Track robots.txt

The **Robots.txt** tab in AI Crawl Control provide insights into how AI crawlers interact with your `robots.txt` files across your hostnames. You can monitor request patterns, verify file availability, and identify crawlers that violate your directives.

To access robots.txt insights:

1. Log in to the Cloudflare dashboard ↗, and select your account and domain.
2. Go to **AI Crawl Control**.

   Go to **AI Crawl Control**
3. Go to the **Robots.txt** tab.

Check managed robots.txt status

The status card at the top of the tab shows whether Cloudflare is managing your `robots.txt` file.

When enabled, Cloudflare will include directives to block common AI crawlers used for training and include its Content Signals Policy in your `robots.txt`. For more details on how Cloudflare manages your `robots.txt` file, refer to Managed `robots.txt`.

Filter robots.txt request data

You can apply filters at the top of the tab to narrow your analysis of robots.txt requests:

- Filter by specific crawler name (for example, Googlebot or specific AI bots).
- Filter by the entity running the crawler to understand direct licensing opportunities or existing agreements.
- Filter by general use cases (for example, AI training, general search, or AI assistant).
- Select a custom time frame for historical analysis.

The values in all tables and metrics will update according to your filters.

Monitor robots.txt availability

The **Availability** table shows the historical request frequency and health status of `robots.txt` files across your hostnames over the selected time frame.

Column	Description
Path	The specific hostname's `robots.txt` file being requested. Paths are listed from the most requested to the least.
Requests	The total number of requests made to this path. Requests are broken down into: - **Successful:** HTTP status codes below 400 (including **200 OK** and redirects). - **Unsuccessful:** HTTP status codes of 400 or above.
Status	The HTTP status code from pinging the `robots.txt` file.
Content	An indicator showing whether the `robots.txt` file contains Content

Signals	Signals ↗, directives for usage in AI training, search, or AI input.

From this table, you can take the following actions:

- Monitor for a high number of unsuccessful requests, which suggests that crawlers are having trouble accessing your `robots.txt` file.
  - If the **Status** is `404 Not Found`, create a `robots.txt` file to provide clear directives.
  - If the file exists, check for upstream WAF rules or other security settings that may be blocking access.
- If the **Content Signals** column indicates that signals are missing, add them to your `robots.txt` file. You can do this by following the Content Signals ↗ instructions or by enabling Managed `robots.txt` to have Cloudflare manage them for you.

Track robots.txt violations

The **Violations** table identifies AI crawlers that have requested paths explicitly disallowed by your `robots.txt` file. This helps you identify non-compliant crawlers and take appropriate action.

How violations are calculated

The Violations table identifies mismatches between your **current** `robots.txt` directives and past crawler requests. Because violations are not logged in real-time, recently added or changed rules may cause previously legitimate requests to be flagged as violations.

For example, if you add a new `Disallow` rule, all past requests to that path will appear as violations, even though they were not violations at the time of the request.

Column	Description
Crawler	The name of the bot that violated your `robots.txt` directives. The operator of the crawler is listed directly beneath the crawler name.
Path	The specific URL or path the crawler attempted to access that was disallowed by your `robots.txt` file.
Directive	The exact line from your `robots.txt` file that disallowed access to the path.
Violations	The count of HTTP requests made to the disallowed path/directive pair within the selected time frame.

When you identify crawlers violating your `robots.txt` directives, you have several options:

- Navigate to the **Crawlers** tab to permanently block the non-compliant crawler.
- Use Cloudflare WAF to create a path-specific security rules for the violating crawler.
- Use Redirect Rules to guide violating crawlers to an appropriate area of your site.

Related resources

- Manage AI crawlers
- Analyze AI traffic
- Cloudflare WAF

Previous
← Manage AI crawlers

Next
What is Pay Per Crawl? →

Last updated: Oct 24, 2025