



CITARI RESEARCH · 2026

The 2026 State of AI Search Visibility

Evaluating AEO Readiness Across the West of Scotland

A primary-research whitepaper · 160 audited businesses

Citari

AI Search Visibility Auditing
Human website. Machine Search.

strategy@citari.co.uk
<https://citari.co.uk>

SECTION 1

Executive Summary

Citari's audit of 160 West-of-Scotland businesses reveals that the regional economy is, for the most part, invisible to the non-rendering fetch bots that increasingly mediate how customers discover suppliers.

This whitepaper is built on Citari's own primary research: every percentage below is computed directly from the Citari Audit Framework's run against 160 deduplicated business domains (deduplicated from 163 audited records by normalising each URL — stripping scheme and a leading `www.`, lower-casing, and trimming any trailing slash).

The headline finding is stark. **Citari's research indicates that 68.1% of audited businesses (109 of 160) returned an empty crawl** — Citari's crawler, reading the raw served HTML exactly as a non-rendering AI fetch bot does, recovered no usable content from more than two thirds of the region's commercial websites. Only 31.9% (51 of 160) returned legible, structured HTML.

The Citari Audit Framework attributes this to two distinct, fixable failures. **An Automated Security Gate (firewall/WAF) intercepted 48.1% of all sites (77 of 160)** — nearly half the region is actively turning AI fetchers away at the door. A further 18.8% (30 of 160) returned a 401/403 lockout or a genuinely blank page, and 1.3% (2 of 160) served a client-side JavaScript shell with no content in the initial HTML response. Crucially, this is overwhelmingly a *network* problem, not a deliberate editorial one: Citari found that only **3 of 160 businesses (1.9%) issue a genuine `robots.txt` directive blocking any AI user-agent**. The region is not choosing to opt out of AI search — it is being silently locked out by its own infrastructure.

Yet a second Citari finding complicates the picture in a commercially important way. In Citari's Stage-2 latent-citation test — three frontier models (Claude Sonnet 4.5, OpenAI GPT-4o, Gemini 2.5 Flash) queried in plain text-completion mode with no web search, no retrieval and no tools — **38.1% of businesses (61 of 160) were named by at least one model from training-data knowledge alone**. More strikingly, **43 businesses (39.4% of the 109 that failed the technical crawl) were still recognised by a model despite being technically illegible**. Citari terms this the *Shadow Citation Paradox*: latent, off-page brand authority is currently carrying websites that their own servers have rendered invisible.

The commercial stakes are concrete. Across the wider market, Citari's published guide *Unlocking the AI Search Frontier* documents that ChatGPT reached roughly 900 million weekly active users by February 2026, that Google's AI summaries have driven click-through on the underlying result down from about 15% to roughly 8% (a 54% drop), and that position-one organic click-through has fallen 61%, from 1.76% to 0.61% (Seer Interactive and Pew Research Center). In an answer-engine

economy, the businesses that an AI cannot read are the businesses an AI cannot recommend — and a 4.4× conversion premium for AI-referred traffic (Semrush clickstream analysis) is accruing to whoever the models *can* both read and recall.

Citari's conclusion is that the West of Scotland's AEO problem is, for most firms, a same-day infrastructure fix rather than a long content programme. The remediation framework in Section 7 maps every finding in this paper to the Four Pillars of the Citari method. To benchmark your own estate against this dataset, contact strategy@citari.co.uk.

SECTION 2

Introduction to the AEO Era

Citari's audit of 160 West-of-Scotland businesses was designed to answer a single strategic question: when a prospective customer asks an AI assistant for a solicitor in Glasgow, an optician in Hamilton or a roofer in Paisley, can that business even be seen — and is it already known? The answer, derived entirely from Citari's primary research, is that for roughly two in three regional firms the honest answer is “no” on the first count, and that a surprising minority survive only on the second.

This matters because search itself is changing shape. For two decades, visibility meant ranking on a page of blue links and earning the click. That model is eroding. The macro-context that frames this paper — drawn from Citari's published guide *Unlocking the AI Search Frontier* and kept deliberately subordinate to our own findings — is that AI answer engines now intercept the query before the click ever happens. ChatGPT reached approximately 900 million weekly active users by February 2026; Perplexity served around 780 million monthly queries; an estimated 31% of Gen Z now reach for an AI tool first. In the United States, around 58.5% of searches already end without a click (zero-click), and Google's AI summaries have roughly halved the click-through that does occur, from about 15% to around 8%.

The mechanism that replaces the click is *Retrieve-and-Synthesise*: an answer engine fetches sources, reads them, and composes a single synthesised recommendation. Two capabilities therefore decide whether a business appears in that answer. First, the engine must be able to **retrieve and parse** the business's page — and many production AI fetchers, like Citari's crawler, do not execute client-side JavaScript and will not negotiate past a hostile firewall. Second, the engine must already **recognise** the business as a credible entity, a function of the brand's footprint across the model's training corpus and the wider web.

Answer Engine Optimisation (AEO) is the discipline of engineering for both. It is not a rebrand of SEO; it is a distinct technical and editorial practice concerned with machine legibility, structured data, answer-first content and factual authority. The remainder of this paper measures where the West of Scotland stands on each — beginning with the most acute failure Citari uncovered: the wholesale, largely accidental blocking of AI crawlers.

SECTION 3

The Great Scottish Bot-Blocking Crisis

The defining finding of Citari's research is the scale of the empty crawl. **Citari's audit of 160 West-of-Scotland businesses found that 109 of them — 68.1% — returned no usable content to a non-rendering fetch.** Before interpreting that number, the method must be stated precisely, because the figure means something specific. The Citari crawler fetches the raw served HTML over HTTPS using a standard HTTP client and parses it with an HTML parser. It does not run a headless browser and it does not execute JavaScript. It deliberately reads each page the way a non-rendering AI fetch bot does: whatever is present in the initial HTML response is the entirety of what it — and a large share of real AI fetchers — can see. A domain is scored `ok` only when that raw response yields at least 400 characters of parsed body text with measurable structure; otherwise it is an `empty_crawl`.

Citari's framework resolves every empty crawl into one of three causes, and the distribution across the 160-business dataset is the heart of this section:

EMPTY-CRAWL CAUSES ACROSS THE 160-BUSINESS DATASET

Empty-crawl cause	Count	% of all 160	% of the 109 empty crawls
Automated Security Gate (firewall/WAF)	77	48.1%	70.6%
Blocked or blank (401/403 lockout or empty page)	30	18.8%	27.5%
JavaScript shell (HTTP 200, body under 400 chars)	2	1.3%	1.8%
Total empty crawl	109	68.1%	100%
Legible (<code>ok</code>)	51	31.9%	—

The dominant failure is the **Silent Firewall**. Citari found that an Automated Security Gate intercepted the request for 77 of 160 businesses — 48.1% of the entire region and more than seven in ten of all empty crawls. These are not low-effort sites; they include some of the region's largest and best-resourced firms. The mechanism is mundane and therefore widespread: an off-the-shelf WAF or anti-bot ruleset returns a 403 or a challenge page to any client whose user-agent or request signature it does not recognise. To a human on a mainstream browser the site looks perfect. To an AI fetcher, it does not exist.

The second cause, **blocked-or-blank**, accounts for 30 businesses (18.8% of all sites, 27.5% of empty crawls): a hard 401/403 lockout or a page that genuinely returns no content. The third, the **JavaScript shell**, is rarer in this dataset than the regional stereotype would suggest — only 2 businesses (1.3%) served an HTTP 200 with a sub-400-character body, i.e. a client-side framework that paints content only after rendering. Citari's crawler, like many AI fetchers, never

triggers that render, so the content is effectively absent. The low JS-shell count is itself a finding: in the West of Scotland the crawler-blindness problem is overwhelmingly about the firewall, not the front-end framework.

The most important nuance in this section — and the one a casual analyst gets wrong — is the relationship between firewall interception and `robots.txt`. A firewall block is a *network* event; a `robots.txt` Disallow is an *editorial* directive. They are not the same thing and must never be conflated. Citari parsed `robots.txt` for six AI user-agents (`ChatGPT-User`, `OAI-SearchBot`, `ClaudeBot`, `PerplexityBot`, `Google-Extended`, `Applebot-Extended`) on every site. Counting only genuine `robots.txt` directives — and excluding firewall-intercepted rows, where the WAF prevented `robots.txt` from being read at all — Citari found the following:

ROBOTS.TXT DIRECTIVES VS. FIREWALL INTERCEPTION, BY AI USER-AGENT

AI user-agent	Genuine robots.txt Block (excl. firewall)	Firewall-intercepted (network block)
<code>ChatGPT-User</code>	0	77
<code>OAI-SearchBot</code>	0	77
<code>ClaudeBot</code>	2	77
<code>PerplexityBot</code>	0	77
<code>Google-Extended</code>	3	77
<code>Applebot-Extended</code>	2	77

In total, **only 3 of 160 businesses (1.9%) issue any genuine `robots.txt` directive blocking an AI bot** — and even those three block only a subset of agents (most commonly `Google-Extended`). By contrast, **77 businesses (47.8%) show their AI bots intercepted at the firewall** across all six user-agents simultaneously, which is the signature of a network-layer block rather than a published policy.

The strategic reading is therefore optimistic. The West of Scotland has not decided to opt out of AI search; almost no one has written a rule to that effect. The region is being locked out *by default* — by security middleware that nobody configured with AI fetchers in mind. This maps directly to **Pillar 1 of the Citari method — Unlock the Security Gate**: the remedy is a WAF allowlist for the named AI user-agents and, where applicable, server-rendered or prerendered HTML so the initial response carries content. It is an infrastructure change, not a rewrite, and Section 7 sets out the steps.

SECTION 4

The Off-Page Paradox

If Section 3 were the whole story, the regional outlook would be bleak: two thirds of businesses unreadable, therefore two thirds unrecommendable. Citari's second major finding shows why the reality is more nuanced — and why technical remediation is an opportunity rather than a lost cause.

In Stage 2 of the audit, Citari queried three frontier models — Claude Sonnet 4.5, OpenAI GPT-4o and Gemini 2.5 Flash — about each business using consumer-intent prompts built from the firm's service and city. These queries were made in **plain text - completion mode: no web search, no retrieval or grounding, and no tool use**. The models answered purely from knowledge baked into their training data, and Citari recorded a citation as a case-insensitive substring match — was the business name present anywhere in the answer. This is a precise and deliberately conservative test. It does **not** measure live, real-time AI search; it measures *latent* brand recognition — whether a model already “knows” a business from its training corpus. Read this throughout as a measure of the **baseline brand - knowledge floor**: the recall a business already has before any live engine retrieves a single page.

Against that floor, **Citari found that 61 of 160 businesses (38.1%) were named by at least one model from latent knowledge alone**. The paradox emerges when this is cross-tabulated against the technical crawl: **43 businesses — 39.4% of the 109 that returned an empty crawl, and 26.9% of the entire dataset — were recognised by at least one model despite being technically illegible to a non-rendering fetch**.

THE SHADOW CITATION PARADOX

Citari Shadow - Citation measure	Count	Denominator	%
Cited by ≥ 1 model (latent)	61	160 (all)	38.1%
Cited by ≥ 1 model and empty crawl	43	109 (empty crawls)	39.4%
Cited by ≥ 1 model and empty crawl	43	160 (all)	26.9%

This is the *Shadow Citation Paradox*, and its interpretation is central to Citari's thesis. Because Stage 2 uses no live retrieval, a latent citation cannot have come from the business's own (unreadable) website. It must have been earned off-page — through directory listings, Google Business Profiles, news coverage, professional registers, review platforms and the broader web that the models trained on. Off-page authority, in other words, is currently doing the work that these firms' own sites cannot.

Two implications follow. First, latent recall is real but uneven and fragile: it favours larger, older or more newsworthy brands and offers nothing to the many smaller firms with no shadow footprint. Of the 109 empty-crawl businesses, the majority — 66 of them — were recalled by no model at all.

Second, and more importantly, latent recognition is the *floor*, not the ceiling. A business that a model already half-knows, and whose site is then made legible and well-structured, gives a live answer engine both the recognition to surface it and the retrievable, parseable content to quote it. The Four Pillars exist precisely to convert this latent floor into live, citable visibility.

The off-page paradox should therefore not be read as permission to neglect the website. It is evidence that brand equity already exists for many West-of-Scotland firms — and is being squandered the moment a live engine tries, and fails, to read the source. Closing the gap between latent recall and technical legibility is the single highest-leverage AEO move available to the region.

SECTION 5

The Battle of the Engines

Citari's latent-citation test also lets us compare how the three frontier models behave when recalling local businesses from training knowledge. The pattern is consistent and, for AEO strategy, instructive. The figures below are Citari's own, computed across all 160 deduplicated businesses; each is the count of businesses whose name appeared in that model's plain-completion answer.

LATENT LOCAL RECALL BY MODEL

Model	Businesses cited (latent)	% of 160
Gemini 2.5 Flash	48	30.0%
OpenAI GPT-4o	24	15.0%
Claude Sonnet 4.5	15	9.4%
Any model (≥1)	61	38.1%

Citari's research indicates Gemini leads decisively on latent local recall, naming 30.0% of the region's businesses — double GPT-4o's 15.0% and more than three times Claude's 9.4%. This ordering should be framed carefully: it is a difference in *latent/training recall of local entities*, not a difference in live local-search integration. None of the three models was given web access in this test. Gemini's lead therefore reflects how readily its training knowledge surfaces specific local commercial entities for a localised consumer query, with GPT-4o more selective and Claude the most conservative — Claude tends to name a local business only when the entity is strongly established.

For practitioners, the strategic reading is threefold. First, the spread across engines means **AI Recommendation Share is not a single number but a portfolio**: a business invisible to Claude may still be recalled by Gemini, and a brand that wants resilient visibility must earn recognition broadly rather than optimise for one assistant. Second, the conservative behaviour of Claude and the selectivity of GPT-4o reward exactly the factual-authority signals that Pillar 4 of the Citari method targets — verifiable statistics, authoritative outbound citations and named attribution — because

these are the signals that move a brand from “plausibly real” to “confidently nameable.” Third, because all three figures describe the latent *floor*, the engines' differing thresholds make the case for live retrievability even stronger: a well-structured, firewall-open site gives the more cautious engines the on-page evidence they need to cite a business they would otherwise omit.

In short, Gemini will currently mention more West-of-Scotland businesses unprompted than its rivals — but no business should rely on a single engine's training-data memory. The durable strategy is to be both broadly recognised off-page and reliably retrievable on-page, so that every engine, whatever its threshold, can both recall and verify the brand.

SECTION 6

The Sector Vulnerability Index

Citari's dataset spans nineteen normalised service sectors across the West of Scotland, allowing a direct ranking of which industries are most and least AEO-ready. The table below is computed entirely from Citari's primary research. For each canonical sector it reports the deduplicated business count (n), the empty-crawl rate, the latent-citation rate (the share named by at least one model), and — critically — the four Citari scores **averaged over ok rows only**. Citari never averages in the by-construction zeros of empty-crawl sites, because doing so would understate the genuine on-page quality of the firms that are legible; the per-sector ok denominator is stated so every score average is traceable.

SECTOR VULNERABILITY — CRAWL HEALTH, LATENT CITATION & ON-PAGE SCORES

Sector	n	Empty-crawl %	Cited ≥1 %	ok rows	Visibility	Comprehension	Trust	Reading-Ease
Solicitors	31	71.0%	48.4%	9	12.3	19.1	12.0	33.6
Estate Agents	26	88.5%	42.3%	3	3.7	22.5	11.6	39.7
Accountants	20	50.0%	25.0%	10	12.2	20.4	9.4	26.8
Dentists	17	76.5%	35.3%	4	0.0	17.5	19.0	55.2
IT Support	10	50.0%	10.0%	5	0.0	5.0	13.4	59.1
Vets	9	55.6%	44.4%	4	8.3	28.1	11.9	50.6
Opticians	8	62.5%	50.0%	3	22.2	16.9	1.5	14.5
Roofers	6	66.7%	33.3%	2	0.0	33.7	8.9	46.6
Electricians	4	75.0%	0.0%	1	0.0	0.0	20.0	72.0
Plumbers	4	25.0%	25.0%	3	11.1	3.6	17.4	52.1
Physiotherapists [Indicative Sample]	3	66.7%	33.3%	1	0.0	42.3	62.3	51.9
Architects [Indicative Sample]	3	66.7%	33.3%	1	11.1	18.0	0.0	14.4
Hotels [Indicative Sample]	3	100.0%	33.3%	0	—	—	—	—
PR & Marketing Agencies [Indicative Sample]	3	66.7%	100.0%+	1	33.3	0.0	16.7	50.0
Financial Advisers [Indicative Sample]	3	66.7%	66.7%	1	11.1	0.0	17.2	82.0
Engineering [Indicative Sample]	3	33.3%	66.7%	2	11.1	24.4	5.9	17.7

Sector	n	Empty-crawl %	Cited ≥1 %	ok rows	Visibility	Comprehension	Trust	Reading-Ease
Contractors [Indicative Sample]	3	66.7%	0.0%	1	0.0	4.1	9.0	27.0
Motor Dealers [Indicative Sample]	2	100.0%	50.0%	0	—	—	—	—
Storage Facilities [Indicative Sample]	2	100.0%	50.0%	0	—	—	—	—

Sectors marked **[Indicative Sample]** have $n \leq 3$ businesses: their rates are directional signals, not robust verdicts on the sector across the region, and should not be generalised from two or three data points. Smaller sectors also carry small ok denominators; Citari reports their score averages for completeness but they too are indicative rather than robust. Hotels, Motor Dealers and Storage Facilities returned no ok rows at all and therefore have no on-page score averages.

† **Common-word substrings caveat.** Citari's latent-citation test is a case-insensitive substring match of the business name against each model's answer. Two of the three PR & Marketing agencies are named **Frame** and **Wire** — single, common English words. A model writing "...build a clear **framework**..." or "...a direct **wire** to your customers..." in an ordinary marketing answer trips the match without ever intending to cite the specific business. The sector's 100% rate is therefore almost certainly inflated by these incidental hits; of the three firms, only **BIG Partnership** — a distinctive multi-word name cited by all three models — is a confident latent citation. Citari flags this in the interest of methodological transparency.

Several patterns emerge from Citari's data. **Estate Agents are the most technically vulnerable of the well-populated sectors:** 88.5% of the 26 audited agencies returned an empty crawl, the highest rate of any sector with meaningful sample size, overwhelmingly because of Automated Security Gates on portal and franchise platforms. Yet 42.3% are still cited from latent knowledge — a textbook Shadow Citation Paradox, where strong brand and portal presence carries firms whose own sites a fetcher cannot read. **Solicitors**, the largest sector at 31 firms, sit at 71.0% empty-crawl with a healthy 48.4% latent-citation rate; their legible sites score reasonably on comprehension and reading-ease, suggesting the profession's problem is the firewall, not the content.

At the healthier end, **Accountants (50.0% empty), IT Support (50.0%) and Engineering (33.3%)** crawl best among populated sectors — though IT Support's very low latent-citation rate (10.0%) shows that being readable is necessary but not sufficient: these firms lack the off-page footprint to be recalled, and need Pillar 4 authority signals as much as Pillar 1 access. **Plumbers** stand out as the most crawl-healthy trade (25.0% empty), a reminder that smaller independent trades on simple, server-rendered sites are sometimes more legible than large enterprises behind heavy security stacks.

The citation column tells its own story, with one important caveat. **Financial Advisers (66.7%) and Engineering (66.7%)** punch above their crawl health on latent recall, reflecting media and B2B visibility. PR & Marketing Agencies show a headline 100% citation rate, but Citari flags this as substring-inflated rather than genuine: two of the sector's three firms are named "Frame" and "Wire", common words that the latent-citation substring test almost certainly matched incidentally

(see the † note above the table) — only BIG Partnership is a confident citation here. At the other extreme, **Electricians and Contractors register 0% latent citation** despite being readable — they are entirely dependent on closing the off-page authority gap.

Taken together, the Sector Vulnerability Index points to a two-track remediation priority for the region: high-empty-crawl, high-latent sectors (Estate Agents, Solicitors, Dentists, Opticians) need **Pillar 1 access fixes first** to convert existing brand equity into live citations; readable-but-unrecalled sectors (IT Support, Electricians, Contractors) need **Pillar 4 authority building** to earn a place in the answer at all.

SECTION 7

The Citari Technical Remediation Framework

Citari's research shows that the West of Scotland's AEO deficit is, for most businesses, an infrastructure problem with a short remediation path. The framework below maps every finding in this paper onto the Four Pillars of the Citari method and is sequenced for CTOs, IT directors and agency owners. The principle is *diagnose specifically, prescribe directionally*: fix access first, because no other pillar matters to an engine that cannot read the page.

Pillar 1 — Unlock the Security Gate (highest priority; addresses 79 of 109 empty crawls)

The Automated Security Gate (77 sites) and JavaScript shell (2 sites) together account for the majority of the region's invisibility. Remediate in this order:

- 1. Audit your WAF / anti-bot ruleset for AI user-agents.** Confirm whether your firewall is returning 403s or challenge pages to `ChatGPT-User`, `OAI-SearchBot`, `ClaudeBot`, `PerplexityBot`, `Google-Extended` and `Applebot-Extended`. In this dataset, 47.8% of businesses were intercepted at this layer.
- 2. Allowlist the named AI fetchers** at the WAF/CDN edge by user-agent and, where your vendor supports it, by published IP range — rather than relaxing security globally. The goal is to let verified AI fetchers through while keeping malicious bot mitigation intact.
- 3. Serve content in the initial HTML response.** If your site is a client-side single-page application, adopt server-side rendering (SSR) or build-time prerendering so the raw HTML a non-rendering fetcher receives already contains your content. Citari applies exactly this technique to its own site: a post-build prerender step replaces the empty SPA shell with fully rendered, semantic HTML.
- 4. Verify the fix the way an AI sees it** — fetch your own pages with a plain HTTP client (no browser, no JavaScript) and confirm at least 400 characters of meaningful body text are present in the response.

Pillar 2 — Official AI Registry Code (JSON-LD structured data)

Among the 51 legible (ok) sites, Citari found that **22 (43.1%) were missing at least one core JSON-LD field**, and only 56.9% carried complete structured data. The most commonly absent fields were `telephone` (missing on 22 of the 51), `address` (20), `url` (7) and `name` (5).

- 1. Publish a `LocalBusiness` (or appropriate sub-type) JSON-LD block** on every key page, populating at minimum `name`, `address`, `telephone` and `url`.
- 2. Mirror your visible contact and location details exactly** in the structured data; inconsistency undermines the trust signal it is meant to send.
- 3. Add `FAQPage` structured data** that mirrors a visible, answer-first FAQ — Citari's published guidance notes that the FAQ format is materially more likely to be cited by answer engines.

Pillar 3 — AI Scanning Layout (Comprehension)

The legible sites in this dataset averaged a Comprehension Score of just 17.5/100, indicating thin semantic structure even where content is readable.

- 1. Lead with answer-first paragraphs:** state the direct answer in the opening sentence, then elaborate.
- 2. Use question-form `<h2>` / `<h3>` headers** that mirror real consumer queries, so an engine can map a question to your answer.
- 3. Structure data as genuine HTML `<table>` s and `` / `` lists**, not as images or CSS-styled `<div>` s — these are the constructs an engine parses most reliably.
- 4. Ensure all crawl-relevant content renders on initial paint** as plain semantic markup, not inside collapsed interactive components that leave the DOM empty until a user interacts.

Pillar 4 — Factual Density & Authority (Trust + Reading Ease)

Legible sites averaged a Trust Score of 12.5/100 and a Reading-Ease Score of 39.9/100. The engine-comparison in Section 5 showed that the more conservative models reward exactly these signals.

- 1. Increase factual density:** include specific, verifiable statistics and figures rather than unquantified claims.
- 2. Cite authoritative outbound sources** (primary research, official bodies, `.gov` / `.edu` / `.ac.uk` domains, recognised industry references). Citari's published guidance, drawing on ACM SIGKDD 2024 research from Princeton and Georgia Tech, documents measurable citation lifts from adding statistics (+31%), quotations (+41%) and source citations (+28%), while keyword stuffing reduced visibility (-8%).
- 3. Add named attribution and quotes** from real, identifiable people to strengthen E-E-A-T signals.

4. Write in plain, concise language. Shorter sentences and simpler syntax raise the Reading-Ease Score and make content easier for an engine to parse and quote accurately.

Sequencing and measurement

Citari recommends remediating strictly in pillar order. A business that opens its Security Gate (Pillar 1) converts its latent brand recognition — the *floor* measured in Section 4 — into live, retrievable visibility; structured data (Pillar 2), comprehension (Pillar 3) and authority (Pillar 4) then raise the ceiling. Because the Shadow Citation Paradox shows that 39.4% of currently invisible firms already enjoy off-page recall, the access fix alone is frequently enough to begin appearing in live answers.

WORK WITH CITARI

Benchmark Your Own Estate

This whitepaper is built on Citari's audit of 160 West-of-Scotland businesses — primary research conducted with the Citari Audit Framework. The same framework can benchmark your own website against this dataset, identify exactly which pillar is costing you AI visibility, and quantify your AI Recommendation Share across Claude, GPT-4o and Gemini.

To commission an audit or discuss your AEO strategy, contact strategy@citari.co.uk or visit <https://citari.co.uk>.

Human website. Machine Search.

Methodology note

All dataset-derived figures in this paper are computed by Citari from a deduplicated set of **160 West-of-Scotland businesses** (163 audited records, deduplicated on normalised URL). Stage 1 (technical crawl) reads the raw served HTML with a non-rendering HTTP client and does not execute JavaScript. Stage 2 (citation simulation) queries Claude Sonnet 4.5, GPT-4o and Gemini 2.5 Flash in plain text-completion mode with no web search, retrieval or tool use; a citation is a case-insensitive substring match of the business name and therefore measures latent (training-data) recognition, not live retrieval. One known limitation follows from this design: businesses whose names are short, common English words can register incidental substring matches when a model uses the word in ordinary prose, so per-business citations for such names are treated as indicative (see the PR & Marketing note in Section 6 for the clearest example). As a sensitivity check, Citari reviewed all 61 cited businesses and found only two — “Frame” and “Wire” — with short common-word names at material risk of incidental matching. Excluding both shifts the headline figures only marginally: businesses cited by at least one model move from 38.1% to 36.9% (59 of 160), Gemini from 30.0% to 28.7% (46 of 160), GPT-4o from 15.0% to 14.4% (23 of 160), Claude unchanged at 9.4%, and the shadow-citation rate from 39.4% to 38.5% of empty crawls. Every finding in this paper holds under that adjustment, so the as-published figures — the literal output of the documented method — are retained throughout. The four 0–100 scores are computed from raw HTML, and all per-sector score averages are taken over `ok` rows only. External market statistics are drawn from Citari's published guide *Unlocking the AI Search Frontier* and its cited primary sources (Seer Interactive, Pew Research Center, Semrush, ACM SIGKDD 2024) and are used for context only.