

Does AI Hate Sponsorships?

Prepared by Nuno¹

September 14, 2026

Abstract

This paper examines whether sponsored AFAR articles have a different probability of receiving an AI citation than non-sponsored articles. We use a Sponsored dummy that equals one for articles classified as sponsored and zero for articles classified as non-sponsored. Unassessed sponsorship status is handled separately. The data contain 171,449 answers across six sources and collection routes. The main estimation sample contains 350,122 response–article opportunities, 748 queries, and 2,022 AFAR articles. The preferred sponsorship coefficient is -0.1 percentage points and is not statistically significant at the 5% level. Alternative comparison counts, query weights, sample restrictions, and source-specific models give the same qualitative result. The analysis does not provide statistically significant evidence of a sponsorship penalty for AFAR in the constructed sample.

1 Introduction

AI assistants are changing how people access information to plan trips. Travellers can use a chat interface to discover destinations, compare hotels, build itineraries, and seek advice during a trip. These interactions place generated recommendations between travellers and the websites that provide travel information. Understanding which sources these assistants research and cite in this process is a central question for the future of travel media.

The need for current information makes this question particularly relevant in travel. Prices, opening hours, availability, and local conditions can change after a model is trained. Knowledge acquired during pre-training alone cannot reliably capture these changes. Retrieval from external sources can help assistants provide current answers. Travel publishers can therefore inform recommendations, but their visibility depends on how the assistant researches an answer. The assistant interprets the context of the query, expands it into several related searches, retrieves candidate pages, and then decides which sources to read and cite.

Brands need to understand whether particular characteristics of their content reduce its probability of citation. Sponsorship is one such characteristic. Commercial partnerships support travel content, and these relationships are part of the commercial model of travel publishers. If models disproportionately exclude sponsored pages, these partnerships could carry a visibility cost. Whether sponsored content receives fewer citations is therefore an important empirical question.

In this paper, we study whether sponsored content is cited with a similar probability to non-sponsored content. We present a case study of AFAR, a travel media brand and magazine. We use Nuno’s records of 171,449 AI answers to 2,914 queries across six sources and collection

¹ Discover Nuno at <https://discovernuno.com/>

routes. These records identify the citations included in each answer. We combine them with an inventory of 22,152 AFAR records containing page information, content categories, and known sponsorship status. We use the recorded Sponsored dummy and handle unassessed sponsorship status separately.

We construct comparison groups of AFAR articles that could address the same query. For each cited article, we identify articles in the same content category that were published before the query’s first recorded answer. We rank candidates using query terms and similarity to the cited article. After applying identity exclusions, we retain up to five comparison articles. The resulting sample contains 350,122 response–article opportunities across 748 queries and 2,022 AFAR articles. We estimate the relationship between sponsorship and citation with query and source fixed effects. Standard errors are clustered by query and article.

We find no statistically significant sponsorship penalty for AFAR at the 5% level. The preferred coefficient is -0.1 percentage points, relative to an article-level citation probability of 1.1% per response–article opportunity. The robustness checks support the same conclusion. We vary the number of comparison articles, hold cited anchors fixed, and restrict the sample to stronger lexical matches. We also give queries equal weight, exclude observations with unassessed sponsorship status, and estimate the relationship separately by source. None of these sponsorship coefficients is statistically significant at the 5% level.

We find no evidence of a statistically significant citation penalty for sponsored AFAR content.

2 Related literature

Research on retrieval-augmented generation explains how external documents can enter a generated answer. Lewis et al. (2020) combine a generative model with an external document memory. Their framework motivates the distinction between stored model knowledge and retrieved information. It does not establish the architecture used by every commercial source in our data. We observe recorded citations, rather than the complete retrieval process.

Research on generative engine optimisation studies how content changes affect source visibility. Aggarwal et al. (2024) introduce a framework and benchmark for evaluating such changes. Our analysis examines existing AFAR pages with different sponsorship status. The comparison therefore differs from an experiment that modifies the same page and holds its other attributes fixed. It measures an observational association in Nuno’s tested query population.

Citation itself is an incomplete measure of source use. Liu, Zhang, and Liang (2023) find gaps in citation completeness and accuracy in generative search. Their findings support a narrow interpretation of our outcome. A recorded link shows that an AFAR page appears in the saved citation extract. It does not establish that the page supports every related statement, determines the recommendation, or receives reader traffic.

These literatures distinguish several stages that our data cannot fully connect. An article can be retrieved, used to construct an answer, and cited. We observe its recorded sponsorship status and whether it receives a citation. Our contribution is to document the association

between these two variables for AFAR under a fixed comparison strategy. Citation does not capture every way an article may influence an answer.

3 Context

Nuno measures visibility and relevance through repeated AI query testing. The project records answers with their query, source, company, and analysis identifiers. A query can produce several answers and can appear under more than one source. This structure supports measurement of citation frequency across repetitions. It also requires clear separation between counts of queries and counts of answers.

4 Data and empirical strategy

4.1 Answers and recorded AFAR citations

The descriptive sample contains all eligible recorded answers. It covers July 1, 2026, inclusive, through September 9, 2026, exclusive, under the saved UTC eligibility rules. The six source labels are ChatGPT, two sources of logged-out ChatGPT data that we call ChatGPT Source A and Source B, Claude, Gemini, and Google AI Overview. The three ChatGPT labels distinguish collection routes. They are not three independently identified model families.

Table 1 distinguishes the main denominators. Queries counts distinct exact query identifiers within a source. Observations counts answers, including repetitions. The external-citation and AFAR-citation columns both use all eligible answers for that source. The conditional page count uses only answers with an AFAR citation.

AFAR citation is much less frequent than external citation. External-citation rates range from 72.9% for Google AI Overview to near-complete coverage for ChatGPT. AFAR appears in 0.6% to 3.5% of answers across sources. Conditional on an AFAR citation, answers cite between 1.08 and 1.36 AFAR pages on average. These differences describe Nuno’s tested queries and collection routes; they are not population market shares.

The cited-page extract contains 523 distinct AFAR pages and 4,317 citation events. A citation event is an answer–page pair. Repeated appearances of a page in different answers produce separate events, while duplicate appearances within an answer do not. Distinct-page counts describe the breadth of cited content. Event counts give more weight to pages cited repeatedly.

Sponsorship is measured by a binary article characteristic. Sponsored equals one when an article is classified as sponsored and zero when it is classified as non-sponsored. Articles whose sponsorship status could not be assessed are identified separately. The classification reflects the saved evidence at collection and may contain measurement error. It does not establish an article’s commercial status at every answer date.

Table 2 reports the number of distinct AFAR pages cited by each source and the share with recorded sponsorship. ChatGPT cites 170 pages, of which 10.6% are sponsored. The

Table 1: AFAR citation by AI source

Source	Queries	Observations	External (%)	AFAR (%)	Pages cited
ChatGPT	2,904	34,537	99.9	2.2	1.10
ChatGPT logged out (Source B)	948	7,191	85.3	1.2	1.08
ChatGPT logged out (Source A)	2,358	28,996	89.7	1.8	1.36
Claude	2,790	32,960	98.6	3.5	1.18
Gemini	2,766	34,063	99.3	2.7	1.12
Google AI Overview	2,818	33,702	72.9	0.6	1.15

Notes: Queries counts distinct exact queries within source. Observations counts eligible AI answers, including repetitions. External is the percentage of all answers with a recorded external-domain citation. AFAR is the percentage citing an AFAR page. Pages cited is the mean number of distinct identified AFAR pages per answer conditional on citing AFAR. All rates use the full eligible answer population for that source. Sources can share queries, so query counts are not additive. ChatGPT has 34,526 externally citing answers out of 34,537; 99.9% is displayed to reserve 100.0% for exact completeness. Source: Nuno, July 1–September 8, 2026 (UTC).

corresponding share is 10.1% among Claude’s 276 pages and 18.8% among Gemini’s 149 pages. These shares describe sponsorship among cited pages. They do not measure the probability that a sponsored article receives a citation.

Table 2: Sponsorship among cited AFAR pages

Title	Pages	Sponsored (%)
ChatGPT	170	10.6
ChatGPT logged out (Source B)	42	9.5
ChatGPT logged out (Source A)	202	9.9
Claude	276	10.1
Gemini	149	18.8
Google AI Overview	70	12.9

Notes: Title identifies the AI source or collection route. Pages counts distinct AFAR articles cited within each source. Sponsored (%) is the percentage of those pages with a Sponsored value of one. The denominator includes all cited pages, including those with unassessed sponsorship status; the percentage therefore measures recorded sponsorship. Articles cited through several sources appear once in each source count. Source: Nuno and the frozen AFAR crawl.

Cited AFAR pages cover several article purposes. Table 3 shows 169 distinct lists and rankings, 144 features and news articles, 69 hotel and resort profiles, and 54 destination profiles. The table retains indexes and unclassified pages. Its sponsorship column is weighted by citation events, unlike the distinct-page percentages in Table 2. This prevents the prevalence of sponsorship among pages from being confused with its prevalence among recorded citations.

4.2 Inventory and matched candidates

The inventory covers AFAR’s published web addresses. Original, redirected, and canonical addresses are stored separately, and verified aliases are linked to a single article identity. Each record carries the article’s content category and its known sponsorship status. No new crawl or rematching is performed for this paper.

Table 3: Cited AFAR pages by category

Category	Unique Pages		Citation Events		Sponsored (% of category)
	Count	%	Count	%	
Lists & Rankings	169	32.3	2,207	51.1	10.7
Guides & Itineraries	40	7.6	434	10.1	10.1
Destinations	54	10.3	150	3.5	8.0
Hotels & Resorts	69	13.2	449	10.4	3.6
Other places	2	0.4	3	0.1	0.0
Tips & Advice	18	3.4	144	3.3	3.5
Features & News	144	27.5	741	17.2	14.4
Indexes & Directories	10	1.9	133	3.1	1.5
Unclassified	17	3.3	56	1.3	0.0

Notes: Unique Pages reports distinct identified AFAR pages. Its % column is each category’s share of all 523 cited pages. Citation Events counts answer–page pairs; a page cited in several answers generates several events. Its % column is each category’s share of all 4,317 citation events. Sponsored (% of category) is the percentage of citation events within that category linked to recorded sponsorship; it is event-weighted. Articles with unassessed sponsorship status remain in the denominator. Category shares may not sum to 100.0% after rounding. Lists compare or rank options; guides organise itineraries; destinations describe geographic areas; hotels profile an accommodation; other places describe individual non-hotel venues; tips provide practical advice; features cover narrative or news content; indexes organise links. Unclassified denotes insufficient evidence. Source: saved AFAR citation extract.

The full AFAR inventory contains 22,152 inventory records. Table 4 retains uncollected content, nonarticles, unclassified records, and records sharing a web address. Features and news account for 32.5% of inventory records, and other-place profiles for 30.4%. Sponsorship status is unassessed for 1,896 records, or 8.6%. These are characteristics of the collected inventory. The inventory is neither a census nor a probability sample of all AFAR output.

The candidate strategy starts from cited query–article pairs, which we call anchors. For each anchor, the saved procedure seeks comparisons from the same publisher and primary-purpose category for the same query. Comparison eligibility requires a publication date no later than the first answer to that query. A lexical search forms a pool of up to 300 candidates. Ranking combines query-token coverage and anchor similarity, with weights of 0.65 and 0.35. Article identifiers resolve ties.

Identity exclusions are applied before selecting the main sample. The saved procedure removes ambiguous identities and verified citation aliases from the original top-ten shortlists. It ranks the survivors without refilling the lists. The main k5 sample takes up to five surviving comparisons per supported anchor. A comparison reused by several anchors within the same query is counted once in the query’s candidate pool. It may still enter under another query.

Table 5 describes the 1,809 distinct AFAR comparison pages in that main sample. The largest groups are features and news, with 580 pages, and lists and rankings, with 567. No indexes or unclassified pages enter the matched profile. A page is reported as a high lexical match if any of its saved query-specific comparison roles passes the existing rule. This aggregation changes neither the stored flags nor the matching.

Table 4: AFAR inventory by category

Category	Observations	Share (%)	Sponsored (% of category)
Lists & Rankings	1,929	8.7	9.7
Guides & Itineraries	983	4.4	6.0
Destinations	1,126	5.1	5.1
Hotels & Resorts	1,731	7.8	1.8
Other places	6,743	30.4	0.0
Tips & Advice	685	3.1	9.6
Features & News	7,192	32.5	7.5
Indexes & Directories	14	0.1	7.1
Unclassified	1,749	7.9	0.1

Notes: Observations counts all 22,152 frozen AFAR inventory records. Share uses all records as the denominator. Sponsored (% of category) is the percentage of records within each category with recorded sponsorship. All records in that category, including observations with unassessed sponsorship status, remain in its denominator. The record retains uncollected content, nonarticles, unclassified records, and records sharing a web address; it is not deduplicated by web address. The collected inventory is not a census or a probability sample of AFAR output. Category definitions follow Table 3. Source: frozen September 11, 2026 inventory.

Table 5: Matched AFAR pages by category

Category	Observations	High match (%)	Partial match (%)	Sponsored (% of category)
Lists & Rankings	567	92.2	7.8	11.8
Guides & Itineraries	187	88.8	11.2	4.8
Destinations	118	51.7	48.3	11.9
Hotels & Resorts	260	75.8	24.2	3.5
Other places	13	15.4	84.6	0.0
Tips & Advice	84	58.3	41.7	9.5
Features & News	580	86.2	13.8	11.0
Indexes & Directories	0	—	—	—
Unclassified	0	—	—	—

Notes: Observations counts 1,809 distinct AFAR comparison pages in the frozen k5 profile. High match is the percentage of pages with at least one query–article pairing that passes the saved high-lexical-relevance rule. Partial match is the percentage of retained pages with no pairing that passes that rule. These groups partition the matched pages; Partial match is a reporting label for the complement of High match, not a separate semantic assessment. Sponsored (% of category) is the percentage of distinct matched pages within each category with recorded sponsorship. All percentages use all matched pages in the category, including observations with unassessed sponsorship status. Dashes denote undefined percentages for empty categories. Match labels are automated proxies, not independently validated semantic relevance. Source: saved AFAR comparison sample.

Coverage is measured before restricting to five comparisons. Table 6 uses all 1,449 cited AFAR query–article pairs and allows up to ten surviving comparisons. Of these, 1,342 pairs have at least one match. The mean number of comparisons, including zero-match pairs, is 7.74. Ten matches survive for 62.3% of pairs, and 81.8% have at least one comparison flagged as high lexical relevance. These rates describe support under the saved rules.

High lexical relevance is not equivalent to semantic suitability. The saved rule combines token coverage with checks for detected geography or subject contradictions. A targeted project review found comparisons that passed this rule despite failing the query’s substantive constraints. The review was not a random sample, so it does not estimate a failure rate for AFAR. It nevertheless shows that the high-relevance restriction cannot be treated as independently validated matching.

Table 6: Matching coverage for AFAR

Measure	AFAR
Cited query–article pairs	1,449
Pairs with at least one match	1,342
Average matched articles	7.74
Pairs with ten matches (%)	62.3
Pairs with a high match (%)	81.8

Notes: The denominator is all 1,449 cited query–article pairs before matching exclusions. An article can appear for several queries, but only once per query across sources and repetitions. Comparisons are counted after identity exclusions and before the five-comparison cap, with a maximum of ten per pair. Averages include zero-match pairs. High match requires at least one comparison with the saved high-lexical-relevance flag; the anchor need not pass. Comparisons can support several anchors. These measures describe support, not semantic accuracy. Source: frozen matching record.

4.3 Observable differences and estimation support

Table 7 compares sponsorship, article length, age, and category shares between cited and matched AFAR articles. The sample contains 1,342 cited and 4,819 matched query–article observations. Each observation receives equal weight. An article can appear under several queries, but only once within each query and role. The table reports unadjusted means and differences, with standard errors clustered by query and article.

Sponsorship shares are similar across cited and matched articles. Among observations with assessed sponsorship status, 10.4% of cited articles and 10.2% of matched articles have recorded sponsorship. The difference is 0.2 percentage points and is not statistically significant at the 5% level. This comparison describes sponsorship among the two groups; it does not estimate the effect of sponsorship on citation probability.

Cited articles are newer than matched articles. Their mean age is 2.7 years, compared with 3.6 years for matched articles. The difference is statistically significant at the 1% level. Mean article length is 1,690.6 words for cited articles and 1,735.4 words for matched articles. This difference is not statistically significant at the 5% level. Similar sponsorship shares therefore do not imply that the groups are comparable in every characteristic.

Category shares are also similar across the two groups. Lists and rankings account for 44.8% of cited observations and 42.6% of matched observations. Features and news account for 22.5% and 23.0%, respectively. None of the tested category differences is statistically significant at the 5% level. These comparisons describe broad content categories and do not establish that individual articles are equally relevant to a query.

Missing values are excluded separately for each balance variable. Sponsorship is assessed for 1,339 cited and 4,815 matched observations. Age is available for 1,253 cited observations and all 4,819 comparisons. The latter difference partly reflects the publication-date requirement for comparison eligibility. The balance table therefore reports variable-specific comparisons rather than a single complete-case sample.

Table 7: Cited and matched AFAR articles

Variable	Cited	Matched	Mean diff.
Overall sponsorship (%)	10.4 (30.51)	10.2 (30.21)	0.2 [0.12]
Length (words)	1,690.565 (1,247.29)	1,735.423 (1,089.79)	-44.859 [-0.40]
Age (years)	2.650 (2.57)	3.640 (2.81)	-0.990*** [-5.23]
<i>By article category:</i>			
Lists & Rankings	44.8 (49.75)	42.6 (49.45)	2.2 [0.53]
Guides & Itineraries	10.1 (30.19)	11.5 (31.93)	-1.4 [-0.55]
Destinations	5.5 (22.83)	4.8 (21.32)	0.7 [0.77]
Hotels & Resorts	14.3 (35.03)	15.4 (36.08)	-1.1 [-0.46]
Other places	0.2 (4.72)	0.3 (5.19)	0.0 [-0.40]
Tips & Advice	2.5 (15.72)	2.5 (15.52)	0.1 [0.09]
Features & News	22.5 (41.78)	23.0 (42.10)	-0.5 [-0.17]
Indexes & Directories	0.0 (0.00)	0.0 (0.00)	0.0 —
Observations	1,342	4,819	

Notes: Means with sample standard deviations in parentheses; cited-minus-matched differences with t-statistics in square brackets. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$. Sponsorship and category means use percentages; their differences and SDs use percentage points. Length and age means/differences use three decimals. Age is years on September 9, 2026. Observations are equally weighted query–article roles, deduplicated within query. Tests have no fixed effects and use query/article two-way clustering with finite-sample corrections. Unknowns are excluded separately by variable. Assessed counts (cited, matched) are 1,339 and 4,815 for sponsorship, 1,332 and 4,819 for length, and 1,253 and 4,819 for age. Category indicators use all observations. A dash denotes an undefined test for a constant variable. No category is omitted as a reference group.

The regression sample has a different unit from the balance table. It contains 25,688 query–article–source cells, representing 350,122 response–article opportunities and 41,390 distinct eligible answers. The same answer supplies an opportunity for each article in its query’s candidate pool. The sample covers 748 queries and 2,022 distinct AFAR articles. It contains 3,825 recorded citation events. Google AI Overview remains outside the frozen estimation sources, although it appears in the descriptive tables.

Within-query sponsorship variation is limited to part of the sample. Of the 748 included queries, 331 contain both assessed sponsored and assessed non-sponsored article states. The remaining queries do not provide the same direct within-query contrast. There are also 401 opportunities with unassessed sponsorship status, or 0.1% of the main opportunity sample. We retain them with a separate indicator and report an assessed-only sensitivity check.

5 Results

5.1 Model and estimand

The analysis focuses on the Sponsored dummy. Let Y_{qas} denote the proportion of answers from source s to query q that cite candidate article a . Let S_a equal one for articles classified as sponsored and zero for articles classified as non-sponsored. Let M_a identify observations whose sponsorship status could not be assessed. For these observations, we code $S_a = 0$ and $M_a = 1$; they are not treated as known non-sponsored articles. The model is

$$Y_{qas} = \beta S_a + \theta M_a + \alpha_q + \delta_s + \varepsilon_{qas}. \quad (1)$$

Query fixed effects, α_q , account for differences across queries. Source fixed effects, δ_s , account for average differences across sources and collection routes. The coefficient of interest is β , the association between sponsorship and citation. The missing-status indicator separates unassessed observations from the assessed reference group. Its coefficient is a control and is not reported in the regression tables.

Each grouped cell is weighted by its number of answers. This gives equal weight to the underlying response–article opportunities. Standard errors cluster by query and article. We use cluster-count corrections and minimum-cluster t degrees of freedom. The grouped-cell finite-observation correction is disabled. An explicit expansion to all 350,122 opportunities reproduces the preferred coefficients and covariance matrix, which verifies that grouping does not change the reported inference.

The coefficient is a controlled association in the constructed candidate sample. Article inclusion depends on an observed cited anchor and surviving comparisons. A zero outcome means no recorded citation in the eligible answer extract. It does not show that the system retrieved and rejected the article. The model therefore does not estimate a causal effect of sponsorship or the citation probability of an arbitrary AFAR inventory page.

5.2 Main sponsorship association

Table 8 reports the sponsorship coefficient across three specifications. Column (1) includes no query or source fixed effects. The coefficient rounds to 0.0 percentage points and is not statistically significant at the 5% level. Sponsored and assessed non-sponsored opportunities both have a mean citation rate of 1.1% after rounding.

The preferred estimate is negative and not statistically significant at the 5% level. Column (3) reports a coefficient of -0.1 percentage points after controlling for query and source fixed effects. All three columns use 350,122 response–article opportunities. The results do not provide statistically significant evidence of a sponsorship penalty for AFAR.

Adding source fixed effects does not change the reported sponsorship coefficient relative to query fixed effects alone. Each query has the same candidate articles across its available sources. Within each query and source, articles share the same answer-count exposure. Source fixed effects can therefore account for average citation differences without changing the common sponsorship coefficient. This does not imply that the sponsorship association is identical across sources.

Table 8: Sponsorship and AFAR citation

Variable	(1) Citation	(2) Citation	(3) Citation
Sponsored	0.0 (0.4)	-0.1 (0.3)	-0.1 (0.3)
Query FE	No	Yes	Yes
Source FE	No	No	Yes
Answers citing AFAR (%)	8.3	8.3	8.3
Observations	350,122	350,122	350,122

Notes: The dependent variable is Citation. The Sponsored row reports the sponsorship coefficient, with query/article two-way clustered standard errors in parentheses. Coefficients and standard errors are in percentage points. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$. Answers citing AFAR (%) is the percentage of distinct answers in the estimation sample that cite any AFAR page: 3,415 of 41,390 answers. Each answer is counted once. This answer-level summary differs from the regression outcome, which records citation to an individual candidate article. FE denotes fixed effects. Observations counts response–article opportunities. Unassessed sponsorship status is controlled for separately where present; its coefficient is not displayed. Unless otherwise stated, grouped means receive answer-count weights, equivalent to equal weights on the underlying opportunities. Inference uses cluster-count corrections, minimum-cluster t degrees of freedom, and no grouped-cell observation correction. The estimates are conditional associations, not causal effects. All columns use the same sample with up to five comparisons. Column (1) is unadjusted. Column (2) includes query fixed effects. Column (3) adds source fixed effects.

5.3 Sample and source sensitivity

Table 9 reports sponsorship estimates across four groups of robustness checks. Panel A varies the number of comparison articles. The sponsorship coefficients are -0.1, 0.0, -0.1, and -0.1 percentage points for one, three, five, and ten comparisons. None is statistically significant

at the 5% level. These samples retain the matching procedure but differ in the number of candidate opportunities.

Panel B holds the cited anchors fixed while varying the number of comparison articles. It uses anchors with ten surviving comparisons in every specification. None of the sponsorship coefficients is statistically significant at the 5% level. These checks reduce differences in anchor composition but do not establish that all comparison articles are equally relevant to the query.

Panel C reports separate estimates for each source. The ChatGPT coefficient is -0.2 percentage points. The coefficients for ChatGPT logged out (Source A) and ChatGPT logged out (Source B) are -0.1 and 0.0 percentage points, respectively. Claude has a coefficient of -0.3 percentage points, while Gemini has a coefficient of 0.1 percentage points. None is statistically significant at the 5% level. Source fixed effects are omitted in these specifications because each uses one source.

Panel D checks sensitivity to matching quality, query weights, and missing sponsorship status. Restricting the sample to high lexical relevance gives a coefficient that rounds to 0.0 percentage points. It is not statistically significant at the 5% level. This restriction uses the saved automated matching rule and does not independently establish semantic relevance.

Changing weights and excluding unassessed sponsorship status give similar results. Equal total weight per AFAR query yields a coefficient of -0.2 percentage points. Excluding 401 opportunities with unassessed status yields -0.1 percentage points, using 349,721 observations. Neither coefficient is statistically significant at the 5% level. The preferred specification retains unassessed observations with a separate control.

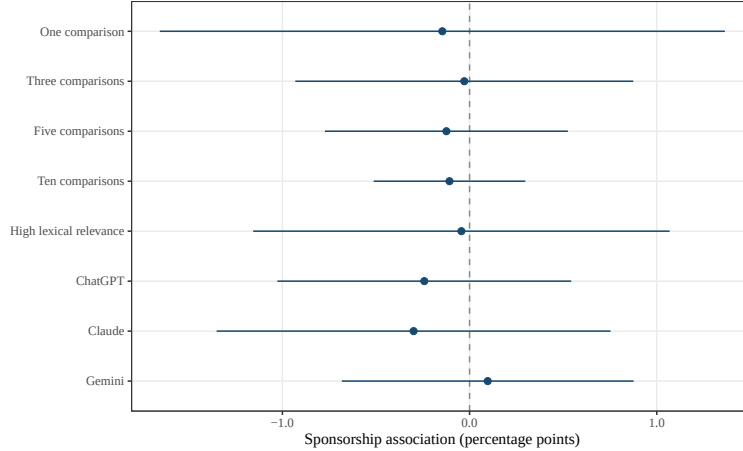
Figure 1 summarises selected sample and source estimates on a common percentage-point scale. None of the displayed sponsorship coefficients is statistically significant at the 5% level. This does not establish that the coefficients are equal across specifications or support a causal interpretation.

Table 9: Robustness of AFAR sponsorship estimates

Specification	Estimate (pp)	p-value	Observations
<i>A. Comparison counts</i>			
One comparison	-0.1	0.850	138,225
Three comparisons	0.0	0.951	249,568
Five comparisons	-0.1	0.708	350,122
Ten comparisons	-0.1	0.603	545,113
<i>B. Common cited anchors</i>			
Common: one	-0.6	0.468	83,593
Common: three	-0.3	0.513	154,027
Common: five	-0.3	0.435	225,081
Common: ten	-0.2	0.364	394,593
<i>C. Sources</i>			
ChatGPT	-0.2	0.545	86,347
ChatGPT logged out (Source A)	-0.1	0.809	73,563
ChatGPT logged out (Source B)	0.0	0.928	21,221
Claude	-0.3	0.577	82,757
Gemini	0.1	0.807	86,234
<i>D. Sample and weighting checks</i>			
High lexical relevance	0.0	0.939	138,900
Equal query weight	-0.2	0.609	350,122
Assessed sponsorship only	-0.1	0.708	349,721

Notes: Each estimate is the Sponsored coefficient in percentage points. *** p<0.01, ** p<0.05, * p<0.10. P-values use query/article two-way clustering and minimum-cluster t degrees of freedom. Observations counts response–article opportunities. All models include query fixed effects and source fixed effects when multiple sources are present. Panel A varies the comparison cap. Panel B holds cited anchors fixed across caps. Panel C uses separate sources. Panel D restricts to high lexical relevance, gives equal total weight to each query, or excludes unassessed sponsorship status. The high-match rule is an automated proxy, not independently validated semantic relevance. Other samples retain a missing-status control where needed. Unless otherwise stated, grouped means receive answer-count weights. These are conditional associations, not causal effects.

Figure 1: Sponsorship associations across AFAR samples



Notes: Points are AFAR-only sponsorship coefficients in percentage points. Bars show pointwise 95% confidence intervals with query/article two-way clustering. The dashed line denotes zero. All models retain query and source effects where source varies. High lexical relevance is the saved automated rule, not independently validated semantic matching. Samples are citation-conditioned. The figure reports selected specifications from Table 9; it does not display every specification.

5.4 Interpretation and limitations

The estimand is the difference in citation probability per response–article opportunity between sponsored and assessed non-sponsored AFAR candidates, conditional on query and source. The design compares articles that compete for the same query, so the coefficient cannot be driven by sponsored content concentrating in topics that AI sources cite more or less often. The results show no distinction between the two groups. Sponsored and assessed non-sponsored opportunities share the same observed citation rate of 1.1%. The preferred coefficient is -0.1 percentage points with a query/article two-way clustered standard error of 0.3 percentage points, which implies a 95% confidence interval of approximately -0.7 to 0.5 percentage points. The estimate without fixed effects rounds to 0.0 percentage points, so the result does not depend on the fixed-effects specification. Every robustness coefficient in Table 9 lies within the preferred interval, and none is statistically significant at the 5% level. The data therefore provide no indication that the sources discount sponsored AFAR content when they select citations.

The estimate is an observational association obtained under a fixed comparison strategy. It describes the relationship between measured sponsorship and citation for AFAR content that competes for Nuno’s tested queries. It does not identify the causal effect of a change in an individual article’s sponsorship status. The confidence interval also states the precision of the result: a penalty smaller than about 0.7 percentage points would not be detectable with the current sample. This paper is the first study in a planned series, and its scope is set by what the current data can support.

Four features of the data define that scope. First, the sample is selected: queries enter through cited anchors, articles enter through matching eligibility and ranking, and the collected inventory is incomplete. The fixed effects absorb query-level and source-level citation propensity but do not reverse this selection, so the estimates describe AFAR pages that

compete for the tested queries rather than every AFAR page. Second, comparison articles are matched on shared query terms and article purpose. This does not guarantee equal fit to every requirement in a query, and the regression omits direct controls for content quality, access, and distribution. The statistically significant age difference between cited and matched pages in Table 7 is one measured example of a remaining difference. Third, sponsorship status is recorded at the time of data collection. Some observations are unassessed, and status may have changed between publication and the answer date. Fourth, the outcome records whether an AFAR page appears in the saved citation extract. As Liu, Zhang, and Liang (2023) show, a citation establishes inclusion in the answer, not that the page determined the recommendation or received reader traffic.

Further studies can test the result under stricter identifying assumptions. Independent semantic assessment of each candidate would validate the comparison sets. Repeated measurements around documented changes in sponsorship status would support within-article comparisons that the present cross-sectional design cannot. Extending the design to other publishers would establish whether the pattern generalises beyond AFAR. Each extension tests whether the absence of a penalty persists under stricter conditions. None is required to support the descriptive finding reported here.

6 Conclusion

This paper documents the relationship between sponsorship and recorded AI citation for AFAR. It compares sponsored and non-sponsored articles within query-specific candidate pools and gives equal weight to response–article opportunities. The analysis keeps every denominator explicit: full-answer citation rates, distinct cited pages, citation events, inventory records, matched pages, and response–article opportunities.

The central finding is that AI sources show no distinction between sponsored and non-sponsored AFAR content. Articles of both kinds that compete for the same query are cited at the same observed rate of 1.1%. The preferred coefficient is -0.1 percentage points and is not statistically significant at the 5% level. The estimate is unchanged whether or not query and source fixed effects are included. Its sign and magnitude are stable across every sensitivity check, including alternative comparison counts, fixed anchors, stricter lexical matching, equal query weights, assessed-only observations, and each of the five sources estimated separately. No specification in this study yields a sponsorship coefficient that is distinguishable from zero at conventional significance levels.

Within this sample, sponsorship status carries no measurable citation cost. In the balance comparison, cited and matched articles differ significantly in age but not in sponsorship share. This paper is the first in a series of studies on how AI sources select and cite travel content.

References

- Aggarwal, P., V. Murahari, T. Rajpurohit, A. Kalyan, K. Narasimhan, and A. Deshpande. 2024. GEO: Generative Engine Optimization. Proceedings of KDD 2024. Research paper.
- AFAR. 2026. About AFAR. Publisher description and editorial standards. Accessed September 14, 2026. Publisher source.
- Lewis, P., E. Perez, A. Piktus, F. Petroni, V. Karpukhin, N. Goyal, H. Küttler, M. Lewis, W.-t. Yih, T. Rocktäschel, S. Riedel, and D. Kiela. 2020. Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks. Advances in Neural Information Processing Systems 33. Research paper.
- Liu, N. F., T. Zhang, and P. Liang. 2023. Evaluating Verifiability in Generative Search Engines. Findings of EMNLP 2023, 7001–7025. Research paper.