Alan Wang

858-539-5564 | Authorized to work in US | zew@seas.upenn.edu | zew013.github.io

Education

University of Pennsylvania — *M.S. in **Data Science*** — 2024 - 2026

University of California San Diego — *B.S. in **Data Science**, B.S in **Applied Math & Economics*** — 2020 - 2024

***Relevant Coursework*:** Algorithms, Statistics, Data Mining, Forecasting, Visualization, Machine/Deep Learning, Computer Vision, Database Recommender System, Big Data Analytics, Optimization, Time Series, Regressions, Large Language Model.
***Programming language*:** Python (NumPy, Pandas, Matplotlib, Scikit-Learn, PyTorch, Spark, Dask, etc.), SQL, R, Java
***Others*:** Tableau, Git, AWS, Unix shell scripting, Microsoft Office Suite

Professional Experience

***Data Modeling Intern* | TE Connectivity** — **May. 2024 – Aug. 2024**

- Launched an Auto-ML pipeline on AWS SageMaker to improve cost estimations. Reduced cost estimation time from hours to 10 minutes, allowing cost analysts to focus on strategic decision-making rather than manual estimates.
- Designed and automated cost models using Excel-VBA, integrated with cloud database for real-time updates, enhancing estimation accuracy with regression models and resulting in a 70% reduction in time spent on calculations for procurement.

***Machine Learning Intern* | Grant Street Group** — **May. 2023 – Aug. 2023**

- Proposed and developed a machine learning powered transaction monitoring system using models like Random Forest, ARIMA, Prophet, and Temporal Fusion Transformer for self-supervised anomaly detection.
- Led a team of four to implement a new system leveraging the Prophet time series model, which improved fraud detection by 30% and anomaly detection by 40% over the previous static threshold-based system.
- Proficient at using SQL, Tableau, Python, and Machine Learning to deliver insights and dashboards for daily operations.

***Machine Learning Head Teaching Assistant* | Halicioglu Data Science Institute, Penn Engineering** — **Mar. 2023 – Present**

***CSE-PACE Program Designer* | UCSD CSE Department |** *Webpage* — **May. 2022 – Sep. 2022**

- Implemented the NSF-supported peer-mentor project as part of the CS curriculum that supported over a thousand students.

Research Experience

***Data Science Capstone Owner* | Prof. Alex Cloninger** — **Oct. 2023 – Apr. 2024**

Diffusion Models for Image and Data Generation | *GitHub, Webpage*

- Investigated how scene representations are generated during the diffusion process. Demonstrated that 3D properties are learned early in the denoising stage before human visual recognition. Achieved by inserting probing classifiers into self-attention blocks.

***Research Assistant* | Rappel Laboratory** — **Feb. 2023 – Mar. 2024**

Image Segmentation and Propagation Analysis Pipeline for cAMP Waves in Cell Aggregation Stage | *Demo*

- Developed a two-stage Python workflow that segments images and videos, extracts signals from cAMP waves, applies unsupervised clustering algorithm for data cleaning, and constructs velocity vector fields.
- Optimized and parallelized the code, reducing processing time from 50 minutes to 4 minutes.

***Research Assistant* | Prof. Richard Carson & Prof. Dale Squires** — **Dec. 2021 – Dec. 2022**

Data-Driven Analysis of Ethical Preferences in UN Membership Policies

- Developed an ETL data mining pipeline using Python and AWS to large dataset spanning 70 years. Sharply improved processing speed and accuracy, especially for handwritten records. Performed fixed-effect regression analysis.

Leadership & Projects

Data Analyst, Tech VP | Lumnus Consulting (Student Enterprise) | *Webpage* — **Nov. 2021 – Feb. 2023**

- Led team projects by building machine learning models and creating visualizations, facilitating clear communication of data insights. Launched and maintained the company website using React.js and Heroku.
- Organized events, including data analysis projects and alumni talks, fostering collaboration and knowledge sharing.

Deep Learning Projects

Language Intention Classification Using BERT & Model Compression for Web Deployment | *GitHub*, *Webpage*

Using CNN and LSTM models for Image Captioning on COCO Dataset | *GitHub, Report*

Building Neural Network from Scratch & Building Transformer in PyTorch | *GitHub, Report*