

ALAN JIAN

San Francisco, CA · New York, NY
(516) 474-0309 | alanjian131@gmail.com | <https://alanjian.github.io/>

Education

University of California – Berkeley
Master of Information and Data Science

Dec 2023
Overall GPA: 4.00

University of California – Berkeley
Bachelor of Arts in Data Science - (concentration in Molecular Cell Biology)

May 2022
Overall GPA: 3.62

Technical Skills

- **Languages (and associated packages):**
 - Python (PyTorch, Pandas, NumPy, Scikit-Learn)
 - R (Stargazer, ggplot, Tidyverse)
 - SQL (MySQL, PostgreSQL, NoSQL)
 - Java
- **Exploratory Data Analysis:**
 - Data Cleaning via Python, R, SQL
 - Data Visualization via Matplotlib, ggplot, seaborn
- **Model Building:**
 - Machine Learning via sklearn, PyTorch, Tensorflow
 - NLP via Huggingface, SpaCy, NLTK
 - Bayesian Inference via PyMC
- **Data Engineering:**
 - Relational Databases (PostgreSQL)
 - NoSQL Databases (GCP Cloud Firestore, MongoDB)
 - Containerization via Docker

Relevant Projects

Real-time Electricity Demand Forecasting (with Uncertainty Quantification) Aug 2023 – Dec 2023

- Utilizing ML-based and classical forecasting approaches (i.e LSTM, Prophet, TBATS) in conjunction with posterior predictive distributions to create interpretable prediction intervals for 5-minute-ahead, hour-ahead, and next-day electricity load prediction
- Consulted with experts at California Energy Commission, California ISO, New York ISO, and CPUC to identify key modeling goals
- Designed a front-end website via Flask to serve a real-time visualization of estimated electricity load and its uncertainty bounds

Bayesian Transfer Learning for Natural Language Processing Aug 2023 – Dec 2023

- Generalized code from *Pre-Train Your Loss: Easy Bayesian Transfer Learning with Informative Priors* for NLP (originally implemented for CV)
- Utilized PyTorch-Lightning Callbacks to adapt the Stochastic Weighted Average-Gaussian (SWAG) technique to learn empirical priors
- Explored the usage of Stochastic Gradient Langevin Dynamics (SGLD) with a cyclic learning rate to handle approximate sampling
- Applied the updated code on XLNet with an upstream task of permutation language modeling using Bookcorpus, and a downstream task of text classification using AG News to achieve a near state-of-the-art accuracy of 94.4% using just 80% of the downstream training data

Experience

College of Computing, Data Science, and Society Berkeley, CA
Data, Inferences, and Decisions – Teaching Assistant Jan 2023 – Present

- Advised student projects in neural networks, causal inference, and Bayesian approaches to modeling and uncertainty quantification
- Reinforced student learning by delivering weekly lectures and lab walkthroughs that distilled challenging topics down to the essentials
- Wrote and deployed code assignments and auto-graders on topics such as Bayesian Hierarchical Modeling via JupyterHub

Division of Computing, Data Science, and Society Berkeley, CA
Lead Applied NLP/ML Researcher Jan 2021– Aug 2021
Applied NLP/ML Researcher May 2019 – Aug 2020

- Developed and codified robust NLP/ML-based metrics to quantify goal congruence in student teams
- Established linkage between pedagogic structure, goal formation, and goal congruence in a large-scale study of multiple team-based engineering classes across the UC Berkeley campus
- Extracted features from real survey data collected from different project-based Berkeley courses to develop production-ready goal categorizers and similarity predictors, and communicated results using data visualization techniques in Seaborn and Matplotlib
- Gained experience working with and fine-tuning transfer-learned deep-learning NLP transformers and embeddings such as BERT
- Published and presented our findings at the 18th International Conference on Design Education, winning best paper for our work

Teaming by Design Berkeley, CA
Data Engineer/Product Manager Aug 2020 – May 2021

- Turned above research into project focused on the development of a web-based teaming platform to improve student teaming experiences
- Designed and implemented end-to-end ELT data pipelines to process, aggregate, and store sensitive survey data via Google Cloud Firestore using a combination of NoSQL, React.js, and Python
- Utilized interactive data visualization elements via Chart.js to provide actionable insights to customers on teaming performance
- Collaborated with full-stack engineers, UX-designers, professors, and business strategists to create a product that meets the needs of both students and instructors while facilitating future research in the area
- Launched a pilot curriculum in classes taught by the Fung Fellowship Program at UC Berkeley that utilized the platform to enhance student learning and detect conflicts

Honors and Awards

Best Paper Award, ASME IDETC-CIE Design Education Conference
Berkeley Data Scholar, Data Discovery Program
SERC Greener DeCal Fellowship Award Recipient
Regional Finalist, Siemens Competition in STEM

Publications

Beckman, S, Jian, A, Sabharwal, A, & Goucher-Lambert, K. "Examining Goal Congruence on Engineering Design and Innovation Student Teams." *Proceedings of the ASME 2021 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference. Volume 4: 18th International Conference on Design Education (DEC)*. August 17–19, 2021. ASME. <https://doi.org/10.1115/DETC2021-71780>