

Can AI Access Your Website?

Your website may contain clear, useful information. However, AI discovery systems still need permission to collect it.

This exercise answers one question:

Does your robots.txt file allow major search and AI discovery crawlers to access your important public pages?

It checks the instructions your website publishes for crawlers.

Two terms used in this exercise

- **Crawler:** An automated visitor used by a search engine or AI provider to find and read webpages.
- **robots.txt:** A public file that tells crawlers which parts of a website they may or may not access.

What you need

- Your website address.
- The addresses of up to two other important public pages.
- A web browser.
- ChatGPT, Claude or another AI assistant. The best here is an assistant like Claude Co-work or OpenAI's Codex
- Approximately five minutes.

Start a new conversation so that previous chat history does not influence the assessment.

Step 1 — Choose three important pages

Choose pages that contain information you would want an AI system to find.

Homepage

Main experience, product or service page

Booking, enquiry or practical information page

Step 2 — Copy your robots.txt file

Open a web browser and add /robots.txt to the end of your website's main address.

For example:

<https://www.yourwebsite.co.nz/robots.txt>

If your website is <https://www.newzealand.com>, this would be <https://www.newzealand.com/robots.txt>

If the file opens, copy all of the text it contains.

If the file cannot be found, write NO ROBOTS.TXT FILE FOUND in the prompt (outlined in the next section).

The AI will explain what that means.

Step 3 — Run the assessment

Replace the placeholders below and paste the complete contents of your robots.txt file into the prompt.

Read the robots.txt instructions below and explain whether major search and AI discovery crawlers are allowed to access my three important public pages.

You do not need to browse my website. This is a robots.txt permission check only.

Website: [PASTE WEBSITE ADDRESS]

Pages to check:

1. [PASTE HOMEPAGE ADDRESS]
2. [PASTE IMPORTANT EXPERIENCE, PRODUCT OR SERVICE PAGE]
3. [PASTE BOOKING, ENQUIRY OR PRACTICAL INFORMATION PAGE]

Here is the complete content of my robots.txt file:

--- START OF ROBOTS.TXT ---

[PASTE ROBOTS.TXT CONTENT HERE]

--- END OF ROBOTS.TXT ---

Check these crawlers:

- Googlebot — Google Search and Google's AI search features
- bingbot — Bing Search and Microsoft Copilot
- OAI-SearchBot — ChatGPT search
- Claude-SearchBot — Claude search
- PerplexityBot — Perplexity search
- Meta-WebIndexer — Meta AI search
- Applebot — Apple search experiences
- Amzn-SearchBot — Amazon and Alexa search experiences

For each crawler:

1. Look for instructions written specifically for that crawler.
2. If there are no crawler-specific instructions, apply the rules under User-agent: *.
3. Check the path of each of my three pages against the relevant Allow and Disallow rules.
4. Apply the most specific matching path rule. If equally specific Allow and Disallow rules conflict, treat Allow as taking precedence.
5. Do not treat the absence of a crawler's name as a block if the wildcard rules allow access.

Use these result labels:

- ALLOWED — robots.txt does not block this crawler from the page.
- BLOCKED — robots.txt contains a rule that blocks this crawler from the page.
- UNCLEAR — the supplied file is incomplete, conflicting or cannot be interpreted reliably.

Use this exact output structure:

1. OVERALL RESULT

Choose one:

- ALLOWED BY ROBOTS.TXT — all eight crawlers are allowed to access all three page paths.
- BLOCKED BY ROBOTS.TXT — at least one crawler is blocked from at least one page path.
- TECHNICAL CHECK NEEDED — the result cannot be determined from the supplied information.

Explain the result in no more than three plain-English sentences.

2. CRAWLER RESULTS

Create one table with these columns:

Search or AI product | Crawler | Homepage | Main offering page | Booking or information page | Rule applied

Use only Allowed, Blocked or Unclear in the three page columns. Quote the exact robots.txt instruction in the final column.

3. WHAT TO DO NEXT

Give me no more than three priority actions. If nothing is blocked, say that no robots.txt change is required based on these pages.

4. MESSAGE FOR MY WEBSITE PROVIDER

Only if something is blocked or unclear, write a short message I can copy and send to my website provider. Name the affected crawler, page path and robots.txt rule, and ask them to review it.

Write for a non-technical business owner. Keep the entire answer concise.

Step 4 — Record your result

Overall result

- ☐ Allowed by robots.txt
- ☐ Blocked by robots.txt
- ☐ Technical check needed

Is any crawler blocked?

Which page or section is affected?

Which robots.txt rule causes the block?

What needs to go to your website provider?

Step 5 — Choose your next action

If all eight crawlers are allowed

No robots.txt change is required for the three pages checked. Continue to the separate page-access exercise to test whether AI products can open and cite them.

If a crawler is blocked

Send the generated message to the person or company responsible for your website. Ask them to confirm whether the block is intentional before changing it.

If the result is unclear

Ask your website provider to review the complete robots.txt file and confirm which of the eight crawlers can access the three page paths.

Retest date

The objective is simple: make sure your website is not telling important search and AI discovery crawlers to stay out.

Crawlers checked by this exercise

Crawler names and provider policies can change. These are up to date as of the 10th of September 2026.

Product or ecosystem	Crawler
Google Search and Google AI features	Googlebot
Bing and Microsoft Copilot	bingbot
ChatGPT search	OAI-SearchBot
Claude search	Claude-SearchBot
Perplexity	PerplexityBot
Meta AI	Meta-WebIndexer
Apple search experiences	Applebot
Amazon and Alexa search experiences	Amzn-SearchBot

Official sources

- [IETF — Robots Exclusion Protocol](#)
- [OpenAI — Overview of OpenAI Crawlers](#)
- [Google — AI features and your website](#)
- [Microsoft Bing — Webmaster Guidelines](#)
- [Anthropic — Web crawling and site-owner controls](#)
- [Perplexity — Crawlers](#)
- [Meta — Web crawlers](#)
- [Apple — About Applebot](#)
- [Amazon — About Amazonbot](#)

