

ISAAC ORESANYA

APPLIED AI ENGINEER

Ibadan, Nigeria | +234 810 742 2878 | isaacjoresanya@gmail.com | Available immediately
isaacjoresanya.com | github.com/TheDataIsaac | linkedin.com/in/isaac-oresanya

SUMMARY

Applied AI engineer who builds retrieval and structured-extraction applications in Python, with source-quote validation. Three public projects: a deployed newsletter assistant, an extraction pipeline run over 8,595 job postings, and a hybrid-search RAG app. Over two years of reporting and data-pipeline work behind them.

SKILLS

LLM applications: OpenAI API with structured outputs, RAG, LangGraph, prompt design, quote checks

Retrieval: BM25, sentence embeddings, reciprocal rank fusion, Chroma, NumPy

Backend: Python, FastAPI, Pydantic, asyncio, SQLite, Docker, Git, Streamlit

Data: pandas, SQL, ETL, data cleaning, duplicate audits, Scrapy, Playwright

INDEPENDENT PROJECTS

What Would Isaac Say? | Python, LangGraph, FastAPI, Chroma, Docker github.com/TheDataIsaac/what-would-isaac-say

- Deployed a public assistant in Docker on a cloud server behind HTTPS. It answers from 60 newsletter posts and reviews a reader's draft against the advice in them.
- Tested it on 88 questions. It answered 58 of the 60 the newsletter covers and refused 12 of 12 unrelated ones. A LangGraph relevance check raised refusals of 10 tricky near-topic questions from 5 to 9.
- Added a Python check that every quote the assistant shows appears in the source text.

Job posting extraction and Q&A | Python, OpenAI API, Pydantic, FastAPI github.com/TheDataIsaac/job-extraction-pipeline

- Ran typed extraction over 8,595 freelance postings, deduplicated from 9,193 rows. Requests run 10 at a time with retries, and each result is saved as it arrives so a stopped run resumes.
- Required a supporting quote for budget, deadline, experience level and red flags, and let the model return null instead of guessing. 97.1% of records passed the automated quote and evidence checks.
- Added a FastAPI endpoint and embedding search that answers questions from retrieved postings.

Hybrid search for freeCodeCamp lessons | Python, NumPy github.com/TheDataIsaac/hybrid-search

- Wrote BM25 and reciprocal rank fusion from scratch and combined them with local sentence embeddings, so both exact-term and similar-meaning questions find the right lesson.
- Indexed 144 lessons as 698 chunks, downloaded at a pinned commit so every run uses the same text.
- Added RAG answers that cite the lessons they came from and link to each one.

PROFESSIONAL EXPERIENCE

Reporting & Operations Analyst | Data Career Jumpstart Jan 2024 to present

- Own weekly reporting across 70+ business metrics. Cut preparation from about eight hours to five and raised on-time delivery from about 85% to 98% over 50+ cycles.
- Cut post-delivery corrections 40% by validating source data before each report goes out.
- Built self-service Streamlit views that cut turnaround on common performance questions from one working day to under 30 minutes.

Data Analyst, Contract | Novae, LLC Jun 2024 to Jan 2025

- Built Scrapy and Playwright scrapers for three platforms and merged two collection runs into 163,229 records. Prepared 141,228 analysis-ready records with pandas.
- Found 60,784 duplicate rows, 43.6%, in the first collection of 139,554 and cut it to 78,770 records. Added duplicate removal to each platform's processing step..

EDUCATION

B.Sc. Physics | Federal University of Agriculture, Abeokuta | Jan 2021 to Sep 2024