CHUN-JU TAO

(646) 894-7186 | ct3354@nyu.edu | linkedin.com/in/chun-ju-tao-3b1485254/ | iridiumtao.github.io/

EDUCATION

New York University **Sep 2023 - May 2026**
MS, Computer Engineering (GPA: 3.83) *New York, NY*
- **Coursework:** ML, MLOps, High-Performance ML, Reinforcement Learning, Software Engineering, Network Security

National Taichung University of Science and Technology (NTCUST) **Sep 2019 - Jun 2023**
BEng, Computer Science and Information Engineering (CSIE) (GPA: 3.79) *Taichung, Taiwan*
- **Coursework:** Deep Learning, Algorithms, Data Structures, Computer Networks

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

EXPERIENCE

Micron Technology **Jul 2025 - Aug 2025**
Data Science Intern *Taoyuan, Taiwan*
- **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.

CARITY AI **May 2024 - Aug 2024**
Software Developer *Ontario, Canada*
- Automated CI/CD for an LLM-based product, containerizing 4 microservices on *AWS ECS* with *GitHub Actions*, **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.

MoBagel **Jan 2023 - Jul 2023**
Software Engineering Intern *Taichung, Taiwan*
- **Re-architected and migrated a legacy Java system to a modern *.NET stack (C#, Vue.js)*** for a critical government inventory system managing assets worth trillions, enhancing performance, security, and scalability.

Mindtronic AI **Jun 2022 - Sep 2022**
Software Engineering Intern *Taipei, Taiwan*
- **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**.

PROJECTS

Taigi (Taiwanese-Hokkien) Medical Advising LLM **Mar 2025 - May 2025**
New York University *New York, NY*
- 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 **Sep 2021 - Jul 2023**
NTCUST *Taichung, Taiwan*
- **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:** 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