CHUN-JU TAO

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

Passionate about creating innovative and scalable solutions in Software. Experienced in designing backend systems, automating CI/CD pipelines, and building cloud-native MLOps platforms. Seeking a backend or software infrastructure role to tackle complex engineering challenges.

EDUCATION

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

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: Algorithms, Data Structures, Computer Networks, Electronic Commerce Security

SKILLS

Languages: Python, JavaScript (React, Vue), Go, Java, C#, Swift, MS SQL, PostgreSQL, C
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.
- Engineered repo-documentation tools for an enterprise application with over one million lines of code using Prompt Engineering with Roo Code Orchestrator, MCP, and Qdrant; produced modular docs and standardized class/method summaries; cut token usage 10x and expected ~3x developer efficiency.

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

- Engineered a critical full-stack system to automate inventory and budgeting for trillions in government assets for the Taiwan Water Corporation, migrating a legacy Java 4 application to a modern .NET stack (C#, MS SQL, Vue.js) to enhance performance, security, and scalability.
- Proactively identified and reported critical security vulnerabilities across legacy and new systems, including SQL injection risks and an exposed database, preventing potential large-scale data breaches.
- Established GitFlow and an Agile-like development model for a 10-person team, fostering a culture of collaboration that improved development efficiency and stabilized team management during a 300% expansion.

Software Engineering Intern, **Mindtronic AI,** Taipei, Taiwan — June 2022 - September 2022

- Spearheaded the backend migration from Node.js to Go, re-architecting and building the new system from the ground up to enhance processing efficiency and system security; Mastered the Go language independently to deliver a robust, production-ready backend.
- Owned the full lifecycle of 53 RESTful APIs in Go, from design and implementation to documentation, proactively identified and eliminated critical SQL injection vulnerabilities across the entire API suite while ensuring the system could reliably process over 480,000 data entries weekly.
- Developed key data-rich features for the React frontend to enable real-time fleet monitoring, delivering complex user-facing functionalities including interactive dashboards, live video streaming, and vehicle trajectory visualization on a map.

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 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.
- Fine-tuned an 8B LLaMA-3.1 into the first Taigi medical advisor using 120K bilingual Q&A pairs with LoRA + mixed-precision on an A100 GPU; tracked all runs in MLflow for full reproducibility.

Loud Plants in Your Area, **New York University,** New York, NY February 2025 - May 2025
Advisor: Professor Dishita Turakhia

- Initiated and led the end-to-end development of a novel iOS application that translates plant bio-acoustic signals into real-time, AR visualizations, defining the project vision and architecting the full technology stack.
- Engineered a custom machine learning pipeline based on academic research to classify plant health. Independently implemented a deep scattering network (ScatNet with Morlet wavelets), extracted Mel-frequency cepstral coefficients (MFCCs), and trained a high-performing SVM classifier for signal analysis.
- Developed a fully functional AR prototype using Swift, RealityKit, and Reality Composer Pro. Owned the entire iOS application development, building custom animated UI overlays that rendered dynamic plant statuses based on live data from the ML pipeline.
- Validated the project's core hypothesis through a user evaluation study that demonstrated the AR interface significantly increased user-plant interaction and emotional connection (mean score increase from 1.67 to 5.33, p=0.018).

AI Editor-in-Chief and Virtual News Presenter, **NTCUST, Taichung, Taiwan** September 2021 - July 2023
Advisor: Professor Jia-Wei Chang

- Led a year-long capstone project from concept to award-winning completion, architecting a fully automated AI pipeline that autonomously generated animated news segments from trending topics.
- 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).
- Pioneered the team's adoption of GitFlow, establishing a structured version control workflow that significantly improved development velocity and collaboration, and served as a foundational experience for implementing Agile methodologies in subsequent professional roles.

HONORS

1st prize & Best Demo, **OpenHCI'25,** presented at **the 11th Annual TAICHI Conference** 2025
Emerging Technology Application Award in Fi-Award 2023 by **the 13th International Conference on Frontier Computing**, Tokyo, Japan 2023
1st prize in **Project Competition** at the College of Information and Distribution Science, NTCUST 2023
Winner of Better Retail, **Level-Up Society Hackathon**, organized by ShowCode, UK 2021