

# Wei-Han (Zoe) Hsu

Brooklyn, NY 11201 | [zoe70416@gmail.com](mailto:zoe70416@gmail.com) | (332) 249-9798 | [Website](#) | [LinkedIn](#) | [GitHub](#)

## EDUCATION

### New York University

*M.S in Data Science* (GPA: 3.7/4.0)

New York, NY

Sep 2022 – May 2024

### Taipei Medical University

*B.S in Medical Laboratory Science and Biotechnology* (GPA: 3.9/4.0)

Taiwan

Sep 2016 – Jun 2020

**Core Coursework:** Data Science, **Probability and Statistics for DS**, Linear Algebra, **Natural Language Processing**, **Machine Learning**, **Big Data**, **Deep learning**, **Deep Learning in Medicine**, **Computer Vision**

## WORK EXPERIENCE

### New York University IT - Anomaly Detection System for Network Security

New York

*Machine Learning Engineer- Student Researcher*

Sep 2023 - Present

- Built a high-performing **unsupervised Pytorch Autoencoder in Python**, achieving a 73% F1 score for **network anomaly detection** associated with malicious behavior.
- Develop scripts for the **ETL process**, ingesting **unstructured** network data through stream processing.
- Implemented **MLOps** best practices with MLflow, streamlining the machine learning workflow.
- Deployed the best-perform model as the **real-world AI solution** following MLOps best practice.

*Data Scientist - Student Researcher*

Feb 2023 - Aug 2023

- Performed **exploratory data analysis**, **preprocessing and feature engineering** on a 100GB dataset.
- Built/Tested different **anomaly detection** algorithms (Autoencoder and Isolation Forest).
- Use **Docker and Kubernetes** to deploy and manage containerized applications.

### Medidata, a Dassault Systèmes company - MedidataAI Department

New York

*Commercial Analytics Intern*

May 2023 - Aug 2023

- Leveraged analytical expertise** to optimize marketing strategies for pharmaceutical companies (customers).
- Created impactful **data visualizations** via **Tableau** to simplify complex information for stakeholders.
- Collaborated cross-functionally** to decode intricate data and extract comprehensive insights.
- Employed **SQL** for efficient data extraction and transform unstructured data into **structured data**.
- Leveraged **tree-based marketing machine learning** models for targeted customer engagement.

### aetherAI- Medical device startup

Taiwan

*Medical Technician/ Medical Data Management*

Jul 2018 – Jun 2022

- Conducted **SQL analysis** on 1M labeled data, elevating prediction model precision.
- Pioneered aetherAI's HEMA project and oversaw clinical trial data management and familiar with **ICD-10**.
- Collaborated across departments to facilitate clinical validation processes for FDA approval.

### Sleep Center of Shuang-Ho Hospital

Taiwan

*Research Assistant*

Aug 2020 – Mar 2021

- Undertook the **preprocessing and statistical analysis of medical data (Python and SPSS)**
- Conduct research on deploying machine learning on **multiclassification** of apnea subtypes
- Achieved over **80% accuracy** in using multiple **machine learning** (including Random Forest) using **sklearn**.
- Improved internal clinical decision-making using the model and **co-authored** an article in **Diagnostics, 2021**.

## PROJECTS

### Life Expectancy Prediction | Python, AWS, S3, Sagemaker

- Uploaded and updated the dataset on **AWS S3** for efficient storage.
- Utilized **XGBoost regression** using **Sagemaker** to train a predictive model, obtaining a remarkable 90% r-squared score for life expectancy forecasting.

### Frequent Item Co-Occurrence Analysis | Hadoop, Map-reduce

- Ensured data integrity by employing **data processing** techniques.
- Develop a multi-stage **map-reduce** algorithm to compute frequently co-occurring items in grocery shopping baskets

## SKILLS

**Libraries:** NumPy, Pandas, ML framework (PyTorch, TensorFlow/Keras, sci-kit learn)

**Programming Languages:** Python, SQL, R

**Tools:** AWS, Spark, Git, Docker, Kubernetes, Tableau, Statistical software (SPSS)