

AI crawler checklist

Which AI agents to let into your website, who runs them, and what to check so the ones you want can read and use your pages.

1. DECIDE BY KIND OF AGENT

AI assistants Allow Fetches pages for a person, in real time, while they use an AI assistant.	AI and search crawlers Allow Indexes pages so a search engine or AI search tool can find, cite and link to them.	Autonomous agents Allow and test Uses a website like a person would, clicking, typing and filling in forms to finish a task for someone.	AI training crawlers Your call Collects public pages that may be used to train AI models.
---	--	--	---

2. WHO RUNS WHAT: THE NAMES TO USE IN ROBOTS.TXT

COMPANY	AI ASSISTANTS	AI AND SEARCH CRAWLERS	AUTONOMOUS AGENTS	AI TRAINING CRAWLERS
Google	Gemini-Deep-Research, Google-NotebookLM, GoogleAgent-URLContext	Google-InspectionTool ✓, Googlebot ✓, Storebot-Google ✓	Google-Gemini-CLI, GoogleAgent-Mariner	Google-CloudVertexBot, Google-Extended, GoogleOther ✓
Anthropic	Claude-User	Claude-SearchBot	Claude-Code, Claude-Web	anthropic-ai, ClaudeBot
Amazon	Amzn-User	Amazonbot ✓	AmazonBuyForMe, NovaAct	-
OpenAI	ChatGPT-User ✓	OAI-SearchBot ✓	ChatGPT Agent	GPTBot ✓
Mistral AI	MistralAI-User	MistralAI-Index	-	MistralAI-Training
Moonshot AI	Kimi-User	-	Kimi-Agent	KimiBot
Apple	-	Applebot ✓	-	Applebot-Extended
Baidu	-	Baiduspider ✓	-	ERNIEBot
ByteDance	DoubaoBot	-	-	Bytespider
DuckDuckGo	DuckAssistBot	DuckDuckBot	-	-
Meta	meta-externalfetcher	-	-	meta-externalagent
Perplexity	Perplexity-User ✓	PerplexityBot ✓	-	-
AI2	-	-	-	AI2Bot
Alibaba	-	-	-	QwenBot
Cognition	-	-	Devin	-
Cohere	-	-	-	cohere-training-data-crawler
Common Crawl	-	-	-	CCBot
DeepSeek	-	-	-	DeepSeekBot
Diffbot	-	-	-	Diffbot
Exa	-	ExaSearchBot	-	-
Kagi	kagi-fetcher	-	-	-
Manus	-	-	Manus-User	-
Microsoft	-	bingbot ✓	-	-
Phind	PhindBot	-	-	-
Yandex	-	YandexBot ✓	-	-

✓ The operator publishes IP ranges or a reverse DNS domain, so you can prove a request really came from it. Anyone can copy the others' names.

3. CHECK THESE

- ☐ robots.txt lets AI assistants and search agents in
- ☐ Bot protection lets AI agents through
- ☐ Verify an agent by its IP address or reverse DNS before you trust its name. ghostagentlab.com/tools/verify-bot/
- ☐ No CAPTCHA or challenge on arrival
- ☐ AI agents can open your product, pricing and cart pages
- ☐ Content loads without JavaScript ghostagentlab.com/tools/ai-view/
- ☐ llms.txt guide for AI ghostagentlab.com/tools/llms-txt/
- ☐ Measure which agents visit, how often, and what they read, from your server or CDN logs. ghostagentlab.com/platform/#observe

4. A ROBOTS.TXT TO START FROM

```
# Everyone, including AI assistants and AI search
User-agent: *
Allow: /
Disallow: /cart
Disallow: /checkout

# Only if you've decided against AI training
User-agent: GPTBot
User-agent: ClaudeBot
User-agent: Google-Extended
User-agent: CCBot
Disallow: /
```

Test yours against every agent: ghostagentlab.com/tools/robots-txt/