

Yes. Once the claims are separated by *what was actually measured*, the evidence base is considerably narrower than much GEO marketing suggests.

One useful anchor is a July 2026 critical survey of 45 GEO studies. Its conclusion is unusually relevant here: the literature shows that already-retrieved content can sometimes be causally manipulated to change citation/use, but it found **no technique with a stable, longitudinal, cross-platform causal effect on organic discoverability or downstream behavior**. The survey is itself a single-author preprint, so I would treat it as a useful synthesis rather than final authority, but its methodological criticism is verifiable against the underlying studies.

I would classify the important claims as follows.

1. Verified by platform documentation or reasonably strong experimental evidence

Claim	Verdict	What the evidence actually establishes
Google AI Overviews/AI Mode still depend on foundational SEO.	Verified for Google	Google says supporting pages must be indexed and eligible to appear in Search with a snippet, and that there are no additional technical requirements for AI Overviews/AI Mode. Crawlability, internal links, textual content, normal technical SEO and people-first content remain applicable. This is authoritative about Google's stated system behavior, although it does not disclose
Google AI Search can fan a user query out into multiple searches.	Verified	weighting or guarantee inclusion. Google explicitly documents query fan-out: the model can issue concurrent related queries across subtopics /data sources and retrieve additional results while constructing an answer.
ChatGPT Search can rewrite a prompt into multiple targeted searches and search again after seeing results.	Verified	OpenAI documents exactly this behavior and gives an example where an initial query is followed by a more specific search after reviewing the first results.
AI-search crawler access is a genuinely new technical consideration outside ordinary Google SEO.	Verified	OpenAI says allowing OAI-SearchBot is important for inclusion in ChatGPT summaries/snippets. Perplexity says PerplexityBot builds its search index and can be controlled via <code>robots.txt</code> . These are real additional
Google Search and Gemini have different publisher controls.	Verified	technical surfaces. Google says Google-Extended controls use of crawled material for future Gemini models and certain Gemini grounding uses, but does not affect inclusion or ranking in Google Search. Google AI Overviews/AI Mode are instead governed through ordinary Google
There is no special Google AI schema, <code>llms.txt</code> , Markdown version or mandatory "chunking" requirement.	Verified—and some common claims are explicitly contradicted	Search/Googlebot controls. Google's July 2026 guidance says Google Search ignores <code>llms.txt</code> , does not require special AI files/markup, does not require tiny content chunks, and does not require a special writing style for generative
Google AI Overview and AI Mode are not simply the same ranking output rendered differently.	Verified at the architectural level	Search. Google says AI Mode and AI Overviews may use different models and techniques and therefore may show different responses and links. That establishes heterogeneity; it does <i>not</i> tell us what specific optimization factor makes one URL win in either

Changing a document can causally change how much a generative model uses it—once that document is already supplied to the model.	Experimentally supported, but conditional	system. The peer-reviewed KDD 2024 GEO paper manipulated source documents and found some treatments could increase its defined visibility measures by as much as ~40%. That is real causal evidence inside the experimental retrieval/context setup . It is not evidence that those edits make the document more likely to be found organically by ChatGPT, Google or Perplexity.
--	---	---

That last distinction is probably the most important one in the entire GEO debate.

The industry often converts:

“If the model already has your document, editing it can affect how it uses it”

into:

“Make these edits and ChatGPT/Google will retrieve and cite your page more often.”

The experiment establishes the first proposition, not the second. The 2026 literature survey specifically flags this distinction.

2. Supported mainly by observational correlation or reasonable inference

These are useful working hypotheses, but I would not present them as established ranking factors.

Claim	Evidence	What we can and cannot conclude
Ranking for the user's exact query is less determinative of AI citation than in ordinary search.	Ahrefs examined 863,000 SERPs and ~4 million AI Overview URLs in March 2026. Only 37.1% of cited URLs appeared in the organic top 10 for the same query; 36.7% were outside the top 100.	Strong descriptive evidence. It shows same-query rankings and citations often diverge. Google's documented fan-out provides a plausible explanation. But the study did not observe Google's internal fan-out queries, so it cannot prove that a particular citation entered through a particular subquery.
Therefore topical/question-space coverage probably matters more than tracking one literal keyword.	Follows reasonably from documented fan-out plus the weak same-query citation overlap above.	Reasonable inference, not a proven optimization treatment. We know multiple queries occur. We do not know that adding N related sections increases citation probability by X%. Google's own advice warns against manufacturing pages for every fan-out/query variant.
Citation and brand mention are different outcomes.	Semrush logged 3,981 domain appearances across 115 prompts, 14 countries and four AI systems. 61.7% of citations in that dataset did not result in the brand being named in the answer; mentions without citations also occurred.	Well-supported descriptively. You should measure them separately. But the study does not establish that any particular “brand optimization” tactic causes mentions.
Citation and actual use/absorption of a source are also different outcomes.	A 2026 study analyzed 602 prompts and tens of thousands of citation records and found citation breadth and estimated influence diverged	Useful measurement evidence, but much of the feature analysis is observational. It supports separating citation from answer influence; it doesn't establish a universal ranking formula.
Pages containing clear definitions, numbers, comparisons and procedures may be easier for AI systems to use as evidence.	The same study found high-influence cited pages were associated with greater semantic alignment, structure and extractable evidence.	Correlation. These characteristics may cause greater use, may be proxies for page quality/relevance, or both. The original GEO experiments offer conditional causal evidence for some edits <i>after context inclusion</i> , but not for organic retrieval.

Third-party/earned media appears disproportionately important in some AI systems.	A 2025 research paper reported a strong preference for third-party earned media versus brand-owned/social sources across its tested AI search systems and query sets.	Good evidence of a source-selection pattern . It does not prove that securing one more PR mention will causally raise your brand's ChatGPT/Gemini recommendation rate. "Do digital PR" is an inference from the sourcing pattern.
Different AI engines/surfaces select substantially different sources.	Ahrefs compared hundreds of thousands of paired Google AI Mode/AIO responses and observed only 13.7% URL citation overlap. Google independently confirms that the two systems can use different models and techniques.	The difference is real . The stronger claim that each therefore requires a known, separate set of optimization tactics is not yet established.
Traditional rankings and AI citations are related, but the relationship is not deterministic.	Ahrefs' 2026 study found substantial but minority top-10 overlap; its 2025 study had found much higher overlap, illustrating how rapidly this relationship can shift.	Correlation supports continuing SEO. It does not show that increasing rank from position 8 to 3 <i>causes</i> an AI citation. Both could be consequences of relevance/authority or other hidden variables.
AI Overviews are associated with lower outbound CTR.	Pew tracked browsing activity from 900 U.S. adults: traditional-result clicks occurred on 8% of visits with an AI summary versus 15% without; AI-summary-source clicks occurred on ~1%. Ahrefs separately found a 58% lower position-one CTR associated with AIO queries in its December 2025 dataset.	Strong observational evidence of association , particularly Pew's behavioral data. But neither is a randomized experiment. Queries that trigger AIOs differ systematically from queries that do not, so "AI Overviews caused exactly X% click loss" is stronger than the designs support. Ahrefs itself initially describes its finding as a correlation.
Adding ordinary schema markup does not appear to be an AI-citation growth lever.	Ahrefs followed 1,885 pages that added JSON-LD, matched them to ~4,000 controls and used difference-in-differences. AI Mode (+2.4%) and ChatGPT (+2.2%) changes were statistically indistinguishable from zero; AIO showed a small negative difference that the authors explicitly declined to interpret causally.	This is better than a simple correlation and provides reasonably strong evidence against a large generic schema!' citation effect . It cannot prove that every schema type is useless in every use case. Google's own position is that schema remains normal SEO infrastructure but there is no special AI schema.

One implication is that my previous statement that content should function well as “evidence” is reasonable, but needs a qualifier:

There is evidence that evidence-rich content is more likely to be used once encountered; evidence that mechanically adding “extractable facts” makes a page organically retrieved more often is much weaker.

That is the scientifically safer formulation .

3. Common GEO/SEO claims that currently outrun the evidence

This is where I would be particularly skeptical of vendor checklists .

Common claim	Evidence assessment
"Add an <code>llms.txt</code> file to improve AI visibility."	Unsupported as a general ranking tactic ; explicitly false for Google Search. Google says it ignores <code>llms.txt</code> for visibility/rankings. Other services may choose to consume such a file, but there is no evidence that it is a cross-engine ranking signal.
	Unsupported for Google; Google explicitly says no. Ordinary

“You need special AI/GEO schema.”	structured data can still serve its conventional Search/product/local purposes.
“Schema markup makes ChatGPT/Gemini/AIO cite you more.”	Not supported. Cross-sectional data can make schema look strongly associated with citation because high-quality sites use more schema. A matched longitudinal Ahrefs analysis found no citation uplift in AI Mode or ChatGPT and no convincing positive AIO effect.
“Break every page into small 40–60 word AI-readable chunks.”	No robust organic-search evidence. Google specifically says there is no requirement to break content into tiny pieces for AI understanding.
“Write in a special LLM style: answer-first paragraphs, FAQs everywhere, rigid Q&A formatting.”	Weak as an AI-ranking claim. Clear writing can obviously help humans and may aid downstream extraction, but no robust cross-platform experiment shows that a specific house style causes greater organic retrieval. Google explicitly says no special generative-AI writing style is necessary.
“Generate landing pages for every likely fan-out query.”	Not supported, and potentially contrary to Google guidance. Fan-out itself is verified; mass-producing exact pages for guessed fan-outs is not. Google specifically warns against creating content for every possible query variation and says this can fall under scaled-content abuse when done to manipulate visibility.
“Add quotations/statistics/citations and AI visibility will rise by 30–40%.”	A major overgeneralization of real research. The KDD GEO study did show large gains for some treatments in its benchmark. But the content had already entered the experimental context. The result demonstrates a downstream generation effect, not a 40% increase in live ChatGPT/Google organic discoverability.
“Unlinked brand mentions are the new backlinks.”	Plausible narrative, no strong causal evidence. AI systems clearly can mention entities without citing them, and third-party sources often appear prominently. But no strong experiment establishes a transferable exchange rate between web mentions and AI recommendations. Google specifically warns that pursuing inauthentic mentions is not useful.
“Get mentioned on Reddit/YouTube/Wikipedia and your AI rankings will improve.”	Mostly correlation. These domains appear frequently in citation datasets. That may reflect user-generated expertise, query type, search rankings, authority, available content, Google's own ecosystem, or other confounders. Observing frequent citations does not establish that creating a Reddit post causes your brand to be recommended. For example, Pew found Wikipedia, YouTube and Reddit prominent in <i>both</i> ordinary and AI search results.
“AI engines prefer long-form content.”	Not established as a causal rule. Some datasets find longer pages associated with greater citation influence, but length is heavily confounded with comprehensiveness, authority and subject matter. Google's guidance explicitly says there is no ideal page length.
“GEO is a solved set of ranking factors analogous to traditional SEO.”	Definitely unsupported at present. The best current synthesis finds heterogeneous metrics, poor transfer of generic heuristics, stochasticity and no demonstrated stable cross-platform causal intervention on organic discoverability.

The claims from my previous answer, regraded

If I re-state the earlier answer under this stricter evidentiary standard:

High confidence: Google still rests on SEO fundamentals; Google uses query fan-out; ChatGPT rewrites/searches multiple queries; AI-specific crawler controls exist; Google's AI surfaces can use different models/techniques; citation/mention are genuinely different observable outcomes; `11ms.txt`, special AI schema and mandatory chunking are not required for Google.

Medium confidence / useful hypothesis: broaden keyword research toward topic and question spaces; make genuinely useful facts and evidence easy to identify; monitor citations separately from mentions; cultivate trustworthy third-party coverage; analyze each AI surface separately. These recommendations fit the observed architecture and data, but their incremental causal effects are largely unquantified.

Low confidence when presented as formulas: AI-specific word counts, rigid answer blocks, FAQ-heavy templates, citation-density targets, schema-as-GEO, `11ms.txt` as a ranking tactic, guessed fan-out landing pages, “X Reddit mentions = Y

ChatGPT visibility,” or universal instructions to add stats/quotes because “research proves a 40% lift.” The literature does not currently justify those claims.

The practical conclusion is narrower than most GEO playbooks : **we have good evidence that the retrieval/synthesis environment is different; we have much less evidence about controllable levers that reliably improve organic retrieval within that environment.** Today, the strongest defensible strategy is therefore to retain conventional SEO as the eligibility/relevance foundation , ensure engine-specific accessibility , publish information that is actually worth retrieving, build legitimate reputation across the web, and experimentally measure AI citations /mentions—rather than treating currently observed correlations as a new catalog of ranking factors .

Suppose a website already ranks well organically in Google.

Why might its page still not be cited in an AI-generated answer, and why might the brand itself still not be mentioned or recommended ?

Separate these stages where possible :

- retrieval
- source selection
- citation
- brand mention
- recommendation

Distinguish documented mechanisms from hypotheses or unknowns .

Yes. A strong Google organic ranking only proves that Google considers the page competitive for a particular conventional search result set. It does **not** guarantee survival through the later stages of an AI answer.

A useful model is:

organic ranking "retrieval "source selection "citation "brand mention "recommendation

Each transition can fail for a different reason.

Stage	Why a strong-ranking page/brand can disappear	Evidence status
Retrieval	The AI searches different/rephrased/fan-out queries, uses different indexes/providers, or doesn't search the web at all	Mostly documented
Source selection	Many pages are retrieved; the system chooses a subset useful for particular claims, potentially seeking different or diverse sources	Partly documented , exact scoring unknown
Citation	Being a candidate/relevant source doesn't guarantee being exposed as a visible citation	Separation is documented/observed ; exact attribution rules mostly unknown
Brand mention	The answer may use information from a source without the source's identity being relevant to the prose	Strongly observed ; causes mostly inferred
Recommendation	The brand must satisfy an additional evaluative task—fit, price, features, reviews, context, preferences, etc.	Some vertical mechanisms documented ; general brand-ranking logic opaque

1. Retrieval: the page may never enter the AI's working set

This is the first and perhaps most important distinction .

A. The AI may not search the same query you rank for

Documented .

Google says AI Overviews and AI Mode can use **query fan-out**: the model generates related queries across subtopics and retrieves results for those queries. Google's 2026 documentation gives the example of a user asking how to fix a weedy lawn producing searches such as "best herbicides for lawns," "remove weeds without chemicals," and "how to prevent weeds in lawn."

ChatGPT Search similarly documents that it can rewrite a prompt into one or more targeted searches and then issue additional searches after inspecting the initial results.

So imagine your page ranks #2 for:

best CRM for architects

The AI may actually retrieve around:

CRM with project billing

architecture CRM QuickBooks integration

CRM under \$50/user

CRM reviews for small architecture firms

Your #2 position on the original query says nothing about where your page would appear for those intermediate searches.

This is a **documented mechanism**, not a GEO hypothesis.

B. "Ranks in Google" does not mean "retrievable by every AI engine"

Documented .

A site ranking in Google is obviously accessible to Googlebot, but ChatGPT has its own search-access requirements. OpenAI says sites should allow **OAI-SearchBot** for content to be discovered and included in ChatGPT summaries/snippets.

Perplexity similarly operates **PerplexityBot** and its own search index, supplemented by third-party crawlers.

Therefore:

Google ranking !' evidence of Google accessibility

but not necessarily:

Google ranking !' ChatGPT/Perplexity accessibility

C. ChatGPT might not search the web at all

Documented , platform-specific.

OpenAI says ChatGPT decides whether web search would improve the response; search can also be manually invoked.

If the system answers a question using model knowledge without a live search, your current Google ranking cannot directly cause inclusion in that response.

That distinction is less applicable to Google AI Overviews/AI Mode, whose documentation describes them as grounded in Google's Search index and ranking systems.

D. User context may change what gets retrieved

Documented .

ChatGPT says its search-query rewriting can incorporate location and, where enabled, relevant Memory.

Google likewise documents personalized AI responses and recommendations based on prior searches, preferences, account information, location and other context when personalization is enabled.

Thus two people asking apparently similar questions may not generate identical retrieval sets.

2. Source selection : retrieval still does not mean the page becomes evidence

Assume your page **was retrieved**.

There can still be a selection stage.

Google's 2026 description of its RAG process is particularly useful here. Its core Search systems retrieve relevant pages, after which Google's systems **review the specific information in those pages** to generate the response and show relevant supporting links.

So there are conceptually at least two decisions :

Is this page relevant enough to retrieve?

and then:

Is something on this page useful enough for this generated answer?

A. Your page can be relevant overall but not best for the particular claim

This follows directly from the documented architecture .

Consider a page that ranks very well because it is an excellent broad guide:

“Complete Guide to Solar Panels”

An AI answer might need five very specific propositions :

- current federal tax-credit rules;
- average installation cost;
- panel degradation rate;
- regional sunlight data;
- particular inverter compatibility .

Your broad page could be retrieved but lose each supporting slot to:

- IRS.gov for the tax rule;
- a current pricing database for cost;
- a manufacturer specification sheet for degradation ;
- NREL for solar data;
- the inverter manufacturer for compatibility .

This does **not** mean your page is lower quality.

It means **“best search result for the overall information need” and “best supporting evidence for one generated proposition” are different selection problems.**

The architecture is documented ; the exact page-level selection rule is not.

B. Google explicitly seeks a broader set of supporting pages

Documented in general, exact mechanism unknown.

Google says query fan-out and its models allow it to identify a “wider and more diverse” set of helpful links than conventional search.

That provides another perfectly legitimate reason why a dominant blue-link result might not dominate citations .

If five high-ranking SEO pages all essentially contain the same secondary information while another specialist source provides unique evidence for one sub-question , the AI system can surface the specialist source instead.

What we **cannot** currently say is something like:

“Google has a citation-diversity penalty that suppresses domains already ranking organically.”

Google has disclosed no such formula.

C. Exact AI source-selection weights are largely unknown

This is where claims should stop.

Platform\$ publicly talk about things such as:

- relevance;
- helpfulness;
- quality;
- trust;
- supporting evidence;
- query context.

Perplexity describes Pro Search as synthesizing a “diverse and high-quality set of sources.”

OpenAI says ChatGPT Search ranking uses multiple factors intended to find reliable and relevant information and explicitly says there is no way to guarantee top placement.

But neither OpenAI, Google nor Perplexity publishes a useful numerical model saying:

organic rank × authority × passage relevance × freshness = citation score.

Any vendor claiming to know such a formula is going beyond the public evidence.

3. Citation : selected /relevant information and visible attribution aren't the same thing

This stage is especially easy to misunderstand.

A **citation is a user-interface/output decision**, not simply a synonym for retrieval.

OpenAI's documentation explicitly distinguishes the sources displayed as inline citations from “**other relevant links**” available in the Sources panel.

Google similarly describes AI Search as generating answers while exposing links associated with/supporting parts of the response.

What is generally *not* exposed publicly is the complete internal set of pages considered during generation.

Therefore, from outside the system:

No visible citation does not prove the page was never retrieved.

And, importantly :

Retrieval does not prove the page influenced the generated text.

Those intermediate states are normally unobservable.

What seems to matter ?

A 2026 academic analysis of 21,143 search-layer citations across several AI systems found that **citation selection** and the degree to which cited pages appear to contribute to the answer (“citation absorption”) diverge. Pages with greater estimated influence tended to show stronger semantic alignment and more extractable evidence such as definitions, numbers, comparisons and procedures.

But that is **observational**, not a disclosed ranking algorithm.

It would be legitimate to say:

Pages used deeply by generated answers often have identifiable supporting material.

It would **not** be legitimate to say:

Adding three statistics to your page causes Google AI Mode to cite it.

The study does not demonstrate that.

4. Brand mention : citation of your website does not imply naming your company

Now suppose the AI **does cite your page**.

The brand can still remain invisible.

This distinction has very good observational evidence.

Semrush's June 2026 study recorded 3,981 domain appearances across ChatGPT, Gemini, Google AI Overviews and AI Mode. In that dataset:

- **61.7%** were citations where the brand was **not** mentioned in the generated answer;
- 13.2%** had both citation and mention;
- 25.1%** were brand mentions **without** a citation.

So empirically:

citation !| mention

and

mention !| citation

Why would the brand not be named?

The most defensible explanation is semantic rather than algorithmic.

Suppose IBM publishes an excellent page explaining a particular machine-learning concept.

An answer might say:

“Gradient boosting builds an ensemble sequentially by correcting errors from previous models.”

and cite IBM.

There is no communicative reason for the generated sentence to say:

“According to IBM, gradient boosting ...”

unless IBM's identity is relevant to the claim.

The page is functioning as an **information source**, not as an object being discussed.

That explanation is a **reasonable inference**, rather than a publicly documented “brand mention algorithm.”

Semrush's data supports it indirectly: informational prompts produced high citation rates but much lower brand-mention rates, while comparative prompts produced substantially more mentions.

But again, that is correlation.

We should **not** transform it into:

“Writing comparative content causes ChatGPT to mention your brand.”

The study does not establish that.

A brand can also be mentioned without its site being used

That occurs frequently in the same study.

Possible inputs include:

- other websites discussing the brand;
- structured/product data;

model knowledge ;
search results from third parties.

Which one caused a particular mention is generally **unknown from the output alone**.

This is one reason traditional “domain citation tracking” is an incomplete proxy for brand visibility.

5. Recommendation : being mentioned still does not mean being chosen

Recommendation is a qualitatively different task.

Consider:

“What CRM tools exist for architects ?”

versus:

“What CRM would you recommend for a 12-person architecture practice that uses QuickBooks ,
needs project profitability reporting and has a \$600/month budget ?”

Your company might easily be mentioned in the first answer but fail the second .

The AI is no longer merely determining :

Is Brand X relevant?

It is evaluating:

How well does Brand X satisfy this particular user's constraints relative to alternatives ?

For shopping , some of these factors are explicitly documented

OpenAI says ChatGPT's product selection can consider :

- the query and conversation context ;
- price ;
- reviews ;
- ease of use ;
- first- and third-party product metadata ;
- third-party content ;
- model-generated information from before the new search ;
- safety policies .

It also says not all available products will be shown .

Google similarly says its AI-assisted “Top recommendations ” in Shopping consider factors such as:

- relevance ;
- ratings ;
- price ;
- product features .

Those recommendations use Google's Shopping data aggregated from brands, retailers and other content providers .

So for commerce , it is demonstrably possible for this to happen:

Brand's article ranks #1 organically

but

Brand's product is not recommended

because the recommendation stage is looking at attributes that are not equivalent to the ranking signals that made the article #1.

Recommendation can also be personalized

Google documents personalization of AI responses/recommendations according to user history and preferences.

OpenAI documents use of context, Memory and user constraints in shopping recommendations.

So there may literally be no globally correct answer to:

“What position does our brand rank in ChatGPT recommendations?”

For some recommendation tasks, the output is conditional on user context.

Outside product shopping, recommendation algorithms are much less documented

For questions such as:

best enterprise cybersecurity consultancy

recommended tax adviser for startups

top marketing agencies for SaaS

the platforms do **not** publish comprehensive weighting systems explaining why company A is recommended over company B.

It is reasonable to hypothesize that signals such as:

- third-party descriptions ;
- reputation ;
- reviews ;
- expertise ;
- location ;
- price ;
- specialization ;
- recent information ;
- evidence matching the requested use case

may contribute.

But unless a platform documents those signals for that particular experience, they should remain **hypotheses**, not “AI ranking factors.”

Put the whole sequence together

Suppose `example.com/best-crm` is currently Google's #2 organic result for “**best CRM for architects.**”

The following sequence is entirely plausible:

1. Organic ranking:

Google thinks the page is highly relevant to the conventional query.

2. Retrieval:

AI Mode fans the question out into five narrower searches. Your page appears for two but not the other three.

Documented mechanism.

3. Source selection:

The answer needs pricing, QuickBooks integration and project-profitability evidence. Other pages contain more direct support for those statements.

Documented general architecture; exact selection criteria unknown.

4. Citation:

Those pages become visible supporting links; your page does not. You cannot infer from the output whether your page was considered and discarded or never retrieved.

Internal state unknown.

5. Brand mention:

Your company might still be named because other sources discuss it—or it might not be, because the answer has no reason to name the publisher of an informational source.

Separation empirically observed; mechanism largely opaque.

6. Recommendation :

Even if named, your CRM may fail the user's \$600/month requirement or lack the requested integration, while another product matches better.

Constraint-based selection is documented in product recommendation systems; general B2B recommendation weighting is unknown.

The most important diagnostic implication

If a site already ranks well, “**why aren't we visible in AI?**” is too coarse a question.

The useful questions are:

- 1.
- 2.
- 3.
4. **Retrieval failure:** Are we absent from the searches/subtopics the AI actually explores?
5. **Selection failure:** Are we retrieved but less suitable as evidence for the claims being generated?
6. **Citation failure:** Are competing sources receiving visible attribution instead?
7. **Mention failure:** Is our content supplying information while our entity is irrelevant or insufficiently surfaced in the prose?
8. **Recommendation failure:** Are we known and mentioned, but losing the actual evaluation against the user's constraints?

Those are five different failure modes. Current public evidence supports the existence of the stages much more strongly than it supports any universal checklist for manipulating them.

And that is why **excellent Google SEO can be necessary yet insufficient for AI recommendation visibility**: conventional ranking gets a document into a strong competitive position, while an AI answer subsequently makes additional decisions about *which questions to search, which evidence to use, which sources to expose, which entities to name, and which alternatives best satisfy the user's decision criteria*.

If you were advising an experienced SEO team for the next 90 days, what would you actually change because of AI Search?

Separate your recommendations into:

- continue doing because traditional SEO remains foundational
- change or add because AI-generated discovery creates a genuinely different requirement or outcome
- test because the evidence is promising but not conclusive
- do not treat as proven despite current industry claims

, AI crawler access, query fan-out and AI visibility measurement where relevant.

For an experienced SEO team, I would make **four real changes over the next 90 days**: add AI crawler governance, make research fan-out-aware, instrument AI visibility as a separate measurement layer, and run controlled experiments on citation/brand visibility.

I would **not** rebuild the content program around “GEO formatting,” add speculative schema, or launch an `llms.txt` initiative.

1. Continue doing because traditional SEO remains foundational

Technical SEO: no separate Google-AI technical stack

Keep crawlability, indexability, canonicals, internal linking, rendering, duplicate-content control, page experience and important information in accessible HTML/text at the center of the program.

Google now states explicitly that its generative Search experiences are rooted in its **core Search ranking and quality systems**. AI Overviews/AI Mode use Google's Search index and core ranking systems for retrieval, and pages need to be indexed and eligible to appear in Search.

90-day action: Don't divert technical SEO resources into a separate "AI technical optimization" backlog for Google. Fix the same crawl/index/rendering problems you would already prioritize.

Schema and structured data: maintain it, but don't relabel it as GEO

Keep implementing **supported structured data where it serves an established purpose**: rich results, products, local businesses, organizations and other normal Search features.

Google says structured data can help Search understand page content, but its current AI guidance is unusually clear:

- structured data is **not required** for generative AI Search;
- there is **no special AI schema.org markup**;
- over-focusing on structured data for AI visibility is unnecessary.

This is reinforced by a reasonably strong quasi-experimental Ahrefs study. It tracked 1,885 pages adding JSON-LD against roughly 4,000 matched controls. It found no statistically distinguishable citation uplift in ChatGPT or Google AI Mode; the AIO result was slightly negative and not something the authors interpreted as a useful causal effect. That does **not prove every schema type has zero AI value**, but it is strong evidence against "add schema and AI citations go up" as a generic claim.

90-day action: Maintain schema quality. Fix invalid or stale markup. Add schema where normal Search documentation supports it. **Do not create an "AI schema" project.**

For ecommerce and local businesses, give particular attention to accurate Merchant Center feeds and Google Business Profiles. Google explicitly says these can help product/service information appear in both generative AI responses and other Search experiences.

Content quality and originality

Continue investing in content that is genuinely distinctive: first-hand expertise, original research, proprietary data, demonstrations, useful tools, expert analysis and information that is difficult to reproduce from generic web summaries.

Google's July 2026 AI guidance specifically emphasizes **non-commodity content** and says its generative systems examine multiple sources; simply rephrasing information already available everywhere is unlikely to be a durable advantage.

This isn't really a new AI tactic. It is traditional quality SEO becoming more important because commodity information is particularly easy for generative systems to synthesize without sending users to the original publisher.

Links, authority and legitimate PR

I would continue doing high-quality digital PR and link earning for all the traditional reasons.

What I would **not** do is tell the PR team, "unlinked mentions are now the new backlinks." We do not have causal evidence for such an equivalence.

Google itself warns against pursuing inauthentic mentions for AI visibility.

2. Change or add because AI-generated discovery creates a genuinely different requirement or outcome

These are the things I would actually add to the operating model.

A. Add AI crawler governance to the technical audit

This is genuinely additive.

For Google AI Overviews/AI Mode, normal Googlebot/Search controls apply. There is no special AIO crawler.

But other systems introduce separate crawlers.

OpenAI says that publishers wanting content included in ChatGPT Search summaries/snippets should make sure they do not block **OAI-SearchBot**. OpenAI also separates OAI-SearchBot from GPTBot, which relates to potential training rather than Search inclusion.

Perplexity likewise says **PerplexityBot** builds/supplies its search experience and recommends allowing it if you want pages

surfaced in Perplexity search.

Days 1–14:

- Audit:
- - robots.txt
 - CDN/WAF bot rules
 - bot-management services
 - IP allow/block policies
 - server logs
 - noindex /snippet controls

for at least:

- Googlebot
- OAI-SearchBot
- PerplexityBot

Then make an explicit business-policy decision about search crawling versus model-training crawling rather than conflating the two.

This is probably the clearest **new technical requirement** created by cross-platform AI search.

B. Change keyword research into query-family/fan-out research

Do **not** abandon keywords.

Instead, add another layer.

Google documents that AI Overviews and AI Mode may issue multiple concurrent **fan-out queries** around the original question. OpenAI similarly says ChatGPT Search can rewrite the user's query into one or more targeted searches, review results, and issue additional searches.

So for strategically important topics, stop representing the opportunity with only:

enterprise payroll software

Also model questions such as:

- alternatives;
- comparisons;
- implementation;
- compatibility;
- pricing;
- use-case constraints;
- problems/risks;
- terminology;
- evaluation criteria;
- follow-up questions.

Important: this does **not** mean publishing a landing page for every guessed fan-out query.

Google explicitly warns against creating pages for every query variation merely to manipulate AI or Search results.

What I'd change operationally

For the top 20–50 revenue-driving topic areas, create a **query-family map**, not another giant keyword spreadsheet.

For each topic capture:

Core intent !' likely subquestions !' comparison criteria !' factual dependencies !' existing best URL !' content gaps

Then decide whether a gap requires:

- improving an existing page;
- creating a genuinely distinct resource;
- or doing nothing because the existing page already answers it.

That is a real response to fan-out without turning fan-out into scaled-content production .

C. Add AI visibility measurement as its own layer

This is probably the biggest organizational change.

Do not replace traditional SEO KPIs. Add another measurement layer.

Google began rolling out dedicated **Generative AI performance reports** in Search Console in June 2026. They expose generative-AI impressions, pages, countries, devices and time trends for AI features such as AI Overviews and AI Mode. The rollout initially covered a subset of sites.

OpenAI also makes ChatGPT Search referrals relatively measurable: referral URLs automatically include `utm_source=chatgpt.com`.

I would create five separate AI metrics

Don't call all of these "AI ranking":

4. **AI feature impressions**
5. **URL citations / visible source appearances**
 - Brand mentions**
 - Brand recommendations**
 - Referral traffic + conversions**

And I would deliberately **not** create a KPI called "retrieval rate" unless you actually have evidence that the system retrieved the page. Usually you do not.

Absence of a visible citation does not establish non-retrieval.

Build a controlled prompt panel

Choose perhaps 50–200 commercially important prompts, stratified by:

- informational ;
- comparison ;
- commercial investigation ;
recommendation ;
branded/non-branded ;
- key vertical/use-case.

Track them separately in:

- Google AI Mode/AIO;
- ChatGPT;
- Gemini where relevant;
- Perplexity.

Repeat measurements rather than treating one run as a ranking report. Recent research emphasizes that generative visibility is stochastic and that citation, content absorption and downstream outcomes should be treated separately.

This gives you something much more useful than:

"We're position 3 in ChatGPT."

There generally isn't a stable equivalent of position 3.

D. Broaden "entity optimization ," but define it carefully

I would use the term **entity data hygiene**, not "entity optimization ," because the latter is often sold as something more deterministic than the evidence supports .

For your own entity, ensure facts are clear and consistent :

- official company/product names;
- ownership/relationships where relevant;
- product/category descriptions ;
- pricing where public ;

- locations ;
- contact information ;
- leadership/authorship ;
- product attributes ;
- Business Profile ;
- Merchant Center ;
- appropriate Organization/Product/LocalBusiness structured data.

That's good SEO/data hygiene regardless.

For Google specifically , Business Profile and Merchant Center are explicitly mentioned as information sources relevant to AI Search experiences .

What is **not** established is that adding more `sameAs` links, changing Organization schema, or hitting some vendor-defined “entity score” causes ChatGPT/Gemini recommendation visibility .

I would therefore invest in **entity accuracy** , not an abstract “entity authority score.”

3. Test because the evidence is promising but not conclusive

This is where I'd put most genuinely experimental GEO work .

A. Test content that is easier to use as evidence

There is reasonable evidence that once a document is in a generative system's context , changes to its content can affect whether/how prominently it is used .

The original GEO research experimentally altered documents and observed visibility effects of up to about 40% in its benchmark . But the important limitation is that the source was already present in the experimental context . It did **not** prove a 40% improvement in organic ChatGPT/Google retrieval .

A 2026 study also found that highly influential cited pages tended to contain more readily identifiable evidence—definitions , numerical facts , comparisons and procedural steps—but those findings are primarily observational .

So I would test, rather than mandate :

- clearer definitions ;
- explicit factual claims ;
- original statistics ;
- methodologies ;
- source provenance ;
- comparison tables where useful to humans ;
- dated information ;
- direct answers to genuinely important subquestions .

Use matched pages or controlled groups where possible .

Measure :

organic rankings + AI citations + brand mentions + conversion, not citations alone .

If citations improve but organic performance or conversions decline , it wasn't a successful optimization .

B. Test clearer content structure—but not “AI chunking ”

Headings , paragraphs , tables and logical sections already improve human usability .

Google recommends well-organized content , but explicitly says there is **no requirement to break content into tiny AI-readable chunks** , no ideal page length and no need for a special AI writing style .

So I'd test :

better human information architecture

rather than :

“every answer must be 40–60 words because LLMs prefer chunks.”

A reasonable experiment might revise 20 pages so that important claims , definitions and comparisons are easier to locate , while

leaving another comparable group unchanged.

That can produce useful internal evidence without pretending the formatting rule is universal.

C. Test legitimate third-party presence

This one deserves attention, but not dogma.

One 2025 research study found that several AI search systems disproportionately surfaced third-party/earned media relative to brand-owned material in its tested query sets.

That makes a sensible hypothesis:

brands that are accurately and independently described by credible external sources may have more opportunities to enter AI synthesis and recommendation.

But it does **not** establish:

gaining ten unlinked mentions increases ChatGPT recommendations by X%.

I would therefore test expansion of **real** third-party coverage:

- authoritative industry publications;
- independent reviews;
- relevant expert roundups;
- customer case studies on partner sites;
- credible forums/community participation;
- analyst or association resources where legitimate.

Then measure whether brand mention/recommendation frequency changes.

Don't manufacture mentions. Google explicitly discourages that behavior.

D. Test recommendation -specific content separately from citation content

If recommendations matter commercially, create a separate test set around decision prompts such as:

Best X for Y

X vs Y

Which product fits condition Z?

Alternatives to X for small teams?

Then audit whether the web contains enough accurate evidence to distinguish your product on the criteria users actually ask about.

That might expose gaps in:

- pricing clarity;
- specifications;
- integrations;
- suitability;
- limitations;
- geographic availability;
- customer type.

This is different from simply trying to get a blog article cited.

The causal evidence for general B2B AI recommendation optimization is weak, so treat this as **product-information/reputation experimentation**, not an established ranking-factor program.

4. Do not treat as proven despite current industry claims

This is the list I would explicitly put in the team's operating guidelines.

`llms.txt` as an AI Search ranking tactic

Do not prioritize it for search visibility.

Google says Search ignores `llms.txt`; it neither helps nor hurts Google Search or its generative Search visibility.

Ahrefs also analyzed 137,210 domains in May 2026. Of roughly 38,000 sites publishing valid `llms.txt` files, **97% received zero requests** during the study period; AI retrieval bots accounted for only a small fraction of requests among files that were fetched. That is observational log data—not proof the file can never become useful—but it provides little reason to make it a 90-day SEO priority.

Exception: if you have a specific non-search agent/tool ecosystem that explicitly consumes `llms.txt`, treat that as a separate agent/developer use case.

“AI-specific schema”

Don't do it.

There is no special Google AI schema requirement.

Likewise, don't extrapolate the fact that AI-cited sites often use schema into:

schema causes AI citations.

The matched longitudinal evidence cited above does not support that inference.

Mandatory 40-word answer blocks / rigid “LLM-friendly” formatting

Not proven.

Google explicitly rejects mandatory chunking and AI-specific rewriting as requirements.

Clear writing is good. A universal LLM paragraph formula is not established.

FAQ sections everywhere

Don't add FAQs merely because someone says “LLMs love FAQs.”

Use Q&A formatting when it genuinely improves the page.

There is no robust cross-platform causal evidence that adding FAQ blocks to otherwise equivalent pages reliably improves organic AI retrieval.

“Add statistics and citations for a 40% GEO boost”

This is a misleading reading of the original GEO experiment.

The research provides useful evidence that modifying already-present source content can change downstream generative visibility. It does **not** show that adding statistics increases the probability that your page is organically discovered by Google AI Mode or ChatGPT by 40%.

“Unlinked mentions are the new backlinks”

Not proven.

Treat legitimate external coverage as potentially useful reputation/discovery evidence, but don't invent a new PageRank-style model for mentions without evidence.

“We need pages for every fan-out query”

No.

Fan-out is real. Mass-producing pages around guessed fan-out variants is not a justified consequence—and Google explicitly warns against it.

“Our GEO tool knows the AI ranking factors ”

Treat such claims skeptically .

Google says third-party SEO/AEO/GEO tools do not have access to its internal ranking data and cannot guarantee performance .

A July 2026 review of 45 GEO studies reached an even broader conclusion : although already-retrieved content can causally affect citation/use in some experimental environments , it found **no reviewed technique demonstrating a stable, longitudinal, cross-platform causal improvement in organic AI discoverability and downstream behavior**.

The 90-day program I would run	
TimingWork	Audit OAI-SearchBot/PerplexityBot/Googlebot access; establish policy for search vs training crawlers; enable/inspect Google's Generative AI Search Console report if available; segment ChatGPT referral traffic; baseline 50–200 important prompts across engines; audit existing schema rather than adding speculative markup.
Days 1–30	Build query-family/fan-out maps for your highest-value topics; identify missing substantive information rather than keyword variants; audit entity/product/company facts across owned surfaces, Merchant Center/GBP where applicable; identify citation vs brand-mention vs recommendation gaps.
Days 31–60	Run controlled tests on clearer evidence presentation/content structure; run a legitimate third-party coverage experiment in selected topic areas; remeasure the fixed prompt panel multiple times; compare AI outcomes with traditional SEO and conversion metrics; kill tactics that don't outperform controls.
Days 61–90	

My resource allocation

For a site that already has a mature SEO program , I would roughly think about the incremental 90-day AI effort as:

~40% **measurement and diagnostics** — because teams currently know much less about where they're failing in the AI pipeline than they think.

~25% **technical/crawler and data hygiene** — mostly a finite audit, not permanent headcount .

~20% **fan-out/topic analysis and substantive content-gap remediation** .

~15% **controlled experiments** around content evidence, third-party visibility and recommendation prompts .

I would allocate **close to zero incremental budget** to llms.txt , “AI schema,” mass FAQ production , rigid AI-writing templates or buying manufactured brand mentions .

The operating principle would be: **treat documented architectural differences as requirements; treat plausible optimization levers as experiments; and require causal or at least longitudinal evidence before turning an experiment into SEO doctrine**.