

Audits that hold up to a second look.

I run technical, on-page, schema and AI-search audits, and every finding is checked by hand before it reaches you. These are real engagements from 2026. Clients are described by sector only, and nothing here claims traffic or ranking numbers — you are seeing what I found, how I checked it, and what changed when fixes shipped.

8

case studies

142

findings documented

INSIDE

1. **D2C packaged-food brand** D2C food e-commerce, India
2. **Web and AI development studio** Web development studio, India
3. **NeoArk Digital (my own agency)** Digital marketing agency, Chennai
4. **B2B software company, energy sector** B2B SaaS, renewable energy
5. **B2B telecom provider** B2B telecom, India
6. **Wealth advisory firm** Financial advisory, India
7. **Design-led e-commerce brand** D2C and B2B e-commerce, India
8. **Events and catering business** Events and catering, South India

Also inside: SEO 100X, my evidence-graded research library, and my keyword method.

Ramakrishnan S

Freelance SEO: technical, content and AI-search audits
ram-resume-nine.vercel.app · ramakrishnan15126@gmail.com

The process, every time

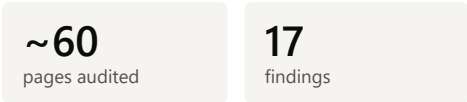
1	Technical audit	An agentic audit of the whole site: indexation, canonicals, schema, performance, images, AI-search readiness — and the on-page copy checked for factual errors.
2	Remediation spec	Prioritised fixes, each with a way to prove it landed. Implemented by your developer, or by AI under my direction.
3	Competitor gap map	Five to twenty brands in your niche: what is working for them, what page types they run, and what is missing on your site.
4	90-day plan	Twenty to fifty target keywords with real demand behind them, grouped into clusters and committed to a roadmap.
5	Execution	New pages, articles, title and meta rewrites, and on-page changes across the ninety days.
6	Re-run	The audit runs again against the baseline, and the loop repeats one cluster at a time so authority compounds.

D2C packaged-food brand

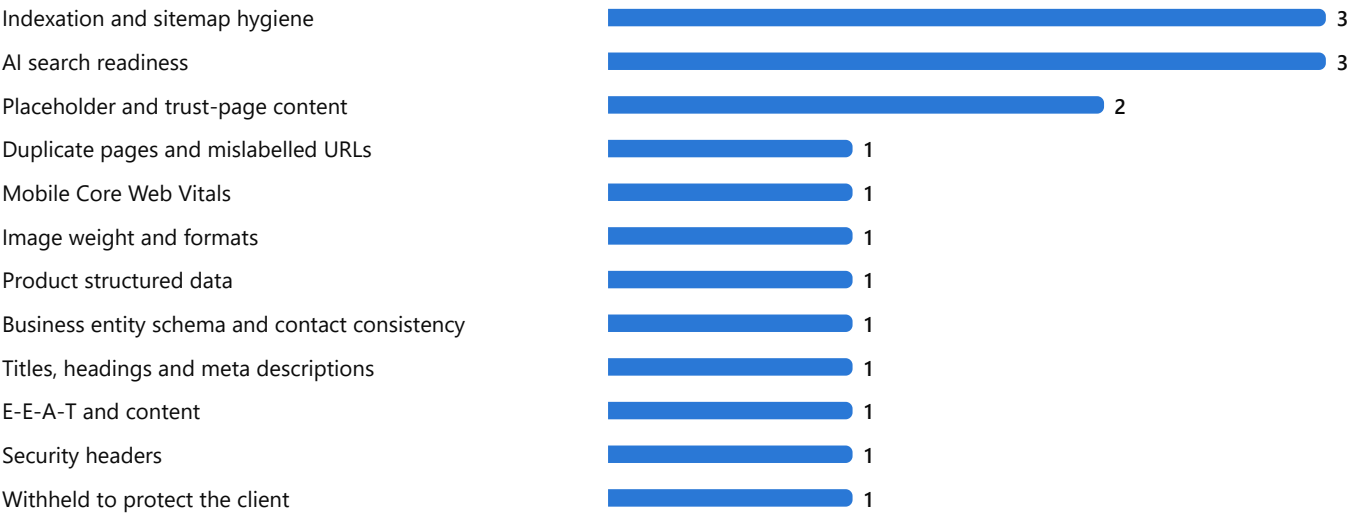
D2C food e-commerce, India · WordPress + WooCommerce · Aug 2026

Full technical, content, schema and AI-search audit, then two re-audits to track what the client's team fixed, followed by keyword research and a competitor gap analysis.

Audit delivered, then re-audited twice.



FINDINGS BY CATEGORY



BY SEVERITY



WHAT THE AUDIT FOUND

- Most of the URLs submitted in the sitemap were near-empty tag archives — indexable, self-canonical, and crowding out the real catalogue.
- A live, indexable FAQ page was still template placeholder text.
- The legal terms page, copied from a template, named a different business as the contracting party.
- Several product URLs described a different product from the one the page sold.
- Mobile LCP ran two to three times over Google's 2.5-second threshold, and the site published four different phone numbers across its text, schema and chat widget.

CHECKED BEFORE IT SHIPPED

Four first-draft claims were withdrawn or corrected before the report was released — including two 'duplicate product' pairs that had only been inferred from similar-looking URLs.

WHAT CHANGED



My audit health score, a weighted baseline across seven categories rather than an external metric.

Fixes implemented by the client's own team — 11 of the 17 findings, fully or largely.

Twelve days later the second re-audit found the score holding at 65, with every fix still in place.

DELIVERED

- Client-facing audit report, findings ordered by severity
- Prioritised action plan with effort sizing
- Two re-audit reports with before-and-after scorecards
- Keyword research — shipped without volume figures when the data was unavailable, and saying so
- Competitor gap analysis

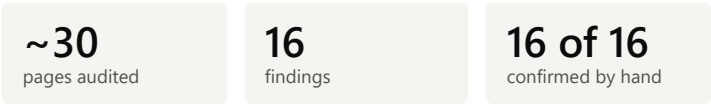
Anonymised: the client is described by sector only. No client name, website address, page URL or quoted copy is included, and no traffic, ranking or conversion figure is claimed.

Web and AI development studio

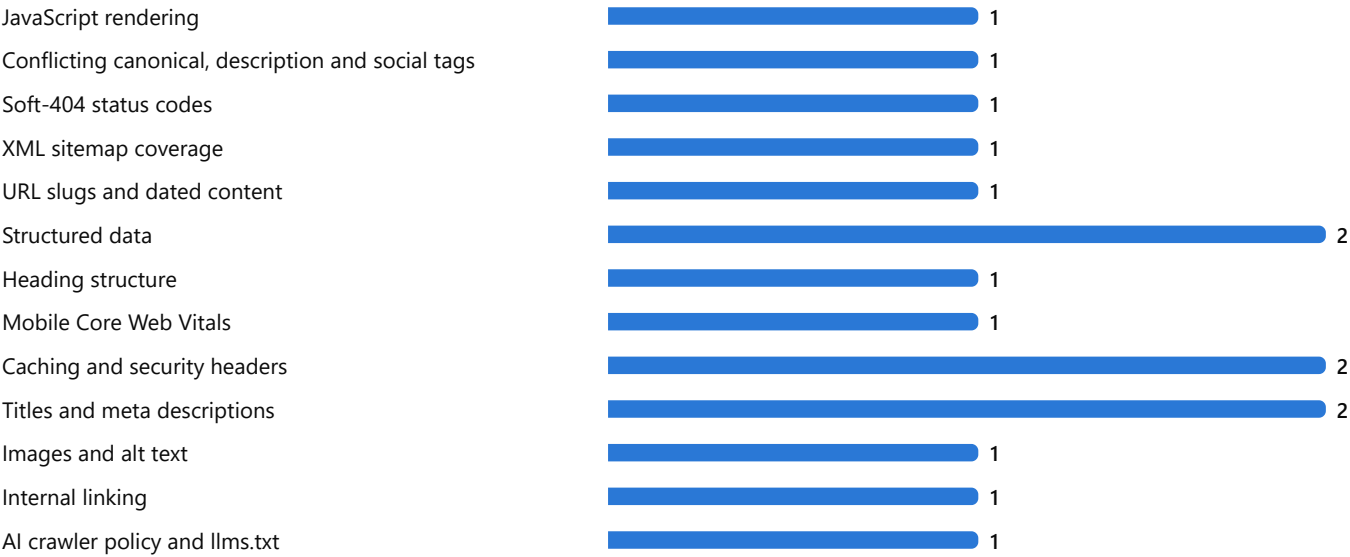
Web development studio, India · React single-page app, no server-side rendering · Jul – Aug 2026

Full technical audit of every reachable URL. My pre-send check caught five false findings, so I withdrew the draft and reissued it corrected — then a full SEO and AI-search report with a 90-day proposal, a topical map and a 30-day blog plan.

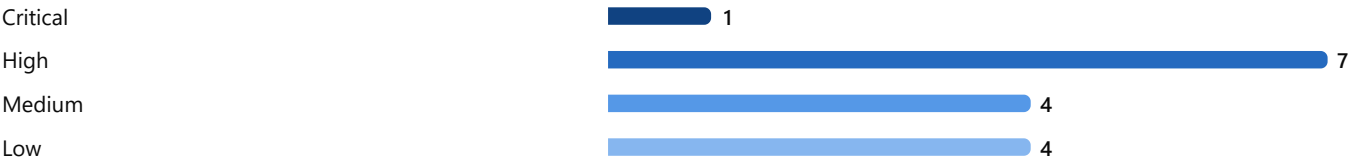
Audit delivered; fixes not yet implemented.



FINDINGS BY CATEGORY



BY SEVERITY



WHAT THE AUDIT FOUND

- Nothing was server-rendered: to any crawler that does not run JavaScript — including most AI crawlers — the site was empty.
- Every page shipped two conflicting sets of canonical, description and social tags.
- Unknown URLs returned HTTP 200 instead of 404, so any mistyped address could be indexed.
- The sitemap left out several indexable pages, including the longest content on the site.
- No page tested passed mobile LCP.

CHECKED BEFORE IT SHIPPED

Before sending, my verification pass found five false findings: the measurement script had read only the first matching tag on each page. I withdrew the finished draft and reissued it corrected.

DELIVERED

- Corrected developer-facing technical report, 16 issues
- Verification and correction log
- Full SEO and AI-search report with a 90-day proposal
- Topical map and blog strategy
- 30-day blog plan
- Keyword volume and difficulty research

Anonymised: the client is described by sector only. No client name, website address, page URL or quoted copy is included, and no traffic, ranking or conversion figure is claimed.

NeoArk Digital (my own agency)

Digital marketing agency, Chennai · React single-page app, rebuilt on Next.js · Aug 2026

A baseline audit of my own agency's site across six specialist passes, cross-checked against each other, then a full re-audit after the site was rebuilt, with an action plan for each.

Audited in August 2026, then re-audited after the rebuild.

5

pages audited

39

findings

AREAS CHECKED

- Technical SEO: crawlability, indexability, redirects
- Content quality and E-E-A-T
- Titles, descriptions and headings
- Structured data
- Performance and Core Web Vitals
- Images
- Site structure, navigation and accessibility
- AI search readiness

BY SEVERITY



WHAT THE AUDIT FOUND

- Unknown URLs returned HTTP 200 and canonicalised to the homepage — a sitewide soft-404.
- No body content in the server HTML: every word was rendered by JavaScript, so non-rendering AI crawlers saw an empty page.
- Blog articles were locked inside button-driven accordions with no URLs of their own, so they could not be indexed or cited.
- Blog index images shipped as oversized PNGs, dragging down mobile LCP.
- The homepage promoted articles that did not exist on the site.

CHECKED BEFORE IT SHIPPED

Across the two audits, more than a dozen automated tool warnings were rejected on verification before anything was written into a report.

DELIVERED

- Full audit with per-finding verification
- Six specialist audit passes, cross-checked
- Re-audit and a fresh action plan after the rebuild
- Proposed structured-data blocks and a draft llms.txt
- Site architecture and keyword map
- Competitor homepage audit

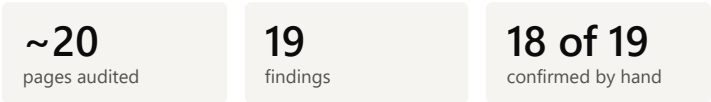
Anonymised: the client is described by sector only. No client name, website address, page URL or quoted copy is included, and no traffic, ranking or conversion figure is claimed.

B2B software company, energy sector

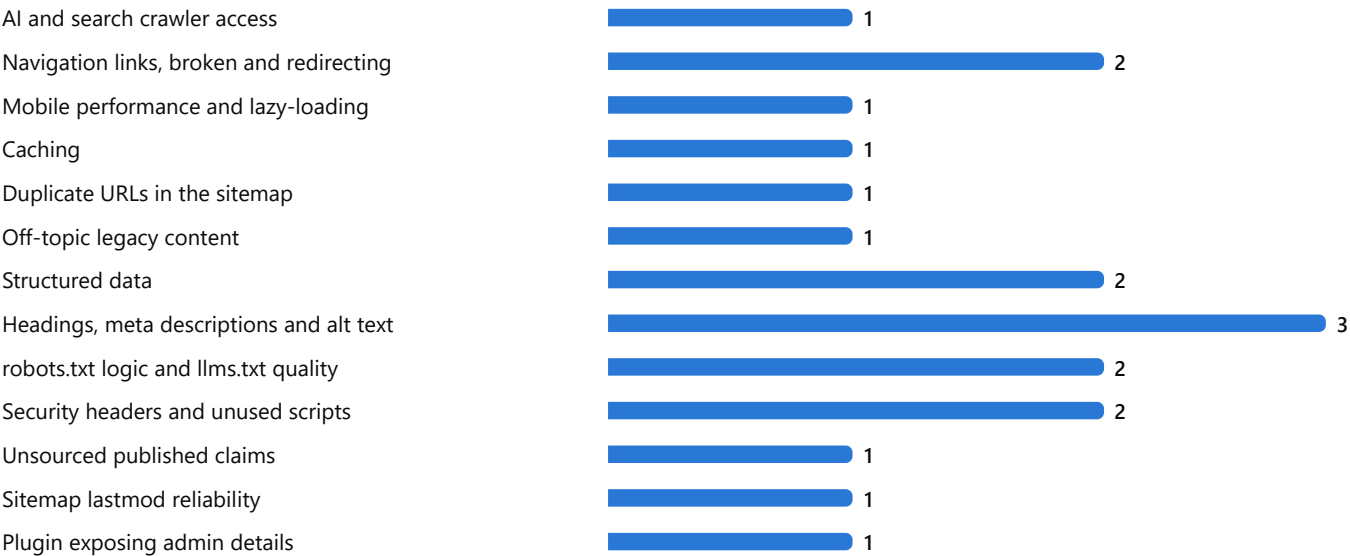
B2B SaaS, renewable energy · WordPress + Elementor · Aug 2026

Full technical, schema, performance and AI-search audit with a phased action plan. When the top finding was challenged, I re-tested it against independent third-party crawl data — it held.

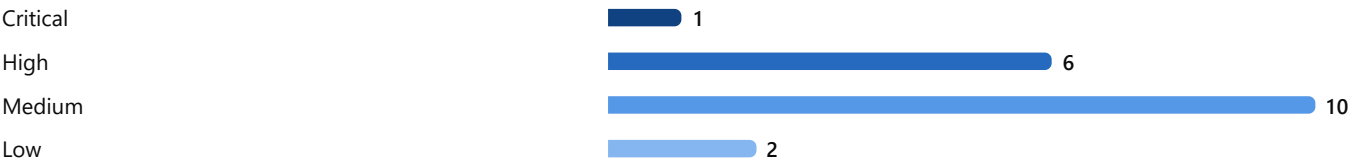
Audit delivered; fixes not yet implemented.



FINDINGS BY CATEGORY



BY SEVERITY



WHAT THE AUDIT FOUND

- robots.txt invited AI crawlers in, but the host's firewall turned them away before they could read it.
- Mobile LCP was far over threshold: every image was JavaScript lazy-loaded, and admin-only templates shipped on public pages.
- A site migration left duplicate URL pairs in the sitemap, each claiming to be canonical.
- The main navigation pointed at a malformed hostname and at retired pages.
- A share of the blog still covered the company's pre-pivot business, diluting its topical focus.

CHECKED BEFORE IT SHIPPED

Two candidate findings were dropped in verification — one was a redirect the checker had followed and mistaken for a live duplicate page.

DELIVERED

- Audit report with severity and confidence labels
- Phased action plan: this week, the next two weeks, this quarter
- Pre-change baseline for measuring what changes

Anonymised: the client is described by sector only. No client name, website address, page URL or quoted copy is included, and no traffic, ranking or conversion figure is claimed.

B2B telecom provider

B2B telecom, India · WordPress with a page builder and SEO plugin · Aug 2026

Full-site technical audit, then an organic growth plan and a topical-authority cluster architecture, with a forecast model built on leading indicators rather than invented targets.

Audit delivered; no live-site changes recorded.

~175

pages audited

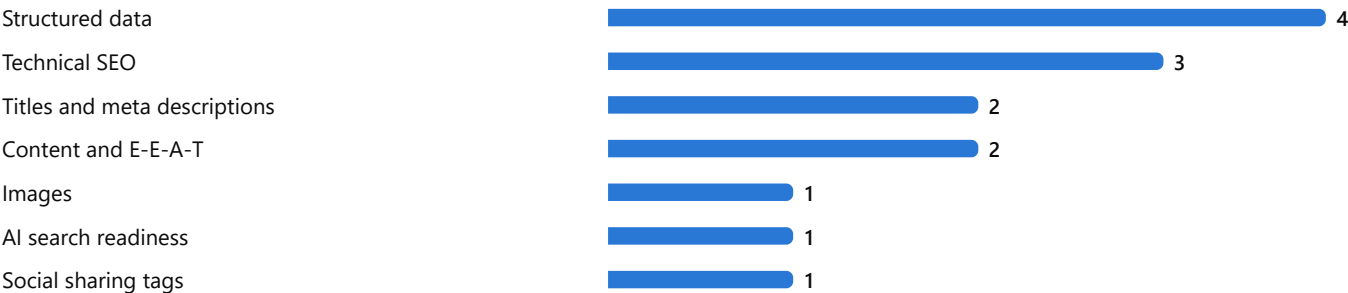
14

findings

13 of 14

confirmed by hand

FINDINGS BY CATEGORY



BY SEVERITY



WHAT THE AUDIT FOUND

- Structured data declared the site's identity as both a person and an organisation, with no logo or contact details.
- A self-serving star rating was marked up on the homepage — a review-snippet policy and manual-action risk.
- FAQ structured data sat on core commercial pages whose questions did not appear on the page.
- Newer, deeper pages were missing from the XML sitemap while retired redirecting URLs were still listed.
- Almost no editorial internal linking between hub pages and the pages beneath them.

CHECKED BEFORE IT SHIPPED

Three automated warnings were rejected before publishing — including a 'duplicate title on every page' that turned out to be the titles inside SVG icons.

DELIVERED

- Technical audit with a prioritised action plan
- Organic growth plan diagnosing four causes
- Topical-authority cluster architecture
- Forecast model built on leading indicators
- Pre-change baseline

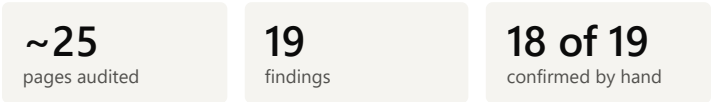
Anonymised: the client is described by sector only. No client name, website address, page URL or quoted copy is included, and no traffic, ranking or conversion figure is claimed.

Wealth advisory firm

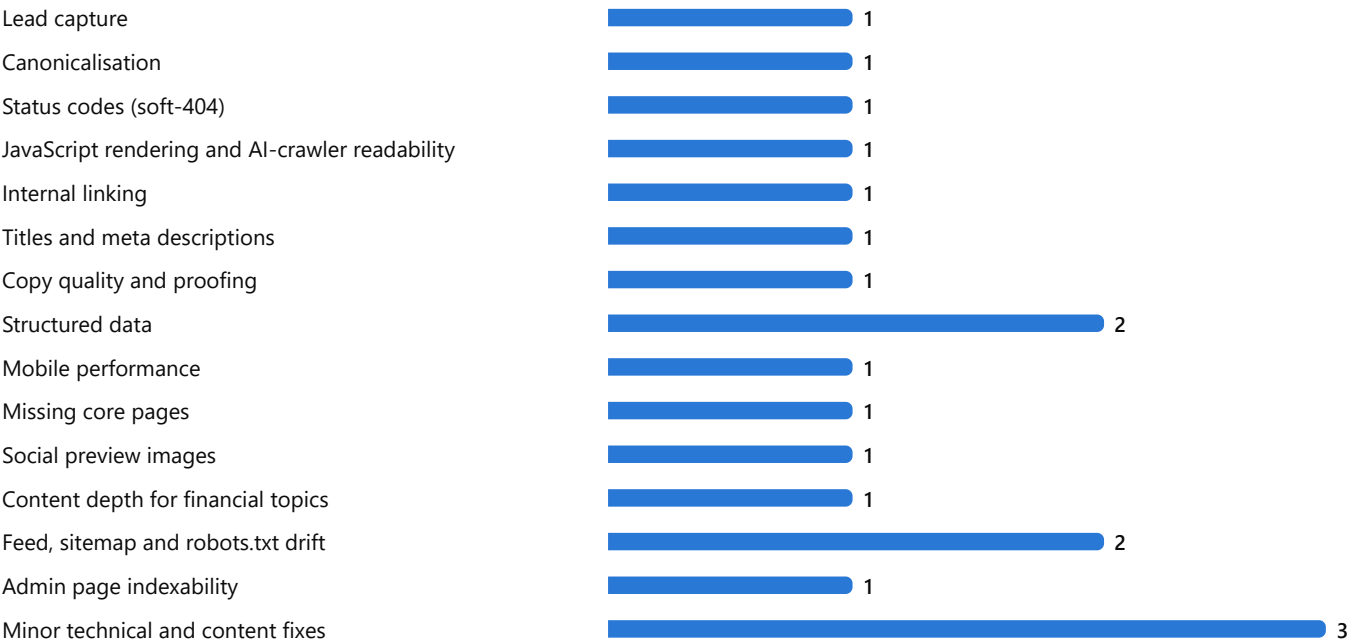
Financial advisory, India · React single-page app · Aug 2026

Technical SEO audit plus a content-accuracy review, turned into a plain-language client report and a call brief for walking the client through it.

Audit delivered; implementation not recorded.



FINDINGS BY CATEGORY



BY SEVERITY



WHAT THE AUDIT FOUND

- The contact and callback forms were silently rejecting every submission — enquiries were being lost.
- Client-side JavaScript overwrote correct canonicals on most articles with stale addresses that led to a 'not found' page.
- Every non-existent URL returned an indexable 200 page, which is why the dead canonicals went unnoticed.
- All body content rendered client-side: crawlers that do not run JavaScript received zero words.
- The homepage had no crawlable internal links — the whole navigation was built from buttons.

CHECKED BEFORE IT SHIPPED

An automated scan flagged the opening-hours schema as malformed. It was the format the specification requires, and it was rejected before the report went out.

DELIVERED

- Plain-language client report, 19 findings
- Content and accuracy addendum
- Call brief for walking the client through the findings
- Sequenced work plan, split into developer work and content work

Anonymised: the client is described by sector only. No client name, website address, page URL or quoted copy is included, and no traffic, ranking or conversion figure is claimed.

Design-led e-commerce brand

D2C and B2B e-commerce, India · Wix · Aug 2026

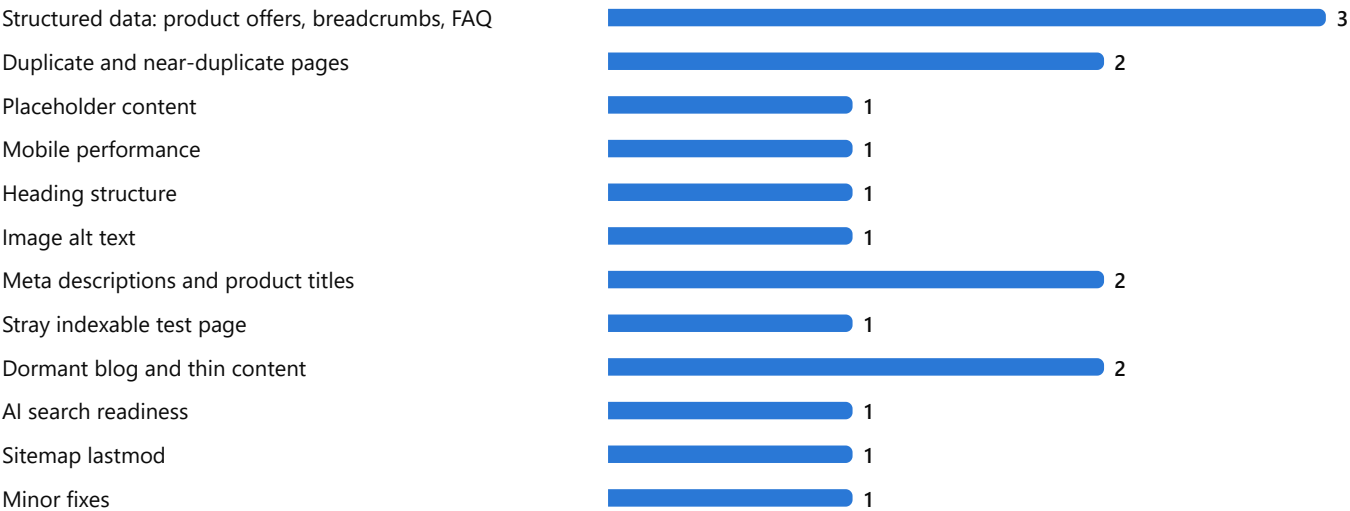
Full-site technical, on-page, content, schema and performance audit, with every content page crawled and a random sample of product pages checked.

Audit delivered; implementation not recorded.

~55

pages audited

FINDINGS BY CATEGORY



BY SEVERITY



WHAT THE AUDIT FOUND

- Product offer markup used the wrong property casing, so price and availability were unreadable on every product — traced to a custom override, not the platform.
- The live FAQ page was unedited site-builder placeholder text, and it was in the sitemap.
- Mobile LCP was well over Google's threshold; the server was fast, so the cause was client-side script weight.
- Most pages had no H1 and most images had empty alt text, including every thumbnail on the main category page.
- Gallery pages carried identical body text, and a retired copy of the main category page was still in the sitemap.

CHECKED BEFORE IT SHIPPED

One apparent crawler block turned out to be rate-limiting caused by the audit itself, and was rejected before it reached the report.

DELIVERED

- Full audit report with a weighted health score
- Prioritised action plan: immediate, quick wins, strategic
- Unknowns and follow-ups table
- Pre-change baseline

Anonymised: the client is described by sector only. No client name, website address, page URL or quoted copy is included, and no traffic, ranking or conversion figure is claimed.

Events and catering business

Events and catering, South India · Hand-coded static HTML · Aug 2026

Full-site technical and on-page audit of a small brochure site, with a plain-language client report and a script for the walkthrough call.

Audit delivered; the site files were not accessible, so nothing has been implemented.

under 10
pages audited

18
findings

18 of 18
confirmed by hand

AREAS CHECKED

- Technical SEO: crawl files, canonicals, duplicate URLs
- Conversion path: the enquiry form
- Titles, meta descriptions and headings
- Structured data
- Performance and Core Web Vitals
- Images
- Content
- Local SEO signals
- Measurement
- AI search readiness

BY SEVERITY



WHAT THE AUDIT FOUND

- The site's only enquiry form could not submit anything — confirmed four independent ways.
- No robots.txt, no XML sitemap, no structured data and no canonical tags, with the homepage reachable at several duplicate addresses.
- No analytics installed at all, so there was no way to measure a before and after.
- Mobile LCP of several seconds on a very heavy page, while desktop scored perfectly.
- No machine-readable address, click-to-call link or map for a business that serves a local area.

CHECKED BEFORE IT SHIPPED

Two tool findings were dropped after byte-level re-testing — apparent text corruption was a missing charset header, not broken content.

DELIVERED

- Plain-language client report with self-check boxes
- Internal technical audit with per-finding verification
- Call-preparation script
- Three-stage action plan
- Dated changelog, including withdrawn findings

Anonymised: the client is described by sector only. No client name, website address, page URL or quoted copy is included, and no traffic, ranking or conversion figure is claimed.

SEO 100X research library

SEO 100X is my evidence-graded SEO, AEO and GEO knowledge base, built for agency work on a free-tool stack in India. Every claim carries an evidence tier from T1 to T4, disagreements between sources are ruled on once and recorded, and advice that fails review is quarantined rather than quietly kept.

532

claims graded

11

validated notes

10

source conflicts ruled on

5

SOPs

EVERY CLAIM CARRIES A TIER

- T1** Primary or controlled: Google documentation, court exhibits, controlled experiments, own measured tests. Can drive a decision alone.
- T2** Large independent observational data with no product stake. Can drive a decision when direction is corroborated.
- T3** Vendor or interested party; incentive flagged, effect sizes treated as upper bounds. Never decisive alone.
- T4** Practitioner consensus or anecdote. Useful for hypotheses, never a basis for a recommendation.

FIVE FINDINGS THAT CHANGE HOW I WORK

1. **Formatting is hygiene, not a growth lever.** Two independent controlled studies find formatting-only edits have little effect on AI citation, and Q&A formatting specifically showed no absorption benefit. The controlled evidence ranges from no significant effect to roughly +17%, far below the several-hundred-percent figures sold by vendors. The library's own audit flagged that presenting +17% alone as 'the real effect' overstates it, so the range is the honest statement.
2. **The famous +40% GEO result is misread.** The widely quoted +40% visibility gain measures answer share for content already sitting in a model's context. A systematic review of 45 studies rules that it establishes neither organic discoverability nor traffic. It should never be quoted as an expected client outcome.
3. **No technique reliably drives durable AI discoverability.** The 45-study systematic review found no technique with a stable, cross-platform causal effect on getting discovered or on downstream clicks and conversions. The defensible offer is eligibility, crawler access and measurement, not promised citations. The library quarantines its own AEO/GEO recommendations until they are re-derived from this evidence.
4. **Google assesses quality at the site level.** Court exhibits describe Google's quality signal (Q*) as largely static and related to the site rather than the query, so sitewide cleanup outranks per-page tinkering. No scale or threshold is evidenced. An earlier version claimed a 0-1 scale with a 0.4 cutoff for snippets; that came from an agency's illustrative example, was not in the exhibit, and has been deleted.
5. **AI citation has decoupled from top-10 ranking.** A large independent SERP study (n=863K) puts the share of AI Overview citations drawn from the organic top 10 at about 38%, down from 76% a year earlier, with the ruling stated as a 20-40% range depending on query mix. Rank tracking therefore no longer measures AI visibility, and the two should be reported as separate KPIs. The underlying figure dates from March 2026 and is past the library's own 6-month staleness window, so it needs re-checking before reuse.

Auditing my own library

I ran an adversarial review of the whole library: four independent AI verifiers, each hunting a different failure mode and told to default to FAIL when unsure, then a judge.

458

items checked

78

findings

9

critical errors fixed

6

still flagged, openly

WHAT I GOT WRONG, AND FIXED

1	A quality-score claim (0-1 scale, 0.4 cutoff for snippets) was tagged as court testimony but came from an agency's illustrative example, and had spread to a headline finding and a client-refusal rule.	Claim deleted, correction recorded in the file, and the headline finding rewritten.
2	A conflict ruling contained a fabricated verbatim quote attributed to Google's spam policies page.	The page was re-fetched and Google's actual wording substituted.
3	One competitor's sitemap file actually profiled an unrelated company (a plastic water-tank manufacturer) because its sitemap was cross-domain and hosts were never checked.	The collector now validates hosts and warns; the file was re-collected at 0 valid URLs.
4	One brand's architecture file double-counted 67 URLs across overlapping sitemaps, reporting 302 URLs and 128 blog URLs instead of 234 and 64.	Deduplication was added to the collector and all files re-collected.
5	The SERP dataset's caveat divided 720 rows by a 389-query universe (317 never searched), forbidding position claims the data supports and permitting frequency claims it does not.	The caveat was rewritten to the verified shape: a gapless top 10 for 72 queries.
6	The free-stack measurement SOP gave a non-AI CTR formula that was biased upward and cannot be computed, since Search Console reports AI impressions but no AI clicks.	Replaced with an explicit bounded method using a lower and an upper bound.
7	Six case studies graded T2 were agency self-published, which matches the library's own T3 definition.	Downgraded to T3 with the reason recorded.
8	The indexing file dropped a hedge from Google's wording and added an editorial line as if it were Google's.	The hedge was restored and the editorial addition removed.
9	The README and the retrieval skill routed to a directory that no longer existed, so the worked example pointed at missing files.	The map and routing were rewritten against the real file tree.

STILL OPEN, FLAGGED RATHER THAN HIDDEN

- The AEO/GEO layer failed logical soundness and is quarantined, not fixed
- A local-search vertical file applies a B2C product-review study to Indian service businesses without testing that transfer
- A citation-mechanics header claims unanimous results across six models; its own table shows one factor significant in 3 of 6
- The +17.3% formatting figure is still promoted as 'the real effect' where the honest statement is a null-to-+17.3% range
- A pricing model rests on a utilisation figure set above the only observed value; its retainer and hiring numbers are arithmetic of chosen inputs
- A collection run concluded there is no authority incumbent in the niche despite collecting no traffic, backlink or authority data

THE LESSON

Gathering the evidence was clean; the failures happened in the step from evidence to recommendation, where a study was graded honestly and then acted on as if the grade were higher. Rigour at intake has to be re-applied at the point of advice.

THEN I WENT TO THE PRIMARY SOURCES

Every ranking-internals claim had rested on trade-press summaries of court testimony, so the four primary exhibits were downloaded and read directly. This confirmed NavBoost's 13-month window, the ABC topicality signals and the site-level nature of the quality signal, added new detail such as location and device segmentation, and demoted five widely repeated claims that do not appear in the exhibits. An exhibit records Google's internal decision that the nosnippet control limits use of content for AI grounding, not just display, and that it chose not to say so publicly. That changes what the control can be promised to do for a publisher.

My SOPs

AI CRAWLER ACCESS CONTROL

An audit of a site's robots.txt against AI training, search and user-triggered crawlers, with a recommended default posture put to the client as a business decision.

1. Audit AI crawler access first on every new client
2. Separate training crawlers from search crawlers; the controls are independent
3. Verify each user agent against the vendor's own documentation, not blog lists
4. Allow the search crawlers by default so AI search visibility is not blocked
5. Put the training-crawler choice to the client explicitly
6. Verify real crawler traffic against vendor-published IP ranges before reading logs
7. Never present robots.txt to a client as a security control

Tools: robots.txt, Vendor crawler documentation and published IP lists, Server logs, Google Search Console

CLIENT KEYWORD WORKFLOW

A chain of five client-reviewable deliverables: a technical audit, a top-100 keyword report, a keyword research workbook, a 30-day content plan and a tracker.

1. Request Search Console exports and the business facts the scoring depends on
2. Collect demand evidence from Search Console, Google Autocomplete and forum threads
3. Build the candidate list with a uniform relevance rubric and written exclusions
4. Score with an objective-weighted model, holding unmeasured inputs neutral
5. Build the workbook from a script and verify it in Excel before sending
6. Plan 30 days in cluster weeks, running page edits and blogs on the same topic
7. Track one row per deliverable; judge edited pages at day 30, new pages at 60-90

Tools: Google Search Console, Google Autocomplete, Reddit, Quora, Excel, Python

COLD PROSPECT AUDIT

A sendable outside-in on-page, technical and claims-consistency audit of a prospect's site, where every finding survives the prospect checking it.

1. Crawl every sitemap URL rather than a sample
2. Test host variants, header-dependent redirects and a control 404 URL
3. For multilingual sites, decide whether a page is translated before judging its canonical
4. Check the site's own claims for inconsistent numbers, restated statistics and dates
5. Admit a finding only with command output, stated scope, a control and a verbatim Google source
6. Rate severity by impact, choosing the lower rating when unsure
7. Write each finding in plain English with a check-it-yourself line

Tools: Python crawl scripts, curl, Browser developer tools, PageSpeed Insights API, Google Search Central and web.dev documentation

KEYWORD RESEARCH ON A FREE STACK

A filterable keyword database narrowed to one cluster and prioritised into a 30-90 day sprint.

1. Order targeting by awareness stage, starting nearest to purchase
2. Prove demand from several sources rather than search volume alone
3. Gather ideas from Search Console, Keyword Planner, forums and People Also Ask
4. Add a location modifier only where the answer genuinely varies by place
5. Classify each keyword by current position and as primary, variation or secondary
6. Score on several variables, dropping inputs that need paid tools
7. Filter hard and sprint on one cluster

Tools: Google Search Console, Google Ads Keyword Planner, Google Autocomplete, People Also Ask, Reddit, Quora, Perplexity

MEASURING AEO/GEO WITH A FREE STACK

A client reporting frame with three separate KPIs: classic rank and traffic, AI citation impressions, and AI answer share.

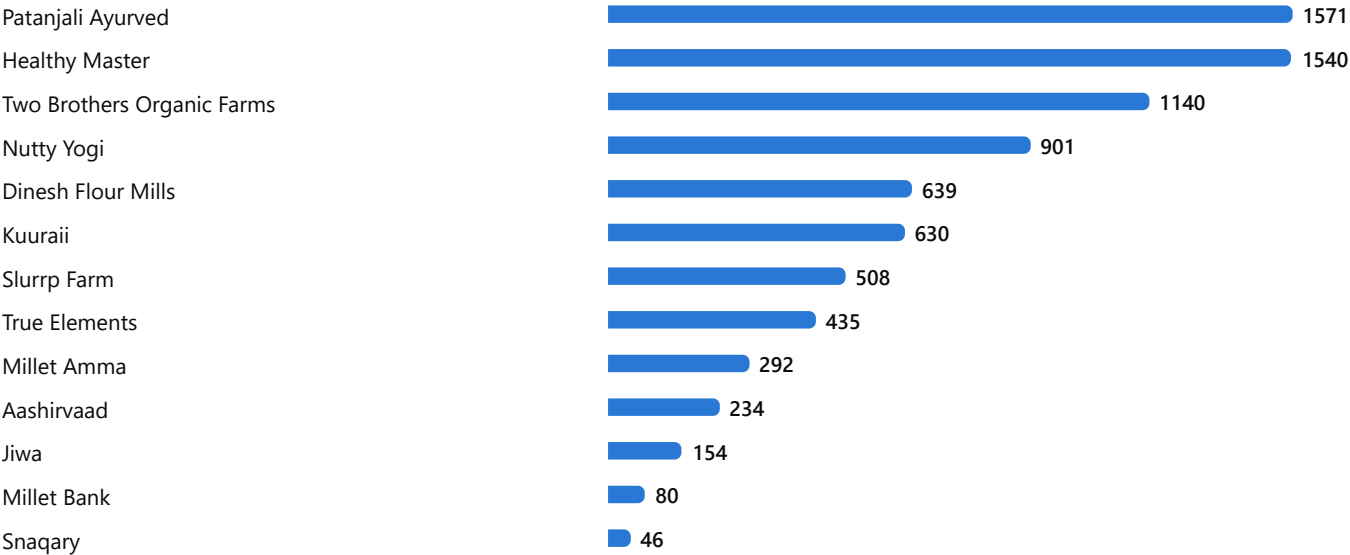
1. Use Search Console's generative AI report for page-level AI impressions
2. State its hard limits: no clicks, no CTR, no position, no query dimension
3. Remember AI impressions are already inside Performance totals, not additive
4. Report non-AI CTR as a bounded range, never a single computed figure
5. Run a fixed prompt panel across AI assistants and use rolling medians
6. Segment server logs by AI crawler user agent
7. Check Bing Webmaster Tools indexing and report the three KPIs separately

Tools: Google Search Console generative AI report, Bing Webmaster Tools, Server logs, Manual prompt panel across ChatGPT, Perplexity, Gemini and Claude

Niche research: millet D2C, India

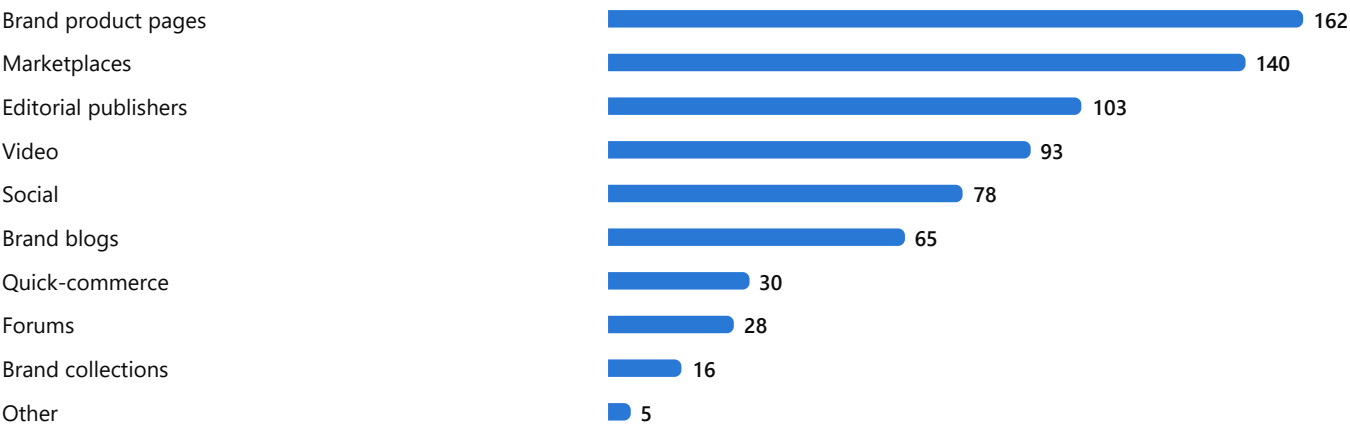
A snapshot of how brands in one niche build their sites and what Google shows for its searches, collected and re-verified on 7 Sep 2026. Brands are named from their public websites. No traffic, backlink, volume or authority data was collected, and no ranking position is claimed for any brand.

SITEMAP SIZE BY BRAND



WHAT GOOGLE SHOWED IN THE TOP 10, BY RESULT TYPE

720 results: a complete top 10 for 72 of the 389 queries collected



WHAT THIS DATA CAN SUPPORT

Supports position-level statements within those 72 queries only, not niche-wide market share. About a fifth of results were international sites despite an India setting, so India conclusions need that filtered.

SINGLE-PAGE TEARDOWNS

- Orggu: Teardown only: millet flour collection page of about 1,440 words; Long-form copy below the product grid, including a nutrient comparison table, buying checklist, regional-names table and FAQ; Richest schema of the pages captured: CollectionPage, ItemList, Product, BreadcrumbList and FAQPage; Collection URLs use keyword-phrase slugs
- Advait Living: Teardown only: multi-millet product page of about 4,380 words; Shows price and per-100g unit price in the first viewport; Includes a numbered making-process section, FAQ, reviews and links to related blog posts; Carried Product, Offer and FAQPage schema
- Manna Foods: Teardown only: multigrain atta product page of about 280 words; Full e-commerce schema stack (Product, Offer, Brand, BreadcrumbList) despite little visible body copy; Product benefit copy appeared in the meta description rather than the page body
- Prakruti Farms: Teardown only: multi-millet flour product page of about 860 words; Opens with labelled specification bullets and lists each millet with regional-language names; Gives specific cooking guidance with mix ratios; carried Product and Offer schema
- Khadar Groups: Teardown only: multi-millet flour product page of about 1,940 words; Uses the full long product title as the H1; Carried Product, Offer and BreadcrumbList schema; health benefits described qualitatively

DATA FIXED BEFORE IT WAS USED

- One brand's sitemap was cross-domain and pointed to an unrelated manufacturer's site; the collector now validates hosts and the file was re-collected at 0 valid URLs, so that brand is excluded here
- One brand's sitemaps overlapped and double-counted 67 URLs; deduplication was added and the corrected figures (234 URLs, 64 blog URLs) are used here
- All other architecture maps were re-crawled after the collector fix and came back unchanged
- The SERP caveat was rewritten from a false 1.85-results-per-query figure to the verified 72-query, gapless top-10 shape
- Six case studies were downgraded from T2 to T3 because they were agency self-published
- One URL on a different host was dropped from a brand's sitemap map during host validation

Keyword research pipeline

How I turn a site's own search data, Google Autocomplete and real forum questions into a scored top-100 keyword list, a page map and a 30-day plan — on free tools, and checked before it ships.

THE PIPELINE

- 00** On-page extraction: every page's title, meta description, canonical, headings, word count and schema.
 - 01** Demand discovery: niche seed phrases run through Google Autocomplete, as proof that people search them.
 - 02** Question mining: the same for question-shaped searches — best, how to, versus, is.
 - 03** Evidence index: Autocomplete proof, Search Console queries and forum threads merged into one lookup.
 - 04** Scoring: every hand-classified keyword scored on six weighted factors and given a verdict.
 - 05** Top-100 report: the ranked list, grouped by cluster, with ties handled at the cut-off.
 - 06** Related searches: what people also search for each keyword.
 - 07** Workbook: a formula-driven scoring engine, page map, clusters and a 30-day tracker, built in Excel.
 - 08** Workbook check: after Excel recalculates, every rank, score and verdict must match the scoring output exactly.
 - 09** Tracker check: every due date recomputed independently, and every task starting as not started.
-

HOW KEYWORDS ARE SCORED

Every keyword gets six sub-scores from 1 to 5, each multiplied by a weight. The heaviest weights go to proof that people actually search it and to where the site already stands in Google, so the plan starts with real demand and quick wins rather than guesses.

CHECKED BEFORE IT SHIPS

- Excel recalculates the workbook, then every rank, score and verdict is checked against the scoring output, row by row.
- Every sheet is scanned for formula errors before anything is sent.
- Every due date in the 30-day tracker is recomputed with an independent working-day calculation.
- The scoring weights live in two places, so the checker catches any drift between them.

Method only. The search data belongs to the site owners and is not published.

The State of Chennai SME Websites 2026

Original research. I sampled 161 Chennai small-business websites from OpenStreetMap and audited 125 of them, after throwing out a first sample from web search that returned businesses on three other continents. Sites that disallowed crawling were excluded, and only aggregate figures are reported — no individual business is named.

