LIUWEI (Liu Wei)

AI & Full-Stack Developer | Computer Science Student

Email: sunflowerslw0607@gmail.com | Phone: +86 18355366842 | LinkedIn: linkedin.com/in/liu-wei-936468338 | GitHub: github.com/SunflowersLwtech

About Me

Solid foundation in AI theory with extensive project experience, specializing in deep learning, computer vision, and large language model applications. Complete experience from model selection, optimization, deployment to full-stack AI application development. Passionate about solving technical challenges in high-performance AI systems, eager to create value in real business scenarios.

Education

Asia Pacific University (APU) & De Montfort University (DMU), Malaysia   2024-2027 (Expected)
Bachelor of Science (Hons) in Computer Science (Dual Degree)
- **CGPA: 3.87 / 4.00** (First Class Honours) | Entrance Scholarship Recipient

Hackathon & Competition

Deriv AI Hackathon 2026 -- Finalist (Top 4-7)   Feb 2026
Team: TradeIQ | Cyberjaya, Malaysia | Core Developer & AI Pipeline Architect
- Built **"TradeIQ"**, an AI-powered intelligent trading platform for retail traders -- awarded **RM 2,000** prize as Finalist among 56 teams (200+ participants)
- Designed a **5-agent AI pipeline** (Market Monitor -> Event Analyst -> Portfolio Advisor -> Behavioral Sentinel -> Content Creator) using **DeepSeek V3** with streaming function calling
- Developed flagship **"Trading Twin"** feature: dual equity curve simulation comparing impulsive vs. disciplined trading, with AI-generated behavioral narratives
- **Tech Stack**: Next.js 16, React 19, Django 5, Django Channels (WebSocket), Supabase PostgreSQL, Upstash Redis, Deriv WebSocket API, Bluesky AT Protocol
- **Delivered full-stack product in 24 hours**: real-time market analysis, behavioral coaching with risk scoring, and social content engine with one-click publishing

Internship Experience

Kuayue Xingkong (Hangzhou) Technology Co., Ltd.   Sep 2025-Present
Large Language Model Algorithm Engineer
- **LLM Application & Optimization**: Prompt engineering, CoT logic; Fine-tuned GPT, Claude, Gemini, DeepSeek; Built RAG with LangChain/LlamaIndex; LoRA/QLoRA/PEFT; Optimized inference (TensorRT, vLLM, DeepSpeed)
- **Agent System Development**: Multi-agent architecture (Planner/Executor/Tool-Using Agent); Agent memory systems; Function Calling/Tool Use integrating APIs, databases, code execution
- **Backend & API**: Python (FastAPI/Flask) + Node.js (Express/NestJS); PostgreSQL/MongoDB/Redis; Async queues (Celery/asyncio); Streaming (SSE/WebSocket)
- **Frontend**: React/Next.js/Vue/TailwindCSS; Chat panels, agent consoles; Real-time streaming & reasoning chain visualization
- **DevOps & Product**: Docker/K8s/Nginx CI/CD; AWS/GCP/Vercel; Prometheus/Grafana; Defined AI product architecture from scratch

Skills

Programming Languages	Python, TypeScript, JavaScript, Java, C, SQL
Backend Development	NestJS, FastAPI, Django, Node.js, Socket.IO, Microservices
Frontend Development	React, Next.js, Redux Toolkit, Tailwind CSS, Streamlit
Database & Cache	PostgreSQL, Supabase, TypeORM, Redis, MongoDB
Deep Learning	TensorFlow, PyTorch, CUDA, GPU Accel., TensorRT, vLLM, DeepSpeed
Computer Vision	DeepFace, FasNet, RetinaFace, MTCNN, Emotion Recognition, Liveness Detection
NLP & LLM	GPT-4, Claude, Gemini, DeepSeek, Deepgram STT/TTS, RAG, LoRA/QLoRA/PEFT
AI Frameworks & Tools	LangChain, LlamaIndex, Ollama, Function Calling, Multi-Agent Systems
Cloud Native & DevOps	GCP (Vertex AI), AWS, Docker, Kubernetes, CI/CD, Git

Project Experience

SynCVE - Real-time Emotion Recognition & Anti-Spoofing System 2025-Present
Core Developer & AI Algorithm Lead
- Integrated **DeepFace** and **FasNet** anti-spoofing model, achieving **85-95% accuracy** for 7 core emotions and efficient liveness detection
- Configured **CUDA 11.8** GPU acceleration for TensorFlow & PyTorch; reduced detection latency from 15s to **1-2s** (**87% improvement**)
- Diagnosed and resolved PyTorch CUDA initialization failure (WinError 127); created environment configuration documentation for team

SynPod - AI-Driven Intelligent Travel Voice Assistant Platform 2025-Present
Core Developer & AI Algorithm Lead
- Integrated **OpenAI GPT-4** (Function Calling) and **Deepgram Nova-2 (STT) / Aura-2 (TTS)** APIs for real-time voice interaction and podcast generation
- Designed production-grade **Token Bucket** rate controller and **FSM** with 5 states for complex generation workflows
- Achieved **<300ms latency** voice interruption via persistent WebSocket, intelligent text chunking, and content prefetching

SynText - AI-Driven Enterprise Real-time Translation SaaS Platform 2024-Present
Project Lead & Chief Architect
- Led NestJS microservices with PostgreSQL and Redis, achieved **millisecond-level API response**
- Self-developed speech-to-text engine; integrated and optimized **DeepSeek AI** for translation; designed subscription tiers for commercialization

Malaysia Smart Tourism AI Assistant 2025-Present
Core Developer & AI Algorithm Lead
- Fine-tuned **Gemini 2.5 Flash** on **Google Vertex AI** for Malaysian context; designed four-stage dialogue engine with multimodal interaction (FastAPI + Streamlit)

AIT Systems - Intelligent Enterprise Management Platform (with Daikin Industries) 2025-Present
Technical Lead
- Designed TypeScript-based microservices + RAG stack with **Ollama** local LLM; RBAC data security; enterprise AI assistant for knowledge retrieval

Certifications

- **Python for Everybody Specialization** -- University of Michigan
- **C for Everyone: Programming Fundamentals** -- UC Santa Cruz
- **Introduction to Networking** -- NVIDIA