How search engines and AI crawlers read your help center

Search and AI

Your help center tells search engines and AI tools what they may read and where to find it, with no setup on your part. Here is what every public help center publishes, what the signals mean, and how to turn it all off.

Plans: All plans

What your help center publishes

These addresses work on your help center's domain, including your custom domain if you have one.

Address	What it's for
/sitemap.xml	Lists your home page, categories and published articles, with links between language versions, so search engines find every page.

/robots.txt	Tells crawlers what they may read, links your sitemap, and states your content signals.
/llms.txt	A Markdown index of your categories and articles for AI assistants, in your default language, with the address of your public MCP server.
/ai.txt	Your AI usage preferences in plain text: AI assistants may read your pages and cite them, and training models on them is not permitted.
/.well-known/ai-catalog.json	Points AI tools to your public MCP server, when it is on.
/mcp	Your public MCP server, which AI assistants use to search and read your articles. See Let readers use your help center in their AI assistant.

A Markdown version of every page

AI tools read Markdown more reliably than a styled web page. Your home page, your list of categories and every category and article page return clean Markdown at the same URL when a tool asks for it with the Accept: text/markdown header:

```
curl -H "Accept: text/markdown" https://help.example.com/content/your-article
```

Readers can also download an article as Markdown from its **Export** menu, when your article page has one. Developers find every address, header and file format in llms.txt, ai.txt and Markdown pages on the developer portal.

What the content signals mean

Your robots.txt includes this line:

```
Content-Signal: search=yes, ai-input=yes, ai-train=no
```

- **search=yes:** search engines may index your pages and show them in results.
- **ai-input=yes:** AI assistants may read your pages to answer questions and cite them.
- **ai-train=no:** your content may not be used to train AI models.

Like robots.txt itself, these signals are instructions that well-behaved crawlers follow, not a lock.

Only what anyone can read

- The sitemap, llms.txt and the MCP server list only published content that a visitor without an account can read. Drafts, private articles and categories, unreleased translations and team notes stay out.
- Search results pages and "not found" pages tell search engines not to index them.
- Private and password-protected help centers are closed to crawlers, because a visitor has to sign in or enter the password first. Help centers limited to specific IP addresses are closed to crawlers from other addresses too. A help center that isn't published yet is closed to every crawler until you publish it.

Turn it off

1. In the left menu, click the gear icon, then **Settings**.
2. In the **General** card, switch off **Search engine indexing**.
3. Click **Save changes**.

Your robots.txt then asks every crawler to stay out and sets all three signals to no. llms.txt stops answering, ai.txt states that your help center has opted out, and the public MCP server turns off. Search engines can take a while to drop pages they already indexed.

The **Hide from search engines (no-index)** switch in the template editor works differently: it adds a noindex tag to your pages, but it doesn't change llms.txt or the public MCP server. See Hide your help center from search engines.

Related articles

- SEO settings for your help center
- Hide your help center from search engines
- Let readers use your help center in their AI assistant
- Make your help center ready for AI
- llms.txt, ai.txt and Markdown pages on the developer portal