、溯源完整度（来源与提炼过程记录、勘误响应时长）、适应响应度（调整边界与改进幅度及其效果）以及公共参与度（证据贡献纳入率、质询响应率、信任指数变化）等。在执行层面，可依据“试点运行—扩大范围运行—常态化运行”的推进路径，使循证科学传播在不确定性的环境中保持快速运行、严格审计、理性对话的稳态，既不过度依赖模型的流利度，也不因谨慎而停滞；既维护证据链条的强度，也为公共意义的生成预留可操作空间。

4 结语

AI 对循证科学传播具有双重效应。AI 可赋能循证中数据收集、证据甄别和效果评测等环节，使许多过去受制于成本与效率的循证实践具备了现实可行性，从而扩展了循证科学传播的覆盖范围，增强了循证科学传播的处理能力。但与此同时，AI 幻觉打破了科学传播中以证据为基础的内容根基，也改变了科学知识的产生过程与逻辑，削弱了科学知识的稳定性与共识形成的可行性，造成失实信息从个案失真扩展为结构性错误。基于此，本研究提出面向 AI 幻觉的以证据性、可追溯性、适应性与公共性为原则的循证科学传播人机协同路径，并明确每项原则下 AI 端干预机制与人机协同的关键节点，形成贯穿“内容生产—流程管控—节奏调控—效果反馈”的循证科学传播体系。

虽然 AI 幻觉对科学传播的影响是负面的，但其对科学的影响不全是负面的。AI 幻觉中所蕴含的脱离现实的幻想也可能成为科学创新的意外催化剂。AI 幻觉的确可能造成事实性错误，但也可能突破现实的束缚，激发超越常规思维的灵感与假设。2024 年诺贝尔化学奖得主大卫·贝克（David Baker）坦言，模型的幻觉给了他无限的创造力[36]。科学发现的早期阶段往往依赖直觉、想象与大胆的猜想。AI 幻觉生成的假设性内容本质上是一种“被算法激发的想象”，这种想象有时可能是错误的，但也可能为科研人员提供新的思考路径或验证方向。因此，从这一视角来看，AI 幻觉虽然是科学传播与研究需要防范的风

险，但也可能成为推动科学范式跃迁的“错误之源”，在虚幻与真实之间，为科学的下一个突破提供灵感。

本研究的创新性和主要贡献集中于两点。（1）从内容与过程两个层面探讨 AI 幻觉对科学传播的影响。在一般传播中，AI 幻觉引发的问题多集中于信息生产结果的错误性，即内容层面的失真[37]，而解决方案通常是事实核查与内容修正，旨在纠正已发布的错误信息。然而，在科学传播中，关注点并不止于内容的真实性，更涉及知识生产过程的合法性。因此，降低科学传播中 AI 幻觉的负面影响不仅需要在内容层面把关，还需要在生产过程层面做好可溯源性的流程管理和适应性的节奏把控。（2）从制度化的证据治理视角建构应对路径。既有研究多聚焦于科学传播中虚假信息的谣言澄清、事实核查、平台审核与媒介素养等维度[38–39]，主要强调内容端纠错与受众端教育，以证据链贯穿内容生产、传播及反馈全过程的研究鲜有报道。鉴于证据是科学传播的基础，且相当一部分失实信息源于证据的选择或使用不当，本研究追溯问题根源，以证据性、可追溯性、适应性与公共性为原则，构建循证科学传播人机协同路径，为循证科学传播平台与机构提供了可操作的方向。

本研究仍存在一定局限性。一是本研究侧重提出治理原则与机制组合，尚未在多议题、多平台和多语种环境中给出系统性对比验证。二是关于“科学知识稳定性”“传播态度极化”等关键结果指标的度量边界与因果识别还需要更长期的追踪数据加以检验，这也是未来进一步研究的方向。

参考文献

[1] Luna S D，Broer I，Bilandzic H，et al. Quality in Science Communication with Communicative Artificial Intelligence：A Principle–Based Framework[J]. Public Understanding of Science（Bristol，England），2025，34(8)：9636625251328854.

[2] 胡泳，王昱昊. 技术过程论视角下 AI 幻觉生成的价值负荷与伦理问题探析 [J]. 南京社会科学，2025(3)：84–94.

[3] Ji Z，Lee N，Frieske R，et al. Survey of Hallucination in Natural Language Generation[J]. ACM Computing Surveys，2023，55(12)：1–38.

[4] Alkaissi H，McFarlane S I. Artificial Hallucinations in ChatGPT：Implications in Scientific Writing[J/OL]. Cureus，2023，15(2)：e35179. https://pubmed.ncbi.nlm.nih.gov/36811129/.

[5] Jones C R，Bergen B K. Lies，Damned Lies，and Distributional Language Statistics：Persuasion and Deception with Large Language Models[J]. arXiv preprint arXiv：2412.17128，2024.

[6] 李乐，周志忍. 英国社会政策循证决策的理论与实践：对中国的启示 [J]. 中国公共政策评论，2016，11(2)：117–131.

[7] 徐宏宇. 证据在科技决策中的应用研究 [J]. 现代情报，2020，40(9)：90–95.

[8] 王溯，董瑜. 科技创新政策循证决策机制研究——以日本第六期《科学技术与创新基本计划》为例 [J]. 中国科技论坛，2024(11)：156–168.

[9] Kahan D M. Making Climate–Science Communication Evidence–Based：All the Way Down[M]//Culture，Politics and Climate Change. Routledge，2014：203–220.

[10] Jensen E A，Borkiewicz K，Naiman J P，et al. Evidence–Based Methods of Communicating Science to the Public through Data Visualization[J]. Sustainability，2023，15(8)：6845.

[11] Jensen E A，Gerber A. Evidence–Based Science Communication[J]. Frontiers in Communication，2020(4)：78.

[12] 魏夏楠，张春阳. “循证决策”30 年：发展脉络、研究现状和前沿挈领——基于国内外代表性文献的研究综述 [J]. 现代管理科学，2021(4)：26–36.

[13] Biermann K，Nowak B，Braun L M，et al. Does Scientific Evidence Sell? Combining Manual and Automated Content Analysis to Investigate Scientists’ and Laypeople’s Evidence Practices on Social Media[J]. Science Communication，2024，46(5)：619–652.

[14] 施锦诚，王迎春. 大模型创新变革：新模式、新挑战与新趋势 [J]. 中国科技论坛，2024(7)：31–40.
[15] 杨克虎，郭丽萍，蔡立群. 证据驱动的循证决策国际进展、趋势与展望 [J]. 图书与情报，2024(2)：102–108.
[16] Ferdinands G，Schram R，de Bruin J，et al. Performance of Active Learning Models for Screening Prioritization in Systematic Reviews：A Simulation Study into the Average Time to Discover Relevant Records[J]. Systematic Reviews，2023，12(1)：100.
[17] Manghi P. Challenges in Building Scholarly Knowledge Graphs for Research Assessment in Open Science[J]. Quantitative Science Studies，2024，5(4)：991–1021.
[18] Durante Z，Huang Q，Wake N，et al. Agent AI：Surveying the Horizons of Multimodal Interaction[J]. arXiv preprint arXiv：2401.03568，2024.
[19] 周志忍，李乐. 循证决策：国际实践、理论渊源与学术定位 [J]. 中国行政管理，2013(12)：23–27，43.
[20] Park J S，O’Brien J，Cai C J，et al. Generative Agents：Interactive Simulacra of Human Behavior[C]//Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology. New York：ACM，2023：1–22.
[21] Scheufele D A. Communicating Science in Social Settings[J]. Proceedings of the National Academy of Sciences，2013，110（Supplement 3）：14040–14047.
[22] Walters W H，Wilder E I. Fabrication and Errors in the Bibliographic Citations Generated by ChatGPT[J]. Scientific Reports，2023，13(1)：14045.
[23] Hong S，Lee S. Overconfident AI：How Artificial Intelligence Navigates Risk and Uncertainty[J]. Journal of Psychology and AI，2025，1(1)：2486974.
[24] 王浩，维克托瑞·道尔可. 科学研究的目的：确定性还是客观性 [J]. 科技进步与对策，2025，42(20)：142–149.
[25] Kraus M，Bingler J A，Leippold M，et al. Enhancing Large Language Models with Climate Resources[J]. arXiv preprint arXiv：2304.00116，2023.
[26] 刘海龙. 生成式人工智能与知识生产 [J]. 编辑之友，2024(3)：5–13.
[27] Xu Z L，Cheng Y，Yang Y Q，et al. AI for Science 2025[R]. Shanghai：Fudan University，Shanghai Academy of AI for Science，Springer Nature，2025.
[28] 叶喜艳，刘蔚，侯春梅，等. 科学研究方法可重复性提升路径及对策建议 [J]. 中国科技期刊研究，2024，35(6)：746–754.
[29] 周慎，陶美丽，刘湘. 人工智能生成科普内容的底层逻辑与参与主体功能分析 [J]. 科普研究，2024，19(4)：14–22.
[30] 吴国林. 论知识的不确定性 [J]. 学习与探索，2002(1)：14–18.
[31] 王聪，陆袁芳洲，何彩妮. 生成式人工智能科普前沿科学成果的特点研究——基于中国主要大模型生成内容与《中国科学报》化学领域相关报道的对比 [J]. 科普研究，2025，20(2)：15–23，42，105–106.
[32] Tamang T，Zheng R. When AI Sees Hotter：Overestimation Bias in Large Language Model Climate Assessments[J]. Public Underst Sci，2026，35(1)：82–99.
[33] Xu Z，Jain S，Kankanhalli M. Hallucination Is Inevitable：An Innate Limitation of Large Language Models[J]. arXiv preprint arXiv：2401.11817，2024.
[34] Gustafson A，Rice R E. A Review of the Effects of Uncertainty in Public Science Communication[J]. Public Understanding of Science，2020，29(6)：614–633.
[35] Gustafson A，Rice R E. The Effects of Uncertainty Frames in Three Science Communication Topics[J]. Science Communication，2019，41(6)：679–706.
[36] 澎湃新闻. 2024 诺贝尔化学奖得主：“模型幻觉”给我无限创造力 [EB/OL].（2025–01–17）[2025–10–12]. https://www.thepaper.cn/newsDetail_forward_29945477.
[37] 周健宇，黄淑愿. AIGC 背景下新闻生产的挑战、机遇与应对——以 ChatGPT 为切入点 [J]. 决策咨询，2024(2)：71–76.
[38] 胡兵，陈可琳. 易引发网络舆情的科学谣言叙事构型探究 [J]. 科学学研究，2025，43(8)：1739–1747.
[39] 彭虹，薛蕾. 科学传播视野中的谣言治理 [J]. 中国出版，2014(22)：24–27.

（编辑 颜 燕 马 赫）

From "Knowledge Transmission" to "Meaning Sharing": On the Paradigm Shift of Science Communication from the Perspective of Philosophy of Scientific Culture

Meng Jianwei[1, 2]

(Tianfu College of Southwestern University of Finance and Economics, Chengdu 610052)[1]
(School of Humanities, University of Chinese Academy of Sciences, Beijing 100049)[2]

Abstract: Contemporary science communication practice is deeply trapped in the dilemma of "knowledge transmission": the public's understanding of scientific conclusions is increasing day by day, but their spiritual identification with science has not improved correspondingly. The deep-seated crux lies in the long-term domination of an "epistemological" view of science, which narrowly understands science as a collection of objective knowledge, thereby severing the intrinsic connections between science and culture, values, and human life. Escaping from this dilemma requires a paradigm shift at the philosophical level. The philosophy of scientific culture provides the theoretical foundation for this shift. It views science, in its essence, as a human-created activity imbued with spiritual pursuits and cultural ideals; thus, the spirit of science and the spirit of the humanities are interconnected at the highest realm. Accordingly, the essence of science communication should shift from "knowledge transmission" to "meaning sharing." This new paradigm contains three layers of implication: first, the sharing of knowledge shifts from transmitting conclusions to presenting the process of exploration; second, the sharing of spirit shifts from informing achievements to empathizing with the spiritual core of scientists pursuing freedom and cultural realm; third, the sharing of value shifts from demonstrating instrumental utility to achieving deep resonance between science and the public's world of meaning. Only in this way can science communication truly bring science, as a "living culture," into the public's spiritual world.

Keywords: science communication; philosophy of scientific culture; knowledge transmission; meaning sharing

CLC Numbers: N4 **Document Code:** A **DOI:** 10.19293/j.cnki.1673-8357.2026.02.010

A Study on Pathway Development for Evidence-Based Science Communication under the AI-Hallucination Challenge

Cheng Xi[1, 2, 3] Su Yue[1, 2, 3] Wu Liting[1, 2, 3] Wang Haoran[4]

(School of Communication, Soochow University, Suzhou 215000)[1]
(Center for Science Communication Research, Soochow University, Suzhou 215000)[2]
(International Joint Laboratory for Science Communication Research, Soochow University, Suzhou 215000)[3]
(Department of Computer Science, Western University, Ontario, Canada N6A 3K7)[4]

Abstract: As an approach grounded in objective facts and systematic research evidence, evidence-based science communication holds potential advantages in mitigating Artificial Intelligence (AI) hallucinations. From the perspective of evidence-based science communication, this study analyzes

the dual relationship between AI and evidence-based communication.（1）The empowering role of AI in evidence-based science communication：AI demonstrates enabling potential in stages such as data aggregation，evidence screening，and impact feedback，thereby enhancing the efficiency and reach of evidence-based communication.（2）The challenges posed by AI hallucinations to science communication：These challenges are reflected not only in content authenticity but also in the procedural legitimacy of knowledge production. They affect dimensions including content foundation，knowledge production，knowledge stability，and the efficiency of consensus-building，causing misinformation to escalate from isolated inaccuracies to structural errors. Based on these findings，this study proposes a human-AI collaboration pathway for evidence-based science communication，guided by the principles of evidentiality，traceability，adaptability，and publicness. By clarifying the AI-side intervention mechanisms and the key nodes of human-AI collaboration under each principle，this study constructs a comprehensive evidence-based science communication system spanning "content production—process control—pace regulation—impact feedback."

Keywords：AI hallucination；evidence-based science communication；knowledge production；Artificial Intelligence

CLC Numbers：G206 Document Code：A DOI：10.19293/j.cnki.1673-8357.2026.02.011

Technological Performance and Public Imagination：Multimodal Science Communication of Embodied AI on Short-Video Platforms：A Case Study of Unitree Robotics

Wu Wenxi Tian Ziyi Hu Ting

［School of Media Science（School of Journalism），Northeast Normal University，Jilin 130024］

Abstract：Short-form video has emerged as a pivotal medium for the dissemination of emerging technologies，positioning technology enterprises as one of the main agents of science communication. Focusing on Unitree Robotics' official Douyin account，this study draws on a sample of 81 videos and 4 114 comments to examine the interaction between corporate technological performance and public sociotechnical imaginaries. The study integrates multimodal discourse analysis with text mining to capture both the semiotic construction of technology and its discursive reinterpretation by the public. The findings reveal that embodied AI is rendered as a visualized，contextualized，and immersive technological performance through the orchestration of textual，visual，and auditory modes. Correspondingly，public discourse generates multiple，intersecting sociotechnical imaginaries structured around functional utility，techno-nationalism，and human-machine relations. These dynamics suggest that technological meaning is not unilaterally transmitted but continuously negotiated and stabilized through platform-mediated interactions between technology enterprises and the public. This study reveals that technological performance goes far beyond mere technology demonstration. It opens up a crucial space for the public to engage in the social construction of technology and to shape sociotechnical imaginaries. Consequently，these practices reconfigure science communication from one-way dissemination toward the participatory co-production of meaning.

Keywords：embodied AI；technological performance；sociotechnical imaginaries；short-video science communication；emerging technology

CLC Numbers：G206；N4 Document Code：A DOI：10.19293/j.cnki.1673-8357.2026.02.012