Robots.txt checklist for AI crawlers

Go through this list once for your own site. You write down your choice per type of bot and put the rules in your robots.txt. Allow 15 minutes. Status as of 10 November 2026.

1. Your choice per type of bot

Type	Bots (name in robots.txt)	Chosen
Training	GPTBot, ClaudeBot, Google-Extended (also controls grounding in Gemini Apps)	☐
Search	OAI-SearchBot, Claude-SearchBot, PerplexityBot	☐
At a user's request	ChatGPT-User, Claude-User, Perplexity-User (robots.txt does not always apply here)	☐

2. The check in five steps

Step	What you do	Done
1	Open yourdomain.com/robots.txt and read what is there	☐
2	Write down your choice per type of bot, one sentence per type	☐
3	Put the rules in your file, with the bot name exactly as the provider spells it	☐
4	Check your firewall or CDN: does it block bots you have just allowed?	☐
5	Put 1 February 2027 and 1 March 2027 in your diary	☐

3. What changes when (providers' own pages, 10 November 2026)

Provider	Position	Done
Anthropic	Reads the shared signal since 30 October 2026. A ClaudeBot rule takes priority.	☐
Google	Planned for 2027. A Google-Extended rule takes priority.	☐
Microsoft	From 1 February 2027. After 1 March 2027 NoArchive and NoCache no longer count.	☐
OpenAI	The token is not on the bot page I read.	☐

4. Starter file (you fill in the choice)

```
# Search: allow
User-agent: OAI-SearchBot
Allow: /

User-agent: PerplexityBot
Allow: /

# Training: your choice
User-agent: GPTBot
Disallow: /

# Shared signal, not yet active at every provider
User-Agent: __usage_ai-training
Disallow: /
```

This starter file switches training off. If you want to allow training, remove the lines for GPTBot and __usage_ai-training.

What robots.txt does not do for you
It is a request to the bot. For bots that a user sends to fetch a page, it may have no effect.
The shared signal is an IETF draft and not yet a standard. It controls the use of your content, not crawling.
A group with User-agent: * does not apply to the token __usage_ai-training. Write the group yourself.

Sources: Google Search Central Blog (30 October 2026), Google's common crawlers (30 October 2026), IETF draft-kelley-aipref-training-preference-signal-00, Microsoft Bing Blog (30 October 2026), OpenAI Overview of OpenAI Crawlers, Anthropic Claude Help Center, Perplexity Crawlers. Consulted on 10 November 2026.