THE FOUNDATIONS

The complete 2026 system for getting found: SEO, AEO, GEO, the tooling under all three, and the field guide for working on client sites, as one connected method. Five parts, cited throughout, built to be read by a human or handed to an AI whole.

RIGHTS AND EDITION

Edition 2026.07. First published July 2026 at selwynuy.dev/guides.

About this edition

Everything here is current as of July 2026 and cited to primary sources throughout. Search, AI answer engines, and the tools around them move fast: prices and platform menus drift, so verify anything time-sensitive against the vendor before you quote it. This book is independent and vendor-neutral. It is not affiliated with, authorized by, or endorsed by Google, OpenAI, Microsoft, or any product named inside.

CONTENTS

THE WHOLE SYSTEM

INTRO ONE FOUNDATION, THREE LAYERS

PART I SEO

PART II AEO

PART III GEO

PART IV THE STACK

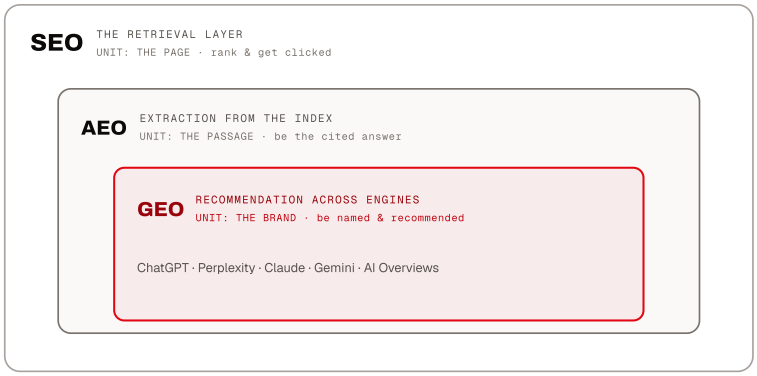
PART V THE CLIENT KIT

INTRODUCTION

ONE FOUNDATION, THREE LAYERS

IN PLAIN WORDS

Three fancy names, one idea. SEO gets your page found by search engines, AEO makes one piece of your page the exact answer a search or chatbot quotes, and GEO gets your brand named and recommended by AI tools. They stack on top of each other, so you build the base first and the rest has something to stand on.



Concentric layers, not rivals: AEO builds on SEO, GEO builds on both. Each layer changes the unit you optimize.

HOW TO READ THIS BOOK

Five parts, one method. **Part I (SEO)** gets you found and ranked. **Part II (AEO)** makes you the passage the engine extracts. **Part III (GEO)** gets your brand named and recommended by AI. **Part IV (The Stack)** is the tooling that measures all three. **Part V (The Client Kit)** is how to work on a client's site safely. Read straight through, or jump to the part that matches where you are stuck.

SEO, AEO, and GEO are concentric layers on one shared foundation, not rival disciplines. **SEO** is the retrieval layer everything else sits on: every major answer and generative engine retrieves from a conventional crawler-built index. Google's index feeds AI Overviews, AI Mode, and Gemini; Bing's feeds ChatGPT search and Copilot; Perplexity runs its own crawler. **AEO** sits on top and changes the unit of optimization from the page to the passage: being the specific chunk an answer surface extracts and attributes. **GEO** sits on top of both and changes the unit again, from the passage to the brand: what generative engines already know about you, plus earned presence in the third-party sources they retrieve and cite. A page cannot win AEO while failing SEO basics, and GEO without AEO-shaped content gets retrieved but not cited. AI Overviews are the surface where all three meet: SEO makes you retrievable, AEO makes you the extracted passage, GEO makes you the named and recommended brand.

PART I

SEO

How search works in 2026, the technical layer you control, and the content system that ranks and gets cited.

THE STATE OF PLAY

SEARCH IS REPRICING, NOT DYING

IN PLAIN WORDS

Fewer people click through to real websites now, because Google and AI chatbots answer a lot of questions right on the results page. Search is not dying, it is changing shape. The new goal is to get your page mentioned inside those answers, and fast, plain, well-built pages are what make that possible.

SEO is going through its largest structural shift since mobile. SparkToro's 2026 clickstream study (Similarweb panel) found that 68.01% of US Google searches ended without a click to the open web in early 2026, up from 60.45% in 2024. Only 276 of every 1,000 searches now send someone to a site that is not Google's own. On queries where an AI Overview appears, Ahrefs measured a 58% lower clickthrough rate for the top-ranking page across 300,000 keywords.

That is the loss column. The other column: Seer Interactive's study of 5.47 million queries showed organic CTR on AI Overview queries rebounding 85% between December 2025 and February 2026, and pages cited inside the AI answer earning roughly 120% more organic clicks per impression than uncited pages. The channel is not dying, it is repricing. Raw traffic per ranking falls; citation becomes the new position one.

68%

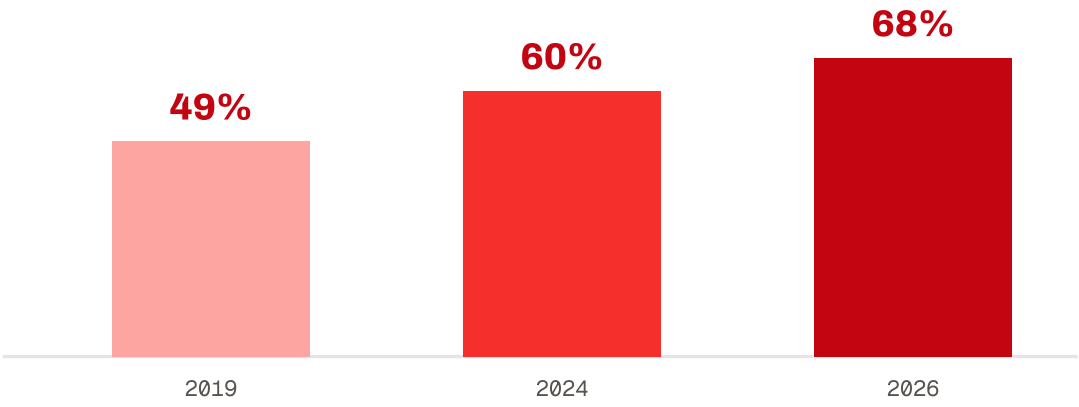
US Google searches ending with no open-web click, Jan-Apr 2026, up from 60.45% in 2024 (SparkToro / Similarweb panel, 2026).

-58%

CTR for the #1 result when an AI Overview is present, across 300K keywords (Ahrefs, Dec 2025 data).

+120%

More organic clicks per impression for pages cited inside the AI Overview vs uncited (Seer Interactive, 2026).



Share of US Google searches ending with no click to the open web (60.45% in 2024, 68.01% in 2026). SparkToro / Similarweb clickstream, 2026.

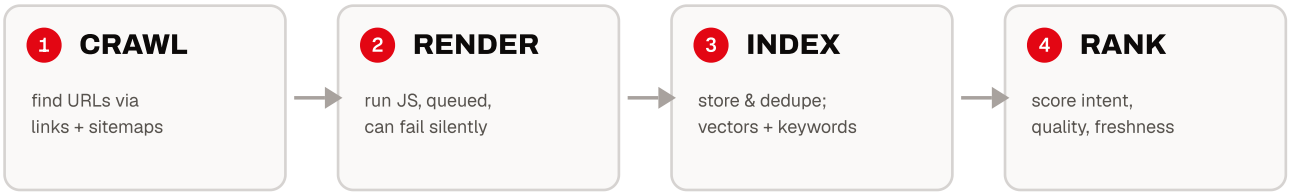
For a developer this is the moment technical fundamentals stop being commodity checklist work. Server-rendered HTML, clean indexing controls, fast pages, and clear entities now decide whether a site is retrievable by classic ranking and by AI answer engines at the same time. Google's own May 2026 guidance says answer and generative optimization "is still SEO". The discipline transfers; understand it deeply and skip the hacks.

THE MACHINE

CRAWL, RENDER, INDEX, RANK

IN PLAIN WORDS

Before Google can show your page to anyone, four things have to happen. A Google robot visits the page, loads it like a web browser would, files it away in a giant library, then decides where it should appear when someone searches. If any of those four steps breaks, your page simply never shows up.



Anything you want found must be in the server HTML: the render step is queued and can fail silently. Source: Google, How Search Works.

Stage 1: Crawling

There is no central registry of URLs. Googlebot discovers pages through links from known pages and submitted XML sitemaps. Crawling is budgeted: crawl capacity (how hard Google can hit your server) times crawl demand (how much it wants your URLs, driven by popularity and staleness). Budget is a non-issue below roughly a million URLs, but infinite URL spaces like faceted navigation, calendar pages, and session parameters can burn it on any site.

Stage 2: Rendering

Googlebot works in two waves: it parses the raw HTML response first (links, metadata, visible text), then queues the page for the Web Rendering Service, an evergreen Chromium that executes JavaScript. The queue delay is variable and rendering can fail silently. The safe engineering position: everything you want indexed lives in the initial HTML response.

Stage 3: Indexing

Google canonicalizes duplicates (one URL per document cluster), extracts content and annotations, and stores documents in an index that pairs the classic inverted index with dense vector representations. Semantically related pages are retrievable without exact keyword overlap.

Stage 4: Ranking and serving

Query understanding models interpret intent; hundreds of signals score topicality, quality, freshness, usability, and context. Two non-documentation sources filled in the picture: in the DOJ antitrust trial, Google VP Pandu Nayak confirmed under oath that Navboost, fed by roughly 13 months of aggregated click data, re-ranks results, and the

2024 Content Warehouse leak showed fields like goodClicks, badClicks, and lastLongestClicks. Leaked fields prove the data is collected, not how it is weighted today.

RULE

Google measures whether your result satisfied the click. Titles and snippets must honestly match page content, and the page must resolve the query fast. A visitor bouncing back to the results page is evidence against you.

YOUR LAYER

TECHNICAL SEO, MAPPED TO NEXT.JS

IN PLAIN WORDS

This is the plumbing part of SEO, the part you fully control in the code. It means making sure your words are actually inside the file the server sends out so robots can read them, telling Google which web address is the real one, and giving every page its own clean title. Get this plumbing right and everything else you do has a chance to work.

Rendering strategy is decision number one

Content must exist in the server HTML response. React Server Components, SSG, and ISR all produce indexable HTML by default. Pure client-side rendering, a "use client" page fetching data in useEffect, gambles on the render queue. Verify it the mechanical way: curl the URL and grep for your content, compare raw vs rendered HTML in Search Console's URL Inspection, and crawl the site twice in Screaming Frog (JS rendering off, then on) and diff the results.

Metadata and indexing controls

Use `generateMetadata` for title, description, canonical (`alternates.canonical`), robots directives, and Open Graph on every dynamic route. Ship `app/sitemap.ts` and `app/robots.ts` . Return real HTTP 404s with `notFound()` ; a 200-status "not found" component is a soft 404 and gets deindexed unpredictably. Moved content redirects with 301/308 in `next.config` or middleware, never a client-side `router.push` .

Duplication and canonicalization

One canonical URL per document: pick a trailing-slash policy, lowercase paths, strip tracking parameters. Faceted navigation needs a deliberate policy (canonicalize filter combinations to the base category, or noindex them). For i18n, emit hreflang via `alternates.languages` and link every language version to all others.

Crawl hygiene

robots.txt controls crawling, not indexing: a blocked URL can still be indexed from links, and a noindexed page must stay crawlable for the directive to be seen. Paginate with real `<a href>` links, not button-only onClick handlers, because Googlebot does not click. Monitor with Search Console's Page Indexing report; "Crawled, currently not indexed" and "Duplicate without user-selected canonical" are the rows that matter.

SPEED AS A SIGNAL

CORE WEB VITALS

IN PLAIN WORDS

Google measures three things about how your page feels to a real visitor: how fast the main content shows up, how quickly the page reacts when someone taps or clicks, and whether things jump around while it loads. Keep all three quick and steady, because a slow, jumpy page annoys people and quietly loses you visitors.

LCP	INP	CLS
≤ 2.5s	≤ 200ms	≤ 0.1
Largest Contentful Paint	Interaction to Next Paint	Cumulative Layout Shift
loading · at p75	responsiveness · at p75	visual stability · at p75

The "good" thresholds, all measured on real Chrome users at the 75th percentile. INP replaced FID in March 2024.

INP is the metric JavaScript-heavy sites fail most, and the one a React developer most needs to understand: long main-thread tasks from hydration, heavy event handlers, and third-party scripts all push it past the 200 ms line.

Honest framing on ranking impact: Google's documented position is that page experience acts more like a tie-breaker among comparably relevant results than a dominant factor. Do not promise ranking jumps from a Lighthouse score. The stronger arguments are that slow pages compound into abandonment and unsatisfied clicks, and conversion rates measurably track speed.

FIELD, NOT LAB

Rankings use field data (CrUX, the Chrome UX Report), not your local Lighthouse run. A perfect lab score can coexist with a failing real-user assessment. Check PageSpeed Insights (it shows both), the Search Console CWV report, and your own RUM via the web-vitals package or Vercel Speed Insights.

The Next.js playbook, by metric:

- **LCP:** next/image with `priority` on the above-the-fold hero, never lazy-load the LCP element, next/font to self-host fonts and kill render-blocking font CSS.

- **INP:** move logic into Server Components to shrink client bundles, code-split with next/dynamic, defer third parties with next/script's lazyOnload, break long tasks with startTransition.
- **CLS:** explicit dimensions on all images and video, reserve space for ads and embeds, animate with CSS transforms only.

THE CONTENT SYSTEM

WRITE, PUBLISH, SCALE

IN PLAIN WORDS

This is how you actually make pages that people and search engines want. Start with what a person is really trying to do when they search, answer their exact question near the top, and add something no other page has, like your own numbers or screenshots. Then turn it into a simple weekly routine so your pages build on each other instead of standing alone.

Start from intent, not keywords

Every query has a dominant intent (informational, commercial investigation, transactional, navigational) and the live SERP is the spec: if the top results are comparison tables and Reddit threads, a 3,000-word essay is the wrong format. The workflow: seed terms from Search Console's real query data, expand with a keyword tool, cluster keywords by shared SERP results (pages rank for clusters, not single keywords), then map one primary cluster to one URL so pages never cannibalize each other.

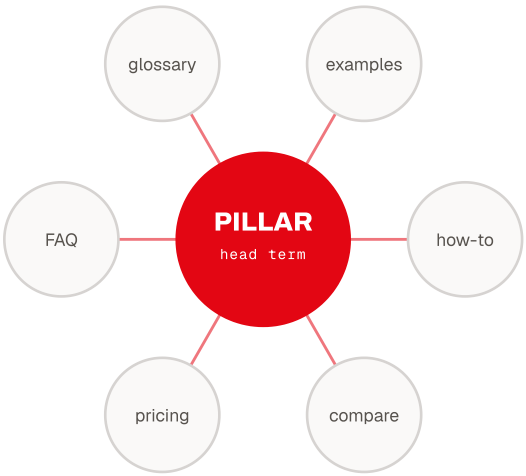
The on-page elements that still matter

Title tag: front-load the topic, keep it around 50 to 60 characters, expect Google to rewrite weak ones. Meta description: no ranking weight, but it is your ad copy for CTR. One H1; descriptive H2s and H3s that mirror the subquestions users ask, because those headings are exactly what snippets and AI systems extract. Readable slugs. Descriptive anchor text on internal links, never "click here".

Information gain is the 2026 differentiator

Generative systems compress consensus content into answers, so a page that restates the top ten results is structurally worthless. Add what does not exist yet: original data (run the benchmark, survey your users, publish the numbers), first-hand experience with screenshots and failure cases, expert quotes, and opinions with reasoning. This is also what Google's helpful content guidance literally asks for: people-first content demonstrating first-hand expertise.

Architecture



Topic clusters: one pillar for the head term, a spoke per subtopic, dense links both ways. Internal linking is the most controllable authority lever you have.

Topic clusters: a pillar page for the head term, spokes for subtopics, dense bidirectional internal links with descriptive anchors. Internal links are the most controllable authority lever a developer has. Every important page within three clicks of the homepage, zero orphans.

NAMED SPAM POLICIES

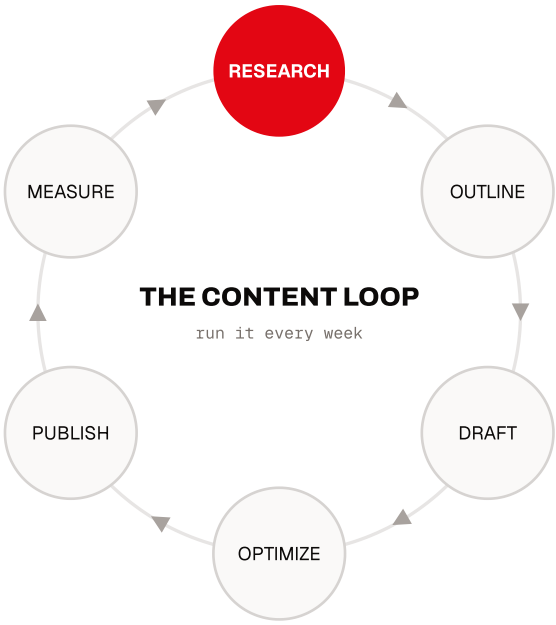
Scaled content abuse (mass-produced pages without value, AI or human), site reputation abuse (parasite content on rented subdomains), expired domain abuse, and doorway pages are all explicitly named in Google's spam policies. Raters are instructed to score low-effort mass-produced content Lowest.

How to write a page that ranks and gets cited

The skeleton of a page built to rank and be cited: one H1, a byline for E-E-A-T, an answer-first intro, question-led sections, your own data, internal links, and schema.

Write the page in that order and it satisfies a human, a ranking system, and an answer engine at once. Lead with the answer, not the backstory. Phrase headings as the questions a reader actually types, then answer each in the first 40 to 60 words before you elaborate. Put one thing on the page that exists nowhere else, your own number, test, or screenshot, because a page that only restates the top ten is worthless to a model that already summarized them. Attach a real byline and date, link out to your primary sources, and link in to three related pages. That is the whole craft; the rest is doing it repeatedly.

The production loop



A repeatable weekly cycle anyone can run: research a question, outline to the live SERP, draft answer-first, optimize, publish, measure, then feed what you learn into the next page. The system is the loop, not any single post.

This is the part that makes it a system rather than a hobby. Each stage is small and handoff-able: research picks one question from the Search Console query data, outlining reads the top three live results and matches their format, drafting fills the page template above, optimizing runs the checklist, and measuring waits a month before judging. Run it once a week and the cluster compounds; run it once and quit, and you never leave the dip.

The pre-publish checklist

The optimize stage in one page. This is the SOP to hand a writer or VA; nothing ships until every box is true.

PRE-PUBLISH CHECK	WHY IT MATTERS
One H1, topic front-loaded, under 60 characters	The title is your click and your topic signal
Every H2 phrased as a real question	Headings are exactly what snippets and AI lift
A 40-60 word answer directly under each question	The extractable unit for snippets and AI Overviews
One thing no competitor has (data, test, screenshot)	Information gain is the 2026 differentiator
Author byline and a visible date	E-E-A-T: proof a real, credible person wrote it
Links out to 3+ related pages, descriptive anchors	Internal links are your top controllable authority lever
Linked in from 3+ existing pages	A new page with no internal links is an orphan
Canonical set, alt text on every image	Avoids duplication dilution, captures image search
JSON-LD valid in the Rich Results Test	Entity clarity for the Knowledge Graph and AI
Content present in the server HTML (curl and grep)	Client-only content may never be crawled

How to scale

Scaling is not writing more; it is running the loop more times without dropping the bar. Three levers. **Cluster, don't scatter:** pick one topic, publish the pillar and its spokes as a batch, and interlink them, so each new page compounds the cluster instead of standing alone. **Templatize:** the page anatomy and the checklist above are the SOP; hand them to a writer or VA and the output stays consistent without you in every draft. **Gate on quality, not volume:** one rule keeps scale from becoming the scaled-content abuse Google penalizes, which is that every page must pass the information-gain check. A page that cannot pass it does not ship, however fast you need the volume.

SMALL MOVES, BIG IMPACT

The highest return per hour, in order: add a 40-60 word answer to the top of your ten best-ranking pages; internal-link every new post from three older ones; refresh the date and the stats on last year's top page; and give your five money pages a real author byline. None takes an afternoon; each one moves a needle.

TRUST

E-E-A-T AND CORE UPDATES

IN PLAIN WORDS

Google wants proof that a real, trustworthy person who actually knows the topic wrote your page, shown through things like a named author, a short bio, and a visible date. E-E-A-T is just shorthand for experience, expertise, authority, and trust. A few times a year Google reshuffles all its rankings in what it calls a core update, and after a drop the fix is usually better content over time, not a quick tweak.

E-E-A-T is not a score and not a direct ranking factor. It is the rubric the roughly 16,000 contracted quality raters use to evaluate results, and their ratings calibrate the machine-learned quality systems, so it shapes what the algorithms are trained to reward. Four letters, with trust as the umbrella the other three feed:

E — **Experience**

First-hand proof you actually used or did the thing. Added in 2022; it matters most for reviews and how-to content.

E — **Expertise**

Depth of knowledge on the topic, shown by who wrote it. A named author tied to a real bio page is the signal.

A — **Authoritativeness**

Whether others in the field treat you as a go-to source. Off-site corroboration outweighs anything you claim yourself.

T — **Trustworthiness**

The umbrella: accurate, transparent, safe. Visible dates, an editorial policy, and honest sourcing build it.

YMYL (Your Money or Your Life) topics carry the strictest bar. The September 2025 rater guidelines update expanded the category to "YMYL Government, Civics & Society" and added rating examples for AI Overview responses. Signals a developer can actually build: real author bylines linked to bio pages, Person and Organization schema tying authors to their sameAs profiles, visible dates, an editorial policy, about and contact pages, and citations to primary sources. Off-site corroboration outweighs on-site claims: independent evidence that the entity is who it says it is.

Core updates: cadence and recovery

2025 shipped three broad core updates (March, June, December); 2026 has already shipped March core, March spam, and May core updates. Google's own recovery doctrine: after a core update there is usually nothing to "fix"; the update re-assessed content broadly, improvements take months, and recovery typically registers at a subsequent core update. The triage protocol: annotate update dates in analytics, segment losses (site-wide vs template vs topic), compare winners and losers in your SERPs, then invest in content quality and pruning rather than technical micro-tweaks.

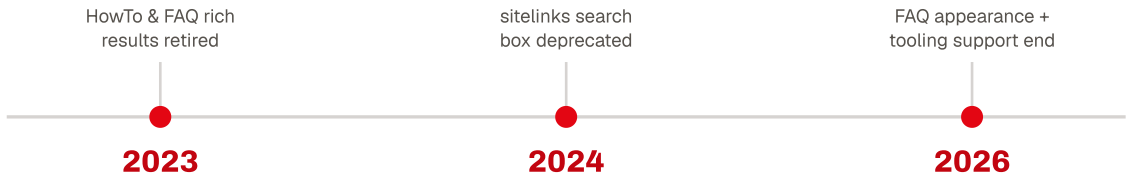
MARKUP AND AUTHORITY

STRUCTURED DATA AND LINKS

IN PLAIN WORDS

Structured data is a hidden bit of code that spells out plainly what your page is, like "this is a product" or "this is a recipe", so machines understand it at a glance. Links are other websites pointing to yours, once treated as gold but now worth far less than simply being mentioned by name. Spend your effort on clear labels and honest mentions, not on chasing a pile of links.

Structured data: fewer rich results, a second job



What Google has removed. Unsupported markup earns nothing, but is not penalized. Ship the schema that still returns a result.

Use JSON-LD, rendered in a Server Component from the same data that renders the page, so the markup never drifts from what a visitor sees. The catch is that most published schema advice is stale: the rich results it promises were retired years ago (above). Unsupported markup is not penalized; it simply does nothing.

What still earns visible rich results: Product, Review, Article, Recipe, Event, JobPosting, LocalBusiness, VideoObject, BreadcrumbList, Organization, Person. The second job in 2026 is entity disambiguation: schema feeds the Knowledge Graph and helps machine consumers, including AI systems, resolve who and what your pages are about. Validate with the Rich Results Test and the schema.org validator.

Links: overrated, not dead

Google's Gary Illyes said links are no longer a top-3 ranking signal, and in 2024: "We need very few links to rank pages." Google's documentation demoted links from "an important factor" to "a factor" in March 2024. What still works: a few relevant, editorially given links from authoritative pages; digital PR built on original data that journalists cite; and unlinked brand mentions, which increasingly matter because LLM-based systems learn brand-topic associations from plain text. The GEO guide carries the anchor dataset here: across 75,000 brands, branded mentions correlate with AI visibility at 0.664 versus 0.218 for backlinks. What fails: paid link networks and exchanges at scale. Internal linking remains the highest-ROI link work a developer can do.

THE NEW SERP

THE AI-RESHAPED SERP

IN PLAIN WORDS

SERP just means the search results page, the list you see after you search. Now AI often writes its own answer at the top and pulls quotes from several pages to build it, sometimes for questions your page never set out to answer. Write clear sections that each fully answer one question, and you become one of the pages it quotes.

AI Overviews (generative answers above organic results) appeared on somewhere between 16% and 48% of queries in 2025-2026 depending on the tracker: Semrush's 10-million keyword panel saw 6.49% in January 2025, a 24.61% peak in July, and roughly 15.69% by November, while BrightEdge's B2B-heavy set reports around 48%. Both are methodology artifacts; the trend is what matters. AI Mode, the fully conversational tab, rolled out to all US users in May 2025.

The key mechanism is **query fan-out**: the system decomposes one query into many synthetic sub-queries, retrieves passages (not whole pages) for each, and synthesizes a cited answer. Full mechanics live in Part III. Three consequences: pages get cited for queries they never targeted; passage-level clarity wins, so write self-contained sections under descriptive headings that lead with the answer; and covering adjacent subquestions increases your retrieval surface. Two independent studies now converge on roughly a third of AI Overview citations coming from top-10 organic results (Ahrefs: 37.9% in early 2026, down from 76% in mid-2025; Surfer: 32%), which means citation retrieval is a partially separate game from classic ranking. The craft of winning extraction is Part II's territory.

GOOGLE'S OFFICIAL RULES

No special requirements beyond being indexed and snippet-eligible. The controls are the existing ones: nosnippet, data-nosnippet, max-snippet, noindex. Google-Extended opts out of Gemini training, not Search AI features. And llms.txt is consumed by no major engine: of roughly 38,000 domains with a valid file, 97% saw zero requests for it in the measured month (Ahrefs). The AEO guide carries the full debate.

INSTRUMENTS

MEASUREMENT IN 2026

IN PLAIN WORDS

These are the free tools that tell you whether any of this is working, like Google Search Console showing which searches led people to you. The big change is what to count: instead of only raw clicks, watch whether AI answers mention you, whether more people search for your brand by name, and whether visitors actually turn into customers.

Search Console is ground truth for queries, impressions, and clicks: segment branded vs non-branded with regex filters, and watch for impressions-up-clicks-down divergence, the signature of AI Overview exposure (AI feature traffic is folded into the "Web" type with no breakout). GA4 for landing-page sessions and conversions, with referral segments for chatgpt.com, perplexity.ai, and gemini.google.com. CrUX and PageSpeed Insights for field vitals. Bing Webmaster Tools, because Bing feeds ChatGPT search. Screaming Frog or Sitebulb for audits, server logs to see what Googlebot and the AI bots actually fetch.

The KPI shift: track citations and share of voice in AI answers, branded search volume, and qualified conversions per session, not raw clicks. AI-citation tracking is a new tool category (Semrush AI toolkit, Ahrefs Brand Radar, Profound, Peec AI); Part III covers the tools and the sampling methodology that keeps their numbers honest. The full stack, and how to wire GA4 and Tag Manager end to end, lives in Part IV.

THE PLAYBOOK

DO THIS, IN ORDER

IN PLAIN WORDS

This is the whole guide turned into a checklist you can follow from the top down. It comes in three levels, from just getting found, to earning a ranking, to staying ahead, and each step names the tool to use and how you know it worked. If you are new, start at level one and do not skip ahead.

Everything above, as an ordered plan. Follow it top to bottom if you are starting out; jump to your level if you are not. Every step names the tool and how you know it worked.

LEVEL 1 STARTER

get found at all

1

Verify your site in Google Search Console.

`search.google.com/search-console` — add your domain, then confirm ownership with the DNS TXT record it gives you.

Done when: the Page Indexing report starts listing your pages (allow a few days).

2

Ship and submit a sitemap.

Add `app/sitemap.ts` and `app/robots.ts`, then submit /sitemap.xml under Search Console > Sitemaps.

Done when: the sitemap status reads "Success" with your page count.

3

Give every page an intentional title and description.

Export `metadata` (or generateMetadata) from each route.

Done when: View Source shows your own title and description, not the layout default.

4

Confirm your content is in the HTML, not just the JavaScript.

Run `curl -s https://yoursite/page | grep "your headline"`.

Done when: the headline appears in the raw response. If not, the page is client-only and crawlers may miss it.

LEVEL 2 PRACTITIONER

earn the ranking

5

Set a canonical on every indexable page.

Add `alternates.canonical` pointing at the clean URL (one trailing-slash policy, no tracking params).

Done when: URL Inspection's "Google-selected canonical" matches the one you declared.

6

Map one keyword cluster to one URL.

Pull real queries from the Search Console `Performance` report, group by intent, split or merge pages so none compete.

Done when: no two URLs rank for the same head term (that is cannibalization).

7

Fix your worst Core Web Vital.

Open `pagespeed.web.dev` on a key page, read the field data, fix the failing metric (LCP: add priority to the hero image).

Done when: the Search Console CWV report moves that URL group to "Good".

8

Add structured data where it earns a rich result.

Render JSON-LD in a Server Component for Product, Article, or Organization as applicable.

Done when: search.google.com/test/rich-results reports a valid, eligible result.

9

Wire real-user monitoring.

Install the `web-vitals` package or Vercel Speed Insights to log LCP, INP, and CLS from actual visitors.

Done when: you have a p75 field trend to watch, not a one-off lab score.

10

Audit rendering at scale.

Crawl the site in `Screaming Frog` twice, JavaScript off then on, and diff the results.

Done when: both crawls surface the same content; any gap is a render-queue dependency.

11

Publish information gain, not consensus.

Ship one page built on original data you own: a benchmark you ran, a dataset you gathered.

Done when: it earns links or citations you did not ask for.

12

Instrument the AI-search divergence.

Build a Search Console filter for impressions-up-clicks-down queries, and add GA4 referral segments for `chatgpt.com, perplexity.ai, gemini.google.com`.

Done when: you can see AI exposure and AI referrals apart from classic organic.

ANTI-PATTERNS

THE MISTAKES THAT COST THE MOST

IN PLAIN WORDS

A plain list of the errors that quietly wreck your results, like hiding your words in code that robots cannot read, or chasing links that no longer count for much. Read it as a list of things to avoid, so you do not pour months into effort that works against you.

- Shipping indexable content only after client-side hydration and trusting the render queue. Anything important goes in the initial server HTML, verified with curl.
- Optimizing lab Lighthouse scores instead of 75th-percentile field data. Rankings use real users.
- Adding FAQ or HowTo schema in 2026 expecting rich results. Those displays are gone; markup that mismatches visible content violates guidelines.
- Publishing scaled AI-generated content without editorial review or added value. That is the named "scaled content abuse" spam policy.
- Soft 404s (HTTP 200 for missing pages), client-side redirects for moved URLs, and robots-blocking a page while expecting its noindex to work.
- Chasing backlink volume when Google needs "very few links". The leverage moved to digital PR, original data, brand mentions, and internal linking.

- Measuring health by rankings and raw clicks in a SERP where 68% of searches end without a click. Track citations, divergence, branded demand, conversions.
- Treating AI search as a separate discipline with hacks like llms.txt instead of extending fundamentals: indexable HTML, answer-shaped passages, entity clarity, and letting the AI crawlers you want citations from actually crawl.

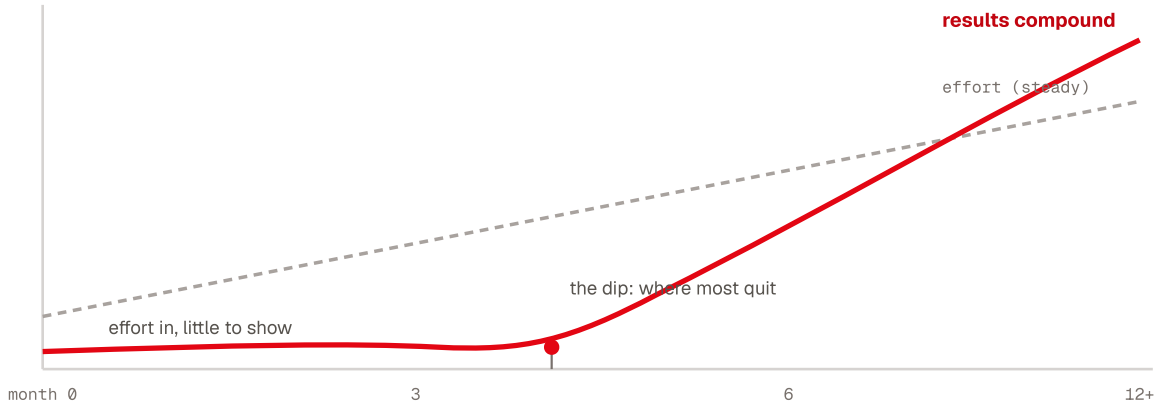
THE HEAD GAME

EXPECT THE DIP

IN PLAIN WORDS

SEO barely shows results for the first few months, and that flat stretch is exactly where most people give up, right before it starts paying off. Knowing the slow start is normal is what lets you keep going. And when a ranking wobbles, do not panic and tear everything apart, wait it out.

The hardest part of SEO is not technical, it is temporal. Results lag effort by months, and that gap is where most people quit. Knowing the shape in advance is what lets you hold the line.



SEO's shape: near-flat months 1-3, long-tail movement 4-6, compounding 6-18. Over 72% of top-10 pages are 3+ years old (Ahrefs), so you compete against age, not just quality.

What to expect

Three to six months for meaningful movement, six to eighteen for a transformation. Only about 1.7% of new pages reach the top ten within a year (Ahrefs). Programs get cancelled between months three and five, almost always just before the compounding would have started. Early signals under six months are impatience talking, not failure.

Traps to avoid

1 **Quitting in the dip**

Months three to five are when the wheel feels stuck and the payoff is closest. Seth Godin's rule: push through the right dip, quit the wrong things fast.

2 **Panicking after a core update**

Do not delete content, rewrite titles, or disavow links in the first two to three weeks; rankings settle. A drop usually means another page now satisfies the intent better, not that you were penalized.

3 **Chasing vanity metrics**

Domain Rating and raw traffic give a dopamine hit while leads stay flat. That is optimization theater. Report the numbers tied to revenue.

THE CADENCE RULE

Watch for anomalies weekly (a traffic crash, a broken index); decide strategy monthly or quarterly. A keyword at #4 one week and #6 the next is normal noise, not a decline. Match how often you check to how often you can actually act.

HONEST UNCERTAINTY

WHAT NOBODY KNOWS YET

IN PLAIN WORDS

An honest list of the things even the experts cannot say for sure, like exactly how much Google weighs clicks, or how big the AI answer effect really is. It is here so you can spot anyone selling you certainty they do not actually have. When studies disagree, trust the overall direction, not one dramatic number.

- **Click signals' exact weight.** Navboost's existence is court-confirmed; its current weighting is not public. Leaked fields are not live weights.
- **AI Overview prevalence and impact numbers disagree wildly** (16% vs 48% prevalence; CTR losses from 19% to 80% by methodology). Trust trends across studies, not any single stat.
- **Whether the early-2026 CTR recovery holds.** Seer's rebound may reflect a temporary UI change; Google iterates the AI Overview link treatment constantly.
- **Whether structured data influences AI citations.** Google says no special markup is needed; practitioner correlation studies are confounded by site quality. The AEO guide's schema section and Part III's contested-tactics list carry this question.

- **Attribution blindness.** Search Console does not separate AI Overview clicks from classic organic, so nobody can precisely measure the effect on their own site.
- **CWV's ranking contribution** stays small and tie-breaker-like per Google, despite agencies marketing it as decisive.

PART II

AEO

Becoming the passage that snippets, People Also Ask, AI Overviews, and voice assistants extract and attribute.

THE INVERSION

WHY ANSWERS BEAT RANKINGS

IN PLAIN WORDS

Search has changed. More and more people get their answer right there on the results page or from an AI, so they never click through to a website. That means being the short answer that gets shown, and named as the source, now matters more than just sitting high on the list of links.

The economics of search visibility inverted between 2024 and 2026. SparkToro's like-for-like clickstream series puts zero-click searches at 60.45% of US Google searches in 2024 and 68.01% by early 2026. Pew Research's behavioral study of 900 US adults found that when an AI summary appears, users click a traditional result on only 8% of visits versus 15% without one, and click a link inside the summary just 1% of the time. Ahrefs measured the CTR penalty for the top result at 58% when an AI Overview is present, worse than the 34.5% it measured in April 2025.

8% vs 15%

Visits where users click a traditional result with vs without an AI summary present (Pew Research, 68,879 searches, 2025).

37.9%

Share of AI Overview citations from top-10 organic results in early 2026, down from 76% in July 2025 (Ahrefs, 863K SERPs).

40-60

Words in the answer-first block that snippet and answer systems most reliably lift (Moz snippet analysis; set-wide standard).

The other side of the ledger, from Part I's canonical narrative: traffic is repricing, not disappearing. Seer Interactive's 5.47-million-query dataset shows organic CTR on AI Overview queries rebounding 85% off its December 2025 floor, and pages cited inside the answer earning roughly 120% more clicks per impression than uncited pages (with the caveat that the rebound may reflect a temporary UI change). Citation is the new position one, and AEO is the craft of winning it.

THE MACHINERY

HOW ANSWER SELECTION WORKS

IN PLAIN WORDS

This explains where those top answers actually come from. Search engines take the boxed answer at the top, called a snippet, from pages that already rank near the top, and AI tools quietly break your one question into several smaller ones and grab a short passage to answer each. Knowing this shows you that you have to rank first, and that every little chunk of your page can get picked on its own.

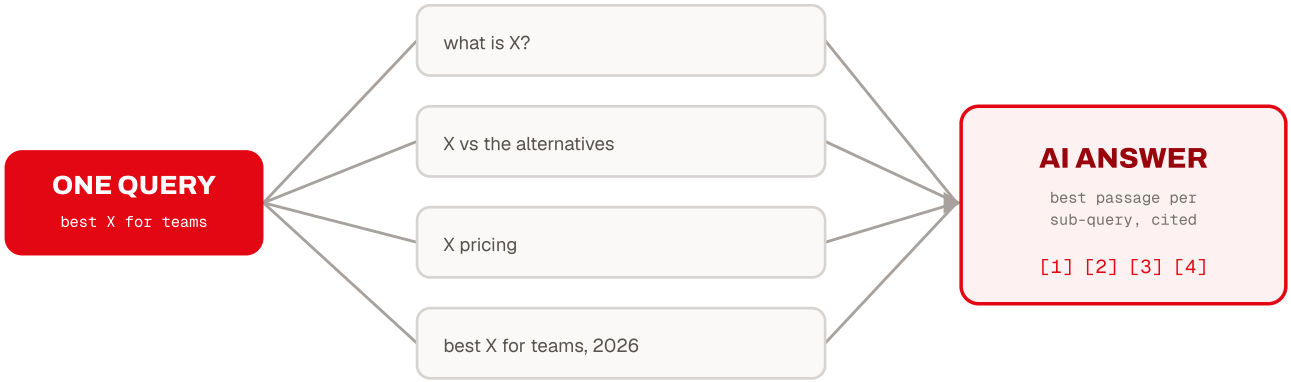
Featured snippets are an elevation, not a separate index

Google's systems decide whether a page "would make a good featured snippet" and elevate it from the normal ranking set. Ahrefs found 99.58% of snippet winners already rank top-10 for the query, so you cannot snippet-optimize your way around weak rankings. Snippets are won by page-one pages that present the most extractable, information-dense passage; winning paragraph answers cluster around 40 to 50 words (Moz, 1.4M snippets).

People Also Ask cascades

PAA is a graph of related questions, each carrying its own snippet. Winning one slot frequently cascades: the same passage serves the PAA accordion, the standalone snippet for that question, and voice answers.

AI Overviews and AI Mode retrieve passages



Query fan-out: one query becomes many, a passage is retrieved for each, and the answer cites the best. You get cited for questions the user never typed.

Google's generative surfaces use query fan-out: one query becomes many synthetic sub-queries, candidate passages are retrieved for each, and a Gemini model synthesizes a grounded, cited answer (Part III owns the full pipeline mechanics). Two consequences: citation happens at the passage level, so a page is a portfolio of independently liftable chunks; and the retrieval net is widening beyond classic rankings, with top-10 results supplying 37.9% of citations in early 2026, down from 76% in mid-2025.

Voice is won as a byproduct

Google Assistant reads featured snippets aloud (Backlinko measured 40.7% of voice answers sourced from snippets); Alexa leans on Bing plus Wikipedia and Wolfram Alpha. Clean, speakable, short-sentence answers win voice without separate work.

> GOOGLE'S POSITION
>
> "There are no additional requirements to appear in AI Overviews or AI Mode" beyond being indexed and snippet-eligible. True but incomplete: eligibility is table stakes. Selection is the competition.

THE TAXONOMY

DRAWING THE LINES

IN PLAIN WORDS

This is a clear table that lays those three jobs side by side, what each one is trying to do and how you measure it. People throw these terms around loosely and mix them up, so having one plain definition to point to keeps you from getting confused or oversold.

The industry blurs these terms constantly, so this table is the set-wide reference (the SEO and GEO guides point here instead of drawing rival taxonomies).

	SEO	AEO	GEO
Objective	Rank and get clicked	Be extracted and attributed	Be known and recommended
Surface	Ranked results, rich results	Snippets, PAA, AI Overviews / AI Mode citations, voice	Generative answers across ChatGPT, Perplexity, Claude, Gemini, AI Overviews
Unit	The page	The passage	The brand / entity (plus the passage)
Primary levers	Crawlability, relevance, links, page experience	Question-first structure, answer-first blocks, format matching	Earned mentions, entity consistency, training-data presence, citability
Measurement	Positions, organic sessions	Answer share per surface	Visibility rate across sampled runs
Time horizon	Months	Weeks to months	Months to model-release cycles

These are concentric circles, not rivals. SEO supplies retrieval; AEO shapes the passages retrieval systems extract; GEO adds the off-site reputation layer generative engines consult. AI Overviews are explicitly the overlap surface where AEO's passage work and GEO's brand work meet on top of SEO's index work. One warning the whole set repeats: vendors use AEO, GEO, LLMO, and AIO interchangeably. Define terms once and hold the line.

THE CRAFT

QUESTION-FIRST ARCHITECTURE

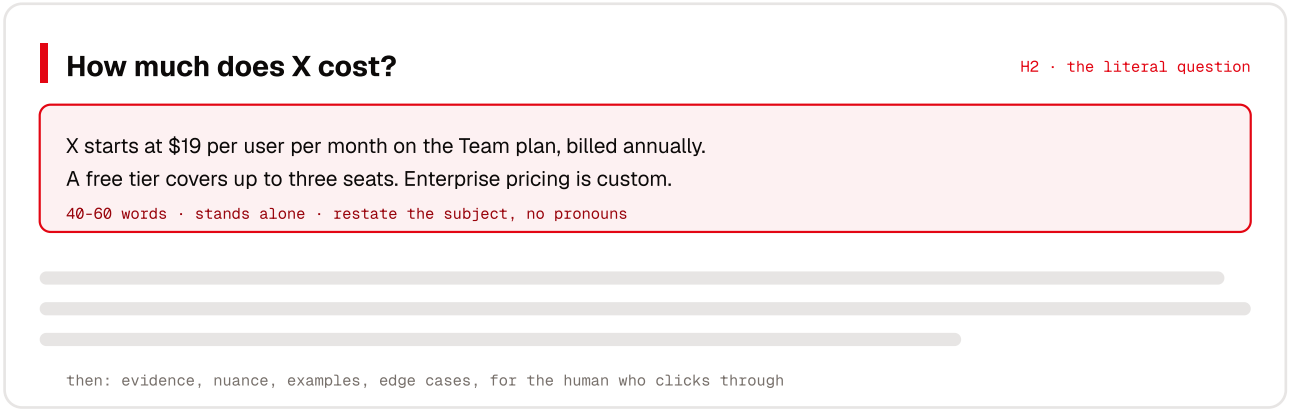
IN PLAIN WORDS

This is the hands-on writing part. Find the real questions people ask, put each question as a heading, then right under it give a short answer of about forty to sixty words that makes full sense on its own, and only after that add the longer details. Write each chunk so it still reads clearly if someone lifts it out by itself, because that is exactly what a search engine or AI does.

Question research replaces keyword research

Mine People Also Ask graphs with AlsoAsked, use keyword tools' question filters, and pull from your own site search and support tickets. Then sample the real answers: ask ChatGPT, Perplexity, and AI Mode the questions your buyers ask, and record who gets cited and why. Cluster questions by intent stage; map each cluster to one URL.

The answer-first block, the atomic unit of AEO



The atomic unit of AEO: a question heading, a 40-60 word answer that stands alone, then the depth. The answer block is what gets lifted into the snippet.

- An H2 or H3 phrased as the literal question a user asks ("How much does X cost?", not "Pricing considerations").
- Immediately below: a direct, self-contained answer of roughly 40 to 60 words, never more than 80, that makes sense with zero surrounding context.
- Then expand: evidence, nuance, examples, and edge cases for the human who clicks through. The inverted pyramid, applied per section rather than per page.

Match format to snippet type

Paragraphs for definitions and "what is" questions; real `<ol>` markup for how-to steps; tables for comparisons, pricing, and specs; short declarative sentences for voice.

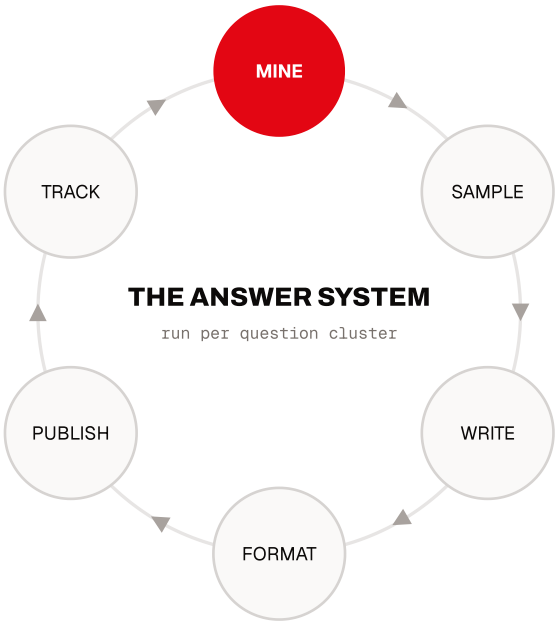
Passage independence

Because fan-out retrieves chunks against machine-written sub-queries, every section should survive being read in isolation: restate the subject noun instead of leaning on pronouns, keep one idea per section, and cover the adjacent sub-questions (cost, alternatives, timeline, risks) on the same page or a tightly linked cluster. Give each question heading a stable anchor link so engines can deep-link the passage.

GOOGLE'S COUNTERPOINT, QUOTED

Google's May 2026 guide says publishers "can ignore tactics like chunking content, creating unnecessary AI text files, or pursuing inauthentic mentions" because its systems understand multi-topic pages. Keep the passage-first writing anyway, for two honest reasons: it is simply clear writing, and non-Google pipelines (ChatGPT, Perplexity) still retrieve chunks. Frame it as craft, not as a Google hack.

The answer system



A repeatable cycle per question cluster: mine the real questions, sample who wins today, write the answer block, match the format the SERP shows, publish, then track citation share and feed it back.

The system is small enough to hand off. Mining and sampling produce a question list with a "who owns it now" note; writing fills the answer-block template; formatting matches the live surface; tracking waits for repeated samples before judging. The work is never a one-off audit because the answer surface shifts under you.

The answer-block checklist

The QA a writer or VA runs on every answer block before it ships.

PER-ANSWER CHECK	WHY IT MATTERS
Heading is the literal question a user asks	Matches the query and is what the engine extracts
Answer is 40-60 words, self-contained	Fits the snippet and survives being lifted out
Subject restated, no pronouns	The passage works when read in isolation
Format matches the live SERP (para, list, table)	The engine picks the format it already shows
A stat, quote, or cited source in the passage	The KDD-validated citation boosts
A stable anchor link on the heading	Engines can deep-link the passage
Organization / Person schema present	Entity clarity for citation eligibility
Page is server-rendered (JS-off check)	Most AI fetchers execute little or no JavaScript

THE TECHNICAL LAYER

SCHEMA AND ACCESS

IN PLAIN WORDS

Schema is a set of hidden labels you add to a page that tell search engines what things are, like this is the company, this is the author. These labels will not magically get you picked, but they help engines trust who you are. The bigger thing here is plumbing: make sure the AI programs that read the web are actually allowed onto your site, because they cannot quote a page they are blocked from reading.

What Google actually says

No special markup, no AI text files, and no special structured data are required for AI Overviews or AI Mode. Schema is not a citation cheat code. Its real value is entity disambiguation and the remaining rich results.

The FAQ and HowTo reality check

The canonical timeline: HowTo rich results ended September 2023; FAQ rich results were restricted to government and health sites in August 2023, stopped appearing in May 2026, lost search-appearance and Rich Results Test support in June 2026, and lose Search Console API support in August 2026. FAQPage markup remains valid schema.org and causes no penalty; it just buys nothing visually. Q&A-formatted *content* remains one of the most extractable patterns. Format beats markup.

Schema that still earns its keep

Organization and Person with sameAs links to Wikipedia, Wikidata, and LinkedIn (these feed the knowledge graphs answer engines consult); Article with author attribution; Product, Offer, and Review for shopping answers; LocalBusiness for voice and local; Speakable (still beta, expectations low); QAPage for genuine forum content.

Crawler access is the real battleground

Answer engines cannot cite what they cannot fetch. The GEO guide carries the canonical crawler taxonomy table; the two facts to carry here: Google-Extended controls Gemini training, not AI Overviews (which use normal Googlebot), and OpenAI's training bot (GPTBot) is a different token from its search bots (OAI-SearchBot, ChatGPT-User). Many sites blocked AI bots at the CDN in 2023-2024 and are still silently opted out of citations. Also: nosnippet, data-nosnippet, and max-snippet limit what Google can quote in snippets and AI Overviews alike. You can protect content from summarization or compete for citation, not both.

Rendering: most AI fetchers execute little or no JavaScript. Server-render the answer blocks. This is Part I's territory; its rendering, soft-404, and canonicalization rules are prerequisites here.

THE PLACEBO

THE LLMS.TXT DEBATE

IN PLAIN WORDS

There is a special file called llms.txt that some people push as the secret to getting picked by AI. The honest truth is that almost no AI actually reads it, so it does next to nothing except for a few sites that are mostly documentation. The real lesson is to trust what search engines actually do, not what salespeople wish they did.

llms.txt is the most hyped and least effective artifact in AEO, and this guide gains credibility by covering it honestly. The proposal (Jeremy Howard, September 2024): a markdown file at /llms.txt listing a site's most important pages in LLM-friendly form. Pitched as robots.txt for the AI era, inverted into an invitation.

The case against, with receipts

- Google does not use it. John Mueller, June 2025: "no AI system currently uses llms.txt", with his comparison to the long-dead keywords meta tag: self-declared, unverified, trivially gameable. Gary Illyes confirmed no support and no plans.
- Ahrefs analyzed ~137,000 domains; of the ~38,000 with a valid llms.txt, 97% received zero requests for the file in the measured month. Adoption is real; consumption mostly is not.
- No major answer engine has committed to it as a ranking or citation input.

The case for, narrower than advocates claim

- Some AI coding tools fetch it opportunistically; documentation-heavy sites report retrieval by coding assistants where a curated markdown index genuinely helps an agent navigate docs.
- Cost is near zero if generated automatically.

VERDICT

A cheap, low-priority hedge for documentation-heavy sites; irrelevant for everyone else. It is not AEO. Anyone selling llms.txt as the centerpiece of AI visibility is selling a placebo. The transferable lesson: always separate what engines document and demonstrably do from what the industry wishes they did.

SCOREBOARD

MEASURING ANSWER SHARE

IN PLAIN WORDS

Answer share is a simple scorecard: out of the questions that matter to your business, how often are you the answer that gets shown or credited. Because AI gives a slightly different answer each time you ask, you have to check the same questions many times over weeks and watch the trend, not trust a single day's reading.

Define **answer share**: of a fixed basket of questions that matter to the business, the percentage where you are the displayed answer or a cited source, per surface (snippet, PAA, AI Overview, AI Mode, ChatGPT, Perplexity), tracked over time against named competitors.

First-party signals

- Search Console folds AI Overview and AI Mode activity into "Web" data with no breakout. State the limitation plainly: you cannot isolate AI Overview traffic in GSC.
- Referral segments for chatgpt.com, perplexity.ai, copilot.microsoft.com, gemini.google.com. Volumes are small; measure your own conversion rates before repeating vendor claims.
- Server logs: AI crawler hits (GPTBot, OAI-SearchBot, PerplexityBot, ClaudeBot) are a leading indicator of retrievability.

Tools

Rank trackers (Semrush, Ahrefs, seoClarity) report snippet and PAA ownership per keyword, still the cheapest reliable answer-share data. AI visibility platforms (Profound, Peec AI, Otterly.AI, Semrush AI Toolkit, Ahrefs Brand Radar) all work the same way: run a prompt basket against multiple LLMs on a schedule, parse who gets mentioned and cited, compute share of voice. The full tool stack, with cost tiers and a setup order, lives in Part IV.

METHODOLOGICAL HONESTY

LLM answers are non-deterministic: Part III's canonical evidence (SparkToro and Gumshoe, 2,961 runs) found under a 1-in-100 chance of the same brand list twice, yet stable 55-77% inclusion rates for top brands. Insist on repeated sampling and trend lines; treat every single-day reading, and every "AI rank #3" dashboard, as noise.

THE PLAYBOOK

DO THIS, IN ORDER

IN PLAIN WORDS

This is the whole thing turned into a checklist you can follow from the top. It is split into three levels, from just getting allowed and eligible, to rewriting your pages to win the answer, to measuring it and keeping it fresh. Each step tells you exactly what to do and how to know it worked, so a beginner can just start at step one.

The whole discipline as an ordered plan. Follow it top to bottom if you are starting out; jump to your level if you are not. Every step names the tool and how you know it worked.

LEVEL 1 STARTER

become eligible

1

Confirm you are snippet-eligible.

Search your codebase and meta tags for accidental `nosnippet`, data-nosnippet, or a low max-snippet on pages you want quoted.

Done when: the pages you care about carry no snippet-suppressing directives.

2

Check which AI crawlers you allow.

Open `robots.txt` and your CDN/WAF rules (Cloudflare). Confirm OAI-SearchBot, PerplexityBot, and Bingbot are not blocked.

Done when: a fetch as each agent (or a log check) returns 200, not 403.

3

Build your first question list.

Mine `alsoasked.com` or your keyword tool's Questions filter, and add the questions from your own support tickets.

Done when: you have 20-plus real questions, written as questions.

4

Sample who wins today.

Ask ChatGPT, Perplexity, and Google AI Mode your top questions and record who gets cited.

Done when: you know who currently owns each answer.

LEVEL 2 PRACTITIONER

win the extraction

5

Retrofit your page-one pages first.

For each page already ranking top-10, add a question-phrased H2 and a self-contained 40-60 word answer beneath it.

Done when: the answer block reads correctly with zero surrounding context.

6

Match format to the answer type.

A paragraph for definitions, a real `<ol>` for how-to steps, a table for comparisons.

Done when: your format matches what the live SERP shows for that query.

7

Give every question heading an anchor link.

Add `id` anchors so engines can deep-link the passage.

Done when: each heading is individually linkable (yoursite/page#the-question).

8

Add entity schema.

Render Organization and Person JSON-LD with `sameAs` links to your real profiles.

Done when: the Rich Results Test validates it.

LEVEL 3 ADVANCED

measure and compound

9

Define an answer-share basket.

Fix 50-100 questions that matter and, per surface (snippet, AI Overview, ChatGPT, Perplexity), record whether you are the answer.

Done when: you have a baseline percentage per surface.

10

Track it with repeated sampling.

Use an AI-visibility tool (`Profound, Peec AI, Otterly`) or your own prompt script, running each prompt many times per week.

Done when: you have a trend line, not a single-day reading.

11

Move off-site where answers are cited.

Earn presence on the surfaces AI answers pull from: `Reddit, YouTube`, review sites, credible publications.

Done when: your brand appears in the cited sources, not only on your own domain.

12

Refresh on a cadence.

Re-sample the basket monthly and update answer blocks when facts age.

Done when: stale facts (pricing, dates) never survive a monthly pass.

Effort allocation, honestly: roughly 60% content restructuring, 20% measurement, 15% off-site, 5% technical files and markup.

ANTI-PATTERNS

THE MISTAKES THAT COST THE MOST

IN PLAIN WORDS

This is the list of common mistakes that burn the most time. The big ones are trying to win the answer on pages that do not even rank yet, hiding your answer under waffle, and trusting a single day's result or shiny vendor claims. Reading this once saves you from the traps most people fall into.

- Chasing answer placement on pages that rank nowhere. Snippets elevate from page one; AEO on unranked pages is building on sand.
- Treating llms.txt, FAQ schema, or any machine-readable file as the strategy. Format and substance beat markup, and the files are unread.
- Burying the answer under preamble, hedges ("It depends on many factors..."), or marketing throat-clearing instead of a self-contained 40-60 word block.
- Writing pages instead of passages: pronoun-dependent, interdependent sections are unextractable even on excellent pages.
- Blocking AI crawlers indiscriminately, or confusing Google-Extended with Googlebot and GPTBot with OAI-SearchBot.
- Reporting single-day AI visibility readings as fact.
- Repeating unverifiable vendor statistics ("FAQ schema lifts snippets 35%") without tracing them to a real study. This is the discipline's biggest credibility trap.
- Measuring success in clicks alone when 68% of searches end without one. Track citation share, mention sentiment, and answer share.

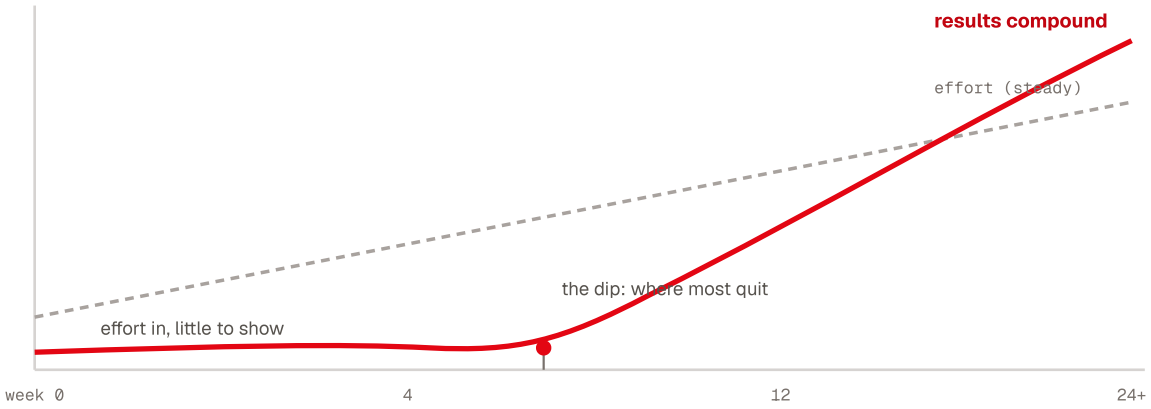
THE HEAD GAME

FASTER SIGNAL, FUZZIER PAYOFF

IN PLAIN WORDS

This is about what to expect and how to stay calm. You will see results faster than with regular search, sometimes in a week or two, but there is no steady rank to watch, so a mention can show up and then vanish. Judge success by whether the AI quotes you accurately and often over time, not by a single good or bad day.

AEO rewards you sooner than SEO, but on a surface with no stable rank to grip. The mental shift is from "what position am I" to "am I the answer, and did it get my facts right".



AEO moves faster than SEO: first citations in one to two weeks on a strong domain, but consistent citation and real business impact still take three to six months. The payoff is presence, not clicks.

What to expect

Technical changes (structure, schema, answer-first blocks) can show in two to four weeks on an established domain; first citations in one to two weeks for strong sites, four to six for new brands; consistent citation patterns in three to six months. Success looks like the machine quoting you accurately and often, not a traffic spike.

Traps to avoid

1 **Importing the rank-check reflex**

An AI citation is not a keyword position; it appears and disappears. Checking it like a rank produces false alarms.

2 **Over-reading a first-week win**

Early citations arrive fast, so teams extrapolate one good week into a trend, then panic when it vanishes. Sample over weeks.

3 **Formatting tricks over trust**

Citation eligibility runs on E-E-A-T and verifiable accuracy. Chasing snippet hacks while ignoring authority underperforms.

THE METRIC THAT MATTERS

Judge AEO on citation consistency and the factual accuracy of the quote, not click volume. Treat it as a recurring workflow, not a one-off audit (Aleyda Solis): the answer surface shifts under you, so the work is never finished.

HONEST UNCERTAINTY

WHAT NOBODY KNOWS YET

IN PLAIN WORDS

This is an honest admission of what is still unknown. Nobody fully knows how the AI picks which sources to credit, the numbers get argued over, and the tricks change fast. So stay humble, note the date on anything you read, and put your effort into the things that last: genuinely helpful content, being trusted, staying readable to crawlers, and keeping a habit of measuring.

- **Selection mechanics are partially opaque.** Google documents fan-out but not how citations are chosen among grounded candidates. The top-10 overlap falling from 76% to 37.9% in under a year shows the ground shifts fast enough to invalidate playbooks mid-year. Date every claim.
- **The zero-click math is contested.** Clickstream panels differ; Google disputes the framing without publishing data. Present ranges.
- **Does schema help AI citation?** Google says no special structured data is needed. Vendor lift numbers rarely share methodology. The defensible claim is entity disambiguation, not bought citations.
- **Non-determinism undermines measurement.** No tool observes real user prompt distributions; all sample. Whether tracked baskets represent actual buyer behavior is unsolved.
- **Attribution of business value is immature.** "AI traffic converts 4-6x better" circulates widely and traces to vendor posts, not audited data.
- **Durability risk.** Google restricted FAQ rich results overnight in 2023 and reshaped citation patterns within months in 2025-2026. AEO tactics have shorter half-lives than SEO tactics. The durable assets: content that genuinely answers questions, entity authority, crawl access, and a measurement habit. The GEO guide's contested-tactics list continues this thread.

PART III

GEO

Getting retrieved, cited, and recommended inside AI answers, with the evidence base and an honest map of the unknowns.

THE SHIFT

THE NEW CONSIDERATION SET

IN PLAIN WORDS

When someone asks an AI for the best option, it hands back a short list of names. If your client's brand is not on that list, the buyer never even hears about them, even when they would be a perfect fit. This section is about why getting your brand onto that list is the whole game now.

Buyers increasingly ask an assistant "what's the best X" and receive a synthesized shortlist. If a brand is not in the model's consideration set or the retrieval layer's citation pool, it is invisible at the moment of highest intent. The traffic side is already measurable: Semrush's clickstream analysis found ChatGPT's outbound referral traffic grew 206% during 2025 even as its own visits plateaued near a billion per month, and 30% of all ChatGPT referral traffic flows to just ten domains. Pew Research found users click a link inside an AI summary on just 1% of visits: the mention itself is most of the value.

0.664

Spearman correlation of branded web mentions with AI Overview visibility across 75K brands, vs 0.218 for backlinks (Ahrefs, 2025).

+206%

Growth in ChatGPT outbound referral traffic during 2025 (Semrush clickstream, 2026).

55-77%

Stable inclusion rate of top brands across repeated AI runs, while exact lists repeat less than 1 in 100 times (SparkToro/Gumshoe, 2026).

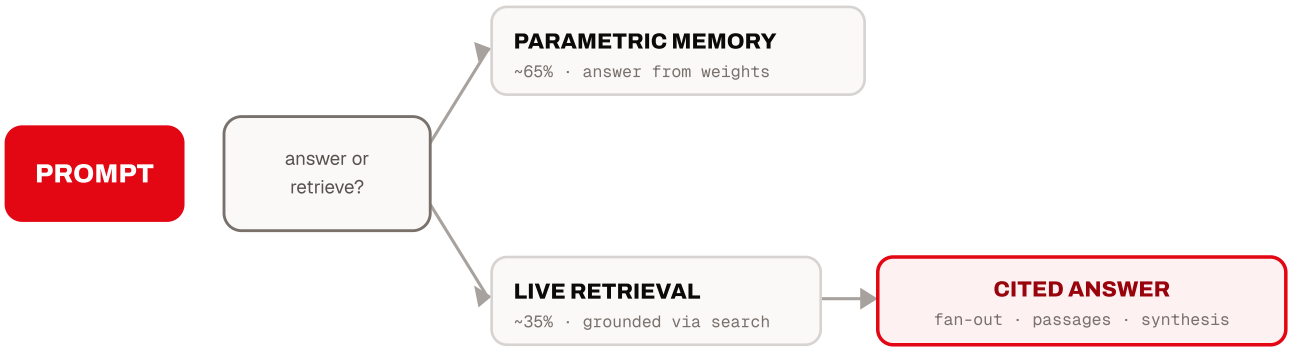
GEO matters not because it replaces SEO but because a growing slice of demand is mediated by generative engines whose selection mechanics differ from ranking algorithms in documented, exploitable ways. This guide covers those mechanics, the evidence for what moves them, and the discipline needed to not fool yourself.

THE MACHINE

THE CITATION PIPELINE

IN PLAIN WORDS

This walks through the steps an AI takes between reading a question and writing its answer. Sometimes it answers from memory, sometimes it searches the web first, and it usually breaks one question into several smaller searches. A citation is just the AI naming its source, and knowing these steps shows you exactly where your brand can slip in.



Every answer forks first: reply from training memory (about 65% of ChatGPT queries) or retrieve live and cite. GEO fights on both paths.

Every major AI answer engine is a retrieval-augmented generation (RAG) system layered on a conventional search index. GEO is impossible to practice well without the four stages.

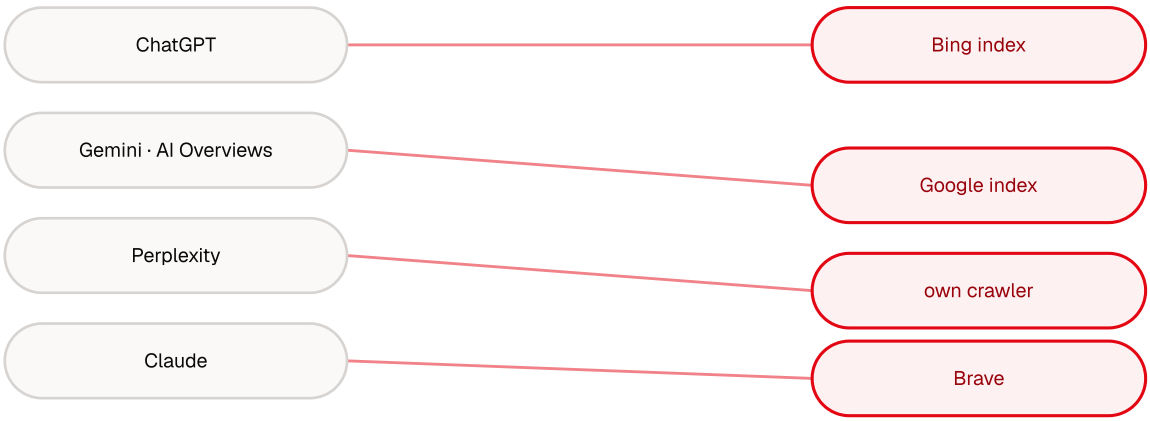
Stage 1: Answer or retrieve

The model first decides whether to answer from parametric memory (knowledge baked into weights at training time) or ground the answer with live retrieval. Semrush's clickstream data shows ChatGPT enabled web search on only 34.5% of queries as of February 2026, down from 46% in late 2024. The split matters enormously: for memory-led answers only your training-data footprint counts; for grounded answers your retrievability counts.

Stage 2: Query fan-out

The engine rewrites the user's prompt into multiple synthetic sub-queries. Google's documentation confirms AI Overviews and AI Mode "may use a query fan-out technique, issuing multiple related searches across subtopics and data sources". Search Engine Land's reverse-engineering of ChatGPT's internal web.run tool found newer models run five to ten-plus rounds of search per response, refining queries from earlier results, and formulating queries around sources the model already knows. Their line worth memorizing: a brand absent from parametric memory won't even be considered. You are optimizing for a cloud of machine-written queries, not one keyword.

Stage 3: Index dependency



Each engine rides a specific index. Absent from Bing means absent from ChatGPT, whatever your Google rank. Verify in Bing Webmaster Tools and adopt IndexNow.

Each engine rides a specific index: ChatGPT retrieves via Bing-derived and third-party search APIs, then fetches selected pages with ChatGPT-User; Gemini, AI Overviews, and AI Mode use Google's index; Perplexity runs PerplexityBot; Claude's web search has used Brave. Consequence: a site absent from Bing effectively does not exist for ChatGPT search, whatever its Google rankings. Bing Webmaster Tools and IndexNow are first-class GEO infrastructure. The foundations each index requires are Part I's territory.

Stage 4: Passage selection and synthesis

Retrieved pages are chunked into passages; the model selects, compresses, and attributes. Citation slots are scarce and shrinking: after a ChatGPT model update, unique domains cited per response fell from 19 to 15 (Search Engine Land). The unit of competition is the passage, not the page. A self-contained, quotable block beats a great page whose answer is smeared across two thousand words.

THE EVIDENCE

THE GEO PAPER AND ITS LIMITS

IN PLAIN WORDS

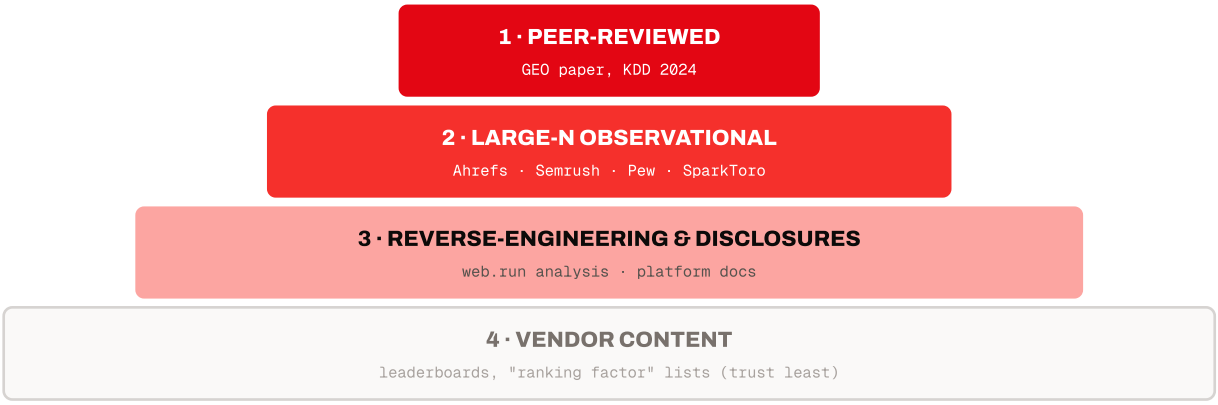
There is one solid research study behind all of this, and it found that adding real quotes, real numbers, and links to sources makes an AI more likely to feature your content. Cramming in repeated keywords actually made things worse. Just remember the study was run in a lab, so treat its findings as a strong hint, not a guarantee.

The discipline's name comes from the one piece of peer-reviewed evidence: Aggarwal et al., "GEO: Generative Engine Optimization", ACM KDD 2024. It tested nine content modifications across GEO-bench (10,000 queries) and found visibility gains of up to 40%. Tactic-level results, measured as position-adjusted word count: quotation addition +41% over baseline, statistics addition +33%, citing sources +28%. Keyword stuffing dropped visibility below baseline, the single most quotable anti-habit finding. And the equity result: citing sources lifted rank-5 sites' visibility 115.1% while rank-1 sites lost 30.3%. GEO is highest-leverage for challengers.

THE LIMITATION, STATED EVERYWHERE THE PAPER IS CITED

The experiments ran on 2023-era models in a sandboxed RAG setup, not live commercial engines; visibility measured answer share, not traffic or revenue; and tactics were applied to sources already retrieved, so the paper says nothing about winning retrieval itself. Treat it as proof that content features causally affect citation in RAG systems, not as a guaranteed percentage.

The evidence hierarchy



Weigh every claim by its tier: strongest and rarest at the top, weakest and most common at the base. Most published GEO content is tier 4 quoting tier 4.

Apply it ruthlessly: every tactical claim gets tagged to a tier, and tier-4 numbers are quoted only with named methodology. Most published GEO content is vendors quoting each other.

THE CRAFT

WHAT MAKES A PASSAGE CITABLE

IN PLAIN WORDS

A passage is just a chunk of your writing that an AI might lift out and use in its answer. To make it easy to lift, open each section with a short answer that stands on its own, back it with real numbers and named sources, and say your brand name plainly. If a chunk only makes sense when it is glued to the rest of the page, the AI will pass it over.

The on-page directives, translated from the validated evidence (the full question-first architecture is Part II's territory):

- A direct, self-contained answer of roughly 40 to 60 words leading every section, written to survive being lifted out of context: subjects restated, no pronoun dependence.
- Specific, sourced numbers ("converts at 7.1%" beats "converts well"), named-expert quotations, and outbound citations to primary sources. These are the three empirically validated boosts.

- Structure for retrieval: descriptive H2s and H3s phrased as questions, tables for comparisons, tight lists, one idea per paragraph. Formatting is retrieval engineering, not decoration.
- Entities by name: your brand, product, and category in the citable passages themselves. Models attribute to named entities, not to "we".
- Freshness: dateline content and update it. Retrieval favors recent documents for time-sensitive fan-out queries.

One honest complication: Google's May 2026 guide says publishers "can ignore tactics like chunking content" because its systems understand multi-topic pages. Keep the self-contained sections anyway, framed correctly: it is clear writing, and non-Google pipelines still retrieve chunks. What the citation-share data adds: Reddit, Wikipedia, YouTube, and LinkedIn dominate cited domains across engines, and 30% of ChatGPT referral traffic flows to ten domains. Being cited *about* on high-trust third-party surfaces is often easier than getting your own domain cited, which is the next section's subject.

THE TWO LAYERS

TRAINING DATA VS RETRIEVAL

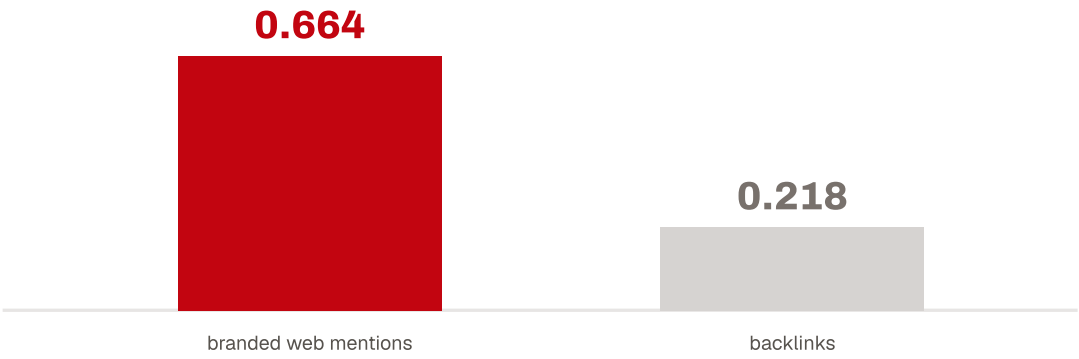
IN PLAIN WORDS

An AI can know your brand in two ways. Either it learned about you when it was first built, or it looks you up fresh the moment someone asks. You help the first by keeping your facts the same everywhere online, and you help the second by earning other trusted websites to mention you by name. This section shows you how to work both.

Layer 1: Parametric presence

Models answer roughly 65% of ChatGPT queries without live search (Semrush). What a model knows about a brand comes from its training corpus: Wikipedia, news archives, Reddit and forums, review sites, GitHub, the general crawl. You cannot inject yourself into a frozen model, but you can influence the next snapshot and, more immediately, the model's query-writing behavior, since engines formulate retrieval queries around entities they already know. The levers: a factually consistent Wikipedia and Wikidata presence where notability genuinely supports it; identical entity facts everywhere (same name, category description, and positioning across your site, LinkedIn, Crunchbase, G2, directories); earned coverage in publications and communities that are demonstrably in training corpora.

Layer 2: Retrieval presence

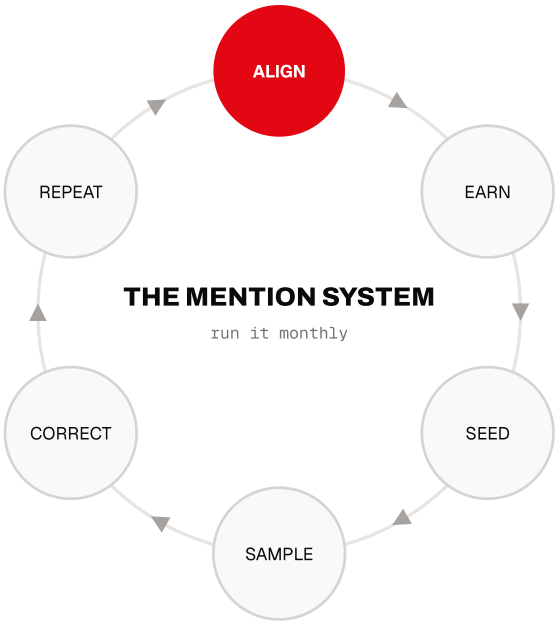


How strongly each signal correlates with AI Overview visibility across 75,000 brands (Spearman). Mentions beat backlinks roughly three to one (Ahrefs, 2025).

For grounded answers the battle is being in the index, matching fan-out queries, and being quotable (previous section). The anchor dataset for where to spend: Ahrefs' 75,000-brand study found branded web mentions correlate with AI Overview visibility at 0.664 versus 0.218 for backlinks, and top-quartile brands for mentions averaged 169 AI mentions versus 14 for the next quartile.

- Reallocate digital PR from link acquisition to named mentions in credible, crawlable publications. A mention with your brand next to category language is the unit of value.
- Be genuinely present on Reddit, industry forums, and YouTube. Astroturfing risk is real: platforms police it, models cross-check, and Google's May 2026 guide now explicitly warns against "pursuing inauthentic mentions".
- Seed best-of and comparison coverage on third-party sites; engines prefer citing neutral-seeming lists for recommendation queries.
- Maintain a machine-consistent fact base: a plain-language About page, sameAs schema linking official profiles, identical descriptions across profiles. One coherent entity for both training and retrieval to encounter.

The mention system



A monthly cycle: align your entity facts everywhere, earn a named mention in a crawlable publication, seed comparison coverage, sample AI visibility, correct what the models get wrong, and repeat. Brand presence is built, not shipped once.

GEO's off-site work only compounds if it recurs. Each pass is a handful of concrete tasks: reconcile any drifted entity fact, land one credible mention next to your category language, add or refresh a third-party comparison, run the prompt panel, and publish a correction for any hallucinated fact. None of it is a campaign; it is a standing routine a VA can own.

The monthly GEO checklist

MONTHLY TASK	WHY IT MATTERS
Ask each engine "what is [brand]" and log the answer	Catches drift and hallucinated facts early
Reconcile any fact that differs across your profiles	One coherent entity feeds both training and retrieval
Earn one named mention in a crawlable publication	Mentions correlate with AI visibility 3x more than links
Add or refresh one third-party comparison / listicle	Engines prefer citing neutral lists for "best X"
Run the fixed prompt panel across engines	Repeated sampling turns noise into a trend
Publish a correction for any wrong fact an engine stated	Authoritative pages are how you fix the record
Confirm your quotable pages are in Bing's index	Absent from Bing means absent from ChatGPT search

PREREQUISITES

TECHNICAL READINESS

IN PLAIN WORDS

AI companies send out little reader programs, called crawlers, that scan the web, and some of them are the ones that let an AI find and quote you. This section makes sure you have not accidentally shut the door on the readers you actually want, and that your pages can be read without any fancy loading tricks. Small settings here can quietly make you invisible.

The crawler taxonomy (canonical for the set)

CLASS	AGENTS	ROBOTS DECISION
Training	GPTBot, Google-Extended, ClaudeBot	Blocking keeps you out of future corpora, not out of search or answers
Search index	OAI-SearchBot, PerplexityBot, Bingbot, Googlebot	Blocking removes you from that engine's answers entirely
User-triggered fetch	ChatGPT-User, Perplexity-User, Claude-User	Blocking prevents live citation at answer time

The two facts every robots.txt audit needs: Google-Extended controls Gemini training only, NOT AI Overviews (those use normal Googlebot, and cannot be opted out of separately except via snippet controls); and blocking GPTBot does not remove you from ChatGPT search, that is OAI-SearchBot and ChatGPT-User. Many sites blocked AI bots wholesale in 2023-2024 and are still invisible in AI search without knowing it.

The rest of the checklist

- **Bing coverage.** Verify in Bing Webmaster Tools, submit sitemaps, adopt IndexNow. Absent from Bing means absent from ChatGPT search.
- **Rendering.** Most AI fetchers execute little or no JavaScript. Server-render anything you want quoted; fetch your pages with JS disabled and check server logs for AI user agents.
- **Structured data.** Google says no special schema is required for AI features. Keep Organization, Article, and sameAs for entity disambiguation; do not sell schema as a citation hack.
- **llms.txt.** No confirmed consumer; Mueller: "no AI system currently uses llms.txt". Optional, low priority. The AEO guide carries the full debate.
- **Performance.** Real-time fetchers work under tight time budgets; slow pages get skipped at answer time. Fast TTFB is retrieval insurance.

SCOREBOARD

MEASUREMENT WITHOUT SELF-DECEPTION

IN PLAIN WORDS

An AI gives a slightly different answer every time you ask, so checking your brand just once tells you almost nothing. You have to ask the same questions many times and track how often your brand shows up, the way you judge weather over a week instead of a single day. Any tool that claims you are ranked number three in an AI is making it up.

SparkToro and Gumshoe.ai ran 2,961 tests across ChatGPT, Claude, and Google AI with 600 volunteers: under a 1-in-100 chance of the same brand list twice for the same prompt, roughly 1-in-1,000 for the same order. Fishkin's conclusion: any tool that gives a ranking position in AI is full of baloney. But visibility percentage across many runs is stable (top brands appeared in 55-77% of responses), so the correct metric is share of responses mentioning the brand, measured across dozens of prompts, each run many times, per platform, as a trend.

The metric stack

- Mention rate and share of voice across a fixed prompt set, weekly, per engine.
- Citation rate: how often your domain is a linked source. A different number from mentions; tools report them differently.
- Sentiment and accuracy of what engines say about you. Hallucinated pricing and stale facts are common and fixable with authoritative pages.
- AI referral traffic, segmented by referrer (ChatGPT appends utm_source=chatgpt.com on many links). Track conversion separately and distrust vendor conversion claims.
- Bot activity in server logs: ChatGPT-User and Perplexity-User hits reveal which pages are being pulled into answers right now.

Tools, and how to buy them

Dedicated: Profound (enterprise leader), Peec AI (mid-market), Otterly.AI (entry tier), Scrunch AI. Incumbent add-ons: Semrush AI Toolkit, Ahrefs Brand Radar. Free: your own prompt panels via API scripts, plus Search Console (knowing AI Mode clicks are folded into web totals) and log analysis. The buyer's checklist: disclosed sample sizes and re-run counts, per-engine breakdowns, a citation-vs-mention distinction, API access to raw runs, and no "AI rank #3" theater. A vendor that cannot explain sampling error is selling decoration. The full tool stack, and how to wire it together, lives in Part IV.

THE PLAYBOOK

DO THIS, IN ORDER

IN PLAIN WORDS

This is the whole guide turned into a to-do list you can work through from top to bottom. It starts with the basics and builds up, and every step names the tool to use and tells you how to know it is done. If you act on only one section of this guide, make it this one.

The whole discipline as an ordered plan. Follow it top to bottom if you are starting out; jump to your level if you are not. Every step names the tool and how you know it worked.

LEVEL 1 STARTER

become retrievable

1

Get into Bing's index.

`bing.com/webmasters` — verify your site, submit your sitemap, turn on IndexNow. This is what ChatGPT search retrieves from.

Done when: Bing Webmaster Tools shows your pages indexed.

2

Audit your AI crawler access.

In `robots.txt` and your CDN/WAF, decide per bot class: training (GPTBot, Google-Extended), search-index (OAI-SearchBot, PerplexityBot, Bingbot), user-fetch (ChatGPT-User, Perplexity-User).

Done when: the search-index and user-fetch bots you want citations from are allowed (verify with a fetch or your logs).

3

Confirm your answers survive without JavaScript.

Load a key page with JS disabled (or `curl` it) and check the passage you want quoted is present.

Done when: your quotable content is in the server HTML.

4

Fix your entity fact base.

Make your name, category, and one-line positioning identical across your About page, LinkedIn, Crunchbase, and G2; add `sameAs` schema.

Done when: an AI asked "what is [brand]" describes you the way your About page does.

LEVEL 2 PRACTITIONER

become citable

5

Lead every section with a quotable answer.

Rewrite section openings as a self-contained 40-60 word answer with the subject restated (no pronouns).

Done when: any single section, copied alone, still makes sense.

6

Embed the three validated boosts.

Add specific sourced numbers, named-expert quotations, and outbound citations to primary sources in the passages themselves.

Done when: each key passage carries a stat, a quote, or a citation.

7

Name the entities in the passage.

Put your brand, product, and category in the citable text, not "we".

Done when: a passage read in isolation attributes to your brand by name.

8

Reallocate PR from links to mentions.

Aim outreach at earning your brand named in credible, crawlable publications and genuine presence on `Reddit`, `YouTube`, not backlinks.

Done when: you are earning mentions next to category language, not just links.

LEVEL 3 ADVANCED

measure without self-deception

9

Build a visibility-rate panel.

Fix a set of prompts, run each 30-100 times per engine on a schedule, and measure the share of responses that mention or cite you.

Done when: you have a mention-rate trend per engine, and you have stopped tracking "AI rank position".

10

Separate mention from citation.

Track how often you are named versus how often your domain is a linked source; they move differently.

Done when: your dashboard reports both, distinctly.

11

Watch AI referrals and bot hits.

Segment referral traffic by `chatgpt.com`, `perplexity.ai`, `gemini.google.com` in GA4, and watch ChatGPT-User and Perplexity-User hits in your server logs.

Done when: you can see which pages are pulled into answers in near-real time.

12

Correct what the models get wrong.

When an engine states outdated pricing or a wrong fact about you, publish an authoritative page stating it plainly, then re-sample.

Done when: the corrected fact shows up in later answers.

ANTI-PATTERNS

THE MISTAKES THAT COST THE MOST

IN PLAIN WORDS

This is a list of the traps that waste the most time and money, so you can steer around them. Things like checking your results only once, forgetting about Bing, or spending on the wrong kind of links. Knowing what not to do can save you more than any single clever tactic.

- Treating GEO as keyword SEO with a new name. Engines fan out into five to ten-plus machine-written sub-queries, and keyword stuffing measurably reduced visibility in the KDD study.
- Tracking "AI rank position" from single prompt runs. Single-sample dashboards measure noise.
- Ignoring Bing. A site unverified in Bing Webmaster Tools can rank #1 on Google and be absent from ChatGPT answers.
- Blocking AI crawlers wholesale, or blocking OAI-SearchBot and ChatGPT-User while meaning to block only the training bot GPTBot.
- Shipping client-side JavaScript content that AI fetchers cannot render, so the passages meant to be quoted are invisible.
- Spending the off-site budget on backlinks (0.218 correlation) instead of earned branded mentions (0.664) and presence where citations concentrate: Reddit, YouTube, Wikipedia.
- Burying the answer: no self-contained, stat-and-source-rich lead in a section means nothing survives passage-level chunking.
- Presenting llms.txt, schema hacks, or single-anecdote case studies as proven tactics instead of labeled experiments. Informed readers notice, and credibility is the product.

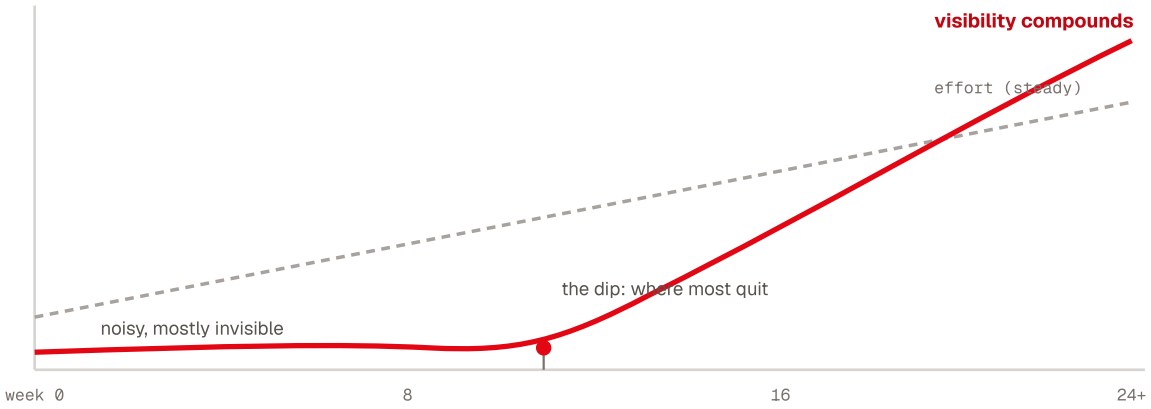
THE HEAD GAME

OPTIMIZING WHAT YOU CANNOT CONTROL

IN PLAIN WORDS

This part is about staying patient, because the results here are slow and jumpy by nature. Your brand might show up one day and vanish the next, and it can take a couple of months before mentions turn steady. That wobble is completely normal, so do not panic and do not quit during the quiet stretch.

GEO is the hardest surface to stay sane on, because the output is probabilistic: the same prompt gives different answers. Volatility is the baseline, not a sign you are failing.



GEO's payoff is the slowest and noisiest: eight to sixteen weeks to consistent citations. In one 2026 study only ~30% of brands stayed visible between two consecutive answers (AirOps). Expect the wobble.

What to expect

Not a rank, a probability. The same prompt varies across sessions, users, and geographies. Consistent citations take eight to sixteen weeks; stable multi-engine visibility four to six months. In that 2026 study only 30% of brands stayed visible between two consecutive answers and 20% across five runs. A single good answer this morning tells you almost nothing.

Traps to avoid

1 **Measuring noise as signal**

Single-session spot checks are noise. Nothing you conclude from one prompt run is real.

2 **Hype in both directions**

Agencies stoke FOMO, yet most expensively marketed "GEO measures" are just good classic SEO. Cynical dismissal during the trough is the same error. Demand evidence either way.

3 **Chasing GEO "hacks"**

Most are unproven or repackaged SEO. The durable levers are retrievability, quotable content, and earned brand presence.

INSTRUMENT FOR NOISE

Fix a prompt library (30 to 50 across brand, category, and comparison queries), run it weekly across engines, report monthly, and smooth with rolling averages. Track share of voice, not rank. Rolling averages are what turn a noisy surface into a trend you can act on.

HONEST UNCERTAINTY

WHAT NOBODY KNOWS YET

IN PLAIN WORDS

This is an honest list of the things even the experts have not figured out yet. Because AI answers change so much, it is genuinely hard to prove what works, and the rules keep shifting as the tools get updated. Knowing these limits keeps you from trusting anyone who claims to have it all solved.

- **Non-determinism makes attribution hard.** Identical prompts produce different brand lists over 99% of the time. Any before/after case study on small samples is indistinguishable from noise, and almost no published case study meets the bar.
- **The flagship research may not generalize.** The paper's 30-40% gains came from a controlled sandbox on 2023-era models. Whether quotation addition still moves citations in today's multi-round fan-out pipelines is publicly untested.
- **Correlation is not causation for mentions.** The 0.664 figure could partly reflect that big, mentioned brands are better at everything. No public controlled experiment isolates "earn 100 mentions, gain X% visibility".
- **The mechanics churn under your feet.** Citation counts per response, search-trigger rates, and fan-out behavior all shifted materially within a year. Principles (retrievability, quotability, entity consistency) age better than tricks.
- **Personalization is unmeasurable from outside.** Memory, account context, and location shape answers per user; tools query from clean accounts and may not see what real prospects see.
- **There is no ChatGPT Search Console.** All measurement is outside-in sampling plus referral crumbs. Prompt-volume estimates sold by tools are modeled, not observed.
- **Contested tactics:** llms.txt (no consumer), schema as a citation lever (weak correlational evidence; Part II's schema section carries the detail), tone hacks for "writing for AI", and blocking crawlers as negotiating leverage (publishers who blocked lost visibility without compensation).

PART IV

THE STACK

The tooling under all three: GA4, Tag Manager, Search Console, research and AI-visibility tools, wired together.

WHY THE STACK

MEASURE BEFORE YOU OPTIMIZE

IN PLAIN WORDS

Before you can make a website better at showing up in search, you have to know how it is doing right now. This section is about the tools that collect that information, and the good news is the whole basic set is free. You cannot fix a problem you cannot see.

SEO, AEO, and GEO all rest on instrumentation. You cannot improve a ranking, an answer share, or an AI mention rate you are not measuring, and you cannot measure it without tools that collect behavior on your own site, report what the engines see, research the market, audit the build, and track how often AI names you. This companion lays out that stack: what each tool is for, what it costs, and the order to wire it in. The good news for a beginner: the entire measurement backbone is free.

$0

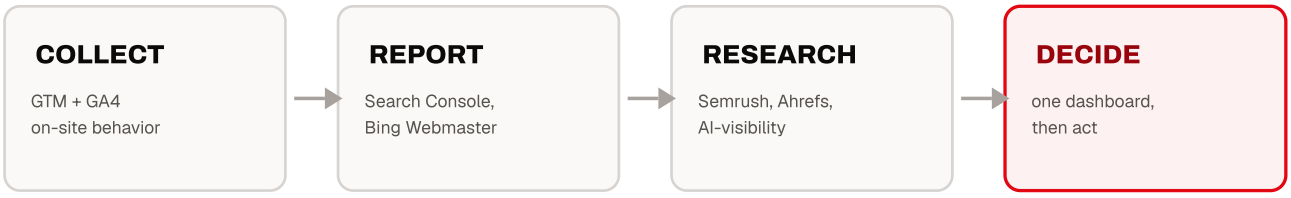
The core stack (Tag Manager, GA4, Search Console, Bing Webmaster Tools, PageSpeed) is free. You can instrument everything before paying a cent.

GA4

The only Google analytics now: Universal Analytics stopped processing data on July 1, 2023, and is event-based, not session-based.

5

Jobs a stack does: collect on-site behavior, measure search, research the market, audit the build, and track AI visibility.



The stack is a data flow: collect what visitors do, report what engines show, research the market, then decide from one place. Most of it is free.

This guide is the tooling layer beneath the trilogy. The **SEO**, **AEO**, and **GEO** guides name these tools in context; here they are gathered into one stack you can build in an afternoon.

THE MAP

WHICH TOOL FOR WHICH JOB

IN PLAIN WORDS

Do not start by picking a famous brand. Start by asking what job you need done, like counting visitors or checking your Google ranking, and then pick the tool that does that job. Every job here has a free way to begin, so you rarely need to pay early on.

Start from the job, not the brand. There are five jobs; each has a free way to start and a paid way to scale. Match the tool to the layer it serves.

JOB	FREE START	PAID UPGRADE	SERVES
Collect on-site behavior	GA4 + Tag Manager	Amplitude, Mixpanel	all three
Measure Google search	Search Console	Semrush, Ahrefs (rank tracking)	SEO, AEO
Measure Bing / ChatGPT surface	Bing Webmaster Tools + IndexNow	—	GEO
Research keywords + competitors	limited free lookups	Semrush, Ahrefs	SEO, AEO
Audit technical + crawl	Screaming Frog (500 URLs)	Screaming Frog paid, Sitebulb	SEO
Measure field performance	PageSpeed Insights / CrUX	Vercel Speed Insights (RUM)	SEO
Track AI visibility	your own prompt script	Profound, Peec AI, Otterly	GEO, AEO
Report and decide	Looker Studio	—	all three

ESTIMATOR VS GROUND TRUTH

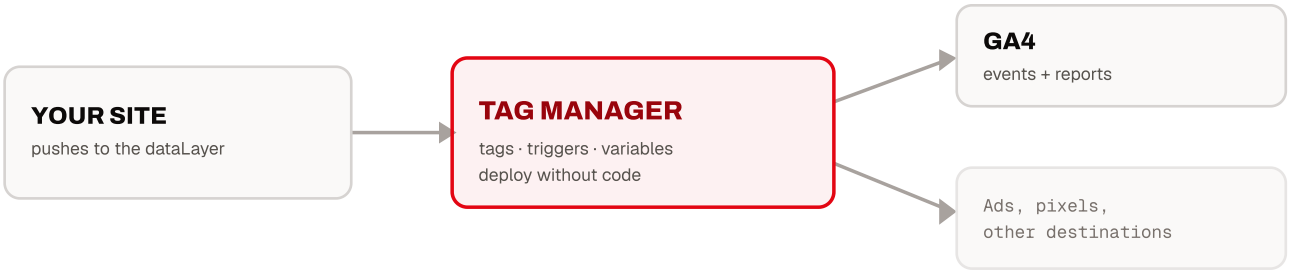
Search Console and GA4 report your real data. Semrush and Ahrefs estimate volumes and positions from a panel and an index. Use estimators to find opportunities; use your own tools to measure what actually happened.

THE BACKBONE

TAG MANAGER AND GA4

IN PLAIN WORDS

Google Analytics 4, or GA4, is a tool that counts what people do on a website, like which pages they open and whether they sign up. Google Tag Manager, or GTM, is a control panel that lets you turn that tracking on without editing the site's code. You set them up once and they quietly gather the numbers for you.



Tag Manager is the container you install once; from it you deploy and change tags (GA4, ads, events) without touching code. GA4 is where the events land.

Google Tag Manager (GTM) is not analytics. It is a container: you add one snippet to your site, then deploy and edit tags from its UI without a code change or a redeploy. Triggers decide when a tag fires; variables pull data from the dataLayer. It is how you manage GA4, ad pixels, and custom events in one place.

GA4 is the analytics itself, and it is event-based: every interaction (page_view, click, a custom "signup") is an event, not a pageview in a session the way Universal Analytics counted. You define the key events that matter (signup, contact, purchase) and GA4 reports them. Wire GA4 through GTM, not by pasting the tag directly, so future changes stay in one container.

THE AI-REFERRAL SEGMENT

In GA4, build a segment or referral view for chatgpt.com, perplexity.ai, and gemini.google.com. Volumes are small but the intent is high, and it is the only way to see AI-assistant traffic apart from classic organic. This is the measurement the SEO and GEO guides point back to here.

Worked example: track a signup end to end

The abstract "GTM fires a tag" becomes concrete when you follow one conversion through the chain. This is the deep setup the SEO and GEO guides point back to.

```
// Your app pushes an event when the signup succeeds
window.dataLayer = window.dataLayer || [];
window.dataLayer.push({ event: "sign_up", method: "email" });
```

From that one line the chain runs itself: **(1)** the app pushes `sign_up` to the dataLayer; **(2)** a GTM trigger listens for that custom event; **(3)** it fires a GA4 event tag named `sign_up`; **(4)** in GA4 you mark `sign_up` a key event; **(5)** it now shows in Conversions and can anchor any report or dashboard. Add a destination or change the trigger later without touching the app: that is the whole point of routing through the container.

BUILD ONE DASHBOARD

In Looker Studio (free), connect Search Console and GA4 and build four tiles: impressions and clicks over time (Search Console), key-event conversions (GA4), AI-referral sessions (GA4 filtered to the chatgpt, perplexity, and gemini referrers), and top landing pages. One link you check weekly beats six tabs nobody opens.

FROM THE ENGINES

SEARCH CONSOLE AND BING

IN PLAIN WORDS

Google and Bing each give you a free tool that shows how your site looks to them: what people typed to find you, and whether your pages are even listed. Because these numbers come straight from the search engines, they are real, not guesses. Bing matters more than you would expect, because ChatGPT's search pulls from it.

These are the tools the search engines give you, so they report ground truth, not estimates. Both are free and non-negotiable.

Google Search Console

The only source of your real Google query data: impressions, clicks, average position, and the queries themselves, plus the Page Indexing report, the Core Web Vitals report, rich-result status, and URL Inspection. Watch for the impressions-up-clicks-down divergence that signals AI Overview exposure (AI feature traffic is folded into the "Web" type with no separate breakout). Start every keyword and content decision from this data, not from an estimator.

Bing Webmaster Tools

The same instrument for Bing, and it matters more than its market share suggests because ChatGPT search is Bing-derived: a page absent from Bing is absent from ChatGPT answers whatever its Google rank. Verify your site, submit your sitemap, and adopt IndexNow so new and changed URLs are picked up fast. This is first-class GEO infrastructure, covered in Part III and instrumented here.

RESEARCH AND AUDIT

SEMRUSH, AHREFS, SCREAMING FROG

IN PLAIN WORDS

Semrush and Ahrefs are research tools that show which words people search for and what your competitors rank for, though their numbers are smart estimates, not exact counts. Screaming Frog is a program that crawls your whole site the way a search engine would and lists problems like broken links and missing labels. This section also covers checking how fast your pages load for real visitors, which affects your ranking.

Research: Semrush and Ahrefs

The two heavyweight suites: keyword research, competitor analysis, backlink data, rank tracking, and site audits. They are estimators (modeled volumes and positions from a panel and a crawl index), so treat their numbers as directional and confirm against your own Search Console data. Both have limited free lookups and now ship AI-visibility add-ons (Semrush AI Toolkit, Ahrefs Brand Radar). For a small site, a single suite on a focused month beats paying for both year-round.

Audit: Screaming Frog and Sitebulb

Desktop crawlers that behave like a search bot. Screaming Frog is free up to 500 URLs (paid for more and for JavaScript rendering); Sitebulb adds guided reports. This is the tool for the rendering check Part I describes: crawl once with JavaScript off and once with it on, then diff, to find content that only exists after hydration. It also surfaces broken links, redirect chains, missing metadata, and orphan pages at scale.

Performance: PageSpeed Insights and CrUX

PageSpeed Insights shows both a lab score and, more importantly, the field data from the Chrome UX Report (CrUX): real-user Core Web Vitals at the 75th percentile. Rankings use the field data, so this is the number that counts. For a continuous trend rather than a spot check, wire real-user monitoring with the web-vitals package or Vercel Speed Insights.

THE NEW CATEGORY

AI-VISIBILITY TOOLS

IN PLAIN WORDS

These are new tools that check whether AI assistants like ChatGPT actually mention your site when they answer a question. They work by asking the AI the same set of questions over and over and counting how often your name comes up. This matters because more people now ask an AI instead of typing into Google.

A category that did not exist before 2024: tools that measure whether AI answers name and cite you. They all work the same way, which is worth understanding before you buy one: they run a fixed basket of prompts against multiple LLMs on a schedule, parse which brands get mentioned and cited, and compute a share of voice over time.

The landscape: **Profound** (enterprise leader), **Peec AI** (mid-market, competitor benchmarking), **Otterly.AI** (entry tier, around $29 per month), and **Scrunch AI**, plus incumbent add-ons in **Semrush AI Toolkit** and **Ahrefs Brand Radar**. The free start is your own prompt panel run through an API script. Whatever you pick, demand what Part III specifies: disclosed sample sizes and re-run counts, per-engine breakdowns, a citation-versus-mention distinction, and no single-number "AI rank" theater.

SAMPLING, NOT RANKING

LLM answers are non-deterministic: the same prompt rarely returns the same list twice. A tool that reports a stable "AI rank" is selling decoration. The honest metric is a mention rate across many repeated runs per engine, read as a trend.

THE BUDGET

FREE, FREEMIUM, PAID

IN PLAIN WORDS

This section sorts every tool into three groups: free, free with limits, and paid. The advice is simple: use up the free tools first, since they cover almost everything a small site needs. Only pay once you hit a wall the free ones cannot get past.

FREE	FREEMIUM	PAID
Tag Manager	Semrush (limited)	Semrush / Ahrefs full
GA4	Ahrefs (limited)	Screaming Frog paid
Search Console	Otterly.AI (entry)	Sitebulb
Bing Webmaster Tools	Vercel Speed Insights	Profound
PageSpeed / CrUX		Peec AI
Screaming Frog (500)		AI Toolkit add-ons
Looker Studio		
your own prompt script		

You can instrument the whole measurement backbone for $0. Pay only when you need research scale, crawls past 500 URLs, or serious AI-visibility tracking.

The order of spend matters. Exhaust the free column first: Tag Manager, GA4, Search Console, Bing, PageSpeed, and a Looker Studio dashboard cover everything a small site needs to see. Move to freemium when a free tier's limits pinch. Buy paid tools only for a specific job the free stack cannot do: research at scale, large crawls, or tracking your brand across AI answers.

THE PLAYBOOK

BUILD YOUR STACK, IN ORDER

IN PLAIN WORDS

This is the step-by-step order for setting everything up, from your first day to a full setup. If you are starting fresh, go top to bottom; if some basics are already running, jump to your level. Each step tells you exactly how to know it worked.

Wire it in this order. Follow top to bottom if you are starting from nothing; jump to your level if the basics are already live. Every step names the tool and how you know it worked.

LEVEL 1 STARTER

free instrumentation

1

Install Google Tag Manager.

Create a container at `tagmanager.google.com` and add its snippet to your site's head and body.

Done when: GTM Preview mode connects to your live site.

2

Set up GA4 through GTM.

Create a GA4 property at `analytics.google.com`, then add a GA4 Configuration tag in GTM firing on all pages.

Done when: GA4 Realtime shows your own visit.

3

Verify in Search Console and Bing Webmaster Tools.

Add both, confirm ownership (DNS or the GA4/GTM tag), and submit your sitemap. Turn on IndexNow in Bing.

Done when: both report your pages indexed.

4

Filter internal traffic.

In GA4 admin, add a data filter for your team's IP so your own visits do not pollute the numbers.

Done when: your own sessions stop appearing in reports.

LEVEL 2 PRACTITIONER

measure what matters

5

Track real conversions as GA4 key events.

Define events for signup, contact, or purchase with `GTM triggers`, then mark them key events in GA4.

Done when: GA4 counts the event when you complete the action.

6

Build AI-referral segments.

In GA4, segment sessions from `chatgpt.com, perplexity.ai, gemini.google.com`.

Done when: you can see AI-assistant traffic apart from classic organic.

7

Add one research tool.

Start on a free tier of `Semrush or Ahrefs`, but seed it with your real queries from Search Console first.

Done when: you have a keyword and competitor list grounded in your own data.

8

Audit with Screaming Frog.

Crawl the site (free to 500 URLs) with `JavaScript off, then on`, and diff the two.

Done when: no surprise gaps between raw and rendered content.

LEVEL 3 ADVANCED

instrument the edge

9

Wire real-user monitoring.

Add the `web-vitals` package or Vercel Speed Insights for a p75 field trend.

Done when: you have a real-user vitals trend, not a one-off lab score.

10

Stand up AI-visibility tracking.

Run a prompt panel (your own script or `Profound, Peec, Otterly`), sampled many times per engine.

Done when: you have a mention-rate trend per engine, not a single reading.

11

Read your server logs.

Track Googlebot and AI-bot (`ChatGPT-User, PerplexityBot`) hits to see what is actually fetched.

Done when: you can name which pages the bots pull.

12

Build one dashboard.

Pull Search Console, GA4, and AI-visibility into a single `Looker Studio` view.

Done when: you check one dashboard, not six browser tabs.

ANTI-PATTERNS

THE MISTAKES THAT COST THE MOST

IN PLAIN WORDS

This is a list of the common slip-ups that waste time and money, like paying for expensive tools before using the free ones, or letting your own team's visits count as real traffic. Read it as a checklist of what not to do. Dodging these mistakes saves you more than any single tool ever will.

- Expecting Universal Analytics metrics in GA4. GA4 is event-based, not session-and-pageview based; learn the event model instead of fighting it.
- Not filtering internal and bot traffic, so your own team's visits inflate every number and every decision built on them.
- Tracking vanity metrics (raw pageviews, bounce rate) instead of defining key events and qualified conversions.
- Buying Semrush, Ahrefs, or Profound before exhausting the free ground-truth stack that already answers most questions.
- Trusting estimator volumes and positions as fact. Search Console and GA4 are your real data; the suites are modeled.
- Reporting single-day AI-visibility readings. Non-deterministic answers demand repeated sampling and a trend, or it is noise.
- Letting the GTM container become an undocumented pile of tags. Name tags, note what each does, and remove dead ones.
- Living in six tool tabs instead of one dashboard, so nobody actually looks and the data changes nothing.

THE HEAD GAME

DATA WITHOUT DECISIONS

IN PLAIN WORDS

Having lots of numbers can trick you into staring at charts instead of actually deciding anything. This section names the mental traps, like only noticing the data that agrees with you, so you can catch yourself doing it. A tool can measure things, but it cannot make the choice for you.

The stack's psychological trap is not impatience, it is the opposite: endless looking. More dashboards do not make better decisions, and the biases that corrupt analysis are predictable. Name them so you can catch yourself.

What to expect

You will have more data than you can act on. The goal is not to see everything; it is to answer the few questions that change what you do. A tool measures; it does not decide.

Traps to avoid

1 **Analysis paralysis**

Chasing every possible insight leads to inaction. Pick the handful of metrics tied to a decision and ignore the rest.

2 **Confirmation bias**

If you believe a change is working, you notice the data that agrees. Dashboards strip context, which makes it worse.

3 **Cherry-picking**

The flattering date range, segment, or channel is always available. Decide the comparison before you look.

4 **Survivorship bias**

Published case studies are the winners; the sites the same trick failed for never wrote a post. Ask "where did this not work?"

5 **Sunk cost**

Effort already spent on a losing page or tool is not a reason to keep spending. Set a kill-criterion up front.

THE ONE HABIT

Before you open the dashboard, write down the number that would change your mind. It defuses confirmation bias, cherry-picking, and sunk cost in one move, and it turns a report from theater into a decision.

WHAT TO PAY FOR

WHAT IS WORTH PAYING FOR

IN PLAIN WORDS

Here is the honest bottom line: the free tools cover what a small or new site actually needs. Pay only when one specific job outgrows them, like researching thousands of keywords or crawling a very large site. Buy the job you need done, not the brand name.

The honest verdict: the free stack measures everything a small or new site needs. Tag Manager, GA4, Search Console, Bing Webmaster Tools, PageSpeed, Screaming Frog's free tier, and a Looker Studio dashboard cover collection, search truth, technical audit, performance, and reporting for zero cost. Pay when a specific job outgrows them: keyword and competitor research at scale (Semrush or Ahrefs), crawls beyond 500 URLs, or serious AI-visibility tracking across engines (Profound or Peec).

Two caveats. Tool pricing and features churn quickly, so buy the job, not the brand, and re-evaluate yearly. And a tool measures; it does not decide. The stack exists to tell you what to fix and build, which is the work Parts I to III describe. Instrument first, then go optimize.

PART V

THE CLIENT KIT

Working on someone else's site: getting access the safe way without a password, reading the CMS, and editing a live site without breaking it.

WHY THIS GUIDE

YOU WORK ON THEIR SITES, NOT YOURS

IN PLAIN WORDS

The websites you work on belong to your clients, not to you, so the first thing to sort out is not the actual work but how to get into their accounts safely. Do that part well, without ever being the person who broke something or leaked a password, and clients start trusting you with more.

Every other guide in this set assumes the website is yours. This one does not. As a VA or contractor, you almost never own the accounts or the site you are hired to improve. The client owns Search Console, the analytics, the domain, and the CMS, and your whole job happens inside their setup. That changes the first question of any engagement from "what do I do?" to "how do I get in, safely, without becoming the person who broke something or leaked a password?"

This guide answers exactly that. It is the practical layer under the SEO, AEO, and GEO work: how to request access the professional way, what to ask a new client for on day one, how to tell what their site is built on, how to edit it without taking it down, and how to hand everything back cleanly when the engagement ends. Get this part right and the client trusts you with more. Get it wrong once and the relationship never recovers.

0

Passwords you should ever hold. Every serious platform lets the client add you as your own user; that is always the right path.

~43%

Share of all websites running WordPress (W3Techs, 2026). If you learn one CMS deeply, learn this one; it is the site you will meet most.

1 click

What it takes the client to remove you when done, if you set access up correctly. That reversibility is what makes them say yes.

Two ideas run through everything here. First, **least privilege**: ask for the lowest level of access that still lets you do the job, and no more. Second, **you are a guest**: it is their house, their keys, their risk. Behave like the professional who is careful with both, and access stops being a hurdle and becomes the thing clients rehire you for.

The words you will see

This guide stays plain, but a handful of terms come up throughout. Here they are once, in one place, so nothing later trips you up. You do not need to memorize them; just know where to look back.

TERM	IN PLAIN WORDS
CMS	The system a site's content lives in and is edited through, like WordPress or Shopify.
Role / permission	How much a given login is allowed to do, from read-only up to full control.
Least privilege	Asking for the lowest role that still lets you do the job, and no more.
2FA (two-factor)	A second login step on top of the password, usually a code sent to a phone.
Slug	The last, readable part of a page's web address, like the about-us in /about-us.
Staging	A private copy of the site where you test changes before they go live.
Cache	A saved copy of a page served for speed; your edit may not show until it is cleared.
Headless	A site whose content and its design are split apart, so a developer controls how the content appears.

THE ACCESS RULE

YOUR OWN LOGIN, LOWEST ROLE, NEVER THEIR PASSWORD

IN PLAIN WORDS

Never take a client's password. Instead, ask them to add you as your own separate user, with only the access you actually need. It is safer for them, and it means you can never be blamed for breaking something you were never able to touch.

There is one rule, and it holds on every platform: **get access under your own login, at the lowest role that does the job, and never take the client's password.** Google, Meta, Shopify, WordPress, Microsoft, and every registrar have a built-in way to add an outside collaborator as a separate user or linked partner. That native path is always the correct one. Asking for the client's password is the amateur move, and security-aware clients notice.

THE RIGHT WAY ✓	THE WRONG WAY ✗
They add YOUR email as a user	You share one password between people
You get a limited role (Editor, not Admin)	One login with full access to everything
Your own 2FA, on your own phone	Their 2FA breaks, or you text them for codes
The logs show exactly what you did	No record of who changed what
They revoke you in one click, keep ownership	They must reset the password to remove you

Same access, two ways to get it. The left column is what a professional asks for; the right column is the pattern that gets sites hacked and VAs blamed.

Why this is not just etiquette

These are the reasons to give the client, because they are the reasons a nervous client says yes:

1 **Two-factor actually works**

If you log in as the client, their 2FA code goes to their phone, so either they disable 2FA (dangerous) or you text them for a code every session (unworkable). Your own login keeps everyone's 2FA intact. This is the most persuasive point in practice.

2 **The blast radius stays small**

A shared password is one leak away from total compromise. Separate logins isolate the damage.

3 **There is an audit trail**

Platform logs show who did what. Under a shared login every action reads as "the owner," so nobody can tell your work from theirs, and a dispute has no evidence.

4 **Offboarding is clean**

When the work ends, the client removes one user. Nothing to rotate, no lingering keys, no "what did they still know?"

5 **They stay in control**

The owner can downgrade or remove you at any time without locking themselves out. Say this out loud: *you can remove me anytime with one click, and you never give up ownership.*

Privilege climbs left to right. The job almost always lives on the Editor rung: edit any page and the SEO fields, but not the plugins, theme, users, or billing. Climb higher only when a specific task forces it.

WHEN A SHARED LOGIN IS TRULY UNAVOIDABLE

Some legacy tools and hosting panels have no multi-user option. Then, and only then, share the credential through a password manager's shared vault (1Password guest access, or a Bitwarden Organization collection), never over email, chat, or a doc. The vault gives you use of the login without the plaintext living in a message, and the client can revoke it instantly. Put the 2FA seed in the same vault, and rotate the password the day the engagement ends.

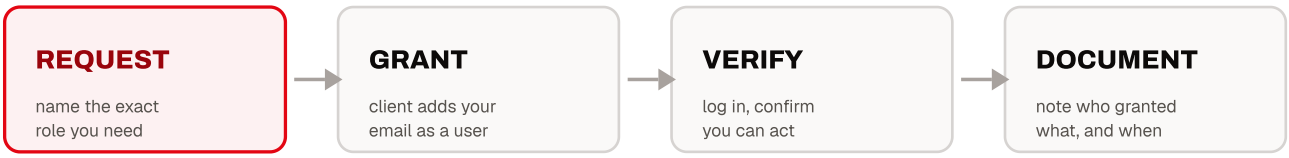
THE ONBOARDING REQUEST

WHAT TO ASK FOR, IN ONE MESSAGE

IN PLAIN WORDS

On your first day, send one clear message that lists every account you need, the exact level of access for each, and the email to add you under. Asking for it all at once, instead of a little at a time over weeks, is what gets you working in an hour instead of dragging on for a fortnight.

Do not drip-feed access requests over two weeks. On day one, send one organized message listing every account you need, the exact role for each, and the email address to add. Clarity here is the difference between access in an hour and a fortnight of back-and-forth. Below is the flow, then a checklist you can copy, adapt to the engagement, and paste.



Every access follows the same four beats. The last one, document, is the step most people skip and later wish they had, when it is time to prove or undo access.

- Search and analytics
- **Google Search Console** — Full user
- **Google Analytics 4** — Editor (or Analyst for read-only reporting)
- **Google Tag Manager** — Publish, on the container
- **Bing Webmaster Tools** — Read/Write
- CMS and hosting
- **WordPress** — Editor (Administrator only if configuring plugins or technical SEO)
- **Shopify** — Collaborator account, scoped to Products, Themes, Online Store, Analytics
- **Domain registrar / DNS** — delegated DNS-only access, or the client makes changes from your exact instructions

- Ads and social
- **Google Business Profile** — Manager (the client stays Primary Owner)
- **Google Ads** — Standard user, or a manager-account link via their 10-digit Customer ID
- **Meta (Facebook Page + Instagram)** — Partner access to the assets, via your Business Portfolio ID
- Research tools
- **Semrush / Ahrefs** — you bring your own; only take a client seat if they choose to add and pay for one

ALWAYS INCLUDE

The email address the client should add. Use the same Google account for every Google product, your Microsoft account for Bing, and your partner IDs for Shopify and Meta. The single most common reason a grant fails is the wrong email, or a personal address added where a work one was meant.

PLATFORM BY PLATFORM

ACCESS, PLATFORM BY PLATFORM

IN PLAIN WORDS

Each website tool, like Google Search Console or Shopify, has its own way to add you and its own name for the access level you want. This section is a lookup list: for each one, the exact role to ask for and the exact words to send, so you never have to guess or fumble.

Each row below is the role to ask for and who grants it. After the table, the exact request to send, per platform, written so a non-technical client can follow it.

PLATFORM	ASK FOR	WHO GRANTS	NOTE
Search Console	Full user	Owner, in Settings	covers all day-to-day SEO
Google Analytics 4	Editor / Analyst	Owner, in Admin	Editor to configure, Analyst to report
Tag Manager	Publish (container)	Owner, User Management	Publish so your tags go live
Business Profile	Manager	Owner, People and access	client stays Primary Owner
Google Ads	Standard, or MCC link	Owner, Access and security	manager link is the agency way
WordPress	Editor	Owner, Users	Admin only to configure plugins
Shopify	Collaborator	You request, owner approves	does not use a staff seat
Bing Webmaster	Read/Write	Owner, Settings > Users	can import from Search Console
Meta (Page/IG)	Partner access	Owner, Business Settings	never via a personal profile
Registrar / DNS	DNS-only, delegated	Owner, or does it for you	often a one-off TXT record

Google Search Console

Ask for **Full user**. It shows all data and lets you submit sitemaps, request indexing, and read every report, without the power to manage other users. If the client set up a **Domain property** (verified by a DNS record, it covers every subdomain and both http and https), you inherit everything. A **URL-prefix property** only covers that exact address, so check which exists.

Copy-paste request

Please add me as a **Full user** in Search Console: open Settings > Users and permissions > Add user, enter [my email], set permission to **Full**, and Add.

Google Analytics 4

Ask for **Editor** if you need to configure the property (create key events, custom dimensions, link Search Console or Ads), or **Analyst** if you only build reports and read data. Avoid Administrator. Grant it at the property level, not the account level, unless the client wants you across every property.

Copy-paste request

In GA4, please go to Admin > Property Access Management > +, add [my email], and set the role to **Editor** (or Analyst if you prefer read-only).

Google Tag Manager

Container permissions run Read, Edit, Approve, Publish. Ask for **Publish** if your tracking changes need to go live by your hand, which is normal when you are wiring GA4 tags, events, or schema through GTM. If the client wants to review first, ask for Approve or Edit and let them publish. You rarely need account-level Admin.

Copy-paste request

In Tag Manager, please add me with **Publish** rights on the container: Admin > User Management > +, add [my email], set Account to User and Container to **Publish**, Invite.

Google Business Profile

Ask for **Manager**. It lets you edit the listing, hours, photos, posts, and reviews, while the client stays **Primary Owner**. Losing primary ownership of a Business Profile is a known way businesses get locked out of their own listing, so never ask to be made owner.

Copy-paste request

Please add me as a **Manager** on your Business Profile (you stay Primary Owner): Settings > People and access > Add, enter [my email], choose **Manager**, Invite.

WordPress

Ask for your own **Editor** user. Editor can publish and edit every page and post and the on-page SEO fields, but cannot touch plugins, the theme, settings, or users, which is exactly the boundary you want. Ask for **Administrator** only if the task is installing or configuring an SEO plugin or technical settings, and say why. Never accept the shared `admin` login.

Copy-paste request

Please create my own user: Users > Add New User, email [my email], Role **Editor** (or Administrator only if I will configure plugins or technical SEO). Please do not share the main admin login.

Shopify

Do not ask for a staff account, which uses up one of the client's limited seats. Ask for a **Collaborator account** instead: you request it from your Shopify Partner dashboard, it does not count against staff seats, and you scope it to only the sections you need. Many stores require a 4-digit collaborator code before a request can be sent, so ask for that too.

Copy-paste request

I will send a **Collaborator request** from my Partner account (it will not use a staff seat). If collaborator requests are locked, please send me your **4-digit code** from Settings > Users and permissions, then approve my request and grant Products, Themes, Online Store, and Analytics.

Meta, Google Ads, DNS

For **Meta**, connect through Business Partner access, never a personal profile: the client adds your Business Portfolio ID under Business Settings and assigns the Page, Instagram, and any ad account. For **Google Ads**, either be added as a Standard user or link their account to your manager (MCC) account using their 10-digit Customer ID. For **DNS**, the safest pattern for a one-off (like verifying a Search Console Domain property) is to send the exact record and have the client paste it; for standing access, ask for a scoped DNS-only role (Cloudflare Members, GoDaddy Delegate Access, Namecheap Share Access), never the registrar login.

READING THE SITE

WHICH CMS IS THIS?

IN PLAIN WORDS

A CMS, or content management system, is just the program a website is built and edited in, like WordPress or Shopify. Before you promise a client any work, figure out which one their site runs on, because a free browser tool can tell you in under a minute and each system is a completely different job to edit.

Before you quote content work, know what the site is built on, because a WordPress site and a headless Next.js site are two completely different jobs. You can identify almost any site in under a minute with two checks, then confirm with a third.

- **Run Wappalyzer or BuiltWith.** The browser extension loads the site and reads its CMS, framework, e-commerce platform, analytics, and host from public signals. This answers 90% of sites instantly.
- **Check for the tells.** WordPress: a `/wp-admin` login and asset paths under `/wp-content/`. Shopify: `cdn.shopify.com` assets and `/products/` URLs. Wix: `wixstatic.com`. Squarespace: `squarespace.com` assets. Webflow: `data-wf-page` attributes.
- **View source for the generator tag.** Many platforms self-identify in `<meta name="generator">`: "WordPress 6.x", "Squarespace", "Wix", "Ghost", "Hugo". WordPress can hide this, so absence is not proof.

CROSS-CHECK BEFORE YOU TELL THE CLIENT

Confirm two signals before you name the platform. A `_next/` asset path with no `/wp-admin` and no CMS generator is the signature of a custom or headless build, which is your cue that a developer is involved and some SEO asks will be dev tickets, not VA tasks. Knowing this before you quote saves an awkward conversation later.

THE LANDSCAPE

WHAT YOU WILL ACTUALLY MEET

IN PLAIN WORDS

Websites split into two groups: ones you can update on your own, and ones where a developer has to handle the structural parts for you. This section walks through the systems you will actually run into and shows which group each one falls in, so you know before you say yes whether you can finish the job alone.

There are really two families of site. On one side are platforms where a VA can be productive alone with a limited role. On the other are developer-built sites where you can fill in fields but cannot change structure or SEO without the developer. The single most useful thing to know about any CMS is which side of that line it sits on.

The line that matters is not "easy" versus "hard", it is "can I finish the job alone?" WordPress sits in the sweet spot: a limited role, free SEO plugins, and no developer needed for routine work.

CMS	ASK FOR ROLE	SEO CONTROL	THE GOTCHA
WordPress	Editor	strong, via Yoast or Rank Math	ask if a page builder (Elementor) is in use
Shopify	Collaborator	title, meta, slug on every item	changing a handle changes the URL, redirect it
Webflow	Content editor	strong and native	never touch the Designer; legacy Editor is retiring in 2026
Squarespace	Content Editor	adequate, per page	little granularity, SEO settings often need Admin
Wix	scoped role	improved, per page	closed platform, cannot swap templates after publish
Ghost	Editor / Author	native per-post meta	theme changes need a developer
Sanity, Contentful, Strapi, Payload	editor (dev-set)	only the fields a developer exposed	no visual page building, structure is code
Next.js / MDX / static	via an editing layer, if any	whatever is hardcoded	a bad edit can fail the build and down the site

WordPress, in a little more depth

Because you will meet it most, know its shape. Content lives in **Posts** and **Pages**, edited in the **Gutenberg** block editor by default. Many sites layer a page builder like **Elementor** on top, which replaces that editor with a drag-and-drop canvas, so always ask "is there a page builder?" before you quote. SEO comes from a plugin: **Yoast** or **Rank Math**, both of which add a per-page panel for the SEO title, meta description, slug, and schema. **Advanced Custom Fields** (ACF) turns the edit screen into labeled fields rather than a free body, which is safer but means you cannot restructure the page. And if you edit a page and the change does not show, the site is probably **cached** (WP Rocket and similar), so clear the cache and hard-refresh.

The headless reality

On a headless CMS like **Sanity**, the content backend and the front end are separate: you edit fields in a studio, and a developer's code decides how they render. You *can* edit the text and images a developer set up as fields. You *cannot* move a section, change the layout, add a content type, or add an SEO field that was never built. If the client wants the hero moved down and the site is headless, that is a developer request, not a VA task. Contentful and Strapi are friendlier for editors; Payload and raw MDX are the most developer-owned.

EDITING SAFELY

EDITING SAFELY ON A LIVE SITE

IN PLAIN WORDS

The website is live and real visitors are on it while you work. This section is a short list of habits that stop a small edit from taking the whole site down, like making a copy first and never touching the code.

It is their live site, and real visitors are on it while you work. These habits keep a routine edit from becoming an outage or a lost afternoon. None of them are advanced; skipping them is what causes the damage.

- **Use staging when the change is non-trivial.** Most managed WordPress hosts, Shopify (a duplicated unpublished theme), and Webflow (preview or branches) let you work on a copy, review, then push live. Never experiment on production.
- **Draft, preview, then publish.** Every serious CMS has a draft state and a preview. Look at the rendered page before you publish; do not edit live and hit save blind.
- **Duplicate before you edit.** Copy the page, product, or template first. If the edit goes wrong, the original is untouched and you revert in seconds.
- **Never touch theme or template code.** WordPress Theme File Editor, Shopify `.liquid` files, and the Webflow Designer can break the entire site with one mistake. Stay in the content and SEO-field layers; code is a developer's job.
- **Know that revisions and caching have limits.** WordPress keeps post revisions you can restore, but a cache or CDN can make a bad edit look fine or a good edit look broken. Clear the cache when you check your work.
- **Confirm a backup exists before anything bulk.** Before mass edits or plugin changes, check the host or a backup plugin has a recent snapshot you can roll back to.
- **Change URLs carefully.** Editing a slug changes the URL. Set a 301 redirect (Shopify auto-offers one; WordPress via a redirect plugin; Webflow in its redirects panel) so you keep the SEO equity and avoid a 404.

THE ONE-LINE SAFETY RULE

If you cannot undo it in under a minute, do not do it on the live site. Duplicate, stage, or ask the developer first. This single test prevents almost every catastrophic edit.

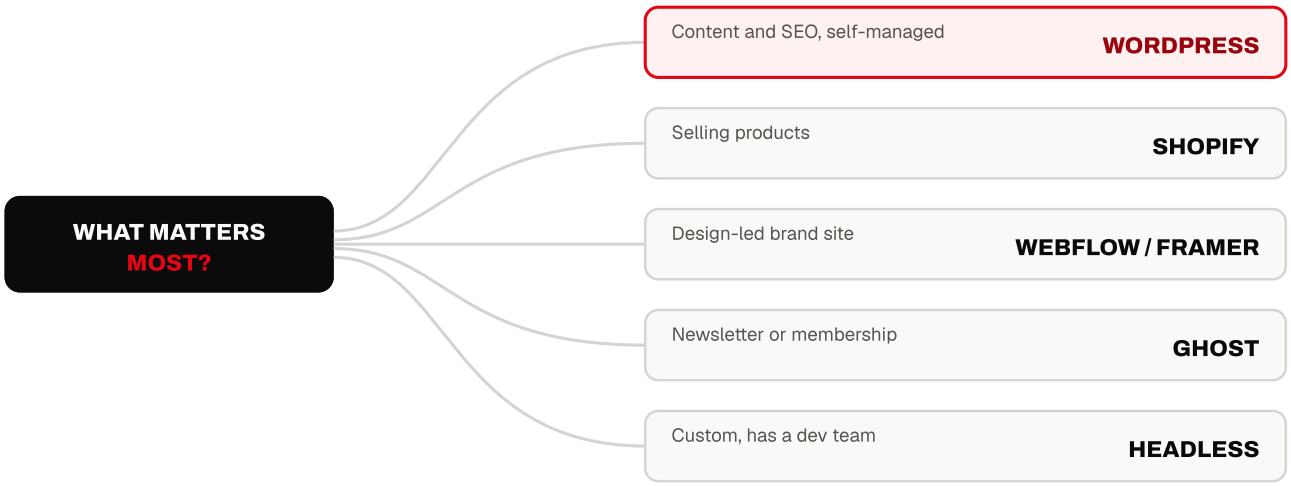
WHAT TO BUILD ON

WHEN THEY ASK WHAT TO BUILD ON

IN PLAIN WORDS

Sooner or later a client will ask which platform they should build their next website on. Start from what matters most to them, whether that is selling, writing, or design, and this section gives you a simple answer for each, with WordPress as the safe pick when they are not sure.

Sooner or later a client asks what they should build their next site on. Start from the priority, not the brand name. There are really five answers, and one safe default.



Five priorities, five answers. WordPress is the safe default whenever content and SEO are the point and the client wants to manage day-to-day and hire help easily.

The line to give a client: *if you want to control content and SEO yourself and hire help easily, WordPress; if you are selling, Shopify; if design is the point and you will not need a plugin ecosystem, Webflow or Framer; if a developer built you something custom, keep that developer on retainer.* Say the last part plainly, because a headless or custom build means some of the SEO work they want will need the developer, and it is better they hear that from you at the start than discover it in an invoice.

FREE VERSUS PAID

THE TOOLS THAT COST MONEY

IN PLAIN WORDS

Almost everything you need is free and good enough, so you do not have to spend a client's money to do solid work. This section lists the few tools that charge a fee and, next to each, tells you whether the free version will do the same job.

Most of what a VA needs is free and genuinely good enough. The free tier of Yoast or Rank Math covers on-page SEO titles, meta descriptions, slugs, sitemaps, and basic schema. Recommend a paid tool only when the client needs a specific paid feature, and say which feature and why. Prices below are approximate 2026 figures and drift, so confirm on the vendor page before you quote them.

TOOL	ROUGH COST	FREE ENOUGH INSTEAD?
Yoast SEO Premium	~$129/yr per site	yes, free tier covers most on-page work
Rank Math Pro	~$108/yr, unlimited personal sites	yes, and far cheaper across many sites
WP Rocket (caching)	~$59/yr, single site	yes, free caching plugins exist
Elementor Pro	~$59/yr up to agency tiers	free Elementor works for basic builds
Shopify SEO apps	~$5 to $40+/mo each, recurring	often yes, native SEO fields are free
Framer editor seats	~$10/mo content editor, ~$20/mo editor	needed to edit at all on paid plans

THE AGENCY-PRICING TELL

If you manage many client sites, Rank Math's unlimited-sites pricing is dramatically cheaper than Yoast's per-site model. Recognizing which plugin a client already runs, and what it costs to scale, is itself a piece of advice worth giving.

THE PLAYBOOK

ONBOARD A CLIENT, IN ORDER

IN PLAIN WORDS

This is the whole start-to-finish routine for taking on a new client, laid out in the order you actually do it: get in safely, learn the site, then settle into a steady monthly rhythm. Follow it top to bottom your first time and you will not miss a step.

Run a new engagement in this order. Follow top to bottom for your first client; jump to your level once the rhythm is second nature. Every step names the exact place to act and how you know it worked.

LEVEL 1 DAY ONE

get in, safely

1

Send one access request.

Paste the onboarding checklist from `section 03`, trimmed to this engagement, with your email and partner IDs.

Done when: the client has one clear message, not a trickle of asks.

2

Verify each grant as it arrives.

Log in to `each platform` and confirm you can actually do the thing you need, not just see it.

Done when: you have acted once in every account, under your own login.

3

Identify the CMS.

Run `Wappalyzer`, confirm with a second signal, and note whether a developer is involved.

Done when: you can name the platform and whether SEO asks are VA or dev tasks.

LEVEL 2 WEEK ONE

get oriented

4

Baseline the site.

Pull the starting numbers from `Search Console and GA4` so you can prove your impact later.

Done when: you have a dated snapshot of impressions, clicks, and key events.

5

Confirm the editing surface.

Find where content and SEO fields live in `the CMS`, whether a page builder is in use, and whether there is staging.

Done when: you have made one safe test edit and seen it render.

6

Start a change log.

Open a shared doc and record every change: what, where, when. This is your record if anything breaks.

Done when: the client could see exactly what you have touched.

LEVEL 3 **ONGOING** run the loop

7

Work the engagement loop.

Cycle `access, audit, plan, edit, measure, report` on a monthly cadence.

Done when: the client gets a monthly report tied to the baseline.

8

Keep access lean.

Drop any role you no longer need, and never escalate `past Editor` without a task that requires it.

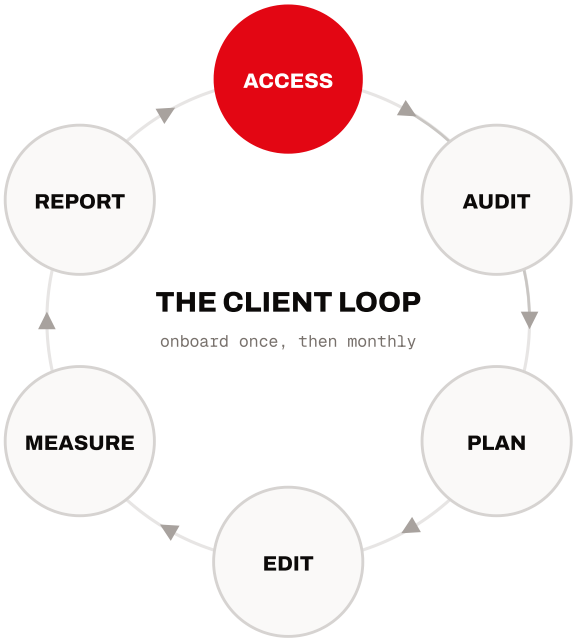
Done when: your access still matches the work, months in.

9

Prepare the clean exit in advance.

Keep the change log and access list current so `offboarding` is a five-minute job, not a scramble.

Done when: you could hand back everything today without loose ends.



After onboarding, the engagement is a loop: audit what is there, plan the changes, edit safely, measure against the baseline, and report. Then round again.

ANTI-PATTERNS

THE MISTAKES THAT COST THE MOST

IN PLAIN WORDS

This is a plain list of the errors that do the real damage, like taking a password or editing code on a live site. Read it once so you can see each one coming, because avoiding these matters far more than any clever trick.

- Taking the client's password to "go faster." It breaks their 2FA, erases the audit trail, and marks you as the person who does not know better. Decline even when they offer.
- Asking for Administrator or Owner when Editor would do. Over-asking makes a careful client nervous and puts you one slip away from breaking something you never needed to touch.
- Editing theme or template code on a live site. One wrong character in a theme file or a `.liquid` template can white-screen the whole site. That is a developer's risk to take, not yours.
- Quoting content work before checking the CMS. A headless or static build can turn a "quick edit" into a developer ticket you cannot fulfill.
- Changing a slug without a redirect, then losing the page's rankings to a 404.
- Forgetting the cache, so you tell the client a change is live when it is not, or think your edit failed when it worked.
- Keeping no change log, so when something breaks a week later nobody, including you, can say what changed.
- Leaving access in place after the engagement ends, so the client is left wondering who still has the keys.

THE HEAD GAME

YOU ARE A GUEST IN THEIR HOUSE

IN PLAIN WORDS

The real job here is being someone a client feels safe handing their website to. They are nervous about letting anyone in, so being careful, clear, and never a mystery is what gets you rehired and recommended, far more than any single clever change.

The psychology of this work is not patience, it is trust. You are handed the keys to something a client has spent years and real money on, and they are nervous about it, whether or not they say so. Everything that makes you easy to trust makes you easy to rehire and to refer. That is the actual career, more than any single ranking.

What to expect

Early on, expect friction on access; it is not distrust of you specifically, it is a client who has been burned or simply does not know how. The VA who sends clear steps and asks for the lowest role is the one who gets access fastest, precisely because they look like they have done this before. Expect, too, that you will be judged less on any one clever change than on whether you are careful, communicative, and never a mystery.

Traps to avoid

1 **The shortcut that erodes trust**

Taking a password or full admin to save ten minutes trades a small convenience for the exact thing that makes you rehireable. Never make that trade.

2 **Touching outside your scope**

"While I was in there I also changed..." is how a small engagement becomes a blame conversation. Stay in the lane you were hired for; suggest the rest, do not just do it.

3 **Going quiet**

Silence reads as either nothing happening or something going wrong. A short, regular update, even "no changes this week, here is what is queued," is worth more than it costs.

4 **Scope creep by osmosis**

Access to everything is not permission to touch everything. More access than the job needs is a liability for both of you.

THE ONE HABIT

Log every change you make: what, where, and when. It is five seconds of typing that turns you from a potential mystery into the person with the clearest record in the room. When something breaks and you can show it was not your last edit, or that it was and here is the fix, you become the professional the client keeps.

THE CLEAN EXIT

OFFBOARDING, AND WHAT ALL THIS BUYS YOU

IN PLAIN WORDS

When the work ends, you neatly hand back every bit of access you were given and send a note confirming it. Because you never took a password and only used your own logins, this is a quick, clean goodbye, and that is exactly what makes a client call you first next time.

When the engagement ends, the whole least-privilege model pays off at once. Because you took your own login everywhere, offboarding is a short, clean list rather than a password-rotation scramble. Walk it deliberately and confirm each item with the client.

- Remove your user from **Search Console, GA4, Tag Manager, Google Ads, and the Business Profile**, or ask the owner to.
- Delete or downgrade your **WordPress** user (reassign your content on deletion), and revoke any application passwords you created.
- Remove the **Shopify** collaborator link and the **Meta** partner link.
- Give up any **registrar or DNS** delegated access.
- Rotate any credential that was ever genuinely shared, and remove any authenticator you registered on a shared account. This is the one case where removal alone is not enough, because you knew the secret.
- Send a short offboarding note listing every access relinquished. It is professional, and it protects you from a later "they still have access" claim.

Do this well and the last thing the client remembers is how easy you were to end with, which is exactly what makes them call you first next time. Careful access, safe edits, a clear record, and a clean exit are not the boring part of the work. For a VA, they are the reputation.

For the tools named throughout, see **The Stack** companion; for the SEO, AEO, and GEO work you do once you are safely inside, see the rest of the Foundations set.

BIBLIOGRAPHY

SOURCES

Every citation from every part, gathered in one place. Grouped by part so you can trace a claim back to where it was made.

Part I — SEO

- Google Search Central: How Google Search Works `developers.google.com/search/docs/fundamentals/how-search-works`
- Google Search Central: AI features and your website `developers.google.com/search/docs/appearance/ai-features`
- Google Search Central: core updates documentation `developers.google.com/search/docs/appearance/core-updates`
- Google Search Quality Evaluator Guidelines, Sept 2025 `guidelines.raterhub.com/searchqualityevaluatorguidelines.pdf`
- web.dev: Core Web Vitals thresholds `web.dev/articles/vitals`
- SparkToro: 2026 zero-click study `sparktoro.com/blog/in-2026-less-than-one-third-of-google-searches-still-send-a-click/`
- Ahrefs: AI Overviews reduce clicks by 58% `ahrefs.com/blog/ai-overviews-reduce-clicks-update/`
- Seer Interactive: AIO impact on Google CTR, 2026 update `seerinteractive.com/insights/aio-impact-on-google-ctr-2026-update`
- Semrush: AI Overviews study `semrush.com/blog/semrush-ai-overviews-study/`
- Ahrefs: 97% of llms.txt files never get read `ahrefs.com/blog/llmstxt-study/`
- Search Engine Land: links as a ranking factor (Illyes) `searchengineland.com/links-google-search-ranking-factor-gary-illyes-432422`
- Kopp Online Marketing: DOJ trial and leak analysis `kopp-online-marketing.com/what-we-can-learn-from-doj-trial-and-api-leak-for-seo`
- Aleyda Solis: query fan-out explained `aleydasolis.com/en/ai-search/google-query-fan-out/`
- Search Engine Journal: Google's guide calls AEO/GEO "still SEO" `searchenginejournal.com/googles-new-ai-search-guide-calls-aeo-and-geo-still-seo/575026/`
- Next.js Learn: rendering and ranking `nextjs.org/learn/seo/rendering-and-ranking`

Part II — AEO

- Google Search Central: AI features and your website `developers.google.com/search/docs/appearance/ai-features`
- Google Search Central: featured snippets `developers.google.com/search/docs/appearance/featured-snippets`
- Google Search Central: optimizing for generative AI features (May 2026) `developers.google.com/search/docs/fundamentals/ai-optimization-guide`
- Pew Research Center: clicks with AI summaries (2025) `pewresearch.org/short-reads/2025/07/22/google-users-are-less-likely-to-click-on-links-when-an-ai-summary-appears-in-the-results/`
- Ahrefs: AI Overviews reduce clicks by 58% `ahrefs.com/blog/ai-overviews-reduce-clicks-update/`
- Ahrefs: 38% of AI Overview citations pull from the top 10 `ahrefs.com/blog/ai-overview-citations-top-10/`
- Ahrefs: 97% of llms.txt files never get read `ahrefs.com/blog/llmstxt-study/`
- SparkToro: 2026 zero-click study `sparktoro.com/blog/in-2026-less-than-one-third-of-google-searches-still-send-a-click/`
- Aggarwal et al.: GEO, ACM KDD 2024 `arxiv.org/pdf/2311.09735`
- Search Engine Journal: Google drops FAQ rich results (2026) `searchenginejournal.com/google-drops-faq-rich-results-from-search/574429/`

- Search Engine Journal: llms.txt has no current implementation `searchenginejournal.com/google-says-llms-txt-is-purely-speculative-for-now/577576/`
- Search Engine Land: query fan-out guide `searchengineland.com/guide/query-fan-out`
- Seer Interactive: AIO impact on CTR, 2026 update `seerinteractive.com/insights/aio-impact-on-google-ctr-2026-update`
- Backlinko: voice search study `backlinko.com/voice-search-seo-study`
- Digiday: WTF are GEO and AEO `digiday.com/media/wtf-are-geo-and-aeo-and-how-they-differ-from-seo/`
- Zapier: AI visibility tools, 2026 landscape `zapier.com/blog/best-ai-visibility-tool/`

Part III — GEO

- Aggarwal et al.: GEO, ACM KDD 2024 `arxiv.org/abs/2311.09735`
- Ahrefs: AI Overview brand visibility factors, 75K brands `ahrefs.com/blog/ai-overview-brand-correlation/`
- Semrush: ChatGPT clickstream study, 17 months `semrush.com/blog/chatgpt-search-insights/`
- Pew Research Center: clicks with AI summaries `pewresearch.org/short-reads/2025/07/22/google-users-are-less-likely-to-click-on-links-when-an-ai-summary-appears-in-the-results/`
- SparkToro: AI brand-recommendation inconsistency study `sparktoro.com/blog/new-research-ais-are-highly-inconsistent-when-recommending-brands-or-products-marketers-should-take-care-when-tracking-ai-visibility/`
- Search Engine Land: inside ChatGPT search, web.run and fan-out `searchengineland.com/inside-chatgpt-search-web-run-fan-out-queries-ai-visibility-477339`
- Google Search Central: AI features and your website `developers.google.com/search/docs/appearance/ai-features`
- Search Engine Journal: llms.txt has no current implementation `searchenginejournal.com/google-says-llms-txt-is-purely-speculative-for-now/577576/`
- Ahrefs: RAG explained for search marketers `ahrefs.com/blog/retrieval-augmented-generation/`
- iPullRank: how AI platforms expand queries with fan-out `ipullrank.com/expanding-queries-with-fanout`
- Digiday: WTF are GEO and AEO `digiday.com/media/wtf-are-geo-and-aeo-and-how-they-differ-from-seo/`
- Zapier: AI visibility tools, 2026 landscape `zapier.com/blog/best-ai-visibility-tool/`

Part IV — The Stack

- Google Tag Manager Help `support.google.com/tagmanager`
- Google Analytics 4 (Universal Analytics sunset, July 2023) `support.google.com/analytics/answer/11583528`
- Google Search Console `search.google.com/search-console/about`
- Bing Webmaster Tools + IndexNow `bing.com/webmasters`
- web.dev: Core Web Vitals and the Chrome UX Report `web.dev/articles/vitals`
- Screaming Frog SEO Spider (free to 500 URLs) `screamingfrog.co.uk/seo-spider/`
- Looker Studio (free reporting) `lookerstudio.google.com`
- Zapier: the best AI-visibility tools, 2026 `zapier.com/blog/best-ai-visibility-tool/`

Part V — The Client Kit

- Google Search Console: managing owners, users, and permissions `support.google.com/webmasters/answer/7687615`
- Google Analytics 4: access and data-restriction management `support.google.com/analytics/answer/9305587`
- Google Tag Manager: managing users and permissions `support.google.com/tagmanager/answer/6107011`
- Google Business Profile: manage owners and managers `support.google.com/business/answer/3403100`
- Google Ads: manager accounts (MCC) and user access `support.google.com/google-ads/answer/7459700`
- WordPress: roles and capabilities `wordpress.org/documentation/article/roles-and-capabilities/`

- Shopify: collaborator accounts `help.shopify.com/en/manual/your-account/users/security/collaborator-accounts`
- Bing Webmaster Tools: how to add users `bing.com/webmasters/help/how-to-add-users-d5d00364`
- Meta: add partners to your Business Manager `facebook.com/business/help/708679622611131`
- Cloudflare: manage account members and roles `developers.cloudflare.com/fundamentals/manage-members/`
- Webflow: legacy Editor deprecation and Content editor role `webflow.com/updates/legacy-editor-deprecation`
- W3Techs: CMS usage and market share, 2026 `w3techs.com/technologies/overview/content_management`
- Wappalyzer: identify the technology a site uses `wappalyzer.com`

NOW GO BUILD.

The guides are the map; the work is repetition. Pick one page or one client, run the loop, keep a change log, and let the compounding do the rest. You have the whole system now, from getting crawled to getting cited to getting hired.