

The AI Visibility Technical Checklist

Twelve checks. Most take under ten minutes. None require a developer.

This is the unglamorous layer underneath everything else. You can write perfect content and still be invisible if the crawler never reaches the page. Do this first, then worry about what you are publishing.

Access • can AI systems reach you at all

☐ 1. Check your robots.txt for AI crawler blocks

Visit **yourdomain.com/robots.txt**. Look for any of these followed by **Disallow:** / • GPTBot • OAI-SearchBot • ClaudeBot • Claude-SearchBot • Claude-User • PerplexityBot • Google-Extended

If you find one, that crawler was told not to touch your site.

☐ 2. Check your firewall or bot-management settings

This is the one almost nobody checks and the one that matters most. Since July 1 2025, every new Cloudflare zone blocks GPTBot, ClaudeBot, and PerplexityBot by default. The block happens at the edge, before robots.txt is even read, so a clean robots.txt proves nothing.

On Cloudflare: Security settings, look for "Block AI Bots" or Bot Fight Mode. On Webflow: Cloudflare runs underneath your dashboard since the April 2026 migration, contact support and ask directly. Any other host: ask whether a CDN or proxy sits in front of your site.

☐ 3. Submit your site to Bing Webmaster Tools

ChatGPT browses through Bing. If you are not indexed there, ChatGPT's live search cannot find you regardless of how well you rank on Google. Most small operators have never done this.

Free, about ten minutes, and the single most-skipped item on this list.

☐ 4. Confirm your site is indexed in Google Search Console

Table stakes, but verify rather than assume. Google's AI Overviews lean heavily on pages that already rank organically.

Structure • can a model extract a clean answer

☐ 5. Answer-first opening on every page

The first two sentences of the page, and of every major section, should answer the question the heading implies. Research on AI Overview citations has found that if your answer is not in the first third of a page, there is a better than even chance it will not be cited.

☐ 6. Paragraphs of two to four sentences

Retrieval systems break your page into chunks before anything reads it whole. A ten-sentence paragraph containing three ideas cannot be extracted cleanly. Short paragraphs are citation units.

☐ 7. Question-style headings that mirror real queries

"Content Structure Tips" tells a model nothing. "How do I structure content so AI cites it?" matches how people actually ask. Models use headings as entry points.

☐ 8. A definition block for every key term

One to three sentences, no jargon, phrased as a direct definition. "X is defined as..." outperforms "X refers to the way...". Write one for any term you introduce.

☐ 9. An FAQ section, five to eight questions, answers that stand alone

Each answer must work with no surrounding context. "As mentioned above, the key is structure" is unusable. State the answer in the first sentence, add one or two sentences of support, end without referring elsewhere in the page.

- ☐ 10. Schema markup on the pages that matter
- Organization schema on your homepage. FAQPage schema on question pages. Product schema on offers. Industry analysis in 2026 found 65% of pages cited by Google AI Mode and 71% cited by ChatGPT carry structured data. Correlation, not causation, but it is an afternoon of work and it stops the machine guessing.
- ☐ 11. An author bio with real, verifiable credentials
- Name, specific professional title, one line of specific expertise, and a link to a verifiable profile. "Passionate about digital marketing" adds no signal. Named credentials and a linked profile do.
- ☐ 12. At least one piece of original data on your site
- If your content only restates what everyone else says, a model has no reason to cite you over any other source. One specific finding, case study number, or observation that exists nowhere else makes you a required citation rather than an optional one. It does not need a research budget. It needs you documenting something you already observe.

Scoring yourself

Score	What to do
10 to 12	Your technical layer is solid. Your gap is almost certainly earned media, where you appear on sites you do not own.
6 to 9	Fix the unchecked access items first, numbers 1 to 4. Those are binary and they gate everything else.
Under 6	Start at item 1 and work down. Do not write another page until items 1 to 4 are clear, because none of it can be retrieved.

Re-check

Run items 1 through 4 again after any hosting change, CDN change, or platform migration. Defaults get reset and nobody sends a notification. Run 5 through 12 whenever you publish something you actually care about being cited.

Want this checked properly?

The free AI Visibility Scan runs six real checks against your live site in about sixty seconds. No email required.

closermethod.com/seo

Sources: Cloudflare Content Independence Day AI crawler policy, July 2025. Documented 2026 Webflow/Cloudflare infrastructure migration reporting. CXL analysis of Google AI Overview citation sources. Stackmatix 2026 structured-data citation analysis. Ankit Prajapati, AI citation writing methodology, June 2026.