# Mosh Oluwashina

moluwashina@gmail.com | LinkedIn | GitHub

## EDUCATION

**Carleton University** Conferred June 2026

*Bachelor of Computer Science (Honours), Artificial Intelligence and Machine Learning stream*

- CGPA 3.0/4.0. Entrance Scholarship 2021. Grades: Database Management Systems A+, Neural Networks A, Generative AI and LLMs A-.

## PROJECTS

**Multi Agent Customer Support System** 2026, personal project

*Python, OpenAI API, Prompt Engineering, FastAPI, SQLite, Pinecone, Streamlit, Docker, GitHub Actions*

- Built a multi agent support system in which a router sends each question to an orders agent or a read only policy agent. The policy agent has no refund tool, so it cannot issue refunds.
- Enforced refund rules in code rather than prompts. Only delivered orders qualify, and refunds over $100 go to a SQLite approval queue, with duplicate refunds and duplicate requests blocked.
- Built a RAG pipeline that embeds 11 handbook chunks with OpenAI embeddings and retrieves the top 5 matches from Pinecone.
- Evaluated routing on 20 labelled questions, improving accuracy from 90% to 95% through prompt engineering of the classifier.

**Portfolio Website with Resume RAG Chatbot** 2026, personal project | oluwashina.ca

*Python, FastAPI, LangChain, Pinecone, OpenAI, Prompt Engineering, Docker, Google Cloud Run*

- Built a RAG chatbot that answers recruiters' questions about my background using only my own documents. It splits the knowledge base by heading, embeds it with OpenAI, and retrieves the closest chunks from Pinecone, tagging every answer with its source section so visitors can see where it came from. It replies "I don't have that information" when the answer isn't in the sources.
- Fixed a prompt rule the model kept ignoring (questions about upcoming projects got "I don't have that information") by moving the decision into a separate intent classifier step. It scored 100% on a 20 question labelled set and 100% on a 20 question held out set, and real misses are added to the test suite.
- Built a FastAPI backend with streaming responses, per visitor session memory, input limits, per IP rate limiting and a CORS allow list. API keys stay on the server.
- Containerized the API with Docker and deployed it on Google Cloud Run with an instance cap and spend limits. Built the responsive, dark mode site in plain HTML, CSS and JavaScript, hosted on GitHub Pages with a custom domain, Cloudflare DNS and HTTPS.

**Text-to-SQL Evaluation Sensitivity** Fall 2025, Carleton, three person team

- Reproduced GPT-4 and GPT-3.5-Turbo inference pipelines across Spider and BIRD, zero shot and few shot.
- Used schema grounded prompts and fixed decoding, logging every prompt and output.
- Found exact match scored 0.0 on BIRD though about 83% of generated SQL was syntactically valid.
- Introduced token level Jaccard similarity and SQL validity rate as relaxed metrics, and analysed failure modes through qualitative error inspection.

**Stroke Risk Prediction (selected for presentation to external sponsors)** Jan to Apr 2026, Carleton, R, three person team

- Prepared 5,110 patient records (4.87% stroke cases): median imputed missing BMI, factored categoricals, stratified 80/20 split.
- Compared logistic regression, random forest, linear SVM and RBF SVM across baseline, upsampling and SMOTE conditions.
- Baseline hit about 95% accuracy but very low recall; upsampling was best (recall about 0.80 to 0.85, Random Forest best F1).
- My contribution: implemented the SMOTE condition (less stable than upsampling) and additional model experimentation.

## EXPERIENCE

**LoveWorld Canada** Ottawa, Canada | Oct 2025 to Apr 2026

*AI Engineer | Python, FastAPI, Claude API (Sonnet / Opus), ChromaDB, LangChain / LlamaIndex*

- Architected and deployed a RAG powered conversational learning assistant backend over a 6 month contract, indexing 500+ educational documents and transcripts for real time semantic search and automated study synthesis.
- Configured ChromaDB with 512 token chunks and 10% overlap, boosting retrieval relevance by 25% and preserving lesson context.
- Built asynchronous FastAPI REST endpoints with Server-Sent Events streaming, bringing time to first token under 800ms.
- Wrote Claude API system prompts enforcing source citations and grounded responses, reaching 95%+ faithfulness to source materials and reducing hallucinations.

**STEM the Gap** Ottawa, Canada | Feb 2024 to Aug 2024

*Data Scientist*

- Built predictive models forecasting demand for STEM programs at 88% accuracy, informing quarterly content and marketing strategy.
- Built automated Tableau and Python dashboards tracking real time website metrics, cutting manual reporting time by 40% and speeding leadership decisions.

## SKILLS

**Languages:** Python, SQL (SQLite), R, C, HTML/CSS/JavaScript; Java from coursework; data structures and algorithms (NeetCode, LeetCode)
**AI and LLM:** OpenAI and Claude APIs, prompt engineering, RAG, Pinecone, ChromaDB, MCP, evaluation, AI governance, QLoRA
**Frameworks and tools:** FastAPI, LangChain, CrewAI, Streamlit, Docker, GCP Cloud Run, GitHub Actions; Claude Code, Codex, Cursor daily

**Volunteer:** Media and Events Volunteer Coordinator, Christ Embassy (2017 to Present); Sports Prefect, Grange School (2019 to 2020).