

# Being found by buyers who research through AI assistants

When a buyer asks an assistant who can help, it names your business accurately, names a competitor, or says nothing. This report sets out what an assistant can read about a business, why businesses drop out of answers, and what a marketing or digital lead can check in an afternoon.

|              |                                                                                                                                                                                                         |
|--------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Author       | Amit Govil, Founder and CEO, XONIK                                                                                                                                                                      |
| Published    | October 22, 2026                                                                                                                                                                                        |
| Reading time | 30 min read                                                                                                                                                                                             |
| Online       | <a href="https://xonik.com/insights/reports/being-found-by-buyers-who-research-through-ai-assistants/">https://xonik.com/insights/reports/being-found-by-buyers-who-research-through-ai-assistants/</a> |

## IN BRIEF

An AI assistant names your business only if it can fetch your pages, read them as plain text and tell you apart from other businesses. The providers document access and eligibility, but not how one eligible page is chosen over another. This report sets out the mechanisms, the fixes in order and an afternoon self-check.

## KEY FINDINGS

### 1. A business can drop out of answers without anyone noticing

The buyer never visits, never fills in a form and never appears in a lost-deal note. Every usual measure of demand starts with a contact, so each reports what it reported before and the shortfall is put down to the market or to timing.

### 2. Access is set in two places, the robots file and the network in front of the site, and a refusal is silent

No provider tells a site owner that a crawler was refused. Search and training are separate decisions, and a block on a search crawler carries a cost the provider names, so each should be decided on purpose.

### 3. Reachable pages go unused when no sentence on them is worth repeating

In our experience, brand language, service names with no statement of what each does, claims with no source and substance kept behind a form give an assistant nothing to quote. We infer that a page that says what the business does, for whom and why is easier to use.

### 4. An assistant has to decide that many mentions are one business

A short name, a registered name, product names and old directory entries are not obviously one entity. Consistent naming, organization markup and corrected outside listings remove the ambiguity. Markup is a way to say who the business is, not a lever for being chosen.

5. The providers document who can be used, not who is chosen

Eligibility is not selection. No provider states a rule for choosing between two eligible pages, and Google says no special markup or AI text file is needed for its features. Anyone who gives a formula is working beyond the documents.

6. Technical fixes get done, and the rest waits unless one named person owns the question

Technical fixes arrive as tickets with an owner, while the work on words and agreement stalls because its owner is marketing or the website, which means nobody. Results lag, so a single test taken before the engines have refetched anything proves little.

## Why does it matter that buyers ask an assistant before they ask a supplier?

A buyer who needs a supplier can now put the whole question to an assistant, in full sentences, and read a composed answer that names businesses and links to the pages it drew on. Google describes the same shift inside its own results: its AI Overviews and AI Mode produce a generated response with links to supporting websites. For any one business the answer takes one of three forms. It names the business accurately. It names a competitor. Or it says nothing about the business at all. Exhibit 1 shows where in the buyer's route that third outcome is decided.

[Google Search Central: AI features and your website \(updated 10 December 2025\)](#)

EXHIBIT 1

### Where a buyer's route loses a business that an assistant cannot use



The loss is decided inside the assistant, before the buyer visits any site, so nothing later in the route records it.

## What does the evidence say about buyers and the answer?

The picture matters for where the loss happens. The assistant reads what it can reach, composes an answer and names suppliers. A business that is not in the answer is not in the shortlist the buyer forms from it, so the buyer never visits, never fills in a form and never appears in a lost-deal note. Every usual measure of demand starts with a contact, so each one reports what it reported before, and the shortfall is put down to the

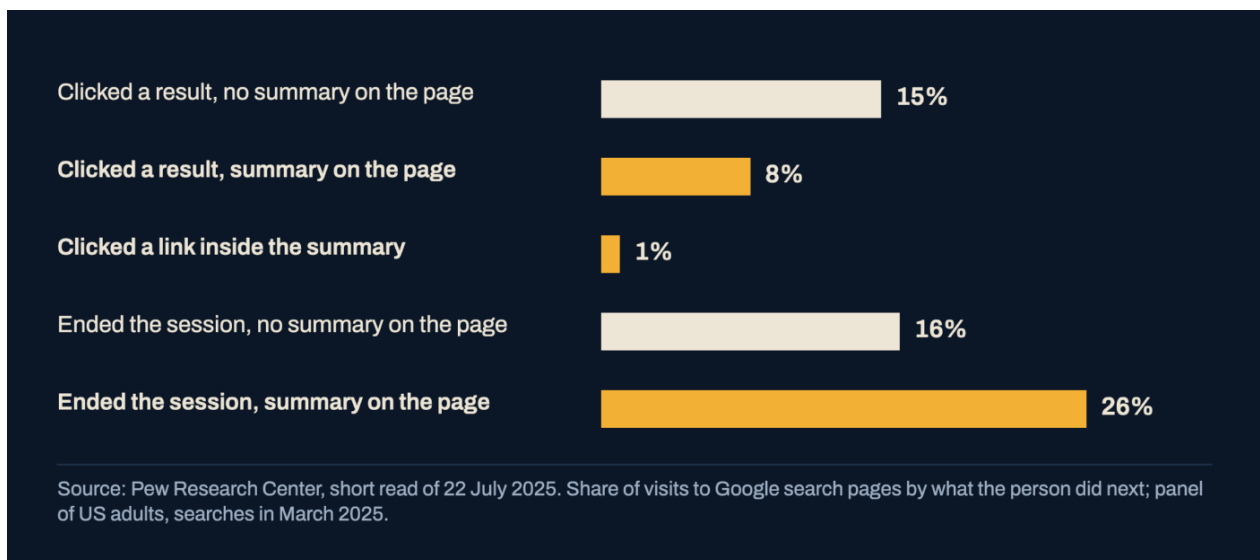
market or to timing.

One measured study shows why the answer itself carries weight. Pew Research Center followed the Google searches of a panel of US adults who agreed to share their browsing activity, for the month of March 2025, and recorded what each person did after a search page loaded. It compared visits to pages that showed an AI-generated summary with visits to pages that did not. Exhibit 2 shows the actions the study reports. Fewer visits ended in a click on an ordinary result when a summary appeared, more of them ended the browsing session, and clicks on the links inside the summary were rare.

[Pew Research Center: Google users are less likely to click on links when an AI summary appears in the results \(22 July 2025\)](#)

#### EXHIBIT 2

#### What people did after a Google search page, with and without an AI summary



When a summary appeared, fewer visits ended in a click on an ordinary result, the links inside the summary drew very few clicks, and more visits ended the session.

## What do we make of it, and who owns the problem?

What follows is our reading, not a statement of the study. When the answer sits on the page and fewer visits continue to a website, attention goes to the businesses the answer names. The study covers Google only, a panel of adults in one country and one month. It does not measure business buyers or any assistant other than Google's. It does report that longer searches and searches phrased as questions produced a summary more often, and a buyer weighing suppliers tends to write that way, which is another inference of ours.

A leader should care now because the failure sits between teams. Marketing writes the pages. The web team owns the robots file and the network rules. Legal or finance owns the registered name. Communications owns the profiles and listings. Nobody owns the question "does an assistant name us correctly?", so every part can be sound while the whole fails. This report takes that question apart by mechanism, keeps what the providers document apart from what we infer, and ends with a self-check a marketing or digital lead can run in an afternoon.

## Method and limits

This report rests on three things. The first is public documents, each read on 20 October 2026 for the passage relied on: the crawler and search documentation of Google, OpenAI, Anthropic and Perplexity, the Robots Exclusion Protocol standard, the llms.txt proposal, the Schema.org vocabulary and one study by Pew Research Center. The second is our delivery experience, described as pattern and mechanism, never as a count or a case, since every engagement is under NDA. The third is one charted snapshot, read from the Pew study and stored with the sentence each figure came from.

Where a statement rests on a document, the text names it. Where it rests on experience or on reasoning from the mechanism, the text says so. We tested no assistant for this report and measured no rate of appearance.

- The providers change their products and documents without notice, and two of the pages read show no date. Check the sources before acting on any line of the crawler table.
- A document states what a provider intends, not what a particular answer did on a particular day.
- Eligibility is not selection. A business that meets every documented condition may still go unnamed, and the providers do not say how one eligible page is chosen over another.
- The Pew study covers Google, adults in one country and one month. It does not say how many business buyers use assistants, and neither does this report.
- Our delivery experience is a source of patterns to test. It is not a measurement, and it cannot say how often any pattern occurs.

Nobody owns the question "does an assistant name us correctly?", so every part can be sound while the whole fails.

## What can an assistant read about a business, and what can it not?

An assistant has three ways to know something about a business, and the providers' own pages describe them as separate jobs. The first is training. A training crawler collects content that may contribute to future models: OpenAI names GPTBot for this and Anthropic names ClaudeBot. Google's Google-Extended token governs use of crawled content for training Gemini models and, in Google's words, for grounding in Gemini Apps and in Grounding with Google Search on Vertex AI. The second is search. A search crawler builds an index ahead of time: Googlebot for Google Search, OAI-SearchBot for ChatGPT search, Claude-SearchBot for Anthropic and PerplexityBot for Perplexity. The third is a fetch at the moment of the question. ChatGPT-User, Claude-User and Perplexity-User each visit a page when a person asks something.

The distinction decides what a business can change. None of the documents read describes a way to read back what a model learned about a business, so we infer that an answer composed from memory can be out of date. What a business can act on is what an assistant finds and fetches when the question is asked, which is why this report concerns pages, rules and names.

Exhibit 3 sets out what an assistant reads as five layers: access, pages, structure, identity and other sources. Each depends on the one above it. A page perfectly written but refused at the door contributes nothing, and a consistent name that no page states plainly gives the assistant nothing to repeat.

[Google Search Central: Google's common crawlers \(updated 15 October 2026\)](#)

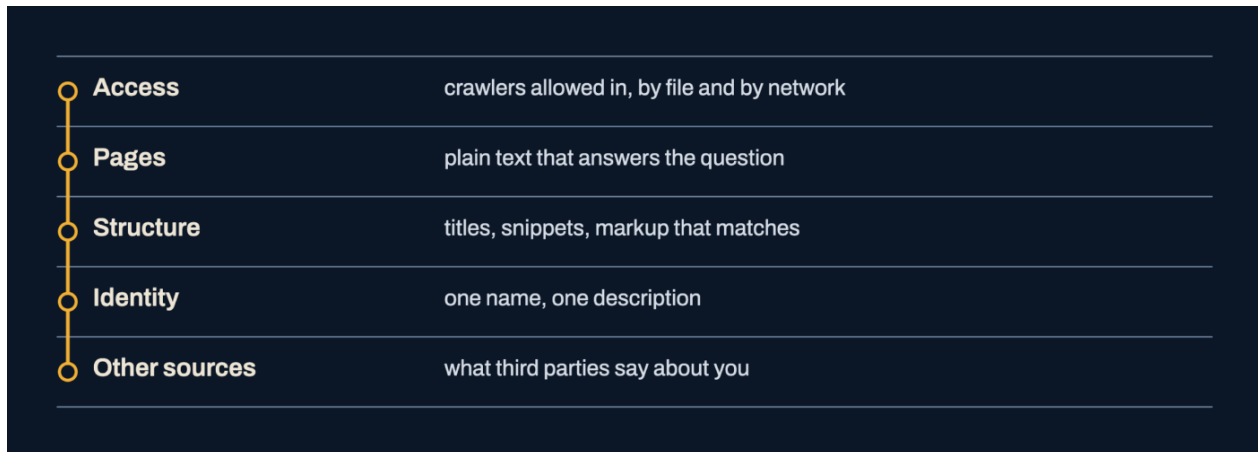
[OpenAI: Overview of OpenAI crawlers \(no date shown\)](#)

[Anthropic: Does Anthropic crawl data from the web, and how can site owners block the crawler \(updated 7 April 2026\)](#)

[Perplexity: Perplexity crawlers \(no date shown\)](#)

#### EXHIBIT 3

### What an assistant reads about a business, in five layers



Each layer is read only if the one above lets it through, so the work runs from the top down.

### What does an assistant fail to read?

Some things an assistant cannot read follow from how the web works. Google's guidance for AI features lists, among the basics, that important content should be available in textual form and that pages should be findable through internal links. Its JavaScript guidance says some sites ship an empty shell and need a script to run before the content exists, that Google renders such pages in a later step, and that server-side or pre-rendering remains a good idea because not all bots can run JavaScript. The OpenAI, Anthropic and Perplexity pages read for this report do not say whether their crawlers run scripts, so a business should not assume they do.

Pages behind a login or a form are a second case. Google's JavaScript guidance says a page behind a login should return a status that tells the crawler it cannot be crawled or indexed, and Anthropic states that its bots respect anti-circumvention technology, giving the example that it will not try to get past a CAPTCHA. Text inside an image is a third. In our experience these share one cause: the substance of the business sits in a place designed for a person with a browser, and the assistant meets only the frame around it.

[Google Search Central: AI features and your website \(updated 10 December 2025\)](#)

[Google Search Central: Understand JavaScript SEO basics \(updated 4 March 2026\)](#)

[Anthropic: Does Anthropic crawl data from the web, and how can site owners block the crawler \(updated 7 April 2026\)](#)

## Which crawlers can reach your pages, and who sets the rules?

Access is governed by two separate mechanisms that firms often treat as one. The first is the robots file. The Robots Exclusion Protocol, published by the IETF as RFC 9309 in September 2022, lets a site owner state which paths automatic clients may fetch, and says plainly that these rules are not a form of access authorization. It works only on a crawler that chooses to honor it. Google states that its common crawlers always obey robots.txt rules when crawling automatically, and Anthropic states that its bots honor the directives. The second mechanism is the network in front of the site: a content delivery network, a firewall or a bot-management rule. Google's own list of basics includes allowing crawling both in the robots file and "by any CDN or hosting infrastructure". A rule at that layer refuses a crawler whatever the robots file says.

Exhibit 4 sets out the crawlers the four providers document, and what each provider says each one is for. The wording follows the providers' pages, and the table records only what they state.

[IETF: RFC 9309, Robots Exclusion Protocol \(September 2022\)](#)

[Google Search Central: Google's common crawlers \(updated 15 October 2026\)](#)

[Google Search Central: AI features and your website \(updated 10 December 2025\)](#)

[Anthropic: Does Anthropic crawl data from the web, and how can site owners block the crawler \(updated 7 April 2026\)](#)

### EXHIBIT 4

#### The crawlers the four providers document, and what each says it is for

| Crawler          | Operator and job            | What the operator's page says                                                                                                                                                                                             | Controls the page describes                                                                                      |
|------------------|-----------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------|
| Googlebot        | Google; crawling for Search | Crawl preferences for it affect Search and all Search features. A page can be linked from AI Overviews or AI Mode if it is indexed and eligible to be shown with a snippet                                                | robots.txt token Googlebot. The nosnippet, data-nosnippet, max-snippet and noindex controls limit what is shown  |
| Google-Ext ended | Google; a control token     | Manages whether crawled content may be used to train future Gemini models and for grounding in Gemini Apps and Grounding with Google Search on Vertex AI. Does not affect inclusion in Search and is not a ranking signal | robots.txt token. It has no user agent string of its own                                                         |
| OAI-Search Bot   | OpenAI; search              | Surfaces websites in ChatGPT's search features. A site opted out is not shown in ChatGPT search answers but can still appear as a navigational link                                                                       | robots.txt. OpenAI recommends also allowing its published address ranges. A change can take about a day to apply |
| GPTBot           | OpenAI; training            | Crawls content that may be used in training generative AI foundation models                                                                                                                                               | robots.txt. Disallowing it indicates content should not be used for training. Independent of OAI-SearchBot       |
| ChatGPT-U ser    | OpenAI; user actions        | May visit a page when a user asks ChatGPT or a custom GPT a question. Not used to crawl automatically and not used to decide whether content appears in Search                                                            | Because a user starts the visit, robots.txt rules may not apply                                                  |

|                  |                           |                                                                                                                 |                                                                                                               |
|------------------|---------------------------|-----------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------|
| Claude-SearchBot | Anthropic; search quality | Navigates the web to improve the quality of search results                                                      | Disabling it prevents indexing for search and may reduce visibility and accuracy in results                   |
| ClaudeBot        | Anthropic; training       | Collects web content that could contribute to model training                                                    | Restricting it signals that future material should be excluded from training datasets                         |
| Claude-User      | Anthropic; user questions | May access websites when a person asks Claude a question                                                        | Disabling it prevents retrieval in response to a user query and may reduce visibility in user-directed search |
| PerplexityBot    | Perplexity; search        | Designed to surface and link websites in Perplexity results. Not used to crawl content for AI foundation models | robots.txt. Perplexity recommends allowing it and permitting its published address ranges                     |
| Perplexity-User  | Perplexity; user actions  | May visit a page to answer a question and include a link to it. Not used for crawling or for training           | Started by a user, so it generally ignores robots.txt rules                                                   |

Each provider separates search, training and user-triggered fetches, and says something different about whether robots rules apply to the last. Wording follows each operator's own page, read on 20 October 2026.

## How do the crawler rules differ, and what does a block cost?

Three things in the table are easy to miss. First, each provider separates the jobs, so a business can allow search and refuse training. OpenAI says each of its settings is independent and gives that very example, and adds that when both its search and training crawlers are permitted it may draw on a single crawl for both purposes, so a business that permits both should not suppose it has agreed to two visits. Google's position differs in kind. Google-Extended is a token with no user agent string of its own, it covers training and grounding for Gemini, and Google says it does not affect inclusion in Search. Inclusion in Search is left to the Googlebot rules, which Google's AI features page names as the control.

Second, user-initiated fetches follow different rules. OpenAI says that because a person starts a ChatGPT-User visit, robots rules may not apply, and Perplexity says its user fetcher generally ignores them. Anthropic says its bots honor robots directives. So the robots file governs the automatic crawlers, and the network layer decides whether a user-triggered visit is welcome. Perplexity's page gives firewall notes, saying a site behind a web application firewall may need to allow its bots explicitly, and OpenAI recommends allowing its published address ranges as well as the robots entry.

Third, a block costs something the provider names. OpenAI says a site that opts out of its search crawler will not be shown in ChatGPT search answers, though it can still appear as a navigational link. Anthropic says that disabling Claude-User or Claude-SearchBot may reduce a site's visibility in user-directed search and its accuracy in results, and that blocking by address is not a reliable way to opt out, because it stops the crawler reading the robots file at all. These statements are the reason to treat the training question and the search question as two decisions.

[OpenAI: Overview of OpenAI crawlers \(no date shown\)](#)

[Anthropic: Does Anthropic crawl data from the web, and how can site owners block the crawler \(updated 7 April 2026\)](#)

[Perplexity: Perplexity crawlers \(no date shown\)](#)

[Google Search Central: Google's common crawlers \(updated 15 October 2026\)](#)

## How does access go wrong, and why does nobody notice?

How access goes wrong is mundane, and in our experience it is rarely a decision anyone remembers making. Four patterns recur. A robots file written when fewer crawlers mattered allows those by name and refuses everything else, and the standard explains why that now works against the business: when no group matches a crawler's name, the crawler follows the group for all agents, so a new retrieval crawler inherits the refusal. A template or plugin adds a blanket refusal of AI crawlers when the intent was to refuse training only. A bot-management rule on the content delivery network, switched on to stop scraping, challenges every unfamiliar agent. And a staging setting that disallowed everything is carried into production at launch.

The standard is also unforgiving about errors. If the robots file cannot be reached because of a server or network error, the crawler must assume complete disallow; if the server says the file does not exist, the crawler may fetch anything. A robots file that errors during a deployment, or on one host name of several, reads as a closed door. The file also applies per host, and Anthropic tells owners to repeat any instruction for every subdomain.

The cost is silence. No provider tells the site owner a crawler was refused, and the buyer sees only an answer without the business in it. The delay hides the cause as well. OpenAI and Perplexity each say a change to the robots file can take about a day to be reflected, and Google says that after a change to preview controls a recrawl can take anywhere from several days to several months. A business that fixes a block on Monday and tests on Tuesday may conclude that nothing changed.

[IETF: RFC 9309, Robots Exclusion Protocol \(September 2022\)](#)

[Google Search Central: AI features and your website \(updated 10 December 2025\)](#)

[OpenAI: Overview of OpenAI crawlers \(no date shown\)](#)

[Perplexity: Perplexity crawlers \(no date shown\)](#)

[Anthropic: Does Anthropic crawl data from the web, and how can site owners block the crawler \(updated 7 April 2026\)](#)

## How do the engines choose and cite their sources?

The providers say more about who can be used than about who is chosen. Google states that for a page to be linked from AI Overviews or AI Mode, the page needs to be in Google's index and able to appear with a snippet, that there are no additional technical requirements, and that no special markup, machine-readable file or AI text file is needed. It adds that meeting every requirement does not make crawling, indexing or serving certain. OpenAI says OAI-SearchBot surfaces websites in ChatGPT's search features. Anthropic says Claude-SearchBot analyzes content to improve the relevance and accuracy of search responses. Perplexity describes PerplexityBot as built to surface and link sites in its results, and Perplexity-User as a visit made to answer a question, with a link to the page in the response.

Google also describes how answers are built. AI Overviews and AI Mode may use what it calls query fan-out: issuing several related searches across subtopics and data sources, and finding more supporting pages while the response is generated. It says this lets it show a wider and more diverse set of links than a classic search, that it shows an AI Overview only where its systems judge it an addition to classic results, and that the two features may use different models, so their responses and links vary. Sites appearing in these

features are counted in the Web search type of the Performance report in Search Console, so a Google-side signal already exists in a tool a marketing lead can open.

None of the documents states a rule for choosing between two eligible pages, the number of sources used, or the weight given to a page against a third party's description of the business. Anyone who gives a formula for those is working beyond the documents. We can reason from the mechanism, and we mark it as inference. A retrieval step has to find passages that match a question. A page whose opening lines state the answer in plain words offers a passage that stands on its own, for the same reason it would in an ordinary search. A page that says nothing a reader could quote offers nothing, however well the site ranks.

Google's helpful-content guidance is the nearest published statement of what its systems are built to reward, and it is written for Search, not for the other providers. It asks whether content provides original information, reporting, research or analysis, and it asks who created a page, how, and why. We take that as a statement of what is worth writing, and we do not claim the other assistants follow it.

The /llms.txt proposal belongs here because it is the best-known attempt to tell assistants what to read. Version 2 proposes a markdown file at /llms.txt with a name, a short summary and links to pages written for machines, and suggests markdown copies of pages. It says the file is used on demand, when an agent needs information, and mainly for inference, not training. It is a proposal by one author. Google states that no AI text file is needed for its features, and the OpenAI, Anthropic and Perplexity crawler pages read for this report do not describe it as an input. If the file is cheap to publish and kept in step with the site it does no harm, but it does not replace pages a person can read, and we do not count its absence as a cause of disappearance.

[Google Search Central: AI features and your website \(updated 10 December 2025\)](#)

[Google Search Central: Creating helpful, reliable, people-first content \(updated 5 October 2026\)](#)

[Jeremy Howard: The /llms.txt file, version 2 \(published 3 September 2024, modified 10 August 2026\)](#)

[OpenAI: Overview of OpenAI crawlers \(no date shown\)](#)

[Anthropic: Does Anthropic crawl data from the web, and how can site owners block the crawler \(updated 7 April 2026\)](#)

[Perplexity: Perplexity crawlers \(no date shown\)](#)

## Why do some reachable pages still go unused?

A reason we meet often when a business is reachable and still not named is that its pages contain no sentence worth repeating. An assistant answering "who helps a clinic group replace its booking system?" looks for text that connects a kind of business, a problem and a capability. Pages written in the language of a brand campaign, with a mission and a list of values, connect none of them. Pages that list service names with no statement of what each does, for whom, or what problem it removes give the assistant a noun and nothing to say about it. We infer that a competitor whose page says what it builds, for whom, and why, is easier to use for the same reason it is easier for a person to understand.

A second pattern is the claim with no source. A page says the firm is trusted, leading or experienced, and carries no method, no named document and no evidence a reader could follow. A person discounts such a claim, and a retrieval system has the same problem in another form: nothing in the page can be checked or attributed, so it supplies no support for a sentence an answer might want to make. Google's helpful-content guidance, which asks about original information, clear sourcing and background on the author, points the

same way. Evidence beside the claim turns a statement into something that can be cited.

A third is the page that answers no question. Sites grow by department and by campaign, and the result is pages organized around the firm's structure instead of the buyer's problem. A buyer asks what a failed data migration costs to repair, and the site has a page called Solutions and a page called Team. A page can also fail when its title and description say something other than its content: Google says it builds snippets mainly from the page content and uses the description element only when that describes the page better. Eligibility for Google's AI features is tied to being shown with a snippet, so the text that produces the snippet is not a minor detail.

The fourth is content behind scripts and forms. The usual version in firms that sell expertise is the gated report: the substance sits behind a form, and the public page carries a title, a button and a line of promotion. A person fills in the form. A crawler cannot, so the assistant sees that a document exists and nothing of what it says. The fix is not to ungate everything. It is a public page that states what the document concludes and how it was made, with only the full asset gated. Otherwise the firm's best thinking is invisible to the one reader that reads everything.

[Google Search Central: Control your snippets in search results \(updated 20 April 2026\)](#)

[Google Search Central: Creating helpful, reliable, people-first content \(updated 5 October 2026\)](#)

## Does the assistant know which business you are?

Naming is the mechanism people underrate, because a human reader fills gaps without noticing. A firm trades under a short name, is registered under a longer one, calls its services by internal product names, was renamed after an acquisition, and appears in directories under whatever a past employee typed. An assistant composing a sentence about the business has to decide that these mentions are one entity. When they are not obviously one, the results are poor: the business is confused with a namesake, the answer is built only from the mentions that match the question's wording, or the description is taken from a directory and not from the firm.

Two kinds of documented help exist. Google says organization markup on the home page can help it understand a business's administrative details and disambiguate the organization in search results. The Schema.org Organization type carries properties for this: a legal name, an alternate name for an alias, and a sameAs property for the address of a reference page that unambiguously indicates identity, such as an official profile. Google adds that structured data must match the visible text, and that no special schema is required for its AI features. We therefore treat markup as a way to remove ambiguity about who the business is, not as a lever for being chosen.

The second kind is profile data outside the site. Google's AI-features guidance lists keeping Merchant Center and Business Profile information up to date among the basics. For dental and clinical practices, hospitality and retail, found through map listings and review sites as much as through their own pages, an out-of-date listing may be the first thing an assistant finds. For legal and real estate firms the equivalent is the directory and the trade body's member page.

Other people's pages matter too, and the Pew study gives the only published evidence we use. It reports that the sources most often listed in Google's AI summaries, and in ordinary results, were an encyclopedia, a video site and a discussion forum: sources that the businesses in an answer do not control. The study covers Google's results for its panel's searches in a single month and says nothing about how any other

assistant weighs sources. It does show that an answer is not built only from the websites of the businesses it names. In our experience the symptom is an answer that describes a firm in a directory's wording, with a service the firm dropped or never offered, and the fix lies outside the website.

[Google Search Central: Organization structured data \(updated 8 September 2026\)](#)

[Schema.org: Organization type \(development version, read 20 October 2026\)](#)

[Google Search Central: Introduction to structured data markup \(updated 10 December 2025\)](#)

[Google Search Central: AI features and your website \(updated 10 December 2025\)](#)

[Pew Research Center: Google users are less likely to click on links when an AI summary appears in the results \(22 July 2025\)](#)

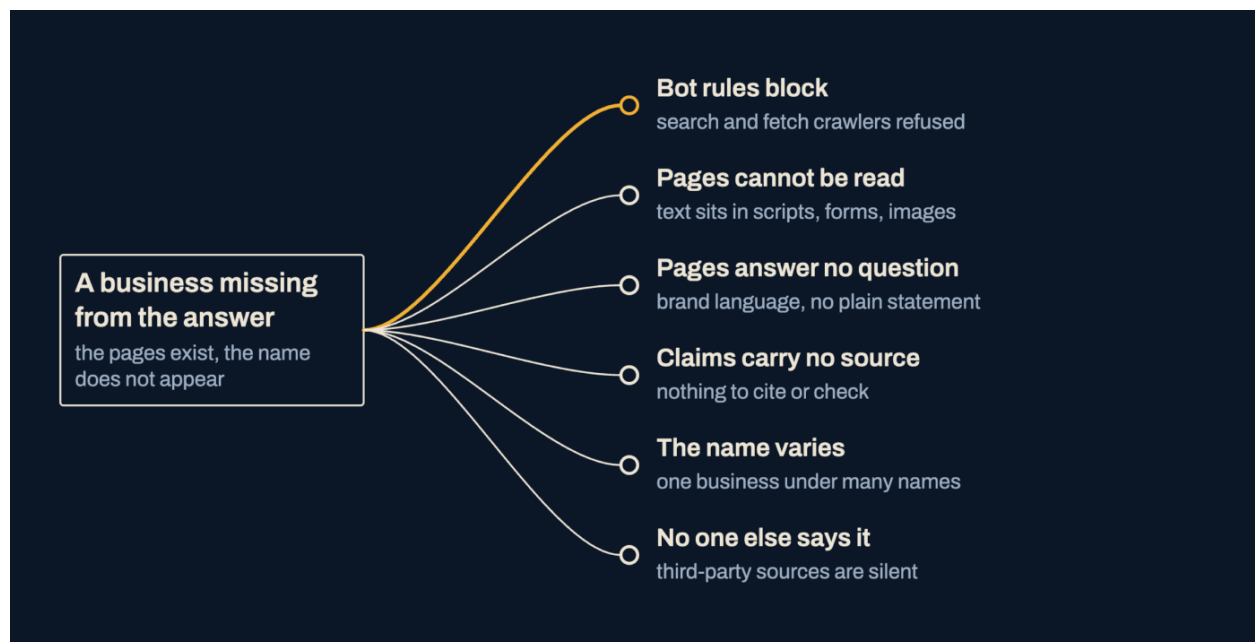
## Why does a business disappear from answers, and how do you find out which reason applies?

Put side by side, the mechanisms give six reasons a business goes missing, set out in Exhibit 5. Bot rules and network rules refuse the crawlers. Pages cannot be read because their text lives in scripts, forms or images. Pages answer no question a buyer asks. Claims carry no source. The name varies, so mentions do not merge. And no one else says the thing the business wants to be known for, so nothing corroborates it. The first is drawn in blue because a firm can settle it with certainty, from documents the providers publish.

We cannot say how likely each reason is, because we hold no count, and the reasons are not independent. A firm refused at the door gives no evidence about its pages, and a firm with sound pages can still be mistaken for a namesake. The order of diagnosis therefore matters more than the list. Exhibit 6 sets it out as three questions in sequence: whether the crawlers can fetch, index and show the pages; whether the pages answer the buyer's question in plain text; and whether the business is named the same way wherever it appears. Only when all three are answered yes does the work move to other sources, and to the next test.

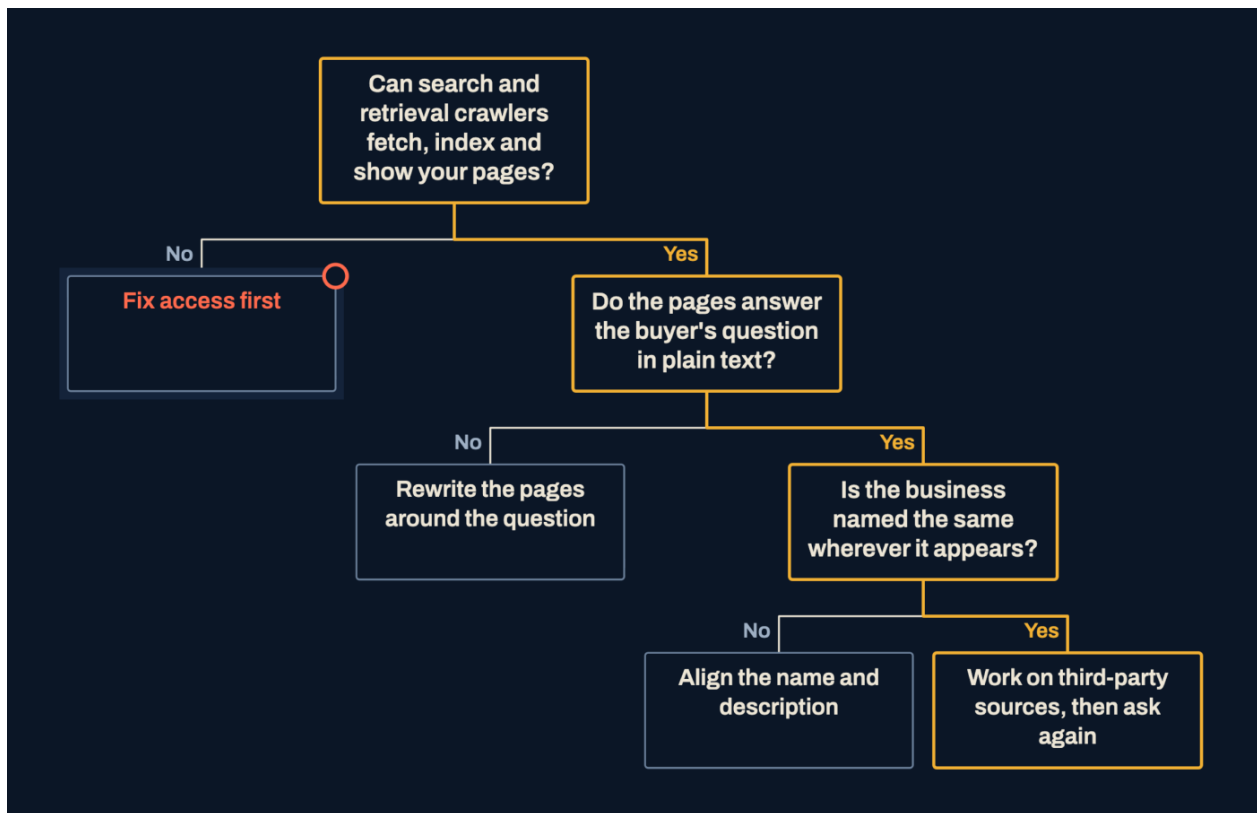
## EXHIBIT 5

## Six reasons a business goes missing from an answer



The first reason can be settled from documents the providers publish. The others need judgment and a test.

## EXHIBIT 6

**Why am I not in the answer? Three questions, in order**

Ask in this order: each question is cheaper and more certain than the one after it.

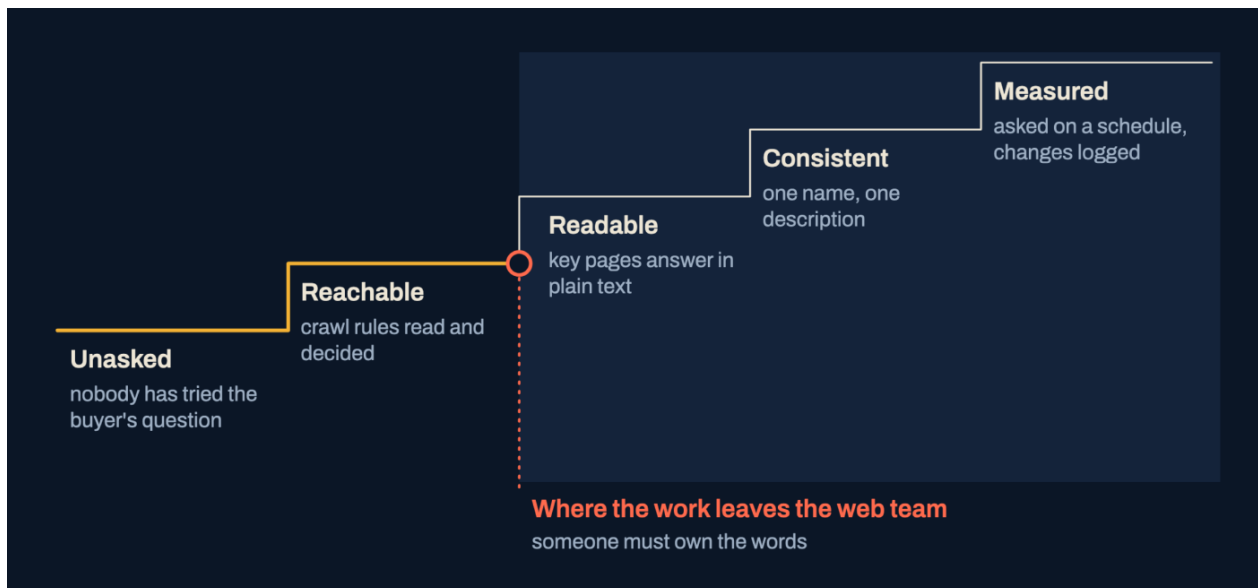
**What changes the outcome?**

Five changes alter the outcome, and they differ in how sure we can be that they work. The first is documented: decide access deliberately, as two decisions, whether search and retrieval crawlers may reach the pages that answer buyers' questions and, separately, whether training crawlers may, and keep the key pages indexable and eligible for a snippet. The second follows from the mechanism: give each page that should be found an opening that states what the business does, for whom and for what problem, in a buyer's words, with the evidence beside the claim. The third is consistency: one name, one short description and one list of services, on the site, in markup, in listings and in profiles. The fourth lies outside the site: the directories, trade bodies and publications that describe the business should say the same thing, and a business can ask for corrections. The fifth is measurement: ask the buyer's questions on a schedule, keep the answers, and change one thing at a time where you can.

Exhibit 7 arranges these as a ladder, and its break sits in the middle. The first two rungs are technical and belong to the web team. From the third the work is words and agreement, and it stalls unless a named person owns it. In our experience the technical fixes get done because they arrive as tickets with an owner, while the rungs above wait because their owner is "marketing" or "the website", which means nobody. The change that makes the rest possible is organizational: one person who owns the question "does an assistant name us correctly?", with standing to ask the web team for a rule change and the communications team for a corrected profile.

## EXHIBIT 7

## A maturity ladder for visibility to assistants



The first rungs are web-team work. From the third rung the work is words and agreement, and it needs a named owner.

## What do these changes not do?

These changes have limits. Fixing access removes a documented reason for absence and does not make a name appear. Clear pages and a consistent identity remove reasons for being skipped or mistaken, but the providers do not document that they cause selection, and no one can promise a citation. Results arrive late, because a recrawl takes time, and a single answer on a single day is a sample of one. Exhibit 8 shows the order of a fix in time and the point at which results lag the work. The error we see at this stage is to call a fix a failure, or a success, from one test taken before the engines have refetched anything.

Nor is there a trick to add. Google states that no special optimization is needed for its AI features beyond the basics of Search, and nothing in the other providers' documents describes a technique that works. The work is access, plain pages, one name and a record of what the assistants say.

[Google Search Central: AI features and your website \(updated 10 December 2025\)](#)

## EXHIBIT 8

## The order of a visibility fix, and where results lag



Results lag the work, so a test taken straight after the changes says little. Ask again only after the engines have refetched.

## WHERE IT GOES WRONG

## A fix judged too soon

The error we see at this stage is to call a fix a failure, or a success, from one test taken before the engines have refetched anything.

## What is the step-by-step response?

## 1. Name an owner and write the questions

The marketing or digital lead takes the question, with a sales lead who knows what buyers actually ask. Together they write the questions in buyers' words: category questions (which firms help a business like this with this problem), comparison questions, questions that name your business, questions for each service and sector you want to be known for, and the question asked after a supplier has failed. Include questions you expect to lose. At the end there is one set of questions, each tied to the page that should answer it.

## 2. Ask, and record the baseline

The owner, or an analyst, puts each question to more than one assistant, in a fresh session without a signed-in history. The log holds the date, the assistant, whether the business is named, how it is described, which pages are linked and who is named instead. At the end there is a dated baseline. It is the only measurement available, and every later step is judged against it.

## 3. Read access as each crawler would

The web or infrastructure lead fetches the robots file on every host name, confirms that it returns a normal page and not an error, and reads each group of rules. For every crawler in Exhibit 4 they record whether it is allowed or blocked and who decided. They then check the content delivery network, the firewall, any bot-management rule and the server log for the crawlers' user agents and published

address ranges. At the end there is a one-page access record: crawler, decision, reason, date.

#### 4. Check eligibility and the plain text of the key pages

The web lead and the content owner take the pages that should answer the questions. In Search Console they inspect each as Google received it, and confirm none carries a noindex or nosnippet instruction nobody remembers adding. They then read each page with scripts off and without signing in: does the substance remain? At the end there is a list of key pages with their indexing, snippet and plain-text status.

#### 5. Rewrite the pages that answer nothing

The content owner and a subject expert revise the pages the list marks. Each opens with a sentence saying what the business does, for whom and for what problem. The home page carries one statement of what the business is. Brand language gives way to plain statement, evidence sits beside each claim, and every gated asset has a public page stating what it concludes. At the end there are revised pages that each contain a sentence worth quoting.

#### 6. Fix the name where you control it, then ask others to fix it

The communications or brand owner, with legal, agrees one name, one short description and one list of services, and applies them to the site, titles, organization markup (legal name, alternate names, and sameAs links to official profiles, all matching the visible text), map and business listings, directories, trade-body pages and the press boilerplate. They then write to the owners of pages that are not theirs and ask for corrections. At the end there is a naming sheet and a set of corrected profiles.

#### 7. Wait, ask again, and keep the change log

The owner waits for the engines to refetch, then repeats step two with the same questions and assistants and compares the answers with the baseline. A change log records what changed, when, and what moved. The check repeats on a schedule and after any change to the robots file, the network rules or a key page, because the products and their documents change. At the end there is a comparison and a note of what to change next.

### What can you check in an afternoon?

- We have opened our robots file on every host name we use, and it returns a normal page and not an error.
- For each crawler in Exhibit 4 we can say whether it is allowed or blocked, who decided, and why.
- Our content delivery network, firewall and bot rules do not refuse the search and retrieval crawlers by user agent or by address range.
- We decided in writing, and separately, whether training crawlers may fetch our content.
- Our key pages are indexed in Google Search and shown with a snippet, and none carries a noindex or nosnippet instruction nobody remembers adding.
- With scripts off and without signing in, each key page still states what we do, for whom and for what problem.

- Each service page opens with a sentence that could be quoted on its own and still be correct.
- Every claim of capability on a key page has its evidence beside it: a method, a document or a named source.
- Anything we gate behind a form has a public page that says what it concludes and how it was made.
- We use one business name, one short description and one list of services on our site, in our markup, in our listings and in our profiles.
- Our map listings, directory entries and trade-body pages show our current name, address and services.
- We hold a written set of buyer questions, in buyers' words, for each service and sector, and have put them to more than one assistant.
- For each answer we recorded whether we were named, how we were described and who was named instead.
- One named person owns the question and the log, and can ask the web and communications teams for changes.
- We repeat the check on a schedule and after any change to the robots file, the network rules or a key page.

## What do the answers tell you to fix?

When the answers are in, read each one for four things: whether the business is named, whether it is described correctly and in its own terms, which pages are linked, and who is named instead. The table maps what you see to the first place to look. It is a set of starting points drawn from the mechanisms above, not a diagnosis, and one answer is one sample.

| What you see                                                                        | What it suggests                                   | What to fix first                                                                                |
|-------------------------------------------------------------------------------------|----------------------------------------------------|--------------------------------------------------------------------------------------------------|
| Not named, and the server log shows no visits from the search or retrieval crawlers | The crawlers are refused or never reach the pages  | The robots file and the network rules (step three)                                               |
| Not named, the crawlers do visit, but key pages are not indexed or show no snippet  | The pages are not eligible                         | Indexing, noindex and nosnippet instructions, internal links (step four)                         |
| A competitor is named and the page linked says plainly what it does and for whom    | Your pages answer the question less directly       | The opening lines of the page, and the evidence beside each claim (step five)                    |
| Named, but described in a directory's or a third party's wording                    | Your own pages are not the easiest source to quote | A plain statement on your own page, then the listing (steps five and six)                        |
| Confused with another business of a similar name                                    | Mentions do not merge into one entity              | Organization markup with sameAs links, and a description that states sector and place (step six) |
| Named by one assistant and not by another                                           | Each product uses its own crawlers and sources     | The access record for that provider's crawlers (step three)                                      |

## What do readers ask about being found through assistants?

### Why does an AI assistant name a competitor and not us?

The report finds several reasons you can check. Crawlers may be refused, pages may not read as plain text or answer no buyer question, claims may carry no source, the name may vary, and nobody else may corroborate you. Providers document eligibility, not selection, so none can promise a mention.

### Should we block AI crawlers on our site?

Treat training and search as separate decisions. Providers let a business allow search and refuse training. Blocking a search crawler carries a cost the provider names, which is less visibility in that assistant's answers. Decide each on purpose and record why.

### Do we need schema markup or an llms.txt file to be named?

No. Google says no special markup, machine-readable file or AI text file is needed for its AI features, and the other providers' pages the report read do not describe an llms.txt file as an input. Organization markup helps say who you are, which is different from being chosen.

### How do we test whether assistants name us?

Write the questions buyers ask in their own words, put each to several assistants, and record the date, whether you are named, how you are described and who is named instead. Repeat on a schedule, because an answer on any given day is only a sample.

## Sources and notes

Each document below was opened and read on 20 October 2026 for the passage cited in the text. Dates are those the publisher shows. Where none is shown, the entry says so.

1. [Google Search Central, AI features and your website \(updated 10 December 2025\)](#)
2. [Google Search Central, Google's common crawlers \(updated 15 October 2026\)](#)
3. [Google Search Central, Introduction to robots.txt \(updated 10 December 2025\)](#)
4. [Google Search Central, Understand JavaScript SEO basics \(updated 4 March 2026\)](#)
5. [Google Search Central, Control your snippets in search results \(updated 20 April 2026\)](#)
6. [Google Search Central, Introduction to structured data markup \(updated 10 December 2025\)](#)
7. [Google Search Central, Organization structured data \(updated 8 September 2026\)](#)
8. [Google Search Central, Creating helpful, reliable, people-first content \(updated 5 October 2026\)](#)
9. [OpenAI, Overview of OpenAI crawlers \(no date shown\)](#)
10. [Anthropic, Does Anthropic crawl data from the web, and how can site owners block the crawler \(updated 7 April 2026\)](#)
11. [Perplexity, Perplexity crawlers \(no date shown\)](#)
12. [IETF, RFC 9309, Robots Exclusion Protocol \(September 2022\)](#)
13. [Jeremy Howard, The /llms.txt file, version 2 \(published 3 September 2024, modified 10 August 2026\)](#)

14. [Schema.org, Organization type \(development version, read 20 October 2026\)](#)
15. [Pew Research Center, Google users are less likely to click on links when an AI summary appears in the results \(22 July 2025\)](#)

**USE THIS**

[Check your visibility with the AI search visibility assessment](#)

[Use the AI search visibility checklist](#)

Start with the cheapest test: open your robots.txt file and read it as each crawler in the table would.

---

This document was made from the article at

<https://xonik.com/insights/reports/being-found-by-buyers-who-research-through-ai-assistants/>. The online article is the current version.