WOODY CHANG

woodychang891121@gmail.com
LinkedIn | GitHub | AI Agent Blog

PROFESSIONAL SUMMARY

Master of Computer Engineering candidate specializing in **Generative AI** and **Agentic Systems** (LangGraph, Semantic Kernel). Proven ability to design and deploy production-grade LLM applications, including a **Hybrid RAG pipeline achieving 93% precision** and a **FastAPI backend serving 1,000+ active users**. Highly skilled in MLOps and orchestrating complex multi-agent workflows.

EDUCATION

University of Toronto Expected May 2026
Master of Computer Engineering – Data Analysis & Machine Learning
University of British Columbia May 2024
Bachelor of Applied Science – Electrical Engineering

TECHNICAL SKILLS

- **AI Agent & LLMs:** LangGraph, Semantic Kernel, RAG, LangChain, Prompt Engineering, OpenAI APIs, MCP
- **MLOps & Backend:** FastAPI, Docker, Redis, AWS (S3), ETL Processes, Git/GitHub
- **Machine Learning & Data:** PyTorch, Scikit-learn, FAISS, MongoDB Atlas Vector Search, SQL

WORK EXPERIENCE

AI Intern June 2025 – Present
Ret[AI]ling Data, Taipei, Taiwan

- Architected multi-agent frameworks using Semantic Kernel and LangGraph, boosting query resolution accuracy by **25%** and reducing response latency by **40%** through dynamic routing and error recovery.
- Developed a Hybrid RAG pipeline integrated with MongoDB Atlas and OpenAI embeddings, achieving **93% precision** on 2,000+ QA pairs using hybrid semantic retrieval and adaptive re-ranking.
- Deployed a production-grade FastAPI + Docker + Redis backend serving **1,000+ active monthly users**, enabling real-time conversation handling and intelligent response caching with **<300 ms** inference time.
- Implemented graph-based workflow orchestration via LangGraph with state tracking and failover recovery, enabling **99% uptime** and seamless task continuity across multi-agent pipelines.

Software Intern Sept 2022 – Dec 2022
Anyload Weigh & Measure, BC

- Designed predictive inventory management system with automated order processing workflows, reducing manual intervention by **60%** through real-time data synchronization and historical trend analysis.

TECHNICAL PROJECTS

KHH Airport AI Analytics Agent Aug 2025 – Present
Skills: LangGraph, LangChain, FastAPI, Redis, AWS S3, Docker

- Architected a production-grade **multi-agent system (Supervisor Pattern)** coordinating 3 specialized agents (SQL Search, Visualization, Analysis) to process queries across **30K+ records** with intelligent routing.
- Built end-to-end NL→SQL visualization pipeline using LangChain SQL toolkit and AI-powered Plotly generation, automatically converting natural language into executable queries and interactive charts.
- Deployed FastAPI-based LINE chatbot with Redis persistence, async streaming, and real-time anomaly detection.

Natural Language to Data Visualization with MCP July 2025 – Aug 2025
Skills: Model Context Protocol (MCP), LangChain, LangGraph, Docker, SQLite

- Developed end-to-end data visualization app by integrating **Tako MCP server** with LangChain agents, enabling real-time querying and interactive embedded visualizations (GDP trends, polling) with automatic iframe rendering.
- Architected MCP-LangChain integration using **MultiServerMCPClient** to connect agents via streamable HTTP, leveraging MCP prompt primitives for dynamic context-aware generation and secure regex-based visualization parsing.
- Engineered async Streamlit frontend displaying step-by-step tool execution flow through **astream_events v2 API**, implementing persistent multi-turn conversations with AsyncSqliteSaver.

Advanced RAG System with Self-Corrective Retrieval June 2025 – July 2025
Skills: LangGraph, LangChain, MongoDB, OpenAI API

- Architected self-corrective RAG pipeline using **LangGraph state machines** with 5+ conditional nodes, implementing hallucination detection and answer grading via structured LLM outputs (Pydantic schemas).
- Built intelligent query routing workflow that dynamically classifies questions, routes between Vector Store and web search, and validates generations against source documents to prevent hallucinations.
- Engineered custom MongoDB Atlas vector search with semantic filtering using OpenAI embeddings, supporting domain-specific tag-based filtering across 13+ categories (parking/membership policies).