

Education

University of Toronto

M.Eng., Computer Engineering

Toronto, Canada

Jan. 2024 - Jun. 2025

Beijing University of Posts and Telecommunications

B.Eng., Software Engineering

Beijing, China

Sep. 2019 - Jun. 2023

Technical Skills

Python, JavaScript, Java, Machine Learning, Deep Learning, MLOps, LLM Agent, PySpark, Azure, Git, Github, Flask, Django, MySQL, PyTorch, Mlflow, Microsoft Fabric, Databricks, Video Encoding, Audio Processing, Databricks

Work Experience

Microsoft

Software Engineer Intern

Articles on [Microsoft Learn](#)

Beijing, China

Feb. 2023 - Dec. 2023

- Initiated the internal version of **FLAML** for **Microsoft Fabric**, contributed **100+ PRs**, resulting in a **267% increase in users**. Also contribute back to **open-source FLAML** on **Github**.
- Enlarged Machine Learning model pool for AutoML by adding **15+** Machine Learning models based on **PyTorch Lightning**, **PySpark MLlib**, **Statsmodel** and **Scikit-Learn**.
- Integrated multiple **MLOps services** provided by Fabric platform via **MIFlow** into FLAML, including **ML experiment tracking, analysis, visualization, artifact logging**, and **model deployment**, reducing inference latency by **37%**.
- Developed the **first-ever autonomous feature engineering module** using **LLM Agent** among AutoML libraries.
- Benchmarked AutoML solutions between **Azure Machine Learning** and **Azure Databricks** on user experience and performance, revealing that **FLAML on Fabric** achieved **30× runtime efficiency** over Azure Machine Learning and a **16% average performance improvement** on benchmark datasets compared to Azure Databricks.

Call Fusion

Research Software Engineer

Try our [Demo](#)

Toronto, Canada

Sep. 2024 - Present

- Redesigned the technical architecture, migrating to an **Audio Agent** using **OpenAI's Real-time API**, creating an **end-to-end speech-to-speech** generation solution, and **reduced reply latency** by **52%**.
- Researched and integrated a noise reduction solution using the neural network-based algorithm **RNNNoise**, which **improved transcription accuracy** by **33%**.
- Benchmarked different vector databases to enhance **Retrieval-Augmented Generation (RAG)** efficiency; migrated from **Faiss** to **ChromaDB**, resulting in a **61% performance boost** in vector retrieval.

National University of Singapore

Undergraduate Research Assistant

Singapore

Apr. 2022 - Feb. 2023

- Participated in a research project about **Event Forecasting** with **Temporal Knowledge Graph** at the NExT++ Centre.

Side Projects

D2Helper, A Full-Stack Web Application for Destiny 2

- A Destiny 2 information and tracking tool with a **4,000+** user base.
- Built a modern user interface using **JavaScript** and **CSS**. Implemented **RESTful API** backend service using **Flask** with **MySQL** cloud database. Used **Docker** for deployment on Tencent Cloud's **Serverless** Application.

AutoGen, A programming framework for Agentic AI

- Active contributor of this **45,000+** stars open-source project, early contributor to **RAG** module in AutoGen.
- Integrated **Anthropic Claude** models as alternative foundation model for LLM Agent.