TAN WEI HAO

+60-11-11199174
weihaotan.77@gmail.com
Kajang, Selangor, 43000
https://tanweihao0418.github.io/

About Me

Motivated Software Engineer with a strong foundation in AI/ML and a focus on building intelligent systems. Proficient in Python, Java, C#, and PHP, with hands-on experience in machine learning and natural language processing. Passionate about designing and implementing AI-driven applications, particularly in AI agent systems. Highly adaptable, self-taught, and eager to contribute to innovative software solutions that push the boundaries of AI and modern software engineering.

Experience

Fullstack Web Solution Engineer (AI)

TimeTec Cloud

- Architected and developed new HRv2 projects using Vue.js, integrated cloud services including AWS S3 bucket for file storage, Kafka for message streaming, and gRPC framework for high-performance microservices communication.
- Maintained and enhanced legacy .NET Core projects, and designed and implemented API logic/endpoints with a RESTful architecture.
- Implemented Model-View-Controller (MVC) architecture using Angular for structured frontend development and maintainable code organization.
- Managed multi-database architecture: MySQL for legacy systems, PostgreSQL for HRv2 application, and MongoDB NoSQL with vector database for chatbot implementation.
- Utilized Git for version control, collaborating effectively within teams to manage codebase, resolve conflicts, and ensure smooth deployment cycles.
- Enhanced HR solutions by implementing OCR technology for automated data extraction
- Utilized GitHub Copilot and MCP tools including Playwright for accelerated development
- Experienced in fine-tuning large language models using Python with Unsloth (Hugging Face) for accelerated training with QLoRA, targeting Deepseek models for NLP applications.
- Developed an advanced AI-powered chatbot using OpenAI Agent SDK, leveraging Retrieval-Augmented Generation (RAG) and write custom prompt to handle complex, context-aware user queries.
- Implemented a custom knowledge base embedding pipeline and stored vector data in Qdrant, enabling efficient semantic search and response generation.

Software Engineering Intern

Resort World Tech Lab - Genting Berhad

- Developed and trained a poker card detection model using YOLOv5 with PyTorch, optimized for real-time inference in casino games like Baccarat.
- Created a WPF-based Baccarat game using .NET Core, integrating AI-based card recognition for gameplay automation.
- Handled multithreading and UI events in WPF with Dispatcher, async/await, and Task for smooth real-time camera processing.
- Built a reusable .NET Class Library for card detection and data collection, and published it as a NuGet package for production use.
- Performed accuracy benchmarking and performance testing on AI model (FPS, latency, precision/recall metrics).

AI Intern

Footfallcam - Meta Research

- Created an image data collection pipeline using FastAPI, enabling efficient ingestion and processing of visual datasets.
- Built automation scripts using Selenium to extract geographic coordinates and integrate OpenAI (ChatGPT), Google Geocoding, and Bing Maps APIs.
- Conducted testing on ESP32-CAM, analyzing image blurriness and estimating age/gender using cloud-based face recognition APIs (e.g., Face++, Inferdo, Kairos).
- Trained CNN models (ResNet50, MobileNet) with Keras & TensorFlow to detect store open/close status from surveillance images.
- Applied PCA for dimensionality reduction and used SVD and regression analysis with Scikit-learn to model campaign effectiveness and predict sales outcomes.
- Resolved customer support tickets, ensuring quick turnaround and issue resolution.
- Experienced with Agile SCRUM methodologies for project planning and sprint-based development.

Projects

AI-Powered ChatBot Platform

- Developed a comprehensive full-stack chatbot platform with Vue.js 3,Python FastAPI, OpenAI API, and Qdrant vector database.
- Built universal Web Component for easy website integration with one-line script implementation.
- Implemented Retrieval-Augmented Generation (RAG) pipeline with semantic search capabilities, multi-format document processing system, and RESTful API with JWT authentication.
- Built custom web crawler with intelligent content parsing and advanced LLM-based document chunking for enhanced RAG system accuracy.

Real-Time Card Detection System

- Developed a sophisticated real-time card detection system using C#/.NET 8, WPF, and multiple YOLO models (v5/v7/v8).
- Implemented advanced computer vision pipeline with frame buffering for 30 FPS processing, multithreading for non-blocking video capture, and CUDA GPU acceleration for optimized inference performance.
- Achieved 80% reduction in false positives through ensemble confidence threshold optimization and dynamic model switching.

Education

2021

New Era University College

Bachelor In Software Engineering - CGPA : 3.84

Certifications

Microsoft Azure Fundamentals (AZ-900)

Language

- Chinese
- English
- Bahasa Melayu

Computer Skills

- C#/.NET
- Python
- Vue.js/ Angular
- Java
- Javascript/Jquery
- PHP
- mySQL/NoSQL/Postgres/Postman
- Github
- Keras TensorFlow/Pytorch
- Selenium/PlayWright
- Github Copilot with MCP tools
- OpenAI Agent SDK
- AWS Bedrock/ Dify.ai
- Qdrant Vector DB

Interpersonal Skill

- Self and fast learner
- Good time management
- Collaboration and Teamwork
- Good communication skill
- Able to work under pressure.