

MICROSOFT COPILOT · BING · AI VISIBILITY · CONTENT STRATEGY · B2B STRATEGY · SEO

Getting cited in Microsoft Copilot: what Microsoft has documented about grounding, and how to see it working

Microsoft says Copilot centres its answers on high-ranking web content, and separately that Copilot runs on Bing's search index and that grounding uses the same crawlers and quality signals as search. Bing's guidelines, as reported by Search Engine Journal, spell out which directives remove a page from Copilot responses. And since February 2026 Bing Webmaster Tools shows the grounding queries behind your citations.

AIVIARA RESEARCH · SEPTEMBER 2026

Microsoft Copilot answers information-seeking questions from web search results, and Microsoft says the experiences that serve them are powered by Bing's search index. Microsoft's Transparency Note for the consumer product, updated in August 2026, says Copilot "centers its response on high-ranking content from the web" and lists hyperlinked citations below the text so users can see which pages grounded the answer. If the answer is assembled from high-ranking Bing results, then getting cited in Microsoft Copilot starts with ranking in Bing, and the rest of the job is done with controls and reports that Microsoft has published itself.

Grounding does not replace search; it runs on the same crawlers and quality signals with a layer on top. And Bing Webmaster Tools now shows the grounding queries behind your citations. What the evidence does not include is an independent study of which page attributes predict a Copilot citation, so most of what follows rests on Microsoft describing its own system.

Copilot is several products, and this guide is about the one that cites the public web

The free consumer product is the cleanest case. Microsoft's support documentation states that "Microsoft Copilot is available at no cost on copilot.com" and that "prompts rely solely on public web data and can't access personal files, emails, or organizational content." The Transparency Note, which covers individuals signed in with a personal Microsoft account

across free and paid tiers, says that for information-seeking conversations "Copilot is grounded in web search results." Microsoft 365 Personal and personal Microsoft accounts add the account owner's own OneDrive files, Outlook.com email and calendar, and documents.

Microsoft 365 Copilot for organisations is a different system, and this guide does not cover it. It grounds on Copilot connectors into Microsoft Graph, and Microsoft Learn says "Search and Copilot only show items to users who have access in the source system." Getting into a customer's private Graph index is an integration problem, and nothing below moves it.

Every tier grounds on the public web. Microsoft's plan comparison page describes the free product as giving "real-time answers based on web search" and the business plans as having "AI chat grounded on the web."

Grounding runs on Bing's index, with a selection layer Microsoft has described but not specified

Two Microsoft statements establish the infrastructure. In October 2025, Krishna Madhavan, a principal product manager at Microsoft Bing, wrote on the Microsoft Advertising blog that "powered by Bing's search index, experiences like Microsoft Copilot, Microsoft Start, and others handle billions of queries each month." In May 2026 the Bing team, in a post by Madhavan, Knut Risvik and Meenaz Merchant, addressed the point directly: "A common misconception is that grounding replaces search. It does not." Grounding, they wrote, "builds on the same foundational infrastructure," the same crawlers, the same quality signals, the same understanding of the web, and "adds a new optimization layer on top."

Grounding does not replace search. It builds on the same crawlers, quality signals and understanding of the web that search uses, with an additional optimisation layer on top.

KRISHNA MADHAVAN, KNUT RISVIK & MEENAZ MERCHANT, MICROSOFT BING · MAY 2026

The May post is the most useful thing Microsoft has published on what that layer does. "Traditional search asks: which pages should a user visit? Grounding asks: what information can an AI system responsibly use to construct a response?" It lists five dimensions where the two diverge, and three bear on content work. Factual fidelity is "critical" for grounding because "chunking/transformations must preserve meaning and claims used in the answer." Source attribution becomes a "core signal" where "evidence needs clear provenance and varying evidentiary weight". And where stale content "mainly degrades ranking usefulness"

in search, in grounding "stale facts can directly produce wrong answers." The post also names not answering as a designed outcome. "Abstention is a valid outcome when support is missing, stale, or conflicting," the Bing team wrote. A page can be indexed, ranking, and still not used.

Madhavan's October 2025 guidance describes what happens to a page that is used. "Assistants like Copilot break content down, a process called parsing, into smaller, structured pieces that can be evaluated for authority and relevance. Those pieces are then assembled into answers, often drawing from multiple sources." Ranking, Madhavan wrote, "still happens, but it's less about ordering entire pages and more about which pieces of content earn a place in the final answer."

Read together, those sources describe what grounding measures and how it treats a page once retrieved. They do not say how one page is selected over another for a given grounding query, and any more specific claim about that logic is inference. No Microsoft source cited here names a separate Copilot crawler, and Bing's own crawler-list help page renders only as a JavaScript shell, so it could not be read for this guide. Microsoft's statement that grounding runs on the same crawlers as search is the closest thing to an answer. Treat "allow bingbot and you have allowed Copilot" as a well-supported inference rather than a documented rule.

Step 1: Getting cited in Microsoft Copilot starts with the directives on your own pages

Bing's February 2026 announcement of its AI Performance report states that "Bing respects all content owner preferences expressed through robots.txt and other supported control mechanisms." Put that beside the shared-crawler statement and a page blocked to bingbot is not crawled, and uncrawled content cannot enter the pool Copilot draws from. That conclusion joins two Microsoft statements; no single Microsoft document states it as one sentence.

The directive-level detail comes from trade press, because we could not render Bing's Webmaster Guidelines page itself. Search Engine Journal's Matt G. Southern reported on February 27, 2026 that Microsoft had rewritten the guidelines to spell out how meta directives affect AI experiences, treating grounding results and citations as an eligibility outcome alongside ranking. In Southern's summary, NOARCHIVE prevents content from being used in Copilot responses and grounding results. NOCACHE limits Copilot to using only the URL, title and snippet. NOSNIPPET and DATA-NOSNIPPET may limit citation quality. A data-snippet attribute lets a publisher specify what text Bing can display or cite, and "Bing

recommends against using NOCACHE on content intended for Copilot if you want richer citations." That is Search Engine Journal's paraphrase of the guidelines, not Bing's wording.

For a B2B site the check is a template audit. If a NOARCHIVE tag was added to pricing, documentation or legal pages during a review of what search engines and AI systems could retain, it also removes those pages from Copilot's grounding pool under the guidelines as Southern summarises them. NOCACHE is subtler, because the page still appears, and a citation built from URL, title and snippet alone looks like a weak page rather than a directive problem. Pull the directives for every page you want selected before touching any copy.

Southern also draws a contrast. Google, Southern writes, "says AI Overviews and AI Mode follow the same preview controls used in Search," but "hasn't published a Bing-style, directive-by-directive breakdown for how other tags (like noarchive or nocache) affect those AI experiences." That is Search Engine Journal's characterisation, and it means this audit has an answer key for Bing that it lacks for Google.

Step 2: Push changes through IndexNow, and expect a faster recrawl rather than a faster citation

IndexNow is an open protocol, launched in October 2021 by Microsoft Bing and Yandex, that lets a site inform participating engines when content has changed. In Wikipedia's account, "IndexNow uses an API to notify search engines about content changes to content in order to trigger a recrawl." It does not put content into the index, and the recrawled page still has to pass Bing's normal evaluation.

Microsoft connects it to Copilot explicitly. The February 2026 AI Performance announcement states that "accurate and up to date content is important for inclusion and citation in AI-generated answers" and that "by enabling faster discovery of content changes, IndexNow helps ensure that AI systems reference the most current version of a page when generating answers." That is Microsoft's stated design intent. We found no independent test isolating IndexNow's effect on Copilot citation timing as distinct from Bing indexing speed, so the defensible claim is that it shortens the gap between a change and Bing seeing it, and Microsoft says that gap matters for AI answers.

Adoption is a 2024 number, and it should be read as one. Fabrice Canel, a principal product manager at Microsoft Bing, said at PubCon in March 2024 that IndexNow had reached 2.5 billion submitted URLs, up from 1.2 billion six months earlier, and that 17% of clicked URLs on Bing came through it, as reported by Search Engine Roundtable's Barry Schwartz, and no later figure could be verified. Google has not adopted the protocol; Wikipedia's list of

supporting engines as of 2025 names Microsoft Bing, Yandex, Naver and Seznam.cz.

Where it earns its place is pages whose facts change. Microsoft's own grounding framework says stale facts produce wrong answers. A pricing page updated on Tuesday that Bing recrawls on its own schedule is, in the meantime, a page Copilot may answer from at the old price. Submit the URL when the price changes, and do the same for release notes, integration lists and anything else a buyer might ask about by date. We found no study of whether Copilot prefers newer pages in general; our guide to [getting cited in Google AI Mode](#) covers what the freshness data shows for Google's features, and none of it was measured on Copilot.

Step 3: Write pages that parse into pieces

If Copilot breaks a page into pieces and evaluates each for authority and relevance, as Madhavan describes, the page has to survive being cut up. Southern's summary of the rewritten Bing guidelines lists what that means in practice: facts stated directly rather than implied, so an AI system can verify them independently; entity names that are clear and consistent, with no ambiguous references; and each URL focused on a single topic with the essential information near the top. Microsoft's October 2025 guidance adds clear titles and headings and question-and-answer formatting.

The entity-name rule is the easiest to fail without noticing. A page that refers to the same product as its full name, a two-letter abbreviation and "the platform" within three paragraphs reads fine to a person and gives a parser three entities. The pair below is illustrative, for an identity and access management vendor selling to mid-market IT teams.

Weak version:

Enterprise-grade provisioning

Our platform makes onboarding effortless. With deep integrations across the identity ecosystem, IT teams can automate the entire lifecycle and stay in control at every step.

Strong version:

Does Northgate IAM support SCIM provisioning with Okta and Microsoft Entra ID?

Northgate IAM supports SCIM 2.0 provisioning with Okta, Microsoft Entra ID and Google Workspace. Users and groups sync every 15 minutes by default, deprovisioning runs within one sync cycle, and SCIM is included on the Business and Enterprise plans.

The second version states four facts a system could verify, names the product the same way twice, and puts the answer in the first sentence under a heading that matches how a buyer would ask.

The guidelines rewrite also changed what Bing says about machine-generated content, and the change reads as a warning about scale. Southern quotes the old policy, under which such content "will result in penalties," and the new wording, under which "large-scale content generated without oversight, quality control, or editorial review often lacks usefulness, accuracy, and originality, and may be excluded from indexing." Exclusion from indexing is exclusion from the grounding pool, on Microsoft's own account of the shared infrastructure. A programme that generates a page per integration without an editor reading each one is running that risk on every page.

Before restructuring a page that already ranks in Google's or Bing's top positions for its keyword, check its organic traffic. Add a question-formatted section to those pages rather than rewriting them. The pages to restructure are the ones that rank nowhere and answer nothing in a retrievable form.

Step 4: Add schema for parsing, and don't credit it with a citation

Microsoft recommends it. Madhavan's October 2025 guidance says "schema is a type of code that helps search engines and AI systems understand your content," that it "can label your content as a product, review, FAQ, or event, turning plain text into structured data that machines can interpret with confidence," and it advises adding JSON-LD matching the page type. Separately, in March 2025, Fabrice Canel reportedly said at SMX Munich that schema markup helps Microsoft's LLMs understand your content, according to a LinkedIn post by David Mihm that Search Engine Roundtable reported. That is a conference remark relayed through a third party, not a Microsoft document.

The counter-argument is one journalist's analysis. Roger Montti, writing in Search Engine Journal in January 2025, argued that AI search engines are grounded through retrieval on search indexes, that "search engine indexes themselves are created from crawled data, not Schema structured data," and that using schema to rank better in an AI search engine is "just fanciful speculation." The exchange has not visibly moved since.

There is one 2026 signal, and it is weak. At SEO Week in April 2026, per Search Engine Journal, Madhavan previewed GEO-focused recommendations for Bing Webmaster Tools whose slide listed "structured data adoption, and structured data quality" among the recommendation areas. The feature had not shipped, and the report rests on attendee screenshots.

These positions are compatible. Microsoft's guidance concerns labelling content that has already been retrieved, and Montti's argument is that markup does not decide whether a page is retrieved, since retrieval works on crawled text.

So add JSON-LD matching the page type, because Microsoft recommends it and it is additive. Then do not attribute a change in citations to it. If citation counts move after a schema deployment that coincided with a content restructure, the restructure is the likelier cause.

Step 5: Read the grounding queries Microsoft now shows you

On February 10, 2026, Bing Webmaster Tools launched an AI Performance report in public preview. Microsoft's announcement says it "shows how publisher content appears across Microsoft Copilot, AI-generated summaries in Bing, and select partner integrations." It reports five things: total citations, average cited pages, grounding queries, page-level citation activity, and visibility trends over time. On the grounding queries, Microsoft attaches a caveat that belongs in every report you build from them: "The data shown represents a sample of overall citation activity."

A grounding query is not the same thing as the question a user typed. Microsoft defines the metric as showing "the key phrases the AI used when retrieving content that was referenced in AI-generated answers." Those are the searches Copilot itself ran, for your site.

In March 2026 Microsoft connected the two views. In a Microsoft Advertising post dated March 23, Madhavan and Merchant wrote that "users can quickly see which pages are

being cited for a specific grounding query, and which grounding queries are driving citations to a specific page." Search Engine Journal reported the feature the following day. This is the query-to-page mapping that the manual panel in our Google AI Mode guide exists to approximate, and here it is a report.

Microsoft Clarity adds a second view. Its Citations feature became generally available on May 13, 2026, per the Clarity blog, and Microsoft Learn describes the dashboard as showing "which of your pages are selected as sources, how often they're cited, and how your domain compares to competitors across the same queries." Clarity itself is free, on Microsoft's own description; the Citations feature's terms are not stated separately.

Dan Taylor, in a contributed column in Search Engine Journal in May 2026 about Taylor's own site, reported over 36,000 Copilot citations across all queries and, for 147 grounding queries Taylor rank-checked, that Bing ranked the site for all but 6, most within the top 20, while Google ranked it for none. Taylor describes Clarity showing pages Copilot reads without citing, and argues the patterns transfer to other AI engines; that last point is Taylor's inference. The single-site, self-reported figures are consistent with Microsoft's "high-ranking content" statement, which is the stronger evidence of the two.

One preview should not enter your reporting yet. At SEO Week in April 2026, Madhavan showed a Citation Share metric, the percentage of citations a site captures for a given grounding query, plus intent labels across 15 categories, topic labels and the GEO recommendations mentioned above. Search Engine Journal's report is explicit that "Microsoft hasn't published an official blog post about these features. The information available comes from attendee screenshots of the presentation." Until it ships, it is a slide.

How to know if it's working

Three readings, taken together over eight to twelve weeks, tell you whether the steps above changed anything.

Start with page-level citation activity in Bing Webmaster Tools for the pages you changed, against similar pages you left alone. Site totals rise and fall with query volume and with Microsoft's own changes to the report, so only the page-level comparison isolates your work.

Alongside that, read the grounding query to page mapping. A page cited for the key phrase you targeted is the result you want; a page cited only for its brand name is a sign the question-formatted section is not being selected. Keep Microsoft's sampling caveat attached to any number you report from this view.

The reading that turns the other two into work is Bing rank for the grounding queries you see. On Microsoft's description of Copilot centring on high-ranking content, and on Taylor's single-site data, a page that ranks in Bing for a grounding query is in the pool and one that does not is unlikely to be. A query in the report for which the page ranks poorly is where content work goes next.

Allow a full recrawl before judging. Submit changed URLs through IndexNow, then give page-level citation counts at least two months to separate from noise.

A worked example

As a hypothetical, take an identity and access management vendor selling to IT teams at companies with 300 to 3,000 employees. Its most valuable Copilot question is how the product handles provisioning and deprovisioning with Okta and Microsoft Entra ID.

The directive audit finds a NOARCHIVE tag on every documentation page, added from a template in 2024. It is removed, IndexNow is wired into its documentation CMS so that every published change submits its URL, and three documentation pages are restructured to open with a direct answer under a heading phrased as the question.

Ten weeks later, two of the three restructured documentation pages are showing page-level citations in the AI Performance report, against none on the untouched control pages. The grounding query view maps a query about SCIM provisioning with Entra ID to the provisioning page, and a Bing rank check puts that page fourth for it. The third page shows no citations, and Clarity shows it being read and not selected.

Common mistakes

Treating Microsoft 365 Copilot answers as evidence about public citations. What a customer's Microsoft 365 Copilot says about you may come from their own Graph-indexed content, not from Bing.

Leaving NOARCHIVE or NOCACHE on the pages you want cited. Under the guidelines as Search Engine Journal summarises them, the first removes the page from grounding and the second limits Copilot to URL, title and snippet. Neither shows up in a content review.

Reading the grounding query sample as a complete log. Microsoft states the data is a sample. A query absent from the report has not been shown to be absent from Copilot.

Building a process around Citation Share. It was previewed from a stage in April 2026 and has not shipped. Report what the dashboard measures today.

Generalising Clarity data to other engines. Clarity is a Microsoft product and reports Bing and Copilot behaviour. Taylor's view that the patterns transfer is a practitioner's inference with no evidence behind it in this guide.

What this playbook won't achieve

It will not tell you how Copilot chooses between two eligible pages. Microsoft has published what grounding measures and how it treats a retrieved page, not the selection logic.

It will not settle whether schema causes citations. The evidence is Microsoft's recommendation, a relayed 2025 conference remark, one journalist's 2025 analysis and an unshipped 2026 preview.

It will not give you a Copilot-specific study of what predicts selection. We found no independent analysis of how freshness, structure or authority signals affect Copilot citation likelihood, of the kind that exists for Google's AI features.

It will not tell you which domains Copilot cites most. The datasets that exist are vendors' own proprietary research, so this guide makes no claim about which sites dominate.

It will not compare how Copilot and Google AI Mode select sources. No sourced comparison of the two selection behaviours was found, and constructing one from separate studies of each would be inference dressed as data.

Microsoft has told publishers which searches its assistant ran before citing them, and connected those searches to pages. That turns Copilot citation work from guessing at retrieval into reading a report and ranking for what it shows. The selection logic inside the grounding layer is still Microsoft's to publish, and whether Citation Share ships as previewed will decide how much further the reporting goes. Until then, the work is Bing ranking, the directives, IndexNow on the pages whose facts change, and a weekly read of the grounding queries.

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