

How to get your page cited by ChatGPT and Perplexity

Answer engines do not rank links, they lift passages and sometimes cite the source. Getting picked is about being easy for a machine to find, parse, trust and quote. Here are the signals that decide it.

Published 2026-06-24 · A Mue guide

More and more buying research now starts in an AI assistant instead of a search box, and the game there is different. An answer engine does not hand back ten blue links; it retrieves a handful of passages, synthesises them into prose, and sometimes credits the sources it used. To be one of those sources your page has to be easy for a machine to find, parse, trust and quote. That is a different job from ranking, and most pages are not set up for it.

What the engine actually does

When someone asks a question, the assistant pulls a few relevant passages from across the web, stitches them into one answer, and may cite the pages it leaned on. So you are not competing for a ranked position, you are competing to be a cleanly liftable passage on the exact thing being asked. Two things follow: write in self-contained chunks, and make the facts unambiguous to a machine.

Let the right crawler in first

None of the signals below matter if the engine cannot fetch the page, and this is where a lot of sites quietly lose. The crawler that reads a page for a live answer is not the same one that reads it to train a model, and the two are controlled separately in robots.txt. OpenAI runs GPTBot to collect training data and OAI-SearchBot to surface and cite pages in ChatGPT search, and its own documentation states the settings are independent, so a site can allow OAI-SearchBot to appear in search results while disallowing GPTBot. The upshot: the common move of blocking GPTBot to keep your content out of training does nothing to your citation odds, but a blanket AI-bot block that also catches the search crawler quietly removes you from the answers.

Perplexity splits the same way. PerplexityBot indexes pages for its answer engine and respects robots.txt, while Perplexity-User fetches a page only when a person asks about it. If you want to be cited, the safe default is to allow the retrieval crawlers, OAI-SearchBot and PerplexityBot, even when you disallow the training ones. So before tuning any of the signals below, read your own robots.txt for a wildcard disallow or a plugin rule that blocks AI user-agents wholesale, because it can shut the front door on the exact engines you are trying to appear in.

The signals that get you quoted

- Descriptive headings: a subheading that names the question it answers splits your page into passages that can each stand alone as an answer, which is exactly what gets lifted.
- FAQ blocks: real questions with short, direct answers map almost one-to-one onto the way people prompt an assistant.
- Citable stats: a specific, attributable figure is the kind of thing an assistant likes to quote, because it is concrete and checkable.
- Structured data (JSON-LD): schema removes ambiguity about what your page describes, so the model is more confident you are relevant.
- Freshness: a visible update date and the current year tell a model the page is not stale, which it weighs when choosing what to cite.
- An llms.txt file: a young convention, but a cheap one, a curated map of your best pages for the crawlers that read it.

Entity clarity comes first

Before any of the formatting, the model has to be sure what and who your page is about. Name the product, company or topic plainly and consistently instead of leaning on pronouns and clever phrasing. Entity clarity, schema and citable stats are usually the three that move the needle most, so fix those before the rest.

Measure it, do not guess

None of this guarantees a spot. The engines change how they retrieve and cite often, results vary from one run to the next and by region, and the honest move is to treat the signals as a checklist of things that genuinely help, then track whether they work. Run a fixed set of buyer prompts each month and record, for each, whether you were mentioned, whether you were cited with a link, and roughly where in the answer. That grid is your AI-visibility trend line. The tool generates those prompts, scores your page on every signal above, and hands you the fixes in priority order.

Frequently asked questions

How do I get my page cited by ChatGPT and Perplexity?

Answer engines lift passages, not links, so make the page easy for a machine to find, parse, trust and quote: put a clear quotable answer near the top, use real headings, add structured data, and let the AI crawlers in.

Does blocking GPTBot stop ChatGPT from citing my page?

No. GPTBot is OpenAI training crawler; OAI-SearchBot is the one that surfaces and cites pages in ChatGPT search, and OpenAI documents the two settings as independent, so you can allow OAI-SearchBot for citations while disallowing GPTBot for training. The real risk is a blanket AI-bot rule that also blocks the search crawler and quietly removes you from the answers.

Which crawlers should I allow to get cited by AI?

Allow the retrieval crawlers even if you block the training ones: OAI-SearchBot for ChatGPT

search and PerplexityBot for the Perplexity answer engine, both of which respect robots.txt. Their user-triggered cousins, ChatGPT-User and Perplexity-User, fetch a page only when a person asks and may bypass robots.txt.

Is getting cited by AI the same as SEO ranking?

No. Ranking returns ten blue links; an answer engine retrieves a few passages, synthesises them into prose and may credit the sources. Being parseable and quotable matters more than classic ranking signals.

Audit your page for AI visibility

Free, no signup: agent.mue.app/tools/ai-search-visibility-auditor

agent.mue.app/articles/how-to-get-cited-by-chatgpt-and-perplexity

