Alan Wang

Open to relocate | 858-539-5564 | zew@seas.upenn.edu
Personal Website: zew013.github.io
Linkedin: linkedin.com/in/alanwang-ucsd
GitHub: github.com/zew013

Education

University of Pennsylvania | 2024 - 2026

M.S. | ***Data Science***

University of California San Diego | GPA:3.99/4.0 | Major GPA: 4.0/4.0 | 2020 - 2024

B.S. | *Primary Major* - ***Data Science*** | *Double Major* – ***Applied Math***

Data Science Courses: Data Structures & Algorithms, Web & Data Mining, Visualization, Machine Learning, Deep Learning, Computer Vision, Image Processing, Recommender System, Network Science, Graph Theory, Database, Big Data.

Math Courses: Calculus, Linear Algebra, Numerical Analysis, Optimization, Probability, Statistics, Hypothesis Testing, Economics, Regressions, Time Series, Operations Research.

Research Experience

Data Science Capstone **| Prof. Alex Cloninger & Prof. Rayan Saab | Oct. 2023 – Apr. 2024**

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

- Reproduced denoising diffusion implicit models and developed experiments upon the latent diffusion repository in Python.
- Our goal was to elucidate whether and how the scene representations such as depth are generated in the diffusion process.
- Used probing classifier to show that an LDM can learn 3D properties like depth from 2D images. Found that 3D information was generated in the early denoising stage before human eyes can recognize.

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

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

- Experimented with computer vision methods and deep learning models to segment videos of cAMP waves.
- Developed a two-stage Python workflow. Used image and signal processing techniques to segment videos and extract the cell signals. Creatively grouped waves using machine learning methods (DBSCAN) and calculated velocity vector fields by performing least squares on activation maps through 3-by-3 kernels.
- Presented my work during lab meetings. Optimized and parallelized the code, reducing average processing time from 50 minutes to 4 minutes. Trained other lab members to use my program.

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

Data-Driven Analysis of Ethical Preferences in UN Membership Policies

- Reviewed documents and collected large dataset focusing on membership assessment policies.
- Developed an ETL pipeline using Python and AWS to prepare dataset spanning 70 years from the UN Digital Library. Sharply improved processing speed and accuracy, especially for handwritten records dating back to the 1950s.
- Applied a Fixed Effect OLS on the processed panel dataset, and deduced the ethical preferences which could serve as a standard to guide and facilitate multilateral cooperation by reducing conflicts and information costs.

Different Conditional Logit Assumptions via Monte Carlo Simulation

- Designed experiments and performed Monte Carlo simulation to evaluate the results under different assumptions.
- Coded an automated program to analyze statistical metrics from experiments and to create academic graphs in Stata.

Professional & Teaching Experience

Data Modeling Intern **| TE Connectivity | Jun. 2024 – Present**

- Reverse-engineered the cost models and learned them using machine learning (random forest and regressions) in Python.
- Upgraded Excel calculators by connecting them to a cloud database, automating manual calculations with VBA, updating them with the most recent data, and enhancing their accuracy using regression models.
- Excellent understanding of business operations, with proficient knowledge in developing models, analyzing data, and compiling reports to drive business solutions.

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

- Proposed an initiative for the ML-powered transaction monitoring system. Experimented with multiple ML models (Random Forest, ARIMA, Prophet, Isolation Forest, etc.) to build an unsupervised anomaly detection solution.
- Led a dynamic team of four and orchestrated our different skillsets. Our new system utilized the Prophet time series model and was tested in the Exasol database, enhancing fraud and anomaly detection by 40% and 70% respectively compared with the previous static threshold-based system. Won the opportunity to present at Summer Company Meeting.
- Experienced at using SQL, Tableau, Python, and Machine Learning to deliver insights and dashboards for daily operations.

***Teaching Assistant* | Halıcıoğlu Data Science Institute, Penn Engineering | Mar. 2023 – Present**

- Automated the grading process by developing test cases and grading systems on Python and Jupyter Notebook.
- Assisted over 400 students by conducting office hours, leading labs and discussions, answering questions on the forum, as well as creating and grading assignments and exams.

***CSE-PACE Program Designer* | UC San Diego CSE Department | May. 2022 – Sep. 2022**

NSF-supported Project: Inclusive longitudinal peer mentoring for community building and retention. | *Webpage*

- Addressed issues that disproportionately affect students from historically marginalized groups by crafting courses that prioritized communication and individual-peer relationships over sheer knowledge acquisition.
- Designed seven courses for the program, which has been integrated as part of the computer science curriculum.

Leadership & Project Experience

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

- Built machine learning models and created visualizations for various team projects.
- Built connections with local businesses. Invited alumni and upperclassmen to talk about their academic and career experiences. Organized events such as data analysis projects, presentations, and case-study competition.
- Developed and maintained our website using React.js and Heroku.

Deep Learning Projects

eBay 2023 ML Competition | May. 2023 – June. 2023

- Achieved a weighted F1 score of 0.86 on the Named Entity Recognition task by finetuning on a pretrained DeBERTa model. Gained a deeper understanding of text data processing, tokenization, and BERT downstream tasks.

Amazon Massive Intent Dataset Classification & Full Stack Deployment | Dec. 2022 – Jan. 2023 | GitHub, Webpage

- Used BERT as encoder and a Neural Network as the decoder to classify text intentions.
- Self-studied Knowledge Distillation. Used the trained BERT-NN as the teacher model and BiLSTM as the student model to compress the model size from over 439MB to 70MB while maintaining a similar accuracy level.
- Deployed the compressed model on my own webpage using Flask-RESTful, Amazon S3, and Heroku.

CNNs and LSTMs with PyTorch for Image Captioning on COCO Dataset | Nov. 2022 – Dec. 2022 | GitHub, Report

Neural Network from Scratch | Sep. 2022 – Oct. 2022 | GitHub, Report

- Implemented a neural network in Python and coded backpropagation, mini-batch gradient descent, and cross-validation using NumPy from scratch. Added early stopping, momentum, and L1 & L2 regularization to enhance the model. Conducted experiments with sigmoid, tanh, ReLU, and softmax as activation functions.

Data Analysis Projects

Analysis of Power Outage Status in the Continental U.S.

- Went through the full process of questioning, data gathering, data mining, explorative data analysis, missingness assessment, hypothesis test, baseline model, final scikit-learn ML pipelines, fairness analysis, and visualization.

Professional Skills

Programming:

Python, R, Stata, SQL, Java, HTML, JavaScript, CSS, LaTeX, MATLAB

Libraries & Frameworks:

NumPy, SciPy, Pandas, sklearn, Matplotlib, Plotly, PyTorch, OpenCV, bs4, D3.js, Dask, GeoPandas, Spark, REST API, etc.

Technology:

Microsoft Office Suite, Git & GitHub, AWS, Basic shell/bash scripting, K8s, vim, Tableau, Heroku, ImageJ, Exasol, Jupyter.