Wei-Chung (Wells) Lu

wellslu.wc@gmail.com | 805 896 7712 | Website | LinkedIn | Github

Skills

Programming: Python, C/C++, Linux, SQL, HTML, Javascript, Java, R, Verilog

SDE Skill: Git, Docker, Singularity/Apptainer, PyTorch, Tensorflow, LangChain, LangGraph, FastAPI, Model Context Protocol(MCP), Google Cloud Platform (GCP), Amazon Web Service (AWS), Azure, CoreML, Snapdragon SDK, cuBLAS

Language: English, Chinese, Japanese

Work Experience

Data Scientist Intern, NXP — Mar. 2026 - Present

- Developed and implemented IEA (Intelligence Engineering Assistant), an AI Agent that assists the design houses in analyzing and testing.
- Deploying and designing FastAPI and Model Context Protocol(MCP) server as backend.

Artificial Intelligence Engineer, Open AI Fab — Mar. 2022 - Nov. 2023

- Independently developed a MobileNetV3-FCOS to identify children's development with over 80% mAP (mean Average Precision), and 25 fps in each test.
- Trained a MobileNetV3-SSD model to detect sock tops with over 95% mAP and inference less than 0.01 seconds.
- Converted model with Core ML and Snapdragon SDK for applying to edge or mobile devices.

Machine Learning Engineer, Merkle — Aug. 2022 - May 2023

- Refined MLOps on GCP to automate model updates of customer tagging projects to reduce 80% working time.
- Enhanced prediction of collaborative filtering to improve sales performance uplift rate by 5% in A/B Testing.
- Mentored two interns to optimize program efficiency to reduce RAM usage and data load time by almost 50%.

Education

University of California Santa Barbara, Master in Electrical and Computer Engineering — Sept. 2024 - June 2026

Soochow University, Bachelor in Data Science — Sept. 2017 - June 2021

Research Experience

Research Intern, Cornell University — June 2025 - Mar. 2026

- Assisted Prof. Saikat Dutta in developing OmniCode, a large-scale benchmark for evaluating LLM-based software engineering agents; **co-first author**. Built a dataset of 1,794 tasks across Python, Java, and C++, derived from 494 real-world GitHub issues spanning 28 repositories.
- Atharv Sonwane, Eng-Shen Tu, **Wei-Chung Lu***, et al. "OmniCode: A Benchmark for Evaluating Software Development Agents," Accepted by ACL Findings, 2026. (*Equal contribution)

Side Projects

Compare the Inference Speed of TVM and C++ on Edge Device Winter, 2025

- Cross-compiled AI models (take ResNet family as example) with TVM for implementing on Raspberry Pi 3 B+.
- Ran models with C++ source code and libtorch on Raspberry Pi 3 B+ to compare these three methods.
- Tools Used: Python, Pytorch, C++, TVM, libtorch

CP-Decomposition in CNN Fall, 2024

- Replaced Conv layer in LeNet model with CP-Decomposition Conv layer and trained it to reduce model weights.
- Tools Used: Python, Pytorch, C++, Cuda

Customer Tracking System Fall, 2020

- Developed a multi-model system for retail stores to identify new customers, customers who just came a few minutes ago, and members.
- Fine-tuned Joint Detection and Embedding (JDE) model to track and identify different customers with over 73% MOTA.
- Tools Used: Python, Pytorch, SQL, Linux, Parallel Computing

Certification

Google Cloud Certified - Professional Data Engineer Mar.2023

The AI training certification from Qualcomm June 2022

Salesforce Certified Administrator Feb. 2023

Japanese-Language Proficiency Test (JLPT) N3 Feb. 2020