arXiv:2601.01750v1 [cs.IR] 5 Jan 2026

当关注度在生成式搜索中转变为曝光

Shayan Alipour[1,2], Mehdi Kargar[1], 和 Morteza Zihayat[1]

[1] 多伦多都会大学
{shayan.alipour,kargar,mzihayat}@torontomu.ca[2] 罗马智慧大学

摘要。 生成式搜索引擎正通过以综合生成的答案和引文取代传统排序列表，重塑信息获取。与此同时，随着 Web3 平台的发展，激励驱动型创作者生态系统已成为企业通过奖励创作者参与共同叙事来建立可见度和社群的重要组成部分。然而，生成式搜索引擎引文中的曝光在多大程度上受到外部注意力市场影响，仍不确定。本研究对 44 家 Web3 企业的曝光情况进行了审查。首先，我们发现围绕各家企业的创作者社区会随时间保持稳定。其次，企业专属查询显示，较受欢迎的声音系统性地获得比其他声音更多的引用曝光。第三，我们发现，粉丝基础较大、创作者核心群体更集中的企业，往往获得排名更靠前的曝光。总而言之，这些结果表明，生成式搜索引擎的引文存在曝光偏差，倾向于已经 prominent voices 的声音，这可能会巩固既有优势者的地位并缩窄观点多样性。

关键词：生成式搜索 · 曝光偏差 · 注意力 生成式搜索 · 曝光偏差 · 注意力 经济

1 引言

生成式搜索引擎（GSEs）正在重塑信息获取方式，使其从基于排序列表的链接转向带有引文面板的综合生成答案 [5,14,20]。不同于传统的搜索引擎结果页（SERPs）——曝光取决于排名位置——在生成式搜索引擎中，可见度取决于哪些参考资料被引用，以及它们在不同回答中出现的频率 [11]。因此，了解哪些来源获得引用，以及曝光如何在各来源之间分配，对于评估这一新检索范式中的可信度与公平性至关重要 [8,2]。以往研究通过将多样化形式化以遏制集中现象 [7,3,17]，并建立分析框架来探讨如何将作为资源的曝光更公平地分配给不同项目和群体 [19,6,10,12]，从而广泛研究了信息检索系统中的选择性曝光。基于这一基础，早期审计表明，生成式搜索引擎表现出寻求一致的倾向，可能会将曝光收窄至受偏好的观点 [18,21]。与此同时，更广泛的数字环境已发展为注意力经济，在其中，可见度本身作为一种货币发挥作用。这种经济通过内容创作者平台和奖励互动的算法排序系统得到进一步强化 [15,16]。Web3 社区——广义而言，即建立在区块链协议之上的去中心化生态系统，其中

所有权和参与通常以代币形式化——代表了一个尤其组织化且可衡量的例子。在这一环境中，开发去中心化协议或应用程序的组织（本研究称为企业 ）依赖由激励型创作者组成的网络来推广更新内容和叙事。近期研究探讨了创作者经济可以采用的新策略，以增加曝光 [2,22]。这些趋势引发了人们对曝光偏差的担忧：生成式搜索引擎会呈现相同的热门声音，从而进一步强化其地位。

在本研究中，我们审查注意力经济如何影响生成式搜索中的多样性与曝光。我们衡量**研究问题1**：Web3 创作者生态系统的集中程度如何；**RQ2:**生成式搜索引擎在多大程度上引用这些创作者，以及外部影响力是否会提升引用曝光；以及**RQ3:** 哪些因素能够解释更高的引用可见度。研究发现表明，创作者社区稳定且呈层级结构，知名创作者获得的引用曝光显著更高，且曝光与粉丝数及内部集中程度相关。我们还发现，生成式搜索引擎的引文面板因访问方式而异：与用户界面（UI）相比，API 调用返回的面板始终更少、内容也更有限。总之，这些研究发现表明，生成式搜索引擎会放大既有的注意力层级，并由此引发有关观点多样性、公平获取信息以及负责任的生成式检索系统设计等方面的思考。我们提供经过整理的查询语料库、创作者排行榜、包含引文面板的生成式搜索引擎响应，以及支持性分析材料，供社区进一步使用。[3]

2 方法论

a) 数据与实验背景。 我们分析了 44 家 Web3 企业，它们代表了主要协议、基础设施提供商和去中心化应用。每家企业都在 X（原 Twitter）上拥有活跃的创作者生态系统，创作者在此推广更新、讨论去中心化技术相关议题并传播相关叙事。为衡量创作者的影响力，我们使用公开的排行榜数据，这些数据来自 $KaitoAI$[4]，，根据专有的 心智份额 评分对贡献者进行排名；该评分衡量内容质量、互动真实性以及创作者在各自企业社区中的活跃一致性。我们从两个相隔三周的排行榜快照（2025-09-28 和 2025-10-18）中收集了每家企业排名前 100 的创作者。这些快照有助于评估创作者社区的时间持续性和排名稳定性（独立创作者$=3$，232）。为评估生成式搜索引擎中的可见度，我们为每家企业设计了三个查询（两个通用查询，一个时效性查询），共得到 132 个独立查询。通用查询关注企业的宗旨或核心技术，时效性查询则涉及过去 10 天内通过官方渠道发布的近期公告。

b) 审计设计。我们使用三条企业专属查询，并考察四种生成式搜索引擎：启用 Live Search 的 Grok-4（Grok）、启用网页搜索的 GPT-5（GPT）、Gemini 2.5 Flash 搭配 Googlegrounding（Gemini），以及启用网页搜索的 Perplexity Sonar- Pro（Perplexity）。在对 GPT、Gemini 和 Per-

[3] https://github.com/shayanalipour/attention_becomes_曝光 [4] https://yaps.kaito.ai

plexity 进行试运行时，我们观察到 API 与用户界面输出之间存在差异，API 调用返回的引文面板有限（社区已注意到此问题[5]）。为量化其影响，我们单独对 Perplexity 进行审计，比较相同查询下用户界面与 API 的引文面板。最后，为单独评估创作者的曝光情况，我们考察了另一种设置，将 Grok 的来源限制为 X 平台的内容。

c) 曝光量化指标。参照推荐系统中的头部/长尾分析 [1], ，我们在企业层级比较排名靠前与排名靠后的创作者所获得的曝光。对于一家拥有 N_e 名排名创作者（排名数值越小越好）的企业 e，将前十名相对于后半部分的曝光优势定义为

$$\Delta_e = \frac{1}{10}\sum_{r=1}^{10} \mathrm{share}_{e,r} \; - \; \frac{1}{N_e - 50}\sum_{r=51}^{N_e} \mathrm{share}_{e,r},$$

其中，$\mathrm{share}_{e,\,r}$ 表示企业 e 的所有引文中归于排名为 r的创作者的比例。 Δ_e 为正值表示，前十名创作者的单排名份额高于后半部分的创作者。我们通过估算各企业的平均 $\mathbb{E}[\Delta_e]$，并采用企业层级的自助重采样（$B = 1,\ 000$）计算其 95% 置信区间，从而汇总各企业的结果。

为衡量与随机基线的偏差，我们采用与位置无关的归一化累积增益（NCG）[13]，该指标根据外部排名赋予分级得分，并将观察到的引文面板与企业特定的期望值进行比较。对于企业 e，每个查询 q 都会产生一个无序的引文面板$S_{e,\,q}$，其大小为$m_{e,\,q} = |S_{e,\,q}|$。我们在式（1）中定义分级增益 g（r）。面板的观察累积增益见式（2），其中 r_e（i）是创作者 i在企业 e内的外部排名。接着，我们通过极端情况进行归一化：所有被引项目均为前十名创作者（式（3））；该情况给出了任意引文面板的上界。据此，我们计算 NCG，如式（4）所示。

$$g(r) = \begin{cases} 3, & r \le 10 \\ 2, & 11 \le r \le 50 \\ 1, & 51 \le r \le 100 \\ 0, & r > 100 \end{cases} \qquad (1)$$

$$G_{e,q}^{\mathrm{obs}} = \sum_{i \in S_{e,q}} g\big(r_e(i)\big), \qquad (2)$$

$$G_{e,q}^{\star} = 3\, m_{e,q}, \qquad (3)$$

$$\mathrm{NCG}_{e,q}^{\mathrm{obs}} = \frac{G_{e,q}^{\mathrm{obs}}}{G_{e,q}^{\star}} = \frac{\sum_{i \in S_{e,q}} g\big(r_e(i)\big)}{3\, m_{e,q}} \in [0,1]. \qquad (4)$$

为量化超出随机机会的放大效应，我们将 $\mathrm{NCG}^{\mathrm{obs}}$ 与从规模为 N_e 的企业候选集中均匀随机抽取引文面板时的期望值进行比较。设 $\bar{g}_e = \frac{1}{N_e}\sum_{r=1}^{N_e} g(r)$ 和 $g_{\max} = 3$。由于在 $m_{e、\,q}$ 次抽样中，增益总和的期望值为$m_{e、\,q}\,\bar{g}_e$，因此期望 NCG 为 $\frac{\bar{g}_e}{g_{\max}}$。对于每个 ($e$、 q)，我们计算提升幅度 $\Delta_{e、\,q} = \mathrm{NCG}_{e}^{\mathrm{obs}}{}_{、\,q} - \bar{g}_e/g_{\max}$。随后，我们对该企业的查询取平均值，以获得企业层面的效应 Δ_e，并通过对企业进行自助重采样，以 $\{\Delta_e\}$ 的均值及其 95% 置信区间汇总各企业结果

[5] https://x.com/nikunjhanda/status/1912387834234958203

（$B = 1$、000）。我们报告平均提升幅度、其区间以及百分点（pp），其中 $\Delta_{\text{pp}} = 100\,\Delta$。

d) 曝光关联建模。为考察哪些因素与较高曝光相关，我们检验以下因素是否与排名相关：(i) 企业内部参与的集中程度，以 Jaccard相似度 j 衡量；以及 (ii) 创作者的受众规模 f （关注者）。我们采用普通最小二乘法，并计算稳健聚类标准误：

$$\log(r_i) \;=\; \beta_0 \;+\; \beta_1\, j_i \;+\; \beta_2\, \log(1 + f_i) \;+\; \varepsilon_i,$$

其中，因变量为创作者排名的对数（排名越低 r 越好），因此负系数表示曝光排名更靠前。样本仅包含与企业排行榜中排名创作者相匹配的引文面板来源。预测变量均经过均值中心化，因此截距是在 j 和 f 取平均值时计算的。

3 研究发现

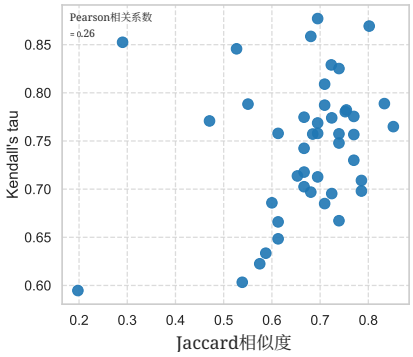
研究问题1：创作者多样性。对于每家企业，我们使用 Jaccard相似度 J 计算前100名集合的成员稳定性，其中 $J \in [0, 1]$，且 0 表示没有共同创作者，1 表示排行榜完全相同。我们使用 Kendall's τ 衡量该集合内的排名顺序稳定性，计算时仅纳入在两个快照中均出现的创作者，其中 $\tau \in [-1, 1]$；1 表示完全一致，0 表示无关联，-1 表示完全逆转。在这 20 天内，创作者社区的成员构成和排名均表现出较高稳定性，表明同一批核心创作者持续活跃，其相对影响力在此期间基本保持不变。图1a总结了这些分布及其相互关系；在所有企业中，平均 Jaccard相似度为 0.67，所有创作者的 Kendall's τ 为 0.75。

表1：各生成式搜索引擎中企业引文面板的分布

	查询 唯一网址 X 网址（n，%）	账号名（n，%）
Grok	132 2,223 1,123（50.52%）	152 (7.02%)
Perplexity（用户界面）	132 3,698 1,255(33.94%)	134（3.62%）
Grok（仅限 X）	132 2,194 2,194（100%）	333（15.12%）

RQ2：创作者曝光。接下来，我们分析生成式搜索引擎如何在不同企业间分配创作者曝光。每个查询-GSE对都会返回一个引文面板，我们对其进行解析，并与排名创作者匹配。首先，我们对比不同的检索配置：能够直接访问社交来源的生成式搜索引擎（例如 Grok 的"Web+ X"和"仅X"模式），其创作者关联引文的报道量高于主要依托网络内容的系统。相比之下，GPT、Gemini 和 Perplexity 的 API 配置不会返回任何来自 X 的来源，即使参数中明确指定了该域名。不过，在 Perplexity 的用户界面中，

（a）创作者的时间持久性



(b) 被引用创作者排名的分布

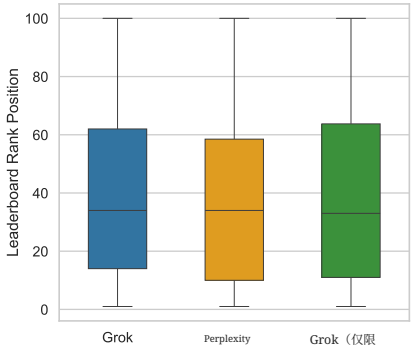


图1：（a）在20天间隔内，不同企业中创作者社区的成员构成与排名顺序稳定性。（b）被引用创作者排名的分布。

每家企业仅使用一个查询的基线运行（未设置域名包含控制）显示，X 报道量为 12.13 %。为使 Perplexity 的来源构成更接近 Grok，我们对每家企业运行了三个查询，并在启用网络访问的同时，允许包含 X 域名。在此设置下，Perplexity 返回了 3,698 个唯一来源，其中 1,255 个来自 X（33.94%），并匹配到 134 个创作者账号名（3.62%）。最后，将 Grok 的来源限制为仅X后，得到 2,194 条 X 引文，其中有 333 个重合的创作者账号名（15.12%）。表1报告了各生成式搜索引擎的引用面板细分情况。

为量化曝光集中度，我们计算企业内部前十名创作者与尾部创作者之间的差异，其中正值表示排名较高的创作者占优。Grok 的跨企业平均差异为 $+4.32$ 个百分点（pp），95%自助法区间为 $[+1.46, +7.51]$。Perplexity 也呈现类似模式，平均优势为 $+4.79$ pp $[+2.42, +7.35]$。当 Grok 仅限于仅X来源时，优势增至 $+7.16$ pp，表明曝光更加集中于最 prominent voices。这些估计汇总于表2a，并与图1b中被引用排名的分布一致：在各生成式搜索引擎中，排名较高（数字较小）的创作者更为集中。

为评估超出随机机会的放大效应，我们将观测到的归一化创作者增益与企业特定的随机基线进行比较；该基线等于从企业排名创作者集合中进行均匀抽样时预期的归一化增益。我们将提升幅度报告为差值（观测值减去基线），单位为百分点。总体而言，在各种配置下，我们都观察到显著超出随机机会的提升幅度。对于 Grok，跨 $n = 37$ 家企业的平均 NCG提升幅度为$+9.39$ 个百分点，95%区间为 $[+2.59, +16.45]$ （排除的企业是由于缺少重叠数据）。Perplexity 和 Grok（仅X）显示出相近的提升幅度，分别为 $+9.94$ 个百分点和 $+8.99$个百分点（见表2a）。

研究问题3：曝光驱动因素。 我们考察创作者和企业属性与引用可见度之间的关联。表2b报告了以引用排名为因变量、以企业层面集中度和创作者

受众规模为自变量的普通最小二乘回归系数。截距（$\beta_0 = 3.05$）提供了一个基准，对应于粉丝数和多样性指标处于平均水平时，预期排名约为21。相较于这一基准，结果表明，曝光倾向于受众规模较大的创作者，以及创作者活动更集中的企业。具体而言，粉丝数与引文面板中排名较高的创作者所获曝光显著相关（$\beta_2 = -0.44$）。企业创作者社群的多样性以 Jaccard相似度 j ($\beta_1 = -1.62$) 衡量，该指标反映成员稳定性；结果显示，创作者生态系统越集中，排名较高的声音获得报道的可能性就越大。

表 2：（a）生成式搜索引擎中的前十优势和 NCG提升幅度，以百分点差值的均值表示。（b）引用曝光的 OLS系数。

	前十名与尾部（95% CI）	NCG提升幅度（95% CI）	企业
Grok	4.32 (1.46，7.51)	9.39（2.59，16.45）	37
Perplexity	4.79（2.42，7.35）	9.94（3.18，16.57）	38
Grok(X)	7.16（3.81,10.93）	8.99（3.83，14.28）	41

(a)

	系数
截距	3.05*
j_i	-1.62*
$\log(1 + f_i)$	-0.44*
观测值	373
伪 R^2	0.30
* $p < 0.05$	

(b)

4 启示与理论变革

生成式搜索引擎以综合生成的答案和引文取代排序列表，从而改变可见度的分配方式。在 Web3 等由创作者驱动的生态系统中，关注度本就集中在少数声音上，这可能会放大现有层级，并削弱观点多样性。针对这些问题，我们作出以下贡献：（i）提出一种将引文面板视为曝光分配的审计框架；（ii）提供 API 与用户界面输出之间存在报道缺口的证据；（iii）提供企业级估计，显示前十优势及超出偶然水平的放大效应；（iv）构建一个涵盖企业专属查询、创作者排行榜和引文面板的语料库。这些贡献共同助力未来研究更严谨地评估曝光偏差。我们呼吁包括生成式搜索引擎公司和研究人员在内的各利益相关方加大力度，推动 API–UI 一致性；与此同时，在设计审计、报告结果以及跨研究比较研究发现时，也应将这些差异纳入考量。

局限性。我们的审计聚焦于Web3，这可能会限制研究结果对其他域名的可推广性，因为其他域名中的注意力动态可能有所不同 [9,4]。尽管我们记录了API与用户界面之间的差距，但由于无法充分了解系统内部机制，我们无法完全归因于其成因。创作者的突出程度通过外部排行榜进行衡量，而其评分标准可能存在自身的偏差。最后，我们的曝光指标假设引文之间的注意力分布均匀，且未纳入交互日志，因此估算的是潜在曝光，而非实际影响。

参考文献

1. Abdollahpouri, H., Burke, R., Mobasher, B.：通过个性化重排序管理推荐系统中的流行度偏差。arXiv预印本 arXiv:1901.07555（2019） 2. Aggarwal, P., Murahari, V., Rajpurohit, T., Kalyan, A., Narasimhan, K., Deshpande, A.：Geo：生成式引擎优化。载于：第30届ACM SIGKDD知识发现与数据挖掘会议论文集。第5–16页（2024） 3. Agrawal, R., Gollapudi, S., Halverson, A., Ieong, S.：使搜索结果多样化。载于：第二届ACM国际网络搜索与数据挖掘会议论文集。第5–14页（2009） 4. Alipour, S., Galeazzi, A., Sangiorgio, E., Avalle, M., Bojic, L., Cinelli, M., Quattrociocchi, W.：跨平台社交动态：对ChatGPT和新冠肺炎疫苗讨论的分析。《科学报告》**14**（1），2789（2024） 5. Allan, J., Choi, E., Lopresti, D.P., Zamani, H.：生成式人工智能时代的信息检索研究未来。arXiv预印本 arXiv:2412.02043（2024） 6. Biega, A.J., Gummadi, K.P., Weikum, G.：注意力公平性：在排名中摊销个体公平性。载于：第41届ACM SIGIR信息检索研究与发展会议。第405–414页（2018） 7. Carbonell, J., Goldstein, J.：MMR的应用：基于多样性的重排序，用于重新排列文档和生成摘要。载于：第21届ACM SIGIR信息检索研究与发展国际会议论文集。第335–336页（1998） 8. Chen, M., Wang, X., Chen, K., Koudas, N.：生成式引擎优化：如何主导人工智能搜索。arXiv预印本 arXiv:2509.08919（2025） 9. Di Marco, N., Bonetti, A., Di Martino, E., Loru, E., Nudo, J., Pandolfo, M.E., Pecile, G., Sangiorgio, E., Scalco, I., Zollo, S., 等：在线社交媒体中的模式、模型与挑战：综述。arXiv预印本 arXiv:2507.13379（2025） 10. Diaz, F., Mitra, B., Ekstrand, M.D., Biega, A.J., Carterette, B.：使用期望曝光评估随机排名。载于：第29届ACM信息与知识管理国际会议论文集。第275–284页（2020） 11. Gienapp, L., Scells, H., Deckers, N., Bevendorff, J., Wang, S., Kiesel, J., Syed, S., Fröbe, M., Zuccon, G., Stein, B., 等：评估生成式即席信息检索。载于：第47届ACM SIGIR信息检索研究与发展国际会议论文集。第1916–1929页（2024） 12. Helberger, N.：曝光多样性作为政策目标。《媒体法杂志》**4**（1），65–92（2012） 13. Järvelin, K., Kekäläinen, J.：基于累积增益的信息检索技术评估。《ACM信息系统汇刊》（TOIS）**20**（4），422–446（2002） 14. Kaiser, C., Kaiser, J., Schallner, R., Schneider, S.：在线搜索的新时代？关于生成式人工智能与传统搜索引擎实际搜索任务期间用户行为和个人偏好的大规模研究。载于：CHI计算机系统中的人为因素会议扩展摘要论文集。第1–7页（2025） 15. Kristoufek, L.：比特币遇上Google趋势与维基百科：量化互联网时代现象之间的关系。《科学报告》**3**（1），3415（2013） 16. Li, S., Ma, J.：Twitter帖子的情感与互动对加密货币价格变动的影响。《金融研究快报》**65**，105598（2024）

17. Santos, R.L., Peng, J., Macdonald, C., Ounis, I.：通过子查询实现显式搜索结果多样化。载于：欧洲信息检索会议。第87–99页。Springer（2010）18. Sharma, N., Liao, Q.V., Xiao, Z.：生成式回音室？大语言模型驱动的搜索系统对多样化信息寻求的影响。载于：2024年CHI计算机系统中的人为因素会议论文集。第1–17页（2024）19. Singh, A., Joachims, T.：排名中的曝光公平性。载于：第24届ACM SIGKDD知识发现与数据挖掘国际会议论文集。第2219–2228页（2018）20. Suri, S., Counts, S., Wang, L., Chen, C., Wan, M., Safavi, T., Neville, J., Shah, C., White, R.W., Andersen, R., 等：生成式搜索引擎在知识工作和复杂任务中的应用。arXiv预印本 arXiv:2404.04268（2024）21. Venkit, P.N., Laban, P., Zhou, Y., Mao, Y., Wu, C.S.：人工智能时代的搜索引擎：事实性且可验证的引证来源响应的虚假承诺。arXiv预印本 arXiv:2410.22349（2024）22. Wu, Y., Zhong, S., Kim, Y., Xiong, C.：生成式搜索引擎青睐什么，以及如何协同优化网络内容。arXiv预印本 arXiv:2510.11438（2025）