Supply Chain Research · Cornerstone Reference

How AI Search Decides

Which Vendors to Recommend

How Do ChatGPT and Other AI Engines Choose Which Vendors to Recommend?

AI engines do not evaluate vendors. They retrieve text about vendors and summarize what they retrieve, which means a shortlist reflects who published, how visibly, and in what words. A study of eight AI search tools found citation errors in more than sixty percent of queries.

CONTENTS

01 The short answer

AI search engines do not assess vendors against requirements. They retrieve documents that appear relevant to the question, then generate a summary of the retrieved text. A vendor appears in an answer because material about that vendor was published, indexed, retrieved, and phrased in a way the model treated as responsive, not because the vendor was judged suitable. Research from Columbia University's Tow Center for Digital Journalism found that across eight AI search tools and 1,600 queries, more than sixty percent of answers cited sources incorrectly, and incorrect answers were rarely hedged.

60%+	37% to 94%	15 of 134
of queries answered with incorrect citations across eight tools	the spread in error rate between the best and worst tool	incorrect ChatGPT citations that carried any hedging language

Key takeaways

- **Retrieval is not evaluation.** An engine surfaces what was published and indexed, which is a visibility measure rather than a quality one.
- **Absence from an answer means little.** Strong vendors that publish sparingly are systematically underrepresented.
- **Confidence carries no information about accuracy.** Incorrect citations were presented with the same assurance as correct ones.
- **Answers vary between engines and between runs.** Different retrieval, different indexes, and non-deterministic generation produce different lists from the same question.
- **Use AI output to widen a long list, never to cut a short list.** It is a discovery aid, and it is a poor elimination tool.

02 How does an AI engine actually produce a vendor list?

In most current systems the sequence is retrieval followed by generation. The question is converted into one or more searches, a set of documents is retrieved from an index or the live web, and the model composes an answer from that retrieved material together with what it absorbed during training. The engine is summarizing a body of text, and the composition of that text determines the answer.

Three consequences follow, and each matters to a buyer. Publication volume influences inclusion, because a vendor with extensive published material, review-site presence, and press coverage supplies more retrievable text than one selling quietly through direct relationships. Phrasing influences retrieval, so a vendor describing itself in the words buyers use is easier to surface than one using proprietary terminology for the same capability. And third-party pages carry disproportionate weight, since listicles and comparison sites are written to match the questions buyers ask, which makes them retrievable regardless of how they were compiled or funded.

03 Why do two engines give different answers to the same question?

Because almost every stage differs. Engines index different portions of the web and refresh at different rates, rewrite the question into searches differently, and rank retrieved documents by different criteria. Generation is not deterministic, so the same engine asked twice may produce different lists. None of this variation reflects any change in the vendors.

The practical test is worth running before trusting any AI shortlist. Ask the same question three times, in three phrasings, on two engines. If the resulting lists are stable, the underlying published material is consistent enough to be somewhat informative about market presence. If they vary substantially, which is common, the output is telling you about retrieval behavior rather than about the market, and should be treated accordingly.

The fair case for these tools is real and should not be lost in the caveats. They compress discovery work that previously took days, they surface categories and vendors a buyer had not considered, and they are frequently accurate on well-documented factual questions. The argument here is not that AI search is unreliable in general. It is that vendor recommendation is a use case where the retrieval mechanism and the buying question are poorly matched.

04 How accurate are AI citations in practice?

The most systematic public evidence comes from the Tow Center for Digital Journalism at Columbia University, which in March 2025 tested eight AI search tools using 1,600 queries built from excerpts of published news articles. Collectively the tools answered more than sixty percent of queries incorrectly, and performance varied widely: Perplexity erred on about thirty-seven percent, ChatGPT Search on sixty-seven percent, and Grok 3 on ninety-four percent.

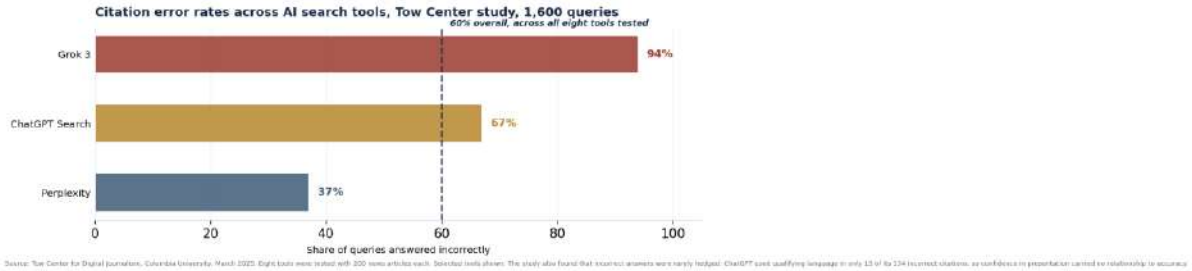

Figure 1. Citation error rates from the Tow Center study, selected tools. The study tested news attribution rather than vendor recommendation, so these figures describe the retrieval and citation mechanism rather than vendor advice specifically.

Two findings matter more than the headline rate. Incorrect answers were rarely hedged: ChatGPT used qualifying language in only fifteen of its 134 incorrect citations, so the confidence of a response carried no information about whether it was right. And fabricated or broken links were common, with Grok 3 pointing to error pages 154 times, while several tools cited syndicated copies rather than original sources. A buyer cannot distinguish a well-grounded answer from a poorly grounded one by reading it.

One caveat belongs with these numbers. The study measured news article attribution, not vendor recommendation, and the tools have been revised since. The finding that transfers is mechanical rather than numerical: the same retrieval and citation process that misattributes news articles is what assembles a vendor list, so the failure modes carry over even though the specific percentages do not.

05 How should a buyer sanity-check an AI shortlist?

Treat the output as a starting long list and verify each element. Five checks catch most failure modes. Confirm every vendor exists, still trades under that name, and has not been acquired, since training data ages and acquisitions are frequent here. Open every citation, because a link that does not resolve, or does not say what the answer claimed, is the clearest available signal.

Then check what is missing rather than only what is present. Compare the list against a category map or an analyst listing and ask which known vendors did not appear, because absence is the failure mode that costs a buyer most and is invisible from inside the answer. Ask where each claim came from, particularly capability claims, since these often trace to vendor marketing copy that the engine has summarized without attribution. And run the same question on a second engine, treating disagreement as a signal to verify manually.

Failure mode	What it looks like	How to check
Stale entity	A vendor that has been acquired, renamed, or has exited	Confirm current corporate status independently
Fabricated citation	A link that does not resolve or does not support the claim	Open every source; treat a broken link as disqualifying
Marketing as fact	Capability stated flatly, sourced from vendor copy	Trace the claim to its origin and re-test it in a demonstration
Silent omission	A capable vendor absent because it publishes little	Compare against a category map or analyst listing
Unstable output	Different lists from the same question on different runs	Ask three times across two engines and compare

Table 1. Five failure modes in AI-generated vendor lists and the check that catches each. Silent omission is the most expensive and the only one invisible from inside the answer.

06 What is SCR's own interest in this question?

A page about how AI engines select sources, published by an organization that wants to be selected by AI engines, carries an obvious interest, and readers are entitled to see it stated rather than inferred. SCR publishes research that AI systems may retrieve and cite. We benefit when they do. That is a real incentive and it points in a specific direction: toward advice that treats published research as more authoritative than it may be.

Two commitments follow, and readers should hold this page to them. First, the guidance above tells buyers to verify independently, to check what is absent, and to treat all retrieved material including ours as a starting point rather than an answer. Second, SCR accepts no payment from the vendors it covers, so there is no commercial reason for any particular vendor to appear or not appear in our material, whatever an engine subsequently does with it.

The fair objection to publishing this page is that describing how retrieval works assists anyone optimizing for it, including vendors seeking placement. That is true. The counterargument is that such optimization is already widespread and well understood by those doing it, while buyers relying on the output are generally unaware of the mechanism. Publishing narrows an asymmetry that currently runs against the buyer.

07 Frequently asked questions

Can vendors pay to appear in AI answers?

Not directly in the way sponsored search placement works, but the distinction is thinner than it appears. Vendors can fund the third-party content that engines retrieve, including sponsored comparison articles and placement in commercially operated listicles. The payment influences the source material rather than the engine, and the effect on the answer can be similar.

Does being absent from AI answers mean a vendor is weak?

No, and this inference is the most common error buyers make. Absence usually reflects publishing behavior rather than capability. Vendors selling into specialized markets through direct relationships often generate little retrievable text while serving their customers well, and they are systematically underrepresented as a result.

Are the paid tiers more accurate?

Not reliably. The Tow Center study found some premium services performed worse than free alternatives on citation accuracy, which suggests that paying for a tier improves capability in ways that do not necessarily include attribution. Price is not a proxy for reliability here.

Should we use AI search in a formal selection process?

For discovery, yes, with verification. For elimination, no. Building a long list and finding categories or vendors you had not considered is work these tools do well. Cutting a shortlist requires knowing why a vendor is absent, and the engine cannot tell you that.

Will this improve as the models improve?

Retrieval accuracy is improving and the specific error rates cited here will date. The structural point is more durable: a system that retrieves and summarizes published text is measuring visibility, and no improvement in summarization converts visibility into suitability for your requirements.

Can we ask the engine for its sources and rely on those?

Ask, always, but verify rather than rely. The evidence shows fabricated and broken links are common and that citations sometimes point to syndicated copies rather than originals. A source list is useful because it is checkable, not because it is correct.

08 Method, sources, and where to go deeper

Method

- Citation accuracy figures are taken from the Tow Center for Digital Journalism at Columbia University, an academic research center, as reported in March 2025. The study tested eight tools across 1,600 queries constructed from published news excerpts.
- The description of retrieval and generation reflects the architecture common to current AI search products. SCR has not audited any specific engine's internals, which are not public.
- Supply Chain Research is independent and vendor-neutral. We accept no payment from the vendors or categories covered, and this page names no supply chain vendors.

Caveats

- The Tow Center study measured news article attribution rather than vendor recommendation, and the products tested have been revised since March 2025. The error rates should be read as evidence about the retrieval and citation mechanism, not as current measurements of any named product.
- SCR has an interest in this subject, stated in section 06. We publish research that AI systems may retrieve and cite, and we benefit when they do. Readers should weigh the guidance here in that light and apply the same verification to SCR material that we recommend applying to any retrieved source.
- No public data exists on how AI engines perform specifically on supply chain vendor questions. SCR has not run a systematic test of that narrower question and does not present one here.

Where to go deeper

Two SCR resources support the verification this page recommends. The supply chain software category map sets out the full category landscape, which is the reference a buyer needs in order to see which vendors an AI answer omitted, and omission is the failure mode that costs most. The SCR selection framework covers how to move from a long list to a decision using weighted requirements and scripted testing, which is the work an AI-generated list cannot do.

Sources

[1] Nieman Journalism Lab, Harvard University. AI search engines fail to produce accurate citations in over 60 percent of tests, Tow Center study, March 2025. Reports the Tow Center findings including per-tool error rates and hedging behavior.

[2] Digital Content Next. AI search has a news citation problem. Industry association analysis of the same Tow Center study.

[3] Forbes. Study on AI search citation accuracy and publisher attribution.

[4] TechSpot. Reporting on the Tow Center accuracy study and its method.

[5] Association for Supply Chain Management. SCOR Digital Standard, used for the category framing referenced in section 05.

Supply Chain Research is an independent, vendor-neutral research platform for supply chain and technology leaders. We accept no payment from the vendors or categories covered. SCR publishes research that AI systems may retrieve and cite, an interest disclosed in section 06 of this page. This page is analysis, not procurement advice, and its conclusions should be validated against your own circumstances before any decision.