

Wei-Han (Zoe) Hsu

332-249-9798 | zoesj3xu4@gmail.com | [linkedin.com/in/weihanz-hsu](https://www.linkedin.com/in/weihanz-hsu) | zoe70416.github.io/

EDUCATION

New York University

Master of Science in Data Science (GPA: 3.7/4.0)

New York, NY

Sep. 2022 – May 2024

Taipei Medical University

Bachelor of Science in Medical Laboratory and Medical Technology (GPA: 3.9/4.0)

Taipei, Taiwan

Aug. 2016 – June 2020

EXPERIENCE

Data Scientist / Machine Learning Engineer

Oct 2024 – Present

Lasa Health

NY

- Built an ML-based clinical support system for 26 chronic pelvic pain conditions using XGBoost and ensemble models, achieving 0.85 AUROC scores and targeting a 30% reduction in diagnosis time.
- Built an automation bot using Selenium to extract standardized medical codes from public sources, cutting data extraction time from one month to three days and significantly accelerating project timelines.
- Preprocessed/analyzed 100 billion EHR records from Mayo Clinic using BigQuery SQL and Python, applying NLP to improve feature extraction, increasing model accuracy by 5%.

Data Scientist

Feb 2023 – May 2024

New York University IT

New York, NY

- Led the development of a TensorFlow Autoencoder-based anomaly detection system for the NYU research network, achieving a 73% FI score and adapting innovative techniques to address network issues.
- Conducted data mining, exploratory analysis, and feature engineering on a 100GB dataset using PySpark on HPC systems, effectively transforming raw data into actionable insights
- Engineered ETL scripts for ingesting unstructured network data, streamlining data gathering and preparation processes for subsequent modeling.
- Implemented MLOps best practices for deployment using MLflow, Docker, and Kubernetes, thereby facilitating the transition from prototype to production.

Data Science / Commercial Analytics Intern

May 2023 – Aug 2024

Medidata Solutions Inc.

New York, NY

- Engineered a Tree-based Marketing Mix Model (Random Forest) to optimize campaign strategies (including phone, in-person, and email outreach) projecting a 30% increase in sales to target customers.
- Implemented A/B testing methodologies to evaluate and enhance sales team performance, collaborating cross-functionally to extract comprehensive insights.
- Utilized SQL, Snowflake, PySpark, and AWS (S3, SageMaker) for data analysis and model development, creating Tableau dashboards to communicate complex information to stakeholders.

Data Science Research Assistant

Aug 2020 – May 2021

Sleep Center of Shuang-Ho Hospital

Taipei, Taiwan

- Engineered a multi-classification model for sleep apnea subtypes, reducing clinical decision-making time from 2-3 months to 3 weeks, enhancing patient care efficiency. And co-authored a publication in Diagnostics (2021).
- Achieved 80 accuracy using ML algorithms (Random Forest, LR, SVM, etc.) with scikit-learn.

SELECTIVE PROJECTS

10-K Analyzer with RAG Chatbot | Python, LLM, RAG, HuggingFace

Feb 2025 – Apr 2024

- Built a financial analysis tool combining SEC 10-K filings with Retrieval-Augmented Generation (RAG), leveraging LLMs to deliver a context-aware chatbot. Integrated NLP techniques to extract key financial metrics.

TECHNICAL SKILLS

Languages: Python, SQL, R, JavaScript, Bash

Frameworks: Pandas, Numpy, Matplotlib, Scikit-Learn, Spark, PySpark, TensorFlow, Pytorch, Hugging Face

Tools: GCP, AWS (S3, EMR, EC2, Sagemaker), HPC, Hadoop, Docker, Kubernetes, Tableau, Git

Machine Learning & Modeling: Statistical Modeling, Anomaly Detection, Large Language Modeling