

CHUN-JU TAO

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

Passionate about creating innovative 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

Sep 2024 - May 2026

MS, Computer Engineering (GPA: 3.73)

New York, NY

- **Coursework:** Software Engineering, Human Computer Interaction, ML, MLOps, Reinforcement Learning

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:** 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

EXPERIENCE

A Day - AI-Powered Personal Management iOS App

May 2025 - Present

Co-Founder

New York, NY

- Architected a production-grade MVVM application with SwiftData and latest SwiftUI features, decoupling the service layer to ensure strict separation of concerns across 7 interconnected data models and 60+ source files.
- Engineered an AI-augmented development workflow (Cursor/Claude) with custom MCPs and skills for context management, accelerating delivery velocity by 4x while strictly adhering to TDD cycles to eliminate regression risks.
- Established a CI/CD Pipeline using Xcode Cloud with a rigorous QA framework using Swift Testing, implementing 700+ unit and integration tests to guarantee architectural integrity and regression prevention across ViewModels and data flows.

New York University

Jan 2026 - May 2026

Course Assistant

Taoyuan, Taiwan

- Built a 5-microservice ML safeguard lab (FastAPI, Docker Compose) routing requests through two custom MobileNetV2 guard models and a Food11 classifier via short-circuit logic; generated synthetic edge-case data and benchmarked median/p95/p99 inference latency to quantify the cost of each safety layer.
- Reviewed end-to-end MLOps systems from 30 teams (110 students) on Chameleon Cloud across design, code, and live demos, assessing them against production criteria: experiment tracking, automated retraining, drift monitoring, secrets hygiene, and container reproducibility; delivered structured feedback that teams used to remediate issues such as data leakage and shortcut/proxy learning.

Micron Technology

Jul 2025 - Aug 2025

Data Science Intern

Taoyuan, Taiwan

- **Architected a production-scale Python pipeline** and self-serve Streamlit app over **33GB of bi-weekly fab-dispatch logs**, surfacing key scheduling drivers via an **explainable LightGBM/SHAP model** for reproducible, evidence-based tuning adopted across teams.
- Engineered repo-documentation tools for **an enterprise application with over one million lines of code** using Prompt Engineering and sub-task agents; **cut token usage 10x** and projected **~3x gain in developer efficiency**.

CARITY AI

May 2024 - Aug 2024

Software Developer

Ontario, Canada

- **Automated CI/CD for an LLM-based product** by containerizing 4 microservices on AWS ECS with GitHub Actions, **reducing infrastructure costs by 40%** and **cutting deployment time by 70%**.
- **Engineered a cloud-native migration** from EC2 to AWS ECS, utilizing k6 and Grafana to benchmark performance and right-size hardware resources, ultimately **reducing infrastructure costs by 40%**.
- Delivered a **Retrieval-Augmented Generation (RAG)** proof-of-concept showing a potential **5x token-usage reduction**, shaping the team's cost-optimization roadmap.

MoBagel

Jan 2023 - Jul 2023

Software Engineer

Taichung, Taiwan

- Engineered a **full-stack material management system** for the Taiwan Water Corporation (national utility, 7M+ households), migrating its legacy Java application to automate inventory and budgeting over millions of historical records.

- **Introduced GitFlow and an Agile workflow** that scaled the engineering process as the team tripled from 10 to 30, enabling a smooth transition recognized by the supervising manager.
- Discovered and **reported critical security flaws**, including SQL-injection vectors and a publicly exposed, unauthenticated test database that risked a large-scale breach of decades of records.

Mindtronic AI

Jun 2022 - Sep 2022

Software Engineering Intern

Taipei, Taiwan

- **Rebuilt the entire backend in Go from a legacy Node.js service**, owning 53 RESTful APIs end to end (design, implementation, and SQL-injection hardening) while reliably processing 480K+ 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.

PROJECTS

NYU Marketplace | Practical Software Engineering

Sep 2025 - Dec 2025

Advisor: Professor Gennadiy Civil (Adjunct, Senior Engineering Manager at Google), New York University

New York, NY

- **Architected an end-to-end CI/CD pipeline** using *Travis CI* and *AWS Elastic Beanstalk*, enforcing strict merge gates (linting, **85%+ test coverage**) that automated deployment and ensured immediate public availability upon PR approval.
- Engineered scalable *RESTful APIs* and real-time *WebSocket* features using *Django 5* and *Channels*, utilizing *Cursor* to **accelerate development velocity by ~3x** while enforcing system design standards as the lead code reviewer
- Developed a high-performance web app using *React 19* and *Vite*, pioneering an **AI-assisted workflow** by leveraging *Figma AI* for rapid wireframing and design-to-code implementation.
- **Led the engineering team using Agile Scrum** methodologies, orchestrating weekly sprints and backlog grooming via *ZenHub*; established a rigorous *GitHub* forking workflow with mandatory peer reviews to maintain high code integrity.

Privacy-First AI Smart Lamp for Ephemeral Night Conversations - Oblivilight

Jul 2025 - Aug 2025

OpenHCI'25, the 11th TAICHI Conference

Taipei, Taiwan

- **Led a user-centric design** process from research to prototype, identifying key user needs for tangible, privacy-preserving "forgetting mechanisms" in AI companions through 11 user interviews and secondary research.
- **Architected a full-stack proof-of-concept integrating an LLM** for conversation, emotion analysis, and a multi-modal interface with voice (*Whisper/TTS*) and gesture controls.
- Designed a **novel interaction model** that visualizes emotional sentiment as colored light and externalizes digital conversations into physical artifacts via a thermal printer, directly addressing AI data permanence anxiety.

Taigi (Taiwanese-Hokkien) Medical Advising LLM

Mar 2025 - May 2025

Advisor: Professor Fraida Fund, 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*.
- 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**.

Real-Time Plant Health & Mood Visualization - Loud Plants in Your Area

Feb 2025 - May 2025

New York University

New York, NY

- 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

Sep 2021 - Jul 2023

NTCUST

Taichung, Taiwan

- **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, Taipei, Taiwan, 2025
- **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