

# HENRY HSU

Houston, TX 77030

+1-510-996-6837   [hsuhengjui@gmail.com](mailto:hsuhengjui@gmail.com)   [linkedin.com/in/yarikama](https://www.linkedin.com/in/yarikama)   [github.com/yarikama](https://github.com/yarikama)

## SUMMARY

**GenAI Full-Stack Software Engineer who optimized 140+ APIs (27.7% faster), built WebSocket real-time systems with Redis for frontend-backend integration, scaled systems to 20M+ records with 10M+ async data pipelines, achieved 67% test coverage with CI/CD, and contributed 14+ PRs to LlamaIndex. Expertise in Python, C++, TypeScript. Pursuing M.C.S. at Rice University.**

## EDUCATION

### Rice University - Top 20 U.S. University

M.C.S., Computer Science (GPA: 4.00/4.00)

Courses: Database Implementation, Web Development, Big Data Tools and ML Models

**Aug. 2025 – Dec. 2026 (Expected)**

Houston, TX

### National Yang Ming Chiao Tung University (NYCU) - Top 3 Taiwan University

Minor: Computer Science (Domain GPA: 4.13/4.30)

B.S., Industrial Engineering and Management (GPA: 4.07/4.30)

**Sep. 2020 – Jun. 2024**

Hsinchu, Taiwan

## WORK EXPERIENCE

### Back-End Engineer & GenAI Team Lead

MaiAgent Co., Ltd. – Award-Winning B2B GenAI Startup

**Sep. 2024 – Aug. 2025**

Taipei, Taiwan

- **API Optimization** – Optimized **140+** RESTful APIs through SQL query refactoring, connection pooling (Elasticsearch, Cohere, OpenAI), and Django caching; reduced response time of 13 high-traffic APIs by **27.7%** overall and prevented N+1 queries
- **Scalable Data Pipeline** – Architected async data processing pipeline using **asyncio**, **Celery**, and batch/generator patterns to handle **10M+** records for multi-level knowledge base indexing, achieving **5x faster** parsing speed
- **Real-Time Communication** – Implemented **WebSocket**-based real-time notification system with **Redis** pub/sub for file parsing status updates; built event-driven architecture for agent state transitions, pushing canvas rendering triggers to frontend for UI updates
- **System Architecture & Refactoring** – Led incremental refactoring with backward/forward compatibility, scaling from **3M to 20M+** text chunks; designed normalized data schema and implemented Singleton patterns to optimize object creation and memory usage
- **CI/CD & Testing** – Built **GitHub Actions** CI pipeline with pytest integration and unit testing, achieving **67%** code coverage from zero baseline; reduced production hotfixes by **90%** initially and **50%** long-term; managed database migrations in CD pipeline

## PROJECTS

**Elderly Care Map Website** | Apache, PHP, MySQL, Google Maps API, HTML/CSS (Bootstrap5), JavaScript

**Sep. 2023 – Jan. 2024**

- **Database Design & Integration:** Designed normalized database schema with 8 tables handling 2,000+ long-term care institutions; implemented complex joins and constraints for user favorites, institution info, and geolocation data; integrated government open data (CSV) with automated ETL pipeline
- **Full-Stack Development:** Built interactive map-based search platform with Google Maps API integration featuring radius-based filtering, bed capacity classification (color-coded indicators), and real-time distance calculation; implemented user authentication system with CRUD operations for favorite institutions management

**Chat Bar - Real-Time Multiplayer Game Server** | C/C++ (Socket Programming, SFML), MySQL, SHA-256

**Oct. 2023 – Jan. 2024**

- **Server-Client Architecture:** Designed and implemented real-time multiplayer game server using **TCP socket programming** with multi-threaded client handling, supporting concurrent connections for chat and gameplay synchronization
- **User Management System:** Built authentication system with SHA-256 password hashing, MySQL database integration for user data persistence, and implemented ranking/leaderboard features with optimized queries
- **Game Engine:** Developed client-side GUI and game logic using SFML library with event-driven architecture for character movement, group chatting, and real-time state updates

## AWARDS

**Presidential Hackathon Winner (2024)** - Urban noise detection with LLM-powered structured analysis

**23rd Golden Peak Award (2025)** - Outstanding Commercial Product, MaiAgent AI Platform

**Two-Time Dean's List Recipient** - Top 5% Academic Performance, NYCU

**AI Workshop Outstanding Award (Top 3/50)** - Multi-Agent RAG tutoring system, NYCU CS

## SKILLS

- **Languages:** Python, C/C++, TypeScript | **Frontend:** HTML, CSS, JavaScript, React | **Backend:** Django, FastAPI, Express.js
- **APIs:** RESTful, WebSocket, Redis Pub/Sub | **Databases:** PostgreSQL, Neo4j, Milvus, Elasticsearch, Chroma, MongoDB
- **Async:** asyncio, Celery | **DevOps:** GitHub Actions, Docker, git, Shell Script | **AWS:** EC2, S3, RDS, ElastiCache | **Methods:** Agile, Scrum