SHAURYA CHANDNA

+1 (226) 749-0721 | shauryachandna13@gmail.com | LinkedIn | GitHub

Professional Experience

Ontario Ministry of Energy and Electrification **Toronto, ON**
Data Analyst Intern January - April 2025

- Built a centralized data automation pipeline using Python and Pandas to extract, clean, and enrich energy datasets for Ontario buildings via REST APIs, using dynamic lookup calls and automating geospatial visualization in ArcGIS.
- Automated 40% of all company maps and dashboards workflows, decreasing manual cleaning time by 97%..

Innova Solutions **Atlanta, Georgia**
Data Science Intern May - August 2024

- Developed a hybrid time series forecasting system to optimize warehouse inventory for an energy-sector client, combining Croston's Method and ARIMA with a decision tree selector trained on demand sparsity and variance.
- Reduced RMSE by 4% on usage prediction for a SaaS product by engineering temporal embeddings, applying Bayesian hyperparameter tuning on XGBoost, and interpreting model behavior with SHAP analysis.

ElevonData **New York, NY**
Data Analyst Intern May - August 2023

- Developed end-to-end ETL pipelines for a telecom cost-reduction initiative, utilizing Airflow for scheduling, Talend DI for data transformation, and SQL for efficiently storing and querying structured data.
- Designed advanced Power BI reports with custom DAX measures, interactive visualizations, and data modeling to track contract usage and network spending, driving a 7.3% reduction in project costs.

Education

University of Waterloo **Waterloo, ON**
Bachelor of Mathematics, Computational Math (Honors) Expected Graduation 2027
Coursework: Applied Linear Models, Computational Statistics & Data Analysis, Logic & Computation, Intro to Optimization, Information Theory and Applications, Intro to Computational Mathematics, Computer Organization and Design
Scholarships: Faculty of Mathematics Entrance Scholarship, President's Scholarship of Distinction (1/200 students awarded)

Projects

QuizMaster: RAG Powered Document Quiz Generator – *Python, FastAPI, OpenAI API, Pinecone, React, Langchain*

- Architected a Retrieval Augmented Generation (RAG) system to transform resource documents into interactive quizzes, enabling natural language querying over educational content with high semantic accuracy.
- Engineered a document preprocessing pipeline, employing semantic chunking.
- Utilized **OpenAI's text-embedding-3-small** model to generate high-dimensional embeddings, fine-tuned the embedding and retrieval pipeline using domain-specific educational datasets.
- Constructed and optimized a **Pinecone** vector database, implementing indexing strategies such as quantization and dimensionality reduction to improve similarity search time-complexity by 40%.
- Developed a hybrid cosine similarity search mechanism, augmented with metadata filtering and query expansion, to retrieve top-k relevant document chunks with enhanced precision and recall metrics.
- Integrated **LangChain** to orchestrate the end-to-end workflow, chaining embedding generation, vector storage, and language model inference for seamless quiz generation from retrieved content.
- Designed a high-performance **FastAPI** REST API to handle real-time quiz generation requests, paired with a **React**-based client interactive layer for intuitive PDF uploads, query inputs, and quiz delivery.

Ads for AI Agents – *Python, QDrant, Langchain, SBERT*

- Selected for and demoed initial stages of a startup with top investors from comma.vc, Sequoia and a16z with 2 classmates.
- Built a RAG-powered model to recommend user personalized and web host-based advertisements, for a relevant and seamless advertisement experience, for hosting on AI agents.
- Optimized **Qdrant** vector database performance through advanced techniques, including hierarchical navigable small world (HNSW**)** indexing for sub-linear similarity search, and payload-based filtering to enhance query efficiency.
- Fine-tuned OpenAI's **text-embedding-3-small** model using contrastive learning and triplet loss on curated ad-domain datasets, incorporating gradient-based optimization and knowledge distillation to enhance embedding quality.

Skills

Languages/Frameworks: Python, SQL, C/C++, React, PyTorch, Tensorflow, FastAPI, Streamlit
Machine Learning Libraries: scikit-learn, matplotlib, pandas, numpy, scipy, keras
Deep Learning: NN, RNN, LSTM, CNN, Regularization