Wendong Li

(647) 233-7096 | levscaut@gmail.com | Website | LinkedIn | Github

Education

University of Toronto — **Toronto, CA**

M.Eng., Computer Engineering — Jan. 2024 - Apr. 2025(Expected)

Relevant Courses: Distributed System in Moden Web-Scale Applications, Video Encoding, Blockchain Development, AI Alignment, Natural Language Processing Application.

Beijing University of Posts and Telecommunications — **Beijing, CN**

B.Eng., Software Engineering — 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, Multimedia Encoding, Distributed System.

Work Experience

Microsoft — **Beijing, CN**

Software Engineer Intern — ***Articles on Microsoft Learn*** — Feb. 2023 - Dec. 2023

- Contributed to **FLAML**, Microsoft's open-source **Automated Machine Learning** (**AutoML**) library, initiated an internal version for the **Microsoft Fabric** data platform. Submitted **100+ PRs**, driving a **263% increase** in user adoption.
- Expanded AutoML model candidate pool by adding **15+ Machine Learning models (PyTorch Lightning, PySpark MLlib, Statsmodels, Scikit-learn)**.
- Integrated multiple **MLOps** services (experiment tracking, visualization, artifact logging, deployment) with custom **MLflow** backend, reducing **inference latency** by **37%**.
- Developed the **first-ever autonomous feature engineering** module using **LLM Agent** within an AutoML library.
- Benchmarked AutoML solutions against **Azure Machine Learning Studio** and **Azure Databricks**; FLAML on Fabric achieved **30× faster runtimes** and a **16%** average performance improvement over competing solutions.

Call Fusion — **Toronto, CA**

Research Software Engineer — ***Try our Demo*** — Sep. 2024 - Present

- Redesigned the technical architecture for an **Audio LLM Agent** using OpenAI's Real-time API, reducing end-to-end speech-to-speech **reply latency** by **52%**.
- Integrated RNNoise, a neural network–based noise-reduction algorithm, boosting **transcription accuracy** by **13%**.
- Evaluated and migrated from **Faiss** to **ChromaDB** for Retrieval-Augmented Generation (RAG), **improving retrieval performance** by **16%**, **reducing 96% of peak cases**.

National University of Singapore — **Singapore, SG**

Undergraduate Research Assistant — Apr. 2022 - Feb. 2023

- Collaborated on an **Event Forecasting with Temporal Knowledge Graph** project at the NExT++ Centre, focusing on research design and implementation.

Side Projects

AutoGen, A programming framework for Agentic AI

- Contributed to **high-level system design** in an Agentic AI framework with **35k+ GitHub stars**.
- Early contributor to the **RAG** module, benchmarking various vector databases and integrating **ChromaDB**.
- Added support for **Anthropic Claude** models as an alternative foundation model for LLM Agents.

D2Helper, A Full-Stack Web Application for Destiny 2

- Developed a **full-stack** web application for the Destiny 2 community with 4,000+ active users.
- Built frontend with **JavaScript and CSS**; implemented a **RESTful** backend with **Flask** and a **MySQL** cloud database.
- Deployed via **Docker** on Tencent Cloud's **Serverless** platform.