

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

王 聪 陆袁芳洲 何彩妮

(中国科学院大学, 北京 100049)

[摘 要] 前沿科学成果科普对于公众理解和参与科学都具有重要意义。随着生成式人工智能的广泛应用, 是否能够将其用于前沿科学成果科普过程已经成为一个重要问题。和生成一般科普内容相比, 生成式人工智能在前沿科学成果相关主题内容生成上具有独特之处: 一方面缺少足够多可以用作训练样本的网络文本, 另一方面其内容具有相对较高的不确定性, 因此需要更谨慎的内容表达方式。本研究通过对比传统科普内容生成者和生成式人工智能大模型产出的前沿科学成果科普内容, 发现两者在科学内容的表述严谨性和来源可靠性方面并没有显著差别, 但是生成式人工智能大模型产出的科普内容相比之下更倾向于呈现出正面情感。基于此, 本研究提出, 生成式人工智能在错误信息产出方面的风险可能并不大, 但是却可能进一步加强已有的刻板印象, 因此生成式人工智能并不能替代已有科普系统中的传统科普内容生成者。

[关键词] 生成式人工智能 前沿科学成果 科普

[中图分类号] N4; TP18 **[文献标识码]** A **[DOI]** 10.19293/j.cnki.1673-8357.2025.02.002

虽然科学的迅速发展已经并可能持续深刻地改变人们的生活, “但是公众对前沿科学的了解却非常有限, 这是我们科学传播系统的重要不足”^[1]。科学家们的科研成果科普是对前沿科学的一种有力展现, 可以使公众更好地了解“档案”之外的科学, 在提高公众科学素质的同时, 有利于公众“思考新研究成果可能带来的社会、伦理、政策方面的影响”^[1], 从而使之更好地参与到与科学相关的社会议题讨论和决策过程中。然而, 目前

能够科普化的科研成果仍然比较有限。根据朱莉·苏莱斯基 (Julie Suleski) 等人的研究, 目前只有 0.013%~0.34% 的研究论文能够得到大众媒介的关注, 其中有关健康和医学的研究论文被报道的比例最高, 其他领域的论文被报道的比例在 0.001%~0.005% 之间^[2]。

生成式人工智能 (Generative Artificial Intelligence) 通过赋能公众, 降低了公众获得前沿科学成果的成本, 提高了公众的主动性, 公众可以通过生成式人工智能获得科普

收稿日期: 2025-02-16

作者简介: 王聪, 中国科学院大学人文学院副教授, 主要研究方向: 科学传播、科学技术与社会, E-mail: wangcong@mails.ucas.ac.cn。

内容,甚至参与科普内容的生产,如利用 ChatGPT 提供的“Science Communicator”“Medical Science Popularization”“PopSci Forge”“SciWrite Assistant”“科普作家”“PopSci Writer Assistant”等科普聊天机器人^[3]。然而,生成式人工智能也会给科普带来一系列的风险,如生成错误信息^[4],冲击已有科普主体组成的系统结构,甚至从根本上动摇传统的内容生产者(信息源和科普内容制作者)的地位。因此,生成式人工智能生成的科普内容有哪些特点,能否取代包括科学家和科普记者在内的传统科普内容生产者就成为受关注的问题。

前沿科学成果的科普相较其他科普具有独特性。其一,前沿科学成果完成时间较近,因此公众语境中的相关内容较少,不能为生成式人工智能提供充足的训练材料。其二,前沿科学成果虽然初步取得了科学合法性,但是还未得到科学共同体的广泛认可,因此内容不确定性较强,在科普过程中需要使用更为谨慎的表达方式。其三,从技术的角度看,生成式人工智能的内容生成包含多个过程:在预训练过程中,对生成式人工智能大模型的训练主要基于已有的网络内容进行,使其具备通用领域的内容创作能力;之后,用户可以利用训练完成的大模型开展科普内容创作。其中,预训练过程存在两种可能。一种可能是大模型缺乏预训练的数据,由于预训练过程需要基于大量已有数据,所以当公共网络资源中不存在相关主题内容时,没有受过相关主题训练的大模型可能在内容生成上存在问题;另一种可能是完成预训练的大模型将具备一定的泛化性能,比如大模型完成 10 以内加减法的训练之后,能够通过泛化,完成 10 以上加减法。

基于以上探讨,本研究以《中国科学报》2024 年 12 月所有化学领域前沿科学成果的报

道内容与生成式人工智能大模型生成的相同主题的报道内容为研究对象,从内容和呈现方式两个角度进行对比分析,以考察生成式人工智能科普前沿科学成果的特点。本研究选择使用的 5 个广受认可的大模型为:讯飞星火认知(科大讯飞)、文心一言(百度)、通义千问(阿里巴巴)、豆包(字节跳动)、元宝(腾讯)。

1 生成式人工智能与科普

目前,生成式人工智能被用于文字、图片、视频等各种形式的内容生成过程中,并在教育、科研、医疗、商业、数字媒介等领域中被广泛应用^[5]。

现有的关于生成式人工智能与科普的研究主要关注两个方面。其一是将生成式人工智能看作一项新技术,探讨加强公众对其理解的必要性、其可能的发展路径、及其带来风险的应对方式^[6-8],比如通过提高全民科学素质的方式使人能够更好地利用人工智能,保护人类创造力,减少对人工智能的过度依赖,提高人工智能使用者的伦理意识并增强其伦理教育^[9-11]。

其二是关注人工智能对科普工作的影响。人工智能有利于激发科普创作者的灵感,降低科普内容的查找难度和所需要时间成本,甚至直接生成科普内容^[12-13],实现个性化内容定制^[14],从而吸引更多更广泛的主体参与到科普创作过程中^[3]。

除此之外,生成式人工智能也可能从科普过程和科普系统两个方面产生一系列的风险。在科普内容方面,生成式人工智能生成的科普内容可能出现虚假信息^[4,15],侵犯隐私、安全性、知识产权^[7],在具体场景中,比如在馆校合作过程中还会造成数字鸿沟、伦理偏见等^[16]。在科普系统方面,生成式人工智能为传统的科普工作组织系统带来了冲击。

一方面，生成式人工智能进一步打破了传统传播者和受众之间的界限，由于降低了技术门槛，一般公众可以在人工智能的辅助下以较低的成本生成科普内容，从而可能削减部分传统传播者，如科学家和专职科普人员的权威性和主体性，动摇已有科普服务系统的根基^[12]；另一方面，生成式人工智能为科普内容监管系统带来了新的压力，由于缺少传统的把关人制度且进一步降低了科普内容制作门槛，科普内容严谨性和可靠性可能进一步降低，其中的价值框架也可能出现问题，从而给监管系统带来了压力^[7]。

已有研究深入探讨了生成式人工智能与科普之间的关系，但是仍存在两方面的问题。一方面，已有研究整体上仍以综述或者质性讨论的方式为主，主要依赖于自身或其他学者的经验判断，以给出可能存在的风险，却较少对人工智能参与科普的实际情况进行量化分析。另一方面，已有研究较少区分生成式人工智能对不同类型科普内容生成的影响，但生成式人工智能在不同类型的内容生成上的性能可能存在较为明显的差异。比如对于已经成为常识的科学知识，或者已经有较多科普文章的科普主题，生成式人工智能能够搜索并学习较为全面的相关信息，从而生成相对准确且通俗的科普内容。但是对于网络上内容较少的前沿科学成果，由于可学习的内容较少，因此有可能影响内容产出性能。而针对前沿科学成果的科普，研究人员的关注比较有限。阿曼达·阿尔瓦雷斯（Amanda Alvarez）等人提到，生成式人工智能可以帮助科学传播者理解科学论文，把其中的科学语言翻译成一般公众了解的形式，还可以帮助科研人员根据不同类型的受众的特点将研究发现生成为不同的科普内容^[4]。

本研究将以实证研究的方式，从内容和呈现方式两个方面，考察生成式人工智能在

前沿科学成果科普方面的特点。在内容方面，将关注科学内容的表述严谨性和来源可靠性 2 个指标；在呈现方式方面，将关注故事性和情感倾向 2 个指标。

2 研究设计

本研究梳理了《中国科学报》2024 年 12 月发表的所有化学领域前沿科学成果的报道，涵盖基础研究、应用实践、技术发明等类型。这些主题具有高度前沿性和多样性，充分体现了化学领域在能源、材料、环境等关键议题中的核心作用，兼具科学价值、社会价值与公众关注度。

研究从中提取了前沿科学成果的主题，并以统一的描述“我是某报社的记者，请帮我写一篇以‘……’为主题的，面向一般社会公众的科普报道，并列岀参考文章”，输入 5 个在中国使用广泛的大模型，分别是讯飞星火认知、文心一言、通义千问、豆包、元宝，使之生成同主题内容。最终，以 24 个主题（见表 1）为内容的 24 篇《中国科学报》的报道和 120 篇 5 个大模型生成的报道作为研究对象开展分析。

本研究聚焦前沿科学成果的文字式科普形式，从“科学内容”与“呈现方式”两个维度切入，分别评估科学报道内容的表述严谨性、来源可靠性、故事性和情感倾向 4 个指标。指标设计兼顾科学传播的“硬标准”与“软实力”，为系统分析和对比专业科学记者撰写内容和大模型科普报道的特点提供了多维度工具（见表 2）。

在科学内容方面，严谨地表达科学成果的内容，减少错误和虚假信息是科普最基本的要求，也是学者们认为生成式人工智能可能出现的问题^[4, 15]。表述严谨性直接关系到信息传播的可靠性，是科学内容传播的核心要求之一，在本研究中主要依靠科学共同体

(化学领域的研究人员)的反馈开展量化研究。而来源的可靠性在一定程度上能够影响所呈现的科学内容,可通过赋分量化来源的权威性和普及性(见表3)。

表1 研究样本主题

序号	报道日期	研究成果类型	研究成果
1	2024-12-25	应用与实践类	基于供体-受体结构的分子纳米光催化剂实现高效光催化分解水制氢,可扩展至其他光化学反应。
2	2024-12-25	其他	单宁酸异相成核策略制备图灵结构 COF 纳滤膜,实现有机溶剂高效分离。
3	2024-12-25	其他	开发具备类人五官感知能力(眼、耳、口、鼻、舌)的人工智能系统。
4	2024-12-23	应用与实践类	金刚石基氧化镓异质集成材料突破,验证优异散热能力与晶圆级应用前景。
5	2024-12-20	基础研究类	我国首个工厂化海水制氢项目建成,绿电制氢结合,每小时产氢 20 立方米。
6	2024-12-19	技术与发明类	纳米黏合剂助力丝网印刷柔性热电材料。
7	2024-12-16	基础研究类	硫同位素分析确定燃煤颗粒物为华北雾霾主因,解释欧洲文物黑色硫酸盐壳成因。
8	2024-12-16	技术与发明类	提出“低品质卤水”概念,分析其在锂资源供应中的潜力及技术路径。
9	2024-12-16	技术与发明类	受“铺盖面”启发开发超薄塑料薄膜加工技术。
10	2024-12-13	技术与发明类	合成新型沸石分子筛 ZMQ-1,解析结构,推动重油催化裂化应用。
11	2024-12-13	技术与发明类	二维半导体单晶材料在c面蓝宝石衬底上高取向外延生长,提高光电探测性能。
12	2024-12-12	其他	观测到石墨中载流子倍增,为激发光子能相关的载流子倍增提供直接证据。
13	2024-12-11	其他	利用 AI 揭示负载型金属催化剂中“金属-载体相互作用”本质,建立方程式。
14	2024-12-11	技术与发明类	研制出国内首座智能化天然气掺氢陶瓷梭式窑,实现相关技术突破。
15	2024-12-10	技术与发明类	提出并验证甲烷导的氨动力发动机构想。
16	2024-12-10	应用与实践类	脆性碲化铋基材料实现塑性转化,开发高性能塑性无机热电材料。
17	2024-12-10	基础研究类	异戊二烯硝酸盐驱动亚马逊对流层形成新粒子。
18	2024-12-06	技术与发明类	制造出世界首块碳-14 钴石电池,有持久供电等应用前景。
19	2024-12-04	技术与发明类	开发定向单层多面体膜分离芳烃-脂肪族烃。
20	2024-12-03	技术与发明类	开发多种缓解现有锂提取方法弊端的新途径。
21	2024-12-03	理论研究类	发现海洋排放的甲烷硫醇增加南大洋气溶胶冷却。
22	2024-12-03	技术与发明类	绿色配体演变调控全无机窄带隙钙钛矿薄膜,制备效率 22.57% 的 2 端全无机钙钛矿叠层电池。
23	2024-12-02	技术与发明类	开发安全高能柔性全聚合物水系电池,突破传统锂离子电池局限。
24	2024-12-02	技术与发明类	制备新型复合球材料用于高效海水提铀,吸附容量显著提升。

表2 指标框架

一级指标	二级指标	定义
科学内容	表述严谨性	科学内容的表述是否反映科学成果的内容,且其他涉及的学科基本知识的表述是否符合科学共同体的表述规范。
	来源可靠性	获得科学内容的来源是否可靠。
呈现方式	故事性	表述中体现的事件发生过程。
	情感倾向	表述中蕴含的情感方向。

表3 来源可靠性指标及编码方式

数据来源	赋分范围 / 分	数据类型
研究成果发表期刊	7	原始研究论文
国际顶级期刊(《自然》《科学》等)	6	高影响力期刊论文
化学领域权威期刊(《先进材料》《美国化学会志》等)	6	领域内高影响力期刊论文
学术数据库(Web of Science、中国知网等)	4-5	学位论文、研究报告
科研机构官网发布(中国科学院化学研究所、中国科学院大连化学物理研究所等)	3-4	高校官网、研究所官网发布的科研报道
主流科技媒体(《中国科学报》《美国科学人》等)	2-3	深度报道且引用权威资料
专业行业资讯网站(化工引擎网、中国能源网等)	2	网络发布的科研报道
综合媒体、自媒体(搜狐网、个人博客等)	1	网络发布的科研报道

在呈现方式方面,本研究主要关注前沿成果相关主题,其本质上是新发生的新闻事件,注重科普的叙事逻辑,以及如何传达信息、方式是否有效,因此关注呈现方式的故事

性。根据故事的组成元素，本研究借鉴了艾玛·弗朗西丝·布鲁姆菲尔德（Emma Frances Bloomfield）提出的叙事网框架（Narrative Web），包括人物、行动、顺序、范围、故事

叙述者、内容 6 个方面，具体定义与赋分方式见表 4^[17]。除此之外，由于内容呈现的情感倾向有可能影响受众对相关科学主题的印象，因此本研究将分析科普内容中的情感倾向。

表 4 叙事网指标操作性定义及编码指标

维度 (Narrative Wedges)	指标	定义、解释
人物 (Character)	主体表述清晰度	通过统计科研单位、科研人员姓名、科研人员身份头衔的提及数量衡量报道中对科研成果产出者信息的明确呈现程度。“主体”指所报道的科研成果的产出者，包括科研人员、团队、机构等。
	主体类型数量	报道中明确出现的科研成果产出团队（科研人员）、推动成果应用于企业生产的人员（企业相关人员）、评价指导科研成果的人员（行业专家与学者）这几类主体的数量。
	主体权威度	衡量报道中主体身份地位所体现的权威性，按照不同头衔对应相应分值。
行动 (Action)	关键事件数量	报道中与该成果相关的、已发生的标志性事件的实际数量。
	行动类型数量	在报道科研成果产生过程中，材料制备、性能测试等各个具体步骤类型的数量总和。
	科研步骤描述所占篇幅比重	科研团队研发过程的叙述内容在整篇报道篇幅中所占的比例，以此衡量研发过程描述的详尽程度。
顺序 (Sequence)	时间顺序提示词	报道中用于清晰呈现事件时间先后顺序的连接词的有无情况。
	情节曲折性	报道中是否存在从研究困境到研究突破这一情节转折的叙事特征。
范围 (Scope)	过去、现在、未来齐全度	报道涵盖历史背景（过去）、最新成果（现在）、应用前景（未来）这 3 个时间维度内容的完整性。
	时间节点表述明确度	报道中对事件发生时间是否有精确到日期的清晰表述情况。
	与化学外的学科关联	报道是否体现所报道的科研成果与化学学科之外其他学科之间是否存在联系的情况。
	提及预期应用领域数量	报道中提到的该成果未来可能应用领域的实际数量。
故事叙述者 (Storyteller)	叙述者数量	报道通过直接引用或间接转述方式阐述观点的人物数量。
	叙述者权威度	依据除记者外的叙述者的身份头衔，赋予相应分值来衡量其在报道中的权威性。
	报道风格亲和度	基于报道中术语频次、修辞手法、提问引导 3 个维度使用的次数，对报道风格是否贴近受众进行评价。
内容 (Content)	受众关注主题数	报道中提到的与受众日常生活、兴趣或利益相关的科学主题的实际数量。
	与自身生活相关度	以报道中呈现出的科研成果对受众日常生活的直接影响程度和引发的情感共鸣程度，衡量其与受众自身生活关联紧密的程度。

在研究方法方面，本研究采用专家评分法分析表述的严谨性，请化学领域的研究人员在参考论文原文或设备和项目相关信息的基础上，对科学内容的表述严谨性进行“高”“低”打分，评分标准是只要出现一处和科学内容相关的不准确之处（不包括少量错别字等问题），则判为“低”，没有不准确之处的科学内容则判为“高”。之后，将“高”赋分为 1，“低”赋分为 0，5 个大模型得分取平均分作为大模型生成内容的最终得分。

本研究采用内容分析法分析科学内容的来源可靠性和故事性。由两位研究人员根据编码本对两篇文章进行独立编码，通过一致性检验之后，由一位研究人员进行编码。在来源可靠性方面，根据赋分框架，研究人员对报道进行赋分，5 个大模型得分取平均分为生成式人工智能大模型生成内容的最终得分。在情感倾向方面，本研究使用了 PyCharm Community Edition 2023.3.4 作为主要开发环境，并结合 Python 3.10 来编写与运行代码。

使用的第三方库有：pandas（用于数据的读取、处理与存储）、SnowNLP（用于中文文本的自然语言处理与情感分析）。

其中，SnowNLP 是一个专门面向中文文本的自然语言处理库，它包含了中文分词、词性标注、情感分析、文本分类、关键字或摘要提取、词频—逆向文件频率、文本相似度等多种功能，能够高效地进行对中文文本的情感分析^[18]。SnowNLP 已被广泛应用于社交媒体文本的情感分析研究^[19]，在情感分析方面可直接对中文句子或段落进行情感倾向计算，并反馈情感得分，得分范围为 [0, 1]，其中数值越大表示情感越正面。

其分析过程主要包括三步。第一，数据加载与预处理：数据文件通过 pd.read_csv() 方法加载，文件中应包含每篇报道的编号（“报道编号”）、标题（“标题”）、内容（“内容”）。第二，情感分析：使用 SnowNLP 对每篇报道的内容进行分句处理，并逐句计算情感得分。第三，统计结果：统计每篇报道中的正面、中性和负面句子数量，并计算句子情感得分的均值作为整体情感倾向。

根据情感得分，本研究将 5 个大模型生成的科普内容的情感得分取平均值作为这一主题生成式人工智能大模型生成文章的情感得分，并与《中国科学报》上报道的情感得分做配对 T 检验，考察两组科普文章是否在情感倾向上存在显著性差异。

3 研究结果

3.1 科学内容

在科学内容方面，《中国科学报》与大模

型生成内容的表述严谨性均值相近（0.71 vs. 0.75），整体表述严谨性较好（见表 5）。总体上，大模型生成报道中的不准确之处主要体现在科学错误和科学表述不准确两个方面。在科学错误方面，存在将紫外光波长表述为 400nm（应为 400nm~10nm），将科学方法裂化表述为蒸馏，以及前后科学内容之间建立不存在的因果关系等问题。在科学表述不准确方面，存在翻译不准确、科学名词书写不规范（如 Bi2Te3 中的数字应使用下角标）、论文发表期刊错误、将其他论文中的结论归入另一篇论文中等问题。

表 5 科学内容表述严谨性统计

	平均值	个案数	标准偏差	标准误差平均值
配对样本 人	.71	24	.464	.095
统计 AI	.75	24	.230	.047

但两者的统计数据并没有显著性差异（见表 6），即传统科普内容生成者并没有在表述严谨性方面表现出统计上的优势。配对 T 检验结果显示，两组内容的表述严谨性差异无统计学意义（ $P=0.674 > 0.05$ ）。这表明传统科普内容生成者与大模型在表述严谨性上表现相当，大模型并未因训练数据限制而显著降低表述严谨性。

在来源可靠性方面，《中国科学报》主要来自一手资料，比如论文原文，但大模型往往参考范围更广，不仅包括论文原文，还包括科研机构网站、科研资助机构网站、科学媒体网站（包括《中国科学报》）、行业媒体网站、自媒体发布的文章。而且大模型之间的参考文献选择范围也有所区别，如文心一言相对更加会参考论文原文，而有的大模型主要参考各大媒体报道。

表 6 科学内容表述严谨性配对 T 检验

		配对差值			t	自由度	Sig. (双尾)
		平均值	标准偏差	标准误差平均值			
配对样本检验	人 - AI	-.042	.479	.098	-.426	23	.674

虽然《中国科学报》的来源可靠性均值更高(5.83 vs. 2.65)(见表7),但是两组科普内容之间没有显著性差别($P=0.127 > 0.05$)(见表8)。这表明差异可能源于数据来源多样性,而非系统性偏差。传统科普内容在来源权威性上更具优势,但与大模型生成内容的来源可靠性差异并不显著。

表7 来源可靠性统计

	平均值	个案数	标准偏差	标准误差平均值
配对样本人	5.83	24	2.665	.544
统计 AI	2.65	24	.575	.117

表8 来源可靠性差异显著性检验

	个案数	相关性	显著性
配对样本相关性 人 & AI	24	.321	.127

综上所述,在科学内容方面,《中国科学报》与大模型生成内容在表述严谨性和来源可靠性方面均较好,并且两者之间不存在统计性差异。

3.2 呈现方式

在故事性方面,从下图(图1)可以看出,两者在各个指标得分上具有类似的特点。比如由于是对前沿科研成果的报道,两者都

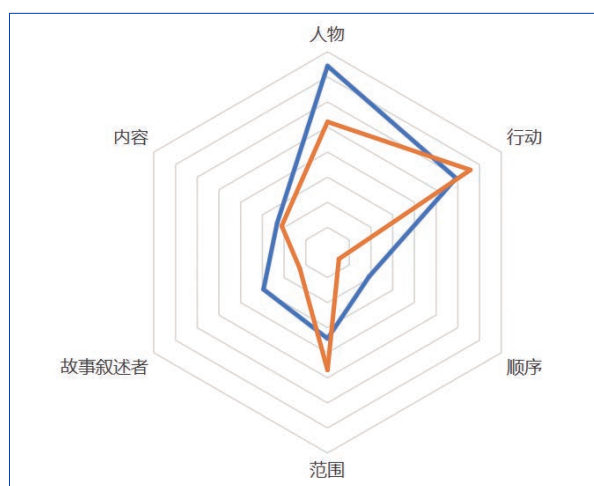


图1 中国科学报与生成式人工智能大模型叙事网

会涉及成果创造者(“人物”)和成果相关事件(“行动”),因此在“人物”和“行动”方面得分较高。而由于前沿科研成果主题与一般公众的日常生活、兴趣或利益(“内容”)距离比较远,两者都较少提到这方面内容,因此,在“内容”方面的得分均较低。

此外,《中国科学报》在叙事网各指标的得分较为均衡,特别是在“人物”“故事叙述者”“顺序”方面。这可能是因为《中国科学报》记者能够直接采访到相关的科学家,从而了解成果获得背后的曲折过程,因此人和事在报道中得到了较好的体现。而大模型在“行动”和“范围”方面表现较好,能够更好地展示研发步骤,并提供理解成果本身的本领域背景性内容和对其他学科意义的内容。

整体上,大模型表现较好的两个指标与《中国科学报》相比也并不表现得特别突出。而《中国科学报》在“人物”“故事叙述者”“顺序”方面的优势较为明显,因而在故事性方面的整体优势更加凸显。

在情感倾向方面,从平均值(见表9)可以看出,无论是《中国科学报》还是大模型生成的内容,在报道前沿科学成果时,基本采用正面的情感(平均值 > 0.5)。然而,配对T检验显示(见表10),大模型生成的同主题内容要比《中国科学报》记者写出的内容具有更高的正面情感倾向性(0.722 vs. 0.687),且两组内容的情感倾向差异具有统计学意义($P=0.008 < 0.05$)。而大模型生成的内容正面

表9 情感倾向样本统计

	平均值	个案数	标准偏差	标准误差平均值
配对样本人	.687	24	.110	.022
统计 AI	.722	24	.066	.014

表10 情感倾向配对T检验

		配对差值			差值 95% 置信区间		t	自由度	Sig. (双尾)
		平均值	标准偏差	标准误差平均值	下限	上限			
配对样本检验	人 - AI	-.035	.059	.012	-.060	-.010	-2.887	23	.008

情感倾向更为显著，可能因其训练数据偏向积极表述，或受传统科普内容的情感基调影响而被进一步放大。

综上所述，在情感倾向方面，大模型生成的内容要比传统报道更倾向于展现正面情感。

4 结论与建议

通过对比传统科普内容生成者和生成式人工智能模型生成的前沿科学成果科普内容，可以看出，大模型在科学内容表述严谨性和来源可靠性方面的表现较好，并且与传统科普内容生产者之间不存在统计学差异。在故事性方面，传统科普内容生成者更有优势。在情感倾向方面，大模型生成的内容要比传统报道更倾向于展现正面情感。

第一，生成式人工智能大模型在前沿科学成果科普内容创作领域，展现出了诸多显著优势。

在科学内容表述严谨性方面，大模型与传统科普内容生产者相比，不存在统计学差异。针对前沿科学成果科普在网络上直接相关内容较少的情况，大模型能够生成表述严谨性较高的内容，较少出现严重误导性内容。在梳理少数不准确内容时发现，同类问题在传统科普作品中同样存在，且出现较多的不严谨内容更多是翻译和科学概念写法上的问题，具有严重误导性的不准确内容比较有限。对于前沿科学成果这类在网络上可学习、可参考、可对比的内容相对较少的科学主题，生成式人工智能模型尚且可以生成较为准确的内容表达，那么对于其他在网络上已经存在较多学习内容的其他科学主题，生成式人工智能模型非常有可能生成表述严谨性更高的科普内容。也就是说，已有研究所担心的

生成错误信息的问题^[4]可能并不会发生，即使发生也可能并不严重。

在科学内容来源可靠性方面，生成式人工智能模型还能够直接参考论文原文（在120篇生成式人工智能生成的文章中，有11篇参考了原论文），因而可靠性较强。因此，虽然理论上生成式人工智能大模型可能会生成错误信息，但随着模型参数规模的扩大和训练数据的增加，相关问题的出现或许在一定程度上能减少。并且，由于生成式人工智能模型生成内容的速度较快、效率较高，并且在生成的过程中，通过人与模型的沟通，大模型可以不断提高科普内容的故事性和通俗性，因此，生成式人工智能技术在科普领域的应用前景较好。尤其对于内容相对来说较为固定的中小学阶段科学类教材来说，利用生成式人工智能产出的前沿科学成果科普文章可以成为教材内容的有益补充。

第二，生成式人工智能大模型在前沿科学成果科普内容创作领域，仍存在一些劣势。

在既有内容依赖方面，由于网络上前沿科学成果内容有限，大模型生成新内容时常参考传统科普内容生成者的已有成果，例如在本研究中，《中国科学报》的内容频繁出现在大模型生成内容的参考文献里。

在情感倾向方面，虽然大模型生成内容与传统报道的情感倾向一致，皆倾向于展现正面情感，但其情感倾向在传统报道的基础上进一步被放大。这可能强化已有刻板印象，加剧公众对科技的盲目乐观，削弱公众对科技伦理的反思能力，从而影响公众对待前沿科学成果的立场，尤其对于不确定性强的前沿科学成果，不利于公众批判性看待并谨慎评判其利弊，可能伤害受众的反思性，进而影响其在科技治理过程中的决策。故对于不

确定性较强的前沿科学成果，更有必要较为中立地开展科普，促进公众批判性地看待前沿科学成果，谨慎评判其可能给科学、社会、一般公众带来的好处和风险，这有利于公众在相关科技治理过程中作出更有利于人类福祉的决策。

第三，生成式人工智能大模型在前沿科学成果科普内容创作领域尚不能完全替代人类传播主体。

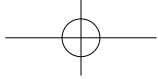
虽然生成式人工智能技术产出的内容在科学内容表述严谨性和信源可靠性方面与传统科普内容生成者的内容没有显著性差异，并且能够让科普内容生成更加便捷，但其尚不能替代已有的传统科普内容生成者。其一，传统科普内容生成者生产的内容为生成式人工智能大模型提供了学习材料，如本研究发现大模型生成内容的参考文献中往往出现《中国科学报》记者的相关报道。传统科普内容生成者不仅为大模型生成内容提供了前沿科学成果的科学内容素材，还能够接触科学家，了解到前沿成果背后的人物与曲折，为大模型的内容呈现提供更多的故事性素材，这些是大模型不可能通过学习获知的内容。其二，传统科普内容生成者的内容更可能在价值观和情感方面发挥根本性的作用。本研究发现，生成式人工智能产出的内容在情感倾向上与传统科普内容生成者是一致的，而且起到进一步加强既有倾向的作用。因此，已有研究所忧虑的生成式人工智能会对当前的科普系统产生冲击，尤其是对传统传播者地位造成威胁的情况^[12]，可能在短期内不会发生。虽然生成式人工智能能够打破传统传播者和受众之间的界限，但在生成式人工智能具有反思性能力之前，将科学内容从科学共同体语境介绍到社会语境过程中必不可少

的科技记者、编辑、科研机构的宣传主管，以及生产科普内容的科学家等科普系统中的角色仍然不可替代。生成式人工智能可以在内容的通俗性、组织形式、故事性等方面推陈出新，但在更根本的方面，比如情感倾向，更可能延续传统科普内容生成者奠定的基调。也就是说，生成式人工智能目前仍然只能起到辅助性的作用，并不能代替传统科普内容生成者。

因此，应该在肯定生成式人工智能大模型优势的前提下，探索平衡大模型和人类内容创作者的方式，进一步提高前沿科学成果科普效率。其一，加强人类内容创作者对大模型的利用能力，通过提高沟通技能、推动大模型内容生成的叙事能力，加强科普内容的通俗性和针对性。其二，提高人类内容创作者对大模型的认识深度，了解大模型在放大已有情感倾向方面的风险，对大模型生成内容有所警惕。其三，制定大模型参与生成科普内容的标注政策，通过作者标注或风险标签的方式，提醒公众对大模型生成内容的风险保持警惕。

但本研究存在一定的局限性。首先，本研究的样本量较少，只关注了《中国科学报》近期报道的24个前沿科学成果主题。其次，本研究仅关注了化学领域的前沿报道，并未涉及其他领域的情况，研究结果可能存在偏差。最后，本研究只聚焦于中国社会情景下的5个大模型，并不能代表国际上通用的大模型，研究结果的普适性可能受到影响。随着生成式人工智能大模型的发展，其对前沿科学成果科普的参与可能进一步提高，但如何更好地利用大模型服务于更广泛的前沿科学成果科普，规避其可能带来的风险，有待进一步的研究。

(下转第42页)



Current Situation and Characteristics in the Development of Scientific Literacy of Chinese Citizens: Main Findings of The 14th National Civic Scientific Literacy Sampling Survey

Gao Hongbin Ren Lei Hu Junping Li Xiuju Huang Yuele Cao Jin

(China Research Institute for Science Popularization, Beijing 100081)

Abstract: To strengthen the national capacity for science popularization and further implement the initiative to enhance the scientific literacy of all citizens, the China Association for Science and Technology (CAST), in collaboration with the National Bureau of Statistics, conducted the 14th national sample survey on Chinese civic scientific literacy in 2024. This survey provided critical data for the evaluation of the “Outline of the National Scheme for Scientific Literacy (2021—2035) ” during the 14th Five-Year Plan period, as well as for achieving the goal of raising the proportion of citizens with scientific literacy to over 15% by 2025. The results show that the scientific literacy of Chinese citizens has been improving at an accelerating pace, with the proportion reaching 15.37% in 2024. The imbalance in scientific literacy development has been alleviated to some extent, and the population with basic scientific literacy is now substantial. The internet has further solidified its position as the primary channel for citizens to access scientific and technological information, while science and technology museums have demonstrated higher visitation and utilization efficiency among various science popularization infrastructures. Overall, Chinese citizens increasingly value science, adopt rational and pragmatic attitudes, and support innovation, with further improvements in critical thinking and scientific awareness. The development of artificial intelligence (AI) enjoys a strong public foundation, with citizens holding positive attitudes and ample confidence toward AI applications, though some concerns remain about job displacement.

Keywords: civic scientific literacy; science and technology information; science popularization infrastructure; science and technology policy-making; artificial intelligence

CLC Numbers: N4 **Document Code:** A **DOI:** 10.19293/j.cnki.1673-8357.2025.02.001

A Study on the Characteristics of Frontier Scientific Achievements in Generative Artificial Intelligence Science Popularization: A Comparison Between Content Generated by Major Chinese Large Models and Chemistry-Related Reports in *China Science Daily*

Wang Cong Lu Yuanfangzhou He Caini

(School of Humanities, University of Chinese Academy of Sciences, Beijing 100089)

Abstract: The popularization of frontier scientific achievements plays a pivotal role in enhancing public understanding and engagement with science. With the widespread application of generative artificial intelligence (GenAI), its potential performance of popularizing frontier scientific achievements has become a critical topic. Unlike the generation of general science popularization content, GenAI encounters distinct challenges when addressing topics related to frontier scientific achievements. Firstly,

there is a paucity of sufficient online textual data for GenAI training. Secondly, these topics often entail relatively higher uncertainty, necessitating a more cautious approach to content formulation. This study conducts a comparative analysis of science popularization content on frontier scientific achievements generated by traditional science popularizers and that produced by GenAI models. The results indicate no significant disparities between the two in terms of scientific accuracy and source reliability. However, the content generated by GenAI models tends to exhibit a more positive emotional tone compared to that produced by traditional science popularizers. Based on these results, this study suggests that while the risk of GenAI disseminating misinformation may be limited, it has the potential to reinforce existing stereotypes. Consequently, GenAI can not entirely supplant traditional science popularizers within the current framework of science popularization.

Keywords: generative artificial intelligence; frontier scientific achievements; science popularization

CLC Numbers: N4; TP18 **Document Code:** A **DOI:** 10.19293/j.cnki.1673-8357.2025.02.002

Evaluation on Science Popularization Writing Ability of Generative Artificial Intelligence: Science Popularization Creation Based on the Theme of “Microbiology”

He Hongpeng Zhang Yaxin Zhang Likai

(Institute for Advanced Studies in Humanities and Social Sciences, Beihang University, Beijing 100191)

Abstract: Generative artificial intelligence (AI) has emerged as a pivotal tool in the creation of science popularization literature; however, its capability in this domain remains subject to skepticism when compared to human authors. To evaluate the sciencepopularization writing competence of generative artificial intelligence, a four-dimensional evaluation framework was developed, comprising readability, engagement, scientific accuracy, and dissemination effectiveness. Using “microbiology” as the thematic focus, sciencepopularization works produced by human creators, DeepSeek, ChatGPT, and ERNIE Bot were systematically examined. The results indicate that both the initial-prompt and deep-prompt versions generated by ChatGPT showed no significant difference in overall scores relative to the human-authored version, with the deep-prompt version exhibiting a significantly superior performance in terms of readability. Moreover, the deep-prompt version of DeepSeek outperformed the human version in the dimensions of engagement and dissemination effectiveness. Notably, evaluators correctly distinguished AI-generated works in less than 55% of cases and showed a tendency to attribute high-scoring works to human creators. These experimental findings suggest that generative artificial intelligence holds the potential to substitute for humanscience popularization creators, thereby prompting the proposal of a novel “human-machine collaborative sciencepopularization creation” model. The study further calls upon the academic community to critically examine the phenomenon of the “Hallucination of Human’s Ability”.

Keywords: generativeartificial intelligence; science popularization writing; science communication; creation ability of artificial intelligence

CLC Numbers: N4; TP18 **Document Code:** A **DOI:** 10.19293/j.cnki.1673-8357.2025.02.003