Wendong Li

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

Education

University of Toronto — **Toronto, CA**
M.Eng., Computer Engineering — Jan. 2024 - Apr. 2025(Expected)

Beijing University of Posts and Telecommunications — **Beijing, CN**
B.Eng., Software Engineering — Sep. 2019 - Jun. 2023

Technical Skills

Programming Language: Python, JavaScript, Java, C++, Solidity, OOD
Frameworks and Tools: Linux, CI/CD, Git, Azure, AWS, GCP, PySpark, Flask, Django, MySQL, Microsoft Fabric, Databricks,
AI&ML: MLOps, LLM Agent, PyTorch, Tensorflow, MLflow, GNN, Transformer, Huggingface, XGBoost, LightGBM, Time-series
Domain Knowledge: Multimedia Encoding, Distributed System, Blockchain(Smart Contract and DApp), Parallel Computing

Experience

Microsoft — **Beijing, CN**
Software Engineer Intern — Feb. 2023 - Dec. 2023

FLAML, an Automated Machine Learning (AutoML) Library

- Contributed to the design and led the development of an internal version of FLAML for the **Microsoft Fabric** platform.
- Authored **articles** to help data scientists onboard this **Python** library, **user number increased by 263%** after internship.
- Benchmarked AutoML solutions against **Azure Machine Learning (AML)** and **Azure Databricks**; FLAML on Fabric achieved **30× faster runtimes** than AML and a **16%** average performance improvement over Azure Databricks.
- Expanded AutoML model pool by adding **15+** models **(PyTorch Lightning, PySpark MLlib, Statsmodels, Scikit-learn)**.
- Integrated multiple **MLOps** services (experiment tracking, visualization, artifact logging, model version control and serving) with custom **MLflow** backend, reducing **inference latency** by **37%**.
- Developed an autonomous **feature engineering** module using **LLM Agent**, reducing human labor for data scientists by automating feature engineering experiments.

Call Fusion — **Toronto, CA**
Machine Learning Engineer — Sep. 2024 - Present

Backend Services and Audio LLM Agent, Live Demo Website

- 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 peak latency by 96%**.
- Used **Celery** to schedule database updates, serving on **AWS Elastic Beanstalk** to improve stability and availability.

National University of Singapore — **Singapore, SG**
Undergraduate Research Assistant — Apr. 2022 - Feb. 2023

- Participated in a **Temporal Knowledge Graph** research project at School of Computing, NExT++ Research Centre.

Projects

AutoGen, A programming framework for Agentic AI — **Github Link**

- 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 **4000+** 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.