NOVAXILABS · WHITE PAPER SERIES

Beyond Rankings

What actually gets a page cited in AI answers — and why the 2026 evidence breaks the assumption that position decides visibility.

Author Dr. Adrian Whitmore **Published** 2026-11-08 **Data as of** 2026-05-31

Edition **v1.0** · Novaxilabs Research

Document Information

TITLE	Beyond Rankings: What Actually Gets a Page Cited in AI Answers
SERIES	Novaxilabs White Paper Series
EDITION	v1.0 — first release
PUBLISHED	2026-11-08
DATA AS OF	2026-05-31 (latest source data cited in this edition)
AUTHOR	Dr. Adrian Whitmore — Independent GEO research consultant. PhD in Information Retrieval (University of Cambridge).
CONTACT	info@novaxilabs.com

> **Disclaimer.** This paper is a **secondary research synthesis**. Every quantitative figure is taken from a published third-party study, and each is cited with its sample size and measurement window. The paper does not present original monitoring data, does not claim access to engine internals, and does not assert that any single study generalises to every category or market. Where two credible studies disagree, both are reported and the disagreement is discussed rather than resolved. Figures are directional and dated; engine behaviour changes faster than studies replicate.

Revision History

EDITION	DATE	AUTHOR	CHANGE
v1.0	2026-11-08	A. Whitmore	First release. Synthesises public measurement data published between May 2024 and May 2026 and argues the ranking–citation decoupling.

Contents

The executive summary alone answers “so what”. Every number in this paper resolves to a citation in Appendix A.

§ Executive Summary

01 The ranking assumption

02 Scope, sources and method

03 Five findings from the 2026 data

04 What this changes for brands

05 Limitations

06 What to do next

A Sources & References

B About the Author

C About Novaxilabs

§ Executive Summary

THESIS

In 2026, being cited by a generative engine is no longer a function of where a page ranks. It is a function of whether the page is **structurally extractable**, **fresh enough to be retrieved**, and **present in enough engine sets** — three properties that can be engineered, measured, and lost.

Twenty years of search practice rest on one assumption: rank determines visibility. That assumption was never tested against answer engines at scale, and the measurements that do test it have only arrived between mid-2025 and mid-2026. They are uncomfortable. Of roughly four

million URLs cited inside Google AI Overviews, only **37.1% also ranked in the top ten** for the query, and **36.7% did not rank in the top hundred at all** (Ahrefs, 863,000 keywords, March 2026) [1]. Across 100,000 prompts, only **11% of cited domains** appeared in both ChatGPT and Perplexity answers (Profound, July 2025) [4]. And in Google AI Mode, **93% of sessions end without a single outbound click** (GEO Hub, 2026) [9].

The conclusion is not that rankings stopped mattering. It is that ranking has become a *relevance signal* inside a much wider candidate pool, not an entry ticket. Brands that keep budgeting against position while the citation decision happens elsewhere are measuring the wrong variable. The first action is to switch the unit of measurement from rank to cited-source share, per engine, and to treat extractability and freshness as the two levers that actually move it.

---

01 BACKGROUND

The ranking assumption

Classic search optimisation is built on a single causal chain: a page ranks, therefore it is seen, therefore it is clicked. Every reporting convention in the industry — average position, share of voice, top-ten count — is a proxy for that chain. Even when AI summaries arrived, most practitioners treated them as one more SERP feature layered on top of the same model.

The behaviour of users says otherwise. Pew Research Center tracked the browsing of 900 U.S. adults across 68,879 Google searches in March 2025. When an AI summary appeared — which happened on 18% of searches, and on 53% of queries of ten words or more — users clicked a traditional organic result on **8% of visits, against 15% where no summary appeared**, and ended the session on 26% of them [10]. The summary did not sit above the ranking. It replaced the ranking as the thing the user read.

> "In roughly one in four searches with an AI summary, the user was finished before reaching any result at all."
>
> — Pew Research Center, 68,879 searches, March 2025 [10]

Traffic data confirms the direction without confirming the mechanism. Semrush's clickstream panel — over one billion rows covering seventeen months of U.S. browsing — shows ChatGPT referral

traffic growing **206% during 2025**, while more than 20% of that outbound traffic returned to Google rather than to the cited site, and over 30% of it concentrated in ten domains [8]. The answer layer is now large enough to be a distribution channel, and narrow enough that its rewards are unevenly split.

That makes the unanswered question expensive. If answers, not results, are what users read, then the only thing that matters is whether your page is chosen as a source. And the choice had not been measured at scale before 2025. This paper reports what those measurements found, including where they contradict each other.

---

02 METHOD

Scope, sources and method

This is secondary research. No measurement in this paper was produced by Novaxilabs. Every figure is a published finding from a named study, reported with its sample size and measurement window, and each claim resolves to a numbered reference in Appendix A. The value added here is not new data; it is cross-study comparison, the reconciliation of conflicting results, and the translation of both into decisions a brand team can act on.

Evidence classes

Claims are graded on the five-level scale used across this series: **Official** (vendor documentation), **Peer-reviewed**, **Independent measurement** (large third-party samples, with sample size and window stated), **Vendor-reported**, and **Analysis** (reasoned inference from observable behaviour, explicitly not a claim of access to engine internals). Seven of the eight studies cited here are independent measurement; the mechanism explanations in Section 3 are Analysis and are labelled as such.

Comparability — and why absolute numbers are avoided

The studies do not share a methodology, and the paper does not pretend they do. Ahrefs measured 863,000 keywords against 4 million AI Overview URLs [1]; an earlier Ahrefs sample covered 1.9 million citations [2]; BrightEdge tracked nine verticals across eighteen months [3]; Profound ran 100,000 prompts across two engines [4]; Wix Studio analysed 1,056,727 citations

drawn from 75,000 AI answers [5]. A percentage from one of these cannot be subtracted from a percentage in another. What can be compared is *direction*: whether each study, on its own terms, finds ranking and citation moving together or apart. They are reported individually for that reason.

What this paper does not claim

It does not claim to know how any engine weights its sources internally. It does not claim these findings hold for every language, market or category — where a study reports category splits, those are shown. It does not claim the trend is permanent: every source here is dated, and the paper's own evidence shows how fast the picture moves (Finding 01 reports a metric that roughly halved within eight months).

---

03 FINDINGS

Five findings from the 2026 data

Each finding below states one claim, gives the study that supports it, and names the sample. Where studies disagree, the disagreement is shown in the figure rather than smoothed away — because the disagreement is itself the finding.

FINDING 01

Citation and ranking are decoupling — and the measurement of that decoupling is unstable INDEPENDENT

The clearest result in the 2026 data is that a large share of cited pages sit outside the ranking positions brands optimise for. Ahrefs matched 863,000 keywords against 4 million URLs cited in Google AI Overviews: only **37.1%** of cited URLs also ranked in the top ten, and **36.7% did not appear in the top hundred at all** (March 2026) [1]. A smaller earlier Ahrefs sample of 1.9 million citations had put top-ten overlap at roughly 76% in July 2025 [2] — a figure that looks like a different world eight months earlier.

FIGURE 1 — TOP-TEN OVERLAP OF CITED PAGES, BY STUDY AND DATE

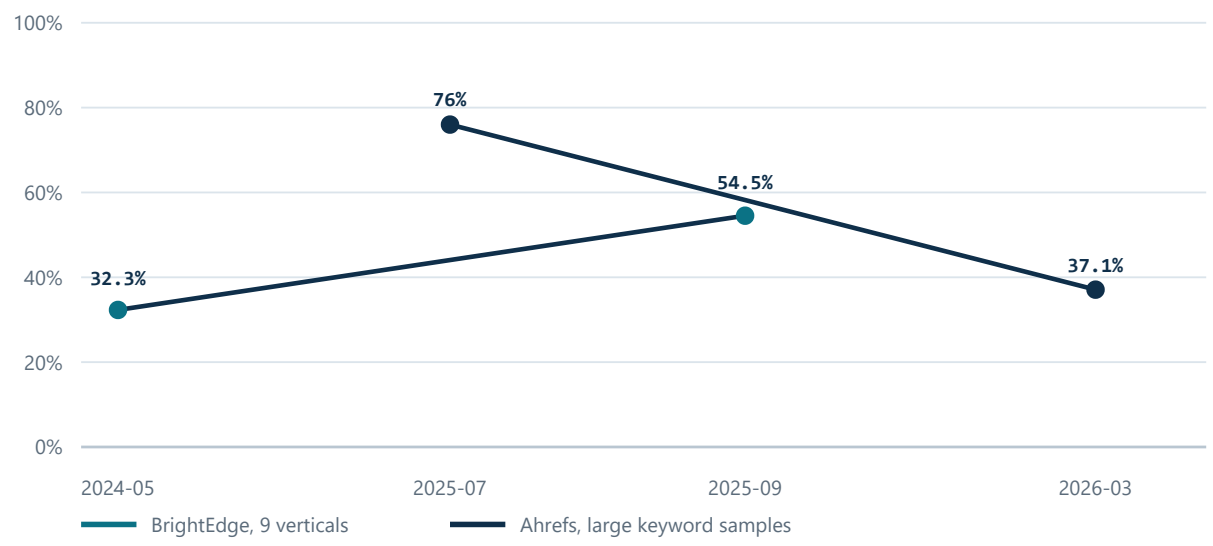


Figure 1 — The two credible studies point in opposite directions, which is the finding: the ranking-citation relationship is real but unstable, and any single measurement of it should be treated as a snapshot, not a law.

Source: BrightEdge, 9 verticals tracked 2024-05 to 2025-09 [3]; Ahrefs, 1.9M citations (2025-07) and 863,000 keywords / 4M AI Overview URLs (2026-03) [1][2]. Samples and methods differ; absolute values are not directly comparable across studies.

The conflict is not a data error; it is a method difference with a lesson attached. BrightEdge tracked a fixed set of nine verticals across eighteen months and saw overlap *rise* from 32.3% to 54.5% [3] — with an extreme spread inside the set, from 75.3% in healthcare to 19.2% in restaurants. Ahrefs used far larger keyword samples and saw overlap collapse in the opposite direction [1][2]. Both are defensible. What survives both is narrower and more useful: **ranking is a relevance signal within a much larger candidate pool, not a precondition for citation** ANALYSIS.

A third study explains why the numbers diverge so violently, and it is the most useful one for planning. Ahrefs tested **15,000 long-tail queries** across ChatGPT, Gemini, Copilot and Perplexity against Google's top ten for the same prompt: only **about 12%** of AI-cited URLs also ranked in the top ten — with Perplexity at roughly 29%, Gemini near 12%, and ChatGPT and Copilot near 8% (August 2025) [18]. The gap between 76%, 12% and 37.1% is largely a *query-type* effect ANALYSIS. Head queries with commercial value still track the ranking closely enough to produce high overlap; long-tail and question-shaped prompts, where the answer is assembled from sub-queries, do not. A brand's exposure therefore depends on which kind of prompt its buyers actually type.

A mechanism explains the gap without claiming knowledge of engine internals. AI Overviews are documented as drawing from Google's core ranking systems, but the pool they draw from is the index, not the rendered top ten — and AI Mode adds query fan-out, which issues derived sub-queries whose result sets are not the original result page [11][12]. A page that ranks twenty-fifth for the head query can be the best answer to a fan-out sub-query. The 2026 numbers are consistent with that: the correlation between ranking and citation remains, but it no longer functions as a gate.

FINDING 02

Engines barely share sources with each other VENDOR-REPORTED

Across 100,000 prompts run through ChatGPT and Perplexity, only **11% of cited domains** appeared in answers from both engines (Profound, July 2025) [4]. The overlap is small enough that a brand can be prominent in one engine's answers and invisible in another's, with no contradiction between the two reports. The source is a citation-tracking vendor publishing its own research, so the direction is credible while the exact figure should be treated as vendor-reported.

FIGURE 2 — DOMAIN OVERLAP BETWEEN TWO ENGINES ON IDENTICAL PROMPTS

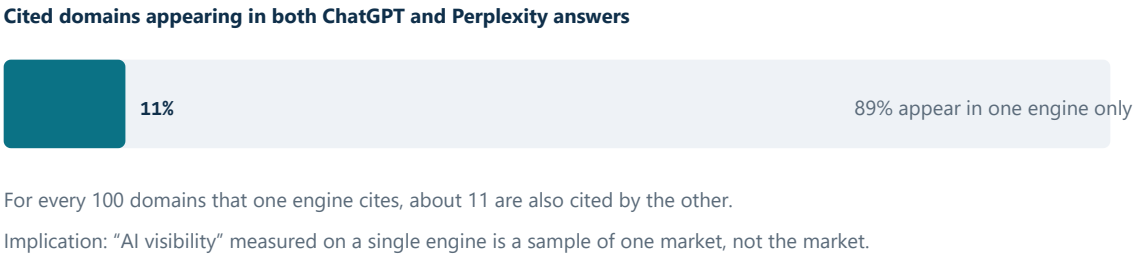


Figure 2 — Cross-engine citation overlap is low enough that a brand's visibility in one engine is largely independent of its visibility in another.

Source: Profound, 100,000 prompts, July 2025 [4]. Single measurement; overlap on other engine pairs is not reported by the study and is not inferred here.

The practical consequence is structural rather than tactical. A single-engine dashboard produces a number that looks like market share but behaves like a sample of one channel. Because the two engines retrieve differently — ChatGPT mixing training memory with live retrieval, Perplexity assembling from a live fetched set [13][14] — the lack of overlap is not random noise. It is the

fingerprint of different retrieval policies, and it means a brand's citation footprint has to be measured per engine and then aggregated.

FINDING 03

What gets cited is extractable structure, not the “best” page

INDEPENDENT

The largest citation dataset available — **1,056,727 citations across 75,000 AI answers** — puts listicles first at 21.9%, long-form articles at 16.7%, and product pages at 13.7% (Wix Studio AI Search Lab, March 2026) [5]. The distribution is not a quality ranking. It is what a summarising model can lift and attribute in a few sentences.

FIGURE 3 — WHAT KIND OF PAGE GETS CITED

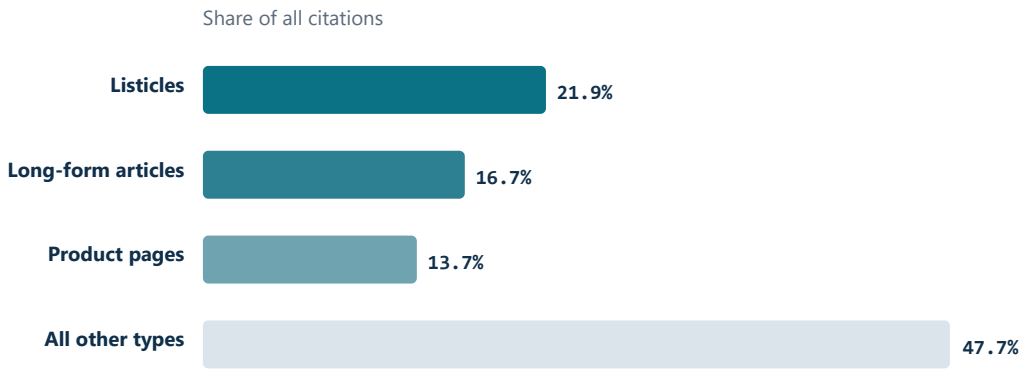


Figure 3 — No single page type dominates; being citable is a property of the page, not of the format.

Source: Wix Studio AI Search Lab, 1,056,727 citations from 75,000 AI answers, March 2026 [5]. Percentages as reported; categories not exhaustive.

Three further measurements sharpen what “extractable” means. Length helps but is not sufficient: pages above 20,000 characters were cited **10.18 times on average** against 2.39 for pages under 500 characters, with the top 4.8% of URLs capturing ten or more citations (Growth Memo, March 2026) [6]. Opening with a direct answer helps more: pages whose titles answered the question directly were cited on **41%** of queries against 29% for loosely related pages (Growth Memo, April 2026) [7]. And a technical shortcut does not work: structured schema markup showed **no measurable lift** in AI Overviews, AI Mode or ChatGPT (Ahrefs, May 2026) [15].

Intent, not industry, predicts the format. Within the same dataset, informational queries drew **45.5% articles** against 21.7% listicles, while commercial queries inverted it — **40.9% listicles** against 6.2% articles [5]. In professional services, 80.9% of cited listicles came from neutral third parties rather than the brands themselves — which means the winning move on buying queries is usually to be *included in* someone else's roundup, not to publish your own.

Two structural facts sit underneath all of this. First, a large share of every answer is spoken for before any brand page is considered: Wikipedia, YouTube and Reddit together supplied **15% of the sources** AI summaries linked to (Pew, 2025) [10], and YouTube alone was the single most-cited domain in AI Overviews — 5.6% of all citations, and **18.2% of citations for pages that ranked outside the top hundred** (Ahrefs, March 2026) [1]. Second, engines differ sharply in how many sources they show: across 126 million U.S. prompts, ChatGPT averaged **15 sources per response** while Gemini averaged **three** (Semrush AI Visibility Index, 2026) [20]. The same page faces very different arithmetic depending on which engine answers.

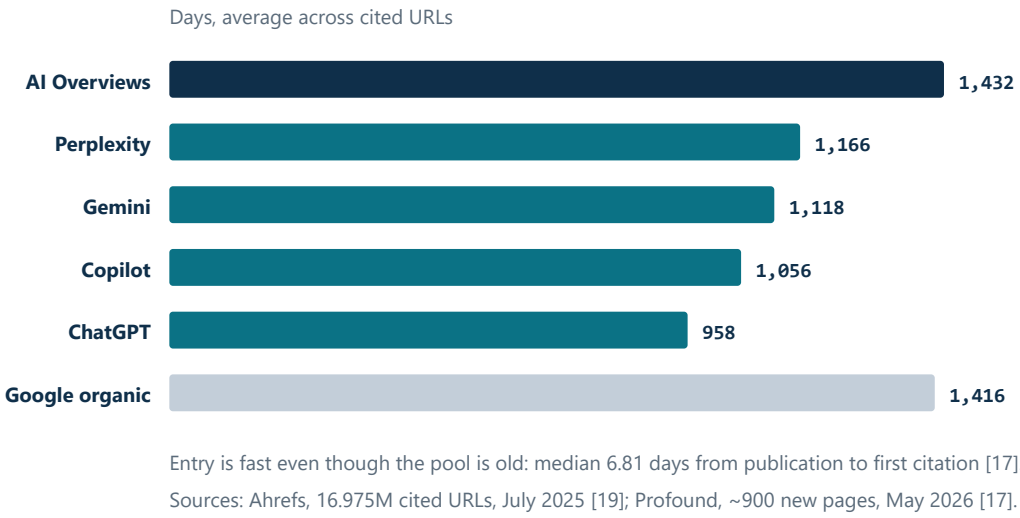
One constraint is infrastructural. In a 14-day controlled test on a single brand's own site, HTML test pages collected **137 AI-bot visits** and appeared as citation sources; the identical content served as Markdown collected **none**, from any crawler, and was cited zero times (OtterlyAI, April 2026) [16]. A page can be excellent and still unavailable if it exists only in a form the retriever does not fetch.

FINDING 04

Freshness is overrated — but the entry window is short INDEPENDENT

The intuition says AI engines prefer fresh pages. The measurement says otherwise: across **16.975 million cited URLs**, the average age of a page cited in Google AI Overviews was **1,432 days** — against 1,416 days for ordinary organic results, a difference of two weeks on a four-year average. ChatGPT's cited pages averaged 958 days, Perplexity's 1,166, Gemini's 1,118, Copilot's 1,056 (Ahrefs, July 2025) [19]. The pool is made of old content. What is fast is *entry*: the median lag from publication to first citation is 6.81 days, and the 90th percentile is 37.10 days — measured by Profound, a citation-tracking vendor, on roughly 900 new pages, so this part of the finding is vendor-reported (May 2026) [17].

FIGURE 4 — AVERAGE AGE OF A CITED PAGE, BY ENGINE



Entry is fast even though the pool is old: median 6.81 days from publication to first citation [17].

Sources: Ahrefs, 16.975M cited URLs, July 2025 [19]; Profound, ~900 new pages, May 2026 [17].

Figure 4 — AI Overviews show almost no freshness preference relative to ordinary search; the fast-moving number is time-to-first-citation, not the age of what gets cited.

Source: Ahrefs, average age of cited pages across 16.975M URLs, July 2025 [19]; Profound, publication-to-citation tracking on ~900 new pages, May 2026 [17]. The two measures are different: page age is not time-to-citation, and the time-to-citation figure is vendor-reported.

The honest reading is narrower than either study alone suggests. “Publish fresh content to win citations” is not what the data shows: relative to ordinary search, AI Overviews exhibit almost no age preference, and the pages being cited are typically years old. Freshness affects *eligibility*, not preference — a new page can be retrieved within days, which matters only when a page is new. The operational conclusion: refreshing an existing page is not a growth lever, but a page that is never published inside the retrieval window has never entered the pool at all. What compounds is continuous existence and repeated retrieval, and a page rewritten every quarter does not reset that clock.

FINDING 05

Being cited does not mean being visited INDEPENDENT

In Google AI Mode, **93% of sessions produce no outbound click at all**, with only 6–8% of sessions reaching any external page (GEO Hub, 2026) [9]. AI Overviews show a comparable pattern at **43% zero-click** on covered queries. Chat-based engines are not a traffic rescue either: only **34.5% of ChatGPT queries** triggered a web search in February 2026, down from 46% at the end of 2024 (Semrush) [8].

FIGURE 5 — SESSIONS THAT END WITHOUT AN OUTBOUND CLICK

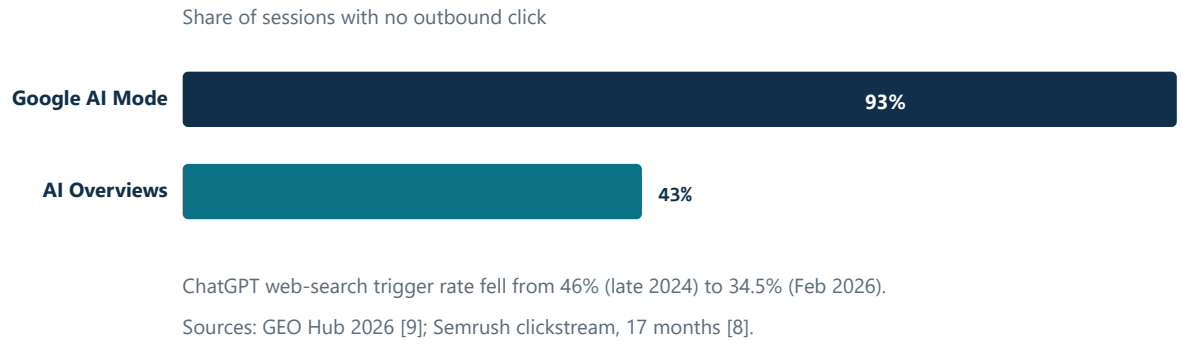


Figure 5 — The answer layer is a place where a brand is described, not a channel through which a visitor arrives. Source: GEO Hub on AI Mode sessions (2026) [9]; Semrush clickstream panel, 17 months to 2026-02, 200M users [8]. Zero-click definitions differ between sources and are reported separately.

The implication is a change of objective function, not a change of channel. If the citation is the endpoint for most sessions, then the outcome worth measuring is whether the brand is named, how early, how accurately, and against whom — not how many visitors arrived. Traffic that does occur is concentrated: over 30% of ChatGPT's outbound traffic lands on ten domains, and over 20% of it returns to Google rather than to the cited publisher [8]. For most brands the realistic prize is being the correct answer inside the answer, with the click as an occasional bonus.

---

04 IMPLICATIONS

What this changes for brands

The five findings converge on one operational shift: the unit of competition has moved from position to presence, and the metrics have to follow.

1. Rank stops being the reporting headline

Average position remains a useful diagnostic — it predicts which pages are plausible candidates — but it is no longer a proxy for visibility. A brand reporting “three keywords in the top ten” has said nothing about whether it appears in answers. The replacement metric is *cited-source share*: of

the answers your category's prompts produce, in what share is your domain one of the sources, and at what position in the source list.

2. Engine selection is a market-structure decision

At 11% cross-engine overlap, “AI visibility” measured on one engine describes one channel [4]. A brand should decide which engines correspond to its actual buying population, then report each separately and only aggregate with the engine mix stated. A blended single number across engines is not a summary; it is a weighted average of incomparable populations.

3. Extractability is the only lever fully inside a brand's control

Ranking depends on competitors and on an algorithm nobody can inspect. Extractability does not: a page that opens with a direct answer, states its claim in the first two sentences, and covers the obvious sub-questions can be improved this week. The measured effect is significant — 41% citation rate for direct-answer titles against 29% for loosely related ones [7], and a 4.3× gap between long and very short pages [6]. The one lever that looked like a shortcut — schema markup — measured no lift at all [15].

4. Freshness is an eligibility gate, not a growth lever

The average cited page is over three years old [19], so “update our content” is not the lever it is sold as. What freshness governs is whether a page can be picked up at all — and that window is short, with a median 6.81 days to first citation [17]. The implication is to publish continuously rather than refresh periodically: a steady cadence of new, answerable pages keeps putting entries into the pool, while a quarterly rewrite of old pages changes almost nothing about their citation behaviour.

5. The outcome to optimise is accuracy, not arrival

At 93% zero-click in AI Mode [9], a strategy whose success condition is a visit will report failure on most of its own wins. The success condition has to be restated: is the brand named, is it described correctly, is it recommended where it should be, and which competitors appear beside it. That is measurable — and it is the only thing the citation layer actually pays out.

05 LIMITATIONS

Limitations

This is a synthesis, not a measurement. Every figure comes from a study with its own sample, window and method, and the paper reports them individually precisely because they cannot be pooled. Three specific weaknesses should travel with the findings:

- **Instability.** Figure 1 shows two credible studies reaching opposite conclusions about the same relationship within eight months. Any single number in this paper is a snapshot, and the paper's own thesis implies it will age.
- **Category spread.** BrightEdge's verticals ranged from 75.3% overlap in healthcare to 19.2% in restaurants [3]. An average across verticals can describe no vertical at all. Readers should assume their own category sits somewhere inside that range, not at its midpoint.
- **Geographic and linguistic bias.** The studies cited are predominantly U.S. and English-language samples. Engine behaviour differs by market, language and login state, and none of the figures here should be transferred to another market without re-measurement.

The paper also does not claim to describe how engines weight sources internally. Mechanism explanations — core-ranking integration, query fan-out, wider candidate pools — are drawn from vendor documentation and from observable inputs and outputs, and are labelled Analysis wherever they go beyond what a vendor has stated.

06 RECOMMENDATIONS

What to do next

Ordered by effort-to-impact, each with an owner and a way to check it worked.

FIGURE 6 — EFFORT VERSUS IMPACT ACROSS THE SEVEN ACTIONS

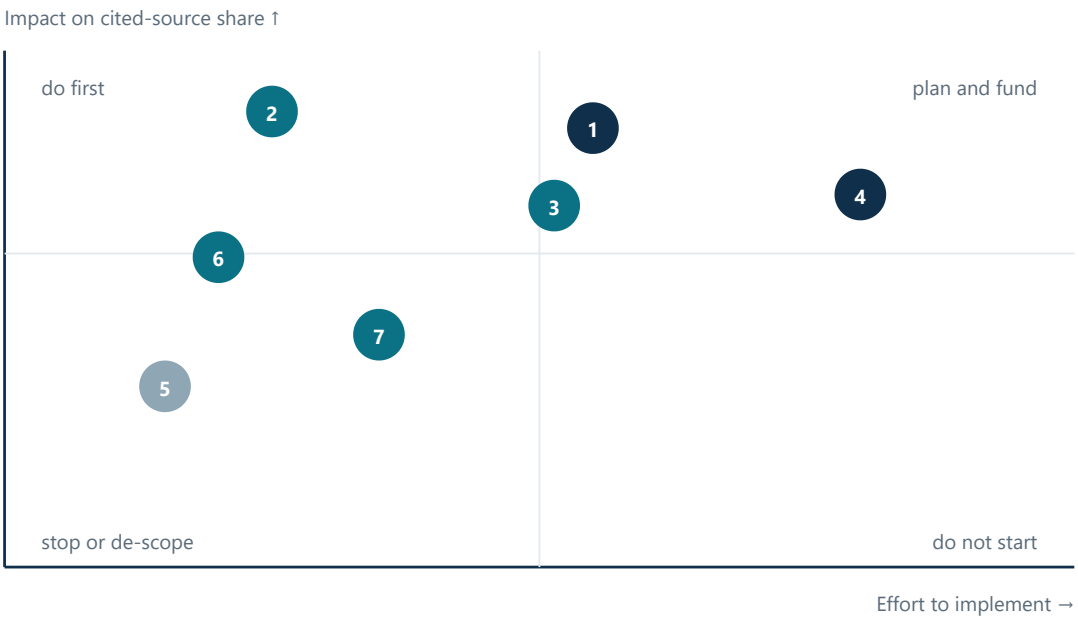


Figure 6 — The same seven actions placed by implementation effort and expected effect on cited-source share, so a team can sequence work rather than argue about order.

Source: the author's classification of the actions described in this section. Positions are an editorial judgement, not a measurement — no per-action effect size has been published, and none is claimed here. Numbers refer to the table below.

#	Action	Owner	How you know it worked
1	Build a per-engine cited-source baseline for 20–30 category prompts. Record domains cited, not your rank.	SEO / content lead	You can name the five domains that dominate your category's answers, per engine.
2	Rewrite the opening of every priority page as a direct answer to the question it serves.	Content editors	Citation rate on those prompts rises toward the 41% direct-answer benchmark.
3	Target 30–50% coverage of fan-out sub-questions per page, not 100%.	Content strategy	Pages stop trying to answer everything; cited-source share holds or improves.
4	Publish new answerable pages every two weeks instead of refreshing old ones.	Editorial calendar owner	New URLs accrue their first citation inside the 6.8-day median window.
5	Stop counting schema markup as a GEO lever; keep only baseline structured data.	Technical SEO	Budget reallocated to extractability work with a measured effect.
6	Verify every priority page is served as crawlable HTML, not only as Markdown or a JS-rendered shell.	Web / dev	Bot logs show retrieval hits on the page itself.
7	Re-measure the baseline quarterly and keep the old numbers visible.	Research	You can show the metric moving — as Figure 1 shows it can.

The sequence matters more than any single row. Actions 1 and 2 are cheap and produce a signal within one publishing cycle; action 4 is the one that converts a one-off improvement into a standing capability. Action 5 is included because removing a plausible-looking but ineffective lever is itself a gain.

APPENDIX A

Sources & References

Source policy. Evidence is graded on a five-level scale: **Official** (vendor documentation or announcement), **Peer-reviewed**, **Independent measurement** (large third-party sample with published method), **Vendor-reported** (a vendor publishing data that also serves its own product claim), **Analysis** (reasoned inference from observable behaviour, not a claim of access to engine internals). Tool-vendor studies (Ahrefs, Semrush, Wix Studio) are graded Independent measurement because their samples and methods are published; where a vendor measures its own headline selling point — Profound's cross-engine overlap — the entry is downgraded to Vendor-reported. Some studies are not stably addressable on their own platform; where that is the case, the entry cites the secondary source that names them, so every figure in this paper can still be checked against a live page.

1. Ahrefs, "AI Overview citations versus organic ranking," 863,000 keyword SERPs and 4 million AI Overview URLs, March 2026: 37.1% of cited URLs ranked top 10, 26.2% ranked 11–100, 36.7% outside the top 100. Reported in SEOValley, "What Gets Cited in AI Overviews," September 2026. seovalley.com (https://seovalley.com/what-gets-cited-in-ai-overviews-and-how-it-changes-by-query/) INDEPENDENT

2. Ahrefs, earlier citation sample, 1.9 million citations, July 2025: top-ten overlap approximately 76%. Same reporting source as [1]. INDEPENDENT

3. BrightEdge, citation-to-organic overlap across nine industries, May 2024 – September 2025: 32.3% → 54.5%; healthcare 75.3%, restaurants 19.2%. Reported in SEOValley [1]. INDEPENDENT

4. Profound, "Cross-engine citation overlap," 100,000 prompts run through ChatGPT and Perplexity, July 2025: 11% of cited domains appear in both. Reported in AI Visibility Software, "Myth: AI Visibility Is Just SEO Rebranded," July 2026. ai-visibility-software.com (https://www.ai-visibility-software.com/blog/myth-ai-visibility-just-seo-rebranded-data-says/) VENDOR-REPORTED

5. Wix Studio AI Search Lab, citation composition study, 1,056,727 citations across 75,000 AI answers from ChatGPT, Google AI Mode and Perplexity, March 2026: listicles 21.9%, articles 16.7%, product pages 13.7%; informational queries 45.5% articles; commercial queries 40.9% listicles. Reported in SEOValley [1]. INDEPENDENT

6. Growth Memo, page-length and citation analysis, March 2026: pages above 20,000 characters cited 10.18 times on average versus 2.39 for pages under 500 characters; top 4.8% of URLs cited ten or more times. As reported, with attribution, by Position Digital. position.digital (https://www.position.digital/blog/ai-seo-statistics/) INDEPENDENT

7. Growth Memo, headline-form analysis, April 2026: pages whose headlines directly answer the question are cited on 41% of queries, against 29% for loosely related headlines. Same reporting source as [6]. position.digital (https://www.position.digital/blog/ai-seo-statistics/) INDEPENDENT

8. Semrush, "ChatGPT search insights," clickstream panel of 200 million users, 17 months / over one billion rows to February 2026: ChatGPT referral traffic +206% during 2025; over 30% of outbound traffic to ten domains; over 20% returning to Google; web search triggered on 34.5% of queries (46% at end of 2024). semrush.com (https://www.semrush.com/blog/chatgpt-search-insights/) INDEPENDENT

9. GEO Hub, "The click that never came: AI Mode's 93% zero-click rate," 2026: 93% of AI Mode sessions produce no outbound click; 6–8% reach an external page; AI Overviews 43% zero-click. geo.alibaba.com (https://geo.alibaba.com/geo-hub/the-click-that-never-came-ai-mode-93-percent-zero-click-rate-open-web) INDEPENDENT

10. Pew Research Center, search behaviour study, 900 U.S. adults / 68,879 Google searches, March 2025: 18% of searches produced an AI summary (53% for queries of ten words or more); 8% organic click rate where a summary appeared versus 15% without; 26% of summary sessions ended with no result visited; Wikipedia, YouTube and Reddit supplied 15% of linked sources. pewresearch.org (https://www.pewresearch.org/) INDEPENDENT

11. Google, "What happened with AI Overviews and next steps," The Keyword, May 2024: AI Overviews draw on Google's core ranking systems and identify relevant results from the search index. blog.google (https://blog.google/products-and-platforms/products/search/ai-overviews-update-may-2024/) OFFICIAL

12. Google, "Expanding AI Overviews and introducing AI Mode," The Keyword, March 2025: AI Mode uses query fan-out, issuing multiple related searches across subtopics and data sources. blog.google (https://blog.google/products/search/ai-mode-search/) OFFICIAL

13. OpenAI, "Introducing ChatGPT search," October 2024: ChatGPT search pulls current information from the web with links to sources. openai.com (https://openai.com/index/introducing-chatgpt-search/) OFFICIAL

14. Kompozy, "How Perplexity selects sources: retrieval, ranking and citation mechanics," 2026 — third-party technical analysis; no official Perplexity documentation describes the selection pipeline. kompozy.io (https://kompozy.io/guides/perplexity-source-selection-mechanics) ANALYSIS

15. Ahrefs, schema-markup test on 1,885 pages that added JSON-LD, May 2026: no statistically meaningful uplift in AI citations across AI Overviews, AI Mode or ChatGPT. Reported in AI Visibility Software [4]. INDEPENDENT

16. OtterlyAI, "GEO Experiment: Markdown vs HTML," controlled test on the vendor's own site, 14 days, two matched page pairs: the HTML pages received 137 AI-bot visits and were cited; the identical .md versions received zero visits and zero citations across ChatGPT, Google AI Overviews and Perplexity. Single-brand experiment; does not generalise. otterly.ai (https://otterly.ai/blog/geo-experiment-html-vs-markdown/) VENDOR-REPORTED

17. Josh Blyskal (Profound), time-to-first-citation study, May 2026: roughly 900 newly published pages tracked across LLM response logs; median 6.81 days, 90th percentile 37.10 days. Published by a citation-tracking vendor. tryprofound.com (https://www.tryprofound.com/blog/research) VENDOR-REPORTED

18. Ahrefs, “Only 12% of AI-cited URLs rank in Google’s top 10 for the original prompt,” 15,000 long-tail queries across ChatGPT, Gemini, Copilot and Perplexity, August 2025: ~12% overlap overall; Perplexity ~29%, Gemini ~12%, ChatGPT ~8%, Copilot ~8%. Reported in AI Visibility Software [4]. INDEPENDENT

19. Ahrefs, average age of cited pages across 16.975 million cited URLs, July 2025: ChatGPT 958 days, Copilot 1,056, Gemini 1,118, Perplexity 1,166, Google AI Overviews 1,432, ordinary Google organic 1,416. Reported in SEOValley [1]. INDEPENDENT

20. Semrush AI Visibility Index, 126 million U.S. prompts, January–April 2026: ChatGPT averaged 15 sources per response, Gemini averaged three. Reported in SEOValley [1]. INDEPENDENT

B

AUTHOR

About the author

Dr. Adrian Whitmore is an independent GEO research consultant and a PhD in Information Retrieval from the University of Cambridge. He previously led SEO research at a UK digital marketing group, where he built the Citation Share framework for measuring how generative engines distribute attribution across competing pages. He is the author of four arXiv papers on generative search visibility and an external contributor to the Novaxilabs White Paper Series. His work focuses on measurement design — what can be counted, what cannot, and what a number changes when it is reported to a board.

This paper was written independently. Its conclusions were not reviewed or directed by Novaxilabs, and no claim in it depends on Novaxilabs' own monitoring data.

C

COMPANY

About Novaxilabs

Novaxilabs measures how a brand appears in AI answers — not how a page ranks. The platform tracks mention rate, recommendation position, sentiment and the sources engines actually cite, across seven engines, on one dashboard.

7 engines · ChatGPT, Perplexity, AI Overviews, AI Mode, Gemini, Claude, Copilot — scanned on one dashboard.

Daily · repeatable, neutral scans with history you can compare week over week.

24/7 · change alerts when rank or sentiment moves.

See your baseline AI visibility.

A free scan of how your brand currently appears in AI answers — mention rate, rank, sentiment, and the gap your competitors are filling.

Start your free scan **Talk to us**

Prefer email? info@novaxilabs.com

© 2026 Novaxilabs. All rights reserved.

This white paper is provided for informational purposes. Figures cited are third-party, dated and directional; engine behaviour changes over time. Studies are graded by evidence class in Appendix A.