

Where AI Engines Actually Get Their Citations & Why It Breaks the "SEO Plus Trust" Argument

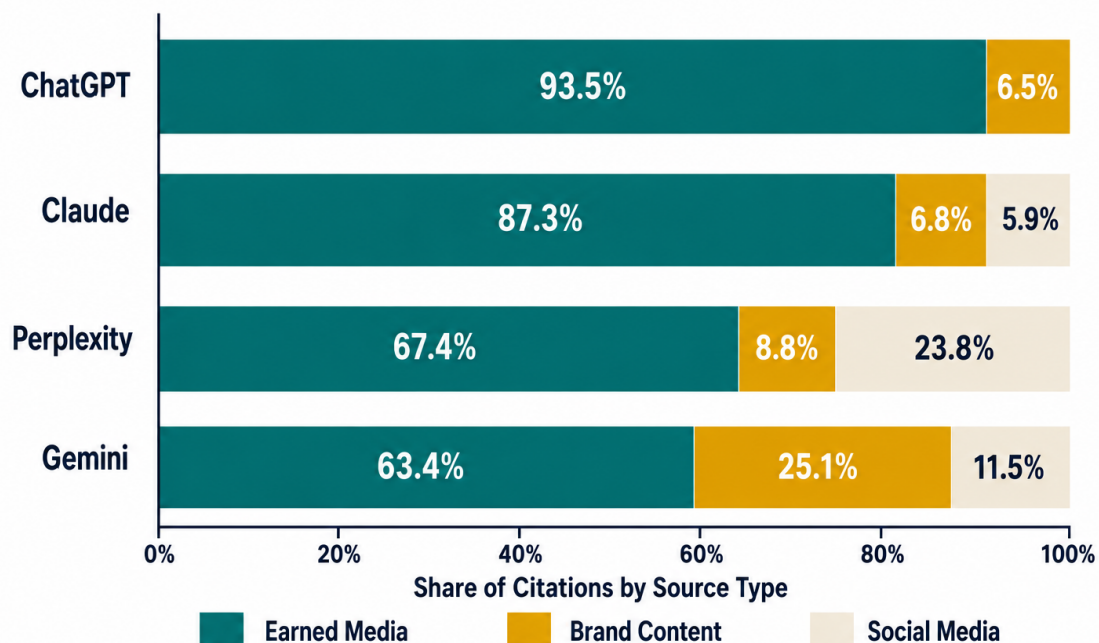
There is a comfortable story doing the rounds in search marketing circles at the moment. It goes something like this: keep doing solid SEO, build up some trust signals, and AI search will look after itself. Citations will follow naturally, the way rankings used to.

It is a tidy story. It is also wrong, and the data on how AI engines actually source their citations shows exactly why.

What the research shows

A University of Toronto study (Chen et al., 2025) tracked how four major AI engines, ChatGPT, Claude, Perplexity and Gemini, source the citations they generate when discussing well-known brands. The researchers classified every citation into one of three buckets: earned media, brand content, or social media.

How Each AI Engine Sources Its Citations (Well-Known Brands)



The pattern is stark. ChatGPT pulls 93.5% of its citations from earned media, third-party editorial coverage, trade press, independent commentary, and just 6.5% from content the brand controls directly. Claude follows a similar shape: 87.3% earned media against 6.8% brand content, with a small 5.9% slice from social media. Perplexity shifts the balance further, 67.4% earned media, 8.8% brand content, and a notable 23.8% from social sources. Gemini is the outlier, drawing 63.4% from earned media but a much larger 25.1% from brand content, alongside 11.5% social media.

Read across all four engines, one thing holds constant: earned media is doing the majority of the work everywhere, and in ChatGPT's case, very nearly all of it.

This is the number that demolishes the "SEO plus trust signals" argument. Trust, in that argument, is usually shorthand for things the business itself puts on its own website: testimonials, credentials, about pages, case studies. All of this lives in the brand content column, the smallest slice of the pie on every single engine measured. The column doing the heavy lifting, earned media, is largely outside a business's direct control and requires an entirely different kind of work to build.

What this means for the OPTIMUM audit

Each of these three source categories maps onto specific dimensions inside the OPTIMUM Seven Dimension AI Readiness Audit. Below is how they align, and what each mapping means in practice for a client's Now Score and Future Score.

Source Type	OPTIMUM Dimension(s)	Audit Implication
Earned Media (63.4%–93.5% share)	Trust and Credibility Signals; Content Quality and E-E-A-T	External, third-party validation is the single largest citation driver across every engine. An audit that finds weak or absent trade press, review coverage, or independent commentary should flag this as a primary invisibility risk, not a secondary one. PR and digital PR activity sits inside the audit scope, not outside it.
Earned Media (secondary read)	Topical Authority	When third parties write about a brand in connection with a specific topic, that is external corroboration of subject-matter relevance. A gap here suggests the brand hasn't been discussed by anyone else in its category, which weakens topical standing regardless of how strong the owned content is.
Brand Content (6.5%–25.1% share)	Entity and Brand Clarity	Owned pages are where entity definition lives: consistent naming, clear service or product descriptions, disambiguation from similarly named competitors. A weak score here shows up as an engine misattributing or under-citing the brand's own material.

Source Type	OPTIMUM Dimension(s)	Audit Implication
Brand Content (secondary read)	Technical Foundation	Since brand content is entirely within the client's control, it is the layer most exposed by poor crawlability, thin markup, or slow indexing. Platforms with heavier reliance on brand content (Gemini at 25.1%, versus ChatGPT at 6.5%) make Tier 1 technical hygiene disproportionately important for visibility on that platform specifically.
Social Media (0%–23.8% share)	Local and Contextual Relevance	Community forums, local groups, and platform discussion threads are where hyper-local and contextual mentions accumulate. Relevant mainly for businesses with a genuine local footprint, estate agency, automotive retail, and largely a non-factor for platforms like ChatGPT that show no measurable social share.
Social Media (secondary read)	Trust and Credibility Signals	User-generated sentiment and social proof contribute a minor, platform-dependent trust signal. Perplexity's 23.8% share makes this worth flagging as a platform-specific behaviour in client reporting, rather than treating social presence as universally decisive.
<i>(No direct source category)</i>	Structured Data and Machine Readability	None of the three tracked source types map cleanly onto this dimension; it is infrastructure sitting beneath all of them. Schema and machine-readable markup remain a separate, foundational audit layer, addressed below.

Two things stand out immediately. First, five of the seven OPTIMUM dimensions map directly onto measurable citation sourcing behaviour. This is not framework marketing. It is a reflection of how these systems actually retrieve and attribute information. Second, the dimension that does not map onto any single source category, Structured Data and Machine Readability, is arguably the most important one to understand, because it explains why strong sourcing alone is not enough.

A note on Structured Data and Machine Readability

The chart's three source categories, earned media, brand content and social media, describe where an AI engine pulls its citation content from. Structured Data and Machine Readability is a different kind of thing entirely. It is not a source. It is the infrastructure that determines whether any of those three sources can be read, parsed and correctly attributed in the first place.

This distinction matters more than it might first appear. A business could have strong earned media coverage, a well-built site full of brand content and an active social presence, and still under-perform on AI citation because none of it carries the schema markup,

structured metadata or machine-readable formatting that lets an AI engine parse it with confidence. Earned media coverage without an Organisation or Person schema tying it back to the correct entity is a weaker signal than it should be. Brand content without Product, Service or FAQ markup is harder for a retrieval system to extract cleanly, even if the copy itself is excellent.

Put simply, the three source types in this chart answer the question "what is being cited?" Structured Data and Machine Readability answers a prior question: "can this be read correctly at all?" One sits above the other. A business can win on source mix and still lose on retrieval if the underlying markup is absent, inconsistent or malformed.

This is precisely why Structured Data and Machine Readability sits as its own dimension within the OPTIMUM audit rather than being folded into Trust and Credibility Signals or Technical Foundation. It is measured separately because it fails separately. A client can score well across earned media, brand content and social visibility, and still find their OPTIMUM Now Score dragged down by a missing schema layer that no amount of PR or content work will fix.

The point this proves

None of this is an argument against SEO or against building trust signals. Both still matter. But the "SEO plus trust is sufficient" position rests on an assumption that AI citation behaves like the old SERP, where a business could largely control its own visibility through its own website. The sourcing data says otherwise. The majority of citation weight sits with third parties the business does not control, and the layer that determines whether any of it gets read correctly sits in technical infrastructure that traditional SEO and trust-signal thinking rarely touches at all.

That is Tier 1 and Tier 2 in miniature. Tier 1 is the gate, technical foundation and machine readability that determine whether a business is legible to these systems at all. Tier 2 is the intent-driven architecture built above it, where earned media, topical authority and contextual relevance do their work. Skipping straight to trust signals without addressing either is why so many otherwise reputable businesses remain invisible in AI search, not because they lack credibility, but because the systems reading them cannot parse what credibility looks like on their site.

Source: Chen et al., University of Toronto, 2025 (arXiv:2509.08919). Graph and context originally published by Ann Coyle, Coyle Emerald Narratives Boston, Mass. July 2026