

What Earns a Citation in Google's AI Overviews? Ten Signals Tested on 370 Ranking Pages

Eric Lancheres · On-Page.ai Research | July 16, 2026

ABSTRACT

Our previous study found that the page-one results Google's AI Overviews cite are no more original than the ones they skip [1], which left the obvious question open: if not originality, what does earn the citation? We tested ten candidate signals on the same 370 page-one results (50 keywords, ten verticals), joining authority and demand metrics from the Ahrefs API [3] and page age from the Internet Archive [4] onto the same-day citation labels: Domain Rating, URL Rating, page backlinks, referring domains, estimated organic traffic, organic keyword count, brand search volume, page age, keyword-matching domains, and originality as the published control. Under the same within-SERP comparison and permutation testing as the previous study, none predicted citation. Cited and uncited pages had median Domain Ratings of 88 and 87 ($p > .99$); backlinks, referring domains, traffic, keyword counts, brand volume ($p = .38$), page age ($p = .60$), and keyword-match domains were all indistinguishable. The SERP's top page by each signal was cited at or below chance rates on seven of nine measures. Two context findings frame the null: page one is an authority monoculture (median DR 87.5; 44% of results at DR 90+), leaving little authority variation to reward, and 66% of the off-SERP pages the Overviews cite have an estimated organic traffic of zero. Within page one, citation selection by Google's AI appears indistinguishable from random with respect to every signal the industry currently optimizes.

KEY FINDINGS

- None of the ten signals predicted which page-one results the AI Overview cited: not domain authority, not backlinks, not brand size, not page age, not keyword-matching domains, and (as previously published) not originality.

- Authority can't separate what ranks: page one is already an authority monoculture, with almost half of results at Domain Rating 90 or above.
- Two-thirds of the pages AI Overviews cite from beyond page one have an estimated organic traffic of zero. The AI's sources are largely pages search traffic never visits.
- Video on page one was the one format the Overview avoided citing (15% vs. 47% for articles), even though YouTube dominates its off-SERP citations.

1. Background

This is the third study in a series. The first measured how much new information top-ranking pages add beyond the pages they rank alongside [2]. The second captured every source cited by Google's AI Overviews and ChatGPT across 50 keywords and found citation decoupled from that originality measure [1]: the cited ranking pages were no more original than the skipped ones, and the SERP's most original page was cited less often than chance.

That result invites a natural response, because the industry holds several other theories of what earns AI citations: authority (“AI trusts high-DR sites”), links (“citations follow the link graph”), brand (“AI cites brands it knows”), longevity (“established pages win”), and keyword-domain-match. Each is a measurable claim about the same event: a page on page one either was or was not cited by the AI Overview shown above it. This study tests all of them on the same data, with the same method, so the answers are directly comparable to the originality result they respond to.

2. Method

Population and labels. The unit of analysis is the page-one organic result: 370 distinct pages across the 50 SERPs captured on July 15, 2026 (US, English) for the second study [1], with the same capture supplying the label: whether the AI Overview shown on that SERP cited the page (173 cited, 197 not). Because authority and demand metrics exist for pages our originality pipeline could not score (video pages, forum threads), this study includes the full 370, which is 58 more pages than the originality analysis could cover.

Signals. For every page we joined: Domain Rating, URL Rating, page-level backlinks, referring domains, estimated organic traffic, and organic keyword count from the Ahrefs API, retrieved July 16, 2026 [3]; brand search volume (the US monthly search volume of the site's registrable domain name as a query, e.g. the volume of “nerdwallet”, as a measurable proxy for brand familiarity); page age (years since the page's first Internet Archive snapshot [4], available for 205 of 370 pages); keyword-match domain (whether the domain name matches the query exactly or partially, computed from the domain label); content format (article, video, forum, social, classified by host); and the Information Gain Score from the previous study as the published control [1, 2].

Analysis. Identical to the previous study: every comparison is within-SERP, so keyword difficulty and vertical drop out. For each signal we report cited and uncited medians, a two-sided permutation test on the median difference that shuffles citation labels within each keyword (10,000 shuffles), and the chance-baseline cut: how often the SERP's unique leader on that signal was among the cited pages, against random selection of the same citation counts. With ten signals and three cuts each, a couple of cells will sit above chance by luck alone; we flag rather than headline them (§3.4).

3. Results

3.1 The league table

Table 1 is the study. Every row is a theory of AI citation the industry currently invests in; every row is a null.

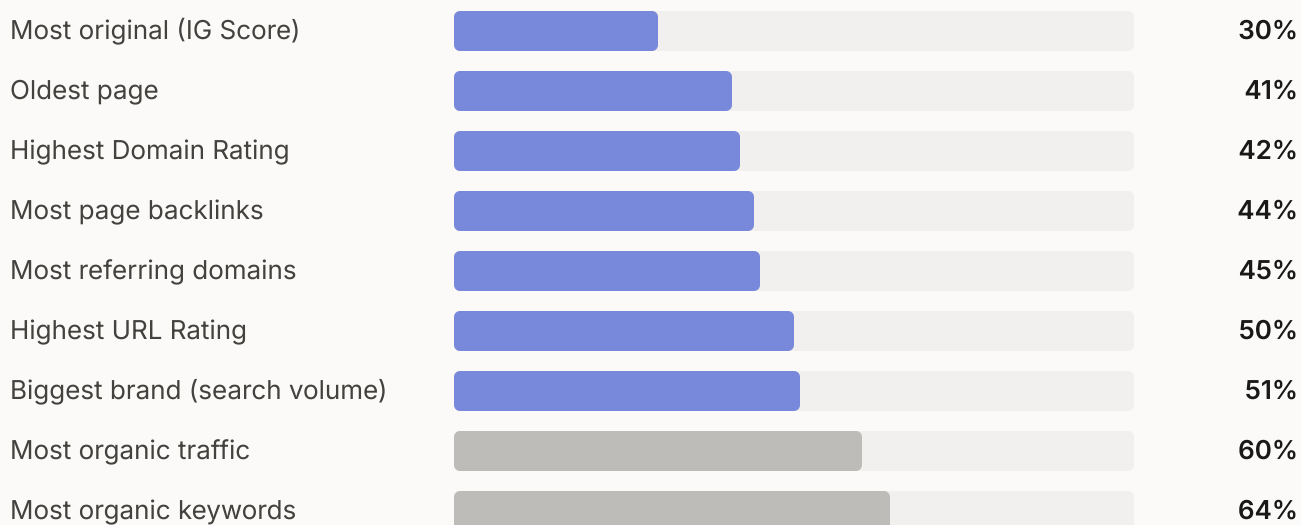
Signal	n	Cited median	Uncited median	p	Top cited (vs chance)
Domain Rating	370	88	87	>.99	42% (vs 46%)
URL Rating	370	4.5	4.5	>.99	50% (vs 45%)
Page backlinks	370	8	10	.73	44% (vs 46%)
Page referring domains	370	5	5	>.99	45% (vs 46%)
Page organic traffic	370	226	214	>.99	60% (vs 46%)
Page organic keywords	370	22	22	>.99	64% (vs 46%)
Brand search volume	353	59,125	35,289	.38	51% (vs 46%)
Page age (years)	205	3.3	3.6	.60	41% (vs 48%)
Keyword-match domain	370	n/a	n/a	>.99	45% (vs 51%)
Originality (IG Score), control	312	52	55.5	.07	30% (vs 46%)

Table 1. Ten signals, within-SERP cited vs. uncited comparison across 370 page-one results. p-values from two-sided permutation tests shuffling citation labels within keywords (10,000 shuffles). 'Top cited' = share of eligible SERPs where the signal's unique leader was among the cited pages (chance baselines 45–51%). Keyword-match domains were too rare for medians; the permutation test uses the full 0/1/2 coding.

Read the table twice. First for the medians: cited and uncited pages are near-identical on every signal: Domain Rating 88 vs. 87, five referring domains each, twenty-two ranking keywords each. Second for the last column: being the SERP's single best page by a signal (its most authoritative, most linked, biggest-brand, oldest) earned citation at roughly chance rates on seven of nine measures, and below chance for originality.

Being the SERP's best page, by any measure, doesn't earn the citation

How often the page-one result that leads its SERP on each signal was among the AI Overview's citations. Random selection of the same citation counts would land near 46%. Bars near that mark mean the signal buys nothing.



Share of SERPs where the signal's unique leader was cited (% of eligible SERPs). Gray bars: the two cells above chance, discussed and cautioned in §3.4.

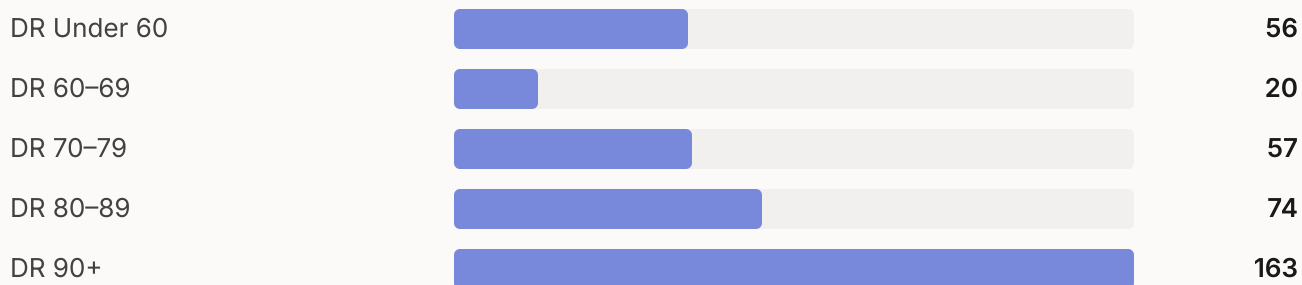
Figure 1. For each signal, the share of eligible SERPs (unique leader on that signal, at least one cited and one uncited scored page) where the leading page was among the AI Overview's citations. Chance baselines per signal sit at 45–51% (mean $\approx 46\%$). The most original page was cited in 30% of SERPs; the highest-authority page in 42%.

3.2 Authority in detail

The authority null deserves its context, because it is conditional: these are pages that already rank. Google's first page, in this sample, is an authority monoculture: median Domain Rating 87.5, interquartile range 73–92, 44% of results at DR 90+. Whatever role authority plays in reaching page one (this study does not test that), there is very little authority variation left within page one for a citation mechanism to reward. And within that compressed range, it rewards none of it: the highest-DR page on the SERP was cited 42% of the time against a 46% chance baseline.

Page one is already an authority monoculture

Domain Rating of the 370 page-one results. Nearly half sit at DR 90 or above. There is little authority variation left for the AI to prefer.



Pages per Domain Rating band (count of page-one results, n=370).

Figure 2. Domain Rating distribution of the page-one results in the sample: median 87.5, interquartile range 73–92, and 163 of 370 pages (44%) at DR 90+. Within this compressed range, cited and uncited pages are indistinguishable on authority (Table 1).

Page-level link metrics, the version of the authority theory closest to “citations follow links,” fare no better. The median cited page has 8 backlinks and 5 referring domains; the median uncited page has 10 and 5. If the AI Overview's citation selection consults the link graph, it hides it well.

3.3 Brand, age, and keyword domains

Brand search volume shows the largest raw gap in the table: cited pages sit on domains whose brand names draw a median 59,125 US searches a month, against 35,289 for uncited pages. It still fails the within-SERP test decisively ($p = .38$), with the biggest brand on the SERP cited at 51% versus a 46% baseline. Big brands are simply everywhere on page one; the AI does not favor them further. Page age is flat (3.3 vs. 3.6 years, $p = .60$; oldest page cited at 41%, below its 48% baseline, so no seniority bonus). Domains matching the query exactly or partially were rare in the sample and showed nothing ($p > .99$): whatever exact-match domains are doing in 2026, earning AI Overview citations is not it.

3.4 Two cells above chance

Honesty requires pointing at the two gray bars in Figure 1: the SERP's top page by estimated organic traffic was cited in 60% of eligible SERPs, and by organic keyword count in 64%, both above their $\approx 46\%$ baselines, while their median comparisons are perfect nulls (226 vs. 214

visits; 22 vs. 22 keywords; both $p > .99$). With thirty cells tested, one or two will land there by luck, and a leader-only effect without any distributional difference is exactly the shape luck takes. We report them as candidates for replication in the next quarterly capture, not as findings. If they replicate, the interpretation would be modest: the page already winning the most search visibility picks up a small citation edge: visibility begetting visibility, not a lever anyone can pull directly.

3.5 Format on page one

Format	Cited	On page one	Citation rate
Article	152	323	47%
Forum thread	18	32	56%
Video page	2	13	15%
Social	1	2	50%

Table 2. Citation rate by content format among page-one results. Format classified by host (video platforms, forum platforms, social platforms, otherwise article).

The one departure from uniformity is format. Video pages ranking on page one were nearly never cited by the Overview (2 of 13), a striking inversion, given that YouTube is the single largest destination of AI Overview citations overall [1]. When the Overview cites video, it overwhelmingly reaches beyond the visible results to do so. Forum threads on page one, by contrast, were cited slightly more often than articles (56% vs. 47%), directionally consistent with the previous study's finding that community content is a favored destination.

3.6 The zero-traffic citations

Finally, the enrichment throws unexpected light on where the other three-quarters of Overview citations go [1]. The off-SERP pages the AI cites carry respectable domain authority (median DR 83), but two-thirds of them have an estimated organic traffic of zero, and 73% draw fewer than ten estimated visits a month. These are pages organic search effectively never sends anyone to. For the “links and rankings are dying” conversation, this cuts both ways: the AI answer layer is surfacing pages the ranking layer ignores, and a citation from it is, on this evidence, not a proxy for the page having search traffic, or any obvious way to earn it.

Two-thirds of the pages AI cites from beyond the SERP get no measurable organic traffic

Share of pages with an estimated organic traffic of zero. Even on page one, deep URLs often carry no measurable traffic of their own, but the off-SERP cited set is in another category.

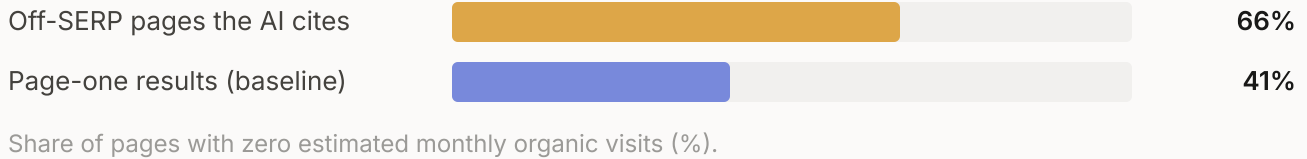


Figure 3. Share of pages whose Ahrefs-estimated organic traffic is zero: 270 of 411 off-SERP AI-Overview-cited pages (66%; 73% under ten visits/month) against 153 of 370 page-one results (41%). Traffic estimates are model-based and undercount low-volume queries. The comparison between the two groups is the point, not the absolute levels.

4. Discussion

Three studies now point at one conclusion from three directions. Originality does not predict which ranking pages the AI Overview cites [1]. Neither, this study adds, does authority, links, traffic, brand, age, or domain naming. And the citations themselves are unstable on the ChatGPT side, where 85% did not recur across identical runs [1]. The simplest model consistent with all of it: conditional on ranking, today's AI citation selection is effectively random with respect to every property of a page that SEO currently optimizes. The mechanism presumably optimizes something: passage relevance to the synthesized answer, retrieval convenience across fanned-out sub-queries [5]. But none of it projects onto the levers practitioners are being sold.

Two boundaries on that claim. First, everything here is conditional on page one: authority may well govern who gets to rank (our data cannot say), and a quarter of Overview citations do land on page one, so ranking remains the one reliable, measurable path toward citation. Second, a null on ten signals is not proof that no signal exists; it is proof that the ten most-recommended ones are not it. The honest reading for practitioners: treat any service promising to “get you cited by AI” through authority, brand, or content tactics as making a claim this data contradicts, and treat the AI answer layer, for now, as weather rather than a market.

5. Limitations

The population is pages that already rank on page one for 50 hand-built US keywords, captured on one day. Range restriction is the price of the clean within-SERP design, and these results say nothing about how the tested signals affect ranking itself. All third-party metrics are estimates: link-index coverage varies by site, traffic estimates undercount low-volume queries (the zero-traffic shares in §3.6 compare groups under the same estimator, but the absolute levels inherit its blind spots), and brand volume uses the domain label as the brand query, a proxy that mislabels some multi-word brands. Page age covers the 55% of pages the Internet Archive had crawled; never-archived pages skew small, so the age null is the least powered row in Table 1. Citation labels are AI Overviews only; ChatGPT's citations overlapped page one too rarely to test [1]. And with thirty tested cells, the two above-chance results in §3.4 should be treated as multiple-comparisons candidates until the next capture replicates or retires them.

6. SEO Implications

For teams deciding where AI-citation effort goes:

- There is currently no buyable lever for AI Overview citations. Authority, links, brand, age, exact-match domains, originality: none separated cited from uncited pages that already rank. Budget accordingly, and ask any vendor claiming otherwise for their within-SERP numbers.
- Ranking is still the one path that measurably helps: a quarter of Overview citations land on page one, and everything upstream of ranking, including the originality gradient from our first study [2], remains yours to optimize.
- If video answers your query's intent, page-one video is not how it gets cited; the Overview reaches off-SERP for video. A channel strategy and a ranking strategy are different investments; this data says don't conflate them.
- Measure citation claims with recurrence and baselines. The two above-chance cells in this study came with flat distributions and thirty tests' worth of chances to appear. The same skepticism applies to any tool's citation dashboard.

Acknowledgments

The authority, link, traffic, and search-volume data in this study were retrieved through API access provided by [Ahrefs](#), whom we thank, both for the access and for maintaining an index that makes measurement like this possible. Ahrefs had no role in the design, analysis, or conclusions of this study, including the null result for its own Domain Rating metric, and did not review this paper before publication.

References

1. Lancheres, E. (2026). “Do AI Assistants Cite Original Content? 793 AI Citations Measured Against the Pages That Rank.” On-Page.ai Research. api.on-page.ai/research/ai-citation-study
2. Lancheres, E. (2026). “Information Gain in Google's Top-Ranking Results: A Measurement of 150 Pages.” On-Page.ai Research. api.on-page.ai/research/information-gain-study
3. Authority, link, and traffic metrics retrieved via the Ahrefs API (batch analysis and keyword volumes), July 16, 2026. docs.ahrefs.com
4. Page first-capture dates via the Internet Archive Wayback Machine availability API. archive.org/help/wayback_api.php
5. “AI Mode in Google Search: Updates from Google I/O 2025,” Google (describes the “query fan-out” technique). blog.google/products-and-platforms/products/search/google-search-ai-mode-update

Cite this study: Lancheres, E. (2026). What Earns a Citation in Google's AI Overviews? Ten Signals Tested on 370 Ranking Pages. On-Page.ai Research. <https://api.on-page.ai/research/ai-citation-predictors>