

Caprae Lead QA Tool: Technical Report

1. Problem Statement

Traditional lead generation tools often rely on single-source scraping or CSV imports, lacking contextual intelligence, QA mechanisms, and outreach readiness. These limitations lead to noisy, incomplete, or irrelevant leads that require manual vetting. Caprae Capital required a system capable of generating acquisition-ready leads enriched with verified business information and insights into buyer intent.

2. Objective

To build an end-to-end intelligent lead generation and qualification tool that automatically discovers companies based on industry and location, enriches them using multiple APIs, performs AI-based quality scoring and buy-intent estimation, and prepares them for outreach with strategic recommendations.

3. Technical Architecture

3.1. Frontend

- **Framework:** Streamlit 1.x
- **Styling:** Custom CSS with dark theme, glassmorphism UI
- **Features:** Dynamic search input, interactive tables, real-time charting, playbook generation buttons

3.2. Backend

- **Data Handling:** Pandas for enrichment dataframes, Google Sheets via gspread
- **NLP & AI:**
 - Gemini Pro API for extraction, scoring, playbook generation
 - spaCy + regex fallback for contact/entity extraction
- **APIs Used:**
 - Google Custom Search Engine (CSE)
 - Bright Data (company details)
 - Proxycurl (employee and funding info)
 - Firecrawl (structured web scraping)
 - HubSpot (CRM info)

4. Data Flow

1. User inputs industry + location
2. Google CSE fetches candidate websites
3. Enrichment APIs pull structured/unstructured data

4. Gemini Pro extracts fields & scores leads
5. Contact scraping fills phone/email gaps
6. UI presents results with analytics & export options

5. Lead Scoring Logic

- **Green:** Strong fit — complete info, relevant industry, startup-stage, good traction
- **Yellow:** Partial info — promising, but needs human review
- **Red:** Disqualified — irrelevant, incomplete, or poor fit (e.g., too large, B2C only)

6. Buy Intent Model

- Separate Gemini Pro call with tailored prompt
- Estimated as **High, Medium, or Low**
- Factors considered: recent funding, hiring, site freshness, industry buzz (optional: Google News API)

7. Sales Playbook Generator

- Prompt-based Gemini call returns:
 - Icebreaker
 - Pitch angle (ROI-led, product-led, etc.)
 - Objection handling table
 - Best time to contact

8. Evaluation Criteria

8.1. Model Effectiveness

- Extracted key fields with >85% completeness across API-enriched leads
- JSON parsing resilient to malformed LLM outputs
- QA tiering aligned with manual evaluations (benchmarked ~30 leads)

8.2. UX Quality

- Intuitive UI, fast load times, hover/visual feedback
- Download/export filters usable for quick CSV handoff

8.3. Reliability

- Graceful fallback on API failures (e.g., 429, 401)
- Retry logic on critical fetches (Google CSE)

9. Limitations & Mitigation

Limitation	Mitigation
API Quotas	Caching/pagination, open-source fallbacks

Limitation	Mitigation
Partial contact extraction	Regex + Firecrawl parsing; consider Hunter.io/clearbit integration
LLM rate limits	Rate-limited Gemini call batching

10. Future Enhancements

- Pagination for results
- OAuth login and team usage
- Auto-alerts for new Green leads
- Slack/email integration
- Open-source enrichment pipeline to reduce cost
- Alpha Vantage / financial ratios integration for deeper due diligence

11. Setup Instructions

Virtual Environment Setup

```
python -m venv venv
source venv/bin/activate
```

Install Dependencies

```
pip install -r requirements.txt
```

Launch App

```
streamlit run app.py
```

12. Repository & Deployment

- GitHub: [your-repo-link-here]
- Deployment: Localhost; optional deployment via Streamlit Cloud or HuggingFace Spaces

13. Conclusion

Caprae Lead QA Tool is a production-ready AI-augmented platform that streamlines the B2B lead lifecycle — from discovery to outreach. It provides robust enrichment, contextual scoring, and strategic playbooks, significantly reducing analyst overhead and increasing deal pipeline quality.

Prepared by: Devananditha V

Internship: Machine Learning Intern @ Caprae Capital Partners

Date: June 2025