

USER GUIDE

AI Bot Access

Three tools, one goal: make sure AI models can read your site and describe it accurately. Here's how to use every tool in the AI Bot Access toolkit.

PUBLISHED BY EBIZ123S

aibotaccess.com

CONTENTS

What's Inside

1

Why this matters (the short version)

2

Tool 1 — robots.txt Scanner

3

Tool 2 — Free llms.txt Generator

4

Tool 3 — AI llms.txt Builder (BYOK)

5

Which tool should I use?

6

The recommended workflow

7

Troubleshooting & FAQ

8

Disclaimer

SECTION 1

Why this matters

AI-generated answers are replacing the top of search. The question now isn't just "Where do I rank?" — it's "Do AI models even know I exist?"

When someone asks ChatGPT, Claude, Perplexity or Google's AI Overviews a question, the models decide which sources to cite based on two things: **whether they're allowed to read your site**, and **whether they understand what your site is about**.

The AI Bot Access toolkit covers both sides of that equation:

- **Access** is controlled by robots.txt. Our Scanner tells you which AI crawlers your current rules are blocking.
- **Understanding** is communicated by llms.txt — a plain-Markdown file that tells AI models what your site does and which pages matter. Our Free Generator and AI Builder help you create one.

QUICK DEFINITION

"llms" stands for Large Language Models — the AI systems behind ChatGPT, Claude, Perplexity, and Google's Gemini. An llms.txt file is a positive signal you put on your site to help these models understand what they're looking at.

TOOL 1

robots.txt Scanner

FREE · PUBLIC · NO SIGNUP

robots.txt Scanner

aibotaccess.com

What it does: Fetches your site's robots.txt, parses it using the same logic Google's robots parser uses, and resolves each major AI crawler against the ruleset. For each bot you see **Allowed**, **Partially allowed**, or **Blocked**, plus an explanation of why.

When to use it

- You just launched or migrated a site and want to confirm AI tools can read it.
- You're not sure whether a security plugin, CDN, or default hosting rule is quietly blocking AI bots.
- You changed your robots.txt and want to verify the new rules do what you intended.
- You're running an audit for a client site before shipping a visibility strategy.

Benefits

- **Covers 18 AI crawlers** — GPTBot, OAI-SearchBot, ChatGPT-User, ClaudeBot, Claude-Web, anthropic-ai, PerplexityBot, Perplexity-User, Google-Extended, Applebot-Extended, CCBot, Meta-ExternalAgent, Bytespider, cohere-ai, DuckAssistBot, Amazonbot, Diffbot, and YouBot.
- **Distinguishes training bots from answer bots** so you know which blocks hurt your live AI citations vs. long-term model training.
- **Copy-paste fix:** one click gives you a corrected robots.txt snippet that explicitly allows the AI crawlers you want.
- **Also checks llms.txt** presence as a positive signal.

How to use it

1. Go to **aibotaccess.com** and enter any domain (e.g. example.com or https://example.com).
2. Click **Check Site**. The scanner fetches /robots.txt via a server-side proxy (no CORS issues) and analyzes it.

3. Review the summary: **Allowed**, **Partial**, and **Blocked** counts tell you at a glance where you stand.
4. Scroll the per-bot list to see exactly which rule (if any) is blocking each crawler.
5. If you have blocks to fix, click **Copy to Clipboard** under the recommended robots.txt section and paste the result into your site's robots.txt at the root of your domain.
6. Re-scan to confirm the new rules behave as expected.

Reading the results

- **Allowed** — the bot is free to fetch the whole site.
- **Partial** — the bot can access the root, but specific paths are disallowed (common for /admin, /cart, etc. — usually fine).
- **Blocked** — the bot is disallowed from the root. This is the problem state; the fix-snippet addresses it.

TIP

If a block shows up that you don't recognize, check for: (1) a bare User-agent: * with Disallow: /, (2) a WAF or firewall rule at your CDN (Cloudflare, etc.) that blocks bots before they even see robots.txt, or (3) a stale cached robots.txt served by your CDN.

Before you publish

The scanner's recommended robots.txt is an allow-everything starting template. That's the right fix for the most common problem — AI crawlers being blocked. But a complete robots.txt usually also blocks some paths that shouldn't be indexed by any crawler. Before you upload, add Disallow: lines under User-agent: * for anything on your site that matches these patterns:

- **Admin and login areas** — /admin/, /wp-admin/, /wp-login.php, /dashboard/
- **Cart, checkout, and user accounts** — /cart, /checkout, /my-account/, /account/
- **Search result pages** — /search, /?s=
- **Thank-you and confirmation pages** — /thank-you, /order-confirmation
- **Staging, dev, or test folders** — /staging/, /dev/, /test/
- **Tracking parameter noise** — /*?utm_, /*?fbclid=, /*?gclid=

These paths shouldn't appear in Google, Bing, or AI answers. If a path doesn't exist on your site, skip the line for it. You don't need the line to have the path exist — it's safe to leave out.

TOOL 2

Free llms.txt Generator

FREE · PUBLIC · NO SIGNUP

Free llms.txt Generator

aibotaccess.com/llms-txt-generator

What it does: Reads your site's sitemap.xml and homepage, ranks the most likely important pages, and drops them into a valid llms.txt template you can edit, copy, and publish. Purely template-based — no AI call, no cost.

When to use it

- You don't have an llms.txt yet and want to ship one today.
- Your site is small-to-medium (roughly under 50 content pages).
- You're comfortable editing the generated descriptions yourself.
- You want a zero-cost baseline before deciding whether to invest in the AI Builder.

Benefits

- **Instant output.** No API keys, no wait.
- **Correct structure.** Follows the llmstxt.org spec exactly — H1, blockquote summary, H2 sections with linked page lists.
- **Sensible page ranking.** Picks up to 10 pages, prioritizing shallow paths and keyword hits like /about, /pricing, /docs, /features, /blog, etc.
- **Editable in place.** Result lands in a textarea so you can rewrite descriptions, reorder sections, or remove links before downloading.

How to use it

1. Go to aibotaccess.com/llms-txt-generator and enter your site URL.
2. Click **Generate llms.txt**. The tool fetches your homepage (for title and description) and your sitemap (for URLs), then builds the template.

3. Review the draft in the editable textarea. It will look something like this:

```
# Your Site
> One-line summary from your meta description

## Key pages
- [Home](https://yoursite.com): replace this with a one-line description...
- [About](https://yoursite.com/about): replace this...
```

4. Replace the “replace this” placeholders with real one-line descriptions of each page.
5. Click **Copy** or **Download llms.txt**.
6. Upload the file to the root of your domain so it's reachable at <https://yoursite.com/llms.txt>.

WHERE TO PUT IT

llms.txt belongs at your domain root, same place as robots.txt. On WordPress, use a plugin or a rule in your theme's functions. On Next.js / Astro / static sites, drop it in the public/ folder. On Nginx or Apache, it's a file in your docroot.

Before you publish

The generated llms.txt is a structural template, not a finished file. Before uploading it to your site, do these three things:

1. **Remove pages you don't want cited.** The generator includes what it finds in your sitemap. Delete any URL that isn't representative of your brand — tag archives, old posts, checkout pages, utility pages.
2. **Replace the placeholder descriptions.** The “replace this with a one-line description...” lines are meant to be rewritten in your own words. AI models pay attention to these, so make each one accurate, specific, and plain-spoken.
3. **Keep it under ~40 links total.** Curation beats completeness. Shorter llms.txt files outperform long ones — list only the pages you genuinely want AI models to reference when describing your site.

If you don't have the time to write descriptions by hand, Tool 3 (the AI Builder) handles this automatically by crawling your actual page content.

TOOL 3

AI llms.txt Builder (BYOK)

FREE · BYOK · NO SIGNUP

AI llms.txt Builder

aibotaccess.com/ai-llms-txt-generator

What it does: Crawls up to 200 of your pages, reads their actual content, and hands everything to Claude or GPT. The AI writes a full llms.txt with real, tailored, brand-aware descriptions — no templates, no placeholders. You bring your own Anthropic or OpenAI API key, so you pay only the provider's per-token cost (pennies per site, typically).

When to use it

- You run a larger site where templated descriptions won't capture the value of each page.
- You're an agency producing llms.txt as a deliverable for multiple clients.
- You want polished, brand-voice descriptions without writing them by hand.
- You need to ship fast and your client's site has 50+ meaningful pages.

Benefits

- **Real descriptions, not stubs.** The model actually reads each page's content and summarizes it in-context.
- **Brand voice.** Reads your homepage to pick up tone and carries it through every section.
- **Scalable crawl.** Slider from 3 to 200 pages — go deep when sites are content-rich.
- **Pay pennies, not hundreds.** BYOK means you're billed by Anthropic or OpenAI directly — typically \$0.05 to \$1.00 per site depending on model and page count.
- **Progress bar + Stop & Save.** Watch each page complete; stop any time to save a draft from whatever has been fetched.
- **Self-documenting output.** The generated file includes an HTML-comment footer recording the date, page count, and model used.

Before you start

- Have an **Anthropic API key** (sk-ant-...) or **OpenAI API key** (sk-...) ready. Keys are entered in the tool and sent only to your chosen provider — we never store them.
- Know roughly how many pages your site has so you can pick a sensible crawl depth.

How to use it

1. Visit aibotaccess.com/ai-llms-txt-generator and click **Launch Builder**.
2. On the builder page, enter your site URL (e.g. <https://example.com>).
3. Pick your AI provider (Anthropic or OpenAI) and model. **Claude Sonnet 4.6** is the recommended default; **Haiku 4.5** is faster and cheaper for simple sites; **Opus 4.6** is overkill for most cases.
4. Paste your API key. Leave “Remember key on this device” checked unless you’re on a shared machine.
5. Use the **Pages to scan** slider — 15 is the sensible default; go higher for large content sites. Watch the **Estimated cost** box update in real time.
6. (Optional) Add brand notes or instructions in the text area — e.g., “we’re B2B enterprise, avoid marketing fluff” or “emphasize our developer audience.”
7. Click **Generate llms.txt with AI**.
8. Watch the progress bar: homepage → sitemap → crawl N/Total → Claude/OpenAI call → done.
9. If you need to abort, click **Stop & save**. Any pages already fetched are saved as a draft template so you don’t lose work.
10. Review the generated llms.txt. Edit freely, then **Copy** or **Download**.
11. Upload the file to your domain root at /llms.txt.

COST ESTIMATE

A typical 15-page site run costs **\$0.05 to \$0.15** on Haiku, **\$0.30 to \$0.80** on Sonnet. A large 100-page site on Sonnet might run **\$2 to \$5**. These are your API costs paid directly to Anthropic or OpenAI — we don’t mark anything up.

Tips for better output

- **Use the instructions field** to tell the AI what kind of business you are, who your audience is, and any descriptor constraints (“no em-dashes”, “British English”, etc.).
- **Scan 20 to 30 pages** for content-heavy sites rather than the default 15 — the model can then group pages into cleaner sections.
- **Re-run after editing** if the output misses a key theme. The Regenerate button reuses your settings.
- **Ship it, then refine.** Even a first draft is a huge upgrade over no llms.txt — don’t let perfect be the enemy of deployed.

Before you publish

The AI generates descriptions from your actual page content, but you know your business better than any model does. Before uploading the result:

- **Scan for inaccuracies.** AI can misread pricing tiers, feature scope, or positioning. A 30-second read catches 95% of errors.
- **Remove pages you don’t want cited.** The crawler includes everything it finds — delete utility pages, staging links, or outdated content.
- **Tighten the voice if needed.** Use Regenerate with additional instructions (e.g., “drop all marketing adjectives”) if the tone isn’t right for your brand.

SECTION 5

Which tool should I use?

Quick lookup table. Find your goal in the left column; the right column tells you which tool to reach for first.

Your goal	Best tool
Check if AI can read my site	robots.txt Scanner
Diagnose why a specific bot is blocked	robots.txt Scanner
Get a quick llms.txt template for a small site	Free Generator
Write llms.txt descriptions by hand with a starting point	Free Generator
Produce a polished llms.txt for a large content site	AI Builder
Run llms.txt as an agency deliverable	AI Builder
Generate multiple client llms.txt files fast	AI Builder

SECTION 6

The recommended workflow

Here's the end-to-end sequence we suggest for any site — new launch, migration, or audit.

- 1. Scan first.** Run the robots.txt Scanner. If AI bots are blocked, paste the recommended snippet into your robots.txt and verify with a re-scan. No point writing an llms.txt if bots can't read the site.
- 2. Generate a starter llms.txt.** Run the Free Generator to get a structurally-correct baseline. For small sites this may be all you need after a round of manual edits.
- 3. Upgrade to AI when it matters.** For client work, marketing-heavy sites, or sites where description quality has real business impact, run the AI Builder with Sonnet 4.6 and 15 to 30 pages.
- 4. Publish.** Upload the final llms.txt to your domain root so it's live at /llms.txt.
- 5. Verify.** Run the scanner one more time — it'll confirm your new llms.txt is discoverable.
- 6. Re-check quarterly.** AI crawlers come and go, and your pages evolve. A 5-minute re-scan every 3 months is a small insurance premium on your AI visibility.

SECTION 7

Troubleshooting & FAQ

The scanner says my robots.txt is blocking GPTBot, but I didn't write that rule.

Most often this is a bare User-agent: * + Disallow: / combination, or a security plugin or firewall that injects rules. Start by viewing your robots.txt in a browser (yoursite.com/robots.txt) and checking the raw content. If it's empty but bots still fail, suspect a CDN or WAF rule blocking by user-agent at a layer above robots.txt.

My robots.txt looks correct but AI bots still can't reach my site. What else could it be?

The most common cause is **Cloudflare Bot Fight Mode**. If your site is behind Cloudflare, its bot protection can silently block AI crawlers at the CDN layer — before they ever see your robots.txt. To check: log into Cloudflare, go to **Security** → **Bots**, and look for Bot Fight Mode or Super Bot Fight Mode. If either is ON, that's likely your problem. The simplest fix is to turn it off entirely — your site keeps DDoS protection and SSL regardless. If you want to keep bot protection for other crawlers, create a **WAF custom rule** (Security → WAF → Custom rules) that matches AI crawler user-agents (GPTBot, ClaudeBot, PerplexityBot, etc.) and sets the action to **Skip**. This lets AI bots through while blocking everything else. After changing settings, wait 2–5 minutes and re-scan at **aibotaccess.com** to verify.

Will allowing AI crawlers hurt my Google rankings?

No. AI crawlers (GPTBot, ClaudeBot, etc.) are separate from traditional search crawlers (Googlebot, Bingbot). Allowing one does not affect the other.

The free generator only uses my sitemap — what if I don't have one?

If there's no sitemap at /sitemap.xml or /sitemap_index.xml, the generator builds a template from your homepage alone. You can still edit the result to add key URLs manually.

What's the daily usage limit?

5 builder runs per IP per 24 hours, which covers most individual users. Agencies running this for multiple clients can request a higher limit via the form on the landing page at aibotaccess.com/ai-llms-txt-generator.

Is my API key safe in the AI Builder?

Your key is sent only to your chosen provider (Anthropic or OpenAI). It's stored in your browser's local storage if you tick "Remember key on this device" so you don't re-enter it. Our serverless proxy relays the one request to the provider and does not log request bodies or keys. If you prefer, untick "Remember" after each run, or click **Sign out** to clear everything.

What if the AI-generated llms.txt isn't quite right?

The output is always shown in an editable textarea. Tighten voice, reorder sections, remove pages you don't want featured, or re-run with different instructions. The Regenerate button reuses your domain and settings for a fresh pass.

Can I use these tools for client sites?

Yes — especially the AI Builder. Nothing in the workflow identifies you or your clients to us, and the output is yours to ship under your own brand.

How often should I re-generate my llms.txt?

When you add or remove major pages, when your site's messaging changes materially, or every 6 to 12 months as a refresh. Unlike robots.txt, llms.txt isn't fetched on every crawl — models pull it occasionally as context, so minor page changes don't warrant re-generation.

SECTION 8

Disclaimer

The Service. The tools, templates, generated files, documentation, and any other content made available on aibotaccess.com (collectively, “the Service”) are provided “as is” and “as available” for informational and educational purposes only, without warranty of any kind, whether express, implied, statutory, or otherwise, including but not limited to warranties of merchantability, fitness for a particular purpose, accuracy, non-infringement, or that the Service will be uninterrupted, error-free, or secure. Use of the Service is entirely at your own risk.

Your responsibility. You are solely responsible for reviewing, testing, editing, and validating any output produced by the Service — including but not limited to generated robots.txt files, generated llms.txt files, AI-written descriptions, and any recommended configuration changes — before deploying or applying them to any website, system, or third-party property. You are responsible for ensuring that any changes you make comply with all applicable laws, terms of service, and the policies of any platforms, crawlers, or services involved.

No liability. To the maximum extent permitted by law, AI Bot Access, eBiz123s, SolutionC LLC, their officers, employees, contractors, and affiliates shall not be liable for any direct, indirect, incidental, special, consequential, exemplary, or punitive damages, or any other loss or damage whatsoever arising from or in connection with your use of the Service, including but not limited to: loss of search engine rankings, loss of organic or AI-driven traffic, loss of revenue, loss of data, indexing errors, misconfiguration of robots.txt or llms.txt, API costs incurred with third-party providers (including Anthropic, OpenAI, or others) through BYOK usage, security incidents, privacy issues, unintended blocking or allowing of any web crawler or bot, or any claim by a third party arising from your implementation of Service output.

Not professional advice. Nothing on this Service constitutes legal, financial, SEO, marketing, security, or technical advice. Information herein reflects our best understanding of AI crawler behavior, search engine policies, and industry standards at the time of writing — all of which change frequently and without notice. You should consult qualified professionals for advice specific to your situation.

Third-party services. When using BYOK (Bring Your Own Key) features, your API key is transmitted to your chosen provider (Anthropic, OpenAI, or others) and all API usage is governed by that provider’s terms of service and billing policies. AI Bot Access bears no responsibility for third-party provider actions, outages, billing, or data handling.

Changes. This disclaimer may be updated at any time without notice. Continued use of the Service constitutes acceptance of any updated version. For the current full version, see aibotaccess.com/disclaimer.

ONE LAST THING

Your toolkit

All three tools live at **aibotaccess.com**. They're free to try, and the first two are completely free forever — the AI Builder uses your own API key so the only cost is what you pay Anthropic or OpenAI directly.

→ **THE robots.txt SCANNER**

Check your site against 18 AI crawlers in 5 seconds. **aibotaccess.com**

→ **THE FREE llms.txt GENERATOR**

Build a starter llms.txt from your sitemap. **aibotaccess.com/llms-txt-generator**

→ **THE AI llms.txt BUILDER**

Free · BYOK · No signup. AI-written llms.txt with your own API key.

aibotaccess.com/ai-llms-txt-generator

Want the deeper strategy behind these tools? Grab **The AI Visibility Playbook** — a free 30-page tactical manual on how to actually get cited in AI answers. Available at **aibotaccess.com**.

— The eBiz123s Team

aibotaccess.com

Tools and guidance provided “as is” for educational purposes, without warranty. You are responsible for reviewing output before publishing. AI Bot Access and eBiz123s bear no responsibility for loss, damage, lost rankings, lost traffic, indexing issues, or other consequences from use of these tools. See Section 8 above, or the full version at **aibotaccess.com/disclaimer**.