

TWENTY AUDITS · JUNE TO SEPTEMBER 2026

AI answer visibility: what twenty audits found

1,474 scored answers. Five engines, eight markets. What the machines tell buyers when they're asking buyer-related queries, looking to convert, not searching for the brand directly.

12%

Overall citation rate across 1,474 buyer answers

5 of 20

Companies never cited at all, on any engine, in any market

20 / 20

Audits where every engine recommended somebody. A zero means a competitor took the slot

Why we put this together

Over the past year, the first question a buyer asks has been moving out of the search box and into a chat window. The answer that comes back is written in natural language, confident and names somebody. We wanted to know who.

So between June and September we audited twenty companies, one at a time: real businesses across software, professional services, finance and hardware, in eight markets. For each one we asked the engines the questions their buyers ask, scored every answer, fetched the sources behind the citations, and read what the machines currently believe about the company itself.

The findings are shared because the pattern matters more than any single result, and because the category answers being written now will be hard to unwrite later.

Forge Together

September 2026

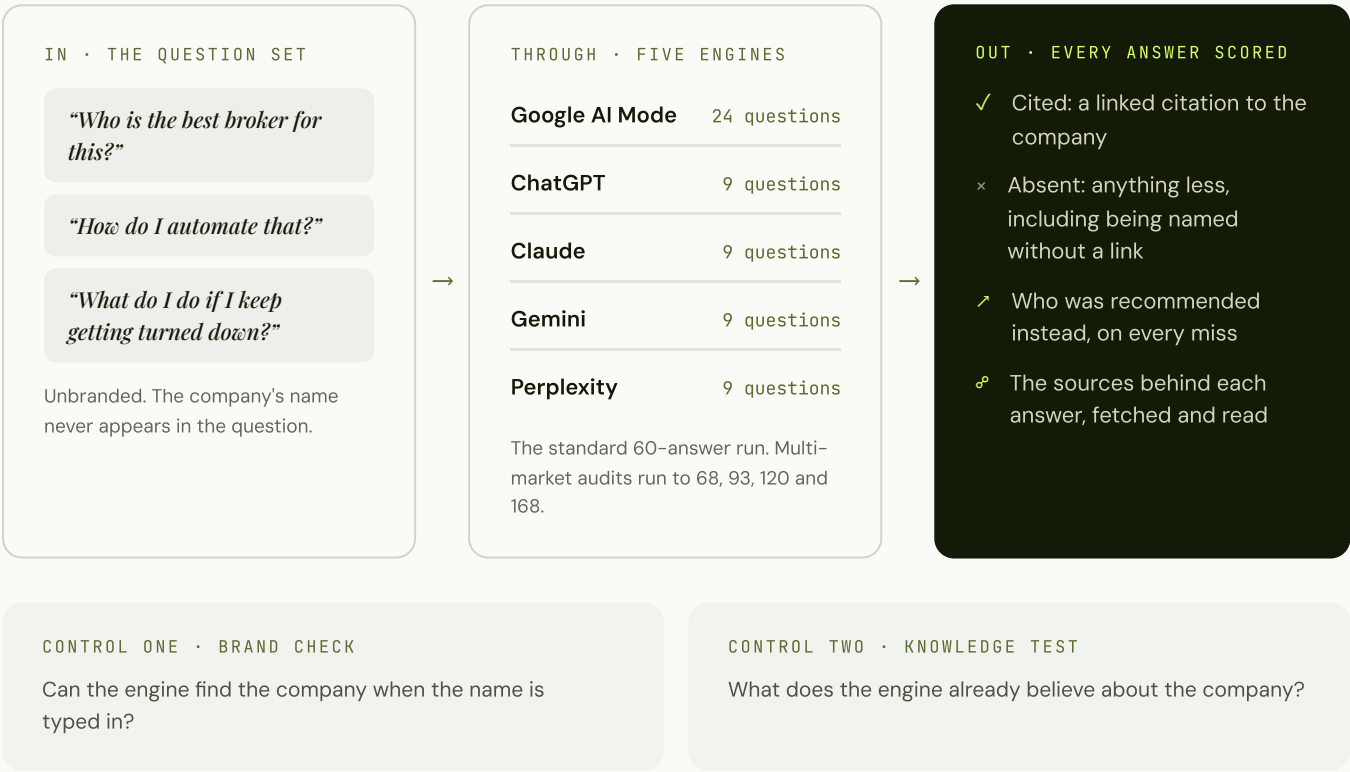
Where the buyer's first question now goes

51% of B2B buyers now start research in an AI chatbot, up from 29% a year earlier (G2, 2026). Between June and September 2026 we measured what those answers say about twenty companies, one at a time, alongside a wider screen of 231 UK professional services firms.

The individual results belong to the companies they were run for, so no audited company is identified here.

01 Audit method

Each audit puts a set of unbranded buyer questions through five engines: the questions someone types when they have the problem and do not yet know who sells the answer.



The two controls matter because a company nobody has heard of and a company the engines know but never recommend look identical from the outside, *and the fixes cost very different amounts.*

02 Citation rates

Across 1,474 scored buyer answers, the audited company was cited in 182 of them. That is **12%**.

SECTOR	MARKETS	RATE	CITED
Legal	UK	0%	0 of 24
Sales agency	UK	0%	0 of 60
Payments	IN, UAE	0%	0 of 68
Insights	UK, US, ES	0%	0 of 93
Sales software	UK, DE	0%	0 of 120
Wealth management	UK	1.7%	1 of 60
Medtech	UK	1.7%	1 of 60
Health data	UK	1.7%	1 of 60
AI software	US, UK	1.7%	2 of 120
Corporate finance	UK	3.3%	2 of 60
Practice management	DE	10%	6 of 60
Waste management	UK	10%	6 of 60
Clinical software	UK	15%	9 of 60
Consumer finance	UK	15%	9 of 60
Infrastructure software	UK, US, DE	19%	18 of 93
Mortgages	UK, UAE, HK, AU	23%	39 of 168
Accountancy	UK	28%	17 of 60
Healthcare workforce	UK	30%	18 of 60
Cybersecurity	US, UK	35%	24 of 68
Logistics hardware	UK	48%	29 of 60

Cited means a linked citation. A company named in the answer text without a link counts as absent, so 12% is a floor. On one audit that distinction was worth 22 further answers.

Half the companies scored 3.3% or less, and five scored nothing at all, on any engine, in any market, on any question. Four scored above 25%, and each of those four had a total miss inside it.

In twenty audits out of twenty the brand-specific terms returned the company. Where the assistants were also asked directly, all four described the business accurately, often down to the founding year, the regulator, the named executives and the published customer results. So the engine holds a correct picture of the company and does not attach it to a buyer's problem. That is a cheaper problem to fix than not being known at all, and it is the one most leadership teams price as the expensive one.

EVERY QUESTION GOT AN ANSWER

In all twenty audits, on every engine, the question produced a written recommendation naming somebody. Not once did an engine decline, hedge or return a list of links instead.

A zero means a competitor took that slot.

03 The 231-firm screen

78%

180 of 231 UK professional services firms never cited once. Screened 3 August 2026, 18 buyer questions each in their own town.

1.35%

Share of the citations their own buyer questions generated. The rest went to national brands and to directories.

2.5×

Questions naming a specialism were cited two and a half times as often as generic head terms, 12.8% against 5.1%.

That screen ran a shorter question set on two engines, Google AI Mode and Perplexity, and did not include ChatGPT. ChatGPT is consistently the weakest of the five engines in our full audits, so 78% is more likely flattering than harsh.

Twenty-four firms appeared only on their specialism and never on the obvious category question. Chasing “best X in my town” targets the most contested and least winnable question in the set.

04 Findings

Seven patterns across sectors and markets.

FINDING 4.1

Pages written as answers

Two blog posts on a penetration testing platform's site carried full article and question markup. Those two pages took seven of the nine blog citations the domain earned. Nine comparison pages, built deliberately for the questions being asked, had half the heading density and thinner markup, and took none. Asked for alternatives to a named rival, Google cited the blog listicle over the comparison page built for that exact question.

One page on a regional accountancy firm's site was quoted by all five engines. It named an industry body, named a real client, and stated a specialism. Nothing else on the site was written that way. One page on an infrastructure software company's site was cited repeatedly across two markets. Seven hundred words, naming the application and stating the constraints. The other 85 pages were not written like that.

It works in reverse. A mortgage broker's homepage carried 23 headings, three shaped as questions, and won ten citations. Its foreign nationals page carried six headings, none shaped as a question, and won nothing across twelve chances.

FINDING 4.2

Rankings that produce no citations

A regional accountancy firm ranked eighth nationally for a term with 5,400 monthly searches, and was cited in none of the four AI answers on that topic. A consumer finance business ranked ninth for a furniture-finance term and was absent from the AI answer on buying a sofa on finance. The penetration testing platform ranked second nationally for its main category term, and half its US citations landed on a single blog post.

Those companies earned their rankings, and the engine reads the answer out with a competitor's name in it. A rankings report cannot show you that.

FINDING 4.3

The sources behind a cited answer

45%

of all English-language citations on one audit came from YouTube, ahead of every vendor and every review site.

The company's own domain is usually a minority of the sources behind a cited answer. The rest is roundups, review platforms, directories, trade publications and video.

We fetch the pages the engines quote and search them for the client. Twelve roundups on a payments audit: none. Twelve third-party guides on a consumer finance audit: none. Twenty-one category pages on an AI research software audit: none, while a direct rival published its own "top ten alternatives to X" list and got cited for it. Ten listicles on a consumer insights audit: none, and two of those ten were written by competitors who ranked themselves into their own lists.

Video was the largest single source in four of the twenty. None of the four had any presence there.

The review layer explains most poor ChatGPT results. Two companies led their category on Google's own answers and lost badly on the standalone assistants, and neither had a listing on G2, Capterra or TrustRadius. One had three review profiles already live and empty, one of them still showing as unclaimed. Of the twenty, one had a working review presence, and the engines cited it in all four of its markets.

FINDING 4.4

Registries filling the gap

In 9 of the 20 the engines returned contradictory facts about the company. Four engines gave four founding years for one, and one told buyers it might be a rebrand of a competitor. Another produced three founding years, three funding totals and four headcounts between them, and surfaced a corporate identity the company had abandoned. One engine put a regional accountancy firm at 150 staff and five offices, another at 220, against a number well above both on the firm's own careers page. A different engine named the wrong professional network. A UK lender was credited with a product it does not sell, because the description belonged to a same-named US business.

In every case the company's own About page said almost nothing checkable: no founding date, no scale, no corporate structure, no regulatory position. So the engine sourced those facts elsewhere.

FINDING 4.5

Who is taking the citations

Two rivals beating an infrastructure software firm in two different countries were both founded within a year of it, and neither was winning on authority. A regional accountancy firm was losing to a competitor holding ten of the 305 references across the whole run, off a handful of well-built pages. An AI research software company held five cited answers. A direct rival held 357. The largest incumbent in the category held 62,950. The first of those gaps is reachable.

In the 231-firm screen the median uncited firm had a local peer taking 14 of its 18 chances.

FINDING 4.6

Where the high scorers are absent

A smart locker company scored 48%, the strongest result in the set, and was absent from all five "best system" questions and from both pricing questions. A workforce software company led the vendor field on Google's answers and scored zero on the four questions about its own published outcomes. The penetration testing platform scored 35% overall and zero across eight UK questions. Across its 96 Google AI Mode questions, a mortgage broker won 13 of the 20 that named the product category and 6 of the remaining 76. A regional accountancy firm scored 3 of 24 on Google, two of them because the buyer typed the name of a town it has an office in.

FINDING 4.7

Second markets

One firm's citation rate in the UAE was more than twice its UK rate on the same questions. The penetration testing platform held six times as many US ranking keywords as UK ones and did not have a single UK service page. A consumer insights business scored zero in Spain, where the answer went to two firms, neither of them Spanish.

A second market runs on its own directories, its own comparison sites and its own unwritten answers.

05 The eleven diagnoses

Two companies can score identically for opposite reasons, and the money goes to different places.

01 Unowned ground

No firm holds the category answer. Directories and generic guidance fill it.

02 A leader forming, the niche still open

One rival breaking through on the broad question, the specific slot unclaimed.

03 Strong overall, zero in one place

Healthy aggregate hiding a total miss on one engine or one buying stage.

04 Pre-answer category

Nobody cited and demand still forming. A chance to set the answer, with a caveat about timing.

05 Ranks but is not quoted

A strong organic estate sitting on the uncited questions.

06 Described but not linked

Named accurately in the answer text, with the click sent to a rival.

07 The citation asset was deleted

A relaunch or migration took the pages the engines were quoting.

08 Cited at scale, out of proportion

Citations in aggregate, near zero on the questions that sell.

09 Answered from the registry

The engines contradict each other about the company, because it never said.

10 Cited to the wrong audience

Real citations, drawn by content that attracts people who will never buy.

11 The crawler is locked out

Everything done right, and the door bolted at the edge. The most expensive single defect in the set.

DIAGNOSIS 11, IN THE FIELD

One company returned a 403 to eight AI crawlers at its CDN edge while serving Google, Bing and an ordinary browser a clean 200. Two of the eight were the live fetchers that open a page while a buyer is asking about the company mid-conversation. Its robots.txt said everything was allowed. It sells AI agents that read the web.

Three others in the same class. One relaunch left 115 of 117 ranking URLs returning 404 with no redirects, and took 687 already-cited AI answers with them. One robots.txt blocked every PDF the company published, while an engine cited a competitor's PDF report in the same answer. One file written specifically to tell language models what the company sells pointed at a server that no longer resolved.

All four were invisible from inside the business.

06 How Forge approaches the fix

1

Stop the self-inflicted damage

Crawler blocks, dead redirects, markup emitted in a form no parser will read, broken machine-readable files, pages that 404. This is days of work and it costs nothing beyond developer time. In several audits it was the entire explanation for the score.

2

State who you are, in a form a machine can read

Founding date, scale, structure, regulatory position, named people, real customers, and the markup that carries them. In 19 of the 20 the site was either missing the structured data engines read or carrying it in a form no parser could use. The single exception had it, and was absent for other reasons. Quick to specify, and the only fix for engines telling buyers the wrong facts about you.

3

Get onto the surfaces the answers are built from

Review platforms, the directories your category runs on, the roundups the engines quote, and video where video leads. Usually the highest-return action on this list, and almost always off your own website, which is why it gets missed.

4

Rewrite the pages that should be winning, and write the questions nobody has answered

This is the longer work. Every audit turned up questions in the category that nobody had answered, and those are the cheapest positions to take.

The proof usually already exists: published outcomes, named clients, accreditations and commendations sitting in press releases, gated PDFs, testimonial carousels and internal project notes. None of it in a form a machine can read. *Rewriting evidence you already own is a very different budget conversation from earning it.*

07 Limits

AI answers vary between runs. A single measurement is a snapshot. We re-run the same question set at 90 days against the same configuration, and the movement between the two runs is the number to trust.

We chose most of these companies ourselves, and we picked them because we expected to find something, so 12% is not a market average. The set does show the range, and what tends to sit behind a low number.

Not every audit ends with a recommendation to spend. One found a category where the answer was unclaimed but the buyers were not there in volume yet, and we said so.

The twenty audits above are baselines. One company in the set has been re-measured, and it had not acted on the findings, so it returned the same score. The evidence that the number moves comes from client programmes rather than from these audits.

£240k → £645k

A healthcare compliance platform's pipeline across a six-month window. The largest contract of that period came through ChatGPT.

2 of 8 → 15 of 24

A digital verification provider's decision-stage AI answers across the four assistants in a month, measured the same way as the audits above, and a seven-figure contract attributed directly to the AI-search work.

The category answer is being written **now**

In most of the sectors we measured it is not yet settled. Three audits found no incumbent at all. Several found a leader whose margin is still narrow enough for a competitor to take. **Being readable by a machine is currently worth more than being big.**

An audit tells you which of the eleven diagnoses applies to you. It produces a number you can re-count and a fix list built from what the run found. It is the first phase of a growth programme rather than the whole of it. The same research feeds the content, the sales material and the organic work underneath.

Email Junior@forgetogether.agency to get your free AI Visibility audit and plan to be discoverable and achieve growth.