



What happens between the question and the answer

How AI answer engines decide which businesses to name, and what it takes to be one of them.

10 August 2026 Pivotal Engineering

Who this is for

You run a website, or you're on the hook for one. You've heard AI is changing search and you'd like to know what's actually true, in enough detail to make a decision instead of a guess.

This assumes you know what HTML is. It doesn't assume you write it. Every technical term gets explained the first time it shows up, and where the ground is genuinely unsettled, we say so instead of smoothing it over.

Answer engines don't rank your site. They read it and quote it. Those are different jobs. The useful part is that the work making you quotable also makes you rank better, so you aren't trading one for the other. Same plumbing, two payoffs.

01 The thing that changed

For twenty-five years search worked one way. You typed words, you got a list of links, you picked one. The whole discipline of SEO grew up around a single question: how do I get my link higher in that list?

Answer engines broke the list. Someone asks ChatGPT which roofer to call, or Google drops an AI summary above the results, and a machine reads a handful of sources and writes the answer itself. The user reads that. Usually they stop there.

We can measure how often. In the first four months of 2026, 68% of US Google searches ended without a click on anything, up from 60% in 2024.¹ When Google shows an AI summary, people click a traditional result 8% of the time instead of 15%.²

That's the part everyone talks about. What matters more, and gets talked about less, is the couple of seconds before the answer appears.

02 What actually happens when someone asks

An answer engine doesn't pull up a ranked list and read from the top. It runs a pipeline, and your site can fall out at any point along it.

- INTERPRETATION** The engine rewrites the question into one or more searches of its own. "Who can fix a sprinkler line this week" might become three: irrigation repair, emergency service, and a place.
- RETRIEVAL** It gathers candidates, from a search index, its own crawl, a live fetch, or some mix. If it can't reach you, the story ends here.
- GROUNDING** It picks the passages that actually support an answer. Not whole pages. Passages. A page can get retrieved and still contribute nothing, because nothing in it cleanly answers what was asked.
- SYNTHESIS** It writes the answer out of what it selected.
- ATTRIBUTION** Sometimes it names its sources. Sometimes it doesn't.

Four places to fall out, and they fail in different ways. Never crawled means you were never in the running. Retrieved but not grounded means your page got read and found unhelpful, which stings more. Grounded but not attributed means your information is in the answer and your name isn't.

Getting cited is a chain, not a switch. Most of what gets sold as "AI visibility" fixes one link and leaves the rest alone.

03 SEO and GEO: same plumbing, two different readers

This is where the confusion usually lives, so let's be precise about it.

SEO optimizes a document to rank in a list. The unit is a URL and a position. The page competes as a whole.

GEO optimizes a passage to get pulled as evidence. The unit is a claim. A sentence competes on its own, lifted out of the page it lives on, stripped of the layout and the context around it.

They're not opposed. They run on the same substrate: a site a machine can reach, parse, and understand. But they reward different shapes. Classic SEO rewards the comprehensive page, long and authoritative, covering a topic end to end. Retrieval rewards the page that comes apart cleanly into self-contained answers.

A page can rank first and still be a lousy thing to quote, because everything good in it depends on three paragraphs of windup.

Where they come apart

Googlebot runs JavaScript. It fetches your page, executes the scripts, waits for the result, and indexes what shows up.

GPTBot, ClaudeBot, and PerplexityBot don't. They grab the raw HTML, take what's in it, and move on. No waiting, no second pass. GPTBot requested JavaScript files in about 11.5% of its requests and ran none of them. ClaudeBot downloaded JavaScript in roughly 23.8% and never executed any of it.³

If your site builds its content in the browser instead of on the server, Google sees a full page and an answer engine sees an empty one. You can pass every SEO audit you own and still be invisible to the systems answering your customers.

That covers most React, Vue, and Angular sites nobody has configured otherwise. It's the finding that surprises capable technical teams more than any other, and it isn't a content problem or a marketing problem. It's a rendering problem. The fix is server-side rendering or pre-rendering, which is well-worn ground with good tooling.

04 Three layers where sites fail

Nearly every retrieval failure comes down to one of three layers. They stack, and the early ones win: a site that fails at access can't be rescued by anything you do further down.

Access. Can the engine get in?

robots.txt is a file at the root of your site telling automated visitors where they're allowed to go. Most were written before AI crawlers existed, so they either say nothing about them or block them by accident, caught in a blanket rule aimed at something else. (Most robots.txt files were last touched by someone who has since left the company.)

Blocking is a legitimate choice. Publishers sitting on paywalled archives make it deliberately, and they're right to. But it should be a decision, and for a business that wants to get recommended it's usually the wrong one made by accident. The agents worth naming today are GPTBot and OAI-SearchBot, ClaudeBot with its search and user-triggered variants, PerplexityBot, and Google-Extended. That list moves. Treat it as something you check twice a year, not something you set and forget.

Identity. Can the engine tell what you are?

An engine reading your page has to work out that you're a business, of a certain kind, serving a certain area. A human gets that from a logo and a phone number in the header. A machine needs it spelled out.

Structured data is how you spell it out. In practice that means JSON-LD, a block of machine-readable description sitting in the page, written in the shared vocabulary at schema.org. It says plainly: this is an organization, it does these things, it works in these places. Skip it and the engine guesses. Guessing is where your competitor gets named instead, or where you get confused with a similarly named outfit three states over.

Extractability. Can the engine find the answer inside the page?

This is the layer almost nobody has thought about, and the one where good SEO habits work against you.

A recipe card and a story about the night you learned to cook can carry the same instructions. Only one of them survives being torn out and handed to someone else. Retrieval systems tear pages up. They cut documents into chunks and judge the chunks, so a page where the answer to “how fast can you get here in an emergency” is spread across four paragraphs and a testimonial gives them nothing clean to lift.

Semantic HTML earns its keep here, for a reason people usually get backwards. Proper heading levels and real list markup aren't only an accessibility concern, though that's reason enough on its own. They tell a chunking algorithm where to cut.

05 Remediation that pays twice

Here's the part that should decide it for you. Almost every fix that helps retrieval also helps classic search, which means none of this work depends on being right about how fast AI search grows.

THE FIX	FOR CLASSIC SEARCH	FOR ANSWER ENGINES
Open AI crawler access in robots.txt	Neutral. Googlebot is governed separately.	Decides whether you're a candidate at all.
Server-side rendering or pre-rendering	Faster indexing, fewer crawl-budget problems, better Core Web Vitals.	Decides whether your content exists at all to a crawler that won't run JavaScript.
JSON-LD structured data	Rich results, better entity understanding, local pack eligibility.	Machine-readable claims that survive being lifted off the page.
Semantic HTML and clean headings	Crawl efficiency, featured snippet eligibility.	Tells a retriever where to cut.
Question-shaped content with direct answers	Featured snippets and People Also Ask.	Passages that lift cleanly and need no windup.
Consistent name, address, phone	Local ranking signal.	Lets the engine work out that two mentions are the same business.
Canonical hygiene, deduplicated pages	Consolidated ranking signals.	Stops contradictory versions of you competing to be the source.

Internal linking and clear structure

Distributes authority, aids discovery.

Helps an engine find the next page when the first one is thin.

Only the first row is specific to answer engines. The rest is work a good technical SEO would have told you to do anyway, now with a second reason and a second place it pays.

So the honest case for this category isn't "abandon SEO." It's that the neglected half of technical SEO just became load-bearing for a second channel, and the shortfall shows up somewhere your analytics can't see it.

06 Measuring it, and what measurement can't tell you

Anyone who guarantees an AI will name you is guessing. Ask them how they measured.

Measuring means running the questions your customers actually ask against the engines answering them, keeping what comes back, and recording whether you showed up. It's sampling, not a census, and it carries real limits that an honest report puts up front rather than in a footnote.

Answers wobble. Ask the same engine the same thing twice and the sources can differ. One run is an anecdote. Patterns across many queries and repeated passes are evidence.

Answers are personalized and local. What an engine tells someone in your city may not match what it tells a tester somewhere else.

Models change without telling anyone. A measurement is a photograph, not a contract, which is why measuring again on a schedule beats any single audit.

Getting cited isn't the same as getting traffic. People click a source cited inside an AI summary about 1% of the time.² If your case rests on citation-driven clicks, the numbers

will let you down. For a local service business the conversion was never a pageview anyway, it was a phone call, and being the name in the answer is what makes one happen. If you're a publisher selling pageviews, the math is genuinely worse, and we'd rather tell you than sell around it.

What we'll commit to: a measured baseline, a documented list of what's structurally in your way, the removal of it, and a second measurement showing what moved. What we won't commit to is how the engines behave. That was never ours to promise.

07 What's still unsettled

A document like this is worth more when it marks its own edges.

llms.txt is a convention, not a standard.

Proposed in 2024, it gives AI systems a plain-text summary of your site. Anthropic has confirmed Claude respects it during retrieval. OpenAI and Google haven't confirmed it as an indexing input, and Google has said outright that no special AI-readable file is needed for its AI features. Deployment sat near 2% of sites as of April 2026, with no formal standardization landed.⁴ Our read: cheap, low risk, real benefit with at least one major engine, and not something anyone should sell you as a ranking factor.

agents.json is earlier still.

It describes how an agent should call your API. That isn't a visibility mechanism, though the two get run together in vendor material often enough to be worth separating.

Attribution behavior is a moving target.

How often engines cite anything, and how prominently, comes down to product decisions inside a handful of companies. It's changed before and it'll change again.

Nobody has a moat here, us included.

The measurement side of this will commoditize as the big SEO platforms ship AI visibility tracking. What holds value longer is reading the results correctly and doing the remediation well. We'd rather say that than pretend otherwise.

08 Five things you can check this week

Roughly in order of how much they matter. You don't need us for any of them.

- 1 Load your homepage with JavaScript disabled. If the content disappears, that's what an answer engine sees.
- 2 Open `yourdomain.com/robots.txt` and look for GPTBot, ClaudeBot, PerplexityBot, and Google-Extended. If they're absent, you're running a default written before they existed.
- 3 Run your homepage through a structured data validator. Check that an Organization or LocalBusiness block exists and that the name, address, and service area are right.
- 4 Pick the question a customer would really type. Ask ChatGPT, Perplexity, and Google with AI Mode on. Write down who gets named.
- 5 Read one of your service pages and ask whether any single paragraph answers a real question on its own, without the ones around it.

If four and five go badly while one through three look fine, your problem is extractability, and that's a content structure job. If one fails, nothing else on the list matters until you fix it.

Sources

1. SparkToro analysis of Similarweb clickstream data, 2026. 68.01% of US Google searches ended without a click in the first four months of 2026, against 60.45% in 2024.
2. Pew Research Center, study of 68,879 Google searches by 900 US adults, fieldwork March 2025. Click-through of 8% with an AI summary present against 15% without; 1% click rate on sources cited inside summaries.
3. Measured AI crawler behavior, 2026. GPTBot requested JavaScript files in approximately 11.5% of requests without executing them; ClaudeBot downloaded JavaScript in approximately 23.8% of requests without executing them.
4. State of llms.txt, 2026. Approximately 2% deployment as of April 2026; Anthropic confirms support in Claude retrieval workflows; no IETF standardization as of that date.

Figures were current at the date of drafting and describe a fast-moving area. Where a claim rests on a single study or one vendor's measurement, it's cited as such. Where we're uncertain, we say so.