

Chenhao Liang

liangchenhao123@gmail.com | liangchenhao.com | linkedin.com/in/chenhao-liang | github.com/LeoLiang-cs | Los Angeles, CA

EDUCATION

University of Southern California (USC)

Los Angeles, CA

Master of Science in Computer Science | GPA: 4.0/4.0 | Coursework: Analysis of Algorithms, Machine Learning for Data Science *Jan 2026 – Dec 2027*

Southern University of Science and Technology (SUSTech)

Shenzhen, China

Bachelor of Science in Computer Science | GPA: 3.60/4.00 *Sep 2021 – Jun 2025*

EXPERIENCE

Figwork | Software Engineering Intern, Part-time Remote

Sunnyvale, CA Sep 2026 – Present

- Developed LLM and backend systems for Figwork's AI recruiting platform (NestJS/TypeScript, PostgreSQL, AWS Bedrock), building the agent that turns a recruiter's chat into a structured hiring spec and the backend services behind candidate matching.
- Co-architected a two-stage candidate recommendation system over 160K+ profiles (embedding recall + LLM reranking against a 106-field job schema), which learns each recruiter's preferences online from like/skip feedback with cold-start exploration.
- Migrated candidate retrieval to pgvector and accelerated hiring-spec insights with caching and an incrementally refreshed inverted index, cutting end-to-end insights latency from 22s to 0.2s.

Segi | Software Engineering Intern, AI Platform Development

Shenzhen, China Jun 2025 – Dec 2025

- Developed core backend services for AI-powered customer support and workflow automation within a property management platform serving 3 enterprise clients across 60+ managed properties; customer-service workflows processed 11,800+ tickets, reducing average ticket creation time from 2 minutes to under 30 seconds.
- Engineered multi-tenant WhatsApp/WeCom message ingestion pipelines in Spring Boot with incremental polling, Redis per-tenant locks and TTL deduplication, and DB-backed cursor recovery to prevent duplicate or lost messages.
- Designed and implemented a modular agent workflow engine using Spring Boot and LangChain4j, with reusable nodes and conditional routing to orchestrate intent recognition, parallel multi-query RAG retrieval with Milvus, and MCP tool execution.
- Built fault-tolerant AI request handling with bounded Java thread pools, timeouts, retries, circuit breaking, and fallback replies, using RabbitMQ to absorb traffic spikes and raising AI reply success from 83% to 96% in 200+ concurrent load tests.

MGI Tech | Software Engineering Intern, Algorithm Development Department

Shenzhen, China Jan 2023 – May 2023

- Built perception algorithms for a robotic ultrasound system: fine-tuned AlphaPose on supine patients to localize the xiphoid process from skeletal keypoints for robot-arm positioning, and cut ultrasound image origin-detection latency from 200 ms to 11 ms.

PROJECTS

NextTier – AI-Powered Custom Merchandise Platform | Industry Project Lead

Los Angeles, CA Apr 2026 – Jul 2026

- Built and deployed an AI-powered custom merchandise platform on AWS using React and FastAPI microservices, integrating design generation, product previews, and ordering through REST APIs; validated through 500 simulated end-to-end workflows.
- Designed PostgreSQL schemas for products and shopping carts and implemented Redis caching to reduce repeated database reads during product browsing and cart interactions, reducing database reads by 69% and improving API response latency by 32%.
- Integrated AWS Bedrock into backend AI generation services and added request validation, duplicate-request protection, and invocation safeguards to prevent malformed inputs and redundant model calls.
- Implemented an asynchronous order-processing pipeline using Amazon SQS to decouple checkout requests from downstream order processing, reducing P95 checkout latency by 77% while sustaining 100 concurrent users with <1% failures in load tests.

RESEARCH EXPERIENCE

CUA-SWE: When Computer-Use Agents Meet Visual Software Engineering

Los Angeles, CA Jun 2026 – Sep 2026

Co-first Author | *FORTIS Lab, University of Southern California* | *Under review at ICLR 2027*

- Curated 33 Web repair tasks through an adversarial curation loop, adapting real GitHub issues and observed UI failures into cases that challenged code-only agents and revealed CUA advantages and limitations.
- Built deterministic behavioral verifiers and analyzed paired code-only and CUA trajectories across screenshots, browser actions, patches, and test outcomes to diagnose partial repairs, protocol violations, and frontier failures.

PUBLICATIONS

Li, Z., Ling, Y., Gu, Y., Huang, B., **Liang, C.**, Islam, S., ... & Seitla, D. FELT: Generating Tactile Signals from Vision for Visuo-Tactile Manipulation. *Tactile Sensing for Robotic Foundation Models Workshop, RSS 2026*.

Lin, H., Li, H., Li, H., Yu, X., Yu, K., **Liang, C.**, ... & Liu, J. Data Heterogeneity-Aware Personalized Federated Learning for Diagnosis. *OMIA Workshop, MICCAI 2024*.

TECHNICAL SKILLS

Languages: Java, Python, JavaScript, TypeScript, C/C++, SQL, HTML, CSS

Frameworks: Spring Boot, FastAPI, NestJS, React.js, Next.js, LangChain, LangChain4j, Prisma

Databases: MySQL, Redis, PostgreSQL, Milvus, MongoDB

Developer Tools: AWS, Docker, Kubernetes, Jenkins, Git, Linux, Maven, Nginx, RabbitMQ