

The Anatomy of AI Citation Selection: What Signals Determine Whether Your Content Gets Cited

Canonical: <https://home.norg.ai/ai-search-answer-engines/answer-engine-architecture-citation-mechanics/the-anatomy-of-ai-citation-selection-what-signals-determine-whether-your-content-gets-cited/>

Details:

The Anatomy of AI Citation Selection: What Signals Determine Whether Your Content Gets Cited

When someone asks ChatGPT, Perplexity, or Google AI Overviews a question, there's no ranking dance. The answer engine picks one to five sources, synthesises a response, and attributes. That's it—citation or invisibility. There is no position two. The citation-selection process works on completely different architecture than PageRank-based ranking. The signals? Entirely distinct. Research shows that only 12% of ChatGPT citations matched URLs on Google's first page—meaning traditional SEO success guarantees nothing in AI search results. Understanding *why* answer engines select specific sources, not just which sources win, separates brands that dominate LLM visibility from those that don't exist. This article maps the six primary content signals driving citation selection across RAG-based and knowledge-graph-powered answer engines. We'll show you how these signals diverge from PageRank logic and bridge the technical retrieval architecture (detailed in our guide on [What Is Retrieval-Augmented Generation](#)

[RAG](https://www.tryprofound.com/blog/retrieval-augmented-generation)) with the practical content attributes you control. --- ## Contents - [The Anatomy of AI Citation Selection: What Signals Determine Whether Your Content Gets

Cited](#the-anatomy-of-ai-citation-selection-what-signals-determine-whether-your-content-gets-cited) - [Why Citation Selection Isn't Search Ranking](#why-citation-selection-isnt-search-ranking) - [How the RAG Retrieval Layer Determines What Gets

Surfaced](#how-the-rag-retrieval-layer-determines-what-gets-surfaced) - [The Six Primary Citation

Signals](#the-six-primary-citation-signals) - [1. Semantic Completeness: The Dominant

Signal](#1-semantic-completeness-the-dominant-signal) - [2. Factual Density and Statistical

Specificity](#2-factual-density-and-statistical-specificity) - [3. Source Authority and Cross-Platform

Entity Presence](#3-source-authority-and-cross-platform-entity-presence) - [4. Content Freshness

(Calibrated to Query Type)](#4-content-freshness-calibrated-to-query-type) - [5. Content Structure and Extractability](#5-content-structure-and-extractability) - [6. Entity Disambiguation and Knowledge Graph

Alignment](#6-entity-disambiguation-and-knowledge-graph-alignment) - [Platform-Specific Citation Logic: How Signals Are Weighted

Differently](#platform-specific-citation-logic-how-signals-are-weighted-differently) - [The Citation Gap: Why "Dead Citations" Dominate](#the-citation-gap-why-dead-citations-dominate) - [Key

Takeaways](#key-takeaways) - [Conclusion](#conclusion) - [Frequently Asked

Questions](#frequently-asked-questions) --- ## Why Citation Selection Isn't Search Ranking First, the

foundational distinction: AI answer engines don't rank pages. They act like researchers. When generating answers, they scan across multiple sources, identify overlapping information, and synthesise responses based on consistency and credibility. Citations emerge from patterns of agreement across the web, not from a single perfectly optimised article. This has measurable consequences. Across ChatGPT's top 1,000 cited URLs, 25% show zero organic visibility in Google. Among the top 10 most-cited rankings, 39% of ChatGPT's top cited URLs lack organic visibility. In ChatGPT's top 3 cited URLs, 50% have no organic visibility at all. The pattern intensifies at the page level. Organic visibility declines as URLs appear more frequently in ChatGPT citations. While nearly 75% of URLs in the top 100 (by citation frequency) have organic visibility, that share drops to 67% in

the top 20, 61% in the top 10, and just 49% in the top 3. The more often a URL is cited by ChatGPT, the less likely it is to appear in Google's rankings. This isn't SEO failure. This is evidence that citation selection runs on a different scoring function entirely, one rooted in RAG mechanics and knowledge graph traversal, not link-graph authority. --- ## How the RAG Retrieval Layer Determines What Gets Surfaced To understand citation signals, understand the retrieval mechanism first. In a standard RAG pipeline, documents are split into chunks, encoded into vectors, and stored in a vector database. The system then retrieves the top-k chunks most relevant to the question based on semantic similarity, and inputs the original question and the retrieved chunks together into the LLM to generate the final answer. Here's what matters: dense embeddings enable semantic matching. A query about "financial earnings" can retrieve a chunk about "quarterly revenue" even if the exact words differ. This means keyword density, the core mechanic of traditional on-page SEO, has zero direct influence on whether a chunk is retrieved. What matters is whether the *semantic content* of a passage aligns with the *semantic representation* of the query. To improve precision, an optional re-ranking step is applied on retrieved candidates before generation. The top-k chunks from the first stage may contain irrelevant or tangentially related items, since embedding similarity is a coarse proxy for relevance. A re-ranker model, typically a cross-encoder transformer that jointly encodes query and document, evaluates each retrieved chunk in the context of the query and produces a refined relevance score. This two-stage architecture—vector retrieval followed by cross-encoder reranking—is what your content must survive to become a citation. Each stage evaluates different properties of your content. --- ## The Six Primary Citation Signals ### 1. Semantic Completeness: The Dominant Signal Semantic completeness measures whether your content provides a complete, self-contained answer requiring no external context or additional clicks to understand. This is the strongest predictor of AI Overview selection ($r = 0.87$, $p < 0.001$) because AI systems prioritise content they can confidently extract. The mechanism is straightforward: GEO optimises at the fact level. Each statistic, definition, or concept needs standalone clarity. An AI engine might cite one 60-word paragraph from your 3,000-word article, ignoring the rest entirely. Content requiring surrounding context to be meaningful fails at the chunking stage. The chunk retrieved will be semantically incomplete, and the reranker will score it below a passage that stands alone. Traditional keyword-stuffed content fails in RAG environments because semantic search identifies concepts, not keyword density. A page with "generative engine optimisation" mentioned dozens of times but lacking conceptual clarity loses to a page that explains GEO thoroughly with supporting examples and clear structure. **Practical implication:** Structure every H2 section as an independently answerable unit. The section heading should mirror a query. The opening paragraph should deliver the answer. Supporting sentences add evidence and context. Don't assume the reader, or the retrieval system, has read the sections above. ### 2. Factual Density and Statistical Specificity Numbers drive source selection. This isn't aesthetic preference—it's a functional requirement of citation-worthy content. Answer engines are asked factual questions and need sources containing verifiable, discrete facts they can extract and attribute. The Princeton University GEO study (Aggarwal et al., published at ACM SIGKDD 2024) empirically validated this: the Princeton University research paper on GEO found that traditional SEO methods like keyword stuffing perform poorly in generative engine environments, while strategies such as adding citations, including statistics, and using authoritative language can boost visibility by up to 40% in AI responses—a finding validated on both Perplexity.ai and a system modelled on Bing Chat. At the content level, the correlation is measurable. Pages with expert quotes averaged 4.1 citations versus 2.4 for those without. Content with 19 or more statistical data points averaged 5.4 citations, compared to 2.8 for pages with minimal data. **Action:** Aim for one statistic, percentage, or numerical data point every 150–200 words. Start paragraphs with relevant stats—it hooks readers and signals fact-based content to AI. Always cite the source. Every statistic should link to its original source. ### 3. Source Authority and Cross-Platform Entity Presence Brand search volume, not backlinks, is the strongest predictor of AI citations, with a 0.334 correlation. Brand-building activities that seemed disconnected from SEO now directly impact AI visibility. The relationship between traditional authority signals and citation probability is more complex than simple correlation. SE Ranking's analysis of 129,000 domains found that the number of referring domains ranked as the single strongest predictor of citation likelihood. Link diversity showed the clearest correlation with citations. Sites with up to 2,500 referring domains averaged 1.6 to 1.8 citations. Those

with over 350,000 referring domains averaged 8.4 citations. But domain-level authority matters more than page-level metrics. Page Trust mattered less than domain-level signals. Any page with a Page Trust score of 28 or above received roughly the same citation rate (8.3 average), suggesting ChatGPT weighs overall domain authority more heavily than individual page metrics. Cross-platform entity presence amplifies citation probability beyond any single domain signal. Even when citations are not visibly displayed, AI models internally rely on source reinforcement to validate answers. Brands that appear repeatedly across reputable sites, forums, reviews, and editorial content are more likely to be referenced. (For deeper insight into how entity recognition drives citation eligibility, see our guide on [*\[Entity Authority and Knowledge Graph Presence: How to Get Your Brand Recognised by AI Answer Engines\]\(https://www.tryprofound.com/blog/entity-authority-knowledge-graph\)*](https://www.tryprofound.com/blog/entity-authority-knowledge-graph).) ### 4. Content Freshness (Calibrated to Query Type) Freshness is a citation signal, but not a uniform one. Its weight depends heavily on the query nature. ChatGPT's citations tend to reflect content freshness when it's relevant, but not at the expense of authority. When queries included terms like "latest" or specific years, ChatGPT consistently cited recent sources. Pages titled "Best X of 2025" from sites like TechCrunch or TechRadar were frequently selected, showing a clear preference for fresh content when the prompt implied recency. For evergreen queries, the calculus reverses. For queries without a time anchor, ChatGPT often preferred authoritative, evergreen sources. A detailed guide on "how GPT-4 works" from 2023 was cited over newer but less in-depth posts. Academic research and technical documentation, regardless of publication date, also remained common citations for conceptual or factual questions. According to MuckRack analysis, the highest AI citation rates occur within seven days of publication, and more than half of all observed citations reference content published within the past 12 months. This makes content maintenance, not just publication, a citation strategy. Updating existing high-authority pages with current data outperforms publishing new thin content. ### 5. Content Structure and Extractability The structural properties of content determine whether a RAG system can cleanly chunk and retrieve it. AI systems extract fragments, not full articles. Content organised into clear headings, concise definitions, numbered steps, comparison tables, and FAQ sections gives AI models discrete, self-contained passages to work with. SE Ranking's large-scale analysis quantified the structural optimum: structure mattered beyond raw word count. Pages with section lengths of 120 to 180 words between headings performed best, averaging 4.6 citations. Extremely short sections under 50 words averaged 2.7 citations. Counterintuitively, broad, topic-describing URLs outperformed keyword-optimised ones. Pages with low semantic relevance between URL and target keyword averaged 6.4 citations. Those with highest semantic relevance averaged only 2.7 citations. The interpretation: ChatGPT prefers sources that describe a topic comprehensively rather than sources engineered to rank for a single keyword. This is the structural expression of semantic completeness. One commonly recommended tactic underperforms in practice. FAQ schema markup, often promoted as a must-have element for LLM optimisation, has shown surprisingly weak results. Pages with FAQ schema markup have 3.6 citations, compared to 4.2 without. The main insight is that structured data is a nice-to-have, not a game-changer. LLMs seem to care more about whether the information is structured via headings than whether it's technically marked up. Focus on content organisation first, schema markup is icing on the cake. (For a step-by-step guide to implementing these structural principles, see our guide on [*\[How to Structure Content for Maximum AI Citation\]\(https://www.tryprofound.com/blog/structure-content-ai-citation\)*](https://www.tryprofound.com/blog/structure-content-ai-citation).) ### 6. Entity Disambiguation and Knowledge Graph Alignment When a RAG system retrieves a chunk and a reranker scores it, both stages benefit from unambiguous entity references. Link to established entities including people, brands, organisations, industry standards, and authoritative bodies. These links provide context signals that help AI models understand your content's relationship to the broader information ecosystem. This matters because knowledge graphs, the structured entity databases underpinning platforms like Google AI Overviews, require unambiguous entity references to make confident associations. Gemini relies heavily on the Google ecosystem and cross-checks information with signals from the Knowledge Graph. Content that names entities clearly, links to their authoritative sources, and uses consistent terminology across a site creates the disambiguation signals that knowledge graph traversal requires. The community-presence dimension of entity authority is also measurable. Domains with millions of brand mentions on Quora and Reddit have roughly 4x higher chances of being cited than those with

minimal activity. For smaller, less-established websites, engaging on Quora and Reddit offers a way to build authority and earn trust from ChatGPT, similar to what larger domains achieve through backlinks and high traffic. (The relationship between knowledge graph presence and citation eligibility is explored in detail in our guide on [*\[Knowledge Graphs Explained: How Structured Entity Relationships Power AI Answers\]\(https://www.tryprofound.com/blog/knowledge-graphs-explained\)*](https://www.tryprofound.com/blog/knowledge-graphs-explained).) --- ## Platform-Specific Citation Logic: How Signals Are Weighted Differently The six signals above operate across all major answer engines, but their relative weighting varies significantly by platform. | Platform | Primary Citation Logic | Top Signal Weight | |---|---|---| | **ChatGPT** | Bing-indexed RAG + parametric memory | Referring domains (link authority) | | **Perplexity** | Live web retrieval, recency-weighted | Content freshness + community presence | | **Google AI Overviews** | Organic index + Knowledge Graph | Semantic completeness + E-E-A-T | | **Bing Copilot** | Bing index, conservative citation count | Domain authority + index presence | The platforms diverge significantly: ChatGPT relies heavily on Wikipedia and parametric knowledge, Perplexity emphasises real-time Reddit content, and Google AI Overviews favour diversified cross-platform presence. For Google specifically, Google's AI-generated answer boxes now appear for a significant portion of searches. These overviews prioritise content that already ranks well organically, has strong E-E-A-T signals, and uses structured data markup. This makes Google AI Overviews the platform most aligned with traditional SEO signals, though even here, 47% of AI Overview citations now come from pages ranking below position #5, proving that AI Overviews operate on fundamentally different ranking logic than traditional search. Perplexity's citation behaviour reflects a different philosophy entirely. Perplexity's citation patterns skew heavily towards Reddit, with nearly half (46.7%) of top sources coming from the platform, alongside recently published content with a strong preference for fresh articles published within the past 90 days. (For a full comparative breakdown of platform-specific source-selection architectures, see our guide on [*\[How Each Answer Engine Selects Its Sources: ChatGPT, Perplexity, Google AI Overviews, and Bing Copilot Compared\]\(https://www.tryprofound.com/blog/answer-engine-source-selection\)*](https://www.tryprofound.com/blog/answer-engine-source-selection).) --- ## The Citation Gap: Why "Dead Citations" Dominate One structural reality content strategies must confront: roughly 67% of citations from ChatGPT's top 1,000 most-cited pages are what industry experts call "dead citations." These include Wikipedia articles, organisational homepages, and app store listings that brands cannot easily influence through traditional marketing or PR outreach. ChatGPT's citations are dominated by informational sources. Across the top 1,000 URLs, citations span 186 unique domains, yet the distribution is highly concentrated: more than 50% of all cited URLs come from Wikipedia, and 9 of the top 10 domains are general education, news, or media sites. This pattern reflects ChatGPT's reliance on broad, reference-oriented content rather than commercial or category-specific sources. However, data including references to over 38,000 unique domains indicates that ChatGPT drew information from an extensive range of sources. This "fat head, long tail" distribution indicates that, although a handful of sites are frequently cited, ChatGPT also draws answers from a broad, long-tail of niche sources for information. Authority sites command a significant chunk of citations, but there's plenty of opportunity in the tail. The strategic implication: competing for the head is difficult. Competing for the long tail, by producing the most semantically complete, factually dense, and structurally extractable content on specific topics, is where most brands have genuine citation opportunity. --- ## Key Takeaways - Semantic completeness is the strongest single citation signal ($r = 0.87$ for Google AI Overviews), measured by whether a content chunk delivers a self-contained, independently meaningful answer without requiring surrounding context. - Factual density materially increases citation probability: pages with 19+ statistical data points average 5.4 ChatGPT citations versus 2.8 for pages with minimal data, according to SE Ranking's analysis of 216,524 pages. - The SEO-to-citation gap is real and structural: among ChatGPT's top 3 most-cited URLs, 50% have zero organic visibility in Google. Citation selection operates on a fundamentally different scoring function than PageRank. - Freshness is query-conditional: content freshness drives citations for time-sensitive queries, whilst depth and authority dominate for evergreen or conceptual questions. The optimal strategy is to maintain both. - Platform signals diverge significantly: ChatGPT weights domain-level link authority most heavily, Perplexity weights recency and community presence, Google AI Overviews weight semantic completeness and E-E-A-T. Optimising for all four requires a platform-aware content architecture. - Structure matters more than schema: section lengths of 120–180 words between headings outperform

both shorter and longer formats, and content organisation via headings outperforms FAQ schema markup in citation correlation. --- ## Conclusion AI citation selection isn't search ranking with a different label. It's a distinct retrieval and scoring process with its own logic, its own signals, and its own failure modes. The content that wins citations isn't necessarily the content that ranks first. It's the content that a RAG system can cleanly chunk, semantically match to a query, rerank as highly relevant, and that an LLM can confidently extract and attribute without risk of misrepresentation. The six signals covered here—semantic completeness, factual density, source authority, content freshness, structural extractability, and entity disambiguation—are the operational levers available to you. None of them is sufficient alone. A factually dense page with poor structure fails at the chunking stage. A well-structured page with thin facts fails at the reranking stage. A fresh page from an unrecognised entity may not survive the authority filter. Understanding how these signals interact with the underlying RAG and knowledge graph architectures (covered in our guides on [*\[What Is Retrieval-Augmented Generation \(RAG\)?\]\(https://www.tryprofound.com/blog/retrieval-augmented-generation\)*](https://www.tryprofound.com/blog/retrieval-augmented-generation) and [*\[Knowledge Graphs Explained\]\(https://www.tryprofound.com/blog/knowledge-graphs-explained\)*](https://www.tryprofound.com/blog/knowledge-graphs-explained)) is the foundation for any serious content strategy targeting answer engine visibility. The practical implementation of these signals, including answer capsule formatting, heading architecture, and schema markup, is detailed in our companion guide [*\[How to Structure Content for Maximum AI Citation\]\(https://www.tryprofound.com/blog/structure-content-ai-citation\)*](https://www.tryprofound.com/blog/structure-content-ai-citation). Ship fast. Learn faster. Become the answer. --- ## References - Aggarwal, Pranjal, Vishvak Murahari, Tanmay Rajpurohit, Ashwin Kalyan, Karthik Narasimhan, and Ameet Deshpande. "GEO: Generative Engine Optimization." *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD '24)*, Association for Computing Machinery, 2024, pp. 5–16. <https://doi.org/10.1145/3637528.3671900> - SE Ranking. "Ranking Factors for ChatGPT: Analysis of 129,000 Domains and 216,524 Pages Across 20 Niches." *SE Ranking Blog**, December 2025. <https://seranking.com/blog/ranking-factors-for-chatgpt/> - seoClarity / Mitul Gandhi. "Analysis of Top-Cited Pages from ChatGPT." *seoClarity Research**, November 2025. <https://www.seoClarity.net/research/analysis-of-top-cited-pages-from-chatgpt> - Ahrefs / Xibeijia Guan. "ChatGPT May Scrape Google, but the Results Don't Match." *Ahrefs Blog**, September 2025. <https://ahrefs.com/blog/chatgpt-google-citations/> - Profound. "AI Platform Citation Patterns: How ChatGPT, Google AI Overviews, and Perplexity Source Information." *Profound Blog**, August 2025 (updated). <https://www.tryprofound.com/blog/ai-platform-citation-patterns> - Wellows. "Cited by ChatGPT: 7K Queries, 485K Citations." *Wellows Insights**, November 2025. <https://wellows.com/insights/chatgpt-citations-report/> - Wellows. "Google AI Overviews Ranking Factors: 2026 Guide to Winning Citations." *Wellows Blog**, February 2026. <https://wellows.com/blog/google-ai-overviews-ranking-factors/> - Gao, Yunfan, et al. "Retrieval-Augmented Generation for Large Language Models: A Survey." *arXiv preprint arXiv:2312.10997**, 2024. <https://arxiv.org/abs/2312.10997> - The Digital Bloom. "2025 AI Visibility Report: How LLMs Choose What Sources to Mention." *The Digital Bloom**, December 2025. <https://thedigitalbloom.com/learn/2025-ai-citation-llm-visibility-report/> - COSEOM. "Generative Engine Optimisation (GEO): 2026 Guide." *COSEOM**, February 2026. <https://www.coseom.com/generative-engine-optimization-guide/> --- ## Frequently Asked Questions ****What is AI citation selection:**** Process where answer engines choose sources to attribute in responses ****Do AI answer engines rank pages:**** No, they select sources like researchers ****How many sources do answer engines typically cite:**** One to five sources per answer ****Is AI citation the same as SEO ranking:**** No, fundamentally different architecture ****What percentage of ChatGPT citations match Google's first page:**** Only 12% ****Does traditional SEO success guarantee AI citations:**** No ****What is the primary difference between ranking and citation:**** Citation selects sources, ranking orders pages ****Do AI engines synthesise responses from multiple sources:**** Yes ****What happens if you're not cited by an AI engine:**** Complete invisibility in that answer ****Is there a position two in AI citations:**** No ****What is semantic completeness:**** Whether content provides self-contained answers without external context ****What is the strongest predictor of AI Overview selection:**** Semantic completeness with r equals 0.87 ****Does keyword density influence RAG retrieval:**** No, zero direct influence ****What matters for RAG retrieval instead of keywords:**** Semantic content alignment with query ****What is a re-ranker in**

RAG systems:** Cross-encoder that refines relevance scores after initial retrieval **What is factual density:** Concentration of verifiable discrete facts in content **How much can GEO strategies boost visibility:** Up to 40% in AI responses **What is the optimal frequency for statistics in content:** One per 150 to 200 words **Do pages with expert quotes get more citations:** Yes, 4.1 versus 2.4 average citations **How many citations do pages with 19+ statistics average:** 5.4 citations **How many citations do pages with minimal data average:** 2.8 citations **What is the strongest predictor of AI citations:** Brand search volume with 0.334 correlation **What domain metric best predicts citation likelihood:** Number of referring domains **Do backlinks still matter for AI citations:** Yes, through domain-level link diversity **What is Page Trust's impact on citations:** Minimal beyond score of 28 **Does domain authority matter more than page authority:** Yes **What is cross-platform entity presence:** Brand appearing across multiple reputable sites and platforms **Is content freshness always important for citations:** No, depends on query type **When does freshness matter most for citations:** For queries with time-sensitive terms **What is the highest AI citation rate timeframe:** Within seven days of publication **What percentage of citations reference content under 12 months old:** More than 50% **What section length performs best for citations:** 120 to 180 words between headings **How many citations do extremely short sections under 50 words average:** 2.7 citations **Do keyword-optimised URLs outperform topic-describing URLs:** No, topic URLs get 6.4 versus 2.7 citations **Does FAQ schema markup improve citation rates:** No, 3.6 with versus 4.2 without **What matters more than schema markup:** Content organisation via headings **What is entity disambiguation:** Clear, unambiguous references to established entities **Which platform relies on Google Knowledge Graph:** Google AI Overviews **What is ChatGPT's primary citation logic:** Bing-indexed RAG plus parametric memory **What is Perplexity's top citation signal:** Content freshness plus community presence **What percentage of AI Overview citations come from below position 5:** 47% **What percentage of Perplexity's top sources come from Reddit:** 46.7% **What are dead citations:** Wikipedia, organisational homepages, and app store listings **What percentage of ChatGPT's top citations are dead citations:** Roughly 67% **How many unique domains are in ChatGPT's top 1000 citations:** 186 unique domains **What percentage of ChatGPT citations come from Wikipedia:** More than 50% **How many total unique domains has ChatGPT referenced:** Over 38,000 **Is there opportunity in long-tail citations:** Yes **What is the citation gap:** Difference between SEO ranking and AI citation selection **Do pages with zero Google visibility get ChatGPT citations:** Yes, 50% of top 3 cited URLs **What is the optimal content maintenance strategy:** Update existing high-authority pages with current data **Should you focus on publishing new content or updating existing:** Updating existing outperforms new thin content **What is RAG:** Retrieval-Augmented Generation **How are documents processed in RAG:** Split into chunks and encoded into vectors **What enables semantic matching in RAG:** Dense embeddings **Can semantic search match different words with same meaning:** Yes **What is the two-stage RAG architecture:** Vector retrieval followed by cross-encoder reranking **Does AI extract full articles or fragments:** Fragments only **What content structure helps AI extraction:** Clear headings, concise definitions, numbered steps, tables **Should each H2 section be independently answerable:** Yes **Where should statistics appear in paragraphs:** At the start **Should every statistic link to its source:** Yes **What is the relationship between brand mentions and citations:** 4x higher citation chances with millions of mentions **Which platforms build authority through community presence:** Quora and Reddit **Does Google AI Overviews align with traditional SEO:** Most aligned, but still fundamentally different **What is the operational foundation for citation strategy:** Understanding RAG and knowledge graph architectures **How many primary citation signals exist:** Six **Are any single signals sufficient alone:** No **What happens to factually dense pages with poor structure:** Fail at chunking stage **What happens to well-structured pages with thin facts:** Fail at reranking stage

Source Data (JSON):

```
"{\n  \"_type\": \"article\",\n  \"title\": \"The Anatomy of AI Citation Selection: What Signals Determine Wh
```

