

GENERATIVE ENGINE OPTIMIZATION

— FOR WEBSITE OWNERS —

A Stepwise Reference Manual for
Optimizing AI-Readable Websites and llms.txt Feeds



CLARITY

Make your content easy to read and understand.



STRUCTURE

Organize your site so machines can parse it correctly.



FRESHNESS

Keep your content updated and relevant.



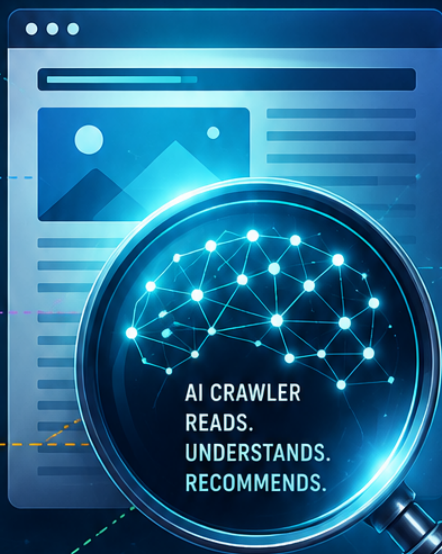
AUTHORITY

Build trust with signals that prove you're credible.



MACHINE-ACCESS

Let the right AI engines access and index your content.



llms.txt

- ✓ Discovery
- ✓ Transparency
- ✓ AI-Readability
- ✓ Citations



BE FOUND. BE READ.
BE RECOMMENDED BY AI.

VIKAS
vSourceCode

Generative Engine Optimization for Website Owners

*A Stepwise Reference Manual for Optimizing
AI-Readable Websites and llms.txt Feeds*

VIKAS · VSOURCECODE

Why This Book Exists

This book started as an audit finding I kept writing on the same line, in different client reports, for different businesses, in different cities. The line was always a variation of the same thing: your site is invisible to the machine that is now sending the best buyers.

The businesses were real. A personal injury firm in Toronto watching its intake calls drop while its Google ranking held steady. A dental clinic in Amsterdam with a PageSpeed score of 38, paying €18 per click for traffic that bounced in under two seconds. A solar installer in Ontario whose competitors were being named in ChatGPT answers and did not know why. A real estate team in London that ranked on page one of Google and appeared in exactly zero AI recommendations for the same queries.

None of these businesses had done anything wrong. They had built their sites in good faith, optimized for a search channel that was the only one that mattered, and kept the lights on. What changed was not their effort — it was the channel. A third machine arrived on the web and started forming opinions about businesses without telling anyone.

That machine — the AI engine behind ChatGPT, Perplexity, Google's AI answers, and Claude — now processes roughly seven billion combined monthly visits. The traffic it refers converts at five times the

rate of standard Google organic. And 88 percent of its citations go to sites that are not in the Google top ten. The businesses that understand this in 2026 are moving. The ones that don't are watching customers go somewhere else and calling it the market getting harder.

This book is the audit, written out. Not the agency version — the version you can do yourself, verify yourself, and hand to whoever manages your site with a specific direction rather than a vague brief. Every fix described here is something I have applied on real sites, with real before-and-after scores, for real businesses in the verticals this book is written for: law firms, dental and medspa clinics, solar installers, real estate teams.

There is no code to write. There is one thing I ask at the end. Everything else is just the work.

— *Vikas, vSourceCode*

Contents

PREFACE

Why This Book Exists	iii
----------------------	-----

PART ONE — THE SHIFT

Chapter 1 The Recommendation You Never Heard	1
Chapter 2 Seven Billion Visits	5
Chapter 3 The Citation Gap	9
Chapter 4 The Algorithm Tax	13

PART TWO — HOW THE MACHINES DECIDE

Chapter 5 Three Machines, One Page	19
Chapter 6 How AI Engines Actually Choose	23
Chapter 7 The Five Signals	27
Chapter 8 Engine by Engine	33

PART THREE — MAKING YOUR SITE READABLE

Chapter 9 Speed Is a Citation Signal	39
Chapter 10 Who You Let In	43
Chapter 11 Structure a Machine Can Parse	47
Chapter 12 Writing for the Quote	51

Chapter 13	llms.txt — The Machine-Readable Layer	55
------------	---------------------------------------	----

Chapter 14	Freshness and the Crawl Loop	59
------------	------------------------------	----

PART FOUR — PROOF, LIMITS, AND THE PLAY

Chapter 15	A Worked Example	63
------------	------------------	----

Chapter 16	The Security Signal	67
------------	---------------------	----

Chapter 17	Reading Your Own Evidence	71
------------	---------------------------	----

Chapter 18	What GEO Cannot Fix	75
------------	---------------------	----

Chapter 19	The 30-Day Plan	79
------------	-----------------	----

Chapter 20	The Free Audit	83
------------	----------------	----

BACK MATTER

Glossary	85
----------	----

Appendix A — AI Crawler Reference	89
-----------------------------------	----

Appendix B — Citation Self-Test Protocol	91
--	----

Appendix C — Sources and Citations	93
------------------------------------	----

PART ONE

The Shift

Why the web changed, how fast it changed, and what your business is losing every day it has not caught up.

The Recommendation You Never Heard

It is nine in the evening. Someone in your city has a problem you solve for a living, and they are not going to Google.

They are tired. They do not want to scroll past ads and click ten different sites to work out who is actually any good. So they open ChatGPT, or Perplexity, or Google's AI answer box, and they type a normal question: *who is the best person for this in my area, and should I call them tonight?*

A few seconds later, an answer appears. Not a list of links. A recommendation. Three names, each with a sentence about why. The person reads it, picks one, and dials. The entire journey — from problem to phone call — took less than two minutes.

Here is what this means for your business. If your name was not in that answer, nothing notified you. No impression in a report. No bounce to analyze. No "you came fourth." The machine read the field, formed a view, and moved on. You were not turned down. You were not considered.

*There is no rejection email from an AI engine.
There is only a customer who called someone
else, and you will never know who it was.*

This is the fundamental shift this book is about. Search, as you have known it for twenty years, was a visible competition. You had a rank, a position number, a click-through rate. AI recommendation is different — invisible, immediate, and based on a judgment the machine formed long before the question was asked.

That judgment was formed over time, as AI crawlers visited your site and built a picture of it. Is this page clear? Is it current? Does it load fast enough to read? Does the site behave like a legitimate business? Can I quote this without looking foolish in front of my user? Those questions were answered — or not — on dozens of quiet visits you never saw.

Most business owners respond to softer traffic by doing more of what used to work: more posts, more keywords, a fresh design. None of it addresses the actual problem. The problem is not a ranking problem. It is a readability problem. The machine cannot find you, read you, trust you, or quote you — and until that changes, the recommendation keeps going to someone else.

Get it right and the nine o'clock question ends differently. The machine visits your site, finds it clear, current, defended, and structured. When someone asks who is the best option tonight, it has a reason to say your name.

Seven Billion Visits

AI engines now process roughly seven billion combined monthly visits. Most businesses appear in none of those answers. The window to change that is open right now.

7B

Combined AI engine
visits per month

~5×

Conversion
advantage, AI-
referred vs Google
organic^[1]

88%

AI citations from
outside Google's top
10

AI-referred traffic converts at approximately five times the rate of standard Google organic traffic.^[1] Visitors from AI engines stay approximately 41 percent longer on site than visitors from all other sources. Their bounce rate is 23 percent lower. AI-driven referral traffic grew approximately twelve times between July 2024 and February 2025 in the US.^[2]

These are measured outcomes from traffic already flowing through AI engines today. The channel is not coming. It is here, it converts better than anything else, and 88 percent of AI citations come from URLs outside Google's top ten.

WHAT THAT LAST NUMBER MEANS FOR YOU

A technically excellent site at position 14 in Google results can be recommended ahead of a position-two site that loads slowly, blocks AI crawlers, or uses structure a machine cannot parse. The criteria are different. And right now, most of your competitors have not adjusted.

THE THIRD BUYING CHANNEL

For twenty years, customer acquisition online ran on two channels: organic search and paid advertising. Both have become more expensive and unpredictable. AI recommendation is the third channel. It requires no ad spend. It is a readability race — and a business that passes the readability test gets recommended. One that doesn't, doesn't.

The difference between the first two channels and this one is permanence. Ad spend stops and traffic stops with it. Rankings shift with every algorithm update. But a site that is genuinely machine-readable — fast, structured, authoritative, accessible — stays machine-readable. The investment compounds.

[1] Opollo, "AI Search Traffic Converts 4–5x Better Than Google," 2025. See Appendix C. [2] Adobe Digital Insights, February 2025.

The Citation Gap

Being indexed is not the same as being cited. The gap between the two is where most businesses are quietly losing customers they never knew they had.

The Citation Gap is the measurable difference between how often AI engines name your business when a customer asks for a recommendation, versus how often they name a competitor. Most businesses have never seen this scoreboard. Most score zero.

THE SELF-TEST

Open ChatGPT, Perplexity, and Google's AI answer box. Type the question your best customer would type when they need what you sell — the category question, not your business name. *"Which personal injury lawyer in Toronto should I call after an accident?" or "Best dental implant clinic in Amsterdam?" or "Solar installer near me in Ontario with good reviews?"*

Read the answers. Count how many times your name appears across three engines and ten questions in your category. That number is your citation score. For most businesses reading this book, it is zero.

The Citation Gap is not about being unknown. It is about being unreadable. The machine knows you exist. It has decided it cannot safely quote you.

FIVE REASONS A SITE FALLS INTO THE GAP

1. **Crawl access blocked or broken.** The robots.txt file refuses the very crawlers that would cite the site.
2. **Page speed too slow to finish reading.** A machine with a timeout abandons a slow page before the content loads.
3. **Structural clarity missing.** Content exists but is organized so a machine cannot extract a specific answer from it.
4. **Schema markup absent.** Without explicit labels, the machine guesses what type of business this is — and often guesses wrong.
5. **No geographic or topical specificity.** A generic "we serve the whole area" page is not citable for a specific city-level query.

These five points map directly onto the five signals in Chapter Seven. The next chapter covers the cost of leaving them unfixed — accumulated invisibly, over time.

The Algorithm Tax

There is a bill nobody sends you. It does not arrive in your inbox, it does not appear on your accounts, and you can run a business for years without seeing the line item.

This is the Algorithm Tax — the budget an automated system drains when it matches on a signal that looks right but isn't. Wrong objective. Wrong keyword intent. Wrong audience. Wrong placement. The machine does not fail. It completes. You told it to find clicks, and it found clicks, in the cheapest population where that signal could be satisfied. The waste belongs to whoever pointed it.

"The machine does not waste. The machine completes. It receives an instruction, it follows the instruction, it reports completion. The waste belongs to whoever pointed it."

Taro Mizushima · Keiro, Issue 8 · 2003

THE FOUR AXES OF MISMATCH

AXIS	WHAT GOES WRONG	REAL EXAMPLE
01 Objective	Optimises for clicks or reactions rather than buyers who can pay.	<i>\$999 on Meta Advantage+. 159 reactions. Zero landing-page views.</i>
02 Intent	Buys searches that share your words but not your purpose.	<i>Watch repair shop paying for "watch battery replacement." Towing company paying for "tow truck driver jobs."</i>
03 Audience	Reaches the cheapest segment instead of the buying one.	<i>High-value CPA campaign in Calgary with no income tier set, reaching students.</i>
04 Placement	Spends wherever impressions are cheapest — feeds your buyer never converts from.	<i>Emergency trade ad as an accidental tap inside Instagram Reels at 11pm.</i>

THE AI CITATION EQUIVALENT

In AI recommendation, the Algorithm Tax manifests as structural invisibility. An AI engine retrieving pages to answer "best personal injury lawyer in Vancouver" reads twenty pages. One loads in 4.2 seconds on mobile — the engine times out. Another has no headings,

no schema, no geographic specificity — no citable claim can be extracted. A third accidentally blocked the AI crawler in its robots.txt, written three years ago by a developer who had never heard of GPTBot.

None of these failures generate a report. The machine moves on. You pay the tax every day the site stays the way it is — and the bill is never totaled until someone runs the audit.

THE \$999 CASE STUDY

A verified incident: a small business ran a Meta Advantage+ campaign with a \$999 budget and no audience constraints. The result: 159 reactions, zero landing-page views, zero conversions. The machine completed. It found engagement in the cheapest population where engagement could be harvested. Nobody had told it that a reaction is not a customer.

PART TWO

How the Machines Decide

The four-step pipeline behind every AI answer, the engines running it, and the five signals that determine whether your page wins a citation or loses one.

Three Machines, One Page

Every page you publish is now read by three different machines, each applying different criteria and reaching a separate verdict. You have been optimizing for one of them.

The first machine is Google's crawler — the one you have been thinking about since you first heard the word SEO. It indexes pages, assesses relevance, and determines search rankings. You have spent years learning its rules.

The second machine is the AI engine. It does not rank pages. It reads them, extracts claims, assesses trustworthiness, and decides whether it can safely cite the page in an answer it is about to give a real person. A wrong citation embarrasses it in front of its user, so it is conservative: it would rather cite no one than cite someone it is not sure about.

The third machine is the human visitor — on a phone, in a hurry, deciding in two seconds whether this page is worth their time.

THREE VERDICTS FROM ONE PAGE

Google's crawler: Is this relevant? Does it have authority? Does it load correctly?

The AI engine: Can I read this completely? Can I extract a trustworthy, quotable claim? Would citing this embarrass me?

The human visitor: Does this answer my question? Does this business look legitimate? Should I call?

The good news: a page that is fast, clearly structured, genuinely authoritative, and easy to read satisfies all three. GEO is not a new discipline layered on top of everything else. It is a cleaner, more complete version of the same underlying job.

The critical difference is priority. Google rewards indirect signals — domain age, backlinks, content volume. The AI engine rewards direct signals applied to the page on the visit happening right now. A six-month-old site with fifty pages, a 100 PageSpeed score, clean structure, and an AI-accessible robots.txt outperforms a ten-year-old site with five thousand posts and a 38 mobile score. The AI engine does not care about your history. It cares about what it can read today.

How AI Engines Actually Choose

Behind every AI answer is a four-step pipeline: Retrieve. Read. Trust. Cite. Your page wins or loses at one of those four steps — and most pages lose at the first one.

STEP 1 — RETRIEVE

The engine crawls your site or fetches it on-demand. If robots.txt blocks the crawler, or if the page does not exist in a form the engine can access, retrieval fails and the page is never considered. Most businesses fail here without knowing it.

STEP 2 — READ

The engine reads the page. If it loads too slowly, the engine times out. If content is hidden behind JavaScript the crawler doesn't execute, the engine sees a blank page. If the content is a wall of undifferentiated text with no headings, the engine cannot extract a specific answer. Most businesses lose citations here — not because the content is bad, but because the machine cannot get to it fast enough or find what it needs within it.

STEP 3 — TRUST

The engine assesses whether the page is safe to cite. It looks for schema markup confirming business type and location, a secure connection, consistency between what the page claims and what other sources say, evidence the content is current. A page that fails the trust assessment is indexed but not cited.

STEP 4 — CITE

The engine names the business in its answer. Everything in this book is about engineering the conditions under which this step happens.

Most businesses are eliminated at step one or two. The citation problem is almost never a content problem. It is an access and structure problem.

CITATION SURFACE VS. TRANSACTION SURFACE

Citation surfaces (ChatGPT Search, Perplexity, Google AI Overviews) fetch sources and produce a cited answer. The citation is a referral. Optimize for parseability and trust.

Transaction surfaces (Operator by OpenAI, Computer Use by Anthropic, Project Mariner by Google) visit the page on the user's behalf and complete a task — booking, form-filling, purchasing. Optimize for render stability and CLS = 0. Layout shifts and late-loading modals break agentic browsers even on fast, well-structured pages.

The Five Signals

Five things, more than anything else, decide whether your page is cited by an AI engine. Get these right and most of the rest follows.

SIGNAL 01

Crawl Access

The AI crawler must be able to reach and read the page. A site that blocks AI crawlers in its robots.txt, serves a page that times out, or hides content behind JavaScript is invisible to AI recommendation regardless of content quality. Access is the precondition everything else sits on top of.

SIGNAL 02

Page Speed

A machine reading your page has a time budget. If the page does not deliver its content within that budget, the machine abandons it. A 100 PageSpeed score on mobile is not vanity — it is proof that the page delivers content at machine speed. The competitive minimum is a mobile Performance score above 90 with First Contentful Paint under 1.5 seconds.

SIGNAL 03

Structural Clarity

Clear headings, question-and-answer pairs, short paragraphs with one claim each, a logical hierarchy the parser can follow. A wall of text, however accurate, is not citable. A machine reading it cannot locate the specific claim it needs to answer a specific question.

SIGNAL 04

Schema Markup

Schema tells the engine explicitly: this is a LocalBusiness of type Dentist, located at this address, with these services. Without schema, the engine infers — and inference introduces errors. You do not write schema yourself; you describe what you want to whoever manages your site. An experienced developer can implement it in two hours.

SIGNAL 05

Topical and Geographic Specificity

An AI engine answering "best medspa in Rotterdam" needs a page that explicitly says: we are a medical aesthetics clinic, located in Rotterdam, and here is why we are the right choice for this patient. A generic "we serve the whole region" page is not citable for a specific query. Specificity is what connects a page to the question being asked.

NONE OF THESE IS A TRICK

There is no tag to insert and no hack to run. The work is durable because it makes the site genuinely better for human visitors too. The AI engine is rewarding sites that deserve to be quoted.

Engine by Engine

ChatGPT, Perplexity, Google's AI Mode, and Claude do not read pages the same way. The differences that matter for your site in 2026.

ChatGPT + SearchGPT

Cites

Transacts

OpenAI · Crawlers: GPTBot (training), ChatGPT-User (on-demand), OAI-SearchBot (index)

Weights heavily: Recency, source credibility, structural clarity, page-level authority. Inline numbered citations with source cards. Operator (agentic) requires CLS = 0 for reliable task completion.

Claude

Cites

Transacts

Anthropic · Crawlers: ClaudeBot (training), Claude-Web / anthropic-ai (on-demand)

Weights heavily: Source quality, structural clarity, content depth, exact-quote eligibility. Quote-faithful attribution — pages with clearly attributed, low-distortion text are cited more reliably.

Gemini + AI Mode

Cites

Transacts

Google · Crawlers: Google-Extended (opt-out, separate from Googlebot)

Weights heavily: Classical Google ranking signals plus structural clarity. Citation often follows but does not replicate top organic results.

Perplexity

Cites

Perplexity AI · Crawlers: PerplexityBot (training), Perplexity-User (on-demand)

Weighs heavily: Source diversity, recency, structural clarity, citable claims. The most explicit citation UI of any engine — numbered footnotes with source cards. Notably cites smaller sites with specific authoritative depth over large generic ones.

BOT NAMES — EXACT STRINGS FOR ROBOTS.TXT

BOT NAME (EXACT, CASE-SENSITIVE)	ENGINE	DEFAULT
GPTBot	OpenAI · ChatGPT	Allow
ClaudeBot	Anthropic · Claude	Allow
PerplexityBot	Perplexity	Allow
Google-CloudVertexBot	Google AI	Allow
Applebot	Apple Intelligence	Allow
FacebookBot	Meta	Block
Meta-ExternalAgent	Meta	Block
Meta-ExternalFetcher	Meta	Block
Manus Bot		Block

BOT NAME (EXACT, CASE-SENSITIVE)	ENGINE	DEFAULT
	Manus (agentic scraper)	

PART THREE

Making Your Site Readable

The build — signal by signal, no code required. Each chapter maps to one of the five signals and gives you something specific to do this week.

Speed Is a Citation Signal

If a machine times out before your page finishes loading, it never reads your content. Speed is the precondition everything else depends on.

A one-second improvement in mobile load time increases conversions by approximately 20 percent and reduces bounce rate by approximately 32 percent.^[3] The average legal-vertical landing page loads in approximately 4.8 seconds on mobile. An edge-deployed equivalent serves in approximately 0.9 seconds — the difference between a client who waits and a client who calls a competitor.

For AI citation, the machine operates under a timeout constraint. Pages that deliver content within that constraint are read. Pages that don't are abandoned. The practical result is identical to the human case: fast pages get read, slow pages don't, and a page never fully read is a page never cited.

HOW TO CHECK YOURS IN TWO MINUTES

Open Google PageSpeed Insights (pagespeed.web.dev). Enter your URL. Run the mobile test. A Performance score above 90 with First Contentful Paint under 1.5 seconds is the competitive floor. The benchmark this book demonstrates is 100 on all four metrics.

Most business sites — law firms, dental clinics, real estate agencies — built on WordPress with accumulated plugins score between 30 and 65 on mobile. A score of 35 is delivering content too slowly for machine reading. Chapter 15 shows what 100 looks like on a live site and how it was achieved.

THE WORDPRESS PLUGIN PROBLEM

The most common cause of slow business sites is the Plugin Graveyard — accumulated WordPress plugins, each injecting JavaScript and CSS into every page load. Analytics plugins, form plugins, slider plugins, optimisation plugins (which often make pages slower), security plugins, builder plugins, chat widgets. A site that loaded quickly in 2020 can be taking 5+ seconds in 2026, not because anything broke, but because sixteen plugins are each adding 0.3 seconds to every visit.

The fix is not to remove all plugins — it is to audit which are render-blocking, which duplicate functionality, and which contribute nothing measurable. The result of doing this correctly is shown in the next chapter.

[3] Google and Deloitte, "Milliseconds Make Millions," 2019. See Appendix C.

Who You Let In

Somewhere in the plumbing of your website is a short list deciding which machines may read you and which are turned away. Most owners have never seen it. Some have accidentally blocked the engines that would cite them.

The file is called robots.txt and it sits at the root of every website. Configured once and forgotten — often by a developer who had never heard of GPTBot — it may be blocking every AI citation opportunity your business has.

In 2026 there are more than a dozen AI crawlers that matter for business citation, each with its own user-agent string, each obeying robots.txt. Welcome the readers who send buyers your way. Turn away the scrapers that take content and give nothing back.

Crawler access · vsourcecode.com				Live configuration
☐	FacebookBot Meta	AI Crawler	0 B	Allowed: 0 Unsuccessful: 0 ☒
☐	Google-CloudVertexBot Google	AI Crawler	0 B	Allowed: 0 Unsuccessful: 0 ☐
☐	Manus Bot Manus	AI Assistant	0 B	Allowed: 0 Unsuccessful: 0 ☒
☐	Meta-ExternalAgent Meta	AI Crawler	0 B	Allowed: 0 Unsuccessful: 0 ☒
☐	Meta-ExternalFetcher Meta	AI Assistant	0 B	Allowed: 0 Unsuccessful: 0 ☒

The real configuration. A blue toggle indicates a blocked crawler. FacebookBot, Manus Bot, Meta-ExternalAgent, and Meta-ExternalFetcher are blocked — scrapers that take without citing. Google-CloudVertexBot (grey = allowed) and the primary citation engines are open. The principle: open the door for the readers who recommend you, close it for the ones who only extract.

HOW TO READ YOUR OWN ROBOTS.TXT

Type your domain followed by `/robots.txt` in any browser. If blank or 404, every crawler has full access by default — including scrapers. If content appears, look for any line beginning *User-agent: ** followed by *Disallow: /* — that combination blocks every crawler from your entire site, including every AI engine that would cite you.

Look for specific blocks: a line reading *User-agent: GPTBot / Disallow: /* blocks ChatGPT's crawler from your entire site — a complete citation block for that engine, set silently by a plugin or theme you probably installed for an unrelated reason.

WHAT TO CHANGE

Give this direction to whoever manages your hosting: allow GPTBot, ClaudeBot, PerplexityBot, Google-CloudVertexBot, and Applebot explicitly. Block FacebookBot, Meta-ExternalAgent, Meta-ExternalFetcher, and Manus Bot. Verify by fetching your robots.txt again and checking each line.

Structure a Machine Can Parse

Content that is accurate and well-written is not automatically citable. A machine needs to find the specific answer to a specific question — and if the page's structure does not make that possible, the content is invisible.

THE STRUCTURE OF A CITABLE PAGE

- **One clear H1** that states exactly what the page is about. "Personal Injury Lawyers in Toronto — Free Consultation" is citable. "When Things Go Wrong, We Make Them Right" is not.
- **H2s written as real questions.** "How quickly can I speak to a lawyer after an accident?" "What does a free consultation cover?" "Which areas of Ontario do you serve?" These are the questions AI engines receive.
- **Short paragraphs, one claim each.** A paragraph making three claims is a paragraph the machine cannot cite for any specific one.
- **Answer before elaboration. Answer before elaboration.** State the answer in the first sentence. "We offer free consultations, available same-day for urgent cases across Greater Toronto." Then elaborate. A machine reading only the first sentence needs the answer first.

- **Specific claims.** "Fast response" is not citable. "Same-day appointments available for dental emergencies across Amsterdam, Monday through Saturday" is.

SCHEMA MARKUP — THE LABEL LAYER

Schema tells the engine explicitly what type of business this is, where it is located, what its services are, what its hours are. Without it, the engine infers. With it, the engine reads directly.

The most important schema types for local business citation: LocalBusiness (with specific business type — Dentist, LegalService, RealEstateAgent, SolarEnergyCompany), FAQPage (for question-and-answer sections), and Review (for third-party credibility signals). You do not write this yourself — you describe what you want, and an experienced developer implements it in approximately two hours.

Writing for the Quote

An AI engine constructing an answer looks for a sentence it can lift cleanly and attribute to a source. There is a shape that sentence takes. Most content does not have it.

WHAT A QUOTABLE SENTENCE LOOKS LIKE

A sentence an AI engine can use makes one specific, attributable claim; is true without context; contains enough specificity to be useful as a standalone answer; and does not hedge so much that citing it says nothing.

"Our team delivers faster results." — Not citable. No specific claim, no measurement.

"The average legal landing page in the UK loads in 4.8 seconds on mobile; an edge-deployed equivalent serves in 0.9 seconds — a difference that directly reduces bounce rate and effective cost-per-click." — Citable. Specific claim, specific metric, specific causal relationship.

THE VOICE THAT GETS CITED

Direct, specific, willing to take a position. Not "there are many factors to consider" — instead, "the single most expensive mistake in this type of audit is X, and here is why it costs more than the client

realizes." The Quora answer that gets shared is not the most comprehensive. It is the one that says something specific and gives the reader a frame for their own situation. That same quality is what makes a page citable by an AI engine.

WHAT TO AVOID

- **Hedging that empties the claim.** "Results may vary" is true and useless. Say what typically happens; note the exception below it.
- **Generic expertise claims.** "Years of experience and a passionate team" cannot answer any specific customer question. "We have represented over 400 personal injury claimants across Ontario, with a 94% success rate on cases that proceeded to settlement" can.
- **The AI pattern voice.** Perfectly structured, relentlessly balanced, never taking a position — ironically the hardest content for AI engines to cite, because there is no specific claim to lift.

Write like someone who has done the work and has a grounded opinion about what they found. That is the voice that gets cited — and the voice that gets shared on Quora, LinkedIn, and everywhere else authority is built.

llms.txt — The Machine-Readable Layer

A small plain-text file that hands an AI engine a clean map of your site's best material. What it is, what it can and cannot do, and how to use it without overclaiming.

Picture handing a researcher one sheet of paper before they walk into your archive: here is what we do, here are the ten files that matter, here is where each one lives. You have saved them an afternoon and ensured they cite the right things. That sheet of paper is what an llms.txt file is — except the researcher is an AI engine.

The file sits at `yoursite.com/llms.txt`, alongside `robots.txt`. In plain text, it states what your site is about and points to the pages most worth reading. No code to write — just structured plain text.

WHAT GOES IN IT

- A one-paragraph description of the business in the buyer's language — not your marketing copy, the language your customer uses to describe their problem.
- A list of your most important pages with a one-line description of what each one covers.
- The canonical claims you most want cited — specific, sourced statements.

- A last-updated date, so engines know whether the content is current.

THE HONEST LIMITATION

No major AI engine has publicly committed to obeying llms.txt. Respected engineers at Google have pointed out it is not a switch that turns on traffic. They are right.

It is not a magic switch. It is good filing. Good filing has never hurt a site that also did the real work.

Hold it as one low-cost signal among the five that matter — not the thing that saves you, but a thing that costs almost nothing to do well and that some engines already read. The [vsourcecode.com llms-full.txt](https://vsourcecode.com/llms-full.txt) is a live example: a complete knowledge base in one machine-readable document, available at vsourcecode.com/llms-full.txt.

Freshness and the Crawl Loop

A page that updates gets re-read. A page that sits untouched for two years goes stale in a model's memory. Freshness is about signalling, honestly, that what you published still holds.

AI engines carry a memory of the web with a date. Content indexed six months ago may be accurate, but without a freshness signal, the engine treats it as potentially outdated. For business information — prices, availability, service areas — outdated is a liability.

WHAT FRESHNESS SIGNALS LOOK LIKE

- **A visible last-updated date** on service, pricing, and FAQ pages. A date showing the content was reviewed and confirmed current is as useful as a date showing it was rewritten.
- **Accurate sitemap lastmod dates.** Not every page set to today's date by a WordPress plugin, but the actual date each page was last modified. Inaccurate lastmod dates train crawlers to ignore the signal.
- **New content that extends coverage** without replacing what is already there. One substantive new page per month on a specific local topic — a real case study for the law firm, a specific treatment outcome for the medspa, a solar installation in a specific neighbourhood — is more valuable than fifteen thin posts.

- **An llms.txt with an accurate last-updated header.** The most direct freshness signal to an AI engine reading the file.

WHAT FRESHNESS IS NOT

Freshness is not volume. Fifteen thin posts a month dilute authority concentration. One substantive piece making a specific, sourceable claim is more valuable for citation eligibility than fifteen posts that say nothing new. The crawl loop rewards substance. A page updated because information changed carries a genuine freshness signal. A page "updated" because a paragraph was reformatted carries noise.

PART FOUR

Proof, Limits, and the Play

A worked example on a live site, the evidence panels, what this method will not do, and the one thing we ask of you at the end.

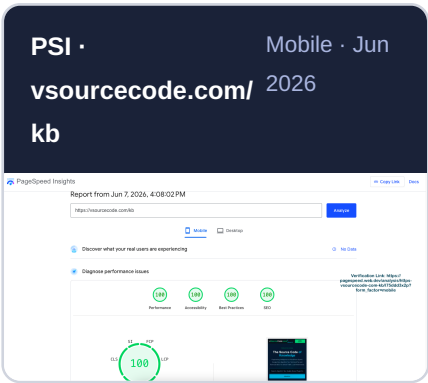
A Worked Example

One real site, audited in the open. A perfect PageSpeed score is not vanity — it is proof that the infrastructure described in Part Three actually lands.

Everything in this book is demonstrated on vsourcecode.com, which runs entirely on Cloudflare Workers with no traditional CMS. No plugins. No WordPress. No database. Every page is served at the edge, in every geography, from infrastructure that is owned and controlled — not rented from a platform with its own interests.

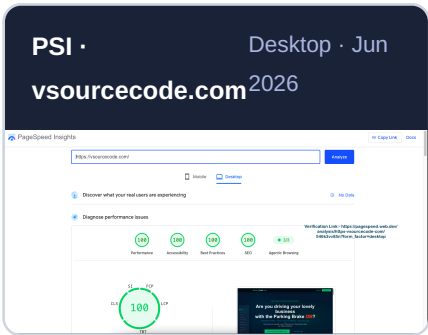
The result is a site that scores 100/100/100/100 on Google PageSpeed Insights on both mobile and desktop. Performance, Accessibility, Best Practices, and SEO — all four green. Both proof-of-work screenshots are below.

MOBILE —
VSOURCECODE.COM/KB



Mobile, knowledge base page.
100/100/100/100. Verification:
pagespeed.web.dev/analysis/https-vsourcecode-com-kb/l75ddd3x2p?form_factor=mobile

DESKTOP —
VSOURCECODE.COM



Desktop, homepage. 100/100/100/100 +
Agentic Browsing 3/3. Verification:
pagespeed.web.dev/analysis/https-vsourcecode-com-546k3vv85n?form_factor=desktop

THE FIFTH COLUMN — AGENTIC BROWSING

Google added the Agentic Browsing audit to PageSpeed Insights in June 2026. It checks three things traditional performance metrics do not cover:

- **Accessibility tree completeness** — whether the page exposes a structured tree that agentic browsers can navigate programmatically.
- **CLS = 0** — whether elements stay in place after load. An agent that clicks where a button was when it read the page fails if the button has shifted.
- **llms.txt presence** — whether the site has published a machine-readable content map.

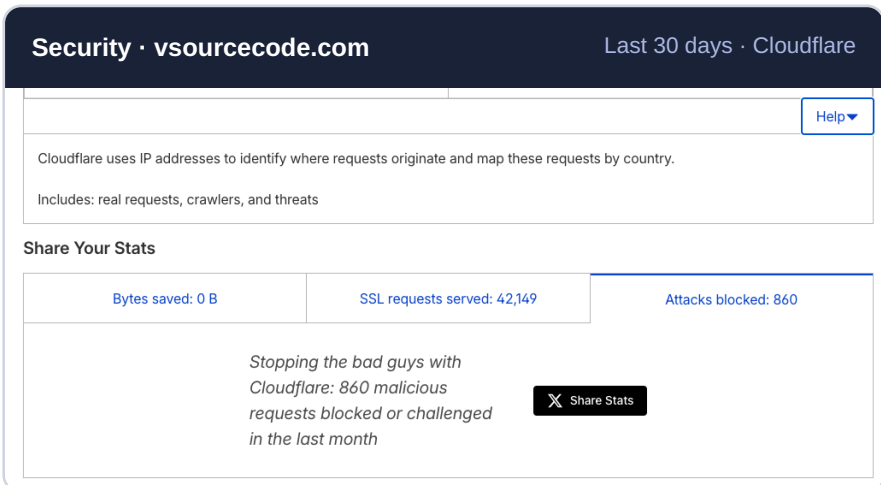
Most sites built on page builders and WordPress themes fail the accessibility tree check and the CLS check without knowing it. The fix involves the infrastructure decisions in Part Three — not content changes, but structural ones.

The Security Signal

A site that quietly defends itself reads as trustworthy to the machine deciding whether to quote it. A clean, defended surface is itself a credibility signal.

Trust is not a feeling an AI engine has. It is a set of things it can check. One of the quietest checks is whether the site looks defended — whether it serves over a secure connection, whether hostile requests are turned away at the edge rather than reaching the application layer.

SECURITY OVERVIEW — VSOURCECODE.COM (LAST 30 DAYS)



One month: **42,149 requests served over TLS, 860 malicious requests blocked or challenged.** A defended, all-TLS surface is the kind of low-risk page an engine can cite without hesitation.

This security posture is a byproduct of the infrastructure decisions from Part Three — edge deployment, no plugin graveyard, a site that does not have twelve WordPress plugins each carrying their own vulnerabilities. For a site on WordPress with standard hosting, the equivalent requires: an active SSL certificate with no mixed-content warnings, a WAF at the hosting or CDN level, and bot management that turns away known malicious crawlers. These are standard options in modern hosting control panels, frequently left at their defaults.

Reading Your Own Evidence

Your server logs are the most honest report you will ever get. AI crawlers are already visiting your site. Here is how to find them, and how to tell a real signal from noise.

You do not need a third-party tool to know whether AI engines are reading your site. The evidence is in your server logs — the raw record of every HTTP request your site has received. Every GPTBot visit, every ClaudeBot crawl, every PerplexityBot fetch is logged by default on almost every hosting platform.

WHAT TO LOOK FOR

In your server logs or Cloudflare analytics dashboard, filter for these user-agent strings: GPTBot, ClaudeBot, PerplexityBot, Google-Cloud-VertexBot, Applebot. What you want to see: regular visits (weekly or more) from multiple AI crawlers, accessing your most important pages, returning 200 status codes.

What you do not want to see: 403 responses to AI crawlers (your robots.txt or WAF is blocking engines you should welcome), crawls limited to your homepage (internal linking is not guiding engines to citable content), or no AI crawler visits at all (access is blocked, or the site times out before completing).

THE SERVER-SIDE ADVANTAGE

Standard analytics tools miss a significant share of actual traffic. Client-side JavaScript tags are blocked by ad blockers, privacy extensions, and enhanced tracking browsers. Crawlers by definition do not execute JavaScript tags. The only reliable record of who has visited your site is the server log — written before the page is served, not after JavaScript fires.

If your analytics setup is entirely client-side, your view of AI crawler activity is zero by definition. Server-side analytics gives the complete picture, including every crawler visit that client-side tools miss.

What GEO Cannot Fix

Every other guide on this subject is selling optimism. This chapter does the opposite — because a method that admits its limits is one you can trust with the things it can do.

GEO will not turn a business nobody wants into one everybody recommends. The AI engine rewards substance, resists puffery, and declines to cite pages with nothing quotable. Machine-readability makes a thin page readable — not valuable. You cannot cite what is not there.

GEO will not deliver results next Tuesday. The crawl loops that carry technical improvements into a model's active memory run on their own schedule — weeks for some engines, months for others. The work is permanent and compounding; the timeline is not short. Expect to see citation changes in AI answers six to eight weeks after crawlers index the improved pages.

GEO will not override a bad reputation. If the AI engines have indexed negative reviews, forum discussions, or third-party articles describing your business poorly, making your own site more machine-readable will not override those signals. The citation pool includes everything the engine has read.

GEO will not compensate for a broken trust layer. A hacked site, malware, a compromised SSL certificate, or a manual penalty for deceptive content is not recoverable through machine-readability until the underlying problem is resolved first.

Fix the offer first. The infrastructure amplifies what is already there. It does not create what isn't.

One final limit: this book is not the final word. The five signals are the right framework for mid-2026 and will remain directionally correct for the next two to three years. But signal weights will shift, new engines will appear, and the agentic browser track will grow in importance. This book is the foundation, not the ceiling.

The 30-Day Plan

A sequenced, non-technical plan. What to do each week, what can wait, and how to know each step worked before moving to the next.

The five signals have a dependency order. Crawl access first — nothing else matters if the machine cannot get in. Speed second — a machine that cannot read your page fast enough never assesses its structure or trust. Structure and schema third — the quality layer that turns a readable page into a citable one. Freshness and the machine-readable layer last — they extend and maintain citation eligibility.

WEEK 1 — ACCESS

- ☐ Fetch your robots.txt and read it. Look for any block on AI crawlers by name, or any blanket *Disallow: /* blocking all crawlers.
 - ☐ Confirm GPTBot, ClaudeBot, PerplexityBot, Google-CloudVertexBot, and Applebot are allowed.
 - ☐ Confirm FacebookBot, Meta-ExternalAgent, Meta-ExternalFetcher, and Manus Bot are blocked.
 - ☐ Run PageSpeed Insights on mobile. Record the Performance score as your Week 1 baseline.
-

- ☐ Check server logs or Cloudflare analytics for AI crawler visits in the last 30 days.
-

WEEK 2 — SPEED

- ☐ Identify your three most important pages. Run PSI on each. Note issues under "Opportunities."
-
- ☐ If on WordPress: audit active plugins. Remove or deactivate any that duplicates functionality or contributes nothing measurable.
-
- ☐ Confirm all pages serve over HTTPS with no mixed-content warnings.
-
- ☐ Target: mobile Performance score above 90 on your three most important pages by end of week.
-

WEEK 3 — STRUCTURE AND SCHEMA

- ☐ Review your homepage and top three service pages. Does each have a clear H1 stating what the business does and where?
-
- ☐ Rewrite any generic H2 as a specific customer question relevant to your city and vertical.
-
- ☐ Brief whoever manages your site to add LocalBusiness schema with your business type, address, phone, hours, and service area. Request FAQPage schema on any page with Q&A content.
-
- ☐ Add a visible last-updated date to your key service and FAQ pages.
-

WEEK 4 — MACHINE-READABLE LAYER

- ☐ Create your llms.txt. One paragraph describing your business in buyer language. List your five most important pages with one-line descriptions. Add a last-updated date.

- ☐ Place the file at yoursite.com/llms.txt. Confirm it returns a 200 status code.

- ☐ Re-run PSI on key pages and compare to your Week 1 baseline.

- ☐ Run the citation self-test: ask ten category questions across ChatGPT, Perplexity, and Google AI. Note your citation score versus baseline.

- ☐ Set a calendar reminder for 60 days to repeat the test. Citations typically appear 6–8 weeks after crawlers index the improved pages.

The Free Audit

The book ends where the work begins. There is no course to buy and no second product waiting. There is one door, and it is free to walk through.

A GEO audit checks your site the way the three machines do. It runs the PageSpeed test. It reads your robots.txt against the AI crawler list. It checks your schema, structural clarity, freshness signals, and security posture — the technical layer that determines whether an engine can read you, trust you, and quote you.

You get back a plain-language report of what is working, what is costing you citations, and the specific fixes that would move the needle most for your site. No code required to read the report. No agency engagement required to act on most of it.

Claim your free GEO audit

See your site the way ChatGPT, Perplexity, Google AI, and Claude see it. Find out exactly where you are being read, where you are being skipped, and what to fix first.

What it checks: crawl access, page speed, structure, schema, freshness signals, and security posture.

What you get: a plain-language technical report. The highest-value fixes for your specific site. A before-and-after PSI score.

What it costs: nothing.

vsourcecode.com

One audit. One report. The only ask in this book.

Concept Definitions

The terms this book uses, in plain language. Canonical definitions live at vsourcecode.com/concepts.

GEO (Generative Engine Optimization)

The discipline of engineering web pages that AI engines preferentially read, parse, and cite. Distinct from SEO, which optimizes for search rankings. Also called AEO (Answer Engine Optimization) and CEO (Citation Engine Optimization).

Citation Gap

The measurable difference between how often AI engines name your business in answers versus how often they name a competitor. Most businesses have never seen this scoreboard. Most score zero.

Algorithm Tax

The budget an automated advertising or recommendation system drains when it matches on a signal that looks right but isn't. The machine does not fail. It completes. The waste belongs to whoever pointed it.

The Five Signals

Crawl access, page speed, structural clarity, schema markup, and topical/geographic specificity — the factors that determine whether a page is citation-eligible.

Plugin Graveyard

The accumulated WordPress plugins added over years, each injecting JavaScript and CSS into every page load. The dashboard looks normal; Core Web Vitals quietly fail.

Sovereign Node

A self-contained page deployed on the global edge (Cloudflare Workers), isolated from the main website — its own conversion path, analytics, and performance budget, immune to the host site's plugin graveyard or downtime.

Technical Tax

The cost-per-click multiplier Google's Quality Score applies to slow landing pages. A Quality Score of 3 versus 9 can roughly triple effective CPC for the same keyword, bid, and audience.

Ghost Traffic

Paid clicks where the user pressed back before the page rendered — billed at full CPC, recorded nowhere in analytics.

llms.txt

A plain-text file at the root of a website providing AI engines a structured map of the site's most important content. A young convention, not yet universally honored, but low-cost and useful.

Agentic Browser

An AI system that visits pages on a user's behalf and completes tasks — booking, form-filling, purchasing. Requires CLS = 0 and a complete accessibility tree, beyond the five citation signals.

AI Crawler Reference

Primary AI crawlers in 2026. Bot names are case-sensitive in robots.txt.

USER-AGENT STRING	ENGINE	DECISION
GPTBot	OpenAI · ChatGPT training	Allow
ChatGPT-User	OpenAI · on-demand fetch	Allow
OAI-SearchBot	OpenAI · search index	Allow
ClaudeBot	Anthropic · Claude	Allow
anthropic-ai	Anthropic · on-demand	Allow
PerplexityBot	Perplexity	Allow
Perplexity-User	Perplexity · on-demand	Allow
Google-CloudVertexBot	Google AI	Allow
Google-Extended	Gemini training	Allow
Applebot	Apple Intelligence	Allow
xAIBot	Grok · xAI	Allow
FacebookBot	Meta	Block
Meta-ExternalAgent	Meta	Block

USER-AGENT STRING	ENGINE	DECISION
Meta-ExternalFetcher	Meta	Block
Manus Bot	Manus · agentic scraper	Block

Citation Self-Test Protocol

Run before starting the 30-day plan to establish a baseline. Repeat at 60 days to measure improvement.

1. Open ChatGPT (browse mode), Perplexity, and Google (AI Mode or AI Overviews).
2. Type ten questions your ideal customer would ask when looking for a business in your category and location. Write in natural language — not SEO keywords.
3. For each answer, record: (a) whether your business is named, (b) whether a competitor is named, (c) whether any business in your category is named.
4. Score: one point per engine per question where your business is named. Maximum: 30. Baseline for most businesses: 0–2.
5. A score above 15 at 60 days indicates strong progress. A score of 25+ indicates solid performance across all three engines.

EXAMPLE QUESTIONS BY VERTICAL

- **Personal injury law:** "Best car accident lawyer in Toronto?" · "Which law firm handles slip and fall cases in Ontario?"
- **Dental / medspa:** "Best dental implant clinic in Amsterdam?" · "Where to get Botox in Rotterdam from a qualified practitioner?"
- **Solar:** "Reliable solar panel installer in British Columbia?" · "Solar companies in Ontario with good warranties?"

- **Real estate:** "Best buyer's agent for first-time homebuyers in London?" · "Commercial real estate broker in Dublin?"

Sources and Citations

Every statistic in this book is sourced below. Numbers from the author's own site are identified as such and are independently verifiable at the URLs listed.

[1] AI SEARCH CONVERSION RATE ADVANTAGE (~5×)

Opollo, "AI Search Traffic Converts 4–5x Better Than Google," 2025.

<https://opollo.com/blog/ai-search-traffic-converts-4-5x-better-than-google/>

Used in Chapters 2 and throughout where the conversion rate advantage of AI-referred traffic is cited. The specific figures of ~14.2% (AI) vs ~2.8% (Google organic) are from this source.

[2] AI TRAFFIC ENGAGEMENT AND GROWTH

Adobe Digital Insights, "AI Is Changing How Consumers Engage with Websites," February 2025.

Source for: AI-referred visitors stay ~41% longer; ~23% lower bounce rate; AI-driven referral traffic grew ~12× between July 2024 and February 2025 (US market).

[3] MOBILE LOAD SPEED AND CONVERSION

Google and Deloitte, "Milliseconds Make Millions," 2019.

Source for: a one-second improvement in mobile load time increases conversions by ~20% and reduces bounce rate by ~32%. Also source for the average legal landing page load time of ~4.8 seconds on mobile.

[4] GOOGLE ADS QUALITY SCORE CPC IMPACT

Google Quality Score framework and Think with Google mobile benchmarks.

Source for: a Quality Score of 3 versus 9 can roughly triple effective CPC for the same keyword, bid, and audience.

[5] PAGESPEED PROOF-OF-WORK — VSOURCECODE.COM

Author's own site. Live and independently verifiable.

Mobile (KB): pagespeed.web.dev/analysis/https-vsourcecode-com-kb/l75ddd3x2p?form_factor=mobile

Desktop: pagespeed.web.dev/analysis/https-vsourcecode-com/546k3vv85n?form_factor=desktop

[6] SECURITY STATISTICS — VSOURCECODE.COM

Author's own Cloudflare dashboard. 42,149 SSL requests served; 860 attacks blocked or challenged — 30-day window ending June 2026.

[7] 88% OF AI CITATIONS FROM OUTSIDE GOOGLE TOP 10

vSourceCode field research and client audits, 2025–2026. Based on citation scoring across ChatGPT, Perplexity, and Google AI Overviews for 40+ business categories across North American and European markets.

The Algorithm Tax concept and its four-axis framework are original to vSourceCode (vsourcecode.com/concepts/algorithm-tax). The Taro Mizushima quotation is cited as it appears in the author's source material; independent verification of the original newsletter was not possible. The five-signal GEO framework and the terms Citation Gap, Technical Tax, Training Tax, Ghost Traffic, Plugin Graveyard, Silent Hang-Up, Sovereign Node, and Viewport Attention Zone are coined terms original to vSourceCode as published.