AI Search Ranker

List of AI Crawlers

AI Web Crawler Directory

There are quite a few companies and organizations that have these robots running 24/7, scraping the internet for training with AI models.

A "web crawler" visits websites and catalogs all the info it can find.

Become familiar with these robots because <u>what these bots seek</u> is the key to ranking your site, or that of a client's...

Noteworthy AI crawlers:

GoogleOther (Google):

Google uses this for AI-related indexing beyond standard search, supporting models like Gemini.

GPTBot (OpenAI):

OpenAI uses this crawler for gathering public text data used to train models like ChatGPT and GPT-4o.

ClaudeBot (Anthropic):

Anthropic uses this general-purpose crawler for training their Claude AI assistant.

Bytespider (ByteDance):

ByteDance (the parent company of TikTok) operates this crawler, which is likely used to train LLMs for TikTok's search function and other AI applications.

Amazonbot (Amazon):

Amazon uses this to gather data that improves services like Alexa.

PerplexityBot (Perplexity):

Perplexity uses this crawler for indexing websites to build their AI-driven search engine.

OAI-SearchBot (OpenAI):

OpenAI uses this crawler specifically for search functions, linking and surfacing websites in their new search engine.

DuckAssistBot (DuckDuckGo):

DuckDuckGo uses this to scrape fact-based snippets for their private AI answer feature, DuckAssist.

Meta-ExternalAgent (Meta):

Meta likely uses this for collecting data to train or fine-tune its LLMs.

Applebot (Apple):

Apple's web crawler for Siri and Spotlight search, with potential use in AI development.

CCBot (Common Crawl):

A non-profit organization operates this crawler that builds an open repository of web data for researchers and developers, potentially used for AI training.

Cohere-ai (Cohere):

Cohere uses this to collect textual data for its language models.