

# AI/LLM-Powered Scrapers

## AI/LLM-Powered Scrapers

### ScrapeGraphAI (MIT / Free)

Describe what you want in plain English, it extracts structured JSON without CSS/XPath selectors. Auto-adapts to layout changes. Supports GPT, Gemini, Groq, Azure, Hugging Face, local Ollama models.

[GitHub](#)

### Crawl4AI (58K Stars, Apache 2.0)

Open-source LLM-friendly web crawler. Outputs clean Markdown for RAG/agents. Heuristic noise filtering, CSS/XPath/LLM extraction. Local-first (no API costs).

[GitHub](#)

### Firecrawl

Developer API that outputs clean Markdown from any URL, optimized for LLM ingestion. Used in enrichment pipelines alongside Maps scrapers.

[firecrawl.dev](https://firecrawl.dev)

## LLM Query Expansion

Not a scraper per se, but a technique: use LLMs to generate category synonyms and related search terms to multiply coverage per geographic area. "dentist" becomes "dental clinic", "oral surgeon", "dental practice", etc. Combined with any scraper for 3-5x more results.

---

Revision #1

Created 2026-06-05 23:48:51 UTC by Ben Adrian Sarmiento

Updated 2026-06-05 23:48:51 UTC by Ben Adrian Sarmiento