

AI search optimisation checklist

Twenty-six checks for being cited by ChatGPT, Perplexity, Claude and AI Overviews -- and how it differs from classic SEO.

Most of this is ordinary good practice weighted differently, not a new discipline. The genuinely new parts are the crawler decisions in group one and the measurement problem in group five: there is no Search Console for AI search, so manual prompt checks are the measurement.

Can the assistants reach you at all?

- ☐ robots.txt allows the crawlers you want: OAI-SearchBot, PerplexityBot, ClaudeBot, Google-Extended
- ☐ You have made a deliberate decision on GPTBot and Google-Extended rather than leaving the default
- ☐ No WAF or bot rule is challenging those user agents -- check your firewall logs, not just robots.txt
- ☐ Content is in the HTML rather than requiring JavaScript to render
- ☐ Pages return 200 quickly; slow responses get abandoned by crawlers as well as people

Is the answer extractable?

- ☐ Each page answers its main question in the first 40-80 words, before any preamble
- ☐ One page answers one question -- a page covering five topics gets cited for none
- ☐ Headings are real questions or plain statements, not clever labels
- ☐ Facts sit in sentences and tables, not inside images or charts
- ☐ Tables have real headers so a row can be read without the surrounding prose
- ☐ Lists are short and parallel, so an extract does not lose the meaning

Is it worth citing?

- ☐ The page contains something not available elsewhere -- original data, a real file, a tested verdict
- ☐ Numbers carry a source and a date
- ☐ The author is named, with checkable credentials on a real profile
- ☐ The page states when it was last updated, and that date is true
- ☐ You say what you do not know, or where the approach fails. Hedged pages get cited more, not less

Machine-readable signals

- ☐ Structured data describes what is visible on the page and validates
- ☐ A single Person or Organization entity, with the same @id used sitewide
- ☐ sameAs links to the profiles that corroborate the entity
- ☐ Canonical, title and description are unique and accurate on every page
- ☐ No FAQ markup invented for content that is not on the page

Measuring it

- ☐ You have a list of the prompts a buyer would actually type
- ☐ You check those prompts across at least two assistants on a fixed schedule
- ☐ You record whether you were cited, mentioned, or absent -- and which URL was used
- ☐ You watch referral traffic from chatgpt.com, perplexity.ai and similar in analytics
- ☐ You accept that impressions here are largely unmeasurable, and do not invent a number

How it differs from classic SEO

Six differences that change what you do. The last row is the important one.

Aspect	Classic search	AI search
What wins	The best page for a query, judged partly on links	The most extractable, best-sourced passage that answers the specific question
Unit of competition	The page, against nine other pages	The passage, against passages from several sites blended into one answer
What a win looks like	A ranking position and a click	A citation, and often no click at all
Measurement	Impressions, position and clicks in Search Console	Manual prompt checks and referral traffic. There is no Search Console for this yet
What does not transfer	Keyword density, exact-match phrasing, thin pages at volume	All of it works against you -- filler dilutes the passages you want extracted
What transfers completely	Crawlability, unique substance, a named accountable author, accurate structured data	The same four, weighted more heavily

The crawlers, and what each one does

Blocking a training crawler does not stop you being cited. These are separate decisions, and they are routinely confused.

User agent	Who	What it does	Advice
OAI-SearchBot	OpenAI	Surfacing pages in ChatGPT search results	Allow. This is how you get cited.
GPTBot	OpenAI	Collecting data used for model training	Owner's decision. Blocking it does not stop you being cited by OAI-SearchBot.
PerplexityBot	Perplexity	Indexing for Perplexity answers, which cite sources prominently	Allow.
ClaudeBot	Anthropic	Crawling for Claude	Allow unless you have a reason not to.
Google-Extended	Google	Controls Gemini training use -- NOT Search indexing	Blocking it does not affect Google Search rankings. Decide separately.
Bingbot	Microsoft	Bing index, which also feeds Copilot	Allow. Blocking it costs you Copilot visibility too.

Your score

Count the unticked boxes per group. Fix the group with the most gaps before starting anything new.

Can the assistants reach you at all?

Is the answer extractable?

Is it worth citing?

Machine-readable signals

Measuring it