

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: Machine Learning, MLOps, Reinforcement Learning, High Performance ML, HCI

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, Software Engineering, E-Commerce Security, Algorithms, Data Structures

SKILLS

Languages: Python, JavaScript (React, Vue), Go, Java, C#, Swift, MS SQL, PostgreSQL, C
Computer Software and OS: Docker, AWS ECS, Terraform, Git, GitHub Actions, MinIO, PyCharm, macOS, Linux
ML Tooling: PyTorch, LlamaIndex, LangChain, MLflow, Airflow, SHAP, LightGBM, Gradio, Streamlit, Label Studio

PROFESSIONAL EXPERIENCE

Data Science Intern, **Micron Technology**, Taoyuan, Taiwan July 2025 - August 2025

- Designed 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.
- Built 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

- Containerized 4 microservices of an existing LLM pipeline product and automated CI/CD with GitHub Actions on AWS ECS, reducing infrastructure costs by 40% and deployment time by 70%.
- Conducted research in implementing Agentic RAG with LlamaIndex, cutting token usage 5x, boosting model efficiency, lowering costs, and enabling complex question-answering.

Software Engineering Intern, **MoBagel**, Taichung, Taiwan January 2023 - July 2023

- Collaborated with an existing team to develop a full-stack material management system for the government-owned Taiwan Water Corporation, automating budgeting and inventory of assets worth trillions.
- 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.

PROJECT

Taigi (Taiwanese-Hokkien) Medical Advising LLM, **New York University**, New York, NY March 2025 - May 2025
Advisor: Professor Fraida Fund

- 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.
- Built a cloud-native MLOps stack (Terraform GPU/CPU nodes, Docker-based microservices with FastAPI/Gradio/MinIO, Prometheus/Grafana) and developed a doctor-in-the-loop retraining pipeline using Airflow and Label Studio across canary, staging, and production nodes.

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

- Leveraged Google Trends and natural language processing (CKIP Transformers, TextRank, TF-IDF, BERT) to find the trending topic for the week, and researched and designed the approach for compiling news articles
- Developed the system infrastructure and designed a pipeline with Docker and Flask for integrating five open-sourced repos into work for project integrity, data transferring efficiency, and future development possibilities

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 2021