

Ilia Alenabi

☎ +1 (778) 708-2776 | @ialenabi@uwaterloo.ca | [in linkedin.com/in/iliall/](https://www.linkedin.com/in/iliall/) | [globe iliall.com](https://iliall.com)

EDUCATION

University of Waterloo

Honours Computer Science; Artificial Intelligence Specialization; Statistics Minor
GPA: 4.0 – President’s Scholarship of Distinction

Waterloo, Ontario

Sep 2022 – Expected

AWARDS & ACHIEVEMENTS

National Mathematical Olympiad – Silver Medalist

Combinatorics Olympiad (ICO) – Silver Medalist

SKILLS

Languages: Python, C/C++, Java, Golang, Rust, JavaScript, TypeScript, Rust

Frameworks/Libraries: Numpy, Pandas, PyTorch, Tensorflow, LLVM, MLIR, React, Django, Next.js

Tools/Platforms: Git, Docker, AWS, Kubernetes, Redis, BigQuery, PostgreSQL, Pinecone, MongoDB

EXPERIENCE

Google

Incoming Software Engineer - Internship

Kitchener, ON

Jan 2027 – Apr 2027

Cerebras

Machine Learning Engineer - Internship

Toronto, ON

Sep 2025 – Dec 2025

- Built a **tensor-dump** comparison tool to validate parity between **CPU** and **CSX** inference runs for **GPT-OSS**
- Developed a **checkpoint-conversion** pipeline for all **inference** models with a peak memory usage of only **7%**
- Refactored the **inference pipeline** configs to enable compatibility with custom **tokenizers** and **image encoders**

Huawei Canada

Compiler Engineer - Internship

Toronto, ON

Jan 2025 – Apr 2025

- Designed an **LLVM** pass for automated software cache creation, tuning memory usage in **distributed systems**
- Analyzed **Redis**’s performance, identified hotspots, and implemented **prefetching** to reduce runtime by **20%**
- Developed an **MLIR** pass to annotate attention layers with **sharding** metadata, optimizing **tensor distribution**

Questrade Financial Group.

Software Engineer - Internship

Remote

Sep 2024 – Dec 2024

- Automated **cloud-based MLOps** pipeline for **fine-tuning** in **Java**, reducing redundant AI inference costs by **8%**
- Integrated **authentication** checks into the internal pipeline, auditing SQL code from **200+** developers for security

Silverberry Group

Data Scientist - Internship

Vancouver, BC

May 2023 – Apr 2024

- Developed an **interactive agent environment** to model **medication-use** behavior prior to product release
- Vectorized 200+ hours of doctor-appointment audio in **Pinecone** using **OpenAI Whisper** and **Hugging Face**

RESEARCH

JetBrains

Research Intern – Model Routing for Cheaper Code Generation

Amsterdam, NL

Jan 2026 – Aug 2026

- Created a pipeline for extracting various OSS issues and generating **agentic** solutions with various **characteristics**
- Developed 6 distinct types of **LLM-as-a-Judge** to predict human preferences towards the answers relative to price

Vector Institute

Research Intern – Interpreting Vision-Language Models

Toronto, ON

May 2025 – Aug 2025

- Developed tooling to extract **intermediate** representations from **vision-language** models for **reasoning** analysis
- Ran **probing experiments** on **CLEVR** to evaluate **VLMs**’ internal representations of **primitive concepts**

Pingoo AI

Research Intern – Building Trustworthy Generative AI for Diabetes Care

Remote

May 2024 – Aug 2024

- Examined the **reliability** of LLMs for diabetes-focused applications using **few-shot learning** and **fine-tuning**
- Created a **self-evaluation** loop to enable models to refine their own outputs, achieving **85%** human-rated accuracy