CHUN-JU TAO

(646) 894-7186 • ct3354@nyu.edu • github.com/iridiumtao • iridiumtao.github.io

EDUCATION

New York University, New York, NY — May 2026
MS in Computer Engineering, GPA 3.83
Relevant Coursework: ML, MLOps, Reinforcement Learning, Software Engineering, Network Security

National Taichung University of Science and Technology (NTCUST), Taichung, Taiwan — June 2023
BEng in Computer Science and Information Engineering (CSIE), GPA 3.79
Relevant Coursework: Deep Learning, Algorithms, Data Structures, Computer Networks

SKILLS

Languages: Python, Go, JavaScript (React, Vue), SQL (MS SQL, PostgreSQL), Java, C#, Swift
Cloud & DevOps: Docker, AWS ECS, Terraform, GitHub Actions, Airflow, Prometheus, Grafana, MinIO, Git, Linux
Data & ML: PyTorch, MLflow, LlamaIndex, Lang Chain, LightGBM, SHAP, Streamlit, FastAPI

PROFESSIONAL EXPERIENCE

Data Science Intern, **Micron Technology,** Taoyuan, Taiwan — July 2025 - August 2025
- Architected a production-scale Python pipeline and Streamlit web app for fab-dispatch analysis, processing 2 weeks' logs (33GB) and delivering a self-serve interface for parameter tuning and rich visuals, enabling fast, reproducible studies and broad cross-team adoption.
- Developed an explainable LightGBM simulation proxy with SHAP analysis for lot-level decision tracing, enabling evidence-based simplification of scheduling parameters by quantifying which factors truly drive selection and reducing tuning overhead for production engineers.

Software Developer, **CARITY AI,** Ontario, Canada — May 2024 - August 2024
- Developed and implemented a CI/CD pipeline with GitHub Actions for an LLM-based product, containerizing 4 microservices on AWS ECS, reduced infrastructure costs by 40% and cut deployment time by 70%.
- Delivered a Proof-of-Concept using Retrieval-Augmented Generation (RAG), demonstrating a potential 5x reduction in token usage and influencing the team's future technical roadmap for cost optimization.

Software Engineering Intern, **MoBagel,** Taichung, Taiwan — January 2023 - July 2023
- Re-architected and migrated a legacy Java system to a modern .NET stack (C#, Vue.js) for a critical government asset management system to enhance performance, security, and scalability.

Software Engineering Intern, **Mindtronic AI,** Taipei, Taiwan — June 2022 - September 2022
- Spearheaded a backend migration from Node.js to Go from the ground up to enhance system security and processing efficiency, owning 53 RESTful APIs processing 480k+ weekly entries.

PROJECT

Taigi (Taiwanese-Hokkien) Medical Advising LLM, **New York University,** New York, NY — March 2025 - May 2025
Advisor: Professor Fraida Fund
- Architected a cloud-native MLOps platform for LLM using Terraform for Infrastructure as Code (IaC), and deployed a suite of Docker-based microservices (FastAPI, Gradio, MinIO) to production.
- Orchestrated a Continuous Training (CT) pipeline with Airflow for human-in-the-loop retraining, and established system observability using Prometheus and Grafana.

AI Editor-in-Chief and Virtual News Presenter, **NTCUST, Taichung, Taiwan** — September 2021 - July 2023
Advisor: Professor Jia-Wei Chang
- Engineered the core system infrastructure to resolve critical dependency and versioning conflicts across 5 disparate open-source microservices; designed and implemented a resilient data pipeline using Docker Compose and Flask to ensure system integrity and enable scalable future development.
- Automated the end-to-end deployment process for the entire stack, creating a reproducible, one-command build that slashed manual setup and deployment time by over 80% (from 2 hours to 20 minutes).

HONORS

Emerging Technology Application Award in Fi-Award 2023 by **the 13th International Conference on Frontier Computing**, Tokyo, Japan — 2023
Winner of Better Retail, **Level-Up Society Hackathon**, organized by ShowCode, UK — 2021