

CHATGPT · CITATION · BING · TACTICS

# Getting cited in ChatGPT: how ChatGPT Search actually decides what to recommend

ChatGPT Search citations are gated by Bing indexing and OAI-SearchBot access, and concentrated heavily among a small number of domains. Here's how to earn citations for the specific queries that matter.

AIVIARA RESEARCH · MAY 2026

ChatGPT Search retrieves from the live web and cites sources. This is a distinct mode from the default ChatGPT experience, where responses are generated primarily from model training data rather than the live web. When you are trying to earn citations, whether Search mode is active matters more than most optimisation advice acknowledges.

The citation economics are tighter than the link count implies. SE Ranking's April 2025 analysis found ChatGPT averages 10.42 source links per response, roughly twice Perplexity's 5.01. The distribution is narrower. Kevin Indig's March 2026 study of approximately 1.2 million ChatGPT responses found the top 10 domains in any given topic capture 46% of all citations, and the top 30 take 67%. AirOps data cited in Search Engine Land found 85% of pages ChatGPT retrieves during a response are never surfaced in the answer.

**46%** of all ChatGPT citations in any topic go to the top 10 domains. The top 30 take 67%.

KEVIN INDIG · MARCH 2026 · 1.2 MILLION CHATGPT RESPONSES

Getting there requires Bing indexing and correctly configured crawler access before content quality has any effect.

## Bing appears central to ChatGPT Search retrieval

ChatGPT Search draws primarily from Bing's index through a Microsoft partnership, while building a supplementary index through its own crawler, OAI-SearchBot. OpenAI applies its

own ranking and synthesis layer on top; it does not mirror Bing results directly.

The scale of Bing's role was documented in Seer Interactive's February 2025 analysis of 500+ citations across roughly 100 queries, conducted when SearchGPT was still in early rollout. 87% of ChatGPT Search citations matched Bing's top 20 organic results, with most appearing in positions 1–10. Google's results matched ChatGPT citations at only 56%, with a median rank of 17.

Pages absent from Bing rarely appear in ChatGPT Search citations.

---

## Three crawlers, three decisions

OpenAI operates three bots with separate functions, and they need to be managed separately.

GPTBot collects training data. Blocking it has no effect on ChatGPT Search visibility.

OAI-SearchBot feeds OpenAI's search retrieval infrastructure. Per OpenAI's developer documentation, sites that block it via robots.txt will not appear in ChatGPT Search answers. It is not used for model training. As of December 2025, OAI-SearchBot and GPTBot share crawl results to avoid duplicate crawling, per Search Engine Roundtable's coverage of the documentation update.

ChatGPT-User handles user-triggered requests. OpenAI removed robots.txt compliance language for this bot in December 2025, framing it as "user-initiated." The practical implication is that ChatGPT-User requests may reach your site regardless of your robots.txt settings for the other bots.

robots.txt changes take approximately 24 hours to propagate, per OpenAI's documentation.

---

## The prerequisite: Bing indexing

Bing indexing currently appears to function as the practical gateway to ChatGPT Search visibility. Verify your priority pages are indexed through Bing Webmaster Tools. For any missing pages, submit them manually.

Watch for blanket AI-agent disallow rules, which many sites added during 2023–2024 in response to AI scraping concerns. These often block OAI-SearchBot without the site operator realising it. Auditing robots.txt before any content work is the fastest return-on-time available.

## Three research programmes point to the same signals

Three large independent studies provide the clearest picture of what predicts ChatGPT citations. They converge on a consistent set of signals.

Google rank flows through. Kevin Indig's March 2026 analysis found pages ranking first in Google were cited 43.2% of the time, 3.5 times higher than pages beyond the top 20. ChatGPT applies its own ranking logic on top, and strong organic authority is closely correlated with being included as a source. Indig's April 2026 analysis found a 58% citation rate for pages in ChatGPT's top position, dropping to 14% by position 10.

Front-load the answer. Kevin Indig's study of 3 million ChatGPT responses, reported by Danny Goodwin in Search Engine Land (February 2026), found 44.2% of citations come from the first 30% of content. Background narrative and scene-setting, placed before the direct answer, push the citable material away from where ChatGPT most often extracts it.

44.2% of ChatGPT citations come from the first 30% of content.

KEVIN INDIG — 3 MILLION CHATGPT RESPONSES

Most B2B pages open with context first: *"Vendor risk management has become a strategic priority as supply chains grow more complex."* Restructured for ChatGPT extraction, the same section leads with the answer: *"Vendor risk scoring models typically evaluate suppliers across financial stability, security posture, compliance exposure, and operational continuity."*

Answer capsules are the strongest content signal. Adam Gnuse's analysis of nearly 2 million sessions, published November 2025 in Search Engine Land, found direct Q&A; formatted sections were the single strongest commonality among cited pages. Proprietary or original data was the second strongest differentiator.

Focused depth over exhaustive coverage. Indig's April 2026 analysis found pages covering 26–50% of a query's sub-topics were cited more than pages attempting comprehensive coverage. The highest-cited pages are focused category guides that answer one question well: what a category is, who uses it, how to choose, and what it costs.

Domain concentration is the structural ceiling. The 46%/67% figures are not easily addressed through content optimisation alone. They reflect sustained domain authority and

publishing track records that compound citation gravity over time. For most B2B brands, the realistic near-term goal is being referenced for specific queries where topical authority is achievable, not broad category dominance.

## Turn 1 is where citations happen

Profound's analysis of 730,000 ChatGPT conversations from October through December 2025 found Turn 1 is 2.5x more likely to trigger citations than Turn 10. Citations occur disproportionately at the start of conversations, before the model shifts into generative back-and-forth.

Turn 1 is **2.5x** more likely to trigger citations than Turn 10. Citations occur disproportionately at the start of conversations.

PROFOUND — 730,000 CHATGPT CONVERSATIONS, OCT–DEC 2025

For content strategy, that means the queries that matter are the initial, category-defining ones: what is this, who uses it, how does it compare, what does it cost.

## The optimisation plan

These steps are sequenced by expected return relative to time invested.

**Step 1: Bing indexing audit.** Open Bing Webmaster Tools and verify your priority pages are indexed. For any absent pages, submit them manually. Check robots.txt for rules blocking OAI-SearchBot, including blanket AI-agent disallow rules. Keep GPTBot permissions separate. Blocking training data collection does not affect ChatGPT Search visibility.

**Step 2: Allow OAI-SearchBot.** Confirm `User-agent: OAI-SearchBot / Allow: /` in robots.txt. This is the technical prerequisite for ChatGPT Search visibility. Without it, the rest of the optimisation stack becomes largely irrelevant.

One thing worth checking: many enterprise CMSs and security hardening templates ship with broad AI-agent disallow rules that were added as defaults in 2023–2024. These can block OAI-SearchBot without anyone having made an active decision to do so.

**Step 3: Restructure existing content to front-load the answer.** Review your highest-value pages. Identify sections that open with background context before reaching the direct answer. Move the answer to the top of each section. H2 and H3 headers should reflect how

a buyer phrases the question, not how a copywriter labels a topic.

Do not rewrite pages that currently hold strong Google rankings wholesale. Add answer-first section intros without replacing the existing content structure. Rewriting a page that ranks well in order to optimise for ChatGPT is a trade that rarely makes sense.

Step 4: Add answer capsule sections. For decision-stage pages, add a Q&A; block with direct, extractable answers. Write the questions the way a buyer would search for them. "What does [product] cost for a team of 50?" is usable. "What makes [product] the definitive solution?" is not. Aim for at least four to six questions per page.

Step 5: Build category-level guide pages. The most frequently cited structure in Indig's data is a focused guide covering what a category is, who uses it, how to choose, and what it costs, on a single URL. If your content library runs toward thought leadership and product narrative, these category guides are likely missing.

Step 6: Surface original data. Proprietary statistics and unique findings are the second strongest citation predictor in the Gnuse study. If your brand runs benchmarking or customer surveys, publishing those findings in a citable format increases citation probability independently of domain authority.

Step 7: Add FAQ schema to priority pages. FAQ schema is associated with higher AI citation rates across multiple sources. The specific evidence for ChatGPT Search is not from a controlled study; the most-cited lift figure applies to Google AI Overviews. Implementation is low-cost and the directional signal is consistent, but treat it as a supporting action rather than a primary lever.

---

## ChatGPT and Perplexity draw from different source pools

The most practically relevant difference is infrastructure dependency. ChatGPT Search requires Bing indexing. Perplexity operates its own independent crawler and is not Bing-dependent. A page that ranks well in Google but poorly in Bing may be visible in Perplexity and absent in ChatGPT Search entirely.

ChatGPT Search activates selectively. Many ChatGPT responses come from training data with no live retrieval. Perplexity's architecture is RAG-first and retrieves on most queries rather than selectively routing to training data. For B2B brands, this means ChatGPT Search citations are concentrated in the queries that trigger Search mode, while base ChatGPT responses may draw from training data regardless of how well the page is optimised.

SE Ranking's April 2025 analysis found only 25.19% of cited domains in common between ChatGPT and Perplexity for the same prompts. The two systems draw from largely separate source pools. Optimisation for one does not reliably produce visibility in the other, and the technical prerequisites need to be managed separately.

Semrush's November 2025 analysis of 230,000 prompts found Reddit's share of ChatGPT citations dropped roughly 50 percentage points in mid-September 2025, a shift isolated to ChatGPT. These patterns can change abruptly. The content signals that have held across studies, answer-first structure and direct Q&A; formatting, have been more durable than domain-level patterns.

---

## Track citations directly, not analytics

ChatGPT Search traffic is difficult to measure through standard analytics. Run the queries your buyers ask most frequently in ChatGPT Search mode and check whether your content appears in the cited sources. Track it weekly, covering category-level questions and competitor comparison queries.

Allow time for re-crawl cycles before expecting measurable movement from any technical or content changes.

At scale, manual monitoring becomes impractical. Aiviara is building infrastructure specifically for monitoring AI citation visibility across LLM platforms. If this is a priority for your team, you can join the early access list at [aiviara.com](https://aiviara.com).

---

## What this playbook won't achieve

Domain concentration is the hard ceiling. With 46% of citations in any topic going to the top 10 domains, consistent broad inclusion is a compounding problem. The brands in those top positions are there because of sustained organic authority built through consistent publishing and inbound links, not through citation-specific optimisation.

This playbook targets specific, actionable improvements for brands that already have some domain presence in their category. It will not close the gap on Wikipedia or Reddit for generic category queries.

ChatGPT Search is also mode-dependent. For queries where Search mode is not triggered, none of these retrieval optimisations apply. OpenAI has not published the exact criteria that determine when ChatGPT switches to live retrieval.

---

## Start with access, then earn the slot

Crawler access is the prerequisite. Content structure determines whether retrieved pages are actually cited. Schema appears to be a secondary signal at best.

A page that OAI-SearchBot cannot reach will not appear in ChatGPT Search answers. A page it can reach but that buries the direct answer in background narrative tends to be retrieved and then skipped over in source selection.

For B2B brands with real authority in a specific category, the opportunity in ChatGPT Search is specific queries where category expertise is demonstrable and where direct, extractable answers are genuinely absent from the current top results. It is a narrower target than broad citation presence, but it is where the available evidence points for brands outside the top 10 domains of their vertical.

---

Aiviara is building infrastructure for monitoring AI brand citations and factual accuracy across LLM platforms. Early access information is available at [aiviara.com](https://aiviara.com).