

CONTENT STRUCTURE · AI VISIBILITY · AEO · B2B STRATEGY

Why AI systems cite some pages and not others

AI search systems retrieve chunks of content, not whole pages. This guide synthesises what converging research across ChatGPT, Perplexity, and Google AI Mode reveals about the structural signals that predict citation: answer-first section openings, query-formatted headings, factual density, and the limits of schema markup.

AIVIARA RESEARCH · MAY 2026

AI systems don't read your content. They retrieve fragments of it.

That distinction changes everything about how you should write for AI citation. A persuasively argued page that buries its main claim in paragraph four is, from a retrieval system's perspective, nearly identical to a page that has no main claim at all. What gets cited is what gets retrieved. What gets retrieved depends almost entirely on how content is structured at the chunk level.

What gets cited is not primarily determined by Google ranking. Analysis of ChatGPT citation patterns found that the majority of cited pages did not rank in the top 10 for the same query. Semrush's multi-platform study identified systematic structural differences between AI-cited pages and pages that rank well in Google. The two optimisation problems have different structural requirements, and treating them as the same task leaves gaps in both.

This guide synthesises what several independent bodies of research reveal about citation-predictive content structure: how retrieval systems work, what the data shows about headings, position, and factual density, where the sources converge, and where the evidence is thinner than the coverage suggests.

How retrieval systems actually process your content

Before addressing what to write, it helps to understand what the system is doing with your page.

AI search tools that cite web sources (Perplexity, ChatGPT Search, Google AI Mode) use a Retrieval-Augmented Generation pipeline. When a page is indexed for retrieval, it isn't

stored whole. It's broken into segments, each converted into a numerical vector that encodes the segment's semantic meaning. When a query arrives, that query is also converted into a vector and matched against the indexed segments. The closest-matching segments are retrieved and passed to the model's context window, from which the response is generated and sources cited.

The model is selecting among chunks, not pages. This has a direct consequence for how content should be structured: the chunk is the unit of competition. A page that has one excellent, dense, citable section buried inside three weak sections will get the weak sections retrieved as often as the strong one.

In most implementations, heading tags function as natural chunk boundaries. The content under an H2 or H3 is treated as a semantic unit: heading, subheading, and supporting paragraphs grouped together and embedded as one. A vague or misdescriptive heading doesn't just fail to help; it actively mislabels the chunk, reducing its match score against the query it's supposed to answer.

Put the answer first, in every section

A March 2026 analysis by Kevin Indig and Amanda Johnson (Growth Memo, March 30, 2026) found that opening with declarative statements ("X is Y" or "X does Z") produced a +14% aggregate citation lift across all seven verticals tested. It was the only writing-level signal with consistent positive correlation across every vertical, including B2B SaaS, Finance, Healthcare, and HR Tech. That finding converges with Semrush's multi-platform study (ChatGPT Search, Google AI Mode, Perplexity), which found Q&A-formatted; content (where each section opens with a direct question and an immediate answer) correlated with AI citation rates at +25.45% compared to Google-ranking pages. The structural principle appears in both datasets: answer first, context after.

This connects directly to how chunking works. When a section opens with contextual framing before reaching the answer, the first retrieved chunk from that section contains low-density material. When it opens with a definition or direct claim, the first chunk contains the most citable content and sits at the highest positional priority.

Most content teams underweight positional priority. Indig's positional analysis of 18,012 verified ChatGPT citations found 44.2% of citations came from the first 30% of page content, 31.1% from the middle third, and 24.7% from the final third. The pattern was statistically significant and consistent across verticals. One explanation, Indig's, is that LLMs trained heavily on journalism and academic writing have internalised "bottom line up front" structure, causing them to weight early framing more heavily. An alternative reading is

simpler: the first sections of a page tend to be the most topic-dense by construction, and the retrieval system finds its best match there. The two explanations are not mutually exclusive and the data does not distinguish between them.

44.2% of ChatGPT citations came from the first 30% of page content. Only 24.7% came from the final third.

KEVIN INDIG / SEARCH ENGINE LAND · FEBRUARY 2026

The practical consequence for any B2B page: opening a pricing page with "Our pricing is designed to flex with your needs as you grow" before listing actual figures pushes the factual content out of the most-cited zone. Opening with "Enterprise plans start at \$X per seat per month" puts the fact where citation probability is highest.

A concrete before/after:

Weak section opening:

"Content structure has become increasingly important as AI systems have evolved. Understanding what these systems are looking for can help content teams make better decisions about how they present information."

Strong section opening:

"AI systems cite pages with question-formatted headings approximately twice as often as pages with abstract topic headings. The heading sets the retrieval context, and the paragraph immediately after it is treated as the answer."

The second version starts with the claim and the number. The first version is two sentences of context that could precede almost any paragraph on any topic.

Heading structure: the chunk boundary problem

78.4% of citations that included question terms originated from H2 headings, according to Indig's February 2026 analysis. In a RAG pipeline, the H2 heading functions as the retrieval query and the content immediately below it is embedded as the candidate answer. In a well-structured page, that's exactly what it is.

Pages with headings that directly answer the query were cited 41% of the time versus 29% for pages with loosely related headings (Indig, Growth Memo, February 2026). A gap of that magnitude attributable solely to heading specificity is not a marginal improvement.

The practical implication for heading structure: write headings as the question the reader would ask, not as a label for the topic. Compare:

Weak (label)	Strong (query)
"Schema markup"	"Does adding schema markup increase AI citations?"
"Heading structure best practices"	"How do headings affect AI retrieval?"
"Factual content"	"Why do pages with statistics get cited more?"

The label heading creates a chunk with low semantic specificity. The query heading creates a chunk that matches the exact retrieval context of someone asking that question.

One failure mode to watch for: what Indig and Johnson describe as the 3–4 heading range. In their analysis, pages with 3–4 headings consistently underperformed relative to pages with either minimal structure or a more comprehensive heading architecture of five or more. The proposed explanation is that 3–4 headings fragment content without creating enough structure to be navigable. The finding comes from one dataset and the mechanism is inferred rather than tested, but it is worth factoring in when deciding how to structure a new page.

Factual density: what the data actually shows

The Princeton GEO paper (Aggarwal et al., ACM KDD 2024, arXiv:2311.09735) provides the strongest controlled evidence on this question. It tested nine content optimisation methods on 10,000 queries using a Perplexity-like retrieval system. Adding statistics produced approximately 30–40% improvement on the Position-Adjusted Word Count metric (which measures how much cited text appears near the top of a document, weighted by position) and 15–30% on the Subjective Impression metric (a human-rated score of how well an AI response satisfies the query). The paper distinguishes these explicitly and the figures should not be collapsed into a single number. Keyword stuffing showed no improvement and in some conditions a marginal decline. That is a controlled experiment on a retrieval system, not observational citation analysis. It points toward factual specificity as a genuine driver.

Indig's linguistic analysis of ChatGPT cited vs. non-cited passages finds consistent patterns in the same direction: cited text averaged 20.6% proper noun density (named people, companies, tools) compared to 5–8% in standard English prose. Definitive language appeared at 36.2% in cited passages versus 20.2% in non-cited content. This is observational, and the causal direction is not established, but it aligns with what the GEO paper's controlled tests suggest.

The Digital Bloom's 2025 AI Citation / LLM Visibility Report adds a format-level dimension. Comparative listicles captured 32.5% of all AI citations, the highest share of any format. Standard product description pages captured 4.73%. The format difference is the density difference made visible: comparative listicles are named entities, specific attributes, and numerical comparisons. Product pages tend to be qualitative, benefit-focused, and short on specifics. That is exactly the content that scores poorly on every citation predictor the research identifies.

32.5% of all AI citations went to comparative listicles. Standard product description pages captured just 4.73%.

THE DIGITAL BLOOM · 2025 AI CITATION / LLM VISIBILITY REPORT

Pages get cited because their chunks are easily retrievable, not because they argue their case well. High factual density increases how distinctively a chunk matches against a specific query. A passage that contains "Supabase charges [detailed pricing per unit] for storage after the free tier" generates a much sharper vector than "Supabase offers competitive pricing for storage-heavy applications." The second sentence may be accurate; it's not retrievable for any specific storage pricing query.

For B2B SaaS pages specifically: feature pages, comparison pages, and technical documentation tend to be the highest-performing formats for AI citation not because they rank well in Google, but because they're dense with named attributes and specific values. A feature page that explains what the feature does in one sentence and spends three paragraphs on the benefits is factually thin. A feature page that names the specific use case, gives a concrete example, and includes at least one real customer scenario gives a retrieval system something to match against.

Adding schema does not lift citation rates

The schema debate in AI visibility coverage is confused. FAQ schema improves AI citations is a claim that circulates widely. The Ahrefs study provides the most methodologically rigorous data on the question and it does not support it.

Ahrefs tracked 1,885 pages that added JSON-LD schema against 4,000 control pages across a seven-month period (August 2025–March 2026), using four separate analytical methods to test whether schema addition changed citation rates. The result: no statistically significant positive effect on ChatGPT or Google AI Mode citations. Google AI Overviews citations showed a statistically significant decline after schema was added. The study also found that in Ahrefs' tests across ChatGPT, Claude, Perplexity, Gemini, and Google AI Mode, the systems extracted visible HTML content only. JSON-LD, hidden Microdata, and hidden RDFa were all ignored.

The correlation between schema-equipped pages and higher AI citation rates that other analyses have found appears to be a quality proxy, not a causal relationship. Well-maintained, technically sound sites use structured data as part of broader good practice. The same sites tend to have better content structure, clearer headings, more factual density. Schema is correlated with those practices without causing the citation improvement.

The distinction that matters: FAQ-formatted content (actual Q&A; structure in visible HTML) does appear to correlate with higher AI citation rates. Semrush's January 2026 analysis of 11,882 prompts across ChatGPT Search, Google AI Mode, and Perplexity, conducted by Luke Harsel, Roma Chereshnev, and Cecilia Meis, found Q&A; format correlated with AI citations at +25.45% compared to Google-ranking pages. That is content structure, not schema. Adding JSON-LD FAQ schema to a page that doesn't actually use Q&A; content structure adds no apparent incremental value.

If you want FAQ-style citation performance, write FAQ-structured visible content. Adding schema to existing prose does not replicate the effect.

What consistently fails

A few failure patterns appear consistently across the available research. The Semrush multi-platform study (ChatGPT Search, Google AI Mode, Perplexity) and Indig's ChatGPT citation analysis converge on most of these, which increases confidence that the patterns are not platform-specific:

Long introductory framing. Content that spends the first 150 words contextualising the topic before reaching the answer pushes the citable material out of the highest positional priority

zone. Given that 44.2% of citations came from the first 30% of pages, that framing prose is burning valuable real estate.

Vague headings. The 12-point gap between specific and vague headings in citation rates is one of the cleanest structural signals in the data. A heading like "Understanding AI Search" gives a retrieval system almost nothing. "How does ChatGPT decide which sources to cite?" gives it a complete query match.

Promotional language. Semrush's January 2026 analysis found AI-cited pages were less promotional than Google-ranking pages by 26.19%. The directional signal is clear: commercial intent language suppresses citation for informational queries. Price entity mentions suppressed citations in 5 of 6 verticals tested by Indig and Johnson. Finance was the exception, where pricing figures function as substantive reference data.

Dense academic prose. Indig found Grade 16 readability outperformed Grade 19.1 in citation rates. Nominalisations, long compound sentences, and passive constructions reduce the semantic density of a chunk relative to its word count. The content becomes harder to match against a query at the embedding level.

Keyword stuffing. The GEO paper found no improvement from keyword stuffing across both metrics, with a marginal decline in some conditions. It is the one finding that requires no metric disambiguation and holds regardless of how the paper is read.

What this won't solve

Structural optimisation for AI citation has a real ceiling.

The strongest empirical data here comes from ChatGPT citation analysis. Findings around retrieval mechanics, schema extraction, and Q&A; structural effects come from multi-platform studies: the Semrush analysis ran across ChatGPT Search, Google AI Mode, and Perplexity; the Ahrefs schema study tested five systems; the GEO paper used a Perplexity-like retrieval architecture. The specific percentages for positional bias, heading citation rates, and entity density are derived from ChatGPT data and have not been independently replicated on other platforms. The structural principles are consistent with how RAG systems work generally, but the precise figures should not be extrapolated across platforms.

Domain authority and source reputation also function as retrieval filters that content structure cannot override. A page from an authoritative domain will be retrieved and cited before a structurally identical page from an unknown domain, all else equal. Structure

improves your odds within your current domain standing. It doesn't substitute for it.

Citation rate also does not equal conversion rate. The research on structural signals measures citation frequency, not what happens after the citation. A page optimised for AI citation that then confuses or loses the referred visitor has solved the wrong problem.

Applying this to existing pages

One note for teams working with pages that currently rank in organic search: structural changes to existing pages carry SEO risk. Before rewriting H2 headings, changing section intros, or restructuring the order of content on a ranking page, check its current organic traffic and keyword performance. The recommended approach is additive: add a direct-answer opening to an existing section rather than rewriting the section; add an FAQ block at the bottom rather than restructuring the page. The structural signals that improve AI citation (answer-first section intros, specific headings, FAQ blocks) can almost always be added without touching existing content.

Pages you're building from scratch have no such constraint. For new content, answer-first section openings, query-formatted headings, and factual density near the top of the page should be considered as part of the initial content architecture.

Aiviara is building infrastructure for monitoring AI brand citations and factual accuracy across LLM platforms. Early access information is available at aiviara.com.