生成式引擎优化实践中的风险与信息生态重塑

2025 年 9 月

郑重声明

本人郑重声明，所提交的论文《生成式引擎优化实践中的风险与信息生态重塑》，为本人在研究过程中独立完成的原创作品。论文中的数据和结论均真实可靠，引用他人文献已遵守相关引用规范，并在参考文献中明确标注来源。如有抄袭、剽窃或侵犯他人著作权的行为，本人愿承担相应法律责任。

Risks in Generative Engine Optimization Practice and the Reshaping of Information Ecosystems

Sep.,2025

摘要

近年来，随着 ChatGPT 等大语言模型的普及，生成式人工智能（Generative AI）对信息检索和分发模式产生了颠覆性影响，传统的搜索引擎优化（SEO）逐步让位于生成引擎优化（Generative Engine Optimization, GEO）。GEO 的核心目标是通过优化内容的可见性、可信度和算法适配性，确保信息能在生成式 AI 的输出结果中被准确学习和展现。本文从新闻传播学、认知心理学等多学科视角，系统分析了 GEO 实践背后的关键机制、伦理困境及风险特征，特别是知识产权归属、算法偏见、可解释性与虚假信息等问题。研究发现，GEO 既可能重塑当前的信息生产格局和传播秩序，也可能加剧信息生态的均衡失衡和权力集中化风险。针对上述挑战，本文提出了五大应对策略，包括技术与伦理深度融合、透明化建设、内容生态去中心化以及公众 AI 素养的提升。本文的研究不仅拓展了生成式传播环境下的理论框架，也为 GEO 实践提供了可操作性的建议。

关键词：生成式引擎优化；GEO；生成式人工智能；人机信任；可解释性

Abstract

In recent years, with the popularization of large language models such as ChatGPT, Generative Artificial Intelligence (Generative AI) has exerted a disruptive impact on information retrieval and distribution models. Traditional Search Engine Optimization (SEO) is gradually giving way to Generative Engine Optimization (GEO). The core goal of GEO is to optimize the visibility, credibility, and algorithm adaptability of content, ensuring that information can be accurately learned and presented in the output results of generative AI.

From an interdisciplinary perspective encompassing journalism and communication studies, cognitive psychology, and other fields, this paper systematically analyzes the key mechanisms, ethical dilemmas, and risk characteristics behind GEO practices—with a particular focus on issues such as intellectual property ownership, algorithmic bias, interpretability, and disinformation. The study reveals that GEO may not only reshape the current pattern of information production and communication order but also exacerbate the risks of imbalance in the information ecosystem and centralization of power.

In response to the aforementioned challenges, this paper proposes five coping strategies, including the in-depth integration of technology and ethics, transparency development, decentralization of the content ecosystem, and enhancement of the public's AI literacy. The research presented in this paper not only expands the theoretical framework in the context of generative communication but also provides actionable recommendations for GEO practices.

Keywords: Generative Engine Optimization;GEO;Generative Artificial Intelligence; Human-Machine Trust; Explainability

目录

1 引言：从“链接为王”到“答案为王”的范式革命

互联网诞生的二十多年后，搜索引擎作为必不可少的检索工具也随之问世。自此，搜索引擎一直是用户获取信息的核心枢纽。用户在搜索框输入关键词，获得十来个蓝色链接清单，点击后跳转至目标网页。企业为争夺这份“目录”的头部位置，发展出 SEO（搜索引擎优化）战术，SEO 围绕着关键词、友情链接、网站加载速度等策略，以期获得更多流量。

然而，随着生成式人工智能技术的崛起，特别是大语言模型（LLMs）的广泛应用，用户与信息交互的方式正在发生根本性变革。人们不再满足于索引列表，而是更倾向于直接向 AI 提问，获取高度整合、概括和定制化的“权威答案”，将浏览过程完全简化。这标志着信息检索从“链接为王”时代向“答案为王”时代的迈进。当你询问：“如何给三个月大的布偶猫驱虫？”GPT-4 会直接生成分步骤指南、药品推荐及注意事项——过去需点击 5 个网页才能拼凑的信息，现在被整合成一份“傻瓜答案”。生成式 AI 正在吞食传统搜索的商业规模：Perplexity 用户量半年增长 10 倍，谷歌的搜索广告收入在 2024 年第一季度首次下滑。在这一背景下，传统的 SEO 策略将面临严峻挑战，甚至可能失去部分效用。一个全新的概念——生成引擎优化（Generative Engine Optimization，简称 GEO）应运而生。GEO 旨在研究和实践如何让品牌、产品、服务或知识内容，能够被生成式 AI 模型有效学习、理解、采纳，并最终以高相关度、高可信度的方式呈现给寻求答案的用户。这不仅是技术层面的优化，也是对内容生产、传播策略、用户认知乃至信息生态系统的全面重塑，更引发了人机信任危机。只要个体或组织理解了的大语言模型的“消化系统”如何评估信息的可信度、完整性与时效性，就能教会 AI 说话。本文将立足于这一范式变革，深入探讨 GEO 的内涵与挑战，并尝试从多学科理论视角，提出一套可能的理解和指导 GEO 实践的理论框架，以期为智能传播领域的学术研究与业界实践提供新的思辨方向。

2 生成引擎优化（GEO）的内涵与技术基石

2.1 生成式 AI 与信息检索的变革

生成式 AI，尤其是大语言模型（LLMs），其核心能力在于能够理解自然语言输入，并生成符合语境、逻辑清晰且富有创造性的文本、代码、图像等内容。“在预测人类文本的过程中，它们既能学习事实，又能模拟（至少某些方面的）人类推理能力。”[1]大型语言模型（LLMs）的出现，也催生了一种全新的搜索引擎范式——利用生成模型收集和总结信息，以回答用户查询的问题。这种新兴技术能生成准确且个性化的回复，正迅速取代谷歌、百度等传统搜索引擎。2025 年 7 月，ChatGPT 的全球市场份额约占 81%（Perplexity 约 8%、微软 Copilot 约 5%、Gemini 约 2%）[2]。而 Perplexity 方面则透露，2025 年 5 月其月查询量达到 7.8 亿次[3]。

生成引擎通常通过整合多个来源的信息，并借助大型语言模型进行总结，从而满足用户的提问需求。当这类模型被应用于问答式智能助手（如 DeepSeek、豆包、腾讯元宝、Kimi 和 GPT 等），用户体验随之改变。如今用户能用自然语言直接提问复杂问题，不再依靠关键词搜索。生成式 AI 会从多个来源提取信息，将全网信息整合加工，生成包含关键要点的总结式答案，而非简单的链接列表。并且它能根据之前的聊天记录调整回答，还能陪你深入追问。这种变革意味着，信息检索到的内容的“可见性”和“优先级”不再仅仅取决于其在搜索结果中的排序，更取决于其内容本身能否被 AI“理解”并“采纳”到其答案中。然而，生成式 AI 的类人修辞伪装模糊了机器逻辑与真实知识的边界，确定性语言修辞则（如“科学研究表明”“众所周知”）掩盖其概率性本质，制造出虚假的专业权威感，同时，其流畅性掩盖了内容的不确定性，其清晰的思考步骤掩盖了算法有限的可解释性和不透明性。生成式 AI 由此成为十分隐秘的滋生欺骗性信息和操纵性信息的温床[4]。

---

[1] Park, P. S., Goldstein, S., O’Gara, A., Chen, M., & Hendrycks, D. AI deception: A survey of examples, risks, and potential solutions [J]. Patterns, 2024, 5(5).
[2] Chen M, Wang X, Chen K, et al. Generative Engine Optimization: How to Dominate AI Search[J]. arXiv preprint arXiv:2509.08919, 2025.
[3] Chen M, Wang X, Chen K, et al. Generative Engine Optimization: How to Dominate AI Search[J]. arXiv preprint arXiv:2509.08919, 2025.
[4] Park, P. S., Goldstein, S., O’Gara, A., Chen, M., & Hendrycks, D. AI deception: A survey of examples, risks, and potential solutions [J]. Patterns, 2024, 5(5).

2.2 生成式引擎优化（GEO）的定义与核心目标

生成式引擎优化（Generative Engine Optimization，后文简称 GEO）是指通过一系列策略和技术手段，优化内容在生成式 AI 模型中的可见性、可信度和被采纳率，从而提升品牌或信息在 AI 生成式答案中的呈现质量和影响力。GEO 的方法能帮助内容生产者从呈现形式、文本风格、内容论述方式等方面优化信息内容。

GEO 的核心目标即，提高内容被生成式引擎作为信息来源的概率，并按照内容生产者和发布者所期望的那样被呈现出来。而要实现这一目标，一是要提升内容可学习性，确保 AI 模型能够有效抓取、理解和学习内容的核心信息。二是要增强内容可信度，建立内容的权威性和真实性，使其成为 AI 信任的知识来源。三是优化内容呈现方式，影响 AI 生成答案的呈现方式、语气和指向，使其更有利于品牌传播。四是管理内容风险，规避 AI 生成错误信息、偏见或不当提及的风险。

2.3GEO 的技术基石与策略方向

GEO 的技术基础在于理解生成式引擎如何处理和发送信息，并在此基础上开发有效的优化策略。当前，GEO 方法可以被视为优化函数，以原始网站内容为输入，输出一个经过优化的版本，导出在生成式引擎中获得更高的可见性[5]。为了简化可见性，目前的 GEO 方法提出了一系列“印象指标”（Impression Metrics），这些指标不仅涉及了引用内容的字数和位置，还加入了额外的因素，例如引用内容与查询的相关性、引用的影响力、信息的独特性和多样性、用户点击引用的可能性以及赋予用户的信息量（主观印象）。这些指标为内容创作者提供了评估其网站在生成式引擎中表现的标准。基于对生成式引擎机制和印象指标的理解，GEO 提出了多种的策略方向。Aggarwal P 等学者率先提出生成式引擎中内容可见性的衡量指标，并通过实验验证其提出的权威性调整（Authoritative）、添加专业术语（Technical Terms）、简化易懂（Easy-to-Understand）、流畅度优化（Fluency Optimization）、添加引用文献（Cite Sources）、添加引言（Quotation Addition）、添加引用数据（Statistics Addition）的 GEO 方法能将内容可见性提升高 40%，而关键词堆砌（Keyword Stuffing）和添加独特词汇（Unique Words）的方法并不能

[5] Aggarwal P, Murahari V, Rajpurohit T, et al. Geo: Generative engine optimization[C]//Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining. 2024:5-16.

生效[6]。GEO 不仅为小型内容创作者提供了与大型机构竞争的机会，也强调了领域特定优化策略的重要性。加拿大的学者们基于实证研究的研究结果，提出了一套具有战略性的 GEO 方案：优化内容结构、主导口碑媒体、采用针对特定引擎和语言环境的优化策略和帮助小众品牌克服固有的“大品牌偏见”[7]。

[6] Aggarwal P, Murahari V, Rajpurohit T, et al. Geo: Generative engine optimization[C]//Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining. 2024:5-16.
[7] Chen M, Wang X, Chen K, et al. Generative Engine Optimization: How to Dominate AI Search[J]. arXiv preprint arXiv:2509.08919, 2025.

3GEO 的风险：原创性、伦理与信息生态风险

3.1 内容原创性与知识产权的困境

当生成式 AI 模型整合多源信息生成答案时，内容的原创性归属将变得模糊。哪些信息来自哪个信源？谁是原作者？这给知识产权保护带来了巨大挑战。GEO 对内容原创性与知识产权的冲击不仅体现在归属与归因的模糊性上，还对新闻传播行业的生态系统产生了深远影响。语言模型就“随机鹦鹉”，只是根据概率将训练数据中的语言形式拼凑在一起，而不涉及意义。因此 Bender 等学者认为，生成式 AI 在综合和呈现信息的过程中，常常无法准确反映每个信息来源的特定贡献[8]。这种“归因缺失”问题，进一步削弱了对于内容创作者的保护，使得付出原创性劳动的作者可能因为难以明确身份而无法获得应有认可与经济收益[9]。

这种局面可能带来以下几方面的影响：首先是内容创作者的激励机制受损。新闻记者或专业内容创作者，投入相当多的时间和精力采集和撰写的高质量内容，却被 AI 模型在几秒之内爬取后生成答案、无偿分发，这无疑减少了原创性内容的市场需求与经济价值[10]。当前，新闻产业对知识付费和广告收入的高度依赖进一步加剧了这一问题，因为生成内容是的读者接触原创新闻的机会减少从而降低了读者对原创新闻的付费意愿[11]。长此以往，高质量的信息生产或许会因收益与成本的倒挂而愈发稀缺。

其次，当前法律法规的滞后性加剧了原创作者与人工智能内容平台权利冲突的复杂性。根据版权法的基本原理，只有具有“独创性”的作品才能获得保护，而生成内容是否满足这一标准存在争议[12]。AI 模型在不同场景下的高效匹配依赖巨量数据集，这些数据集通常源自网络上的文本[13]。然而，许多原始文本的创作者在

---

[8] Bender E. M., Gebru T., McMillan-Major A., Shmitchell S. On the dangers of stochastic parrots: Can language models be too big?[A]. Proceedings of FAccT[C], New York：Association for Computing Machinery, 2021: 610-623.
[9] Martin Kretschmer, Bartolomeo Meletti, Luis H Porangaba. Artificial intelligence and intellectual property: copyright and patents—a response by the CREATe Centre to the UK Intellectual Property Office’s open co nsultation[J]. Journal of Intellectual Property Law & Practice, 2022, 17(3):321-326.
[10] Pickard V. Democracy without journalism? Confronting the misinformation society[M]. 第 1 版. Oxford University Press, 2022.
[11] Chanakya, C.N. AI and the Newsroom: The Impact of Artificial Intelligence on Journalistic Practices and Ethics [J]. Journal of Business, IT, and Social Science, 2022.
[12] Samuelson P. Generative AI meets copyright law[J]. Communications of the ACM, 2023, 66(3):30-33.
[13] El Mestari, S. Z., Lenzini, G., & Demirci, H. Preserving data privacy in machine learning systems[J]. Computers & Security, 2024, 137:103605.

未经告知或补偿的情况下被采集，这种“知识搬运”行为进一步扩大了与版权保护行动的冲突[14]。再次，从融合新闻传播学的视角来看，GEO 的兴起正在挑战传统的公共知识信任体系与信息生态平衡[15]，高质量原创新闻的减少可能破坏媒体的公信力和信息生态的稳健发展。

3.2 算法偏见与信息茧房的加剧

算法偏见的形成与数据偏倚密切相关，这种偏倚不仅源自训练数据中某些特定群体或观点的过度代表，也可能因模型对于多元群体中的表现不均衡而进一步恶化[16]。在特定语境中，偏见同样来源于算法设计过程中的隐形偏见，同质化的开发团队可能因无意识偏好，在算法优化中优先考虑部分群体需求[17]，生成式 AI 可能习惯性地重复训练语料中主流的叙事模式，而隐匿少数群体的声音，这种现象可能被解读为“算法歧视”[18]。

在个人层面，用户对 AI 依赖程度也同样推动信息茧房的加剧。生成式 AI 的目标是生成流畅、连贯且对话上下文相关的回答。这意味着它会根据用户的语境和输入给出尽可能多样化或顺应性的回答。如果用户倾向于某种观点，AI 更有可能避免争议性立场，用相似的话语响应，以确保对话的连贯性和用户体验的满足。当对话不断产生时，生成式 AI 输出的内容可能变得单一化，随着时间推移，用户会逐步丧失探究多元化信息来源的动力[19]。另一方面，在与生成式 AI 的交互中，用户可能逐渐接触到与其已有观点趋同的内容，排斥了其他视角。这种现象在社交平台的算法推荐中早有表现[20]，而 AI 模型可能存在的顺应性特质将进一步加剧该问题。生成式 AI 的开发者通常会为模型制定中立性策略，防止其参与传播错误信息或极端立场。然而，这种中立性在实践中并非完全“可控”。例如，AI 在某些情况下可能会误判用户的语境，从而无意中强化某种主观偏见[21]。

---

[14] Lemley M. A., Casey B. Fair learning[J]. Texas Law Review, 2021, 99(4): 743-786.
[15] Lewis S. C., Westlund O. Big data and journalism: Epistemology, expertise, economics, and ethics[J]. Digital Journalism, 2015, 3(3):447-466.
[16] Noble, S. U. Algorithms of oppression: How search engines reinforce racism[M]. New York: New York University Press, 2018.
[17] Radanliev, P . AI Ethics: Integrating Transparency, Fairness, and Privacy in AI Development[J]. Applied Artificial Intelligence, 2025, 39(1).
[18] Barocas, S., Hardt, M., & Narayanan, A. Fairness in machine learning[R]. NIPS Tutorial, 2017.
[19] Tufekci, Z. Algorithmic harms beyond Facebook and Google: Emergent challenges of computational agency[J]. Journal on Telecommunications & High Technology Law, 2015, 13:203-218.
[20] Pariser, E. The filter bubble: What the Internet is hiding from you[M]. London: Penguin Press, 2011.
[21] Noble, S. U. Algorithms of oppression: How search engines reinforce racism[M]. New York: New York University Press, 2018.

长期来看，过度依赖存在偏见和顺应性的 AI 模型这种机制不仅影响个人层面的信息获取，还可能在社会层面加剧群体之间的认知割裂。值得注意的是，这种依赖也可能对公共领域的舆论空间造成干扰。一些研究发现，生成式内容优化的目标，本质上是追求高参与度或点击量，这可能驱使模型更加倾向于生成易引发争议或极化讨论的内容[22]。其结果是，AI 生成内容可能无意间加剧社会分裂，影响对公共问题的理性讨论。同时，部分学者已经提出，生成式 AI 可能通过优先推荐某些品牌或平台的信息，导致信息资源的垄断与竞争不公平[23]。

3.3 “黑箱”问题与可解释性危机

“黑箱”问题自人工智能出现之时就被广泛讨论，黑箱寓意着 AI 在处理信息和输出过程中的随机性和不可见性，任何人都无法预测某一次程序运行的结果，从连续两次生成的答案不相同即可得知。生成式 AI 模型的“黑箱”问题不仅让内容创作者和用户对其决策过程感到困惑，也让其的公平性和可靠性受到质疑。用户依赖生成式 AI 获取信息时，往往难以辨别生成内容的可信度及其来源，这进一步加剧了错误信息的传播风险[24]。随着开发人员对模型回答的可信度和可解释性的重视，多家 AI 对话助手（如 ChatGPT、Perplexity）在生成答案时，会引用来源链接。即便如此，有实测数据发现，这些引用常常不能完全支撑其答案，一些生成式 AI 应用的引用错误率甚至超过 50%[25]。并且，这些应用仍未明确说明“为何选择这些信源而非其他”，也未披露对不同信源的权重分配逻辑。举例而言，在汽车、消费电子等垂直领域，AI 优先选择“第三方权威媒体”，但未解释“权威”的判定标准（如域名权重、内容可信度评分维度），导致用户与开发者无法追溯信息筛选的底层规则[26]。探索发现生成式 AI 隐匿的筛选规则正是从事 GEO 的研究或是营销人员的首要也是最关键的一步，GEO 的兴起也反向验证了用户心中对 AI 回答公平性和可靠性的质疑。可以说，GEO 的出现使得生成式 AI 的可解释性危机变得难

[22] Tufekci, Z. Algorithmic harms beyond Facebook and Google: Emergent challenges of computational agency[J]. Journal on Telecommunications & High Technology Law, 2015, 13:203-218.
[23] Helberger, N., Karppinen, K., & D'Acunto, L. Exposure diversity as a design principle for recommender systems[J]. Information, Communication & Society, 2019,21(2):191-207.
[24] Diakopoulos, N. Accountability in algorithmic decision-making[J]. Communications of the ACM, 2016, 59(2):56-62.
[25] 夏昊扬，李林杰，宋语阳，等。实测数据告诉你：带引用的 AI 也不可靠[EB/OL]. https://www.thepaper.cn/newsDetail_forward_31251274, 2025-07-29.
[26] Chen M, Wang X, Chen K, et al. Generative Engine Optimization: How to Dominate AI Search[J]. arXiv preprint arXiv:2509.08919, 2025.

以控制。

搜索引擎作为信息流动的中介，一方面它通过搜索算法提供内容，另一方面它也通过广告算法为商业内容提供曝光度。这种双重中介性特征使搜索引擎成为典型的“平台”，既在技术层面优化信息流动，也在经济层面获得广告收入。如何平衡平台化的搜索引擎的中立性和商业性是亟待解决的问题。

Gillespie 的研究指出，生成式 AI 对权威来源的系统性偏好，可能导致信息的集中化，这不但对品牌自有内容及社交内容造成了压制，也可能弱化用户选择信息的自主性，进而形成“算法统治”的风险[27]。这种系统偏好甚至可能在无意中强化既有的权力结构，而非推动信息生态的平等[28]。

更令人担忧的是，操纵一旦有了条件，便不会仅出现在商品选择这一个场景中。这种操作可能不仅在商业领域产生影响，还可能在政治传播中对社会舆论构成威胁[29]。同时，由于生成式 AI 决策的不可解释性，用户很难追溯并质疑其生成的内容，这可能进一步瓦解人机信任关系[30]。

3.4 虚假信息与“幻觉”风险

生成式 AI 有时会产生“幻觉”，即生成看似合理但实际上是虚构或错误的信息。这是因为生成式 AI 的原理是基于大规模数据训练，通过概率分布模型生成语法和语义上合乎逻辑的词汇或句子，而非事实性信息，且其本身不具备事实核查的程序和能力。另外，当超出知识范围时，模型可能为了满足用户体验而通过语言预测虚构信息以“补充空白”。

如果品牌信息被 AI 错误地整合或演绎，可能对其声誉造成损害。AI 生成内容的真实性和准确性问题存在风险，一旦错误信息通过 GEO 传播放大，可能对社会舆论产生深远且负面的影响[31]。当 AI 在缺乏准确验证的情况下生成偏颇或虚假的信息，这种信息可能因其表面的逻辑合理性和语言流畅性而被用户信任。一些研

[27] Gillespie, T. The relevance of algorithms[A]. Crawford, T., Gillespie, H. (Eds.). Media Technologies: Essays on Communication, Materiality, and Society[C]. Cambridge: MIT Press, 2014:167-194.
[28] Pasquale, F. The black box society: The secret algorithms that control money and information[M]. Cambridge: Harvard University Press, 2015.
[29] Tufekci, Z. Algorithmic harms beyond Facebook and Google: Emergent challenges of computational agency[J]. Journal on Telecommunications & High Technology Law, 2015, 13(1):203-218.
[30] Noble, S. U. Algorithms of oppression: How search engines reinforce racism[M]. New York: New York University Press, 2018.
[31] Vosoughi, S., Roy, D., & Aral, S. The spread of true and false news online[J]. Science, 2018, 359(6380):1146-1151.

究指出，AI 生成的虚假信息比传统的信息误导形式更难被察觉，因为其拥有模仿专业语言风格和内容结构的能力，从而提高了认可信度[32]。

GEO 可能进一步加剧虚假信息的生成与传播。恶意攻击者利用 GEO 策略操纵关键词和内容架构，使虚假的生成信息被 AI 优先推荐并放大[33]。这一现象尤其在实时热点事件中表现得更为明显，例如在政治竞选时对候选人的负面信息操纵，或者在突发事件中散布虚假引导性叙事[34]。虚假信息如果通过算法传播，其扩散速度往往超过真实信息，并且具有更广泛的传播影响[35]。

这一问题将来带对用户的误导、对品牌形象的威胁和对公众舆论的操纵。由于缺乏识别 AI 结果的能力，普通用户可能会误信 AI 生成的虚假或错误内容，尤其是在专业领域（如医学、法律、金融）中，这种误导后果更为严重[36]。当 AI 幻觉将虚假或误导信息与品牌结合，它可能削弱公众对品牌的信任。如果品牌被错误地关联到某些负面事件或错误事实，即使后续进行澄清，也可能难以完全弥补声誉损失[37]。另一方面，新闻行业也面临生成式 AI 所带来的精准操纵风险。虚假新闻的生成和传播不仅破坏了新闻的公信力，还可能通过放大偏见、制造对立情绪等方式，影响公众舆论走向[38]。

4GEO 的理论重构：多学科视角融合

理解与指导 GEO 实践，需要我们跳出传统传播学的单一框架，积极吸纳认知心理学、信息科学、伦理学、社会学等多元学科的理论成果，并结合生成式 AI 特性进行创新性转化。结合认知心理学的用户行为研究，可以分析生成式 AI 如何通过内容精准推送影响用户认知过程[39]；结合社会学理论，可以进一步探究 AI 生成

---

[32] Brown, T., Mann, B., Ryder, N., et al. Language Models are Few-Shot Learners[J]. Advances in Neural Information Processing Systems, 2020, 33:1877–1901.

[33] Rini, R. Deepfakes and the Epistemic Backstop[J]. Philosophy & Technology, 2021, 34(2):349–371.

[34] Ferrara, E., Cresci, S., & Luceri, L. Misinformation, manipulation, and abuse on social media[J]. ACM SIGKDD Explorations Newsletter, 2020, 22(2):1–9.

[35] Vosoughi, S., Roy, D., & Aral, S. The spread of true and false news online[J]. Science, 2018, 359(6380):1146-1151.

[36] Siau, K., & Wang, W. Building trust in artificial intelligence, machine learning, and robotics[J]. ACM Transactions on Management Information Systems (TMIS), 2018, 9(3):7.

[37] Wardle, C., & Derakhshan, H. Information disorder: Toward an interdisciplinary framework for research and policy making[R]. Strasbourg: Council of Europe, 2017.

[38] Zhou, X., & Zafarani, R. Fake news: A survey of research, detection methods, and opportunities[J]. ACM Computing Surveys (CSUR), 2020, 53(5):1–40.

[39] Sundar, S. S. The MAIN model: A heuristic approach to understanding technology effects on credibility[A]. Metzger, M. J., & Flanagin, A. J. (Eds.). Digital Media, Youth, and Credibility[C]. Cambridge: MIT Press, 2008:73-100.

信息对社会关系网络与舆论生成机制的重构[40]。更重要的是，对于信息国际化传播的研究，生成式 AI 作为多语言生成工具，为全球化传播提供了新可能，但同时也加剧了文化语境与思想基础的单一化[41]。这些视角的融合将推动传播学从文本传播研究向交互性、生成性信息传播研究的深度拓展。本文在此主要探讨对于“把关人理论”的拓展和对于认知心理学理论的结合。

4.1 传播学理论的再审视与扩展

生成式 AI 俨然已经变身成为“超级把关人”，拥有传统媒体和搜索引擎把关人所不具备的多重职能。它不仅过滤信息，还主动生产内容，从而打破了信息筛选与内容生成的传统二元划分。这种角色转变对信息传播环境带来了深远的影响。首先，从信息权力的角度看，生成式 AI 具有更强的权力集中性。其算法训练依赖海量数据，而这些数据的选择、使用以及逻辑偏向，决定了 AI 把关的信息流向及优先级[42]。更重要的是，AI 的生成能力使其不再是简单的“信息筛选者”，而成为信息的“建构者”，直接塑造用户的认知与观念。而这一过程中的隐性偏见可能加剧信息不对称或形成新的权力失衡格局[43]。

其次，从传播过程看，生成式 AI 形成了以用户需求为核心的反馈循环。当用户输入特定的信息需求后，生成式 AI 不仅在已有数据中寻找答案，还可能根据用户的语境和立场生成新的内容，以满足“精准响应”目标。这种以数据驱动的生产与分发模式，使得经典的把关理论（Gatekeeping Theory）可以扩展为包含生成特征的“生成式把关理论”（Generative Gatekeeping Theory）。新理论需要关注的，不仅是信息筛选过程中的标准与逻辑，还需探讨生成过程中偏见放大的风险和生成信息对认知的潜在影响[44]。

此外，生成式 AI 的传播机制也揭示了把关权力的多层结构。从表面看，生成式 AI 作为“技术性把关人”依赖算法进行操作，带有中立属性；但实际上，技术中的隐性设计（如模型参数、训练数据选择）背后隐藏着开发者的价值观与商业目

[40] Couldry, N., & Hepp, A. The mediated construction of reality[M]. Cambridge: Polity Press, 2017.
[41] Finn, E., et al. What algorithms want: Imagination in the age of computing[M]. Cambridge: MIT Press, 2018.
[42] Gillespie, T. The relevance of algorithms[A]. Crawford, T., Gillespie, H. (Eds.). Media Technologies: Essays on Communication, Materiality, and Society[C]. Cambridge: MIT Press, 2014:167-194.
[43] Noble, S. U. Algorithms of oppression: How search engines reinforce racism[M]. New York: New York University Press, 2018.
[44] Shoemaker, P. J., & Vos, T. Gatekeeping Theory[M]. London: Routledge, 2009.

标[45]。例如，生成式 AI 可能优先显示与商业利益或主流意识形态相关的内容，从而影响信息的多样性与用户体验。这种“算法隐形价值”的作用可能在无意中改变新闻传播领域的信息平等原则，进一步重塑公共舆论的生态结构[46]。

在生成式 AI 逐渐抢占主导地位之一的信息传播时代，把关人理论的延伸还需结合技术可解释性进行深入探讨。研究者提出，“算法透明性”已成为 AI 技术发展的重要伦理议题之一[47]。透明化的生成过程不仅能够提升用户对内容来源与可信度的了解，也有助于减轻生成式 AI 对信息流造成的偏见和误导。然而，透明性本身也会带来新的困境，即生成式 AI 背后的复杂算法往往难以向普通受众清晰解释，可能导致信息不对称与权力疏离的加剧[48]。

4.2 认知心理学理论的融入：理解人机互动与决策

信任与信念形成理论（Trust and Belief Formation Theory）是人机互动研究中的核心议题，它关注用户如何在非人类实体中构建信任，以及这种信任的动态变化[49]。在生成式 AI 的应用情境中，用户对 AI 生成内容的信任建立往往与其对技术能力、输出内容的准确性和外部交互体验相关联[50]。特别是在 GEO 出现后，品牌的信任塑造不仅依赖 AI 生成的信息是否符合用户期待，还需要通过可持续的互动强化用户的信念系统，即在反复接触中形成对品牌稳定性、可靠性的认知联想[51]。

在生成式 AI 中，信任的建立与信息的透明度密切相关。此前的研究表明，当用户能够清晰理解 AI 的工作原理以及生成内容的来源时，信任形成的可能性大幅提高[52]。然而，这种信任的维持是动态且易受挑战的。当 AI 被发现存在“幻觉”或生成偏颇信息时，用户对 AI 的信任可能迅速崩塌，并转化为对技术系统和品牌的

---

[45] Pasquale, F. The Black Box Society: The Secret Algorithms That Control Money and Information[M]. Cambridge: Harvard University Press, 2015.
[46] Zuboff, S. The Age of Surveillance Capitalism: The fight for a human future at the new frontier of power[M]. New York: PublicAffairs, 2019.
[47] Kroll, J., Huey, J., Barocas, S., et al. Accountable algorithms[J]. ACM Transactions on Information and System Security, 2017, 21(1):1-37.
[48] Lipton Z C. The mythos of model interpretability: In machine learning, the concept of interpretability is both important and slippery[J]. Queue, 2018, 16(3): 31-57.
[49] Hoff, K. A., & Bashir, M. Trust in automation: Integrating empirical evidence on factors that influence trust[J]. Human Factors, 2015, 57(3):407-434.
[50] Sundar S ,Shyam.The MAIN Model: A Heuristic Approach to Understanding Technology Effects on Credibility[J]. 2008.
[51] Lewicki, R. J., McAllister, D. J., & Bies, R. J. Trust and distrust: New relationships and realities[J]. Academy of Management Review, 1998, 23(3):438-458.
[52] Siau, K., & Wang, W. Y. Building Trust in Artificial Intelligence, Machine Learning, and Robotics[J]. Cutter Business Technology Journal, 2018, 31:47-53.

不信任[53]。这种信任破裂不仅破坏了人与 AI 间的交互关系，还可能对信息的传播和品牌声誉形成负面的长期影响[54]。

为了在生成式 AI 的语境下重塑信任，品牌及技术设计者需要采取多维度策略。其一，增强可解释性是重建信任的重要途径。通过设计透明的内容生成过程，让用户了解 AI 生成信息的逻辑框架及语料来源，这可以有效缓解因“幻觉”或偏见问题导致的不安全感[55]。其二，品牌需要主动优化与用户的交互体验，提供替代性验证渠道帮助用户检测生成内容的准确性，从而降低误解与不信任的风险[56]。此外，研究还表明，通过引入情感设计，例如在用户体验中融合情感化交互元素，可以使 AI 更好地与用户建立信任关系，为品牌的长远策略打下基础[57]。

更进一步，认知心理学对信任系统的研究也揭示了人类在面对技术实体时的“认知偏误”。“科技赋能偏见”表明用户往往倾向于信任由技术生成的内容，而忽略其中可能的错误及不完整性[58]。在 GEO 实践中，这种偏误可能让用户高度依赖 AI 生成的品牌内容，而降低对真实品牌信息的独立判断能力。因此，品牌在应用生成式 AI 时，除了致力于建立信任，还需投入资源提高用户的“认知抗性”，帮助他们识别内容偏见，增强媒介素养[59]。

从传播学的视角来看，信任不仅是个体性认知的产物，也是一个社会化互动过程中不断构建的动态关系。在生成式 AI 的应用语境中，信任关系的维系与破裂发生于两个层面：个体用户对内容的初步感知，以及通过群体讨论或平台反馈等机制形成的社会化共识[60]。这种双重信任机制提示我们，GEO 实践不仅要在技术设计上优化用户体验，还需结合传播机制促进信息的多元呈现与真实验证，从而构建持续的品牌信任基础。

---

[53] Kadre, P., & Dam, F. AI hallucination and trust in conversational agents: The new cognitive dissonance[J]. Journal of Human-Computer Interaction, 2020, 12(2):89-104.
[54] Mayer, R., Davis, J., & Schoorman, F. An integrative model of organizational trust[J]. Academy of Management Review, 1995, 20(3):709-734.
[55] Lipton, Z. C. The mythos of model interpretability: In machine learning, the concept of interpretability is both important and slippery[J]. Queue, 2018, 16(3):31-57.
[56] Shin, D. H. The role of explainability and causability in trust formation of AI-based systems[J]. Social Science Computing Review, 2020, 38(4):448-468.
[57] Picard, R. W. Affective computing[M]. Cambridge: MIT Press, 1997.
[58] Paecht, N. P. Automation bias in intelligent technology[J]. Digital Societies, 2004, 3(5):112-130.
[59] Schmidt, A., et al. Media literacy education in the era of computational propaganda: Design, challenges, and opportunities[J]. Computers in Education, 2019, 13(4):97-115.
[60] Fogg, B. J. Persuasive technology: Using computers to change what we think and do[M]. Burlington: Morgan Kaufmann, 2003.

5GEO 的未来展望与构建负责任的智能传播

GEO 的未来是充满无限可能的，它将不仅仅是营销工具，更可能成为构建未来信息生态系统的核心要素。但要实现这一愿景，必须在技术发展的同时，高度重视伦理、原创性与社会责任问题。

5.1 技术与人文的深度融合：人机共创与伦理约束

未来的 GEO 发展将是 AI 与人类智能的深度融合。在这一过程中，AI 和人类智能的角色分工愈发明确。AI 通过其强大的数据处理、模式识别和效率优化技术充当知识生产的工具，承担内容生成、信息筛选等任务，而人类则在策略制定、情感表达、创意构思及历史与文化价值传递中塑造更深层次的指导作用。这种人机共创模式结合了技术的效率与人类智慧的不可替代性，为新闻传播、文化生产及社会治理领域带来了巨大的创新潜力[61]。

值得注意的是，人机共创不仅仅是技术和效率的组合，还涉及复杂的伦理约束和价值选择。在未来的 GEO 发展中，如何通过技术规范和道德框架限制 AI 行为，将是保障社会利益的一个关键问题。[62]指出，在人机共创中，伦理约束的核心关注点是保护公众权益和社会价值，例如防止内容生成的歧视性偏见、不透明决策和虚假信息等问题。伦理约束应以三方面为基础：技术透明性、决策可解释性和责任归属。通过设计公平、公正的技术治理框架，可以减少 AI 在应用过程中可能出现的伦理失当。

在技术实践层面，人机共创的伦理约束可以通过嵌入“伦理设计”的模式来实现。例如，Samuelson 认为，生成式 AI 在与人类智能共创时，需保证原创内容的产权保护和创作者的合理权益归属。这可以通过强制性的归因标识，确保 AI 生成内容中引用的原始数据和创意来源能够被追踪[63]。这种机制不仅提升了技术的透明度，也有效回应了日益复杂的知识产权挑战。

---

[61] Bender E. M., Gebru T., McMillan-Major A., Shmitchell S. On the dangers of stochastic parrots: Can language models be too big?[A]. Proceedings of FAccT[C], New York：Association for Computing Machinery, 2021:610-623.
[62] Helberger N., Karppinen K., D'Acunto L. Exposure diversity as a design principle for recommender systems[J]. Information, Communication & Society, 2020, 21(2): 191–207.
[63] Samuelson P. Generative AI meets copyright law[J]. Communications of the ACM, 2023, 66(3): 30-33.

5.2 透明化、可解释性与归因机制的建立

为应对“黑箱”问题和知识产权挑战，未来的 GEO 系统和 AI 模型将更加注重透明化和可解释性。用户和内容生产者将有权了解 AI 生成答案的来源、逻辑和权重，并建立明确的归因机制，确保原创内容得到应有的尊重和回报。

首先，从技术层面推动专门的可解释 AI 算法（Explainable AI, XAI）的研发尤为重要。这些算法可帮助用户了解生成式 AI 内容的来源和决策逻辑，从而提升其可信度。学者 Lipton 提出通过局部可解释模型（如 LIME）提供交互式数据溯源功能，以揭示 AI 决策过程中使用的关键信息[64]。这种技术不仅具有科学价值，还能够让公众更加直观地理解 AI 内容，提高技术的接受度。此外，建立“事实核查层”也是 AI 透明化的重要保障。通过事实验证工具与实时外部数据库的对接，生成式 AI 可以在输出内容前验证信息的真实性，从而降低虚假信息的传播[65]。

知识产权保护也是生成式 AI 发展中不可回避的重要议题。Handke 等学者提出，数据许可和版税协议的改进可以构建更加公平的补偿机制，让原创作者从 AI 内容的使用中受益[66]。针对 AI 生成内容实施透明度机制，有学者建议强制标识生成内容与原创内容的区别[67]。

5.3 内容生态的去中心化与多元化

虽然 AI 倾向于整合信息，但为了避免“信息茧房”和单一化，未来的信息生态需要鼓励多元化和去中心化的内容生产。GEO 策略应同时关注如何在 AI 的整合中保持内容的独立性和多样性，甚至可能出现新的、基于 AI 的“长尾内容”发现机制。Anderson 的长尾理论提出，当生产和分销成本降低后，过去被忽略的非主流和小众内容可以获得相比传统主流内容更大的累计市场[68]，这一观点在新闻传播领域具有重要启示。通过 AI 驱动的个性化推荐和内容发现技术，GEO 可以挖掘长尾内

[64] Lipton Z C. The mythos of model interpretability: In machine learning, the concept of interpretability is both important and slippery[J]. Queue, 2018, 16(3): 31-57.
[65] horne, J., Vlachos, A., Christodoulopoulos, C., & Mittal, A. FEVER: a Large-scale Dataset for Fact Extraction and VERification [A]. Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)[C]. New Orleans: Association for Computational Linguistics, 2018:809–819.
[66] Handke, C., Guibault, L., & Vallbé, J. J. Is Europe falling behind in data mining? Copyright’s impact on data research in academia and industry[J]. Journal of the Association for Information Science and Technology, 2020, 71(9):1087-1099.
[67] Gutierrez, C., Stewart, M., & Knight, J. AI transparency and attribution in global information systems[J]. Journal of Information Ethics, 2022, 28(2):123-135.
[68] Anderson C. The long tail: Why the future of business is selling less of more[M]. Hyperion, 2006.

容的潜力，使用户接触主流信息之外的多样化和小众化内容，从而打破“信息茧房”的困局。这不仅能够丰富新闻内容的多样性，也可以激励更多原创者参与内容生产，为传播生态注入新的活力。

此外，去中心化的信息生态能够帮助缓解对少数平台的大规模依赖，并通过开放 API、内容联动社区和去中心化的分发算法增强小型创作者的表达能力[69]。这种机制的社会意义体现在它对话语权分配的调整，能够为边缘化群体提供更多曝光机会，让信息传播更加公平和包容。

5.4 提升公民的 AI 素养与批判性思维

面对 GEO 带来的机遇与挑战，教育用户如何与 AI 有效互动、识别 AI 生成的虚假信息、理解算法逻辑、保护个人数据，将有助于构建一个更为成熟和健康的智能传播环境。为了缓解生成式 AI 的幻觉和虚假信息风险，首先是加强 AI 内容生成过程中的真实性验证机制，其次就是提高公众对生成式 AI 内容的认知素养，让用户能够对 AI 生成内容的准确性进行批判性分析，而不盲目信任其语言表达的表面可信性[70]。公众在面对 AI 生成内容时，往往容易判断失误或错误解读，部分原因在于用户对 AI 生成过程的理解有限[71]。因此，教育用户如何与 AI 有效互动，理解语言生成背后的逻辑及算法运行机制，将有助于培养他们对内容准确性的批判性意识。公众不仅需要了解 AI 生成内容可能存在偏差或错误，更需要具备正确的内容甄别能力。用户应主动查证内容的来源，而不是对生成式 AI 的语言表达盲目信任。

6 结论

生成式人工智能技术给新闻传播学及相关领域带来了深远影响，其核心能力正在重塑信息检索、内容分发与用户交互模式。本研究着眼于 GEO 的内涵、技术逻辑以及实践应用，明确了其作为一种新模式的独特作用，同时揭示了其带来的多维挑战。研究表明，GEO 通过优化内容被生成式 AI 采纳的概率、展现效果及可

[69] Pickard, V. Democracy without Journalism? Confronting the Misinformation Society[M/OL]. New York: Oxford University Press, 2020 (19 Dec. 2019 online edn); https://doi.org/10.1093/oso/9780190946753.001.0001, 2025-09-04 (accessed).
[70] Lazer D., Baum M., Benkler Y., et al. The science of fake news[J]. Science, 2018, 359(6380): 1094–1096.
[71] Shin D. H. The effects of explainability and causability on user acceptance of AI[J]. Journal of Human-Computer Interaction, 2020, 35(6): 486-491.

信度，推动了品牌传播与个性化信息服务的全面升级。然而，这种优化实践也暴露了诸多风险：例如，内容原创性无法得到充分保证，算法偏见可能加剧信息单一化，虚假信息通过自动化分发迅速扩大其影响力，AI 模型的“黑箱”特性进一步削弱了内容生成与分发过程的透明性。

针对上述问题，本文从多学科角度提出了以下解决思路：一是加强 GEO 的伦理设计与技术规制，确保内容生成和分发过程的公平性与透明性；二是推动分布式内容生产环境，减少信息茧房化及平台垄断的风险；三是建立人机共创机制，挖掘技术与人文的深度融合潜力；四十引入系统化的虚假信息监测和应对机制，提升用户对生成内容的甄别能力；五是提高公众的 AI 素养与批判性思维，为负责任的智能传播生态奠定用户基础。

尽管本文提供了关于 GEO 风险与信息生态重塑的重要见解，但依然存在一些局限性。首先，案例分析主要基于理论推演，缺少大规模实证数据的支持；其次，研究未能深入讨论不同文化、地域对 GEO 实践适配性的差异。未来研究可结合多国数据，对比 GEO 在不同传播环境中的表现与挑战，为全球性传播政策的制定提供决策依据。此外，还应进一步探索算法治理的技术路径，兼顾创新性与伦理责任，推动生成式人工智能在信息生态系统中的可持续发展。

参考文献

[1] Bakshy, E., Messing, S., & Adamic, L. A. Exposure to ideologically diverse news and opinion on Facebook[J]. Science, 2015, 348(6239):1130-1132.

[2] Brown, T., Mann, B., Ryder, N., et al. Language Models are Few-Shot Learners[J]. Advances in Neural Information Processing Systems, 2020, 33:1877–1901.

[3] Chanakya, C.N. AI and the Newsroom: The Impact of Artificial Intelligence on Journalistic Practices and Ethics[J]. Journal of Business, IT, and Social Science, 2022.

[4] Diakopoulos, N. Accountability in algorithmic decision-making[J]. Communications of the ACM, 2016, 59(2):56-62.

[5] El Mestari, S. Z., Lenzini, G., & Demirci, H. Preserving data privacy in machine learning systems[J]. Computers & Security, 2024, 137:103605.

[6] Ferrara, E., Cresci, S., & Luceri, L. Misinformation, manipulation, and abuse on social media[J]. ACM SIGKDD Explorations Newsletter, 2020, 22(2):1–9.

[7] Gutierrez, C., Stewart, M., & Knight, J. AI transparency and attribution in global information systems[J]. Journal of Information Ethics, 2022, 28(2):123-135.

[8] Handke, C., Guibault, L., & Vallbé, J. J. Is Europe falling behind in data mining? Copyright's impact on data research in academia and industry[J]. Journal of the Association for Information Science and Technology, 2020, 71(9):1087-1099.

[9] Helberger, N., Karppinen, K., & D'Acunto, L. Exposure diversity as a design principle for recommender systems[J]. Information, Communication & Society, 2019, 21(2):191-207.

[10] Hoff, K. A., & Bashir, M. Trust in automation: Integrating empirical evidence on factors that influence trust[J]. Human Factors, 2015, 57(3):407-434.

[11] Kadre, P., & Dam, F. AI hallucination and trust in conversational agents: The new cognitive dissonance[J]. Journal of Human-Computer Interaction, 2020, 12(2):89-104.

[12] Kretschmer, M., Meletti, B., Porangaba, L. H. Artificial intelligence and intellectual property: copyright and patents—a response by the CREATe Centre to the UK Intellectual Property Office's open consultation[J]. Journal of Intellectual Property Law & Practice, 2022, 17(3):321-326.

[13] Lazer, D., Baum, M., Benkler, Y., et al. The science of fake news[J]. Science, 2018, 359(6380):1094–1096.

[14] Lemley, M. A., Casey, B. Fair learning[J]. Texas Law Review, 2021, 99(4):743-786.

[15] Lewis, S. C., Westlund, O. Big data and journalism: Epistemology, expertise, economics, and ethics[J]. Digital Journalism, 2015, 3(3):447-466.

[16] Lipton, Z. C. The mythos of model interpretability: In machine learning, the concept of interpretability is both important and slippery[J]. Queue, 2018, 16(3):31-57.

[17] Noble, S. U. Algorithms of oppression: How search engines reinforce racism[J]. Journal of Information Ethics, 2018.

[18] Park, P. S., Goldstein, S., O'Gara, A., Chen, M., & Hendrycks, D. AI deception: A survey of examples, risks, and potential solutions[J]. Patterns, 2024, 5(5).

[19] Radanliev, P. AI Ethics: Integrating Transparency, Fairness, and Privacy in AI Development[J]. Applied Artificial Intelligence, 2025, 39(1).

[20] Rini, R. Deepfakes and the Epistemic Backstop[J]. Philosophy & Technology, 2021, 34(2):349–371.

[21] Samuelson, P. Generative AI meets copyright law[J]. Communications of the ACM, 2023, 66(3):30-33.

[22] Shin, D. H. The effects of explainability and causability on user acceptance of AI[J]. Journal of Human-Computer Interaction, 2020, 35(6):486-491.

[23] Shin, D. H. The role of explainability and causability in trust formation of AI-based systems[J]. Social Science Computing Review, 2020, 38(4):448-468.

[24] Siau, K., & Wang, W. Building trust in artificial intelligence, machine learning, and robotics[J]. ACM Transactions on Management Information Systems (TMIS), 2018, 9(3):7.

[25] Siau, K., & Wang, W. Y. Building Trust in Artificial Intelligence, Machine Learning, and Robotics[J]. Cutter Business Technology Journal, 2018, 31:47-53.

[26] Tufekci, Z. Algorithmic harms beyond Facebook and Google: Emergent challenges of computational agency[J]. Journal on Telecommunications & High Technology Law, 2015, 13:203-218.

[27] Vosoughi, S., Roy, D., & Aral, S. The spread of true and false news online[J]. Science, 2018, 359(6380):1146-1151.

[28] Zhou, X., & Zafarani, R. Fake news: A survey of research, detection methods, and opportunities[J]. ACM Computing Surveys (CSUR), 2020, 53(5):1–40.

[29] Bender, E. M., Gebru, T., McMillan-Major, A., Shmitchell, S. On the dangers of stochastic parrots: Can language models be too big?[A]. In Proceedings of FAccT[C]. New York: Association for Computing Machinery, 2021:610-623.

[30] Binns, R. Fairness in machine learning: Lessons from political philosophy[C]. Proceedings of the 2018 Conference on Fairness, Accountability, and Transparency, 2018:149-159.

[31] Binns, R. Fairness in machine learning: Lessons from political philosophy[A]. Crawford, T., Gillespie, H. (Eds.). Media Technologies: Essays on Communication, Materiality, and Society[C]. Cambridge: MIT Press, 2018:149-159.

[32] Gillespie, T. The relevance of algorithms[A]. Crawford, T., Gillespie, H. (Eds.).

Media Technologies: Essays on Communication, Materiality, and Society[C]. Cambridge: MIT Press, 2014:167-194.

[33] Thorne, J., Vlachos, A., Christodoulopoulos, C., & Mittal, A. FEVER: A large-scale dataset for fact extraction and verification[A]. In Proceedings of NAACL-HLT[C], 2018:809–819.

[34] Anderson, C. The long tail: Why the future of business is selling less of more[M]. Hyperion, 2006.

[35] Couldry, N., & Hepp, A. The mediated construction of reality[M]. Cambridge: Polity Press, 2017.

[36] Finn, E., et al. What algorithms want: Imagination in the age of computing[M]. Cambridge: MIT Press, 2018.

[37] Fogg, B. J. Persuasive technology: Using computers to change what we think and do[M]. Burlington: Morgan Kaufmann, 2003.

[38] Noble, S. U. Algorithms of oppression: How search engines reinforce racism[M]. New York: New York University Press, 2018.

[39] Pariser, E. The filter bubble: What the Internet is hiding from you[M]. London: Penguin Press, 2011.

[40] Pasquale, F. The black box society: The secret algorithms that control money and information[M]. Cambridge: Harvard University Press, 2015.

[41] Picard, R. W. Affective computing[M]. Cambridge: MIT Press, 1997.

[42] Pickard, V. Democracy without Journalism? Confronting the Misinformation Society[M]. Oxford University Press, 2022.

[43] Shoemaker, P. J., & Vos, T. Gatekeeping Theory[M]. London: Routledge, 2009.

[44] Sunstein, C. R. Republic.com[M]. Princeton: Princeton University Press, 2001.

[45] Tufekci, Z. Twitter and tear gas: The power and fragility of networked protest[M]. New Haven: Yale University Press, 2018.

[46] Zuboff, S. The Age of Surveillance Capitalism: The fight for a human future at the new frontier of power[M]. New York: PublicAffairs, 2019.

[47] Barocas, S., Hardt, M., & Narayanan, A. Fairness in machine learning[R]. NIPS Tutorial, 2017.

[48] Wardle, C., & Derakhshan, H. Information disorder: Toward an interdisciplinary framework for research and policy making[R]. Strasbourg: Council of Europe, 2017.
[49] 夏昊扬，李林杰，宋语阳，等。实测数据告诉你：带引用的 AI 也不可靠[EB/OL]. https://www.thepaper.cn/newsDetail_forward_31251274, 2025-07-29.
[50] Pickard, V. Democracy without Journalism? Confronting the Misinformation Society[M/OL]. New York: Oxford University Press, 2020 (19 Dec. 2019 online edn); https://doi.org/10.1093/oso/9780190946753.001.0001, 2025-09-04 (accessed).