Data & Policy (2026), 8: e15
doi:10.1017/dap.2026.10064

CAMBRIDGE UNIVERSITY PRESS

RESEARCH ARTICLE

The attribution crisis in LLM search results: Estimating ecosystem exploitation

Ilan Strauss[1,2,3], Jangho Yang[4], Tim O'Reilly[3,5], Sruly Rosenblat[3] and Isobel Moure[3]

[1]Institute for Innovation and Public Purpose (The Bartlett School of Architecture), University College London, UK
[2]DSI/NRF South African Research Chair in Industrial Development (SARChI-ID), University of Johannesburg, South Africa
[3]AI Disclosures Project (Code for Science and Society), USA
[4]University of Waterloo, Canada
[5]O'Reilly Media, Inc., USA
Corresponding author: Ilan Strauss; Email: ilanstrauss@gmail.com

Received: 18 September 2025; **Revised:** 27 March 2026; **Accepted:** 30 March 2026

Keywords: citations; content monetization; ecosystem exploitation; large language models; LLM search

Abstract

Web-enabled large language models (LLMs) frequently answer queries without crediting the web pages they consume, creating an "attribution gap" in responsible artificial intelligence (AI) usage—defined as the difference between *relevant* URLs read and those actually cited. Drawing on approximately 14,000 real-world LMArena conversation logs with search-enabled LLM systems, we document three exploitation patterns: (1) *no search*: 34% of Google Gemini and 24% of OpenAI GPT-4o responses are generated without explicitly fetching any online content; (2) *no citation*: Gemini provides no clickable citation source in 92% of answers; (3) *high-volume, low-credit*: Perplexity's Sonar visits approximately 10 relevant pages per query but cites only three to four. A negative binomial hurdle model shows that the average query answered by Gemini or Sonar leaves about three relevant websites uncited, whereas GPT-4o's tiny uncited gap is best explained by its selective log disclosures rather than by better attribution. Citation *efficiency*—extra citations provided per additional relevant web page visited—varies widely across models, from 0.19 to 0.45 on identical queries, underscoring that retrieval design, not technical limits, shapes ecosystem impact. To advance auditing and monitoring of AI systems, we recommend a transparent LLM search architecture based on standardized telemetry and full disclosure of search traces and citation logs.

1. Introduction

> The misery of being exploited by capitalists is nothing compared to the misery of not being exploited at all.—Joan Robinson, *Economic Philosophy*, 1962.

The rapid integration of large language models (LLMs) into information seeking workflows has fundamentally transformed how users access and interact with content on the World Wide Web. Search-augmented LLMs, which combine generative capabilities with real-time web retrieval, promise to deliver more accurate, up-to-date, and comprehensive responses than traditional chatbots limited to just their training data (Huang and Chang, 2024). But questions have been raised about responsible AI use, including concerns about source accuracy (Smith et al., 2025), intellectual property rights, unauthorized

This research article was awarded Open Data and Open Materials badges for transparent practices. See the Data Availability Statement for details.

© The Author(s), 2026. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/4.0), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

https://doi.org/10.1017/dap.2026.10064 Published online by Cambridge University Press

website scraping ("access violations") (Cool-Fergus, 2025; Rosenblat et al., 2025), and what we term "ecosystem exploitation" (we use "consume," "read," and "search" interchangeably for a relevant website visit logged by the LLM).

We define *ecosystem exploitation* as *the systematic gap between relevant web content consumed by an LLM system when answering a query and the sources it credits in its output—a measurable shortfall in attribution that undermines the economic reciprocity on which the web's content ecosystem depends.* We call this measurable shortfall the *attribution gap.*

This attribution gap has serious implications for the digital ecosystem on which AI's ongoing utility depends. Content creators, publishers, and knowledge producers rely on appropriate attribution and licensing agreements when their information is used to answer queries. When LLMs systematically consume relevant content without adequate citation or remuneration, they undermine the incentive structures that support high-quality information production and threaten the economic viability of content creation at scale (O'Reilly, 2024); for previous empirical research on attribution practices across commercial LLM systems, see Gao et al. (2023), Yue et al. (2023), Li et al. (2023), Huang et al. (2025), Fayyazi et al. (2025), and Profound (2025). This systematic lack of attribution of the content sources consumed by LLMs during web search is a widely known issue (AutoGPT, 2024; Hacker News, 2025; Reuters, 2025). As recently as June 2025, the BBC threatened Perplexity's search system with legal action for using its content during search (Reuters, 2025); Perplexity has since set up a revenue-sharing program with some publishers and increased citation of sources in response to accusations against it (AutoGPT, 2024). Yet, major LLM vendors reveal little about how their retrieval-augmented generation (RAG) pipelines choose and ingest web content, and how they cite appropriate web sources (Stox, 2025).

Research questions: *To what extent do search-augmented LLMs exploit the web—using content without citing it? And what distinguishes today's best and worst attribution practices?*

Method and data. Using an LMArena dataset of ~14,000 real-world answers to around 7,000 multiturn queries taken between March and April 2025, we analyze the "attribution gap" exhibited by 11 LLM search-enabled models across three provider families: OpenAI (GPT-4o), Perplexity (Sonar), and Google (Gemini). We define the attribution gap as the number of *relevant* URLs visited by the LLM system when answering the query minus the number of URLs cited in the model's output, providing a direct measure of ecosystem exploitation. To isolate this quantity, we remove all hallucinated citations (numbered citations provided in-text that do not have a corresponding source in the search log) and all ungrounded citations (URL citations provided that do not appear in the search logs but may link to a valid site). Our statistical model is a negative binomial hurdle model with bootstrapped confidence intervals. This allows us to quantify both when attribution gaps occur and how severe they are, while accounting for differences in query type (e.g., "Data Science" and "Current Affairs"). In a second regression, we leverage the dataset's head-to-head design, looking at differences between model citation behavior for the same query.

Key findings:

(1) *Search-enabled LLM systems exploit by, surprisingly, not searching at all*, relying instead on their pretraining data or simply not disclosing relevant search logs accurately. Despite the models being in search mode, 15.6% of LLM answers skipped web search entirely. This was highest for Google's Gemini (34%), followed by OpenAI's GPT-4o models (24%).

(2) *LLM search systems exploit by providing no citations (zero attribution)*, 30% of answers provided no citations. This is driven less by query topic and more by model-specific behaviors. Gemini provided no citations for a striking 92% of queries, undermining claims that its impact on third-party traffic will be negligible.

(3) Our zero-hurdle statistical model shows that, for a typical query, *Google's Gemini models and Perplexity's Sonar models have sizable attribution gaps*: of approximately three relevant websites visited but not cited. Perplexity exhibits much higher volume ecosystem exploitation, visiting ~10 relevant websites per query, but with a similar overall attribution gap to Gemini.

(4) *The full extent of ecosystem exploitation may be underestimated* because models appear to selectively disclose which websites they visit, especially GPT-4o models. By design, GPT-4o

https://doi.org/10.1017/dap.2026.10064 Published online by Cambridge University Press

models appear to have a near-perfect correspondence between relevant websites drawn on and those cited, leading to a small attribution gap. Crucially, this selective log disclosure is itself a form of opacity that constitutes a distinct policy problem, because it renders independent auditing of attribution practices effectively impossible.

(5) *Refining a search LLM's RAG pipeline can almost double the number of citations it provides for each extra web page it consumes.* In "head-to-head" model regressions, comparing citation differences between model pairs for identical queries, we find that citation efficiency—the extra citations shown per additional website visited—ranges from 0.19 to 0.45. This indicates that retrieval design (reasoning modules, search context size, and geolocation), rather than technical limits, determines AI's relationship with the World Wide Web's ecosystem.

The classical political economists defined *exploitation* as a category of production, whereby an owner–producer appropriates the difference between the cost of an input and its realized value contribution to output (Zwolinski et al., 2022). The classical economists focused exclusively on the labor input (Fine, 1989; Hollander, 1992), but we can easily extend this framework to the data inputs consumed by LLMs during inference when producing a relevant response (the output).

Policy implications. Advancing a healthy web ecosystem requires *transparent search telemetry* (logs, traces, and metrics) to support real-time evaluation and auditing of deployed AI models (Barredo Barredo Arrieta et al., 2020). Developers, enterprise buyers, and potentially regulators should insist that LLM APIs expose a standard trace of every retrieval step and the sources ultimately cited. The tooling already exists to implement this: observability stacks such as LangSmith, Langfuse, Phoenix, and the GenAI semantic conventions in OpenTelemetry (see https://opentelemetry.io/docs/specs/semconv/gen-ai/) can record an end-to-end *search trace*—query, retrieval, reranking, and citation—so long as each web document is tagged with a stable source ID (typically a URL hash). Per-document relevance scores can travel in the same span (LangChain, 2025). If all providers adopt common definitions (e.g., llm.retrieval.ids, llm.retrieval.scores), third parties could compare across models the exact ratio of "information consumed" to "information cited"; and equitable business models could be built on top of this.

Limitations. Our study does not explicitly test citation accuracy or relevance, beyond removing hallucinated and ungrounded citations (Gao et al., 2023). Additionally, our study does not account for access violations (Cool-Fergus, 2025; Rosenblat et al., 2025)—whether the LLM had permission to visit specific websites, which may be governed by licensing agreements. Our measurement of the attribution gap also relies on provider-disclosed search logs, which introduces a potential two-directional bias: logs may include marginal results the model did not meaningfully draw upon (inflating the measured gap) or, conversely, omit pages the model actually consumed (deflating it), as appears to be the case with OpenAI's models. The net direction of this bias is ambiguous and likely varies across providers. Despite these limitations, our study represents the first systematic, cross-model audit of attribution behavior in commercial search-augmented LLM systems, focusing specifically on their search tools. Our goal is to provide a structured framework for assessing attribution in empirical LLM studies (Elliott and Archer, 2025).

Section 2 describes our data and variables; Section 3 details key empirical features of our data; Section 4 outlines our two regression models; Section 5 presents our core findings from the regressions; Section 6 discusses policy implications; and Section 7 concludes. The Appendix (provided as Supplementary Material) contains more detailed model results.

2. Data and method

We conduct a large-scale empirical audit of the attribution practices of commercial search-augmented LLMs in real-world user interactions. *Attribution* refers to identifying the source material or input features that contributed to a model's generated output or decisions (Li et al., 2023). It emphasizes *relevant* content "consumed" (read and used) by the LLM when answering, instead of all websites visited that may not be relevant and therefore not drawn upon. We assume the search results provided in an LLM's search log

https://doi.org/10.1017/dap.2026.10064 Published online by Cambridge University Press

were already determined to be relevant by the provider. Other forms of exploitation, especially ones involving unauthorized access, are not examined here (O'Reilly, 2024; Cool-Fergus, 2025).

Dataset overview. Our analysis uses LMArena's search dataset (see https://blog.lmarena.ai/blog/2025/search-arena/), containing pairs of model answers to the same user query across 11 major commercial LLMs. Our final sample size (n) is 13,929 observations. We construct all variables from the raw logs rather than using the variables provided by LMArena due to various errors and inconsistencies.

The initial dataset before filtering contains 14,000 conversations from 3,642 users, covering 7,000 queries each given to a pair of models between March (44%) and April (56%) 2025. Each query represents a potential multiturn conversation, including the model's final response. Crucially, this dataset captures actual deployment behavior via application programming interface (API) calls.

Model coverage. We analyze 11 commercial variants grouped by provider:

- OpenAI: api-gpt-4o-mini-search, api-gpt-4o-search, api-gpt-4o-search-high, api-gpt-4o-search-high-loc
- Perplexity: ppl-sonar-pro, ppl-sonar-reasoning, ppl-sonar, ppl-sonar-pro-high, ppl-sonar-reasoning-pro-high
- Google: gemini-2.0-flash-grounding, gemini-2.5-pro-grounding

The model configurations are detailed further in the LMArena documentation (accessed 16 June 2025). For Perplexity and OpenAI, the default "search context size" parameter is set to medium, which controls how much web content is retrieved and passed to the model. We also explore specific features by changing the default settings: for OpenAI, we test their geolocation feature in one model variant by passing a country code extracted from the user's IP address; for Perplexity and OpenAI, we include variants with search context size set to high. Gemini defaults to Google Search Tool enabled.

2.1. Key variables and measurement

We do not rely on LMArena's citation variable construction, which we found did not adequately distinguish between search logs and citations for our use case. We instead reconstruct them from scratch using the model logs in their dataset.

Ecosystem exploitation (attribution gap). Our primary dependent variable measures the difference between unique relevant pages visited by the LLM during search and unique pages actually cited to the user in the API response. This captures the extent to which a model consumes relevant web content without providing appropriate attribution.

Citations. We define citations as any in-text URL reference that is grounded in the model's search results log, including both explicit URLs and numbered references that link to specific sources. We exclude "hallucinated citations" (numbered citations, such as [13], that link to an empty source) and "ungrounded citations" (citation of URLs that exist but are not in the search log). Unique URLs (websites) for citations or attribution are those seen by the model that do not link to the same web page when stripped of parameters (e.g., example.com/?tracking_id=23222 is turned into example.com). This amounts to transforming URLs into their base URL and then checking for duplicates.

Relevant sites visited (consumed). We define relevant sites visited by the LLM as those listed in their search log, regardless of whether they were cited in-text or not. We treat every URL that appears in the search log as relevant, on the understanding that the provider has already filtered out nonrelevant visits. Note, however, that this log may itself be incomplete: some vendors—OpenAI, in particular—seem to pretrim their traces, returning only a subset of the relevant pages the model actually visited.

Conversation classification. We categorize all conversations into 12 topic areas using o4-mini with high reasoning to enable analysis of topic-specific attribution patterns. This classification covers areas from technical queries (software engineering, data science) to consumer-oriented topics (shopping, health, finance). We provided the 12 categories to GPT, drawing on an unsupervised visualization of clusters in the query data on Nomic Atlas (see https://atlas.nomic.ai/data/srulyrosenblat/ai-model-search-comparison-dataset/map).

https://doi.org/10.1017/dap.2026.10064 Published online by Cambridge University Press

Data cleaning. After removing conversations with classification failures or misaligned search traces, our final sample contains 13,929 observations, all successfully classified and with clean attribution data.

Prefiltering of logs. Most models appear to filter their logs for relevant websites visited only—rather than all sites visited. But OpenAI appears to filter them more strictly. Hallucinated citations (numbered citations provided in-text that do not exist in the search log) appear for Gemini and to a lesser extent Sonar (they are by definition zero for GPT-4o models since OpenAI's models do not use numbers in square brackets for their citations, which is what we track for this metric). Moreover, in terms of *citations provided that do not appear in the search logs* (ungrounded citations): 7% of citations by GPT-4o models were not found in their search logs, indicating either an overly restrictive prefiltering of the search logs or simply the model citing from pretrained knowledge.

3. Descriptive patterns

Our data show stark differences in attribution practices across model families, highlighting that design choices, rather than technological limitations, drive model behavior.

The attribution crisis. 15.6% of LLM responses involved no website visits despite being in search mode, yet 30% provided no citations whatsoever—a substantial gap between content consumption and content recognition. This pattern varies dramatically by provider; 39% of responses show perfect attribution alignment (zero gap), while 61% exhibit some degree of ecosystem exploitation. For Gemini, a "zero" gap still tends to signal exploitation though, because the model usually answers without visiting any external site even when its Google search tool is enabled (Table 1).

Provider-specific patterns. Table 1 illustrates three distinct patterns of exploitation: (1) *Not visiting relevant sites* at all, according to the model's logs, when answering a question (exploitation through pretraining reliance)—24% of GPT search answers and 34% of Gemini models. (2) Not providing *any citations* at all—25% of GPT-4o answers and 92% of Gemini answers. (3) Having a *large relative gap* between sites visited and sites cited—as in Perplexity's Sonar.

- *GPT-4o—Limited disclosure of sites visited*: On paper it shows almost perfect alignment between relevant pages searched according to its logs and pages cited. But this is likely an artifact of aggressive log filtering. The model seems to disclose only those URLs it ultimately cites, omitting any additional relevant pages it read (consumed). Support for this view comes from its high share of *ungrounded* citations—links that appear in the answer but not in the trace—suggesting that many visits are simply withheld from the log.
- *Gemini—No citations provided*: Systematic attribution failure with 92% zero-citation responses despite some relevant website searches.
- *Sonar—High-volume, low-credit*: Extensive crawling (10 median sites) with large attribution gaps (5.0 median) despite providing more citations than Gemini or OpenAI.

Table 1. *Attribution statistics by model family*

	GPT-4o	Gemini	Sonar
Median gap	0.0	4.0	5.0
Median citations	2.0	0.0	5.0
Median sites visited	2.0	4.0	10.0
Zero citations	1,309	2,198	663
Zero site visits	1,242	817	120
Sample size (*n*)	5,272	2,394	6,263

Note: "Gap" refers to the attribution gap = relevant websites visited (content consumed) minus unique websites cited.

https://doi.org/10.1017/dap.2026.10064 Published online by Cambridge University Press

Topic-specific vulnerabilities. Attribution failures concentrate in economically and legally sensitive domains, including:

- *Software engineering*: 33% zero-citation rate (700 queries), especially for Gemini.
- *Games/creative writing*: 40% zero-citation rate (494 queries).
- *Education*: 43.6% zero-citation rate (228 queries).
- *Health information*: For mental and physical health and relationships, systematic gaps for Gemini (94%) and GPT (26%).

4. Regression models

4.1. Hurdle model specification

The dependent variable $Y_i\,(i=1,\ldots,n)$ is the *attribution gap*: the number of relevant websites a model visits but fails to cite, for query *i*, where *i* runs from 1 to $n=13{,}929$. Because most answers are perfectly attributed ($Y_i=0$), while the rest show a skewed count distribution, we use a *hurdle model.* It breaks up $E[\mathit{Attribution\ Gap}]=P(\mathit{Gap}>0)\times E[\mathit{Gap}\,|\,\mathit{Gap}>0]$ into two sequential components:

(a) *Hurdle stage:* What is the probability that the gap is exactly zero $\rightarrow P(\mathit{Gap}>0)$? This is a Bernoulli outcome with probability π_i.
(b) *Count stage:* If an attribution gap exists ($Y_i>0$), how large is the gap: $E[\mathit{Gap}\,|\,\mathit{Gap}>0]$? We model positive counts with a *zero-truncated* negative binomial distribution, which has mean λ_i and over-dispersion parameter α (if $\alpha\rightarrow 0$, the model reduces to a Poisson hurdle; see Supplementary Material for the full log-likelihood specification and model diagnostics).

Putting the two parts together gives the probability mass function:

$$\Pr(Y_i=y|\mathbf{x}_i)=\begin{cases}\pi_i, & y=0,\\ (1-\pi_i)\,\dfrac{f_{\mathrm{nb}}(y;\lambda_i,\alpha)}{1-f_{\mathrm{nb}}(0;\lambda_i,\alpha)}, & y\geq 1,\end{cases}\tag{1}$$

where

$$f_{\mathrm{nb}}(y;\lambda,\alpha)=\frac{\Gamma(y+\alpha^{-1})}{\Gamma(\alpha^{-1})y!}(\alpha\lambda)^{y}(1+\alpha\lambda)^{-\left(y+\alpha^{-1}\right)},\ \alpha>0,\tag{2}$$

where $f_{\mathrm{nb}}(y;\lambda,\alpha)$ is the mean-parameterized negative binomial probability mass function.

Linear predictors. Each of the two regression stages has its own regression equation with predictors, consisting of: model family, the query category or "classification" (e.g., "Data Science"), the log of LLM output character count, and the number of unique search results. This leads to the following two equations (we note that the number of conversation "turns" had poor predictive power in our model and was therefore excluded):

$$\mathrm{logit}(\pi_i)=\gamma_0+\boldsymbol{\gamma}_{\mathrm{fam}}^{\top}\mathbf{1}\{\mathrm{model_family}_i\}+\boldsymbol{\gamma}_{\mathrm{cls}}^{\top}\mathbf{1}\{\mathrm{classification}_i\}+\gamma_{\ell}\log(\mathrm{response_character_count}_i),\tag{3}$$

$$\begin{aligned}\log\lambda_i=\beta_0&+\boldsymbol{\beta}_{\mathrm{fam}}^{\top}\mathbf{1}\{\mathrm{model_family}_i\}+\boldsymbol{\beta}_{\mathrm{cls}}^{\top}\mathbf{1}\{\mathrm{classification}_i\}\\&+\beta_{\mathrm{sr}}\,\mathrm{unique_search_results_count}_i+\beta_{\ell}\log(\mathrm{response_character_count}_i)\\&+(\mathbf{1}\{\mathrm{model_family}_i\}\otimes\mathbf{1}\{\mathrm{classification}_i\})^{\top}\boldsymbol{\beta}_{\mathrm{fam}\times\mathrm{cls}}\\&+(\mathbf{1}\{\mathrm{model_family}_i\}\otimes\mathrm{unique_search_results_count}_i)^{\top}\boldsymbol{\beta}_{\mathrm{fam}\times\mathrm{sr}}.\end{aligned}\tag{4}$$

https://doi.org/10.1017/dap.2026.10064 Published online by Cambridge University Press

Equation (3) is the logit probability, where $\exp(\gamma_{\text{fam}})$ represents the odds ratio for producing *any* attribution gap by model family, $\exp(\gamma_{\text{cls}})$ represents the odds ratio by classification topic, and $\exp(\gamma_{\ell})$ is the odds ratio for answer length. The equation contains only main effects (no interactions): the odds of *any* gap depend separately on model family, topic, and answer length. For example, $\exp(\gamma_{\ell}) = 1.4$ would mean a one-unit increase in log(response length) multiplies the odds of a gap by 1.4.

Equation (4) is the negative binomial count component, whereby $\exp(\boldsymbol{\beta}_{\text{fam}})$ represents the multiplicative changes in the expected number of uncited website visits by model family, *given that a gap exists.* $\exp(\boldsymbol{\beta}_{\text{cls}})$ represents the effects by classification topic, and $\exp(\beta_{\text{sr}})$ is the multiplicative effect of additional search results. A value of 1.30 implies a 30% larger gap, for example. Equation (4) adds interaction terms $\boldsymbol{\beta}_{\text{fam}\times\text{cls}}$, so the effect of a model family can differ by query topic once a gap exists, and $\boldsymbol{\beta}_{\text{fam}\times\text{sr}}$, to account for the varying impact of searching (number of URLs) on the attribution gap by model family.

4.2. Head-to-head model comparison regression

To isolate the number of URLs *each* LLM cites from one additional website search visit, while controlling for question types, we run a second regression model. This exploits the LMArena head-to-head design, whereby every question is answered by exactly two model systems. This means that any unobservable query-specific factors driving citation behavior cancel out. This design also ensures that β_{1m} is not biased by systematically "easy" or "hard" opponents. We collapse each model pair that answers the same query into a single observation (*focal model − opponent*) and run a separate OLS regression for every focal model m, where i now runs from 1 to $n = 6{,}951$, and m contains 11 focal models:

$$d_{im} = \beta_{0m} + \beta_{1m}\Delta s_{im} + \beta_{2m}\Delta \ell_{im} + \sum_k \gamma_k^{(m)} D_{ik} + \sum_j \delta_j^{(m)} O_{ij} + \varepsilon_{im}. \tag{6}$$

In Equation (6), d_{im} denotes the *citation-gap advantage*—the difference in unique citations produced by the focal model compared to its "opponent," $\text{citations}_m - \text{citations}_{\text{opp}}$. The term Δs_{im} captures the *search retrieval difference*, defined as the gap in unique URLs visited by the two systems in its logs, while $\Delta \ell_{im}$ measures the length difference in characters between their answers. The vector D_{ik} comprises topic dummies controlling for the classification of question i (reference category: "Current Affairs and Factual Information"), and O_{ij} contains dummies identifying which rival model the focal system faces (baseline = the first alphabetic opponent). The slope β_{1m} measures *citation efficiency*: the expected additional citations that model m produces per additional URL it opens, conditional on topic, verbosity, and opponent. Thus, $\beta_{1m} = 0.40$ implies that 10 extra retrievals yield roughly four extra citations. The coefficient β_{2m} captures the effect of answer length net of retrieval, while topic dummies ($\gamma_k^{(m)}$) and opponent dummies ($\delta_j^{(m)}$) control for domain-specific and match-up effects, respectively.

Each model's β_{1m} is our key quantity of interest and summarizes model m's retrieval-to-citation efficiency or yield: *the impact of a marginal website visit on citations.*

5. Regression findings

5.1. Attribution gap regression model

The regression results, transformed into probabilities, show notable differences in attribution reliability across AI model families (Table 2 and Figure 1). To obtain policy-interpretable results, we used the fitted model to make predictions for a standardized query covering Current Affairs and Factual Information with median characteristics (5 search results, 2,089 character responses). The hurdle component estimates the probability of having any attribution gap, while the count component estimates the expected gap size when gaps occur; multiplying these gives the unconditional expectation. The "expected attribution gap" is the total prediction from our model: $E[\textit{Attribution Gap}] = P(\textit{Gap} > 0) \times E[\textit{Gap} \mid \textit{Gap} > 0]$, whereas the first row ("Probability of a Gap") is only the first term of this equation, $P(\textit{Gap} > 0)$.

https://doi.org/10.1017/dap.2026.10064 Published online by Cambridge University Press

Table 2. *Expected total attribution gap by model family*

	GPT	Gemini	Sonar
Probability of a gap (%)	10.5	70.3	99.3
Expected attribution gap (total)	0.18	3.04	3.12
95% Confidence interval	[0.15, 0.23]	[2.80, 3.28]	[2.99, 3.25]
Standard error (SE)	0.02	0.12	0.06

Note: Results from negative binomial hurdle model for Current Affairs and Factual Information queries with median characteristics (5 search results, 2,089 characters, per query), with bootstrapped mean. Confidence intervals from parametric bootstrap (n = 1,000) for total expected attribution gap. Attribution gap is the missing web page (URL) citations per search query relative to relevant websites consumed, measured as a web page gap. This pattern holds consistently across all topic classifications (see Supplementary Table A3).

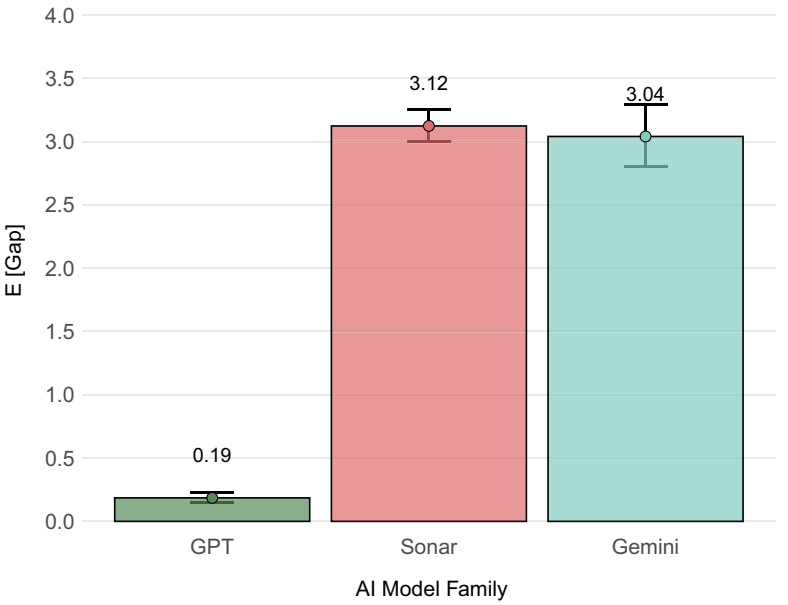


Figure 1. *Expected attribution gaps, predicted (by model family). Note: Predicted values for number of citations missing relative to web pages consumed. Based on negative binomial hurdle model regression coefficients. Bars show the model's expected citation gap (websites visited in the logs minus websites cited in the output), estimated at the median conversation length and median website visits, without interaction effects included. Showing 95% confidence intervals calculated with the* emmeans *package in* ℝ.

The GPT-4o models appear to show less exploitative attribution behavior, but this is most likely due to them being more circumspect in their disclosures by not showing the full extent of their logs (using OpenAI search in the user interface shows far more website visits based on limited testing). OpenAI's models have only a 10.5% probability of having any citation gaps and an expected 0.18 missing citations (relative to relevant websites consumed) per query (Table 2). Sonar exhibits citation gaps in 99.3% of queries with 3.12 expected missing citations per query in total. Gemini falls somewhere between these extremes with citation gaps in 70.3% of queries and 3.04 expected missing citations in total per query.

Sonar's near-universal citation gap (99.3% of queries) makes it particularly concerning for applications requiring source transparency. Gemini is arguably relying too heavily on internal knowledge given that a large portion of its zero attribution gap is due to it having zero (disclosed) website searches, at 34% of queries.

https://doi.org/10.1017/dap.2026.10064 Published online by Cambridge University Press

Finally, in terms of other factors driving *how large* the gap is once it appears (the count component), we can look at the incidence rate ratio (IRR), which tells us the multiplicative change in the expected count when the predictor increases by one unit, holding everything else constant. Longer answers actually *shrink* the gap size: When logged response character count doubles (from, e.g., 100 to 200 characters), the expected number of missing citations decreases by about 11% (IRR 0.89 or so). Reading more web pages from search inflates the gap: every extra page the LLM system reads raises the expected uncited count by 13% (IRR ~1.13).

5.2. Head-to-head model results: citation efficiency

In this section, we focus on the β_{1m} coefficient from our head-to-head regressions (Equation (6)). This coefficient tells us how many more citations the focal model generates for each additional search result compared to its opponent. A smaller coefficient signals poorer performance: the focal model converts each extra relevant page it opens into fewer citations than its comparator does. For example, if Sonar's coefficient is 0.4, this means that when it undertakes five additional relevant web page visits than GPT-4o on a question, it is expected to have $5 \times 0.4 = 2$ more citations than GPT-4o. 264 coefficients are estimated in total, running a separate regression for each of the 11 models with 24 parameters each: 1 intercept, 11 topic dummies, 1 focal-search-diff, 10 opponent-model dummies, and 1 focal-length-diff.

We plot β_{1m} in Figure 2 (see Supplementary Table A4). Almost all coefficients from the regression are positive and highly significant: the focal_search_diff coefficient is significant and positive for every model, indicating that extra retrieval consistently translates into more citations, though the payoff varies. Answer length matters for the basic GPT and Sonar variants but not for "high" search variants or Gemini. Topic and opponent effects are sparse and model specific.

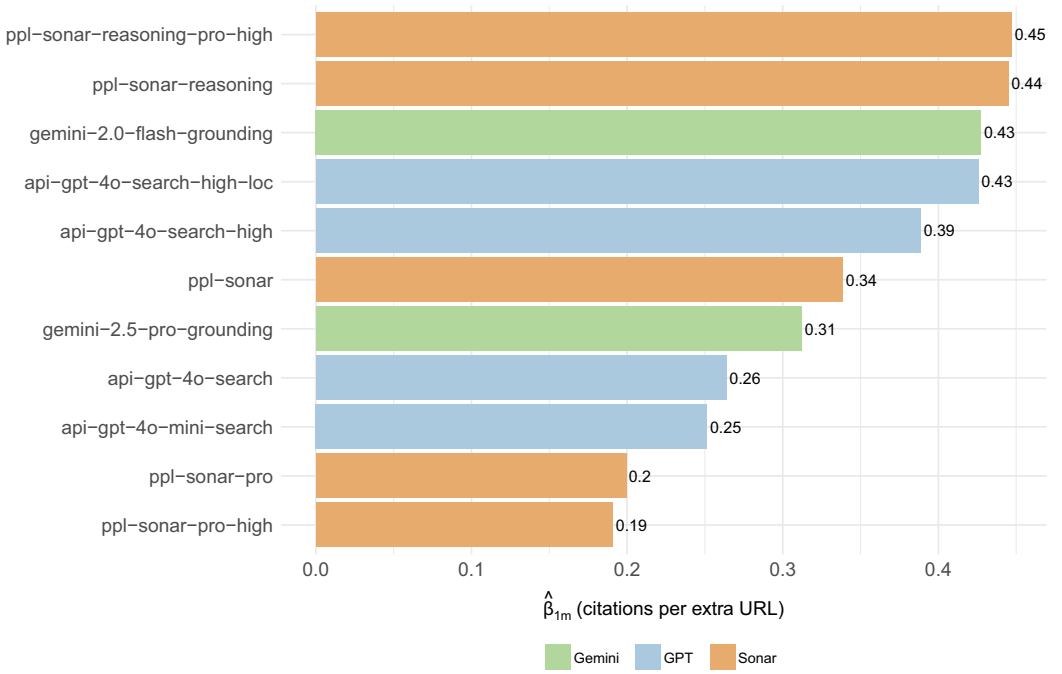


Figure 2. *Focal model: citation difference per extra URL visited. Note: Extra citations gained for each additional URL the focal model opens (differences between models). This holds match-up effects constant, isolating technology effects. Regression coefficient β_{1m} shown, predicting differences in citations for model pairs, for a given query. See equation before.*

https://doi.org/10.1017/dap.2026.10064 Published online by Cambridge University Press

Figure 2 shows that best-in-class variants (Sonar-reasoning-pro-high, Gemini-flash-grounding, GPT-4o-search-high-loc) yield ~ 0.43 citations per extra URL, whereas baseline variants (ppl-sonar-pro-high) return only ~ 0.19. This implies that the best model converts every extra retrieved URL into ~0.45 additional citations (compared to its competitor models), whereas the weakest model variant converts that same extra URL into ~0.19 citations. The span is therefore about $0.45 - 0.19 \approx 0.26$ citations per URL, showing that RAG implementation choices can more than double the payoff from each additional page the model visits. This illustrates just how wide the performance window is, and thus how much room developers (or regulators) have to raise low performers up to the current best practice.

More specifically, our results show that RAG implementation type (technology) is more important than model family (e.g., OpenAI vs. Gemini):

- An ANOVA test shows that within-family variation (the variance) of search coefficients β_{1m} is almost 8 times larger than between-family variation ($MS_{within} = 0.0116$ vs. $MS_{between} = 0.0015$; $F_{(2,8)} \approx 0.13$, $p = 0.88$). This means that differences in model behavior are far greater *within* model families than between them: implementation choices explain roughly 8 times more of the spread in citation efficiency than the provider family label itself.
- Within Sonar alone, upgrading to the *reasoning* tier more than doubles efficiency (from 0.19–0.20 → 0.44–0.45), with Sonar-reasoning (0.44) versus Sonar-pro-high (0.19) = 0.25 difference. For GPT models: GPT-4o-search-high-loc (0.42) versus GPT-4o-mini (0.25) = 0.17 difference.
- *Location signals* matter too. Adding a country code to GPT-4o raises search-citation coefficient efficiency by roughly 10 % (0.39 → 0.43), confirming that retrieval relevance translates directly into better attribution (though it is unclear why this is the case). Practically, this effect size is moderate since for a model getting 10 extra search results, location signals would generate ~0.37 additional citations. But we know that local search operates differently, both in traditional search and increasingly with LLMs (Rollison, 2025).

Finally, the regression results show that more basic models (GPT-mini, GPT-search, and basic Sonar models) compensate for lower search efficiency through verbosity—they need longer answers to achieve similar citation performance. Advanced ("high" variant and Gemini) models achieve higher citation rates through superior search utilization, making verbosity unnecessary to achieve improved citation rates.

6. Policy implications

Without standardized telemetry—comprehensive logs and traces of what an LLM retrieves and cites—no transparent and competitive market for licensing, revenue-sharing, or other content-monetization schemes can easily emerge. Publishers need hard numbers on how often their pages power an answer in order to automate royalty or revenue-share flows; regulators need the same auditable data to enforce forthcoming disclosure rules in jurisdictions such as the European Union (EU) and California. In short, richer telemetry is *the* prerequisite for both commercial remuneration and public oversight.

Our findings underscore that log opacity is a first-order policy problem for the AI ecosystem as a whole. If AI providers can selectively curate which retrievals appear in their traces, the attribution gap becomes unmeasurable, rendering any external audit inaccurate. GPT-4o's near-perfect attribution score, which we argue reflects aggressive log-filtering rather than genuinely superior citation behavior, illustrates the danger: a system can appear fully compliant while its true content consumption remains hidden. Log transparency must therefore be treated as a prerequisite for meaningful attribution standards, not an optional disclosure.

The technical pieces already exist for full disclosure of an LLM's search and citation trace; what remains is coordination. The key challenge is to persuade providers to adopt a common telemetry standard —and to ensure buyers and regulators can incentivize its provision.

Modern observability frameworks—LangSmith, Langfuse, Phoenix, and the GenAI semantic conventions in OpenTelemetry—allow for recording an end-to-end *search trace*, which can detail the search

https://doi.org/10.1017/dap.2026.10064 Published online by Cambridge University Press

activities that an LLM RAG system undertakes when trying to find the most relevant context for the user. One way to think of traces is as a collection of structured logs with context, correlation, and hierarchy baked in. Each web page retrieval, reranking, and generation step by the LLM can be logged as an OpenTelemetry span (a unit of work that forms the building blocks of traces; see https://opentelemetry.io/docs/concepts/signals/traces/), provided every document is tagged with a stable *Source ID*, typically a hash of the original URL or file.

Because hashes are one-way fingerprints, they enable later verification without revealing copyrighted text. The same span can carry the numerical relevance score ("how relevant was this web page to the LLM's answer") produced by the vector store, BM25 index, or cross-encoder, as long as that score is preserved in the trace (LangChain, 2025; Ryaboy, 2025). If each retrieved document carries two stable fields—say llm.retrieval.ids (a hash or URL that uniquely identifies the page) and llm.retrieval.scores (its relevance score or rank)—*and* those fields are propagated from the retrieval step all the way to the final API response, then anyone inspecting the trace can, in theory:

(1) enumerate every page the model actually saw;
(2) check which of those pages were later cited in the answer; and
(3) compare the relevance scores of cited pages with those that were ignored.

In short, the full provenance of "pages viewed" versus "pages credited" becomes auditable.

OpenTelemetry-like open protocols can be a lightweight standard that enables comparison and validation of LLM behavior across models. And adopting telemetry protocols (standards) requires only incremental changes to today's open-source stacks: in LangChain, a single line of code drops the hash and score into Document.metadata, which LangSmith then records automatically in each trace span (LangChain, 2025). Langfuse and Phoenix perform similar mappings, meaning dashboards that plot "high-score pages not cited" can be deployed without further engineering.

The real challenge, perhaps, lies in constructing the market and regulatory incentives to advance widespread adoption and disclosure. Transparent traces unlock clear, quantifiable monetary benefits for providers. API buyers in legal, medical, and financial sectors increasingly require provenance guarantees. This means that model developers that expose richer evidence trails could, therefore, command premium pricing and benefit from greater demand in these compliance-sensitive sectors. However, the model developers may be reticent to undertake such additional disclosures without liability safeguards or more reliable RAG pipelines.

Yet, the incentive landscape is not symmetric across providers. Vertically integrated platforms—most notably Google, which controls both the dominant search index and a leading LLM family—face a structural disincentive to adopt full transparency. Disclosing the true scale of content consumption could empower publishers to demand higher licensing or pay per content fees and provide regulators with the quantitative basis for its negotiation. Smaller or search-dependent providers such as Perplexity, by contrast, could find transparency a viable differentiation strategy, signaling trustworthiness to both users and content partners. This asymmetry suggests that voluntary adoption alone is unlikely to produce a universal standard.

Closing this gap likely requires coordinated pressure from multiple stakeholders. *Enterprise buyers* can mandate telemetry-compliant APIs as a procurement requirement, analogous to existing SOC 2 or data-governance certifications. *Regulators*—particularly under frameworks such as the EU AI Act and California's proposed disclosure rules—can require search-trace transparency as part of broader AI accountability obligations. And *publishers*, armed with a standardized attribution gap metric, can collectively bargain on the basis of auditable evidence rather than opaque bilateral licensing deals. The combination of demand-side pressure (buyers), supply-side differentiation (smaller providers), and regulatory mandates is more likely to achieve widespread adoption than any single lever alone.

Our findings also point toward several specific paths forward that can incentivize actors to engage in proper attribution. Shopping and commercial-intent queries present the starkest case: they exhibit attribution gaps 76% larger than other categories, despite being the domain where attribution carries

https://doi.org/10.1017/dap.2026.10064 Published online by Cambridge University Press

the most direct economic value. Every uncited product review page or price comparison site represents a lost referral, a lost affiliate fee, or a lost click-through—in other words, a concrete market failure, not simply a technical shortcoming. Paradoxically, commercial queries are also where a mutually beneficial business model is most feasible. Early web advertising thrived precisely because attribution (click-through tracking, conversion measurement) aligned the incentives of platforms and content creators. An analogous model for LLM search—where transparent citation enables revenue-sharing, affiliate attribution, or direct-purchase referrals—could transform the category with the largest exploitation gap into the first site of sustainable content monetization.

7. Conclusion

This study provides one of the first systematic empirical audits of attribution practices in commercial search-augmented LLMs. When LLMs exploit content from platforms without proper attribution, they undermine the economic incentives that sustain high-quality information production. Returning to our opening quote by the late economist Joan Robinson—on the notion that exploitation under capitalism involves the absence of commercial relations as well as its presence—we find similar twin forces at work in our analysis. Gemini systematically excludes the World Wide Web's content ecosystem when answering questions as its form of exploitation, while Perplexity exploits through the opposite behavior, overly zealous consumption of web content without commensurate attributions.

We find a substantial "attribution gap" between relevant content consumed from websites and attribution practices that threatens the sustainability of the digital content ecosystem. Our analysis of ~14,000 real-world interactions demonstrates that leading AI systems systematically consume web content without adequate attribution, with Gemini providing no citations in 92% of its responses and Perplexity visiting approximately a dozen relevant websites while crediting only a few.

A negative binomial hurdle model shows that the average query answered by Gemini or Sonar leaves about three relevant websites uncited, whereas GPT-4o's tiny uncited gap is best explained by its selective log disclosures rather than by better attribution. GPT-4o models remain difficult to audit due to what appears to be stricter prefiltering of their model search logs.

The dramatic variation in citation efficiency between models when answering the same question—from 0.19 to 0.45 citations per additional web page visited—illustrates that attribution gaps result from identifiable design choices, not immutable technical limitations. Within Perplexity's own family, upgrading the reasoning module more than doubles citation efficiency; adding geolocation signals to GPT-4o raises it by roughly 10%; and increasing search context size consistently improves attribution across providers. The best-performing systems show that transparent, comprehensive attribution is technically feasible today, even if accurate citations and output remain a real ongoing issue.

Closing the attribution gap is, therefore, less a technical hurdle than one of proper coordination and market incentives. Standardizing two telemetry fields—document hashes and relevance scores—would allow anyone to verify which information an LLM consumed and how faithfully it credited that information. Observability tools already provide the necessary plumbing. What remains is a collective decision by model providers to enable it and by developers and buyers to reward those who do. The transparency this would produce extends beyond epistemic trust in the accuracy of individual answers; it is the foundation for *systemic* trust—confidence that the information ecosystem sustaining AI-generated responses is fair, economically sustainable, and capable of supporting high-quality content production over the long term.

Our results demonstrate that transparency in the web sources consumed and cited by LLMs when answering user queries is fundamentally an engineering choice. This aligns with emerging research on LLM citation evaluation frameworks (Gao et al., 2023) and attribution methods in scientific literature (Saxena et al., 2024; Najjar et al., 2025), which show that technical solutions are available but underutilized. Source-aware training and fine-tuning have also been shown to improve citation behavior (Borgeaud et al., 2022; Asai et al., 2024; Khalifa et al., 2024). We call on LLM providers to adopt standardized search telemetry, on enterprise buyers to require it in procurement specifications, and on

https://doi.org/10.1017/dap.2026.10064 Published online by Cambridge University Press

regulators to mandate it as part of emerging AI transparency frameworks. The web's content ecosystem—and the AI systems that depend on it—cannot afford to wait for voluntary action alone.

Supplementary material. The supplementary material for this article can be found at http://doi.org/10.1017/dap.2026.10064.

Data availability statement. All replication materials (code to construct variables, model scripts, and figure notebooks) and the cleaned dataset derived from the raw LMArena's March–April 2025 Search Arena logs, and are available at https://github.com/AI-Disclosures-Project/Ecosystem_Expl oitation_In_Search_Results. The repository includes instructions to regenerate the analysis dataset from the public source logs and notes any preprocessing and filtering steps.

Acknowledgments. We extend our appreciation to the people with whom we have had conversations that have shaped our thinking.

Author contribution. Conceptualization: I.S., T.O'R.; Data curation: S.R.; Formal analysis: J.Y., I.S.; Methodology: J.Y.; Supervision: I.S., T.O'R.; Writing—original draft: I.S., J.Y., T.O'R., S.R., I.M.; Writing—review and editing: I.S., J.Y., T.O'R., S.R., I.M.

Funding statement. This research was supported by the Alfred P. Sloan Foundation, Omidyar Network, and the Patrick J. McGovern Foundation. The funders had no role in the study design, data collection and analysis, the decision to publish, or preparation of the manuscript. The views expressed are those of the authors and do not necessarily reflect those of the funders.

Competing interests. T.O'R. is the founder and board chair of O'Reilly Media, a publisher whose business may be indirectly affected by policies governing online attribution and web traffic. O'Reilly Media had no role in the study design; data collection and analysis; decision to publish; or preparation of the manuscript. The other authors declare no competing financial interests. *Nonfinancial competing interests*: The authors work on AI transparency and disclosure standards in their academic and policy research.

References

Asai A, **Wu Z**, **Wang Y and Sil A and Hajishirzi H** (2024) Self-RAG: Learning to retrieve, generate, and critique through self-reflection. In *Proceedings of the Twelfth International Conference on Learning Representations (ICLR 2024)*. Vienna, Austria: OpenReview.net. Available at: https://openreview.net/forum?id=hSyW5go0v8.

AutoGPT (2024) *Perplexity AI Faces Attribution Controversy.* Available at https://autogpt.net/perplexity-ai-faces-attribution-controversy/ (accessed 10 August 2025).

Barredo Arrieta A, **Díaz-Rodríguez N**, **Del Ser J**, **Bennetot A**, **Tabik S**, **Barbado A**, **García S**, **Gil-López S**, **Molina D**, **Benjamins R**, **Chatila R and Herrera F** (2020) Explainable artificial intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion 58*, 82–115. Amsterdam: Elsevier. https://doi.org/10.1016/j.inffus.2019.12.012

Borgeaud S, **Mensch A**, **Hoffmann J**, **Cai T**, **Rutherford E**, **Millican K**, **van den Driessche G**, **Lespiau J-B**, **Damoc B**, **Clark A**, **de Las Casas D**, **Guy A**, **Menick J**, **Ring R**, **Hennigan T**, **Huang S**, **Maggiore L**, **Jones C**, **Cassirer A**, **Brock A**, **Paganini M**, **Irving G**, **Vinyals O**, **Osindero S**, **Simonyan K**, **Rae JW**, **Elsen E and Sifre L** (2022) Improving language models by retrieving from trillions of tokens. In Chaudhuri K, Jegelka S, Song L, Szepesvári C, Niu G and Sabato S (eds), *Proceedings of the 39th International Conference on Machine Learning. Proceedings of Machine Learning Research*, Baltimore, vol. 162, pp. 2206–2240. Maryland: PMLR.

Cool-Fergus S (2025) *What happens when the news starts answering back? In Conversation with Miso.ai CEO, Lucky Gunasekara.* Interview with Lucky Gunasekara, CEO of Miso.ai, from AI Frontrunners in News webinar series. *Twipe.* Available at https://www.twipemobile.com/what-happens-when-the-news-starts-answering-back-inconversation-with-miso-ai-ceo-lucky-gunasekara/ (accessed 16 June 2025).

Elliott O and Archer P (2025) *Representation of BBC News content by AI assistants.* Research Report. *BBC.* Available at https://www.bbc.co.uk/aboutthebbc/documents/bbcresearch-into-ai-assistants.pdf (accessed 10 August 2025).

Fayyazi R, **Zuzak M and Yang SJ** (2025) LLM embedding-based attribution (LEA): Quantifying source contributions to generative model's response for vulnerability analysis. *arXiv preprint*, arXiv:2506.12100.

Fine B (1989) *Marx's capital*. 3rd edition. London: Macmillan Education. pp.116. ISBN: 978-0-333-49457-8.

Gao T, **Yen H**, **Yu J and Chen D** (2023) Enabling large language models to generate text with citations. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*. Singapore: Association for Computational Linguistics, pp. 6465–6488. http://doi.org/10.18653/v1/2023.emnlp-main.398.

Hacker News (2025) *Citations on the Anthropic API.* Discussion thread on Claude citation behavior. Available at https://news.ycombinator.com/item?id=42807173 (accessed 10 August 2025).

Hollander S (1992) *Classical Economics*. Toronto: University of Toronto Press.

Huang J and Chang KC-C (2024) Citation: A key to building responsible and accountable large language models. In *Findings of the Association for Computational Linguistics: NAACL 2024*. Mexico City, Mexico: Association for Computational Linguistics, pp. 464–473.

https://doi.org/10.1017/dap.2026.10064 Published online by Cambridge University Press

Huang B, **Chen C and Shu K** (2025) Authorship attribution in the era of llms: Problems, methodologies, and challenges. *ACM SIGKDD Explorations Newsletter 26*(2), 21–43.

Khalifa M, **Wadden D**, **Strubell E**, **Lee H**, **Wang L**, **Beltagy I and Peng H** (2024) Source-aware training enables knowledge attribution in language models. In *Proceedings of the First Conference on Language Modeling (COLM 2024)*. OpenReview.net. Available at: https://openreview.net/forum?id=UPyWLwciYz.

LangChain (2025) *How to Get a RAG Application to Add Citations.* Available at https://python.langchain.com/docs/how_to/qa_citations/ (accessed 24 June 2025).

Li D, **Sun Z**, **Hu X**, **Liu Z**, **Chen Z**, **Hu B**, **Wu A and Zhang M** (2023) A survey of large language models attribution. *arXiv preprint*, arXiv:2311.03731.

Najjar AA, **Ashqar HI**, **Darwish O and Hammad E** (2025) Leveraging explainable AI for LLM text attribution: Differentiating human-written and multiple LLMs-generated text. *Information 16*(9), 767. https://doi.org/10.3390/info16090767.

O'Reilly T (2024) How to fix AI's original sin. *O'Reilly Radar. O'Reilly.* Available at https://www.oreilly.com/radar/how-to-fix-ais-original-sin/ (accessed 16 June 2025).

Profound (2025) *AI platform citation patterns: How ChatGPT, Google AI overviews, and perplexity source information.* Research analysis from August 2024 to June 2025. Available at https://www.tryprofound.com/blog/ai-platform-citation-patterns (accessed 10 August 2025).

Reuters (2025) BBC threatens legal action against AI startup perplexity over content scraping, FT reports. *Reuters.* Available at https://www.reuters.com/business/media-telecom/bbc-threatens-legal-action-against-ai-start-upperplexity-over-content-scraping-2025-06-20/ (accessed 24 June 2025).

Rollison D (2025) *How does ChatGPT conduct local searches? Search Engine Land.* Available at https://searchengineland.com/how-does-chatgpt-conduct-local-searches-454894 (accessed 24 June 2025).

Rosenblat S, **O'Reilly T and Strauss I** (2025) *Beyond public access in LLM pre-training data: Non-public book content in OpenAI's models.* Technical Report. SSRC AI WP 2025-04. Social Science Research Council. https://doi.org/10.35650/AIDP.4111.d.2025.

Ryaboy M (2025) *Cross–Encoders, ColBERT, and LLM–Based Re–Rankers: A Practical Guide.* Available at https://medium.com/@aimichael/cross-encoders-colbert-and-llm-basedre-rankers-a-practical-guide-a23570d88548 (accessed 24 June 2025).

Saxena Y, **Tilwani D**, **Mohammadi A**, **Raff E**, **Sheth A**, **Parthasarathy S and Gaur M** (2024) Attribution in scientific literature: New benchmark and methods. *arXiv e-prints*, arXiv:2405.02228.

Smith J-M, **Intagliata C and Kelly ML** (2025) *The Trump administration's report on kids' health cites made-up scientific studies. National Public Radio.* Available at https://www.npr.org/2025/05/30/nx-s1-5417962/ (accessed August 2025)

Stox P (2025) *Websites with more organic search traffic get mentioned more in AI search. Ahrefs.* Available at https://ahrefs.com/blog/websites-with-more-traffic-have-more-mentions/ (accessed 24 June 2025).

Yue X, **Wang B**, **Chen Z**, **Zhang K**, **Su Y and Sun H** (2023) Automatic evaluation of attribution by large language models. In *Findings of the Association for Computational Linguistics: EMNLP 2023*. Singapore: Association for Computational Linguistics, pp. 4615–4635. http://doi.org/10.18653/v1/2023.findings-emnlp.307.

Zwolinski M, **Ferguson B and Wertheimer A** (2022) Exploitation. In Zalta EN and Nodelman U (eds.), *The Stanford Encyclopedia of Philosophy.* Stanford: Metaphysics Research Lab, Stanford University.

Cite this article: Strauss I, Yang J, O'Reilly T, Rosenblat S and Moure I (2026). The attribution crisis in LLM search results: Estimating ecosystem exploitation. *Data & Policy*, 8: e15. doi:10.1017/dap.2026.10064

https://doi.org/10.1017/dap.2026.10064 Published online by Cambridge University Press