

SAMPLE EDITION · COMPOUND PARTNERS

The AI Citability Snapshot

Your server logs, decoded

ABOUT THIS SAMPLE

This is a sample of the report an AI Citability Snapshot delivers. The organisation below is a composite: every figure is drawn from real, aggregated data across the professional-services websites we monitor, anonymised and blended so that no single client is identifiable. The structure, the definitions and the advice are exactly what a real edition contains. A real edition also comes with access to the live dashboard behind the numbers, and a walkthrough call on the findings.

The composite is a specialist advisory organisation whose marketing leans on published insight and technical content. If that sounds like you, expect your numbers to rhyme with these.

AI SYSTEMS ARE YOUR BIGGEST READER AFTER GOOGLE

3,100

reads a month, machines training, indexing and answering questions live, roughly 85 machine reads for every 100 pages a person viewed.

2,600

live fetches: an assistant pulling a specific page at the moment a real person asked it a question.

70%

of genuine machine reads come from a single operator, OpenAI, a comfortable position right up until it isn't.

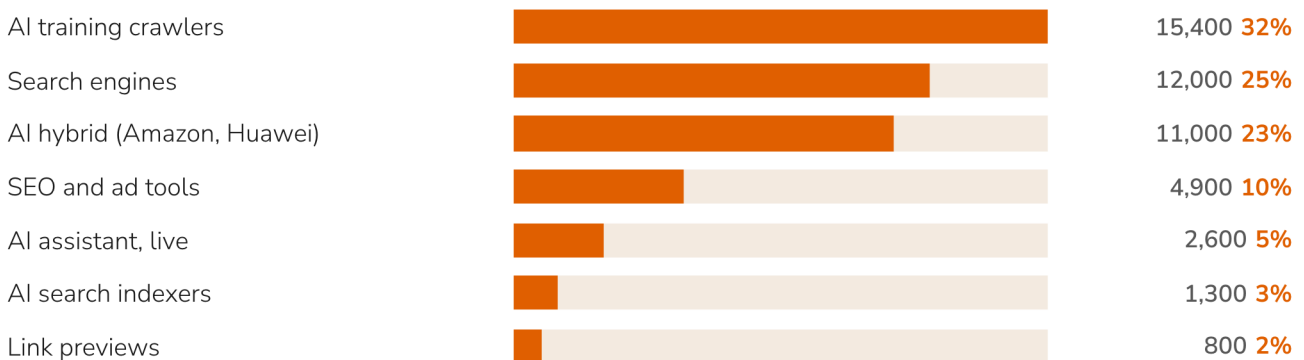
WHAT WE COUNT, AND WHAT WE LEAVE OUT

We check the source IP of every ChatGPT and Perplexity request against the official server lists those companies publish, and sort each hit into a confidence tier. This report runs on Verified and Plausible only. This month we excluded **30% of live AI fetches** as Unverified-suspect and quarantined them rather than counted them, a deliberately conservative line, because "ChatGPT-User" is exactly the kind of user-agent a scraper fakes. Every number below is lower than the raw logs suggest, and more trustworthy for it.

- ◆ **Verified:** user-agent and IP both check out.
- ◆ **Plausible:** a recognised AI user-agent we can't IP-verify: over IPv6, or from Claude and Gemini, which publish no list.
- ◆ **Unverified-suspect:** a recognised user-agent from an IP we can't tie to the operator: a possible spoof. Excluded from this report.

WHO'S ACTUALLY CRAWLING YOU

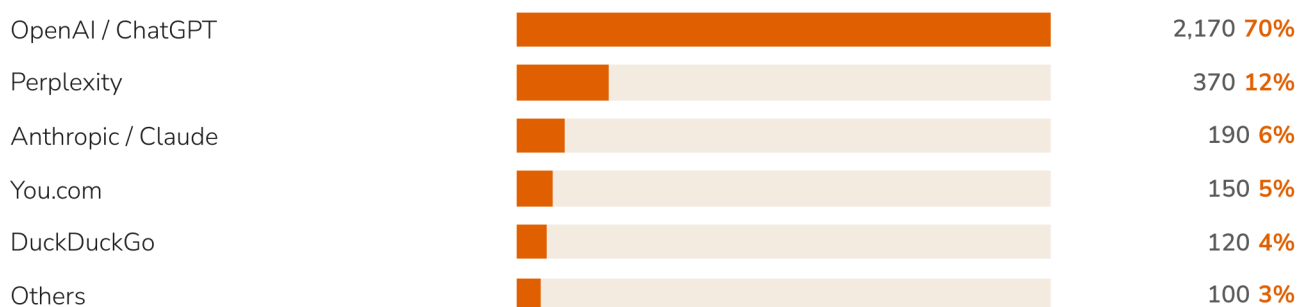
The site took around **48,000** bot requests over the month, roughly **1,550 a day**. The split:



Three of every five requests now come from AI systems of one kind or another. "Google plus a few SEO tools" no longer describes this site's bot traffic, and hasn't for months. These are request counts, a load view, so they include redirects and errors; genuine reads below are lower.

YOUR CITABILITY IS, IN PRACTICE, MOSTLY OPENAI

Of the AI traffic that reads and reaches for this site (genuine 200-OK reads, live plus search), the split by operator:



Citability by AI operator, share of genuine 200-OK reads.

Seven in ten machine reads come from one operator. That concentration is common across the sites we monitor, at two of them a single operator carries over 85% of all machine reading, and it is comfortable right up until it isn't: we watched one engine cut its daily reading of a client site by around 80% inside a single month, verified against the operator's own published servers. If your entire machine presence is one operator's habit, you are one product decision away from starting again. Whether anything is holding the other operators back is one of the three actions at the end of this report.

AI READS YOU FAR MORE THAN IT SENDS YOU

140:1

AI assistants read the site around 2,600 times this month (live fetches) and sent 19 people through, counted from the server logs.

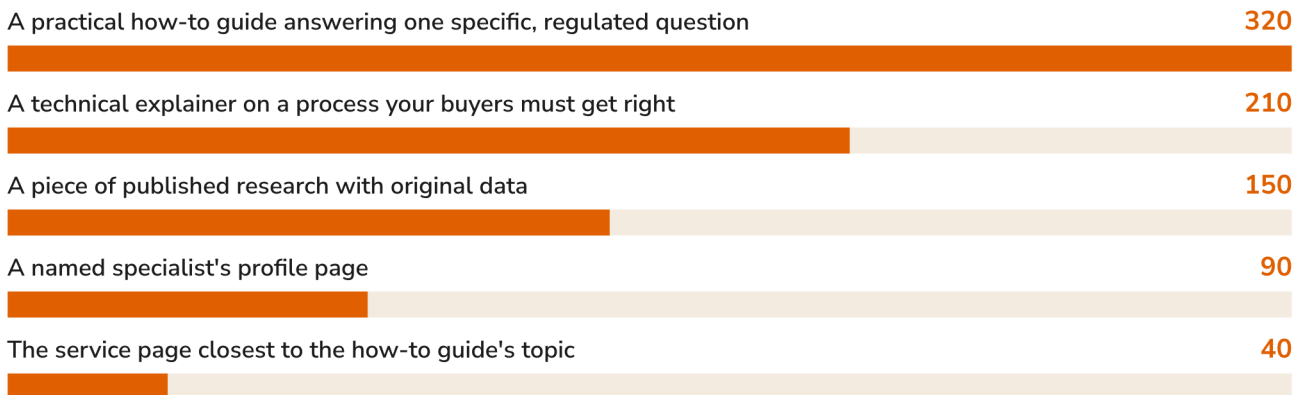
Google Analytics recorded 9 of those 19: GA4 only sees what cookie consent allows, and it often files an AI-referred visit as "direct", so the analytics number is an undercount, here by about half. The server-log count sees every arrival regardless of consent, which is why we report it.

THE READING IS THE REACH

That ratio matches the pattern across our whole monitoring portfolio. Judged as a channel, a ratio like that looks like failure. It is not a channel. Someone asks an assistant a question, your content helps answer it, and most of the time they get what they need without leaving the chat. The click was never the point.

THE MACHINES READ YOUR THINKING, NOT YOUR SERVICE PAGES

Four in five live reads this month were insight content: technical guides, explainers, research and commentary. The service pages built to convert human visitors drew a small fraction of the machine's attention. This is the finding that should change budgets, and it repeats on every website we monitor.

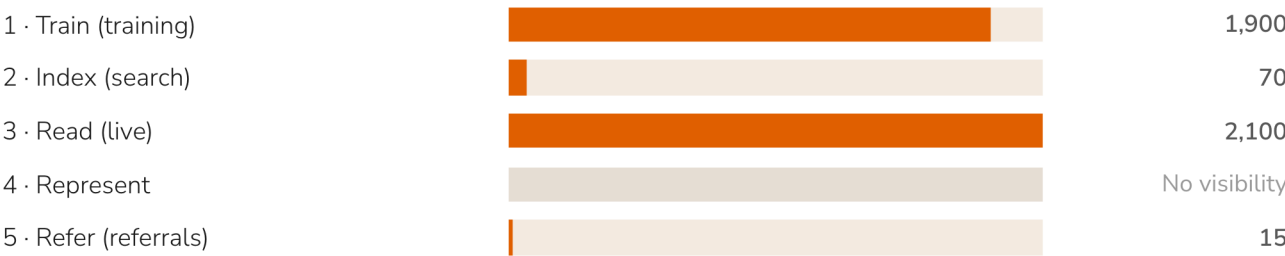


AI reads (live fetches). Described rather than named here; a real edition names the exact pages, links them and ranks the top 25. The machine arrives with a question, not a brief: pages that answer one question well are read dozens of times a month, pages that describe what you sell are not.

WHERE YOU SIT ON THE AI AUDIENCE JOURNEY

OpenAI's month on this site

Each operator sends a fleet of bots, and the fleet is a journey through your content: train, index, read, represent, refer. Four of the five stages are observable:



Stage 4 is a genuine blind spot, not a gap in the logging: when an assistant recommends you inside an answer, there is no fetch, no referral, nothing that reaches a server anyone can measure. We show the gap rather than leave it out, because the step before it and the step after it are both visible, and both are in this table.

The diagnostic that matters is the gap between stages. An operator with heavy training crawl but few live reads has learned you and does not yet cite you. On this composite, OpenAI runs the full journey; the operator to watch is the one stuck at stage 1.

WHAT THIS DOESN'T CAPTURE

This measures AI tools that crawl and fetch your pages, which we can see in the logs. It does not capture Google's AI Overviews (they run on the ordinary Google index, indistinguishable from normal search; with Search Console access, we include Google's own count of how often your pages appeared in its AI results), answers an assistant gives from memory without re-fetching you, retrieval brokered through Bing, or visits where consent or a stripped referrer hides the source. User-agents can be faked, which is why we IP-verify and quarantine what we can't. Low-volume figures are directional. None of this undermines the headline picture, but it is why we report what we can prove and flag what we can't.

SOME OF THE TRAFFIC IS JUST NOISE, AND SOME OF IT IS REDIRECTS

Amazon's and Huawei's bulk crawlers account for most of the "AI hybrid" load on page 2. We would not block them: today's bulk crawler is potentially tomorrow's answer engine, and this kind of expertise content is exactly what is worth being in a training set for.

Redirects are a different matter: **22% of all bot requests** this month were 301s and 302s, mostly self-inflicted by the site's own URL rules. A bot asks for a URL, the site bounces it, a well-behaved crawler follows, and one visit is logged as two hits. Fixing the worst URL groups cuts bot load, frees AI crawl budget for real pages, and makes more fetches land on content. On most sites this is the highest-value cleanup, and a real edition names the exact URL groups to fix.

WHAT WE'D DO ABOUT IT

Three actions, in order

01 · BUILD ON WHAT IS WORKING

Keep feeding the how-to and insight content

It's doing the work. Keep it clean, fast and structured, and write the next piece to answer a question, not to describe a service.

02 · TURN EXPERTISE INTO ANSWERS

Give service pages a question-shaped companion piece

The service pages the machines ignore still hold the expertise buyers ask about. A guide or explainer that answers the question the service solves is how that knowledge becomes something an assistant can read and cite.

03 · CHECK FOR ACCESS GAPS

Make sure nothing is blocking the quiet operators

You cannot choose which operators read you, but you can make sure none of them is locked out. An engine near zero is usually an access gap, not a content problem: check robots.txt, WAF and CDN rules per user-agent before assuming it needs new content.

The real question underneath all three: when a buyer asks the machine, are you the source it reaches for, and can you prove it? Build the position now, while you own it, or read about a competitor who did.

WHAT A REAL EDITION ADDS

Your operators, your pages, your numbers. Every figure above becomes specific: the exact pages the machines read, ranked and linked; your operator split and where each one sits on the journey; your redirect groups, named; and access to the live dashboard so you can see the detail behind every headline.

The AI Citability Snapshot

Your server logs, decoded.

One month of your server logs, £1,000, fixed.

[Request a snapshot →](#)