# Customize your robots.txt

Last Modified on 04/30/2026 12:23 pm EDT

The robots.txt file lets you tell AI crawlers and search indexing crawlers whether they're allowed to crawl your public knowledge base. You can customize the robots.txt to block specific crawlers from crawling your knowledge base completely or from specific categories only. You can ignore this file if your knowledge base is completely private, since the login process prevents tools from crawling your site.

**Most common use**
Customize the robots.txt file if you want to prevent AI companies from training their models on your knowledge base content.

KnowledgeOwl generates a robots.txt file for all knowledge bases automatically. Here's what our default robots.txt file looks like:

```
User-agent: *
Disallow: /api/
```

If you have set your **SEO settings** to **Generate a publicly available sitemap** checked, robots.txt also includes that, for example:

```
User-agent: *
Disallow: /api/

Sitemap: https://myknowledgebase.knowledgeowl.com/help/sitemap.xml
```

## How to customize robots.txt

To customize your robots.txt file:

1. Go to **KB settings > Domain.**
2. If this is your first time customizing your robots.txt file:
   a. Select the **robots.txt** link in **Custom robots.txt** to open your robots.txt file in a separate tab:


   b. Copy the text from that separate tab.
   c. Paste it into the **Custom robots.txt** text box in your **Domain** settings.

Copyright © 2015 – 2025 Silly Moose, LLC. All rights reserved.
Powered by KnowledgeOwl knowledge base software.

3. Enter your new customizations in the **Custom robots.txt** text box.
4. Be sure to **Save** your changes.
5. Select the **robots.txt** link in **Custom robots.txt** in your **Domain** settings to verify that your robots.txt file looks as you'd like.

Check out the Quickstart below for more information on formatting your rules and to copy some rules to prevent the scraping used to train AI models.

# Quickstart to robots.txt customizations

The most common customization our authors like to do is to block certain AI-related crawlers from accessing their entire knowledge base or part of their knowledge base or to allow specific crawlers or bots to have access.

## Allow examples

To allow a specific crawler access to your knowledge base, you specify the `User-agent` you want to have access and then set what you want to `Allow` (or allow them to scan).

To allow a User-agent to scan your entire knowledge base:

```
User-agent: {User-agent name}
Allow: /
```

## Disallow examples

To block a crawler from accessing your knowledge base, you specify the `User-agent` you want to restrict and then set what you want to `Disallow` (or prevent them from scanning).

To prevent a User-agent from scanning your entire knowledge base:

```
User-agent: {User-agent name}
Disallow: /
```

To prevent a User-agent from scanning a specific category only:

```
User-agent: {User-agent name}
Disallow: /{category-name}/
```

## Combining allow and disallow

For example, ChatGPT has two different User-agents:

- `GPTBot`: Used for AI training data collection for GPT models

Copyright © 2015 – 2025 Silly Moose, LLC. All rights reserved.
Powered by KnowledgeOwl knowledge base software.

- `ChatGPT-User`: The AI agent for real-time web browsing when users interact with ChatGPT

If we wanted to prevent ChatGPT from crawling our knowledge base for AI training data collection, but still allow the AI agent to browse it to surface answers to questions in ChatGPT, we'd disallow GPTBot and allow ChatGPT-User. We'd add these entries to our **Custom robots.txt**:

```
# Block ChatGPT training bot
User-agent: GPTBot
Disallow: /

# Allow ChatGPT agent/browsing for answering questions
User-agent: ChatGPT-User
Allow: /
```

## Block all AI training scraping

If your goal is to block the scrapers that are used to train AI models, here are the key entries you'll want to add as of May 2026:

```
# Block major AI training bots
User-agent: GPTBot
Disallow: /

User-agent: Claude-Web
Disallow: /

User-agent: Google-Extended
Disallow: /

User-agent: CCBot
Disallow: /

User-agent: PerplexityBot
Disallow: /

User-agent: Bytespider
Disallow: /

User-agent: Applebot-Extended
Disallow: /
```

# Learn more

For more information on the User-agent strings for common AI tools, check out these resources:

- Chris Lever's User Agent String Database & Bot Directory: Includes quick filters for AI Crawlers versus search and other crawlers with detailed entries explaining whether the crawler respects robots.txts directives and descriptions of what
- SearchEngineJournal's Complete Crawler List for AI User-Agents [December 2025]
- LLMS Central's Complete Guide to AI Bot User-Agents, which has some quick copy-paste examples for robots.txt

Copyright © 2015 – 2025 Silly Moose, LLC. All rights reserved.
Powered by KnowledgeOwl knowledge base software.

Copyright © 2015 – 2025 Silly Moose, LLC. All rights reserved.
Powered by KnowledgeOwl knowledge base software.