

AI OVERVIEWS

The GEO Playbook: Earning Citations in AI Overviews, ChatGPT Search, and Perplexity

Google says there is no special markup for AI Overviews. The "40% GEO lift" study measured something narrower than most pitches claim. Here is what the research actually supports.

Contents

01 Is there special markup that gets a page into AI Overviews?

02 Does the research really show a 40% visibility boost from GEO?

03 Which on-page techniques actually have evidence behind them?

04 What does that rewrite look like in practice?

05 Does FAQ schema still help a page get cited?

06 How do you check whether any of this is working?

07 What are the limits of this, honestly?

Key Takeaways & Sources

01

Is there special markup that gets a page into AI Overviews?

No. Google's Search Central documentation is direct about this: "There are no additional requirements to appear in AI Overviews or AI Mode, nor other special optimizations necessary." No new file format, no AI-specific tag, no schema.org type built for generative surfaces. AI Overviews and AI Mode draw on the same crawling, indexing, and ranking systems as classic Search results.

That has a practical consequence most vendor pitches skip past: a page that isn't already indexable and eligible to appear as a normal snippet in regular Search results cannot appear in an AI Overview either. There is no side door. If a page is blocked by robots directives, thin, duplicate, or simply not ranking for the query in question, no amount of "GEO formatting" changes that. Everything in this guide is refinement layered on top of that floor, not a replacement for it — and it's worth checking the floor first, because a surprising share of "why aren't we getting cited" problems trace back to a page that was never eligible to begin with.

If a page is blocked by robots directives, thin, duplicate, or simply not ranking for the query in question, no amount of "GEO formatting" changes that.

02

Does the research really show a 40% visibility boost from GEO?

The number comes from one paper: "GEO: Generative Engine Optimization" (Aggarwal et al., arXiv:2311.09735, accepted at KDD 2024). It tested interventions like adding statistics, quotations, and citations to source documents, then measured a position-adjusted word-count metric the authors call visibility. The best-performing intervention, adding quotations, raised that metric by roughly 41% in their benchmark — which is where the "up to 40%" line comes from.

The detail that gets dropped in the retelling: this was measured among five sources that were already placed in a language model's retrieved context for a given query. The metric asks "of the words in the model's answer, how many trace back to this source, adjusted for where it appears," not "did this source get retrieved at all." A 2026 critical survey of the field (arXiv:2607.14035) makes the distinction explicit, writing that the original paper's "widely cited gains are valid within its experimental setting but conditional on a source already being present in a fixed context; they establish neither organic discoverability nor durable traffic effects." Across the 45 studies that survey reviewed, none showed a stable, cross-platform causal effect on whether a page gets discovered in the first place.

So the honest reading is narrower than the headline: once a document is already one of the sources an LLM is drawing from, certain edits change how much of the final answer gets attributed to it. Whether those edits helped the document get selected as a source, and whether that selection produces a click or a lead, is a separate question the paper was not designed to answer.

Claim commonly repeated	What the study actually measured
"GEO increases visibility by 40%"	A ~41% relative gain in a position-adjusted word-count metric, from adding quotations, among 5 sources already retrieved into an LLM's context for a benchmark query set
"This proves GEO gets you cited more often"	The paper did not test retrieval odds — whether a source makes it into the model's context in the first place — only how sources already there get weighted
"These techniques drive traffic and conversions"	Not measured in the original paper. A separate 2026 survey found no reviewed GEO technique with a demonstrated, durable effect on organic discoverability or downstream behavior

03

Which on-page techniques actually have evidence behind them?

A narrower set of edits does have real backing, and the mechanism explains why: generative engines built on retrieval-augmented generation don't retrieve whole pages, they retrieve chunks — a paragraph, a list item, a table row — that get embedded and ranked independently of the page around them. A page can rank well overall and still lose every retrieval contest if none of its individual passages can stand alone. Writing for the chunk, not just the page, is the practical shift this implies.

Four techniques follow directly from that mechanism and show up consistently across independent analyses of what gets quoted in AI-generated answers, even where a controlled study of the exact magnitude doesn't exist:

- Answer-first paragraphs: state the direct answer in 40-60 words near the top of a section, before elaboration, so a retrieval system can lift a complete answer without needing the surrounding page.
- Self-contained paragraphs: write each passage so it makes sense with no context from the sentence before or after it — no "as mentioned above," no pronouns referring outside the paragraph.
- Descriptive, question-shaped headings: a heading a reader (or a model parsing the page) could paste back as the query it answers, not a generic label like "Overview" or "Learn More."
- Real specificity: named entities, dated events, and statistics attributed to a checkable source, rather than vague categories like "many experts" or "recent studies."

04

What does that rewrite look like in practice?

Generic version: "Businesses today need to think carefully about their content strategy. There are many factors that go into creating content that performs well, and it's important to consider your audience, your goals, and the competitive landscape. Getting this right can lead to better results over time."

That paragraph fails on every count above. It has no direct answer, no named entity, no date, and no claim specific enough to be checked or lifted out as a standalone fact. A retrieval system scanning it would find nothing to extract, because there is nothing concrete in it.

Rewritten: "A content strategy succeeds or fails on one early decision: whether each page targets a single, well-defined search intent or tries to cover several at once. Google's own Search Central guidance on AI features, updated in 2026, states plainly that ranking in AI Overviews requires no new markup or file format — only the same indexing and relevance signals classic Search already uses. A page trying to serve a comparison-shopper and a how-to reader at once usually satisfies neither well enough to earn a place in either result type."

The second version opens with a direct claim, names a specific source and a specific number, and could be lifted into an answer on its own without the reader needing anything from the paragraphs around it. Notice what it does not do: it does not invent a statistic to sound more specific. If a comparable named study for a given topic doesn't exist, the honest move is to describe the mechanism plainly rather than manufacture a citation.

It has no direct answer, no named entity, no date, and no claim specific enough to be checked or lifted out as a standalone fact.

05

Does FAQ schema still help a page get cited?

FAQPage structured data is still valid, parseable schema.org markup, and Google has said structured data in general helps its systems understand entities and page content even when it produces no visible rich result. Adding it to a page with genuine, visible Q&A; content is reasonable on that basis alone.

What it will not do anymore is win a rich result in Google Search. Google confirmed the FAQ rich-result search feature stopped appearing on May 7, 2026, with the rich-result report and testing-tool support following in June 2026 and API support ending in August 2026. The removal applies broadly, including the narrow set of government and health sites that had kept eligibility after an earlier 2023 restriction. Keep FAQ schema where the content is genuinely Q&A-shaped; just don't tell a client, or write in a strategy doc, that it buys a visible snippet in Google Search. That channel is gone.

The removal applies broadly, including the narrow set of government and health sites that had kept eligibility after an earlier 2023 restriction.

06

How do you check whether any of this is working?

Because AI Overviews and chat-assistant citations don't come with the referral clarity of a normal search click, monitoring has to combine a few imperfect signals rather than lean on one clean number. None of these establishes a causal line to revenue on its own, and claiming otherwise is exactly the overreach this guide is arguing against.

- Run branded and topic queries directly in Google (watching for AI Overview panels), ChatGPT search, Perplexity, and Copilot on a recurring schedule, and log which pages get cited and how the excerpt is worded.
- Watch referral traffic from chat-assistant domains in analytics (chatgpt.com, perplexity.ai, and similar) as a directional signal of downstream visits, understanding that most AI-surface exposure never produces a trackable click at all.
- Track whether pages already ranking on page one see their AI Overview appearance rate change after a rewrite — this isolates the AI-surface question from the separate problem of whether the page ranks in classic Search at all.
- Treat any citation-rate tool built on a small, licensed sample of prompts as directional, not exhaustive; it reflects that tool's prompt set, not the space of queries real users actually type.

07

What are the limits of this, honestly?

The 2026 survey behind much of this guide proposes a workable ethical line for this kind of optimization: keep facts true, keep statistics and citations real and verifiable, keep the page addressed to human readers rather than hiding instructions aimed at a model, and disclose rather than manipulate. Hidden or invisible text intended to steer a language model's output, and fabricated statistics dressed up to look "specific" for a checklist, both fail that test regardless of how well they might work in the short term — and they undermine the exact trust signals that make a page worth citing in the first place.

The larger limitation is structural, not ethical. The C-SEO Bench study (arXiv:2506.11097, NeurIPS 2025) tested 54 conversational-SEO methods and found only 3 produced a statistically significant positive effect; most were neutral, and several actively hurt. The same study documented a "congested and zero-sum" dynamic: as more sites adopt a given tactic, its benefit for any one of them shrinks toward zero, because the model is still choosing among a fixed number of sources to cite regardless of how many pages now use the same trick. No single tactic in this guide, including the ones with genuine mechanical backing, guarantees a citation. The realistic goal is removing the reasons a page gets

skipped — not finding a hack that beats a system built to resist exactly that.

Key Takeaways

- Google's own documentation states there is no special markup, schema, or file format for AI Overviews or AI Mode — a page must already be indexable and snippet-eligible in classic Search before it can appear in either.
- The widely repeated "GEO boosts visibility by up to 40%" figure measured a relative gain among five sources already retrieved into a model's context window — not whether a page gets retrieved, clicked, or converts.
- A 2026 benchmark of 54 "conversational SEO" tactics found only 3 produced a statistically significant improvement, and gains shrink toward zero as more competitors copy the same tactic.
- FAQPage schema is still valid markup, but Google discontinued the FAQ rich-result search feature for nearly all sites starting May 7, 2026 — it will not produce a visible rich result anymore.
- llms.txt shows no measured correlation with AI citation frequency; major AI crawlers largely ignore it and read HTML directly, so treat it as developer tooling, not a citation lever.

Sources

1. AI features and your website — Google Search Central —
<https://developers.google.com/search/docs/appearance/ai-features>
2. GEO: Generative Engine Optimization (Aggarwal et al., arXiv:2311.09735, KDD 2024) —
<https://arxiv.org/abs/2311.09735>
3. Optimizing Visibility in Generative Engines: A Critical Survey of Generative Engine Optimization (arXiv:2607.14035) — <https://arxiv.org/abs/2607.14035>
4. C-SEO Bench: Does Conversational SEO Work? (arXiv:2506.11097, NeurIPS 2025) —
<https://arxiv.org/abs/2506.11097>
5. Intro to structured data — Google Search Central —
<https://developers.google.com/search/docs/appearance/structured-data/intro-structured-data>
6. Google Drops FAQ Rich Results From Search — Search Engine Journal —
<https://www.searchenginejournal.com/google-drops-faq-rich-results/574429/>
7. Does llms.txt Do Anything Yet? The 2026 Evidence — Mekanik —
<https://mekanik.dev/en/posts/does-llms-txt-do-anything-yet/>
8. Generative engine optimization — Wikipedia — https://en.wikipedia.org/wiki/Generative_engine_optimization