



Ritik Namdev

Free Checklist

AI Source Checklist: Get Picked by Perplexity, ChatGPT and Google AI

How each tool finds sources, side by side, then 24 checks that make one page eligible and easy to quote in all three.

You get a side-by-side of what Perplexity, OpenAI and Google say publicly about sources, and a checklist that covers all three. Use it on your most important page today.



Ritik Namdev

SEO since 2018 · I help founders get customers from Google + AI search
ritiknamdev.com · Updated 2026-10

How to use it

1. Read the comparison table to see how each tool finds pages.
 2. Run the access checks first. If a crawler is blocked, nothing else matters.
 3. Go through the page checks on your top 5 buyer pages, then run the side-by-side test monthly.
-

How each tool finds sources (what the companies say)

	Perplexity	ChatGPT search	Google AI Overviews / AI Mode
How it searches	Searches the web in real time for each question	Searches live pages when search is used; otherwise answers from training data	Uses Google's index and ranking systems; may 'fan out' into related searches
How it shows sources	Numbered citations linking to each source	Inline citations and a Sources list	Supporting links inside and beside the answer
Crawler to allow	PerplexityBot	OAI-SearchBot	Googlebot
Training-only token (blocking does not remove you from search)	Not used for training, per Perplexity	GPTBot	Google-Extended
User-triggered fetcher	Perplexity-User (generally ignores robots.txt)	ChatGPT-User (robots.txt may not apply)	n/a
Official requirement	Allow PerplexityBot and its IP ranges	OpenAI recommends allowing OAI-SearchBot	Indexed and eligible for a snippet; no special optimization

Access checks (do these first)

☐ **robots.txt allows Googlebot and Bingbot.**

Google AI features only use indexed pages. Microsoft Copilot is grounded in Bing.

☐ **robots.txt allows OAI-SearchBot.**

OpenAI: sites that opt out of OAI-SearchBot won't be shown in ChatGPT search answers, apart from navigational links.

☐ **robots.txt allows PerplexityBot, and your firewall/CDN doesn't block its published IP ranges.**

Perplexity says this is how you appear in its search results.

☐ **Decide separately on GPTBot and Google-Extended.**

These control training (and for Google-Extended, Gemini grounding). Google says Google-Extended doesn't affect inclusion in Google Search.

☐ **Your CDN or bot-protection doesn't challenge AI crawlers by default.**

Google lists 'crawling is allowed in robots.txt, and by any CDN or hosting infrastructure'. Check your Cloudflare or host bot settings.

☐ **Key pages are indexed in Google Search Console and Bing Webmaster Tools.**

Use URL Inspection. If it isn't indexed, it can't be a Google AI source.

☐ **Important content is in the HTML text, not only in images, PDFs or behind a login.**

Google: make sure important content is available in textual form.

☐ **No nosnippet or max-snippet:0 on pages you want quoted.**

Google uses these controls to limit what is shown, including in AI features.

Page checks (what cited pages tend to share)

These are not official ranking factors. They're the patterns to check when you look at which pages get cited for your buyer questions.

- ☐ **The page answers one buyer question, and the heading says that question.**
A clear match between question and page makes the page easy to pick for that query.

- ☐ **The first sentence under the heading is the direct answer.**
Easy to lift. Add detail after it.

- ☐ **Facts are specific: prices, numbers, dates, places, names.**
Vague claims are hard to quote. The GEO paper found adding statistics and citations raised visibility in its benchmark.

- ☐ **Claims link to their source.**
The GEO paper tested citing sources as one of its strongest methods. It also helps people trust you.

- ☐ **Author or business name is visible, with an About page that says who you are and since when.**
Google's helpful content guidance asks 'who created the content' and recommends clear bylines.

- ☐ **Updated date is shown and true.**
AI answers about prices and 'best X' are time-sensitive.

- ☐ **Structured data matches what's visible on the page.**
Google lists this in its AI features best practices. Schema doesn't replace visible text.

- ☐ **Internal links point to the page from your menu or related pages.**
Google: make content findable through internal links.

☐ **Comparison and pricing questions have their own pages.**

Buyers ask AI these constantly. A page that answers them directly is a candidate source.

Off-site checks

☐ **Google Business Profile is complete, verified and has recent reviews with replies.**

Google says complete info and reviews help local results, and lists 'Business Profile information is up-to-date' in its AI features best practices.

☐ **You appear on the pages that rank for 'best [category]' in your market.**

When AI tools search, those list pages are often what they read. Ask for a fair listing or a correction.

☐ **Your brand facts match everywhere: site, profiles, directories.**

Conflicting facts give AI tools conflicting answers.

☐ **Buyers discuss you on forums like Reddit, honestly.**

Google and OpenAI both have Reddit data partnerships. See the Reddit playbook for the safe way.

Featured pick

theStacc

theStacc's SEO Auditor crawls your site, suggests fixes and checks whether AI crawlers can access it, and its analytics track visits from AI assistants alongside Search Console and GA4 data.

Disclosure: I work at theStacc as a product growth marketer, so I am biased. Compare it against the free options on this page and pick what fits you.

The side-by-side test (monthly, 20 minutes)

1 Pick 5 buyer questions

Questions a buyer asks before paying, e.g. 'best [category] in [city]', '[your product] vs [competitor]', 'how much does [service] cost'.

2 Run each in all three tools

Perplexity, ChatGPT with search on, and Google (look at the AI Overview or open AI Mode). Use fresh, logged-out sessions.

3 Screenshot and log the cited URLs

Sheet columns: date, tool, question, brands named, cited URLs, were you cited (Y/N).

4 Find overlap

URLs cited by two or three tools are the most valuable. Study them: format, facts, author, date.

5 Fix one page

Apply the page checks to the one page that should answer that question. Retest next month.

Sources

- Perplexity: How does Perplexity work? — <https://www.perplexity.ai/hub/faq/how-does-perplexity-work>
- Perplexity: Perplexity crawlers — <https://docs.perplexity.ai/guides/bots>
- OpenAI Help: ChatGPT search — <https://help.openai.com/en/articles/9237897-chatgpt-search>
- OpenAI: Overview of OpenAI crawlers — <https://developers.openai.com/api/docs/bots>
- Google Search Central: AI features and your website — <https://developers.google.com/search/docs/appearance/ai-features>

- Google: Common crawlers (Google-Extended) — <https://developers.google.com/search/docs/crawling-indexing/google-common-crawlers>
- Aggarwal et al., GEO: Generative Engine Optimization (arXiv 2311.09735) — <https://arxiv.org/abs/2311.09735>
- Google: Creating helpful, reliable, people-first content — <https://developers.google.com/search/docs/fundamentals/creating-helpful-content>
- Google Business Profile Help: Tips to improve your local ranking — <https://support.google.com/business/answer/7091>
- Bing Webmaster Tools — <https://www.bing.com/webmasters/about>

Want the next one?

Every week I send one practical SEO + AI search lesson to founders. You're already on the list.

Follow along on Instagram and YouTube **@ritiknamdev**.

<https://ritiknamdev.com/resources/ai-sources/>