

# Jay Chiehen Liao

+886-988818967 | jay.chiehen@gmail.com | linkedin.com/in/jc-liao | github.com/jayenliao

5+ years of experience in **Data Science & Machine Learning** • Proficient in **Python** & **R** programming • 10+ **AI/ML competitions**  
7+ years of experience in **Data Analysis & Statistics** • **Git**, **Linux**, **GPU**, **GCP** • 10+ academic presentations in various domains •  
Skilled in communication, model explanation, and problem-solving • **Interdisciplinary** backgrounds

## WORK EXPERIENCE

### FAZZ (FINTECH STARTUP, Y COMBINATOR ALUMNI)

#### DATA SCIENTIST (FULL-TIME)

Sep 2022 - Sep 2023 | Taipei, Taiwan (Hybrid)

- Implemented the system of transaction anomaly alerts based on the time series analysis.
- Strengthened the anomaly detection model and enhanced the true positive rate by 25%.
- Headed the project of identifying micro-loan fraud patterns with graph analysis.
- Developed and maintained detection models for fraudulent transactions with cryptocurrencies.

### NETWORKS & AI LAB, NCKU

#### RESEARCH AND TEACHING ASSISTANT (PART-TIME)

Aug 2020 - present | Tainan, Taiwan (Hybrid)

- Research topics: Machine learning with graphs, graph neural network ( GNN ), deep learning for tabular data.
- Managed Linux servers with GPUs.

### HEALTH BEHAVIOR LAB, NCKU

#### RESEARCH ASSISTANT (PART-TIME)

Jan 2017 - Jul 2020 | Tainan, Taiwan

## EDUCATION

### RENNES SCHOOL OF BUSINESS (RSB)

#### MASTER IN DATA & BUSINESS ANALYTICS (GPA=4.25/5)

Sep 2021 - Apr 2022 | Rennes, France

Graduation Project: *Identifying Key Features of Standard Essential Patents in Information and Communication Technologies with Explainable Machine Learning*

### NATIONAL CHENG KUNG UNIVERSITY (NCKU)

#### MASTER IN DATA SCIENCE (GPA=4.25/4.30)

Aug 2020 - Aug 2022 | Tainan, Taiwan

- The 1st Place, Academic Performance of The 1st Year
- Thesis: *Learning Graph Structures from Tabular Data for Downstream Classification Tasks* (awarded by Taiwanese Association Artificial Intelligence and TOPCO Scientific)

#### DOUBLE BACHELOR IN STATISTICS AND PSYCHOLOGY

Sep 2015 - June 2020 | Tainan, Taiwan

## SKILLS

**LANGUAGES** **ENGLISH** Professional working proficiency (IELTS 7.5, R=9) **MANDARIN/CHINESE** Native **FRENCH** Elementary

**TECHNICAL SKILLS** **PROGRAMMING** Python • R • Ruby on Rails **MARKUP LANGUAGES** Markdown • R Markdown •  $\text{\LaTeX}$   
**VERSION CONTROL** Git **DATA MANIPULATION** NumPy • Pandas • dplyr • tidyr • SPSS • MS Excel  
**DATA VISUALIZATION** ggplot2 • Matplotlib • Seaborn **DATA MANAGEMENT** MySQL • Big Query • metabase  
**DEVELOPMENT ENVIRONMENT** VS Code • Jupyter • RStudio **OPERATION SYSTEMS** macOS Intel/M1 • Linux • Windows  
**MACHINE LEARNING** PyTorch • Scikit-Learn **GRAPH ANALYSIS** NetworkX • PyTorch-Geometric **CLOUD COMPUTING** GCP

## KEY PROJECTS & AWARDS

### TabGSL: Graph Structure Learning for Tabular Data Prediction

Jay Chiehen Liao AND CHENG-TE LI (2023)

Presented a GNN framework with graph contrastive learning and a transformer-based feature extractor to enhance tabular data prediction by simultaneously learning instance correlation and feature interaction.

### Graph Neural Networks for Tabular Data Learning

TUTORIAL AT THE WEB CONFERENCE (WWW) 2023, AUSTIN, TX

TUTORIAL AT THE IEEE INTERNATIONAL CONFERENCE ON DATA ENGINEERING (ICDE) 2023, ANAHEIM, CA

### Predictive Modeling for 14-day Unplanned Hospital Readmission Risk by Using Machine Learning Algorithms

BMC MEDICAL INFORMATICS AND DECISION MAKING (2021)

- Constructed ML models with promising performance to predict patients' readmission.

- Figured out influential features with model explanations.

**MY ROLES:** Co-first author of the paper • Model constructor

**SKILLS USED:** Python • Catboost • Feature Engineering

### A Mobile App (UPrEPU) to Monitor Adherence to Pre-Exposure Prophylaxis in Men Who Have Sex with Men: A Protocol for a User-Centered Approach Mobile App Design and Development

THE JMIR RESEARCH PROTOCOLS (2020)

**SKILLS USED:** Pseudo-code • Python • R

### TOP 1 out of 401 teams

#### CHUNGHWA POST BIG DATA COMPETITION 2019

- Established ML models and proposed commercial solutions from the big data (>30GB) to reduce failed delivery rate by 65%.

- Evaluated the proposed system's ML performances, flexibility, and estimation of annual cost reduction (e.g., USD 376K in Taipei City.)

**MY ROLES:** Leader • Model constructor • Presenter

**SKILLS USED:** Python • R • GBDT • Visual machines

### TOP 3 out of 859 teams

#### RECOMMENDATIONS OF CREDIT CARD SPENDING

CATEGORIES 2021 (E.SUN BANK)

**SKILLS USED:** Python • Transformer • GBDT

### TOP 25%

#### AI CUP FALL 2020 - MEDICAL INFORMATION

DE-IDENTIFICATION (DEPT. OF COMPUTER SCIENCE AND INFORMATION ENGINEERING, NCKU)

**SKILLS USED:** Python • NLP • Transformer

### EXCELLENT AWARD

CERTIFICATE OF STATISTIC ANALYSIS (THE CHINESE APPLIED STATISTICS ASSOCIATION)