Caprae Capital Lead Enrichment Tool

1-Page Project Report

Overview

Caprae Capital needed a fast, intelligent way to qualify startup leads. Inspired by Cohesive AI's company scraper, this browser-based tool enriches companies by extracting structured insights from their website and LinkedIn—aligned to Caprae's SaaS investment thesis.

Approach

The tool was built with Streamlit and performs:

- Website + LinkedIn scraping using BeautifulSoup
- Cleaning + trimming text content
- Sending content to LLM via OpenRouter API
- Extracting structured data (executives, model, stage, revenue)
- Exporting results to JSON/CSV

Model Used

The model selected was `tokyotech-llm/llama-3.1-swallow-70b-instruct-v0.3`, known for strong JSON formatting and performance. Two specialized prompts were used: one for enrichment, one for revenue estimation.

Data Processing

HTML is parsed and stripped of scripts/styles. Text is truncated and validated before LLM calls. Invalid outputs trigger retries (up to 3).

Performance

- Runtime: under 15 seconds end-to-end
- Accurate executive, funding, model detection
- High-quality structured JSON returned consistently

Challenges

- LLM hallucinations required prompt tuning
- LinkedIn blocking and bot detection
- Retry logic for broken/incomplete outputs
- Session state bugs in Streamlit during dev

Walkthrough

User inputs a URL → scraper extracts text → text sent to LLM → JSON parsed → insights displayed → downloadable as JSON/CSV.

Outcome

Delivered a fast, intuitive enrichment MVP:

- Reduces manual research significantly

- Custom-branded UI

- Structured outputs compatible with CRM

- Fully local and privacy-preserving

Demo video: https://youtu.be/Ya0G8pMd1GM