

CHATGPT · OPENAI · AI VISIBILITY · CONTENT STRATEGY · B2B STRATEGY · SEO

Getting cited in ChatGPT: what OpenAI has actually documented

OpenAI has published the eligibility rule for ChatGPT search in full and, for organic web pages, the selection logic in one sentence. On its product and shopping surfaces it names the inputs that improve ranking. For ordinary web pages it says 'multiple factors' and 'placement is not guaranteed.' This guide covers what sits on each side of that line.

AIVIARA RESEARCH · OCTOBER 2026

ChatGPT searches the web for a share of the questions it is asked, and when it does, OpenAI has published one rule for which web pages are eligible and, for ordinary web pages, one sentence about which of those get cited. The rule is crawler access, at robots.txt and at the network edge. The sentence is that ChatGPT "ranks search results using multiple factors intended to help users find relevant, reliable information" and that "placement is not guaranteed." Getting cited in ChatGPT starts with the rule, because it is mechanical and checkable, and then runs into the sentence. That split matters most to B2B marketers deciding where to put content effort, since one half of it is an audit and the other is not. Our earlier guide covers [getting cited in ChatGPT](#) from a different angle; this one goes deeper on the crawlers, the API, and what OpenAI has and hasn't disclosed about selection.

The audience behind that sentence is OpenAI's own figure. In an October 5, 2026 post about its advertising products, OpenAI said ChatGPT "reaches 1.2 billion people each week." OpenAI is more forthcoming elsewhere. In its commerce documentation, where merchants hand it product feeds, it names the attributes that improve ranking and relevance, and its shopping help page lists what the product carousel considers under a heading that says so. For ordinary web pages, the disclosure stops at "multiple factors." None of the shopping disclosure is evidence that those signals apply to ordinary web citations, because OpenAI has not said that.

1.2 billion

people reached by ChatGPT each week, on OpenAI's own figure. For ordinary web pages, OpenAI's account of how it selects what to cite from that audience runs to one sentence.

OPENAI · OCTOBER 5, 2026 · ADVERTISING PRODUCTS ANNOUNCEMENT

ChatGPT decides when to search, and the API lets you see the decision

OpenAI's help documentation says ChatGPT "may search the web automatically when your question would benefit from current information," and that a user can force a search from the tools menu or with a slash command. The page gives no list of which questions qualify. Web search is available on "ChatGPT Free, Go, Plus, Pro, Business, Enterprise, and Edu," per the same page, and to signed-out users.

The feature is nearly two years old. SearchGPT, announced July 25, 2024, was "a temporary prototype" for "a small group of users and publishers." ChatGPT search launched on October 31, 2024, reached all logged-in users on December 16, 2024, and on February 5, 2025 became "available to everyone in regions where ChatGPT is available. No signup required." OpenAI now flags its launch post as a historical record, so its dates are reliable and its product description is not.

Before any ranking happens, ChatGPT rewrites the question. OpenAI's help page documents two inputs. If memory is on, saved details feed the query, so a user who has said they are vegan and live in San Francisco may produce the search "good vegan restaurants San Francisco," and location inferred from IP address does the same. And when ChatGPT "sometimes partners with other search providers," it typically rewrites the prompt into one or more targeted queries for them. OpenAI's own worked example turns a biotech question into "CCR8 immunotherapy drug development 2025" and a follow-up into "CHS-114 conference 2025." The page you want cited has to be findable for a query like that, whatever the user typed.

Which providers depends on the surface. For consumer ChatGPT search, OpenAI's help page names Microsoft and Shopify, by linking to each one's privacy policy under the heading "Information shared with search providers." It does not name Bing. Bing appears on a different page, titled "ChatGPT search for Enterprise and Edu," which says that for users in those workspaces ChatGPT "may share disassociated search queries with the Bing search engine to return web results." So the defensible statement is narrow. Enterprise and

Edu search is documented as using Bing; consumer search is documented as sometimes using Microsoft and Shopify. OpenAI's documentation does not disclose what share of queries go to a partner, or give a site owner any way to tell which path produced a citation.

On the API, the decision is explicit and inspectable. In the Responses API, OpenAI's documentation says "the model can choose to search the web or not based on the content of the input prompt," while in the older Chat Completions integration "the model always retrieves information from the web before responding to your query." OpenAI documents three Responses API modes, from a non-reasoning model that "simply passes along the search tool's responses," through a reasoning model that "actively manages the search process," to Deep Research, which reaches "hundreds of sources" and "can run for several minutes." The API documentation does not say how Deep Research selects among those sources.

Four crawlers, and only one of them controls search eligibility

OpenAI's crawler documentation lists four named user agents. They are independent, version stamped, and each has its own published IP list. The conflation to avoid is GPTBot with OAI-SearchBot.

User agent	Purpose, in OpenAI's words	What it does and does not control
OAI-SearchBot	"used to surface websites in search results in ChatGPT's search features"	Disallowed sites "will not be shown in ChatGPT search answers, though can still appear as navigational links"
GPTBot	Crawls "content that may be used in training our generative AI foundation models"	Disallowing it "indicates a site's content should not be used in training"; no effect on search
ChatGPT-User	"certain user actions in ChatGPT and Custom GPTs," including GPT Actions	"not used to determine whether content may appear in Search"; "robots.txt rules may not apply"
OAI-AdsBot	Validates landing pages submitted as ChatGPT ads; "only visits pages submitted as ads"	May use landing-page content to decide when an ad is relevant; data "is not used to train generative AI foundation models"

The settings are independent, and OpenAI says so directly. A site "can allow OAI-SearchBot in order to appear in search results while disallowing GPTBot to indicate that crawled

content should not be used for training." OpenAI has said this since July 2024, when its SearchGPT announcement stated that "sites can be surfaced in search results even if they opt out of generative AI training." If both are allowed, OpenAI "may use the results from just one crawl for both use cases." A robots.txt change "can take ~24 hours" to register, per the same documentation, which has its own change feed at developers.openai.com/api/docs/bots/rss.xml.

ChatGPT-User is the one to read carefully. It represents a live fetch triggered by a person, and OpenAI says that "because these actions are initiated by a user, robots.txt rules may not apply." A log full of ChatGPT-User requests tells you people are opening your pages from ChatGPT. It tells you nothing about whether OAI-SearchBot can see them.

Step 1: Clear OAI-SearchBot at robots.txt and at the network edge

Eligibility has two halves in OpenAI's statement of it. "To make a website eligible for inclusion, allow OAI-Searchbot to crawl the site and confirm that the website host or content delivery network allows traffic from OpenAI's published searchbot IP addresses." The second half is the one an audit of robots.txt alone will miss. A bot-management rule at the CDN or WAF that challenges or drops unverified crawlers by user agent or IP reputation blocks OAI-SearchBot regardless of what robots.txt says, and the failure is silent, because nothing in your analytics records a crawl that never arrived.

So the check lives in two places. In robots.txt, confirm that no rule disallows OAI-SearchBot, including a blanket wildcard disallow on paths you want cited and any AI-agent block list added during a security review. If the goal is to stay out of training but in search, disallow GPTBot and leave OAI-SearchBot alone. Disallowing OAI-SearchBot is OpenAI's documented general opt-out from search, and sites that do so can still appear as navigational links. At the edge, allowlist the IP ranges OpenAI publishes at openai.com/searchbot.json, then look for successful OAI-SearchBot requests in server logs instead of taking the configuration on trust. Allow a day for a robots.txt change to propagate before concluding anything.

Step 2: What OpenAI has disclosed about getting cited in ChatGPT, and where it says more

For organic web pages, the disclosure is the sentence quoted at the top. Everything else OpenAI says about selection is for its product, shopping and ads surfaces.

Its Agentic Commerce documentation describes product feeds as the way merchants "share structured product data with OpenAI so ChatGPT can accurately surface their products in search and shopping experiences," and says that recommended attributes such as "rich media, reviews, and performance signals" together "improve ranking, relevance, and user trust." Feeds carry a flag, `is_eligible_search`, and OpenAI's description of it ends with "Eligibility does not guarantee display," the same principle as the organic sentence but attached to a named mechanism. OpenAI's live help page on shopping has a section headed "How Product Results are Selected," which says ChatGPT "may consider available options, price, reviews, and ease of use" and weighs "structured metadata from first-party and third-party providers" alongside "OpenAI safety standards and product policies." Its November 24, 2025 shopping research announcement adds that results are "based on publicly available retail sites" while "avoiding low-quality or spammy sites," and that merchants can follow an allowlisting process to ensure they are available to appear.

So on its product and shopping surfaces, OpenAI names structured data, reviews, price, quality and trust as selection inputs. For ordinary web pages, it says "multiple factors." That contrast is not a licence to import the commerce list into an organic content plan, because OpenAI has not said the same signals apply to ordinary web citations, and the two are documented on different pages for different surfaces.

For organic web-page citation in ChatGPT search, we found no OpenAI statement connecting schema.org structured data, content freshness, or llms.txt to ranking or eligibility, across OpenAI's crawler documentation, its search help page, and its now-withdrawn publisher FAQ. On llms.txt specifically, OpenAI does publish one, for its own developer documentation, alongside Anthropic and Gemini, according to the llms.txt specification page maintained by Jeremy Howard. That makes OpenAI a publisher adopting the convention for its own docs, and says nothing about whether ChatGPT reads anyone else's. Our [llms.txt field guide](#) covers what the broader evidence shows.

OpenAI documents that ChatGPT search typically rewrites a question into one or more targeted queries before sending them to a search partner, and its example queries are dense with named entities and a year. A page that answers a buyer's question has to contain the terms such a query would carry. That is an inference from a documented mechanism, and it is the basis for the pair below, written for a procurement software vendor.

Weak version:

Smarter spend control for growing teams

Our platform gives finance leaders complete visibility and puts them in control, with integrations across the tools you already use.

Strong version:

Does Larkfield Procure sync purchase orders with NetSuite and Xero?

Larkfield Procure syncs approved purchase orders to NetSuite and Xero every 15 minutes, posts receipts against them automatically, and includes both integrations on the Growth and Enterprise plans.

The second version names the product the same way twice, names the two systems a buyer would ask about, and would match a rewritten query like "Larkfield Procure NetSuite purchase order sync." The first contains nothing a targeted query could land on. Before restructuring a page that already ranks in Google's top positions for its keyword, check its organic traffic and add a question-formatted section to it instead of rewriting it.

Step 3: Instrument citations in the app, in your logs, and through the API

In the consumer app, OpenAI's help page says "responses that use web search may include citations," that a user can "select a citation to open its source," and that a Sources control, "when available," lists cited sources and other relevant links. A search-backed answer is not guaranteed to cite anything, and OpenAI attaches a caveat that belongs in any report built from the product: "Search results and citations can be incomplete, outdated, or incorrect."

The referral signal is the easiest to capture. ChatGPT appends `utm\_source=chatgpt.com` to referral URLs. OpenAI documented this in a publisher FAQ it withdrew in October 2026, and the parameter remains directly observable in server logs and analytics. Build a segment on it, then check that it survives your own infrastructure. A redirect chain or consent layer that strips query parameters erases the signal before it reaches analytics, and the resulting zero looks like an absence of citations.

The API is where a citation becomes structured data. In the Responses API, a response that used search contains a ``web_search_call`` item recording the action taken, which for a plain search "will usually (but not always)" include the queries used, and which for reasoning models can also be ``open_page`` or ``find_in_page``. The message that follows carries ``annotations`` of type ``url_citation``, each with the cited URL, its title, and the character offsets of the span it supports. The call can also return the complete list of sources the model read, which is distinct from what it cited. OpenAI's documentation adds an obligation on anyone displaying the results. "Inline citations must be made clearly visible and clickable in your user interface."

Two more parameters turn the API into a test rig. The ``filters`` parameter accepts up to 100 ``allowed_domains`` or 100 ``blocked_domains``, in the Responses API only, so a search can be restricted to your own domain to see what the model extracts from your pages when nothing else competes. And Chat Completions exposes, through ``gpt-5-search-api``, what OpenAI describes as "the fine-tuned models and tool used by Search in ChatGPT." Neither is the consumer product, but they are the closest documented view of it that a brand can run on a schedule.

Step 4: Treat partnerships, ads and the shopping carousel as three separate surfaces

Partnerships put content into retrieval without closing the door to anyone else. At launch in October 2024, OpenAI said "any website or publisher can choose to appear in ChatGPT search" and described the retrieval as using "third-party search providers, as well as content provided directly by our partners." The launch post named fourteen news partners, among them the Associated Press, Axel Springer and Reuters. OpenAI's live eligibility rule, which asks only for crawler and IP access, is consistent with that.

The individual deals describe the same mechanism in near-identical words. OpenAI's December 13, 2023 Axel Springer announcement said ChatGPT's answers "will include attribution and links to the full articles." Its December 4, 2024 Future announcement said users would access Future's content "with attribution and links to the full original articles." The Associated Press licensing agreement dates to July 2023, per Adweek's reporting. We found no OpenAI figure for how much more often partner content is cited than crawled content, so "partnering boosts citations" is a plausible reading of "content provided directly," and no more documented than that.

The shopping carousel is documented separately and says the opposite about partnerships. OpenAI's shopping help page states that "product results are selected independently by

ChatGPT and are not ads, nor influenced by any OpenAI partnerships." That does not contradict the partner-retrieval statement. It describes a different surface, the product carousel shown for shopping queries, and the two should be kept apart in any plan.

Ads are a third surface with their own crawler. OpenAI first announced ad testing on February 9, 2026, with the commitment that "ads do not influence the answers ChatGPT gives you" and that they "are always clearly labeled as sponsored and visually separated from the organic answer," and by August 18, 2026 the product had expanded to 31 European markets. Buying an ad does not buy an organic citation, on OpenAI's own account of the separation.

One surface has gone. OpenAI retired ChatGPT Atlas, its standalone browser, around October 10, 2026, with a notice saying it "no longer receives updates," and the publisher FAQ that documented the referral parameter was withdrawn with it.

Step 5: Account for Enterprise and Edu workspaces, where a search may never happen

If your buyers use ChatGPT inside a workspace their employer administers, OpenAI documents controls there that no amount of site work can override. Workspace owners and admins can turn web search off, and OpenAI's Enterprise and Edu page says that when they do, "ChatGPT will not use web search for affected users, even if a user asks ChatGPT to search." Access can also be switched off per role. Some restricted configurations mean "ChatGPT may use cached web results instead of live searches." Healthcare workspaces "do not use Bing for web search." And where a workspace has connected sources, "ChatGPT may prioritize connected sources and use web search only when connected sources cannot answer the request."

The consequence for a vendor is that a prospect's ChatGPT may answer a question about your category from their own connected documents, from cached results, or without searching at all. None of that shows up in your logs. It is a ceiling on what any citation programme can reach.

How to know if it's working

Three readings, kept separate, tell you whether anything changed.

Start with the referral segment. Sessions carrying `utm\_source=chatgpt.com`, by landing page, are the only signal the consumer product gives you without further work, and they

measure clicks. A page can be cited and never clicked.

Alongside that, run a fixed panel of buyer questions through the Responses API on a schedule and store the ``url_citation`` annotations and the sources list for each. Citation count per page, and the gap between pages read and pages cited, are the two numbers to track. Then run the same panel with ``allowed_domains`` set to your own site, to see what the model extracts when it has to use you. The API tool is not the consumer app, and OpenAI's caveat that results can be incomplete or outdated applies to both.

Finally, check a handful of those questions by hand on chatgpt.com, with memory off, and open Sources. This is the only view of the actual product, and the least repeatable one, because memory and location rewrite the query and a logged-in colleague's answer is not yours.

The only timing figure in the OpenAI documentation this guide draws on is the roughly 24 hours a robots.txt change can take. Allow that day and a recrawl before reading any of this, then give the readings several weeks to separate from noise.

A worked example

As a hypothetical, take a procurement software vendor selling to finance teams at companies with 200 to 2,000 employees. Its most valuable ChatGPT question is which procurement tools sync purchase orders with NetSuite.

The access audit finds robots.txt clean but no OAI-SearchBot requests in the logs for six months. The CDN's bot-management profile, set to challenge unverified crawlers during a 2024 hardening pass, has been dropping the IP ranges OpenAI publishes. The ranges are allowlisted, and successful OAI-SearchBot crawls appear within two days. GPTBot stays disallowed, since the company does not want its documentation in training data.

The team then runs 40 buyer questions weekly through the Responses API, storing the ``url_citation`` annotations and the sources list for each, and segments its analytics on ``utm_source=chatgpt.com``. Three integration pages get a direct answer under a question heading, added as a new section above the existing copy, because two of them rank in Google's top five for their keywords.

Eight weeks later, the NetSuite integration page appears in the sources list for 11 of the 40 questions and in the citations for 4, up from none. The allowed-domains run shows the model quoting the page's first paragraph verbatim. The referral segment records 23 sessions in the month. Those are clicks. What the team still cannot see is whether those

citations came through OpenAI's own index or through a search partner, because nothing OpenAI documents would tell them.

What this playbook won't achieve

It will not tell you how ChatGPT ranks two eligible pages against each other. For organic web citation, OpenAI's disclosure is "multiple factors" and "placement is not guaranteed," and we found nothing more specific in its documentation for that surface.

It will not tell you which path produced a citation. OpenAI documents that consumer search sometimes uses Microsoft and Shopify and that Enterprise and Edu search uses Bing, but not the proportion, and gives a site owner no way to attribute a citation to one route.

It will not make structured data, freshness or llms.txt a ChatGPT signal. We found no OpenAI statement behind any of the three for organic web citation. Add schema if other engines reward it, and do not credit it with a ChatGPT citation.

It will not give you third-party citation data for ChatGPT. This guide rests on OpenAI's own documentation and announcements and cites no independent study of which page attributes predict a ChatGPT citation. It also leaves Deep Research alone, since OpenAI documents its scale and nothing about its selection.

What to do with that

OpenAI has told site owners exactly how to get into the pool, has described how it picks organic pages from that pool in one sentence, and has been specific about selection only on its product, shopping and ads surfaces. Until that changes, the work for an organic site is the two-layer access check and pages that carry the terms a rewritten query would carry, measured through the API's citation annotations rather than the referral parameter alone. If OpenAI ever publishes for the open web what it already publishes for product feeds, the content plan changes. The access check does not.

Aiviara is building infrastructure for monitoring AI brand citations and factual accuracy across LLM platforms. Early access information is available at aiviara.com.