

Caprae Capital Lead Enrichment Tool

Project Overview

The goal of this project was to build a web-based lead enrichment tool for Caprae Capital Partners. Inspired by tools like Cohesive AI's company scraper, the tool aims to help investors quickly evaluate companies using publicly available data. It extracts structured intelligence from company websites and LinkedIn profiles, focusing on metrics relevant to Caprae's investment thesis.

Approach

We developed a browser-based application using Streamlit. The tool performs the following:

- Scrapes textual data from a company's website and LinkedIn (if provided)
- Passes the text through a large language model (LLM) via the OpenRouter API
- Extracts structured data such as executive info, funding stage, business model, competitors, and more
- Estimates revenue using a separate LLM prompt
- Displays the results and allows download as JSON or CSV

Model Selection

We used the Swallow LLaMA 3.1 70B model from OpenRouter. It was chosen for its balance of performance and efficiency in generating structured, high-quality output. Two prompts were used: one for general company enrichment and one for revenue estimation.

Data Preprocessing

The scraper collects and cleans content from the website and LinkedIn page using BeautifulSoup. It removes scripts, styles, and other non-relevant elements. The cleaned text is then sent to the LLM. If the JSON structure is invalid, we retry parsing up to three times.

Outputs

The model returns a structured JSON object including:

- Fit with Caprae's investment thesis
- Key executive name and title
- Executive email (if available)
- Business model and confidence level
- Funding stage and confidence level
- Estimated revenue and confidence level
- Key qualitative insights
- List of competitors

Final Product

The final MVP is a user-friendly, local web application that allows Caprae to input a URL and receive an enriched company profile in under 15 seconds. Features include:

- Custom-branded UI with clean layout
- Streamlined enrichment workflow
- Instant export to CSV or JSON
- Compatible with CRM platforms for deal tracking

Outcome

The project successfully delivered a working MVP that reduces manual research and enables faster evaluation of inbound leads. It automates lead qualification using LLMs and lays the foundation for future CRM or automation features. All scraping is done ethically from publicly available pages.