生成式人工智能赋能科学普及：技术机遇、伦理风险与应对策略

王 硕 阎 妍 李正风

（清华大学社会科学学院，北京 100084）

[摘 要] 生成式人工智能有望极大地赋能科学普及，带来更广泛的主体参与、更丰富的内容呈现、更有效的知识转译以及更精准的科普研究。然而，这类技术也易引发两类潜在的伦理风险：一方面是信息真实性、公平性、隐私安全性、偏见以及版权等生成式人工智能带来的技术伦理问题；另一方面是传统科普的过度通俗化、过少科学沟通、科学权威削弱和民粹主义等问题可能被放大。鉴于此，本文认为，未来需要深入挖掘生成式人工智能在科普实践中的应用潜力，提升公众、科普工作者与监管者的技术认知、应用能力与监管水平，确保科普内容的准确性、隐私安全性及公平性。

[关键词] 生成式人工智能 科学普及 技术机遇 伦理风险

[中图分类号] N4；TP18 [文献标识码] A [DOI] 10.19293/j.cnki.1673-8357.2024.04.001

以 ChatGPT 为代表的生成式人工智能（Generative Artificial Intelligence，以下简称生成式 AI）的迅猛发展为科普带来了众多机遇与挑战。一方面，生成式 AI 有望大幅提升科普内容生产效率，使公众能够轻松创作出高质量的科普文章、视频等多模态内容。另一方面，这种技术也伴随着信息准确性存疑、算法偏见等诸多伦理挑战。党的二十届三中全会强调，“完善生成式人工智能发展和管理机制”[1]。抓住生成式 AI 赋能科普的新机遇，准确理解和有效应对生成式 AI 与科普融合面临的潜在伦理挑战，不仅有助于加深公众的科学认知与参与、提升全民科学素质，还能为科技创新注入新活力，为我国实现高水平科技自立自强与中国式现代化夯实社会基础。

1 研究述评

关于人工智能（Artificial Intelligence，AI）时代的科普，学界已经展开了较为充分的讨论，主要涉及对传统科技传播模式的冲击[2]、与科普活动的深度交互[3]、科学传播体系的重构[4]、科普数字化转型[5]以及面临的伦理挑战[6]等议题。这些研究对把握科普的新发展趋势提供了有益的见解，但是并未充分关注

收稿日期：2024-05-22

基金项目：国家社会科学基金重大项目“深入推进科技体制改革与完善国家治理体系研究”（21ZDA017）；中国科普研究所项目“科普服务文化建设研究”（240103）；清华大学自主科研项目“大学创新体系相关理论与政策问题研究”（20241080002）；浙江省哲学社会科学规划课题“知识分配力视域下新质生产力科普的实践路径研究”（24BMHZ048YB）。

作者简介：王硕，清华大学社会科学学院博士研究生，研究方向：人工智能治理、科技传播，E-mail：s-wang21@mails.tsinghua.edu.cn。李正风为通讯作者，E-mail：lizhf@tsinghua.edu.cn。

到生成式 AI 作为一种颠覆性技术，已经悄然对科普产生了更为系统而深刻的影响。生成式 AI 在许多方面表现出与传统 AI 的本质区别，生成式 AI 能够自动生成内容，这使得它在科普领域不仅仅是辅助工具，还有可能深入改变科普的范式与生态体系。

国内外学界逐渐开始关注生成式人工智能对科学传播与科学普及的影响。国外著名科学传播学者迈克·S.·谢弗（Mike S. Schäfer）在《科学传播杂志》（*Journal of Science Communication*）上发表的论文指出，科学传播研究者在很大程度上忽视了生成式人工智能。他呼吁研究者应该分析“关于人工智能的传播”（about AI）以及“与人工智能的传播”（with AI），探讨生成式 AI 对科学传播本身及其生态系统的影响[7]。国内的科普领域学者则主要关注生成式 AI 对人类智能和创造力的重新定义[8]、生成式 AI 对人类智能的冲击[9]、AI 科普的特性及其与社会的关系[10]、公众在生成式 AI 治理中的能动性和集体义务[11]、生成式 AI 时代的使用者伦理[12]等问题。然而，现有研究主要集中在技术应用和宏观伦理层面，较少直接关注在微观层面上生成式 AI 为科普工作与实践带来的具体机遇及其伴随的潜在伦理挑战。为此，本文旨在对这一议题展开探索性讨论，希望能为科普政策的制定提供启示，也希望能够激发更多同仁的关注与探讨。

2 生成式 AI 为科普带来新的技术机遇

生成式 AI 作为一项新兴且迅速发展的技术，在科普领域具有巨大发展潜力。然而，我们对于这一新兴技术的认识和理解仍然处于初步探索阶段，对于很多前景“看不准”“摸不清”，相关落地实践也具有一定滞后性。所以，目前尚难以形成一个全面和系统的分析框架，来一窥生成式 AI 赋能科普的全貌。但是，从目前已有的实践来看，生成式 AI 在科普领域展现出的应用潜力至少涉及以下 4 个维度，可能成为未来科普研究和实践的关键方向（见图 1）。

2.1 共创：更广泛的科普参与

生成式 AI 有望通过赋能公众，使公众成为科普的重要主体。在传统的科普模式中，知识的生产和传播通常集中在科技工作者、教育工作者、科普工作者等少数专家和学者的手中，大部分公众处于被动接受的信息消费者地位，缺乏参与和创造的主动性。虽然自媒体时代的到来在一定程度上改变了这种权力关系，但创作高质量内容的生产能力依然掌握在少数人手中，公众参与内容生产的能力依然受限。

然而，生成式 AI 凭借其强大的数据检索、集成、分析与自主生成能力，降低了生产高质量科普内容所需要的“文化资本”（cultural capital），赋予公众前所未有的创作自由与能力。无专业背景的普通用户也能够获取有价值的知识，并根据自身需求和兴趣，在社交媒体上创作与分享各种类型、更具创新性和多样性的高质量科普内容，如科技新闻、科普视频等。例如，ChatGPT 开源共享了很多由用户自主探索并创建的定制机器人，其中包括“Science Communicator”“Medical Science Popularization”“PopSci Forge”“SciWrite Assistant”“科普作家”“SciHelper 科普神器”“小卤蛋科普”“科普数字分身”“科普论文写作助手”“PopSci Writer Assistant”等科普聊天机器人，可以根据用户提供的科普主题、目标受众和期望风格等信息生成结构完整的科普文章草稿，包括吸引人的标题、引人入胜的导语、信息丰富的正文和结语，大大降低了高质量科普创作的知识与技术门槛。

在这一意义上，生成式 AI 或许有望降

低知识生产与传播的门槛，使公众能够跨越传统的科普能力障碍，推动其从信息消费者向内容生产者转变。通过这样的“共同生产”（co-production），科学普及将变得更加“平民化”“去中心化”与“共享化”。

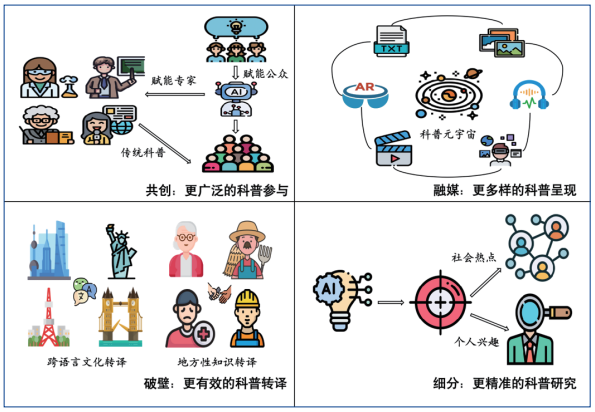

图 1 生成式 AI 为科学普及带来的关键机遇

2.2 融媒：更多样的科普呈现

生成式 AI 有望丰富科普内容的多模态表达。直到今天，我们对于生成式 AI 的技术想象还经常停留在文本生成任务上。事实上，多模态组合才是生成式 AI 未来最具颠覆性的领域。以 OpenAI[①]为例，ChatGPT[②]通过高效的文本生成开启了大模型时代，引发了广泛的应用和关注。DALL-E[③]进一步扩展了生成式 AI 的边界，其可以将文本描述转换为高质量图像，展示了令人惊叹的创造力。而 Sora[④]则代表了文本生成视频的前沿技术，其结合文本生成的语义特征提取和图像生成了每一帧画面，实现了更加复杂和逼真的内容创作。

传统的科普内容主要以文字和图片为主，虽然文字传播具备严谨性和逻辑性，但其抽象性和单一的表现形式限制了公众对复杂科学概念和数据的理解。这种较为单一的传播形式在当前多元化的信息消费环境下显得尤为不足。生成式 AI 的发展打破了这一局限，使得科普能够跨越文本、图像、视频、音频等多种媒介，形成多模态、跨模态的内容，从而能更全面地呈现复杂的科学概念和数据，其通过激活“多重感官”（multisensory）来增强科普效果和“信息保留”（information retention）[13]，丰富科普的形式，增加科普的深度，提升公众的参与度和学习兴趣。

第十三次中国公民科学素质抽样调查数据显示，QQ、微信、微博等社交媒体平台和抖音、快手等短视频平台，已经成为中国公民获取科技信息的主要网络渠道[14]。清华大学科学技术与社会研究中心进行的全国抽样调查显示，超过三分之二的受访者通过微博、知乎、抖音、快手、微信视频号等自媒体平台获取科学技术类信息⑤。随着社交媒体和短视频平台的广泛使用，跨模态表达的重要性愈加凸显，多模态生成技术可以将复杂的科学知识通过生动的图片、动态的视频，以及音频内容呈现，极大地提升用户的理解效率和参与兴趣。我们或许可以期待科普元宇宙（metaverse for science popularization） 的出现与应用，它可以利用多模态生成技术来推动虚拟世界中的科普内容创作，使复杂的科学知识在一个高度互动和沉浸式的环境中得以更生动、具体地呈现，带来更加智能的“人—机”交互式科普体验。

2.3 破壁：更有效的科普转译

生成式AI有望打破跨文化科普的“转译”

①一家位于美国的人工智能研究企业，它开发了包括 ChatGPT 和 DALL-E 在内的多个知名人工智能模型。
②由 OpenAI 开发的一种先进的自然语言处理模型，能够理解和生成人类语言，广泛应用于聊天机器人和文本生成。
③由 OpenAI 于 2021 年 1 月份推出的图像生成系统，可以根据书面文字生成图像，展现了 AI 在视觉艺术领域的潜力。
④由 OpenAI 发布的文本生成视频模型，能够根据文本提示生成连贯的视频内容。
⑤该数据来源于清华大学科学技术与社会研究中心开展的“2023 年科技与社会晴雨表调查”。该调查采用配额非概率抽样方法，基于全国第七次人口普查的地域和年龄分布设计抽样方案，于 2023 年 12 月开展在线问卷调查，共收集了 9 906 个样本，经过筛选得到 3 000 个有效样本。

边界，例如，许多用户切身体验到了 ChatGPT 等生成式 AI 带来的“地道”翻译。所谓的“地道”，不仅指其用词和语法更接近母语者，更在于它能在翻译过程中考虑到文化背景、习俗和语境的细微差别 [15]。生成式 AI 的翻译实现了“动态等价”（dynamic equivalence）和“转译”（translation），能够根据不同地域的文化特点进行微调，例如调整所使用的专业术语、提供与当地文化相关的案例，从而使翻译内容不仅准确，还能让目标受众感到自然和贴切，促进跨文化交流和理解。

在传统认知上，全球前沿科技成果的传播与普及主要依赖于国内专业人士进行翻译和推介。这种传播模式不仅耗时，而且受制于专业人士的科普意愿与能力、激励机制、商业考量以及国内社会对特定领域前沿研究的信息消费需求等多种因素。生成式 AI 的发展成功突破了这些边界，通过实时翻译和文化调适功能，降低了非英语背景的研究者和公众获取全球最新科技进展的门槛，从而推动了知识的全球化与共享。例如，在健康科普领域，“上火”和“湿气”是中国传统医学中常用的概念，但在西方医学中并没有直接对应的术语。生成式 AI 通过文化调适和翻译功能，可以将这些概念准确地传达给外国受众，使不同文化背景的受众能够更准确地理解和接受这些健康信息，促进跨文化的健康交流与理解。

此外，科学知识与地方性知识（local knowledge）的互动是跨文化科学传播中的一个关键问题。地方性知识指特定文化和社区中的传统知识和实践，它们在科学传播中往往被忽视或误解 [16]。生成式 AI 有助于在科普实践中促进科学话语向地方性知识转化，能够通过个性化和情境化的沟通，适应不同群体的需求。这里我们以四类重点科普人群为例进行说明。对于老年群体，生成式 AI 可以将复杂的科学概念以简洁、易懂的语言传达，并结合他们的生活经验加以解释；对于农民，AI 能整合农作物种植、气候变化等科学信息与当地农耕传统，提供有针对性的实用建议；对于患者，生成式 AI 可以将医学知识转化为通俗的健康管理建议，结合个体病史和地方医疗条件，帮助他们理解治疗方案；对于工人，AI 能将工业安全、工作效率等科学概念结合实际工作场景，提供易于理解和应用的安全知识和技术建议，从而更好地促进科学与地方性知识的融合与传播。

2.4 细分：更精准的科普研究

生成式 AI 有望推动科普研究的精准化与个性化。在传统认知上，报刊、电视节目和讲座是科学家与公众沟通的主要途径，这些方式的反馈机制较为缓慢且模糊，难以准确把握公众的兴趣点和需求。然而，随着数字技术的进步，科普者可以通过分析公众在互联网上的活动来更精确地了解他们的兴趣和需求。生成式 AI 在发现和分析公众的科学兴趣与需求方面展现了巨大潜力。例如，生成式 AI 可以快速地从社交媒体上的评论、疑问和反馈中提取出社会对某一科学话题的关注度、情绪和态度。通过分析和整合大量的社交媒体数据，生成式 AI 能够精确捕捉和反映出公共领域中的科学讨论热点和舆论动向，为科普工作者提供宝贵的反馈，响应公众的兴趣和需求，更好地实现科普的互动性和动态性。

这不仅限于社会热点科学话题的热度分析，还可以应用到个性化的科学兴趣点发掘。通过分析个体用户的历史浏览记录、搜索关键词和互动行为，生成式 AI 能够构建个性化的兴趣模型。这种模型不仅能够识别用户当前感兴趣的科学话题，还能够预测他们未来可能感兴趣的领域，并通过用户常用的平台

进行精准推送。通过对受众进行精细化分类，科普工作者能够制定更有针对性的传播策略，提高传播效果和受众参与度。尼古拉斯·尼葛洛庞帝（Nicholas Negroponte）在 1995 年出版的《数字化生存》（*Being Digital*）中预言了“我的日报”（the Daily Me）的诞生，他设想未来每个人的新闻和信息都会根据个人兴趣和需求进行定制，从而提供个性化的阅读体验 [17]，生成式 AI 正在让“我的日报”逐渐变为现实。

3 生成式 AI 与科普融合伴随的潜在伦理风险

生成式 AI 在为科普带来各类机遇的同时，也伴随着一些潜在的伦理风险。为了有效应对这些风险，我们需要准确把握其本质，从而避免陷入对技术的简单否定或盲目崇拜。当前，我们观察到的伦理挑战可以被分为两类。一类是生成式 AI 本身的伦理风险，这些风险在其他领域（如教育、医疗）等同样存在，并且会映射到科普活动中。另一类则是传统科普本身所面临的挑战，但在生成式 AI 技术条件下被放大。分类看待有助于我们更清晰地理解生成式 AI 带来的各种挑战，并制定针对性的治理方案（见图 2）。

图 2 生成式 AI 与科学普及融合的伦理风险

3.1 生成式 AI 本身的技术伦理风险映射到科普活动

首先，信息的准确性与可靠性面临挑战。生成式 AI 能够快速生成大量文本内容，但其信息来源和生成过程并不透明。生成式 AI 依赖训练数据，其数据来源相对多样，包括维基百科、电子书籍、期刊和社交媒体链接等。然而，这些来源很多缺乏严格审查，可能包含虚假和不道德的内容 [18]。这些数据中可能包含错误或偏见，生成的内容也可能反映这些问题。事实上，有关实验表明，公众并不能准确识别深度伪造的科学谣言 [19]。广泛传播未经验证或具有误导性的信息，可能会影响公众的科学理解。并且，生成式 AI 的运作机制对公众来说是一个“黑箱”（black box），其内部过程难以被理解和验证，导致内容出现错误或引发争议时难以追究其责任。2024 年 5 月，Anthropic[①]在理解生成式 AI 模型内部运作机制方面取得重大进展，发现了与代码漏洞、欺骗、偏见等相关的特性，进一步凸显了其缺乏透明性的风险 [20]。长期而言，这些问题可能削弱科学在公众决策中的权威性，影响公众对科学的信任和支持，甚至可能导致社会分裂和对立。

其次，偏见与不公平性的问题同样显著。生成式 AI 在训练过程中可能会吸收和放大现有数据中的偏见，如种族、性别、文化等 [21]，导致不公平或歧视性的信息被传播。尽管算法本身不具有偏见，但它们往往会从训练数据中学习并反映这些数据的特点。例如，一些研究发现，生成式 AI 在提供职业建议时存在性别刻板印象，使用典型的女孩和男孩名字提问时，其提供的 STEM（科学、技术、工程与数学）与非 STEM 建议比例显著不同 [22]。此外，2022 年，一位 YouTube 博主通过训练 GPT–4chan 机器人，生成了 3 万多个包含歧视性言论的帖子，引发了巨大的争议 [23]。偏见性内容不仅可能误导公众，还可能对特定群体造成伤害，加剧社会不公。此外，生成式 AI 的“多轮对话”机制促使机器生成更加贴近用户预期的回答，这可能进一步强化用户已有的知识、观念，形成信息茧房（filter

① Anthropic 是一家美国的人工智能企业，专注于开发通用 AI 系统和语言模型。

bubble），使个体只接触到与自身观念一致的信息，导致观念愈加极端和固化。

最后，生成式 AI 也伴随着潜在的版权归属风险。生成式 AI 通过大量数据训练生成内容，而这些数据通常来自已有作品和资料，因此 AI 生成内容的原创性问题会导致版权归属变得复杂。2023 年 9 月，美国作家协会（The Authors Guild）对 OpenAI 发起集体诉讼，声称其在未经授权的情况下使用原告作家的版权作品训练其大语言模型，生成基于这些小说的总结、复述及模仿作品，这一事件凸显了数据来源的版权问题在 AI 训练中的重要性 [24]。生成式 AI 在生成科普作品时，可能会参考和重组已有的科学文献和文章片段。那么，生成的内容是否属于原创作品？其版权应归属于谁？这些问题在法律和伦理层面尚未有明确的答案。推动科学期刊的开放获取（open access）实践或许是解决科普创作中版权问题的有效手段。开放获取的研究成果经过严格的学术审查，对公众开放，允许自由使用和再创作，从而减少版权纠纷的风险。然而即便如此，在相当长的一段时间内版权问题仍然是不容忽视的挑战。

3.2 传统科普曾经面临的挑战被生成式 AI 放大

首先，生成式 AI 技术加剧了“过多科学通俗化”与“过少科学沟通”的风险。一方面，生成式 AI 时代的到来意味着更多的科学观点和信息在公众场域中竞争，将复杂的科学概念和研究成果以简化和易懂的方式传达给公众，可能存在“简化过度”的风险，导致科学信息的娱乐化和肤浅化。另一方面，生成式 AI 还放大了虚假公众观点和虚假互动的风险。当生成式 AI 被应用于控制社交机器人时，许多内容和评论可能是在平台上伪造的“公众观点”。这些虚假的信息和互动会误导公众，使其误以为这些观点是真实且被广泛认同的。尽管从表面上看，科学信息的传播量增加了，但这些内容和评论并不真正反映公众的真实需求和理解情况，反而会导致信息的同质化和失真，造成“过少的科学沟通”，不仅浪费公众的注意力，还会阻碍公众获取多样化和相互矛盾的信息。

此外，生成式 AI 技术进一步削弱了科普的权威性。一方面，人们总是倾向于将晦涩难懂的语言和大量概念等同于高学术水准 [25]，这可能导致非专家撰写的、基于误导性信息的内容被广泛传播和接受。例如，许多无专业背景的普通科普人士在生成式 AI 的帮助下，能够生成更多看似专业、权威的信息，获得更大范围的信任和传播。另一方面，生成式 AI 能够生成更新颖的信息，而新颖的信息更容易引起公众的注意和信任，很多虚假新闻都是因其新颖性而被转发 [26]。如果公众频繁接触这些看似“权威”“新颖”但未经科学验证的信息，他们可能会逐渐失去对真实科学研究的兴趣与信任，导致“科学民粹主义”（Science-Related Populism）[27]，质疑经过严格验证的科学事实，反而对迎合他们预期设想的“伪科学”深信不疑，进而削弱了科普的权威性。

4 生成式 AI 赋能科普的对策建议

生成式 AI 技术的飞速发展极大赋能了科学知识的传播和普及，带来了更广泛的科普参与、更多样的科普呈现、更有效的科普转译以及更精准的科普研究。然而，随着生成式 AI 技术的快速迭代和广泛应用，其潜在的风险也逐渐显现。当前，社会各界需要协同努力，抓住生成式 AI 技术赋能科学普及的新机遇，同时深入理解生成式 AI 技术带来的伦理挑战，以促进相关技术更好赋能科普事业健康、向善发展。本文将从技术应用、素养提升以及风险管控三个方面提出具体对策建议（见图 3）。其中，技术应用是基础，它为生成式 AI 技术赋

能科普提供了坚实的支撑；素养提升是关键，它决定了各利益相关方能否有效利用这一技术推动科学普及；风险管控是底线，它确保在技术赋能科学普及过程中，信息真实性、隐私安全性以及公平性等方面能够得到有效保障。

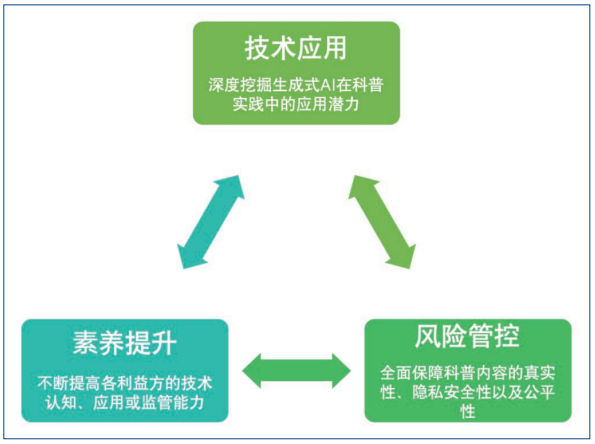

图 3 生成式 AI 赋能科学普及的对策建议

4.1 技术应用：深度挖掘生成式 AI 技术在科普实践中的应用潜力

提升科普内容的可读性与趣味性。鼓励科技期刊和科普平台积极利用生成式 AI 技术，整合多种媒体资源，如视频、音频、动画和 3D 模型等，使科技内容更加直观、易懂。当前，一些医学期刊已经开始在论文中整合 3D 模型、360°全景视图、音视频资料、图片以及网页链接等资源。《实用临床医药杂志》2022 年第 14 期的《颅脑手术定位中三维影像体表投影技术的应用价值》一文，便是利用 AR 技术和“多模态数智内容编辑器”编辑的首个案例 [28]。通过这种多模态整合，复杂的科学概念能以更加生动、形象的方式呈现，从而提高用户的学习兴趣和科学素质。

强化科普工作的个性化与互动性。生成式 AI 技术在科普工作中应充分挖掘用户数据，实现精准的用户画像构建，从而提供高度定制化的科普内容。例如，对于青少年用户，平台可以提供互动性强、富有挑战性的科学游戏和实验，激发他们的探索欲和求知欲。同时，进一步开发基于生成式 AI 技术的互动科普体验，提供沉浸式的科普环境，促进用户与科普内容的深度互动。例如，在我国“十三五”科技创新成就展上，科普机器人“小科”受邀来到现场，全面展示了海量科普知识问答、诗词创作、开放式对话等“技能”[29]。

推动开放科学与生成式 AI 技术的结合。开放科学能够基于全员参与、全过程协同、高度透明性以及共享性，促进科学知识与研究成果的充分共享 [30]。充分利用开放科学资源，有助于提供丰富的科学数据、实验资源和素材库，支持大众用户利用生成式 AI 技术生成更高质量、更准确、更权威的科普内容，从而激发用户的创造力和参与意愿。例如，开放科学交流平台 ResearchGate 和 Academia.edu，允许科研人员就具体的研究问题进行交流和讨论，这些交流内容可以为生成式 AI 系统提供实时的科学讨论和观点，增强其生成内容的时效性和互动性。

4.2 素养提升：不断提高各利益方的技术认知、应用或监管能力

提升社会公众科学素质和数字伦理素养。在大语言模型技术迅速发展的今天，公众的角色已经从信息接收者转变为信息的生产者和分发者。这要求公众提升科学素质与技能的同时增强数字伦理意识 [31]。一方面，建议进一步提升社会公众的科学素质，以“求真”精神辨别科技信息的真伪，通过在教育体系中强调科学方法论和批判性思维的培养，让公众能够评估信息的准确性和可靠性，并审慎辨别信息来源，保持独立思考。另一方面，建议增强公众的数字伦理意识，使公众能以“向善”的原则正确使用新兴技术，了解数字信息的产生和传播过程，明确自身的角色和责任，审慎使用数字工具，维护内容的透明度和可追溯性。

提高科普工作者对生成式 AI 技术的理

解和应用能力。引导科普工作者理解大语言模型工具的工作方式，以及这些工具何时以及如何影响科普活动，为合理、规范使用大语言模型技术提供清晰、全面且适配的信息。为了增强科普工作者对生成式 AI 技术的认识与实践，科技期刊和科普平台应积极开展技术培训和合作，推动生成式 AI 技术在科普中的标准化和规范化应用。

提高监管者对生成式 AI 技术的认知和监管能力。生成式 AI 技术的迅猛发展给监管者带来了“科林格里奇困境”（Collingridge Dilemma）①。生成式 AI 技术发展过快，科普内容的形式和表达方式不断创新，传统的监管框架往往滞后，难以及时跟上技术的飞速变化，导致其在管理风险时捉襟见肘。这种滞后不仅导致潜在风险管理的缺失，还可能由于监管的不确定性或过度严格的管控，压制了生成式 AI 在科普领域的创新活力，削弱了我国在该领域的技术竞争力 [32]。为此，监管者也需要不断提高有关生成式 AI 技术的专业知识，增强对复杂技术趋势的预测和响应能力，从而能够及时更新和调整监管策略，平衡技术创新与风险管理。

4.3 风险管控：全面保障科普内容的准确性、隐私安全性以及公平性

提升科普信息的真实性和准确性。成立由科学家、教育专家和媒体专家组成的审核委员会，对生成式 AI 生成的科普内容进行双重审核，确保其科学性。建立科普内容认证体系，例如在短视频平台通过认证标识增强公众对内容的信任，并定期复查以维持标准。同时，建议有关平台明确标注科普内容生成方式和数据来源，并实施透明化报告制度。

确保数字科普过程中的数据隐私保护。相关部门需要进一步制定严格的科普数据相关保护监管政策，规范生成式 AI 技术在数据收集和处理中的行为，明确要求科普平台在收集用户数据前必须获得用户同意，并告知用户数据使用与保护方式。在技术层面可以采用先进的加密技术和安全措施，防止数据泄露和未授权访问。此外，建议平台定期公开数据使用情况，并成立独立的隐私保护审查委员会，监督生成式 AI 技术的数据使用合规性，确保数据处理过程的透明性和安全性。

减少生成式 AI 技术在科普应用中的算法偏见与歧视。算法是生成式 AI 技术的核心，必须优化发展，遵循公平性原则 [33]。监督技术研发平台使用多样化的训练数据，涵盖不同性别、种族、文化和社会经济背景，以建立多元化的公共数据集和科普数据集，确保科普内容的广泛性和代表性。鼓励公众和专家参与生成式 AI 技术的开发与监督，通过设立开放的反馈和举报机制，收集并分析反馈以及时修正偏见问题。此外，应当制定算法透明度和责任制度，公开生成式 AI 的数据来源，明确算法偏见造成误导信息的责任主体，并制定惩处措施，以增强公众对科普内容的信任，确保算法开发者和应用者对其生成内容负责。

加强对关键科普应用领域的监管。当前需要进一步制定全面的监管框架，确保生成式 AI 技术在短视频等高影响力领域中的应用既符合伦理和法律要求，又能保证科普内容的质量和公信力。一方面，政府应当完善法律框架，明确规定生成式 AI 生成内容的责任主体，制定明确的法规和标准，要求平台和个人遵守特定标准，并对违规行为实施严厉惩罚。另一方面，政府应推动抖音、快手、微信视频号等重点平台自律，形成内部行为准则和道德标准，推动社交媒体平台成立行

①科林格里奇困境（Collingridge Dilemma）指在技术发展早期，因对其潜在影响缺乏了解，很难对其进行有效监管或引导；而一旦技术广泛应用，其影响变得清晰，但调整或控制的难度也随之增大。

业联盟，制定自律公约，定期提交自查报告，确保内容生成和传播的规范性和透明度。

5 结语

生成式 AI 不仅标志着科普领域的一次技术革新，更是提升全民科学素质、厚植科学文化沃土的关键契机。因此，一方面，需要在未来积极抓住生成式 AI 赋能科普的新机遇，深入探索生成式 AI 在科普实践中的创新应用场景，有效提升科普的质量与广度。另一方面，也要准确理解和把握生成式 AI 在科普中带来的伦理挑战，加强对生成式 AI 应用的有效规范与监管，促进生成式 AI 更好地赋能科普事业健康、向善发展。

参考文献

[1] 新华社 . 中共中央关于进一步全面深化改革 推进中国式现代化的决定 [EB/OL].（2024-07-21）[2024-07-24]. http://www.news.cn/politics/20240721/cec09ea2bde840dfb99331c48ab5523a/c.html.

[2] 宋微伟 . 互联网时代人工智能对科技传播的冲击及其融合 [J]. 科技传播，2019，11(3)：131-132.

[3] 黄时进 . 技术驱动与深度交互：人工智能对科学传播的跨世纪构建 [J]. 科普研究，2020，15(3)：16-19.

[4] 董洪哲 . 科学传播在人工智能时代的变革与思考 [J]. 中国广播，2020(10)：29-32.

[5] 李正风，张徐姗 . 走向“数字社会”进程中的科学普及 [J]. 科普研究，2023，18(4)：8-17.

[6] 王硕，李秋甫 . 数字伦理：数字化转型中科学普及的新使命与新规范 [J]. 科普研究，2023，18(3)：57-64.

[7] Schäfer M S. The Notorious GPT：Science Communication in the Age of Artificial Intelligence[J]. Journal of Science Communication，2023，22(2)：1-15.

[8] 段伟文，余梦 . 生成式人工智能时代的终身创造力培养 [J]. 科普研究，2024，19(2)：13-22.

[9] 刘永谋，伍铭伟 . AIGC 对人类智能的冲击与提升公民科学素质的可能路径 [J]. 科普研究，2024，19(2)：39-44，63.

[10] 杨庆峰 . 人工智能科普及其问题 [J]. 科普研究，2024，19(2)：32-38.

[11] 李健民 . 促进公众理解生成式人工智能的集体责任分析——基于集体行动的视角 [J]. 科普研究，2024，19(2)：45-53.

[12] 闫宏秀，杨映瑜 . 生成式人工智能时代的使用者伦理研究 [J]. 科普研究，2024，19(2)：23-31.

[13] Broadbent H J，Osborne T，Mareschal D，et al. Withstanding the Test of Time：Multisensory Cues Improve the Delayed Retention of Incidental Learning[J]. Developmental Science，2019，22(1)：e12726.

[14] 中国公民抽样调查课题组 . 我国公民科学素质的对策发展现状——基于第十三次中国公民科学素质抽样调查的分析 [J]. 科普研究，2024，19(2)：5-12.

[15] Lee T K. Artificial Intelligence and Posthumanist Translation：ChatGPT versus the Translator[EB/OL].（2023-08-01）[2024-07-09]. https://doi.org/10.1515/applirev-2023-0122.

[16] Anaafo D. Between Science and Local Knowledge：Improving the Communication of Climate Change to Rural Agriculturists in the Bolgatanga Municipality，Ghana[J]. Local Environment，2019，24(3)：201-215.

[17] 尼葛洛庞帝 . 数字化生存 [M]. 胡泳，范海燕，译 . 北京：电子工业出版社，2017：182.

[18] Thompson A D. What’s in My AI? [EB/OL]. [2024-07-09]. https://s10251.pcdn.co/pdf/2022-Alan-D-Thompson-Whats-in-my-AI-Rev-0b.pdf.

[19] Doss C，Mondschein J，Shu D，et al. Deepfakes and Scientific Knowledge Dissemination[J]. Scientific Reports，2023，13(1)：13429.

[20] Scaling Monosemanticity：Extracting Interpretable Features from Claude 3 Sonnet[EB/OL]. [2024-07-07]. https://transformer-circuits.pub/2024/scaling-monosemanticity/index.html.

[21] Gillespie T. Generative AI and the Politics of Visibility[J]. Big Data and Society，2024，11(2)：1-24.

[22] Due S，Das S，Andersen M，et al. Evaluation of Large Language Models：STEM Education and Gender Stereotypes[EB/OL].（2024-06-14）[2024-07-09]. https://arxiv.org/abs/2406.10133.

[23] Vincent J. YouTuber Trains AI Bot on 4chan’s Pile O’ Bile with Entirely Predictable Results[EB/OL].（2022-06-08）[2024-07-09]. https://www.theverge.com/2022/6/8/23159465/youtuber-ai-bot-pol-gpt-4chan-yannic-kilcher-ethics.

[24] Gerken T，McMahon Liv. Game of Thrones Author Sues ChatGPT Owner OpenAI[EB/OL].（2023-09-21）[2024-07-09]. https://www.bbc.com/news/technology-66866577.

（下转第 22 页）

[7] Biyela S，Dihal K，Gero K I，et al. Generative AI and Science Communication in the Physical Sciences[J]. Nature Reviews Physics，2024，6 (3)：162–165.
[8] De Angelis L，Baglivo F，Arzilli G，et al. ChatGPT and the Rise of Large Language Models: the New AI–Driven Infodemic Threat in Public Health[J]. Frontiers in Public Health，2023(11)：1166120.
[9] 朱永新，杨帆 . ChatGPT/ 生成式人工智能与教育创新：机遇，挑战以及未来 [J]. 华东师范大学学报（教育科学版），2023，41(7)：1–14.
[10] Nerlich B. Artificial Intelligence：Education and Entertainment[EB/OL].（2023–01–06）[2024–05–28]. https://blogs.nottingham.ac.uk/makingsciencepublic/2023/01/06/artificial–intelligence–education–and–entertainment/.
[11] Maas P，Carey F，Chris Wheeler C，et al. To Infinity and Beyond：SHOW–1 and Showrunner Agents in Multi–Agent Simulations[EB/OL].（2023–06–24）[2024–06–20]. https://fablestudio.github.io/showrunner–agents/.
[12] 李亚玲，覃缘琪，魏阙 . 人工智能生成内容的潜在风险及治理对策 [J]. 智能科学与技术学报，2023，5(3)：415–423.
[13] Schafer M S. The Notorious GPT：Science Communication in the Age of Artificial Intelligence[J]. Journal of Science Communication，2023，22(2)：Y02.
[14] Alvarez A，Caliskan A，Crockett M J，et al. Science Communication with Generative AI[J]. Nature Human Behaviour，2024，8(4)：625–627.
[15] Huang J，Tan M. The Role of ChatGPT in Scientific Communication：Writing Better Scientific Review Articles[J]. American Journal of Cancer Research，2023 13(4)：1148–1154.
[16] 周慎，房曼琳，陈蔚浩 . 科技叙事的生成式变革：人类智能与人工智能的互构与平衡 [J]. 数字出版研究，2024，3(2)：125–132.
[17] Nesari T M. The Emergence of ChatGPT：An Opportunity or Adversity for Scholarly Communications[J]. International Journal of Ayurveda Research，2024，5(1)：1–4.
[18] 卢卫红，杨新福 . 人工智能与人的主体性反思 [J]. 重庆邮电大学学报（社会科学版），2023，35(2)：85–92.
[19] 周慎 . 新文本间性：生成式人工智能的文本内涵、结构与表征 [J]. 新闻记者，2023(6)：39–45.
[20] 腾讯云联合信通院发布《行业大模型标准体系及能力架构研究报告》[EB/OL].（2023–09–25）[2024–05–28]. https://cloud.tencent.com/developer/article/2332984.
[21] 王唯滢，王丽慧，王挺 . 加强国家科普能力建设：时代使命、基本内涵与实践路径 [J]. 科普研究，2024，19(1)：5–16.

（编辑 颜 燕 和树美）

（上接第 13 页）

[25] Sokal A. A Physicist Experiments with Cultural Studies[J]. Lingua franca，1996，6(4)：62–64.
[26] Vosoughi S，Roy D，Aral S. The Spread of True and False News Online[J]. Science，2018，359(6380)：1146–1151.
[27] Mede N G，Schäfer M S. Science–related Populism：Conceptualizing Populist Demands Toward Science[J]. Public Understanding of S cience，2020，29(5)：473–491.
[28] 游慧超，孙登江，欧阳和平，等 . 颅脑手术定位中三维影像体表投影技术的应用价值 [J]. 实用临床医药杂志，2022，26(14)：14–17.
[29] 机器人“小科”亮相“十三五”科技成就展 [EB/OL].（2021–10–27）[2024–09–21]. https://www.stdaily.com/index/kejixinwen/2021–10/27/content_1228536.shtml.
[30] 陈劲，阳镇 . 数字化时代下的开放科学：伦理难题与推进路径 [J]. 吉林大学社会科学学报，2021，61(3)：116–128，236.
[31] 李正风，王硕 . 数字素养、数据权利与数字伦理 [J]. 科普研究，2022(6)：8–14.
[32] 童云峰 . 走出科林格里奇困境：生成式人工智能技术的动态规制 [J]. 上海交通大学学报（哲学社会科学版），2024，32(8)：53–67.
[33] 李伟鑫，王晓丽 . ChatGPT 类生成式人工智能的伦理问题、原因分析与规制路径 [J]. 昆明理工大学学报（社会科学版），2024，24(1)：54–62.

（编辑 颜 燕 和树美）

Empowering Science Popularization through Generative Artificial Intelligence: Technological Opportunities, Ethical Risks, and Response Strategies

Wang Shuo Yan Yan Li Zhengfeng

(School of Social Sciences, Tsinghua University, Beijing 100084)

Abstract: By facilitating greater participation, richer content presentation, more efficient knowledge translation, and more focused science outreach, generative AI has the potential to enhance science communication and popularization significantly. This technology also introduces two different kinds of ethical risks. On the one hand, generative AI raises technical, ethical questions about information authenticity, fairness, security, bias, and copyright. On the other hand, problems historically linked to science popularization may get worse, such as oversimplification, a lack of scientific communication, the erosion of scientific authority, and populism. Consequently, it is crucial to investigate the application potential of generative artificial intelligence in science popularization practices going forward, as well as to raise the technical awareness, application capabilities, and supervision level of the general public, science popularization professionals, and regulators. Additionally, it is necessary to guarantee the accuracy, privacy security, and fairness of the content produced for science popularization.

Keywords: generative AI; science popularization; technological opportunities; ethical risks

CLC Numbers: N4; TP18 Document Code: A DOI: 10.19293/j.cnki.1673-8357.2024.04.001

Analysis of the Underlying Logic and Participating Subject Functions of Artificial Intelligence-Generated Science Content

Zhou Shen[1,2] Tao Meili[1] Liu Xiang[3]

(Department of Science and Technology Communication, University of Science and Technology of China, Hefei 230026)[1]

(Anhui Key Laboratory of Science Education and Communication, School of Humanities, University of Science and Technology of China, Hefei 230026)[2]

(College of Literature and Art, Southwest University of Science and Technology, Mianyang 621002)[3]

Abstract: The content creation in science popularization led by artificial intelligence means that science popularization content production has entered a brand new stage of development. This paper analyzes the current situation and challenges in the development of AI-Generated science popularization content, the underlying logic of AI-Generated science popularization content from the technical paths of large model pre-training, vertical domain model fine-tuning, and popular science content generation, as well as the roles and functions of the three types of science popularization content producers, namely professional content producers, occupational content producers, and the public, in the process of collaborating with AI to create science popularization content. Based on the results of the study, this paper suggests that in order to realize the high-quality development of science popularization, it is necessary to carry out the "AI +" action in the field of science popularization, promote the in-depth fusion of AI and science popularization, and build a new paradigm of human-machine collaboration and multiple co-creation of