

Aarya Shah

☎ 647-767-8243 | ✉ a268shah@uwaterloo.ca | 🔗 [linkedin/aarya](#) | 📅 Study Term Completed: 4A

TECHNICAL SKILLS

Languages: Python, SQL, R, Java, JavaScript, React, Node.js, C, C++, HTML/CSS, Scala
Tools: Azure, GCP, AWS, Snowflake, Excel, Tableau, PowerBI, Git, Airflow, DLT, Kubernetes, Docker
Frameworks: PyTorch, TensorFlow, XGBoost, Langchain, Kafka, Hadoop, Spark, DBT

EDUCATION

University of Waterloo September 2021 – April 2026
Bachelor of Mathematics in Statistics & Computer Science Waterloo, ON
• **Relevant Coursework:** Neural Networks, Data Types & Structures, Linear Models, Stochastic Processes

EXPERIENCE

Globys April 2025 - Present
Machine Learning Engineer Intern Seattle, WA

- Integrated active training & fine-tuning to **10+ multi-agent RAG LLMs** using Langchain/OpenAI embeddings
- Managed **vision transformers** models in **Python** to verify customer info with **98%** satisfaction from **8+ clients**
- Utilized CosmosDB (Azure) indexing in **C++ REST APIs** to optimize backend & reduce query latency by **70%**

Health Canada September 2024 – December 2024
Data Scientist Intern Toronto, ON

- Enforced a **feature store in GCP** that improved machine learning model scalability & performance by over **15%**
- Established time series models such as LSTM, Prophet & XGBoost for forecasting projects with **97%+ accuracy**
- Created **5+ deep learning** projects using **neural networks** (CNN/RNNs) for projects in image/text domains

Gore Mutual Insurance January 2024 – April 2024
Data Engineer Intern Cambridge, ON

- Implemented an **NLP model** to assess query complexities saving hosting costs by **\$500,000+** in Databricks
- Deployed **20+ data pipelines (ETL) & 10+ CI/CD workflows** to **Azure** using Python, Spark, SQL & Git
- Integrated a **Kafka-based** ingestion system reducing data latency by **40%**, feeding **10+ DLT pipelines**

RBC May 2023 – August 2023
Data Engineer Intern Mississauga, ON

- Identified various methods to transition data foundation from **DB2** to **Snowflake** to increase user flexibility
- Accelerated data pipelines by validating 25+ data sources in **Data Lake, Kafka, Hadoop (Hive) & SQL**
- Led the shift of converting legacy **ETL scheduling scripts** to **Airflow** which resulted in better customizability

IBM July 2023 – August 2023
Machine Learning Intern New York, NY

- Conducted **prompt tuning & prompt engineering** on watsonx **GEN AI** models to enhance human interaction
- Developed file processors in **Git, C++ & Docker** which improved RAG model performance for 10+ file types
- Assessed design variations of pre-existing LLMs in fraud detection & Q/A bots which **increased 3% accuracy**

RBC May 2022 – August 2022
Data Analyst Intern Mississauga, ON

- Performed **regression modelling** & statistical analysis in **R, Python & Excel** using macros/VBA
- Designed **15+ Tableau & PowerBI Dashboards** & wrote **20+ SQL queries** to influence business decisions
- Performed **regression and logistic modeling** & statistical analysis in **R, Python & Excel** using macros/VBA

PROJECTS

Policy Q/A chatbot (10000+ users) 🤖 | Python, Angular, Falcon-40b LLM, Langchain, HuggingFace
• Applied RAG for Q/A chatbot with ability to give quotes, check insurance coverage, and give tailored advice

Pharmaceutical (iOS) App (4000+ users) 🤖 | Node.js, React, HTML/CSS, Python, AWS, Nvidia LLMs
• Engineered robust app using ML to read prescriptions, check coverage, corresponding price and availability

Driver Drowsiness Detection (1500+ users) 🤖 | Python, CNN, Computer Vision (OpenCV)
• Constructed a system that processes pupil images & attention span to detect drowsiness with correctness of **92%**

Tensorflow Enhancements 🤖 | C++, Python
• Enhanced features such as compiling libtensorflowlite with SVE or restoring training capability for LiteRT models

Flight Path Optimization 🤖 | Python, SQL, PostgreSQL, Machine Learning
• Built an optimization algorithm using unsupervised learning methods to find most fuel efficient flight path