

llms.txt: a map of your site for AI models

Level: beginner · Reading time: ~9 minutes

`robots.txt` tells crawlers where they may go. `sitemap.xml` tells them what exists. `llms.txt` is a newer, third idea: a short, human-written **map of your best content, in Markdown**, addressed to language models rather than search engines.

It is worth understanding — and worth being honest about, because the evidence on whether anything actually reads it is mixed. We will cover both.

The problem it tries to solve

When a language model reads a web page, it receives HTML: navigation, cookie banners, sidebars, footers, analytics scripts. The actual article might be 10% of the bytes. Models have limited context, so a lot of that budget is wasted on chrome before the content is even reached.

`llms.txt` proposes a shortcut: one Markdown file at your root that says *this is what this site is, and here are the pages that matter, with a sentence about each*. A model can read that one file and know where to go.

The convention was proposed by [Jeremy Howard](#) of Answer.AI on **3 September 2024**. It is a community convention published at llmstxt.org — **not** an IETF or W3C standard, and not something any search engine is obliged to honour.

What goes in it

The spec is deliberately tiny. In order:

1. **An H1 with the site or project name.** This is the **only required element**.
2. **A blockquote** with a short summary — "key information necessary for understanding the rest of the file". Optional but strongly conventional.
3. **Free-form Markdown** — paragraphs, lists, anything **except headings** — for more detail.
4. **Zero or more H2 sections**, each containing a list of links.

5. Optionally, a section literally titled `## Optional`, which by convention holds "links an agent can skip when a shorter context is needed".

Each link line follows one shape: a required Markdown hyperlink, then optionally a colon and a note.

```
- [Link title](https://example.com/page): Optional description
```

A complete example

```
# Example Docs

> Developer documentation for the Example API, covering authentication,
> endpoints, and client libraries.

Written for developers integrating Example into a production application.
All examples use the v2 API.

## Getting started

- [Quickstart](https://example.com/docs/quickstart): Make your first call
in five minutes
- [Authentication](https://example.com/docs/auth): API keys, OAuth, and
token rotation

## Reference

- [Endpoints](https://example.com/docs/api): Every route with request and
response shapes
- [Errors](https://example.com/docs/errors): Status codes and how to
recover from each

## Optional

- [Changelog](https://example.com/changelog): Version history
- [Blog](https://example.com/blog): Longer-form articles
```

That is the entire format. If you know Markdown, you already know it.

Where it goes

At `https://yourdomain.com/llms.txt`. The spec also allows it "at any path within it, covering the pages under that path", with agents preferring **the**

most specific one — so a large site can publish `/docs/llms.txt` alongside the root file.

The honest part: does anything read it?

This is where most guides stop being useful, so here are the facts as of 2026.

Against it:

- Google's Gary Illyes said in July 2025 that Google does **not** support `llms.txt` and has no plans to. John Mueller compared it to the long-discredited keywords meta tag.
- Published crawl analyses report that AI crawlers overwhelmingly fetch HTML directly and rarely request `/llms.txt` at all — one widely-cited study of roughly 500 million AI bot visits over 90 days found only a few hundred hits on the file.
- Adoption studies suggest a large share of the `llms.txt` files that do exist are empty stubs generated by plugins, which does not help the convention's credibility.

For it:

- Anthropic has stated support for the format, and developer tools like **Cursor** use `llms.txt` to understand a project.
- Documentation platforms such as **Mintlify** generate it natively, so it is already standard in developer-docs circles.
- It costs about an hour to produce and nothing to maintain if generated from your content.

The realistic framing: `llms.txt` is a cheap option on a future where agents route on machine-readable surfaces — not a traffic tactic that pays today. Treat it as insurance, not as SEO.

If your goal is being **cited by AI answers right now**, the evidence points much more strongly at structured data (schema.org `Organization`, `Article`, `FAQPage`), clean HTML, and content that states a direct answer in the first 40–60 words. `llms.txt` is a complement to that, not a substitute.

How to generate one

Without code

- llmstxt.studio — generates from a URL
- llmstxtgenerator.org — paste a site, get a file
- **Mintlify** docs sites publish `/llms.txt` automatically
- **WordPress**: several SEO plugins now emit one; check before installing another, since duplicate files at the same path cause confusion

Be careful with crawler-based generators. They tend to dump every URL they find, which produces a long, undifferentiated list — the opposite of the curation the format is for. **A hand-picked file of 15 links beats a generated file of 400.**

Static sites: just write `public/llms.txt` by hand. For a site with a dozen pages this is genuinely the right answer.

How to test it

```
curl -sI https://example.com/llms.txt # expect 200 and text/plain
curl -s https://example.com/llms.txt # read it as a model would
```

Then check it against the spec:

- ☐ Starts with a single H1 — nothing above it
- ☐ Blockquote, if present, comes immediately after the H1
- ☐ The free-form area contains **no headings**
- ☐ Only H1 and H2 are used
- ☐ Every list entry is `- [name](url)` with an optional `: note`
- ☐ Links are absolute, not relative
- ☐ `## Optional` holds genuinely skippable links
- ☐ Served as `text/plain`, `200`, no redirect

Should you add one?

Yes, if you publish documentation or technical content, you can generate it from existing data, and you accept it may sit unread for a year.

Skip it, if you would be hand-maintaining it, or you are hoping for a traffic increase. Put that hour into structured data instead.

References

- llmstxt.org — the specification

- [Jeremy Howard's original proposal](#)
- [llms-txt on GitHub](#)
- [Semrush: What is llms.txt and should you use it?](#)
- [Google Search Central Blog](#) — for Google's own position on emerging formats
- [schema.org](#) — the structured-data vocabulary that does measurably affect AI citation