WUCHEN (AUBREY) LI

New York, NY | (669) 302-7032 | wl758@cornell.edu
GitHub/ LinkedIn/ Website

EDUCATION

Cornell University, New York, NY — Aug. 2024 – May 2026
Master of Science in Information Systems
Relevant Coursework: Data Science, Building Startup Systems (Full-Stack), Natural Language Processing, Computer Vision

ShanghaiTech University, Shanghai, CN — Aug. 2018 – Jun. 2022
Bachelor of Engineering in Computer Science
Relevant Coursework: Software Engineering, Database System, Algorithm & Data Structure, Optimization & Machine Learning

PROFESSIONAL EXPERIENCE

Amazon Web Services (AWS), SDE Intern in AI/ML, New York, NY — May 2025 – Aug. 2025

- **Built GenAI Powered Data Infrastructure:** Developed Model Context Protocol (MCP) system to analyze Bedrock capacity data and provide natural language insights for management team.
  - **MCP Client:** Deployed an MCP client on AWS Fargate using Strands API to receive user queries, extract relevant data with the MCP servers, and generate context-aware answers.
  - **MCP Server:** Implemented Lambda-based MCP servers with AI-powered SQL query generation, AWS Glue crawler automation, and standardized API endpoints for historical data analysis; with AWS CloudWatch query for real-time data analysis.
- **Designed ETL Data Pipelines:** Implemented scalable end-to-end ETL pipelines using S3, Aurora Serverless DSQL, DynamoDB, AWS Glue, Lambda, and EventBridge to process and transform metrics from CloudWatch logs into AI agent-consumable formats.

Florens Asset Management Company, Data Engineer, Shanghai, CN — Sept. 2022 – Jul. 2024

- **Developed Scalable Data Analysis Platform:** Designed and maintained data mid-platform to support business decisions.
  - Built ETL pipelines with HIVE, Greenplum, and PostgreSQL; created **40+** Tableau BI and FineBI dashboards.
- **Developed Asset Selection System:** Engineered a high-performance system to streamline portfolio asset selection.
  - Developed a linear optimization engine using Python, SQL, and CPLEX, supporting selection from **3+ million** assets.
  - Reduced selection time from several days to **5 minutes**, automating **90%** of the selection workflow.
- **Built AI-Powered Automation for Logistics:** Developed CV and LLM solutions to improve operational efficiency.
  - Developed a Mask RCNN based CV model with **95+%** accuracy to detect floor damages in the shipping container return process.
  - Designed an LLM-driven DAG workflow with ChatGPT/Llama and LangChain for automated order booking email processing.

Intel, Software Engineer Intern, Shanghai, CN — Nov. 2021 - Feb. 2022

- **Contributed to Open-Source Recommender Systems:** Contributed to **DeepRec**, co-developed with Alibaba's AI Team.
  - Conducted performance evaluations and testing of BST, DIEN, and DSSM models to identify optimization opportunities.
  - Optimized model efficiency with BF16 precision and self-attention, leveraging Kubernetes for scalable training and testing.
- **Optimized LSTM Model Inference Speed:** Improved the inference performance of Intel's PyTorch LSTM Operator.
  - Optimized LSTM operator inference through profiling, memory alignment, integrating Intel dgemm library (C++/C) , and exp() approximation, achieving **3.5x** speedup. Ensured reproducible benchmarking with Docker.

TECHNICAL SKILLS

- **Coding Language:** Python, C++/C, C#, SQL, Shell Scripting, JavaScript, Java, HTML, CSS, R
- **Tools & Frameworks:** Git, Docker, Linux, AWS Cloud, PyTorch, React, Node.js, Spark, Unity3D, OpenCV
- **Professional Tools:** PostgreSQL, HIVE, Greenplum, CPLEX, Pandas, JSON, Figma, Tableau BI, Excel, CI/CD
- **Other Relevant Course:** Software Engineer, HCI, Building Startup Systems, Unity Game Development, Cryptography

PROJECTS

Cornell Tech: MiniTorch Machine Learning Framework Project (NumPy, Numba, Pytest) — Aug. 2024 – Dec. 2024
Course Project: Developed a PyTorch-like ML framework based on Python with auto-differentiation and GPU acceleration.

- Implemented broadcasting, backpropagation, and auto-differentiation for neural network training.
- Integrated GPU acceleration using Numba and operator fusion, achieving **100x** speedup in training and inference.
- Established a Python modular architecture with Pytest unit tests, ensuring reliability and maintainability.

Awards & Certificate

Cornell University: Certificate of Accomplishment in Computer Science
MICCAI: 2023 CBCT Semi-Supervised Tooth Segmentation Challenge Winner