

Guanghan Li

(415) 604-7028 | Guanghanli90@gmail.com | [LinkedIn](#) | [GitHub](#) | Sunnyvale, CA

EDUCATION

Santa Clara University

Bachelor of Science in Computer Science (GPA: 3.9/4.0)

Santa Clara, CA

Expected Dec 2025

- Relevant Coursework: Machine Learning, Deep Learning, Data Science, Web Mining, Data Structures, Algorithms, Database Systems, Python Programming

EXPERIENCE

AI Solutions Intern

Jun. 2025 – Aug. 2025

Hi-Think Corporation

Liaoning, China

- Fine-tuned Moonshot-Kimi-v2 LLM on 240,000+ feedback entries using supervised fine-tuning (SFT) and parameter-efficient fine-tuning (PEFT), achieving 97.3% sentiment classification accuracy through optimized prompt engineering and few-shot learning strategies
- Engineered end-to-end NLP pipelines using Pandas, LangChain, and DashScope for feature extraction, sentiment analysis, and multi-dimensional segmentation across 20+ organizational variables, building AI-driven correlation models to identify 5+ key satisfaction drivers
- Deployed production ML system with FastAPI, implementing model versioning, A/B testing, and monitoring pipelines with Docker containerization, reducing manual analysis time by 85% through automated workflow integration using n8n

Data Engineer Intern

Jan. 2025 – May. 2025

Santa Clara University

Santa Clara, CA

- Engineered feature extraction pipelines to process multivariate time-series data from Catapult GPS sensors, extracting kinematic features (acceleration, velocity) and physiological metrics (heart rate, player load) for 25+ athletes with data validation and cleansing workflows
- Designed predictive readiness model using ensemble methods (Decision Trees, Random Forest, Gradient Boosting) in Scikit-Learn, achieving RMSE of 0.85 through comprehensive feature engineering, EDA, and statistical hypothesis testing, contributing to 20% injury reduction
- Built interactive data visualization dashboards using Matplotlib and Seaborn for real-time athlete monitoring and workload management, implementing PostgreSQL backend for efficient time-series data storage and query optimization

PROJECTS

Full-Stack Real Estate Analytics Platform | Python, ML, React, FastAPI, PostgreSQL

Mar. 2025 – Jun. 2025

- Architected end-to-end machine learning pipeline integrating web scraping (Selenium, BeautifulSoup) to extract 1000+ property listings, comprehensive data preprocessing (missing value imputation, outlier detection via IQR, feature scaling), and association rule mining using Apriori algorithm to discover property attribute relationships
- Trained and optimized ensemble models (Random Forest, XGBoost, Linear Regression) through hyperparameter tuning (GridSearchCV), improving R² from 0.56 to 0.89, and analyzed feature importance using SHAP values for model interpretability

NFT Marketplace Platform | PHP, MySQL, Solidity, Web3.js, Data Architecture

Ongoing

- Designed and implemented normalized database schema in MySQL for storing NFT metadata, transaction history, user profiles, and blockchain event logs, optimizing query performance through indexing strategies and implementing data integrity constraints
- Built data integration pipelines using PHP and Web3.js to synchronize on-chain blockchain data (smart contract events, token transfers) with off-chain relational database, implementing ETL processes for data consistency and developing analytics queries for marketplace metrics and user behavior patterns

TECHNICAL SKILLS

Languages: Python, Java, C++, SQL, R

ML/DL Frameworks: TensorFlow, PyTorch, Keras, Scikit-learn, XGBoost, LightGBM

NLP & LLMs: Transformers, Hugging Face, LangChain, LLM Fine-tuning (LoRA, PEFT), Prompt Engineering

Data Science: Pandas, NumPy, SciPy, Matplotlib, Seaborn, Plotly, Feature Engineering, SHAP, A/B Testing

Deep Learning: CNN, RNN, LSTM, Attention Mechanisms, Transfer Learning, Model Optimization

MLOps & Tools: Docker, AWS (EC2, S3, SageMaker), FastAPI, PostgreSQL, MySQL, Git, ETL Pipelines