Leo Lee

: linkedin.com/in/thl-leo/ : github.com/THL-Leo : thleo.co/
Email: leolee.developer@gmail.com Mobile: (949) 209-7607

Education

University of Southern California — Los Angeles, CA
Master of Science in Computer Science - Data Science — *August 2023 - May 2025*

University of California, Santa Cruz — Santa Cruz, CA
Bachelor of Science in Computer Science — *October 2020 - June 2023*
Honors: Cum Laude with Highest Honor in the Major.

Experience

Software Engineer Intern — San Jose, CA
AMD (Advanced Micro Devices) — *May 2024 - August 2024*

- Developed dashboard using Python, Flask, MySQL, and JavaScript enabling engineers to monitor test pattern health and device performance
- Optimized dashboard loading time from 13 seconds to 0.2 seconds using SQLAlchemy and asynchronous data loading
- Utilized Alembic for automated database migration and maintained schema consistency across multiple environments
- Implemented Celery notification system for automated alerts and device performance insights
- Created interactive web pages using JavaScript for dynamic content updates, HTML for structured layouts, and CSS for responsive styling and visual enhancements

Machine Learning Research Intern — Los Angeles, CA
University of Southern California — *September 2023 - May 2024*

- Developed comprehensive benchmark platform for outlier detection algorithms using PyOD on large textual corpus datasets
- Implemented feature extraction pipeline using BERT and OpenAI GPT-3 embeddings for anomaly detection model evaluation
- Automated model comparison scripts across multiple outlier detection algorithms, analyzing performance metrics and generating comparative results
- Coordinated team workflow and presented IEEE format research paper to classmates

Projects

Agentic RAG — *Python, LangGraph, LangChain, OpenAI, ChromaDB, Streamlit*

- Built an Agentic RAG for Japanese history using LangGraph with multi-turn conversationality and document retrieval via ChromaDB vector database
- Implemented LLM-as-Judge pattern with GPT-4 for relevance assessment and query rewriting
- Developed Streamlit frontend with real-time streaming and interactive chat interface
- Processed documents using OpenAI embeddings and recursive text splitting for context summarization

Graph Neural Network Traffic Forecasting — *Python, PyTorch, GNNs, Pandas, Scikit-learn, Jupyter*

- Developed DCRNN model using PyTorch and GNN architectures for traffic speed prediction on METR-LA dataset with 207 sensors across Los Angeles
- Implemented model experimentation comparing Euclidean distance, road-network, and correlation-based graph structures
- Built PyTorch pipeline with custom dataset classes, normalization, and training frameworks

MapReduce Inverted Index System — *Python, mrjob, Hadoop, Text Processing*

- Developed distributed MapReduce programs using Hadoop for building unigram and bigram inverted indexes from unstructured text collections
- Implemented text preprocessing with regex cleaning, case normalization, and tokenization
- Built combiner and reducer functions to aggregate term frequencies across documents
- Created selective bigram indexing for terms like "computer science" and "information retrieval"

Skills

Languages: Python, C++, MySQL, PostgreSQL, NoSQL, JavaScript, TypeScript, Java, Go

Machine Learning & AI: PyTorch, Transformers, HuggingFace, GNNs, Scikit-learn, LangChain, LangGraph, OpenAI GPT, BERT, ChromaDB, PyOD, YOLO

Data Tools & Libraries: Pandas, NumPy, Matplotlib, Jupyter, Hadoop, MapReduce, GeoPandas, OSMnx, RTrees, Shapely

Data Storage & Databases: ChromaDB, Weaviate, MySQL, PostgreSQL, NoSQL, SQLite

Frameworks & APIs: Flask, FastAPI, Docker, RESTful API, Celery, Streamlit, mrjob

Developer Tools: Git, Docker, Jupyter, CI/CD, Google Cloud Platform (GCP), Swagger, Pylint, MyPy, Alembic, Playwright, Agile, JIRA, Confluence, Make