ALAN JIAN

San Francisco, CA · New York, NY
(516) 474-0309 | alanjian131@berkeley.edu | https://alanjjian.github.io/

Education

University of California – Berkeley — Expected Graduation: Dec 2023
Master of Information and Data Science — Overall GPA: 4.00

University of California – Berkeley — May 2022
Bachelor of Arts in Data Science — 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 and Hypothesis Testing:**
  - Machine Learning via Scikit-Learn, PyTorch
  - Natural Language Processing via Huggingface, SpaCy, NLTK
- **Software Engineering Tools and Techniques**
  - Version control via Git
  - Data structures and algorithms via Java, Python
- **Statistical Methods**
  - Hypothesis testing
  - Generalized Linear Models via PyMC3/Statsmodels
  - Causal inference Techniques for Observational Data
  - Bayesian Statistics via PyMC3
- **Data Engineering:**
  - Relational Databases (PostgreSQL)
  - NoSQL Databases (GCP Cloud Firestore, MongoDB, Redis, Neo4j)

Work and Research Experience

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 a robust NLP/ML-based framework 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 cutting-edge NLP techniques/embeddings from context-based encodings such as Word2Vec to transfer-learned deep-learning language models and transformers such as BERT
- Published and presented our findings at the 18th International Conference on Design Education, winning best paper for our work

TeamingxDesign — 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 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
- Worked 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
- Helped launch a pilot program in classes taught by the Fung Fellowship Program at UC Berkeley

Relevant Teaching Experience

Principles and Techniques of Data Science — Jun 2023 – Present

- Taught data science topics to both technical and non-technical students, including: loss optimization, linear and logistic regression, decision trees and random forests, and unsupervised methods around dimensionality reduction and clustering

Data, Inference, and Decision-Making — Jan 2023 – May 2023

- Advised and guided student projects in neural networks, causal inference, and Bayesian and frequentist approaches to modeling and uncertainty quantification
- Worked closely with course staff and professors to create, dockerize, and deploy autograders for labs and homework to automate grading flows in a class of over two hundred students

Human Contexts and Ethics of Data Science — Aug 2022 – Dec 2022

- Led thought-provoking conversations around ethical considerations on topics like algorithmic bias, representational and allocative harm, data capitalism, and fair ML

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