

# Leo Lee

 : [linkedin.com/in/thl-leo/](https://www.linkedin.com/in/thl-leo/)  : [github.com/THL-Leo](https://github.com/THL-Leo)  : [thleo.co/](https://thleo.co/)

Email: [leolee.developer@gmail.com](mailto:leolee.developer@gmail.com) Mobile: (949) 209-7607

## EDUCATION

---

### University of Southern California

*Master of Science in Computer Science - Data Science*

Los Angeles, CA

*August 2023 - May 2025*

### University of California, Santa Cruz

*Bachelor of Science in Computer Science*

*Honors: Cum Laude with Highest Honor in the Major.*

Santa Cruz, CA

*October 2020 - June 2023*

## EXPERIENCE

---

### Software Engineer Intern

*AMD (Advanced Micro Devices)*

San Jose, CA

*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
- Implemented Celery notification system for automated alerts and device performance insights

### Machine Learning Research Intern

*University of Southern California*

Los Angeles, CA

*September 2023 - May 2024*

- Conducted benchmarking on outlier detection algorithms using PyOD for large textual documents
- Utilized SkLearn to preprocess and label datasets for machine learning analysis
- Coordinated team workflow and presented IEEE format research paper to classmates

## PROJECTS

---

### Spatial Indexing k-Nearest Neighbor Query — *C++, Python, Pandas, matplotlib*

- Implemented spatial data structures in C++ including Grid Index, kD-Tree, and brute force algorithms for 825K+ geographic coordinates
- Developed k-nearest neighbor and range query algorithms with performance benchmarking across varying dataset sizes
- Created Python testing framework with visualization tools demonstrating 100x speedup over brute force methods
- Built modular C++ architecture with custom hash functions and priority queues for spatial query processing

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

- Built agentic RAG system for Japanese history using LangGraph with document retrieval and ChromaDB vector database
- Implemented LLM-as-Judge pattern using GPT-4 for relevance assessment and query rewriting with loop prevention
- Developed Streamlit frontend with real-time streaming and chat interface
- Processed documents using OpenAI embeddings and recursive text splitting with context summarization

### GNN Spatiotemporal Traffic Forecasting — *Python, PyTorch, NumPy, Pandas, Scikit-learn, Jupyter*

- Developed DCRNN for traffic speed prediction using METR-LA dataset with 207 sensors across Los Angeles
- Implemented graph structure experiments including Euclidean distance, road-network, and correlation-based graphs for forecasting accuracy
- Built PyTorch pipeline with custom dataset classes, normalization, and training frameworks for spatiotemporal analysis

## SKILLS

---

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

**Machine Learning:** PyTorch, Scikit-learn, PyOD, YOLO, LangChain, LangGraph, ChromaDB

**Data Tools & Libraries:** Pandas, NumPy, matplotlib, GeoPandas, OSMnx, RTrees, Shapely

**Frameworks & APIs:** Flask, FastAPI, Celery, RESTful API

**Developer Tools:** Git, Docker, Google Cloud Platform (GCP), Swagger, Pylint, MyPy, CI/CD, Jupyter

**Other Tools:** Playwright, Agile, JIRA, Confluence, Make