

AI-Powered Scrapers & Alternative POI Sources

ScrapeGraphAI, Crawl4AI, Overture Maps, AllThePlaces, OSM

- [AI/LLM-Powered Scrapers](#)
- [Alternative POI Data Sources \(No Scraping\)](#)

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

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.

Alternative POI Data Sources (No Scraping)

Alternative POI Data Sources

Free or commercial POI databases that don't require scraping Google Maps.

Overture Maps Foundation (64.8M POIs Free)

Backed by Meta, Microsoft, Amazon, TomTom, Foursquare. The strongest free alternative to Google Maps data.

Size	64.8M places — Meta (~59.2M), Foursquare (~6.7M), Microsoft (~7.4M), AllThePlaces (~1.7M)
Fields	Names, categories (64+), phones, emails, websites, socials, addresses, brand, operating_status, confidence score, coordinates
Format	GeoParquet on S3 and Azure Blob. Python CLI, DuckDB SQL, browser Explorer
License	CDLA-Permissive-2.0 / ODbL (commercial use OK)
Limits	Monthly updates (not real-time). No reviews/photos. Thinner outside Western countries

Warning: Reddit reports Places layer stopped updating as of Sept 2024 release. Verify current state.

[Overture Places Guide](#) | [Downloads](#)

Overture-Based API (Community-Built, 200x Cheaper)

A developer built a Places API using Overture data + Rust/Axum + PostGIS. Free 5K/mo, \$10/100K, \$30/500K, \$80/2M — vs Google's ~\$1,700 for 100K.

Other Alternative Sources

Source	POIs	Free Tier	Key Strength	Reviews?
OpenStreetMap (Overpass API)	Varies	Unlimited, no key	Free, query any tag combo. Overpass Turbo	No
Foursquare Places	100M+	Commercial	Richest venue data, behavioral insights, check-ins	Tips only
HERE Technologies	Global, 400+ cats	250K tx/mo	TripAdvisor ratings, EV/fuel data, chain ID	Via TripAdvisor
TomTom	~100M, 180+ countries	50K daily tx	Navigation-optimized, relevance scoring	No
Mapbox	Global	100K req/mo	Polished SDKs	No
Geoapify	OSM-based, 400+ cats	3K credits/day	Transparent pricing, can cache/store	No
Yelp Fusion API	Millions	5K calls/day	3 reviews max via API. Open Dataset: 8.6M reviews (academic)	3 (API), 8.6M (dataset)
AllThePlaces	20M+	Unlimited (CC-0)	4,100+ Scrapy spiders, weekly updates	No
MapQuest	Various	Limited	Search API v5: radius/rect/polygon/corridor	No
Nominatim	OSM geocoding	Free	Forward/reverse geocoding	No
Photon (Komoot)	OSM geocoding	Free	Typo-tolerant, multilingual	No

OSM Reality Check (from Reddit): OpenStreetMap is poor for business/POI data. Business coverage is ~75% at best, biased toward bigger/popular locations.

Data Providers & Marketplaces

Provider	Dataset	Pricing	Google Maps?
Bright Data	200.7M+ records	\$0.0025/record	Yes
Datarade	60M+ US (varies)	By provider	Yes

Veridion	134M+ businesses	~\$99/user/mo	Partial
SafeGraph	52M+ POIs	Commercial	Partial (foot traffic)
Dataplор	Various	Commercial	Partial (LatAm strong)
Xtract.io	6M+ locations	Commercial	Partial
Coresignal	N/A	N/A	No (employee data only)
Data.world	N/A	N/A	No (data governance)

Export & Data Liberation Tools

- **Google Takeout:** Export saved/starred places (CSV/JSON, no coordinates — needs geocoding)
- [Takeout Tools](#): Adds coordinates, converts to GeoJSON/KML/GPX
- [json2kml](#): Python converter for saved places to KML
- [Export-Google-Maps-Saved-Places](#)
- **Firefox GPX Exporter:** Export saved lists as GPX with lat/lon