

Cross-Source Evidence Analysis: SEO for AI Search

Method and source limitations

I treated the four files as a closed evidence set. I did **not** independently verify any cited study, platform document, statistic, domain, or mechanism against the web. Therefore, below:

VERIFIED / DOCUMENTED means the supplied panel points to first-party documentation or reasonably strong

- experimental evidence for the narrowly stated claim.
 - It does **not** mean I have independently verified that external source.
- Agreement among the four systems is treated only as agreement, never as proof.

There are two important completeness problems in the supplied panel. The ChatGPT PDF begins with the evidence-challenge response rather than the original initial answer, although ChatGPT later explicitly “regrades” claims from its missing earlier answer. The Claude PDF contains the initial response, evidence challenge, and retrieval recommendation follow-up, but ends before a separate response to the final 90-day prompt; its initial response does, however, contain several 90-day recommendations.

Those gaps matter particularly for Section 6 and Section 9.

1. Semantic Vocabulary Comparison

Concept	ChatGPT	Gemini	Claude	Perplexity	Interpretation	
SEO				Core/default discipline; uses “conventional SEO,” “technical SEO,” “SEO fundamentals.”	Explicitly contrasts SEO with GEO but says GEO builds on SEO.	Treat SEO found descri chan
AI SEO				Not used as its own category.	Appears in a cited source title, not as Gemini's principal category.	“rewe Appe sourc SEO not it categ
GEO				Used, but often critically: “GEO marketing,” “GEO playbooks,” “GEO tool.”	Primary category label: “Generative Engine Optimization.”	Seco more “AI di “AI-s
AEO				Appears only in a late reference to “SEO/AEO/GEO tools.”	Not naturally used.	Not n used.
LLM SEO				Not used.	Not used.	Not u
AI Search optimisation / optimization				Uses “AI Search,” not this as a defined discipline.	Uses “AI Search execution,” but GEO is the category.	Uses searc disco
AI visibility				Frequent measurement/output	Used for	Uses meas

	term.	outcomes/SOV.	and “ platfo
Citation	Extremely explicit; separates citation from retrieval, influence, mention	Uses citation throughout and later	Uses throu calls i
	and recommendation . Central architectural concept and	separates it as Stage 3.	and u docu
Retrieval	explicitly separate from citation .	Central RAG/vector concept .	Centr conce
Recommendation	Treated as a separate evaluative outcome ; especially careful about product vs general	Explicit fifth stage; emphasizes off-site sentiment and attribute matching.	Explic stage emph comp party but la
	B2B recommendation .		mech larget
Entity optimisation	Explicitly says it would prefer “ entity data hygiene ” because “entity optimization”	Uses entity associations/Knowledge Graph language rather than making “entity optimization” its main label.	Prefe establ “entit
	sounds too deterministic .		
Query fan-out	Central documented architectural term.	Explicitly uses fan-out and hidden subqueries.	Explic query

Gemini’s initial framing is explicitly “SEO !” Generative Engine Optimization (GEO),” with a shift toward becoming “a citable data node.” Perplexity instead uses “AI search (GEO/AEO)” as a combined category. Claude concludes that the change is “not a new SEO discipline” but a reweighting. ChatGPT goes further on one term and explicitly substitutes “entity data hygiene” for “entity optimization.”

Vocabulary conclusion

The vocabulary is **fragmented at the discipline/category level** but **much more stable at the mechanism/outcome level**.

Stable concepts are:

SEO !’ retrieval !’ citation !’ mention !’ recommendation !’ measurement

Less stable labels are:

GEO / AEO / AI SEO / LLM SEO / entity optimization

That distinction itself is useful: the panel agrees more about *what processes exist* than about *what the new discipline should be called*.

2. What Does Not Change
Practice / PrincipleChatGPTGeminiClaudePerplexityConsensus StrengthEvidence Cited
Crawlability / indexability

Foundational; Google AI pages still need Search eligibility.	Technical access and crawlability foundational.	Technical crawlability/index hygiene foundatio	
---	--	--	--

Technical site health / accessible HTML	Keep rendering, canonicals, internal links, page experience, accessible HTML. Strongly retained; originality and	Emphasizes blocking, JavaScript and crawler access. E-E-A-T, original	CWV, mobile, rendering, internal linking remain.	
High-quality / useful / original content	proprietary information emphasized. Continue link/PR	research, expert credentials. Backlinks	Accuracy, genuine expertise/experience	
Authority / links / reputation	work but reject "mentions replace backlinks."	and authority remain non-negotiable. Later explicitly rejects	Link building/domain authority retained reweighted.	
Keyword / topic / intent research	"Do not abandon keywords"; add query-family research.	"keywords don't matter anymore."	Keyword/topic research remains a base layer.	
Internal architecture / linking	Explicitly retained.	Not emphasized as strongly.	Explicitly retained.	
Structured data for established SEO purposes	Keep it for valid Search/product/local purposes; not an AI citation hack.	Calls schema foundational and entity-parsing infrastructure.	Keep for rich results/KG disambiguation, but not generic citation lift.	
Organic ranking still matters	Related to AI visibility but non-deterministic.	Strong top-10 overlap claim.	Says ranking remains strong but not sufficient.	

Interpretation

The four-system panel does **not** support "replace SEO with GEO."

The stronger shared position is:

Traditional SEO establishes eligibility, relevance, crawlability, quality and authority. AI-generated search adds further decisions above that foundation.

The main dispute is therefore **not whether SEO remains necessary**. It is **how much additional optimization can reliably influence the later stages**.

3. What Actually Changes				
Proposed Change	ChatGPT	Gemini	Claude	Perplexity
Mechanism	Agreement	Claimed		
Evidence	Quality			
Ranking: multiple AI outcomes	Separates impressions, citations, mentions, recommendations, traffic/conversions.	Moves from ranking/clicks toward citations, mentions and SOV.	Citation frequency	
Query fan-out / query rewriting	Google + ChatGPT	Says AI Mode/Perplexity	Google	

	documented .	create hidden subqueries .	treats
	Strongly additive		
AI-specific crawler access	outside normal Google SEO.	Recommends AI crawler audit.	Traini separ
Retrieval "" organic rank	Explicit.	Explicit passage/vector	Explic
Source selection after retrieval	Explicit second decision; exact weights unknown .	retrieval distinction . Calls consensus/fact density selection mechanisms	Says r weigh
Passage-level usefulness / extractability	Testable, not established as organic retrieval	"documented ." Strongly recommends answer-first/fact-	Stron recom formul
Third-party evidence becomes more important	lever. Test legitimate external coverage.	density formatting . Strongly emphasizes Reddit/reviews/news consensus.	Initial citatio chann
Entity clarity	Reframes as data hygiene, not deterministic	Strong Knowledge Graph/entity framing.	Calls import qualifi
Platform-specific optimization	optimization . Strongly emphasizes surface	Gives engine-specific crawler/retrieval	Differ prefer
Separate AI visibility measurement	differences. Fixed prompt panel, repeated runs, separate	behavior. Prompt/SOV tracking .	Repe becau are un
Personalization / user constraints	outputs . Explicitly important for recommendation .	Attribute matching emphasized.	Googl menti recom

Recommendation becomes a separate evaluative problem

Strong distinction ; shopping factors documented , general B2B opaque .

Explicit Stage 5.

Explic endor menti

The most important distinction appears in ChatGPT's challenge response: the KDD GEO experiment can show that editing a document **already present in context** changes how a model uses it, but that does not establish that the edit increases **organic retrieval**.

That distinction is the key fault line running through much of the panel.

4. Evidence Classification

ClaimSystems Making ClaimClassificationEvidence Cited in PanelSystems That				
Qualify / DisagreeConfidence				
Foundational SEO remains necessary for Google AI Overviews/AI Mode	All	VERIFIED / DOCUMENTED , scoped to Google	Repeated attribution to Google's own AI Search guidance .	

Query fan-out exists in Google AI Search	All	VERIFIED / DOCUMENTED	Google platform documentation repeatedly invoked.	
ChatGPT Search can rewrite/search multiple queries	ChatGPT, Claude; implied by others	VERIFIED / DOCUMENTED as represented	OpenAI documentation.	
AI search/retrieval crawler controls create new cross-platform technical requirements	All	VERIFIED / DOCUMENTED, bot-specific	OpenAI/Perplexity crawler documentation; Claude additionally cites Anthropic bot split.	i t r
Google-Extended is required for Google AIO/AI Mode visibility	Perplexity initial language can imply this; Gemini's crawler list is ambiguous	CONTRADICTED WITHIN SUPPLIED MATERIAL	ChatGPT says Google-Extended does not control Google Search/AIO inclusion.	
GPTBot controls ChatGPT Search inclusion	Gemini initial / Perplexity initial may imply it	CONTRADICTED / LATER QUALIFIED	Later answers distinguish GPTBot training from OAI-SearchBot search inclusion.	l
Passage/chunk-based processing exists in generative retrieval	All	VERIFIED / DOCUMENTED at architectural level, but platform scope varies	RAG/search architecture + platform descriptions.	i
“40–60 words” / “130–170 words” / “250–500 words” are optimal citation units	Gemini, Perplexity initially; vendor material cited elsewhere	UNPROVEN	Mostly playbooks/inference; Claude and ChatGPT explicitly reject precise universal formulas.	
Answer-first formatting causes more organic AI citations	All recommend/test to differing degrees	OBSERVATIONAL/ INFERENCE	Page characteristics + RAG reasoning; conditional GEO experiment.	
Adding statistics/quotes can change model use once source is already in context	ChatGPT, Gemini, Claude	VERIFIED / EXPERIMENTALLY SUPPORTED, CONDITIONAL	KDD GEO benchmark.	i
Adding statistics/quotes raises live ChatGPT/Google organic discoverability 30–40%	Gemini comes closest; common industry claim discussed by others	UNPROVEN / OVERGENERALIZED	Same KDD experiment did not establish retrieval lift.	t
Schema correlates with AI-cited pages	ChatGPT, Claude, Perplexity	OBSERVATIONAL/ CORRELATIONAL	Multiple cross-sectional industry studies.	
Adding generic schema causes higher AI citation rates	Gemini is most positive; others discuss it	CONTRADICTED WITHIN SUPPLIED MATERIAL	Ahrefs matched/longitudinal experiment reported little/no positive lift.	
Product/Review schema with rich attributes may help some AI contexts	Claude	INSUFFICIENT INFORMATION / OBSERVATIONAL	Separate cross-platform study claimed higher citation rates.	r
FAQ schema itself raises AI citations	Perplexity initial/final test; indirectly Gemini	UNPROVEN	No strong controlled panel evidence.	
llms.txt improves Google AI visibility	Industry claim discussed by all	CONTRADICTED FOR GOOGLE	Google guidance as represented by ChatGPT/Claude/Perplexity.	
llms.txt helps non-Google agents/search systems	Gemini/Claude/Perplexity leave possibility open	INSUFFICIENT INFORMATION	No causal panel evidence.	

Brand mentions correlate with AI visibility	ChatGPT, Claude, Perplexity; Gemini similar	OBSERVATIONAL / CORRELATIONAL	Ahrefs-style brand mention studies.	
Increasing unlinked mentions causes more recommendations	Initial Claude/Gemini language can imply it	UNPROVEN	No isolated intervention.	
Third-party/earned media is disproportionately selected by some AI systems	ChatGPT, Claude, Gemini, Perplexity	OBSERVATIONAL / CORRELATIONAL	One or more source-selection datasets.	r l
Posting on Reddit/YouTube/Wikipedia causes AI visibility lift	Industry implication discussed by several	UNPROVEN	Source-frequency datasets only.	
Organic ranking and AI citation are related but not deterministic	All	OBSERVATIONAL / CORRELATIONAL	Several overlap studies.	
Exact top-10 overlap rate is ~76%	Gemini / upper end Perplexity	CONTRADICTED WITHIN PANEL AS A UNIVERSAL NUMBER	ChatGPT cites 37.1%; Claude reports highly divergent studies.	
Freshness matters in Perplexity retrieval/reranking	Perplexity, Gemini; echoed elsewhere	DOCUMENTED / STRONGLY SUPPORTED for Perplexity as represented	Perplexity documentation invoked.	
Freshness matters more than traditional SEO across AI search	Gemini/Perplexity	INFERENCE	No controlled side-by-side panel evidence.	l
Citation and brand mention are different outcomes	ChatGPT, Perplexity; conceptually all four	OBSERVATIONAL / STRONGLY SUPPORTED	Semrush-style datasets show citations without mentions and vice versa.	
Citation and actual source influence/use are different	ChatGPT	OBSERVATIONAL / STRONGLY SUPPORTED	Citation-absorption study.	
Recommendation is not equivalent to mention or citation	All after pipeline follow-up	STRONGLY SUPPORTED conceptually; exact algorithm mostly unknown	Shopping/product documentation + observed behavior.	
AI answer visibility varies significantly run to run	Claude, ChatGPT	OBSERVATIONAL / PREPRINT EVIDENCE	Repeated-prompt study/preprint.	
AI visibility tools provide a precise universal “ranking” metric	Industry claim challenged by ChatGPT/Claude	UNPROVEN	Methodologies /prompt sets differ and are often undisclosed. Gemini calls Bing	
ChatGPT visibility requires Bing indexing / Bing Webmaster Tools	Gemini strongly; Claude partially	INSUFFICIENT INFORMATION / CONTESTED WITHIN PANEL	dependence mandatory; Claude says Bing-primary but notes possible Google-index evidence; ChatGPT doesn't support the mandatory conclusion.	

“Consensus filtering” is a documented source-

Gemini

INSUFFICIENT

Asserted, but direct supporting platform

selection rule

INFORMATION

evidence is not exposed in the supplied Gemini citations.

Perplexity itself eventually states that answer-first structure lacks a large controlled causal experiment and that schema correlation does not establish schema causation. This is materially narrower than some of its initial recommendations.

5. Source Ecology

Recurring / analytically important sources

Source / Domain	ChatGPT	Gemini	Claude	Perplexity	Number of Systems Source Type Claims It Supports
Google official Search documentation	'	Invoked in prose	'	'	' , direct developer
OpenAI crawler/Search documentation	'	Invoked in prose	'	'	' , platform.openai.com/
Perplexity documentation	' /invoked	Indirect	' /invoked		
Princeton/Georgia Tech KDD GEO work	'	'	'		R
Ahrefs	'	—	'		
Semrush	'	—	'		
Pew Research	'	—	'		
FirstMotion	—	'	—		
OptimizeGEO.ai	—	'	—		
Passionfruit	—	'	—		
QuickSEO	—	'	—		
Blck Alpaca	—	'	'		

Anagram

— — '

Digital Applied

— — '

Machine Relations

— — '

Important caveat: the ChatGPT and Claude PDFs are flattened in a way that exposes many **source names/provenance descriptions but not recoverable citation URLs**. Gemini and especially Perplexity expose more clickable/printed URLs. Consequently, raw domain counts would unfairly make Perplexity look more sourced simply because its PDF preserves its link list better.

Complete Gemini external -source inventory visible in the file

Gemini's four answer segments expose these secondary sources:

Initial answer: growthproai.com , naturaily.com , llmrefs.com , optimizegeo.ai , clickforest.com .

Evidence challenge: [blckalpaca.at](#) , [elementera.com](#) , [centralcoastseo.com.au](#) .

Pipeline answer: [seocrawl.ai](#) , [firstmotion.com](#) .

90-day answer: [getpassionfruit.com](#) , [derivatex.agency](#) , [quickseo.ai](#) .

Notably, Gemini's prose repeatedly invokes official platform documentation , but **direct Google/OpenAI documentation is not among the visible supporting links in these answer sections**.

Complete Perplexity linked-domain inventory, grouped by claim family

Perplexity provides an unusually large numbered source list running to source **#163**. Rather than treat citation volume as quality, the domains below are grouped by what they were being used to support .

Claim familyVisible linked	
domains	
Broad SEO/GEO framing	kreativ.de , stackmatix.com , strapi.io , synfinitydynamics.com , w3era.com , witscode.c
Google AI guidance / "still SEO"	connascent.com , digitalapplied.com , ecorpit.com , getpassionfruit.com , ivabot.xyz ,
OpenAI crawler controls	agentsurge.io , ai-advisors.ai , crawloria.com , crawlreadyai.com , icoda.io , plus inline
Perplexity source selection	citegeo.ai , datastudios.org , firstmotion.com , kompozy.io , singularity.digital , zip
Answer-first / structure / freshness	digitalapplied.com , infinitemediaresources.com , ritnerdigital.com , signals.sh , w3er
Schema / structured data	aio-rankings.com , deepsmith.ai , heybuffy.com , isagentready.com , marcodiversi.com ,
Brand mentions / visibility	ai-visibility-software.com , citeflow.io , lumengeo.co , machinerelations.ai
AIO clicks / freshness Retrieval /	claritydigital.agency , parse.gl , sqmagazine.co.uk
ranked pages missing AI answers	aiboost.co.uk , cite.solutions , contently.com , digitalstrategyforce.com , maxaio.ai ,
Source selection / citation	authoritytech.io , cited.md , discoveredlabs.com , firstmotion.com , gist.github.com ,
mechanics Ghost	
citation / mention gap	aiboost.co.uk , answerlab.au , authoritytech.io , digitalstrategyforce.com , elmohq.com ,
Entity / crawler optimization	anagram.ai , authoritytech.io , digitalapplied.com , geo.tenten.co , geoxylia.com , get
Query fan-out	ahrefs.com , astiva.ai , llmranks.io , nobori.ai , sanbi.ai , wellows.com , plus citeflow ,
AI visibility measurement	airankchecker.net , cloro.dev , flint.com
Additional Google	

guidance / AI developers.google.com , hyperfx.ai , infinacode.com , jdsupra.com , margen.net , meowapp tooling

The numbered list itself demonstrates how heavily Perplexity's evidence chain depends on industry/SEO/GEO sources; direct Google documentation is visible at #84 and #160, while most entries are secondary interpretations or commercial/industry research.

Claude identifiable source ecology

Claude's flattened PDF exposes names including Google, OpenAI, Anthropic, Ahrefs, Semrush, Pew Research, arXiv, SSRN, Vercel/MERJ, Anagram, Machine Relations, Digital Applied, BetterAISearch, Shadow, Layer3Labs, MADX Digital, Evertune, SEO-Kreativ, Arfadia, Blck Alpaca, Lemniscate Growth, Discovered Labs, Yoast, Vistrix Labs and others. Its evidence-challenge response noticeably shifts toward platform documentation and explicit evidence tiers.

ChatGPT identifiable source ecology

ChatGPT repeatedly names or invokes Google documentation , OpenAI documentation , Perplexity documentation , the KDD GEO paper, Ahrefs, Semrush, Pew, a 2025 third-party/earned-media study, a 2026 citation-influence analysis and a July 2026 review of 45 GEO studies .

Cross-system sources

The strongest recurring provenance layer is:

Google/OpenAI documentation !' KDD GEO research !' Ahrefs/Semrush-style original industry research.

The most frequently recurring secondary industry sources include FirstMotion , Passionfruit, QuickSEO, OptimizeGEO, Digital Applied, Machine Relations and Anagram.

System-specific sources

Gemini has a compact , largely secondary-industry source set.

Perplexity has by far the widest source ecology, but much of that width is a very long commercial/industry tail rather than primary platform or academic documentation .

Claude incorporates some unusual sources not prominent elsewhere—Vercel/MERJ crawler logs, arXiv, SSRN and several specialist AI-search analyses.

ChatGPT puts unusually high analytical weight on methodological distinctions in the KDD work and the 45-study review.

Primary-source preference

On the evidence visible in the PDFs:

- ChatGPT** most consistently reasons from the limitations of platform documentation and research design.
- Claude** moves strongly toward first-party/platform evidence after being challenged.
- Perplexity** does include direct official sources, but surrounds them with a very large secondary-industry citation network.
- Gemini** invokes official documentation in prose but exposes almost entirely secondary links in its citation lists.

This is about **evidence sourcing**, not answer quality.

Commercial -source dependence

Perplexity has the **highest absolute dependence** on agency/vendor/industry sources because of the size of its source list. Gemini appears to have the **highest proportional dependence among its visible external links**, because its exposed sources are almost entirely secondary rather than direct platform documentation . Citation count alone therefore gives a misleading impression of evidence strength.

6. Follow-Up Stability Analysis				
SystemClaimInitial PositionPosition After Evidence ChallengeChanged?				

Significance			Explicitly says evidence base is narrower than GEO marketing and regrades claims into	Ca re re
ChatGPT	Overall causal confidence	Initial answer is missing from file.	high/medium/low confidence . “Useful hypothesis”; conditional evidence	
ChatGPT	Content structure/evidence-rich copy	Earlier claim apparently stronger.	once document is already retrieved, weak evidence for organic retrieval.	Do
ChatGPT	Third-party coverage	Apparently part of earlier advice.	Medium-confidence hypothesis; causal lift unquantified .	Do
ChatGPT	Exact formats/schema/ 1lms.txt	Earlier claims referenced in regrade.	Exact word counts, rigid templates, schema-as-GEO and 1lms.txt are low confidence .	Do
Gemini	Schema	Initially: schema “remains essential.”	Still calls schema-streamlined entity parsing Tier 1 / verified.	Es m
Gemini	Statistics/quotes/content restructuring	Initially claims up to 40% visibility lift.	Continues treating KDD experiment as Tier 1 evidence.	Ma
Gemini	Third-party consensus	Initially says recommendation often depends more on third parties than own site.	Later classifies brand mentions/consensus evidence as observational and warns Reddit	Do so
Gemini	AI bot blocking	Stage answer lumps GPTBot/PerplexityBot/ClaudeBot into candidate access.	posting is not sufficient. Final advice correctly distinguishes GPTBot training	Q
Claude	Entity establishment	Initially: “strongest genuinely new” finding; “load-bearing.”	from OAI-SearchBot retrieval. Challenge says mention/citation correlation has major confounding and does not	Ma do
			establish causation.	

Claude	Third-party mentions	Initially calls them a “citation-acquisition channel.”	Pipeline answer treats source preference as observational and causal explanation as hypothesis.	Do
Claude	Passage structure	Initially strongly recommends passage-level structure.	Maintains mechanism but rejects exact chunk/word formulas as unsupported .	Q re
Claude	Generic schema citation lift	Already cautious in initial answer.	Remains cautious; cites near-zero controlled lift.	St
Claude	llms.txt	Near-zero-cost pilot, not a project.	Remains unproven/low priority. Challenge says correlation exists	St
Perplexity	Schema !’ citation	Initially says schema helps extraction and structured data is foundational .	but adding schema alone does not reliably increase citations .	Ma do
Perplexity	Answer-first exact formulas	Initial gives 40–60-word answers and 130–170-word passages.	Challenge explicitly says no large controlled experiment proves citation lift.	Do
Perplexity	Brand mentions	Initially presents mentions/provenance as citation signals.	Later calls the relationship strong correlation, not isolated causation.	Do
Perplexity	Bot controls	Initially groups GPTBot, OAI-SearchBot, ClaudeBot and Google-Extended together under citation access.	Final 90-day advice separates search/retrieval bots from training bots.	Co
Perplexity	Freshness	Initially presents it as an important citation factor.	Later says relative importance vs traditional ranking is not rigorously quantified .	Q

Perplexity's evidence challenge explicitly downgrades answer-first correlation without causal uplift. Yet its final plan again specifies 40–60-word answers and 200–500-word standalone sections. That is one of the clearest within-system stability problems.				
Claude likewise moves from calling entity establishment and third-party mentions major new levers to emphasizing confounding and uncertainty when challenged.				
7. Retrieval !’ Citation !’ Mention !’ Recommendation				
System	Retrieval	Citation	Mention	Recommendation
Clearly Distinguishes Them?Important Explanation				

			outcome ;	
ChatGPT	Explicit candidate -pool stage; query rewriting/index/access.	Explicit UI/output attribution stage; no citation does not imply no retrieval.	cites dataset showing citation without mention and mention without citation .	S t fi
Gemini	Explicit Stage 1 passage retrieval.	Explicit Stage 3.	Explicit Stage 4.	E
Claude	Explicit candidate pool.	Explicitly says citation is distinct and under-documented .	Explicitly separate from citation and potentially training-memory-driven.	S
Perplexity	Explicit Stage 1.	Explicit Stage 3.	Explicit “ghost citation” stage.	E

ChatGPT explicitly states that a visible citation is not synonymous with retrieval and that the complete candidate/internal set is usually not externally observable. It then uses Semrush-style data to show that citation and mention can diverge in both directions .

Perplexity also later separates the stages and calls out “ghost citations ,” where a domain is cited but the brand is not named.

Which systems collapse the stages ?

None of the four completely collapses them after the explicit follow-up.

However:

Gemini's initial answer moves quickly from retrieval to “being recommended ” via third-party discussion , making the causal chain sound more continuous than its later stage analysis supports .

Claude's initial answer calls third-party mentions an “input to the citation pipeline,” then later treats the mechanism as observational/inferential.

Perplexity's initial answer talks about becoming a source the model chooses to “quote or recommend ,” but its later answer correctly breaks this into five distinct failure points .

ChatGPT is the most conceptually disciplined in the supplied material about not using citation , mention and recommendation as synonyms .

8. Technical Claims Audit				
TacticSystems Recommending ItClaimed				

Outcome	Evidence Provided	Evidence Classification	Safe Conclusion	
				Google Search purposes + correlation studies + Ahrefs matched experiment showing little/no generic citation lift.
Structured data / schema	All	Entity clarity, Search features; some claim easier AI extraction/citation.		DOC for or Sear citati caus estab
FAQ schema	Perplexity; indirectly Gemini	Easier Q&A extraction/citation.	Mostly playbooks/correlation.	UNP
	Claude/Gemini/Perplexity		Google guidance says	CON
llms.txt	permit low-cost experimentation ; ChatGPT deprioritizes .	Machine/agent ingestion or possible visibility .	no Search/AIO benefit; weak/nonexistent live evidence elsewhere.	for G INSU else
robots.txt	All	Determines access to particular crawlers.	First-party crawler docs repeatedly invoked.	VERI DOC
AI crawler access generally	All	Eligibility for non-Google search/retrieval systems.	OAI-SearchBot/PerplexityBot etc.	VERI bot-s
OAI-SearchBot	All eventually	ChatGPT Search inclusion/retrieval.	OpenAI crawler docs .	VERI DOC
GPTBot	Gemini/Perplexity initially blur it; later answers distinguish it	Training, not equivalent to Search inclusion.	Later crawler-function separation.	DOC distin panel
Google-Extended	Perplexity/Gemini initially include it in broad bot audits	Gemini/training/grounding controls rather than Google Search ranking.	ChatGPT explicitly distinguishes it from Googlebot Search controls .	CON as an crawl requi
Entity optimization	Gemini/Perplexity strongly; Claude initially; ChatGPT as data hygiene	Better entity recognition, trust, mentions/recommendations .	Correlations, Knowledge Graph reasoning, platform data hygiene.	INFE incre visibi
Machine-readable information	Gemini/Perplexity; implicit others	Reduce ambiguity and make facts/attributes easier to interpret.	Structured-data/entity reasoning.	DOC as ge clarit INFE citati
Answer-first / structured formatting	All to differing degrees	Better passage use/extraction/citation .	RAG architecture, observational patterns, conditional GEO experiment.	OBS / INF live o retrie
Exact short answer/chunk lengths	Gemini, Perplexity; industry sources	Better vector/chunk retrieval.	No replicated causal evidence; exact numbers differ.	UNP
Long-form	None confidently recommend length alone;		Confounded observational evidence;	

content	Gemini explicitly attacks the “longer is better” myth.	Authority/comprehensiveness.	Google no-ideal-length guidance invoked elsewhere.	UNP facto
Freshness	Gemini, Perplexity; ChatGPT treats as experiment; Claude less prominent	Better selection for time-sensitive queries.	Perplexity platform descriptions + observed recency.	DOC for s retrie conte comp weig INFE

ChatGPT's 90-day recommendation is particularly clear: maintain schema for established Search purposes, add AI crawler governance, but do not create an “AI schema” project.

9. The 90-Day Advice Test

Because Claude's separate final 90-day answer is absent from the supplied PDF, its initial 90-day recommendations are used here rather than reconstructing an unseen response.

Strong cross-system consensus

Keep conventional SEO operating normally. Crawl/index/rendering quality, content quality, authority and intent matching survive scrutiny across all four.

Audit AI crawler access. This is one of the few genuinely additive technical requirements with clear platform-level support. ChatGPT's final advice calls it probably the clearest new cross-platform technical requirement.

Add AI-specific measurement rather than replace SEO measurement. Separate citations, mentions, recommendations and traffic. Repeated sampling is preferable to a one-run “rank.”

Make research fan-out aware without manufacturing one page per guessed subquery. The fan-out mechanism is substantially better supported than the optimization tactics built around it.

Consensus but weak evidence

Answer-first / evidence-extractable formatting. All four lean toward it operationally, but the panel does not establish a universal causal lift in live organic retrieval.

Third-party mentions / earned coverage as an AI lever. Strong observational rationale, weak causal intervention evidence.

Entity work beyond normal data hygiene. Accuracy is defensible; incremental AI recommendation effects are not established.

Freshness cadence. Most defensible for time-sensitive topics and Perplexity-like retrieval; not established as a universal AI advantage.

Platform -specific

OAI-SearchBot: ChatGPT Search.

PerplexityBot / Perplexity retrieval controls: Perplexity.

Googlebot and normal Search controls: Google AI Overviews/AI Mode.

Google-Extended: should not be conflated with Google Search/AIO inclusion.

Merchant Center / Google Business Profile: specifically relevant to Google shopping/local use cases in ChatGPT/Perplexity's panel answers.

Experimental

Controlled content-structure tests.

Adding original statistics/evidence while measuring organic retrieval separately from downstream use.

Schema intervention specifically for AI outcomes.

Third-party coverage intervention.

Recommendation -oriented comparison/product-information experiments.

llms.txt only where a specific non-search agent ecosystem actually consumes it.

ChatGPT explicitly puts evidence presentation, content structure and third-party presence in the experimental bucket rather than doctrine.

Marketing -heavy / inadequately supported

"Schema gives a predictable AI citation lift."

"llms.txt is the new sitemap/ranking factor."

Exact 40–60/130–170/250–500 word or token formulas.

"LLMs love FAQs."

"Add statistics and get a 40% GEO boost."

"Unlinked mentions are the new backlinks."

"Publish one page for every fan-out query."

"Reddit mentions cause recommendations."

"Long-form content is inherently favored."

"Our GEO tool knows the AI ranking factors."

ChatGPT explicitly rejects most of these in its final operating guidelines.

10. Contradictions

Between systems

ContradictionWhat the Panel SaysWhat

Evidence Would Resolve It

Organic top-10 overlap

Gemini starts around **76%**; Perplexity cites **54–76%**; ChatGPT cites **37.1%**; Claude explicitly notes studies ranging from a majority outside top 10 to 75–92% in top 10/top 12.

Same platform, same date, same locale, same query set, same citation definition, same organic SERP snapshot, and explicit handling of fan-out queries.

Generic schema strength

Gemini calls schema essential/Tier-1 entity parsing; ChatGPT, Claude and Perplexity emphasize no demonstrated generic citation lift.

Randomized or matched schema intervention by schema type, platform and query class.

Answer-first structure

Gemini treats it close to an evidence-backed requirement; ChatGPT says organic retrieval benefit remains weak; Perplexity challenge downgrades it; Claude keeps only the broad mechanism.

Controlled live-search experiment where structure alone changes while content, authority and URL remain constant.

Bing dependence for ChatGPT

Gemini says Bing access makes Bing WMT effectively mandatory. Claude says Bing-primary but explicitly flags possible Google-index use as contested. ChatGPT does not make the mandatory-Bing claim.

Current OpenAI first-party architecture documentation or controlled index-presence experiments across Bing-only/Google-only pages.

Google-Extended

Some initial bot lists imply it belongs with citation/retrieval crawlers. ChatGPT explicitly says it does not govern Google Search/AIO ranking.

Direct Google publisher-control documentation, plus crawl/log experiments.

Exact chunk size

Gemini uses 250–500 units; Perplexity gives multiple different ranges; Claude and ChatGPT reject precise formulas.

Replicated controlled tests across platforms and content types.

Freshness

Gemini calls broad recency preference documented; Perplexity eventually narrows that mainly to certain retrieval systems; ChatGPT treats content freshness as a test rather than a general factor.

Time-controlled refresh experiment with matched static controls across multiple engines.

Entity optimization strength	Gemini/Perplexity often imply Knowledge Graph/entity clarity directly influences selection; ChatGPT explicitly refuses deterministic “entity optimization” language.	Staged entity intervention with repeated pre/post measurements and untreated control entities.
Consensus filtering/source-selection rules	Gemini labels majority-consensus filtering “documented”; ChatGPT/Claude say exact source-selection weights are opaque.	First-party technical documentation or auditable source-selection experiment.
Citation display rules	Gemini says common-knowledge suppression/UI citation limits are documented /observed; ChatGPT and Claude describe citation attribution as one of the least transparent stages.	Product-specific documentation or experiments comparing retrieved/used sources with displayed citations.

Within the same system

Claude: Initially calls entity establishment “the strongest genuinely new finding” and third-party mentions a citation-acquisition channel. After challenge, it explicitly says mention correlations are confounded and do not establish causation.

Perplexity: Initially gives concrete answer/passage lengths. Its evidence challenge says there is no large controlled experiment proving these structures cause more citations. Its 90-day answer then reintroduces 40–60-word answers and 200–500-word sections.

Perplexity: Initially groups GPTBot/OAI-SearchBot/ClaudeBot/Google-Extended together as bots to allow for citation visibility. Its final advice separates training and search crawlers.	
Gemini: Its stage analysis treats blocking GPTBot as removing a site from AI candidate pools; its final answer correctly states GPTBot controls training while OAI-SearchBot controls ChatGPT Search retrieval.	
ChatGPT: The original initial answer is unavailable, so exact reversals cannot be audited. Its explicit “regraded” section nevertheless says many earlier recommendations should now be treated as medium-confidence hypotheses or low-confidence formulas.	

11. What's Missing From All Four?

Gap What Is Missing	
Causal organic retrieval evidence	None provides a convincing replicated intervention showing that a specific content/SEO change causes a live page to be retrieved more often across ChatGPT, Google, Gemini and Perplexity.
Retrieval observability	The panel repeatedly acknowledges that a missing citation does not establish that the page was never retrieved. There is no robust external method for observing the complete candidate set.
Fan-out !' citation provenance	Query fan-out is documented, but none can show which hidden/generated subquery caused a particular citation in a live answer.
Source-selection weights	Relevance, trust, freshness, authority, diversity and extractability are discussed, but no system provides a validated weighting model.
Citation attribution rules	Why one used source gets a visible citation while another does not remains largely opaque.
Citation !' brand mention transition	The panel proves observational separation, but not what intervention causes the brand itself to be named.
Mention !' recommendation transition	General commercial/B2B/local-service recommendation criteria remain especially weakly documented.
Local/commercial query evidence	The examples are mostly generic informational/software/product examples. There is little systematic evidence for local agencies, clinics, real estate, restaurants, professional services, etc.
Geography	No controlled comparison of India vs US/UK/other markets, city-level geography, locale, language or local search context.
Industry differences	No credible evidence that the same lever works similarly in SaaS, ecommerce, finance, healthcare, local services, B2B agencies, etc.

Repeated-run volatility by stage	Some answers cite volatility studies, but none supplies a stable model of variance separately for citation, mention and recommendation.
User-context volatility	Personalization is acknowledged, but not measured systematically.
Business outcomes	AI citation/mention !' branded search !' lead !' revenue causality remains largely unresolved.
Schema intervention by type	Generic schema has weak causal evidence; Product, Review, Organization, Person, FAQ etc. have not been isolated sufficiently in this panel.
llms.txt outside Google	Google is relatively clear in the panel; behavior across other search/agent ecosystems remains poorly established.

Entity intervention

Third-party coverage intervention

Freshness intervention

Index-provider dependence

Measurement standardization

No controlled experiment establishes that improving sameAs/Wikidata/entity consistency causes more citations, mentions or recommendations.

The panel has correlation/source-preference evidence, but no clean causal study of earned-media intervention.

Recency is discussed, especially for Perplexity, but updating otherwise equivalent pages has not been causally isolated.

The Bing/Google/native-index question for ChatGPT and other systems is not satisfactorily resolved by the panel.

"AI visibility," "share of voice," "citation rate," "mention rate" and "recommendation rate" lack a shared sampling protocol.

ChatGPT summarizes this gap unusually well: the panel has much better evidence that **the retrieval/synthesis environment differs** than it has evidence for **controllable levers that reliably improve organic retrieval inside that environment**.

12. Information -Gain Opportunities for KickAss

Information GapWhat the Four Systems Currently SayWhy That Is InsufficientWhat KickAss Could Test/MeanurePotential Value				
Same-query rank vs citation	Organic rank correlates with citation, but reported overlap ranges wildly.	Studies use different samples/methodologies and may ignore fan-out.	KickAss could test a fixed prompt panel, record citations, then check every cited URL's rank for both the original query and every observable related/fan-out query.	Very high
Retrieval vs citation	They agree these are distinct but retrieval is mostly invisible.	Citation absence cannot diagnose retrieval failure.	KickAss could test crawler/server-log activity, exposed search queries where available, Bing/Google index presence and visible	Very high
Citation vs mention vs recommendation	Panel says outcomes differ.	No clear transition probabilities by query type.	citation outcomes as separate observations. KickAss could test the same entities across informational, comparative, commercial and recommendation prompts and measure stage-by-stage conversion rates.	Very high

	Fan-out exists; “optimize for fan-out” remains inference.	No evidence that covering more subqueries causes citation lift.	KickAss could test matched topic clusters where one cohort receives substantive missing subtopic coverage and another remains unchanged. KickAss could test matched URLs with structure-only changes while preserving facts,	Very high
Query fan-out coverage	Answer-first formatting is repeatedly recommended but weakly causal.	Current evidence frequently confuses already-retrieved use with retrieval.	URL, author, links and topic coverage. KickAss could test matched cohorts by Organization, Product, Review and FAQ schema	Very high
Content structure	Correlation vs no generic causal uplift.	Schema types/query types may differ.	separately, measuring citation, mention and recommendation rather than one combined score. KickAss could test a legitimate earned-media intervention for one topic/entity	High
Schema	Strong correlation/source-selection patterns.	Brand size/authority confounds the relationship.	cohort against comparable untreated entities/topics. KickAss could test staged entity clean-up—canonical naming, sameAs, profiles, structured data—while repeatedly measuring	Very high, though operationally harder
Third-party mentions	Widely recommended; deterministic claims unsupported.	No causal intervention evidence.	brand mention/entity-confusion errors. KickAss could test server requests to llms.txt, identifiable agent	High
Entity data hygiene	No Google Search benefit; uncertain elsewhere.	Non-Google behavior remains unclear.	access and downstream citation changes before/after deployment, without treating deployment itself as proof. KickAss could test time-sensitive	Medium
llms.txt	Especially emphasized for Perplexity.	No clean update-vs-control evidence.	matched pages with genuine data refreshes versus unchanged controls. KickAss could test identical commercial/local	High
Freshness		AI recommendations		

Local/geographic effects	Barely examined by panel.	may differ by city, country, language and user context.	prompts across Indian metros and controlled account/location states.	Very high
Run-to-run volatility	Variance is acknowledged .	One-shot screenshots still dominate industry reporting.	KickAss could test 20–30 repeated runs per prompt over multiple days and calculate confidence intervals for citation .	Very high
Business outcome	AI visibility metrics are mostly intermediate outcomes .	Citation lift may not produce traffic or leads.	mention and recommendation rates. KickAss could test AI referral conversions, branded-search changes, assisted conversions and lead-quality differences alongside visibility metrics.	Very high
Recommendation - specific evidence	Product shopping has some documented factors; general B2B/local recommendations do not.	Existing GEO tactics mostly target citation , not choice .	KickAss could test “best X”, “X for Y”, “X vs Y” and constrained recommendation prompts separately from informational prompts .	Very high

The highest value opportunities are therefore ~~not more explainers about GEO terminology~~. They are experiments that separate the stages the panel itself says should not be collapsed .

13. Candidate Article Implications

Candidate thesis:		
SEO for AI Search does not replace the SEO foundation; what changes is the discovery, retrieval, source-selection, synthesis and measurement environment above it.		
Thesis ComponentAssessmentWhy “SEO for AI Search does not replace the SEO foundation”	Strongly supported by the four-system panel	All four retain crawl/index/content/authority foundations .
“Discovery environment changes”	Partially supported	Directionally correct, but “discovery” is broad and can include training memory, retrieval and recommendation . More precise terms are preferable.
“Retrieval environment changes”	Strongly supported	Fan-out/query rewriting, engine-specific access and different candidate pools recur across all four.

“Source-selection environment changes”

Strongly supported that the stage exists; contested regarding its controllable factors

The panel agrees retrieval does not automatically imply use/citation . Exact selection weights remain opaque.

“Synthesis environment

Strongly supported architecturally;

Multiple-source generation is central, but

changes”

optimization levers only
partially supported

tactics such as answer-first formatting are not
proven retrieval levers.

“Measurement environment
changes”

Strongly supported

Ranking alone cannot represent citations,
mentions, recommendations, zero-click
exposure or stochastic output.

The candidate thesis therefore survives, but it can be made more precise.

Better thesis emerging from the evidence

AI Search is not a replacement SEO discipline; it adds a multi-stage visibility system above SEO eligibility. The architectural changes—query expansion, engine-specific access, retrieval, source selection, synthesis and separate citation/mention/recommendation outcomes—are much better established than the optimization tactics currently claimed to manipulate them.

That is stronger than simply saying “SEO changes for AI,” because it preserves the most important information gain in the panel:

the architecture is becoming clearer faster than the ranking factors are.

14. Final Cross-System Findings

Strongest Consensus

Traditional SEO remains the eligibility foundation. Crawlability, indexability, useful content, relevance and authority do not disappear.

1. **AI Search introduces additional stages beyond organic ranking.**
2. **Query fan-out/query rewriting is a real architectural change for at least major documented search experiences.**
3. **Engine-specific crawler/access policy is one of the clearest genuinely new technical requirements.**
4. **Retrieval, source selection, citation, brand mention and recommendation should not be measured as one outcome.**
5. **Organic ranking and AI citation are related but non-deterministic.**
6. **Generic schema should not be promised as an AI-citation booster.**
7. **llms.txt is not established as a Google AI Search ranking/citation lever.**
8. **AI visibility needs measurement beyond conventional rankings and traffic, ideally with repeated sampling.**
9. **Most claimed AI-specific editorial formulas have substantially weaker evidence than the architectural changes they are supposedly optimizing for.**

Most Important Disagreements

1. The actual percentage overlap between top organic rankings and AI citations.
2. Whether schema/entity markup has a direct AI source-selection benefit beyond normal Search/data clarity.
3. How strongly answer-first formatting affects live organic retrieval rather than downstream source use.
4. Whether ChatGPT Search depends strongly enough on Bing to make Bing optimization mandatory.
5. How much freshness matters relative to ordinary SEO.
6. Whether entity optimization is a causal visibility lever or primarily data hygiene.
7. Whether third-party mention volume itself causes citation/recommendation lift.
8. Whether source-selection mechanisms such as “consensus filtering” can currently be called documented.
9. Whether exact chunk/passage lengths have practical predictive value.
10. How much of citation attribution is algorithmically knowable from public evidence.

Claims We Should Not Repeat Without Qualification

1. “Schema makes AI systems cite you more.”
2. “llms.txt improves AI Search rankings.”
3. “Write every answer in 40–60 words.”
4. “LLMs prefer 130–170/250–500-word chunks.”
5. “Add statistics and get a 40% GEO visibility boost.”
6. “Unlinked brand mentions are the new backlinks.”

7. "Get more Reddit mentions and ChatGPT will recommend you."
8. "Ranking no longer matters for AI citations."
9. "Create a page for every query fan-out variation."
10. "GEO now has a known set of ranking factors comparable with traditional SEO."

Genuine New Questions Worth Testing

- Which original or fan-out SERP actually supplies each AI citation?
- Does improving fan-out topic coverage causally increase retrieval/citation?
- Which interventions change **citation without changing mention**, and vice versa?
- What changes a brand from merely being mentioned to actually being recommended?
- Does content restructuring improve live retrieval, or only use after retrieval?
- Do schema interventions behave differently by schema type, platform and query class?
- Does legitimate third-party coverage causally increase recommendation frequency?
- How much do city, country, language and account context change brand recommendations?
- How volatile are citations, mentions and recommendations when measured separately over repeated runs?
- Which AI visibility changes produce measurable leads, conversions or branded demand?

Strongest Information -Gain Opportunity

Build a **stage-separated citation-origin dataset** that links ordinary SERP rank, observable fan-out/subqueries, cited URLs, brand mentions and final recommendations for the **same repeated prompt set**. This directly investigates the largest unresolved gap shared by all four systems: *how a page moves from search eligibility into an AI answer, and where it drops out*.

Recommended Next Research Step

Run a **citation-origin / fan-out mapping study before testing any GEO tactic**.

Use one tightly defined query family—ideally this candidate topic, **SEO for AI Search**—and create roughly 20–30 semantically varied prompts covering informational, comparative and commercial intent. Run each prompt repeatedly across Google AI Overviews/AI Mode, ChatGPT, Gemini and Perplexity under a controlled location/account setup.

For every run, record separately:

- the original prompt;
- any observable query rewrites/fan-out/follow-up searches;
- every cited URL and domain;
- whether each cited domain is actually named in the generated answer;
- every uncited brand mention;
- every explicit recommendation;
- the source type—official, academic, publication, vendor, agency, community, brand-owned;
- the page's Google rank for the **original query**;
- where possible, its rank for each **observable fan-out query**;
- repeated-run stability.

Do **not** change schema, page formatting, `llms.txt`, entity data or content yet.

The first empirical question should be:

Are AI citations better explained by the page's rank for the user's literal query, by its visibility across the fan-out/query family, or by neither?

Only after establishing that baseline should the next phase introduce controlled interventions such as content restructuring, schema, entity clean-up or third-party evidence.