

SEO & GEO

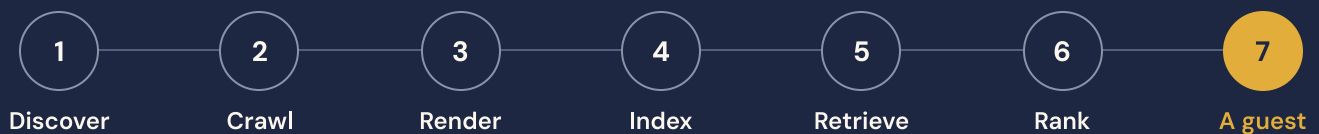
without the folklore

A field guide for short-term rental hosts

What Google, Bing and the AI answer engines have actually put in writing about being found, read closely, labeled by how firmly they said it, and translated for the website of a vacation rental.

Dustin Hofer

Founder of Book With Haven



EVERY STAGE FAILS IN ITS OWN WAY. DIAGNOSE THE EARLIEST ONE FIRST.

Why this guide exists, and how to read it

Search advice arrives with unusual confidence. One guide insists a practice is essential; the next declares it obsolete; neither points to anything a search engine has published.

Much of what circulates as SEO knowledge began life as a correlation study, a vendor's opinion, or a remark repeated until it hardened into a rule. Hosts are a favourite audience for it, because a rental website lives or dies on whether a traveler can find it, and fear is easy to sell to.

This guide takes a narrower road. It reports only what the companies operating the engines have committed to in writing: Google's Search Central documentation, Microsoft's Bing publications, Schema.org, the web standards bodies, and the crawler policies of the AI companies whose answer engines now sit between a traveler and your site. Where those documents run out, the guide writes *unknown* instead of filling the gap.

That restraint is the point. Knowing where the evidence ends is what allows you to decline the next expensive promise with a clear conscience.

There are no secrets here that'll automatically rank your site first in Google.

GOOGLE · SEO STARTER GUIDE

HOW FIRMLY THE SOURCE SAID IT

DOCUMENTED

The engine states the behavior outright, or names it as something its ranking systems use.

RECOMMENDED

The engine explicitly encourages the practice. A recommendation is not a requirement.

DISMISSED

The engine says, in so many words, that it does not use or need this.

UNKNOWN

No first-party document settles it. Not disproved; simply not established.

PRACTICAL

Our reading of what the documented rules mean for a rental site. Labeled so it is never mistaken for an engine rule.

THREE RULES THAT HOLD ON EVERY PAGE

- 1 A recommendation is not a requirement.
- 2 Eligibility for a feature is not a guarantee of it.
- 3 What Google documents says nothing about Bing, ChatGPT or Claude unless the source names them.

The ladder that SEO advice keeps collapsing

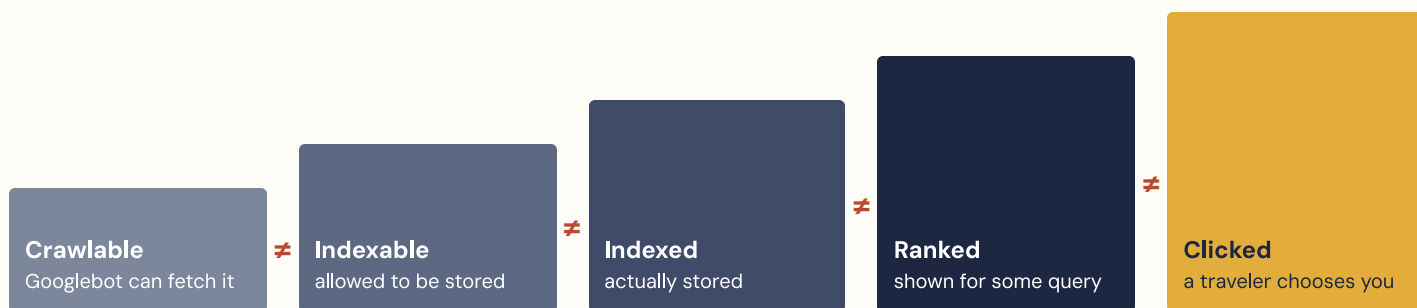


Figure 1. Each rung is a separate outcome. Google says not to expect every URL to be indexed, "only the canonical pages," and an indexed page "is not guaranteed" to be served. Position, in turn, does not decide clicks; result features and query demand intervene. A crawler fetching your page proves the first rung and nothing above it.

The journey of a page, and where it falls off

Consider a lake house with a perfectly decent website. The description is honest, the photographs are strong, the calendar works. Three months later a search for a lake house near that lake returns two marketplaces, a property manager, and no trace of the host.

The instinct is to hire whoever promises to rank it. The more useful instinct is to ask where, exactly, the page fell out of the process. An engine has to learn the URL exists, fetch it, run its scripts, decide whether to store it, match it against a query, order it against rivals, and finally draw it on a results page. Each step fails in its own manner. A failure upstream makes every downstream fix cosmetic; no rewrite rescues a page that was never indexed.

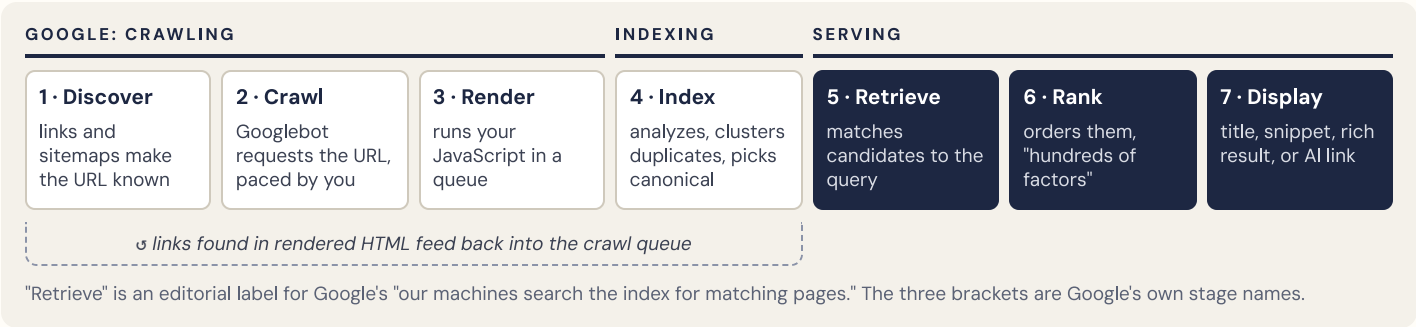


Figure 2. Where you have real influence: stages 1 to 4 respond to how the site is built; 5 to 7 respond to what the page says and who refers to it. Google confirms that a URL "can only" be crawled through an `<a href>` link, that rendering happens in a queue that "can take longer than" a few seconds, and that your canonical preference "is a hint, not a rule."

Find the earliest break, then stop looking

SEARCH CONSOLE SAYS	STAGE	WHAT IT USUALLY MEANS	DO NEXT
The page is not in any report	Discover	An orphan: nothing links to it that Google can follow	Link it from navigation or a hub; list it in the sitemap
Discovered, currently not indexed	Crawl	Google found it, but "rescheduled the crawl," often to avoid overloading the site	Fix 5xx errors; cut the flood of date and filter URLs (page 5)
Crawled, currently not indexed	Index	Fetches, then not selected; usually duplicate or thin	Give the page a distinct reason to exist
Duplicate, Google chose different canonical	Index	Google treats another URL as the real one	Investigate before "fixing"; often the date-parameter problem
Excluded by noindex · Blocked by robots.txt	Crawl / Index	You, or your platform, told it not to	Check templates and the robots file
Indexed, no impressions	Retrieve / Rank	In the system, not winning	Improve the answer, the title, internal context, outside references

ON PATIENCE

Google: "Some changes might take effect in a few hours, others could take several months." It suggests waiting "a few weeks" before judging any change. A fix evaluated after two days may simply not have been processed.

What "ranking factor" means, and what the engines actually publish

The trade speaks of ranking as a scoreboard: discrete factors, each with a weight, tuned one at a time. Google's documentation describes something else.

The unit Google names is the system, each with a stated purpose and no published inputs, and it calls its own inventory a guide to "some of our more notable ranking systems." *Some* is the operative word. Google adds that "the weight applied to each factor varies depending on the nature of your query": there is no fixed formula to reverse-engineer.

SYSTEMS GOOGLE NAMES, AND WHY A HOST WOULD CARE

Link analysis & PageRank

"understand how pages link to each other"

Links matter. Their value is unpublished.

Freshness systems

"fresher content for queries where it would be expected"

Query-dependent. Not a reason to redate pages.

Original content

"original content prominently... ahead of those who merely cite it"

Your own description, not the OTA's copy of it.

Passage ranking

"identify individual sections or 'passages'"

A well-built section can carry a page on its own.

Site diversity

"generally won't show more than two web page listings from the same site"

Two of your pages per query, at most.

Spam detection

"content and behaviors that violate our spam policies"

See the rental-site traps on page 9.

Google also confirms one input everyone argues about: it uses "aggregated and anonymized interaction data to assess whether search results are relevant to queries." **DOCUMENTED** It names no metric. Bounce rate, time on site and anything from Google Analytics remain **UNKNOWN**; the statement concerns behavior on Google's results page, not on yours.

BING PUBLISHES AN ORDER. GOOGLE DOES NOT.

Microsoft lists "the main parameters Bing uses to rank pages" in "general order of importance," then qualifies the order: it "may vary from search to search and evolve over time."

1. **Relevance**
2. **Quality and credibility** · reputation, transparency of ownership
3. **User engagement** · did users click, "spend time," or "quickly return to Bing"
4. **Freshness**
5. **Location and language**
6. **Page load time**

Describes Bing only. **BING**

THE ONE QUALITY INPUT GOOGLE NAMES BY KIND

One of the factors used to determine quality is understanding if other prominent websites link or refer to the content. This is generally a good sign that the information is trustworthy.

GOOGLE · HOW SEARCH WORKS

The strongest documented quality signal, in other words, is one you cannot add to your own page.

Dismissed by Google, by name

the meta-keywords tag

content length as such

keywords in your domain name

heading order

E-E-A-T as a ranking factor

payment for organic rank

quality raters' scores on your page

In Google's words: the meta keywords tag has "no effect on indexing and ranking at all"; "the length of the content alone doesn't matter"; domain keywords "have hardly any effect beyond appearing in breadcrumbs"; on E-E-A-T as a factor, "No, it's not"; raters' ratings "do not directly impact ranking"; ads have "no impact on your organic search ranking." **DISMISSED**

Shape the site so an engine can hold it

Google names exactly three conditions for a page even to be considered, and then closes the loophole: meeting them "doesn't mean that a page will be indexed; indexing isn't guaranteed." Everything in this part removes a documented obstacle. Nothing in it supplies relevance.

1 Googlebot isn't blocked

by robots.txt, a firewall, or your CDN's bot rules

2 The page returns HTTP 200

"Google only indexes pages that are served with an HTTP 200"

3 The page has indexable content

text Google can read, in a supported format

A system of pages, not a pile

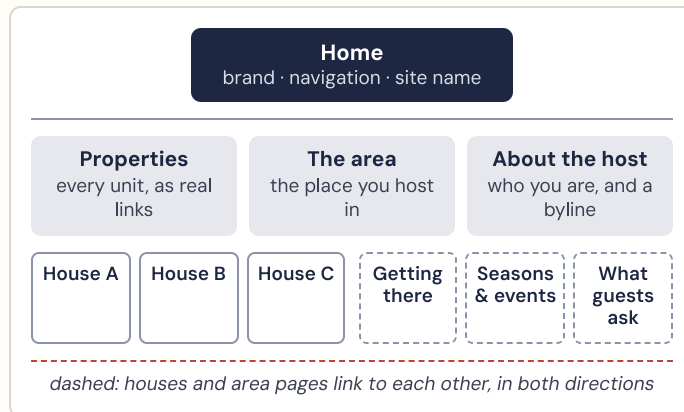


Figure 3. Google: "Every page you care about should have a link from at least one other page on your site." The worth of this shape is in crawlable links between related pages, not in folder names. Do not canonicalize a house to the Properties hub because the hub is broader; canonicals are for duplicates.

The rental-site URL problem

One house, dozens of addresses: /lake-house, /lake-house?checkin=..., ?guests=6. Google clusters near-duplicates and elects one. Left to itself it may elect inconsistently, and Microsoft reports the same clustering in AI systems, which "choose one page to represent the set."

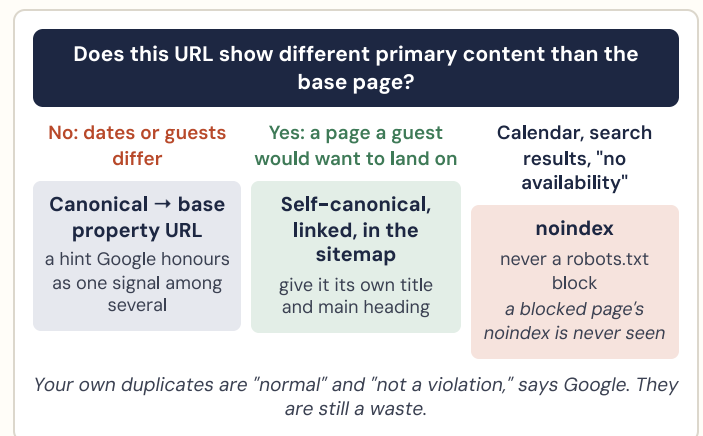


Figure 4. A **PRACTICAL** reading of Google's canonicalization and noindex documentation for booking sites.

TRAP · ROBOTS.TXT IS NOT REMOVAL

A disallowed URL "can still be indexed if linked to from other sites." To keep a page out, Google says to "block indexing with noindex or password-protect the page." And a noindex on a page you have also blocked in robots.txt is never read.

TRAP · THE EMPTY SEARCH THAT RETURNS 200

An availability search that says "no properties match" while returning a success code is reported as a soft 404. Return a real 404, or noindex those pages. Indexed URLs that begin returning 4xx "are removed from the index," which for those pages is the goal.

WHAT GOOGLE RECOMMENDS A URL LOOK LIKE

"When possible, use readable words rather than long ID numbers in your URLs," and "We recommend using hyphens (-) instead of underscores (_)." Google also notes its URL handling "is case sensitive," so pick lowercase and keep to it. A house at /the-lake-house follows every one of those; /Listing.aspx?id=48213 follows none. For ranking, Google says words in the path "alone have hardly any effect beyond appearing in breadcrumbs"; the case for readable URLs is the guest reading them, not the algorithm. **RECOMMENDED**

The checklist, with the engine's own words beside each item

Hand this page to whoever builds or maintains your site. Every row carries the strength of its source; none of them is a ranking trick, and Google says so about several.

ITEM	STATUS	WHAT THE SOURCE SAYS
HTTPS on every page; HTTP redirected to it, never the reverse	RECOMMENDED	Google "strongly recommends using HTTPS." A redirect from HTTPS back to HTTP, or a bad certificate, causes Google "to prefer HTTP very strongly."
One host form (www or not), the other 301-redirected	RECOMMENDED	A redirect is "a strong signal" for choosing the canonical. Bing recommends 301s "to consolidate variants into a single preferred URL."
Every property card, menu item and pagination control is a real <code><a href></code>	DOCUMENTED	Google "can only crawl your link if it's an <code><a></code> HTML element with an <code>href</code> attribute." A card that opens by JavaScript click handler is not a link to Googlebot.
Important content present as HTML text, not only inside a booking widget	RECOMMENDED	"Making sure important content is available in textual form" is on Google's own list of fundamentals for AI features. Rendering is queued and may be skipped on non-200 pages.
Photographs as <code></code> with alt text and readable filenames	RECOMMENDED	Alt text is "the most important attribute" for image metadata and "useful as anchor text." Google's own filename example: <code>my-new-black-kitten.jpg</code> . "Google doesn't index CSS images," so a hero photo set as a CSS background does not exist to Image search.
A mobile layout with the same content as desktop	DOCUMENTED	Google "primarily indexes the mobile version of the content"; most crawl requests come from the smartphone crawler.
Core Web Vitals in the "good" range (below)	DOCUMENTED signal	"Core Web Vitals are used by our ranking systems." Also: Google "always seeks to show the most relevant content, even if the page experience is sub-par," and chasing a perfect score "just for SEO reasons may not be the best use of your time." No weight is published.
Lazy-loading that does not wait for a scroll	RECOMMENDED	Recommended methods "don't rely on user actions, such as scrolling or clicking." A gallery that loads on scroll may never load for the renderer.
XML sitemap of canonical, indexable, 200 URLs only; accurate lastmod; no priority or changefreq	RECOMMENDED · optional	A site of "about 500 pages or fewer" that is well linked may not need one. Google uses lastmod only "if it's consistently and verifiably" accurate; both engines ignore priority and changefreq. Bing calls an honest lastmod "a key signal" for recrawl priority.
IndexNow notification when a page changes	BING supported	Tells Bing and partner engines about added, changed or removed URLs. Google does not participate. "Submitting a URL does not guarantee immediate indexing."

CORE WEB VITALS · THE "GOOD" THRESHOLDS AT THE 75TH PERCENTILE OF REAL VISITS

LCP · LARGEST CONTENTFUL PAINT

≤ 2.5 s

how long until the main photo or heading appears

INP · INTERACTION TO NEXT PAINT

≤ 200 ms

how quickly a tap on the date picker responds

CLS · CUMULATIVE LAYOUT SHIFT

≤ 0.1

how much the page jumps as it loads

Defined by Chrome's web.dev; the ranking role comes from Google's page-experience documentation. A new site has lab data only: pages enter the real-user report once they have "a minimum number of visitors," a number Google does not disclose. Bing lists "page load time" last among its six parameters.

The words guests use, and where Google says to put them

Underneath all the keyword folklore sits one plainly documented signal. Google: "The most basic signal that information is relevant is when content contains the same keywords as your search query." **DOCUMENTED** Read it at its own strength. "Most basic" does not mean strongest, and the same page immediately supplies the limit, using dogs as its example: someone searching for dogs does not want a page with the word "hundreds of times." Google's systems look instead for "other relevant content beyond the keyword," such as pictures, video, or a list of breeds. Repetition is the counter-example. Completeness is the point.

Where the words belong is equally specific. Search Essentials names four locations and no others: "the title and main heading of a page, and other descriptive locations such as alt text and link text." **RECOMMENDED** Not the first hundred words. Not every subheading. Not bold. No density, no percentage, no "LSI keywords"; none of those appears in any Google or Bing document.

Nor do you need every phrasing. RankBrain returns relevant pages "even if it doesn't contain all the exact words used in a search," and Google's 2026 guide to its AI features says outright that "you don't have to worry that you don't have enough 'long-tail' keywords." Building a separate page for each variation, if done "primarily to manipulate rankings or generative AI responses," is named as a violation of the scaled content abuse policy. **DOCUMENTED**



Figure 5. The four locations Google's Search Essentials names. Everything else is inference.

PRACTICAL · THE WORDS YOUR GUESTS ACTUALLY TYPE

Newcomers say "cabin"; regulars say the lake's name. Google's own illustration contrasts "charcuterie" with "cheese board" and suggests writing for the reader you actually have. Your most winnable query is your property's name: travelers who saw you on a marketplace search it to find you directly. Put it in the title, and make sure the page that answers it is yours.

How your result will look

Bear Cove Lake House
bearcovelakehouse.com › the-house

Lake House at Bear Cove · Sleeps 8, private dock

Three-bedroom house on the west shore of Bear Lake with its own dock, a wood stove, and a plowed road in winter. Ten minutes to the village; forty to the ski hill ...

- site name (WebSite markup)
- title link · no length limit
- snippet · mostly from page content

No character limit. "While there's no limit on how long a <title> element can be, the title link is truncated... typically to fit the device width." The same sentence exists for descriptions. Sixty and one hundred sixty are folklore. **DISMISSED**

Descriptive, concise, site name once. Google's incorrect example is *Foobar, foobar, foobars, foo bars*. Let the visible heading agree with the title; competing headings "can be confusing." **RECOMMENDED**

The description is a suggestion. Snippets "are primarily created from the page content itself"; Google "will sometimes use" yours. Write one per page regardless.

What Google means by helpful, on a property page

Google's ranking systems "are designed to prioritize helpful, reliable information that's created to benefit people."

What it hands creators is not a scored checklist but a set of questions to "gauge" their own work against. **RECOMMENDED**

Several of them translate directly into the writing of a rental page.

GOOGLE'S QUESTION, VERBATIM	ON A PROPERTY PAGE, THIS MEANS
"Does the content provide original information, reporting, research, or analysis?"	The particulars only you know. Which bedroom takes the morning sun. How far the dock sits from the water in a dry August. Where to park when the festival closes the main street.
"Does your content clearly demonstrate first-hand expertise... (for example... visiting a place)?"	You have slept there. Say what you learned, and show it: Google's advice for reviews is to "provide evidence such as visuals, audio, or other links of your own experience."
"Does the content provide substantial value when compared to other pages in search results?"	The marketplace already lists your amenities, often in your own words. What does your page add that theirs cannot carry?
"After reading your content, will someone leave feeling they've learned enough about a topic to help achieve their goal?"	Can a guest decide to book without emailing you first?
"Are you writing to a particular word count because you've heard or read that Google has a preferred word count? (No, we don't.)"	Write until the guest's questions are answered. Then stop.

Say who you are

Google's phrasing here is unusually firm: "We strongly encourage adding accurate authorship information, such as bylines to content where readers might expect it." A host page with your name and a way to reach you is exactly that. **RECOMMENDED**

What Google does not say is that a byline moves rankings; it says clear authorship leaves you "likely aligned with the concepts of E-E-A-T." E-E-A-T itself is a framework for its human quality raters, not a score its systems compute, and "trust is most important" within it.

Commodity, and the other thing

Google's 2026 guide names the target: "non-commodity content" offers "unique expert or experienced takes that go beyond common knowledge and the ordinary." It gives an example title of its own: a look inside a sewer line. **RECOMMENDED**

COMMODITY

"Top 10 Things To Do in Bear Lake," assembled from the tourism board.

VS

NON-COMMODITY

"What we tell every guest about the road in February," with the photograph of the plow.

Writing with AI

Google's position sits in one place: "Appropriate use of AI or automation is not against our guidelines," and "Using AI doesn't give content any special gains." The violation is a matter of purpose and scale. Using AI "to generate many pages without adding value for users may violate Google's spam policy on scaled content abuse," a policy that applies "no matter how it's created." Draft with it if you like. The first-hand facts still have to come from you. **DOCUMENTED**

Dates and the freshness reflex

Google's freshness systems show "fresher content for queries where it would be expected." A festival weekend qualifies; how many the house sleeps does not. **DOCUMENTED** Changing a page's date "to make them seem fresh when the content has not substantially changed" is a listed warning sign, and bulk-adding or deleting content to seem fresh draws Google's own parenthetical: "(No, it won't)." **DISMISSED** When a page genuinely changes, show a visible date and keep any dateModified in markup consistent with it.

PRACTICAL · THE MARKETPLACE COPY PROBLEM

Pasting your listing description onto your own site is not a penalty. It is a lost opportunity: Google's original-content systems aim to show original work "ahead of those who merely cite it," and the marketplace publishes your words with far more links behind them. Write the site version fresh, and make it the fuller one.

Traps that catch rental sites, and markup that does less than promised

Google's spam policies define violations by purpose, and its examples map neatly onto a rental website. The consequence is conditional: violating sites "may rank lower in results or not appear in results at all."

POLICY	GOOGLE'S OWN EXAMPLE	THE RENTAL-SITE VERSION
Doorway abuse	"multiple domain names or pages targeted at specific regions or cities that funnel users to one page"	Twenty "cabin rental near [town]" pages that say the same thing and all point at one house. Area pages are fine when each carries distinct, useful content.
Keyword stuffing	"Blocks of text that list cities and regions that a web page is trying to rank for"	The footer listing every town within sixty miles.
Scaled content abuse	"many pages... generated for the primary purpose of manipulating search rankings"	Auto-generated "[Property]" for [Holiday]" pages, however they were produced.
Link spam	"Buying or selling links for ranking purposes"	Paid directory listings without rel="sponsored". Paid links are permitted when so marked.
Hidden text	white-on-white text, text "positioned off-screen"	Tabs and accordions are explicitly not violations. Long amenity lists belong in them.

Structured data: three layers people run together

Structured data (JSON-LD using Schema.org vocabulary) makes a page eligible for a richer appearance. Google: "Using structured data enables a feature to be present, it does not guarantee that it will be present." A structured-data manual action costs eligibility but "doesn't affect how the page ranks in Google web search." It is not a documented ranking factor, and for AI features "there's no special schema.org markup you need to add." **DOCUMENTED** Bing calls markup "one of the clues" it uses to understand a page.



THE VACATION-RENTAL TYPE MOST HOSTS CANNOT USE

Google does document VacationRental markup. Its own page says it is "intended for sites that have already connected with a Google Technical Account Manager and have access to the Hotel Center." Partner-gated, in other words. Adding it to an ordinary site produces nothing the documentation describes.

Rules that apply to all of it: "Don't mark up content that is not visible to readers of the page." "Items that are missing required properties are not eligible." Test in the Rich Results Test, remembering that a pass "does not guarantee that your page will appear exactly as shown."

TYPE	ON A RENTAL SITE	WHAT GOOGLE DOCUMENTS
WebSite · name, url	Yes, on the home page	Eligibility for your site name in results. Domain or subdomain level only.
Organization	Yes, home or about page	No required properties. Logo "must be 112x112px, at minimum." Use the same name as your site name.
BreadcrumbList	Where a visible trail exists	Rich result is desktop-only since January 2025.
LocalBusiness	Only with a real business address	Requires name and address. No documented effect on local ranking or on a Business Profile.
Review · AggregateRating on your own house	No	"If the entity that's being reviewed controls the reviews about itself... ineligible for star review feature." Show reviews as text; do not mark them up for stars.
FAQPage	No longer produces a result	"FAQ rich results are no longer appearing in Google Search" (May 2026); documentation removed. An on-page FAQ can still help readers.
ImageObject · IPTC photo metadata	Optional	Eligibility for the Licensable badge in Google Images.

Links, reviews, and the Business Profile rule most hosts never hear

Links

Google: "We have various systems that understand how pages link to each other as a way to determine what pages are about and which might be most helpful." PageRank remains "one of the fundamental algorithms at Google." **DOCUMENTED** The value of any one link, how it divides among a page's links, how it fades: all **UNKNOWN**. No first-party tool reports a "domain authority"; that number is a vendor's invention.

What the documentation does support: earning references from prominent sites (the one quality input Google names by kind); linking out to cite sources, which "can help establish trustworthiness"; and marking anything paid as `rel="sponsored"` and anything a guest posts as `rel="ugc"`. Buying links "for ranking purposes" is link spam.

PRACTICAL · LINKS A RENTAL CAN HONESTLY EARN

The chamber of commerce. The wedding venue whose guests you house. The race, regatta or festival you are the nearest bed to. The regional tourism site. A local journalist writing about the lake. Ask that the link describe you ("the lake house on the west shore"), not "click here."

Reviews, on Google's platforms

The Maps and Business Profile policies are strict and written down. No incentives: offering "free or discounted goods or services, in exchange for customers to post reviews" is "strictly prohibited." No gating: merchants may not "selectively solicit positive reviews from customers." Asking every guest, with a link or QR code, is permitted; Google itself observes that "a mix of positive and negative feedback often feels more trustworthy." Replies "can help your business stand out"; Google does not say they affect ranking.

ELIGIBILITY · A RENTAL IS NOT A BUSINESS LOCATION

"To qualify for a Business Profile on Google, a business must make in-person contact with customers during its stated hours." The policy lists as ineligible "Rental or for-sale properties such as vacation homes, model homes, or vacant apartments." **DOCUMENTED**

An individual rental cannot have a profile. A management company with a staffed office that receives customers may. A virtual office the business "doesn't operate out of" does not qualify, and putting a town or "vacation rentals" into the business name is prohibited and "could result in the suspension of your Business Profile."

If you are eligible: how Google describes local ranking

Relevance

how well the profile matches the search

Distance

how far from the searcher

Prominence

how well-known: links to the business, "how many reviews you have"

Local results on Maps and Search

Figure 6. From Google Business Profile Help. No weights are published. "There's no way to request or pay for a better local ranking on Google." Separately: "The categories you select affect your local ranking."

Bing Places is a separate, free listing. Microsoft recommends it so "key details such as address, hours, and contact information remain current" in AI answers. Its eligibility rules could not be verified from readable documentation. **BING**

Marking the links you publish, as Google documents it

Paid, sponsored, affiliate

`rel="sponsored"`

Preferred; nofollow "is still an acceptable way to flag them." Unmarked paid links "for ranking purposes" are link spam.

Posted by a guest or user

`rel="ugc"`

Or nofollow. Google says you might remove it for users "who have consistently made high-quality contributions."

You do not vouch for it

`rel="nofollow"`

For links behind a login, untrusted content, or when "you'd rather Google not associate your site with" the page.

An ordinary citation

`no attribute`

Linking to the ferry timetable or the trail map is what "citing your sources" means. It needs no qualifier.

Google, 2019: "There's no wrong attribute except in the case of sponsored links." Qualified links "will generally not be followed," but the page they point to "may still be crawled" by other routes; a `rel` value is not an access control. **RECOMMENDED**

GEO, as the engines describe it

"Generative engine optimization" and "answer engine optimization" are new names, and Google's 2026 guide takes them on by name: "From Google Search's perspective, optimizing for generative AI search is optimizing for the search experience, and thus still SEO." Its closing priority reads, "Prioritize effective SEO strategies over 'AEO/GEO hacks'."

The reason is architectural. Google defines grounding as "relying on our core Search ranking systems to retrieve relevant, up-to-date web pages from our Search index." Microsoft: "Many LLMs rely on data grounded in the Bing index or other search indexes." An AI answer is assembled from search results. A page that cannot be indexed and retrieved cannot be cited, and the preceding nine pages are the GEO work.

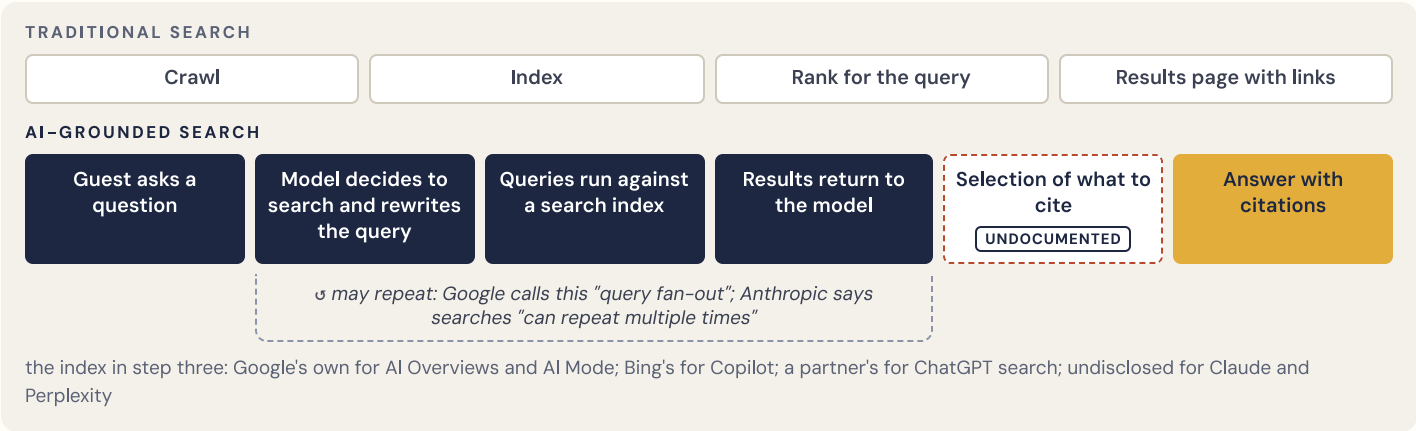


Figure 7. Documented steps only, composited from Google's fan-out and grounding definitions, OpenAI's query rewriting, Anthropic's web-search description and Microsoft's developer grounding flow. No provider, anywhere, documents how a retrieved page becomes a cited one.

Google's own mythbusting list

SOLD AS "AI OPTIMIZATION"	GOOGLE'S DOCUMENTED POSITION
llms.txt, AI text files, Markdown copies of pages	"Google Search itself doesn't use them." Keeping one for other services "will neither harm nor help your site's visibility or rankings in Google Search." DISMISSED
Chunking content into small pieces	"There's no requirement to break your content into tiny pieces for AI to better understand it." Also: "There's no ideal page length." DISMISSED
Rewriting pages in an "AI-friendly" voice	"You don't need to write in a specific way just for generative AI search." DISMISSED
Buying mentions across the web	Seeking inauthentic mentions "isn't as helpful as it might seem." DISMISSED
Special schema for AI	"Structured data isn't required for generative AI search, and there's no special schema.org markup you need to add." DISMISSED

What Google does list as worthwhile for its AI features is the ordinary list: crawling allowed, including by your CDN; content reachable through internal links; a good page experience; "important content... available in textual form"; quality photographs and video; structured data that matches the visible text; a current Business Profile. To be a supporting link in AI Overviews or AI Mode, a page "must be indexed and eligible to be shown in Google Search with a snippet." Then: "There are no additional technical requirements." **DOCUMENTED**

While specific optimization isn't required for AI Overviews and AI Mode, all existing SEO fundamentals continue to be worthwhile.

GOOGLE · AI FEATURES AND YOUR WEBSITE

Bing's advice, the other answer engines, and the crawlers that decide

Bing is the one engine offering content advice for AI

When Microsoft launched its AI Performance report in February 2026 it attached recommendations. They are recommendations in a product announcement, not ranking statements, and they belong to Bing alone.

RECOMMENDED BING

- "Clear headings, tables, and FAQ sections help surface key information"
- "Examples, data, and cited sources help build trust when content is reused"
- "Regular updates help ensure AI systems reference the most current version"
- "Align text, images, and video so they consistently represent the same entities"
- Use IndexNow; register with Bing Places for Business

Microsoft also states a mechanism Google does not: "LLMs group near-duplicate URLs into a single cluster and then choose one page to represent the set." The URL cleanup on page 5 is GEO work by Bing's own account. Google, for its part, makes no headings-and-tables recommendation; clear headings are not chunking, and each position stays with its engine.

Each provider runs several bots, for different jobs

PROVIDER	SEARCH INCLUSION	TRAINING	DOCUMENTED EFFECT OF BLOCKING
OpenAI	OAI-SearchBot	GPTBot	Sites opted out of OAI-SearchBot "will not be shown in ChatGPT search answers." Blocking GPTBot alone does not remove you: "each setting is independent."
Anthropic	Claude-SearchBot	ClaudeBot	Disabling Claude-SearchBot "may reduce your site's visibility and accuracy in user search results."
Perplexity	PerplexityBot	none documented	Surfaces and links sites; "not used for foundation-model training."
Google	Googlebot	Google-Extended, a control token rather than a crawler	Does not affect Search or AI Overviews. Search inclusion and AI features ride on Googlebot together.
Microsoft	bingbot	the noarchive meta tag	Per 2023 documentation, noarchive keeps a page out of Copilot answers and training; nocache limits use to URL, title and snippet.

User-triggered fetchers such as ChatGPT-User and Google-Agent sit alongside these; providers say robots.txt "may not apply" to them because a person asked for the page.

THE BOUNDARY EVERY PROVIDER SHARES

Each documents how to become eligible and how to opt out. OpenAI adds, "Placement is not guaranteed." None documents how it chooses what to cite from what it retrieved; OpenAI notes only that "the number of sources is often greater than the number of citations." Crawler access is necessary. It is not citation. UNKNOWN

Controls, and what they do not control

Google treats its AI features as part of Search, so the levers are the ordinary ones: nosnippet prevents a page's use "as a direct input for AI Overviews and AI Mode," max-snippet limits it, noindex removes it. Since August 2026 a property-level "Search generative AI control" in Search Console excludes a whole site from AI Overviews and AI Mode without touching the rest of Search. DOCUMENTED

The most misread token is Google-Extended. It governs Gemini training and "does not impact a site's inclusion in Google Search nor is it used as a ranking signal." Blocking it does not remove you from AI Overviews.

PRACTICAL · READ YOUR ROBOTS.TXT AND YOUR CDN'S BOT PRESETS

Many "block AI" presets block the search crawlers along with the training crawlers, which removes you from ChatGPT and Claude search answers without your knowing. Decide the two questions separately: allow the search bots; make your own call on training. OpenAI says a change takes roughly a day to register.

What the tools show, and what they cannot

Both engines give site owners free reporting, and both attach documented limits to it. Most Search Console reports "only cover a representative sample of URLs," tables stop at 1,000 rows, and the counts will not match your analytics for reasons Google lists. Read the reports for direction, not for decimals.

REPORT	USE IT FOR	DOCUMENTED LIMIT
Search Console · Performance	Clicks, impressions and queries by page. Traffic from AI Overviews and AI Mode is included under the "Web" search type; every link inside an AI Overview shares one position.	Sampled; 1,000-row cap; Search Labs experiments excluded; no filter isolates AI clicks.
Search Console · Generative AI performance	Impressions of your links inside AI Overviews and AI Mode by page, country, date and device. Worldwide since 31 August 2026.	No query dimension, no clicks. Does not say why a page was or was not chosen.
Search Console · Page indexing, URL Inspection	Which stage a URL failed (page 3). The live test shows what Googlebot received, including whether a noindex or canonical is actually visible to it.	"URL is on Google" "doesn't actually guarantee that your page will appear." Request-indexing is rationed daily and "does not guarantee" indexing.
Search Console · Manual actions, Security issues	The only first-party record of enforcement or a hack.	An empty report rules out one cause. It does not rule out an algorithmic re-evaluation, which has no report.
Bing Webmaster Tools · AI Performance BING	Citations of your pages in Copilot and Bing's AI summaries; the "grounding queries" the AI used; citation share against competitors.	"Reflects how often pages are cited, not page importance, ranking, or placement." A sample. Preview since February 2026.
Bing Webmaster Tools · URL Inspection, IndexNow	Bing's index status and crawl problems; instant notification of changes.	IndexNow "does not guarantee immediate indexing."
Your own analytics	ChatGPT appends <code>utm_source=chatgpt.com</code> to the links it sends; build a segment on it.	Tagging from other providers is not documented.

When traffic drops

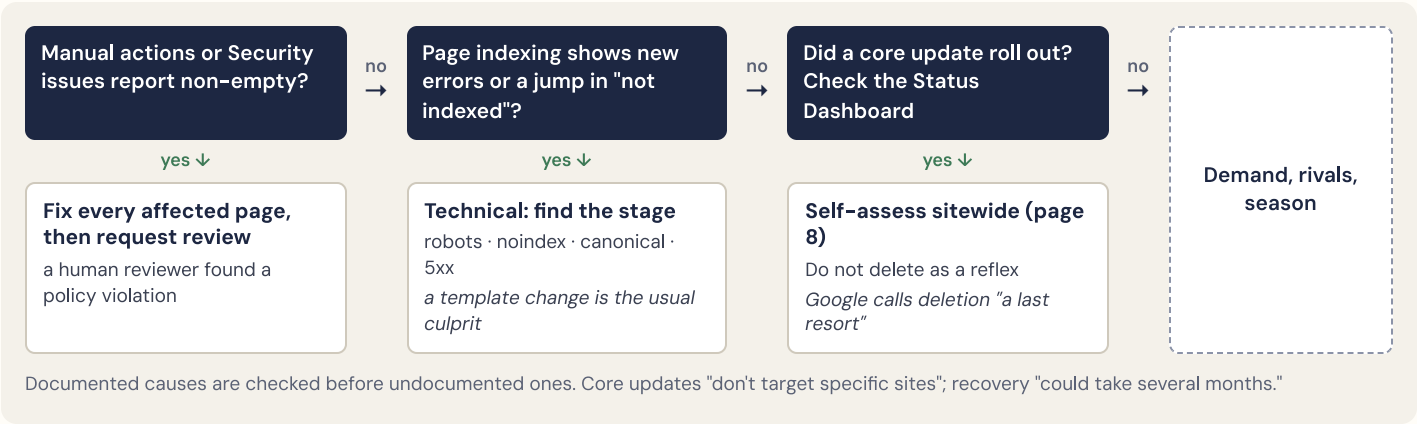


Figure 8. Google's own counsel after a core update is conservative: for small movements, "there's no need to take drastic action (in fact, we recommend avoiding making changes to content that's already performing well)."

What nobody outside the engines knows, and the claims hosts hear anyway

Google explains the silence itself: it must "be careful not to reveal too much detail that would allow people to game our search results." It also ships thousands of ranking changes a year (4,781 launches in 2023), so a published formula would be stale by the time it was typeset. When a consultant quotes a number for any item at right, they are quoting a hypothesis. Patents and leaked documents do not close the gap; a patent proves a technique was patented, not that anyone runs it.

UNKNOWN

The weight of any Google signal

UNKNOWN

The value of any link, or any "authority" score

UNKNOWN

Any threshold at which a classifier changes treatment

UNKNOWN

What a given core update changed, or why a site moved

UNKNOWN

Which interaction metrics are used, at what aggregation

UNKNOWN

For every AI provider: how a retrieved page becomes a cited one

Tear-out: the claim, the verdict, the source

CLAIM	VERDICT	WHAT THE PRIMARY SOURCE ACTUALLY SAYS
Google has 200 ranking factors	NOT SUPPORTED	"hundreds of factors"; no count, no list, weights vary by query
E-E-A-T is a ranking factor	CONTRADICTED	"No, it's not." A "mix of factors" can identify content with good E-E-A-T
Author bios raise rankings	NOT SUPPORTED	Accurate bylines "strongly encouraged"; no ranking effect stated
Titles must be 60 characters; descriptions 160	CONTRADICTED	"no limit"; truncated to fit the device width
The meta keywords tag helps	CONTRADICTED	"no effect on indexing and ranking at all"
Longer pages rank better	CONTRADICTED	"The length of the content alone doesn't matter"
Schema markup boosts rankings	NOT SUPPORTED	Enables rich-result eligibility; a structured-data manual action "doesn't affect how the page ranks"
FAQ schema earns an FAQ rich result	CONTRADICTED	Feature retired May 2026; documentation removed
Duplicate content is penalized	CONTRADICTED	"normal... not a violation." Scraping other sites is
Changing the date makes a page fresh	NOT SUPPORTED	Listed as a search-engine-first warning sign
AI-written content is penalized	CONTEXT	Purpose and value are the test, "no matter how it's created"
A perfect PageSpeed score guarantees rank	CONTRADICTED	"may not be the best use of your time"
Bounce rate or Analytics affect rank	NOT SUPPORTED	No source names them; Google's interaction data concerns its own results page
Buying Google Ads lifts organic results	CONTRADICTED	"no impact on your organic search ranking"
Submitting a URL guarantees indexing	CONTRADICTED	"does not guarantee"; requests are rationed daily
llms.txt is required for AI visibility	CONTRADICTED	Google "ignores them"; no provider documents it as a factor
Blocking GPTBot removes you from ChatGPT search	CONTRADICTED	Separate crawler; "each setting is independent"
Replying to reviews lifts local ranking	NOT SUPPORTED	Review count and ratings "can help"; replies help you "stand out"
Each rental can have a Business Profile	CONTRADICTED	"vacation homes" listed as ineligible

"Not supported" means no reviewed first-party document establishes the claim. That is not proof it is false; only a claim a source directly contradicts is marked contradicted.

Thirty days, in the order the pipeline demands

1

Find the broken stage

- Verify the site in Google Search Console and Bing Webmaster Tools
- Read the Page indexing report; inspect the home page and one house
- Check robots.txt and the CDN's bot rules for Googlebot, bingbot, OAI-SearchBot, Claude-SearchBot
- Confirm every property card is a real link

2

Fix the URL space

- One canonical URL per house; date and guest parameters point back to it
- Calendar, search-results and no-availability pages carry noindex, not a robots block
- Invented URLs return a real 404
- HTTP and the spare hostname redirect once to HTTPS

3

Make the pages worth keeping

- Rewrite each house in your own words with the first-hand specifics
- One descriptive title per page carrying the property name; a heading that agrees; a written description
- Photographs as with alt text
- A host page with your name and contact
- Remove city lists, doorway pages, stuffed footers

4

Reputation, then patience

- Ask every recent guest for a review; no incentives, no gating
- Request one or two honest local links
- Add WebSite and Organization markup; remove self-review and FAQ markup
- Build a `utm_source=chatgpt.com` segment
- Judge in a few weeks, not a few days

SEO can be a helpful activity when it is applied to people-first content, rather than search engine-first content.

GOOGLE · CREATING HELPFUL, RELIABLE, PEOPLE-FIRST CONTENT

SOURCES · EVERY CLAIM ABOVE TRACES TO ONE OF THESE FIRST-PARTY DOCUMENTS, ACCESSED 28 SEPTEMBER 2026

Google Search Central. SEO Starter Guide · In-depth guide to how Google Search works · Google Search technical requirements · A guide to Google Search ranking systems · Creating helpful, reliable, people-first content · Spam policies for Google web search · AI features and your website · Optimizing your website for generative AI features on Google Search · Guidance on using generative AI content · Influencing your title links · Control your snippets · SEO link best practices · Qualify your outbound links · What is URL canonicalization; How to specify a canonical URL · Block Search indexing with noindex · Introduction to robots.txt · What is a sitemap; Build and submit a sitemap · Understanding page experience · Google image SEO best practices · General structured data guidelines · Local business, Organization, Vacation rental and Review snippet structured data · Influence your byline dates · Core updates and your website · Latest documentation updates.

Google Crawling Infrastructure. Google's common crawlers (Google-Extended) · How HTTP status codes affect Google's crawlers · Google's user-triggered fetchers.

Google Search Console Help. Page indexing report · URL Inspection tool · Generative AI performance report · Search generative AI control · What are impressions, position, and clicks? · About Search Console data.

Google Business Profile and Maps Help. Tips to improve your local ranking on Google · Overview of Google Business Profile policies · Guidelines for representing your business on Google · Prohibited and restricted content.

Google. How Search Works: Ranking results; Detecting spam; Rigorous testing.

Microsoft and Bing. How Bing delivers search results (Microsoft Support, March 2025) · Introducing AI Performance in Bing Webmaster

Tools (February 2026) · New AI Visibility Insights (June 2026) · Does Duplicate Content Hurt SEO and AI Search Visibility? (December 2025) · Keeping Content Discoverable with Sitemaps in AI Powered Search (July 2025) · Announcing new options for webmasters to control usage of their content in Bing Chat (September 2023).

AI providers. OpenAI, Overview of OpenAI Crawlers; ChatGPT search help · Anthropic, Does Anthropic crawl data from the web · Perplexity, Perplexity Crawlers.

Standards. WHATWG HTML Living Standard · Schema.org · IndexNow documentation · web.dev, Web Vitals.

Documentation changes without notice. Before acting on a specific requirement, open the cited page and confirm its current wording.