Alan Liu Wu

Livingston, NJ | +1 862-485-9127 | alanlwu@seas.upenn.edu | linkedin.com/in/alanlwu | github.com/alan-w25 | alan-wu.me

Education

University of Pennsylvania — Philadelphia, PA
Master of Science in Engineering in Data Science — *Expected May 2026*

Rutgers University- New Brunswick — New Brunswick, NJ
Bachelor's of Science in Computer Science and Data Science, minor in Statistics — *Sep. 2021 – May 2024*
Activities: HackRU (Team Lead), Rutgers Men's Ultimate Frisbee (President)

Technical Skills

Languages: Python, SQL, R, Java, JavaScript, HTML, CSS
Libraries/Frameworks: Numpy, Pandas, Pytorch, Matplotlib, Seaborn, Sci-kit Learn, Tableau
Certifications: Machine Learning Specialization (Deeplearning.ai), IBM Data Science Professional Certificate

Experience

Graduate Research Assistant — July 2024 – Present
Wharton School of Business — *Philadelphia, PA*

- Scraped and inserted over 12 million records of google review data into PostgreSQL database
- Employed **OpenAI API** on 500 samples to benchmark GPT-4o on unifying and multi-modal representation learning tasks including modality fusion and visual question answering
- Developed sentiment analysis and visualization techniques to extract multi-modal relationship insights from 3.5 million restaurant reviews with python

Undergraduate Research Assistant — Aug 2023 – May 2024
Rutgers Cancer Institute of New Jersey — *New Brunswick, NJ*

- Scraped data from the Human Microbiome Project using **beautiful soup** and used principal components analysis and non-negative matrix factorization to reduce 2,000 microbes across 18 distinct body sites
- Utilized **Python**, **matplotlib**, **seaborn** to build heat maps, cluster diagrams, and survival curves to measure feature statistical significance. Hypothesis tested features with log-rank tests and Kaplan-Meier curves
- Analyzed viability of new DNA sequencing technique for lung cancer prognosis to model patient survival times and outcomes with survival support vector machines (ssvm), and random survival forests on 108 cancer patients
- Tuned L2 strength, learning rate, and kernel of ssvm model with stratified k-fold cv, yielding 0.83 C-Index

Software Engineer Co-Op — Jul 2023 – Dec 2023
ASICS Digital — *Boston, MA*

- Automated previously manual process to verify metadata overlap with **Python** scripts to inform key design decisions during migration to a new Customer Data Platform. Verified over 2 million records
- Authored and optimized **SQL** queries, assisting in GDPR-user compliant deletion for over 1.7 million customers
- Utilized **Postman API** to A/B test time thresholds for email delivery in a global marketing campaign

Projects

Kitchen Pantry Manager | *Next.js, MaterialUI, Firebase* — Aug 2024

- Developed a kitchen pantry inventory manager that adds, deletes, and updates pantry items and categories
- Implemented email and google authentication, shipping product to 100 users

Chinese MNIST Digit Detection App | *Next.js, Fastapi, Pytorch, Docker, Azure* — May 2024 – June 2024

- Deployed full stack web application to accomplish classification of Chinese hand-written digits
- Built custom CNN architecture that achieved validation accuracy of 0.92
- Containerized app with **docker** compose and used **Azure** web services and container instances for deployment

Online Retail Analysis | *Python, Pandas, Numpy, Matplotlib, Sci-kit Learn* — Nov 2023 – Dec 2023

- Conducted data pre-processing on UCI online retail dataset to ensure quality and consistency, implementing RFM (Recency, Frequency, Monetary) analysis to segment 4,000 customers based on purchase behavior
- Developed a Random Forest predictive model, fine-tuned with hyperparameter optimization, achieving a 0.76 R-squared value. This robust model was trained on a dataset with 100,000+ transactions

Clinical Heart Failure Bayesian Analysis | *R, RMarkdown, rstanarm, ggplot2* — Nov 2023 – Dec 2023

- Executed bayesian logistic regression techniques, implementing an MCMC simulation-based model to predict patient survival from a dataset of 299 heart failure cases
- Achieved 0.87 accuracy using Bayesian logistic regression model, incorporating prior knowledge
- Applied posterior prediction and prior prediction checks to ensure validity of statistical assumptions

Interests

Hobbies: Cooking, Ultimate Frisbee, Reading, Golf, Learning New Languages, Running