

# Half of the Inc 5000's llms.txt files were written by a plugin, not the company

Installmap, 2026-09-04

48.3% of the 1,815 llms.txt files we found on Inc 5000 domains name the software that produced them, and the two biggest names are Yoast SEO at 17.8% and Shopify at 17.4%.

## Key findings

- 48.3% of the 1,815 Inc 5000 llms.txt files were written by a WordPress plugin or by Shopify. Yoast SEO wrote 17.8% of them, Shopify's template 17.4%, All in One SEO 7.4% and Rank Math 5.0%.
- 36.8% of Inc 5000 domains serve an llms.txt, against 33.6% of unicorns, 29.3% of YC companies and 12.2% of the Fortune 500.
- Counting only the files with no generator signature, Inc 5000 adoption falls to 19.0% and the Inc 5000 drops to last: unicorns 28.3%, YC 27.3%, Fortune 500 10.7%.
- Consumer Products is the highest Inc 5000 industry cut at 80.1%, and 86.0% of the files in it are the Shopify template.
- Inc 5000 robots.txt files disallow CCBot on 3.8% of domains and PerplexityBot on 0.1%, though 5.7% name PerplexityBot.

We asked 12,713 company domains for three files, llms.txt, llms-full.txt and robots.txt, expecting to count a decision: somebody at the company wrote a file for crawlers and put it up. Then we read the files. Half of the Inc 5000 ones say which software produced them, and the software is Yoast, Shopify, All in One SEO and Rank Math.

## 48.3% of the Inc 5000 files carry a generator signature, against 7.1% of the YC files

| Generator       | Inc 5000 (n=1,815 files) | YC (n=1,652) | Unicorns (n=355) | Fortune 500 (n=56) |
|-----------------|--------------------------|--------------|------------------|--------------------|
| Yoast SEO       | 17.8%                    | 1.6%         | 6.8%             | 8.9%               |
| Shopify         | 17.4%                    | 2.8%         | 4.8%             | <5                 |
| All in One SEO  | 7.4%                     | 0.6%         | <5               | <5                 |
| Rank Math       | 5.0%                     | 0.5%         | 2.0%             | <5                 |
| Other generator | 0.8%                     | 1.6%         | <5               | <5                 |
| No signature    | 51.7%                    | 92.9%        | 84.2%            | 87.5%              |

Yoast SEO, All in One SEO and Rank Math each print their own name in the file, usually in a comment near the top, so we match that string in the first 3,000 characters. Shopify prints no name, so we match the shape of its file: an agent instructions heading together with either the word Shopify or the string UCP, the store's Universal Commerce Protocol discovery block above its MCP endpoints. The other generator row holds Webflow, HubSpot, Mintlify, SiteSEO and anything saying it was generated by or with something. The last row is files that name nothing, hand-written or otherwise.

Inc 5000 companies serve an llms.txt at 36.8% and the Fortune 500 at 12.2%

| Group           | Base  | Serves llms.txt | Serves llms-full.txt | Blocks an AI bot | Median llms.txt size |
|-----------------|-------|-----------------|----------------------|------------------|----------------------|
| Inc 5000 (2026) | 4,931 | 36.8%           | 8.4%                 | 4.4%             | 4,526 bytes          |
| Unicorns        | 1,058 | 33.6%           | 6.5%                 | 5.6%             | 7,852 bytes          |
| Y Combinator    | 5,630 | 29.3%           | 7.2%                 | 6.9%             | 4,920 bytes          |
| Fortune 500     | 459   | 12.2%           | <5                   | 3.3%             | 7,376 bytes          |

Take the signed files out and the order changes. Inc 5000 adoption falls from 36.8% to 19.0%, behind unicorns at 28.3% and YC at 27.3%; the Fortune 500 goes from 12.2% to 10.7%.

Fewer than five Fortune 500 domains serve an llms-full.txt.

17.4% of the Inc 5000 files are the Shopify template, and 86.0% of the Consumer Products files are it

The Shopify file is identical from store to store apart from the store name and the domain, and it lands in one industry cut. 80.1% of the 267 Inc 5000 Consumer Products domains serve an llms.txt and 86.0% of those files are the template, so that 80.1% is one platform rollout rather than 214 decisions. The same file is 2.8% of the YC files and 4.8% of the unicorn files.

Inc 5000 robots.txt files disallow CCBot on 3.8% of domains and PerplexityBot on 0.1%

| Crawler            | Inc 5000 (n=4,931) | YC (n=5,630) |
|--------------------|--------------------|--------------|
| CCBot              | 3.8%               | 6.3%         |
| Bytespider         | 3.7%               | 6.3%         |
| GPTBot             | 3.4%               | 5.9%         |
| Amazonbot          | 3.3%               | 5.3%         |
| Google-Extended    | 3.2%               | 5.7%         |
| ClaudeBot          | 3.1%               | 5.5%         |
| Applebot-Extended  | 3.1%               | 5.4%         |
| meta-externalagent | 3.0%               | 5.4%         |
| PerplexityBot      | 0.1%               | 0.2%         |

Perplexity gets written into robots.txt far more often than it gets blocked. 5.7% of Inc 5000 domains name PerplexityBot and 0.1% disallow it. The crawlers that do get shut out are the training and scraping ones. These shares run over the domains whose robots.txt we could read, 92.1% of the Inc 5000 and 72.3% of YC.

Fortune 500 adoption is 12.2%, and it includes nvidia.com, salesforce.com and

## target.com

Also serving the file: cvshealth.com, paypal.com, dell.com, gm.com, att.com, metlife.com, labcorp.com, thermofisher.com, carvana.com, autozone.com, acehardware.com, paloaltonetworks.com and usbank.com.

## Data funnel

| Step                                   | Inc 5000 | YC    | Fortune 500 | Unicorns | All    |
|----------------------------------------|----------|-------|-------------|----------|--------|
| Domains on the list                    | 4,969    | 6,133 | 500         | 1,111    | 12,713 |
| Fetch attempted                        | 4,969    | 6,133 | 500         | 1,111    | 12,713 |
| robots.txt returned a status, the base | 4,931    | 5,630 | 459         | 1,058    | 12,078 |
| Dropped                                | 38       | 503   | 41          | 53       | 635    |

The dropped domains carry a recorded reason:

| Drop reason         | Inc 5000 | YC  | Fortune 500 | Unicorns | All |
|---------------------|----------|-----|-------------|----------|-----|
| Connection failed   | 20       | 306 | 20          | 29       | 375 |
| Timeout or deadline | <5       | 85  | 19          | 20       | 128 |
| TLS failure         | 12       | 104 | <5          | <5       | 122 |
| Redirect loop       | <5       | 8   | 0           | 0        | 10  |

A connection failure covers a dead domain, a refused connection and a name that did not resolve; the fetcher records one kind for all three. A domain that answered with a 403 is in the base and counts as not serving the file.

## How we counted

We requested three paths from every domain over plain HTTP on 2026-09-04, llms.txt, llms-full.txt and robots.txt, through a rotating proxy with a browser user agent, trying HTTPS first and HTTP second, with a 15 second budget per domain and the body capped at 200 KB. Nothing was rendered, since these files are served as text. The base for each group is the domains whose robots.txt request came back with any HTTP status.

The generator is read from the first 3,000 characters of the body, lowercased, in a fixed order: the Shopify shape first, then the literal strings "all in one seo", "sitemaps", "yoast", "rank math" or "rankmath", "webflow", "framer", "hubspot" and "mintlify", then a match on "generated by" or "generated with". The first rule that matches wins, so a file that mentions a plugin for another reason is misread as that plugin.

Two floors apply. We do not print a share on a base under 20 domains, and a cell reading "<5" is a count under five. We read the files at the apex host only, and a file at that host says nothing about whether its contents are accurate, current, or of any use to a model.

## Dataset

The per group metrics and the per domain rows behind every share are published under

/static/data/llms-txt-adoption/ as data.csv, which carries a generator column on every domain row, a README.md listing the columns and denominators, and a METHOD.md covering the crawl date, the fetch method, the signature rules, the floors and the funnel. Data: CC BY 4.0, Installmap, 2026-09-04. Cite: Source: Installmap, September 2026.

## Source

<https://installmap.com/research/llms-txt-adoption>

Dataset: data.csv, README.md, METHOD.md

## Denominators

| Group                  | Domains scored | On the list |
|------------------------|----------------|-------------|
| Inc 5000 (2026)        | 4,886          | 4,969       |
| Y Combinator companies | 5,386          | 6,133       |
| Fortune 500            | 472            | 500         |
| Unicorns               | 1,000          | 1,111       |