James Fu

jamesfup@gmail.com | 925-875-8886 | jamespfu.com | jamespfu | jamespfu

Education

University of Texas, Austin Expected Dec 2025
M.S. in Data Science

- Cumulative GPA: 3.9/4.0
- Coursework: Machine Learning, Natural Language Processing, Probability and Simulation-Based Inference

University of California, Los Angeles Jun 2024
B.S. in Computational Biology (Data Science Track)

- Cumulative GPA: 3.74/4.0, Dean's List
- Coursework: Machine Learning, Data Science, Data Structures and Algorithms, Discrete Mathematics, Probability

Relevant Experience

UCLA Semel Institute Nov 2022 – Feb 2024
Student IT Technician *Los Angeles, CA*

- Imaged and encrypted 100+ HIPAA-compliant Windows and Mac PCs and upgraded computer hardware.
- Enhanced network infrastructure for 150+ offices by utilizing batch scripts and Excel to gather PC specs, identify units needing replacement, and optimize connectivity and hardware utilization.

Silicon Valley Education Foundation (Computer Science Institute) Jun 2022 – Aug 2022
Teaching Assistant *San Leandro, CA*

- Led 30+ high school students through basic to intermediate Python activities using Adafruit's Circuit Playground Express.
- Developed 10+ interactive lessons on basic data structures, OOP, and project deployment in GitHub; conducted live coding demonstrations, debugged student code, and supervised project development for a final showcase.

Technical Skills

Programming Languages: Python (scikit-learn, numpy, pandas), R (dplyr, ggplot2), C/C++, JavaScript, MATLAB, HTML/CSS

Data Science and Miscellaneous Tools: Data science pipeline (cleaning, wrangling, visualization, modeling, interpretation), Statistics, Experimental design, Hypothesis Testing, Data Structures and Algorithms, NLP, OOP, ETL (SQL), APIs, Excel, Git, React

Projects and Leadership

Enhancing Robustness in Natural Language Inference Models Oct 2024 – Dec 2024
University of Texas at Austin *Austin, TX*

- Created contrast sets using OpenAI's GPT-4o-mini API to generate syntactic distractors modifying the SNLI dataset, addressing model sensitivity to dataset artifacts and enhancing generalization.
- Reproduced dataset cartography by Swayamdipta et al., extending compatibility for ELECTRA, as original codebase was limited to GLUE-style models. Classified training data by difficulty to optimize fine-tuning on challenging examples, resulting in an 8% F1 increase in NLI model robustness on novel contrast sets.

K-Means Clustering for Astrocyte Subtype Quantification Sep 2023 – Jun 2024
Zipursky Lab | Advisors: Dr. Larry Zipursky, Dr. Fangming Xie *Los Angeles, CA*

- Identified six layer-specific astrocyte subtypes by developing a machine learning pipeline that clustered spatial cell data (MERFISH) from the Allen Mouse Brain Atlas and validated subtype feasibility by implementing Support Vector Machines (SVM).
- Discovered 104 astrocyte-specific genes via differential gene expression analysis on 1,122 total genes across 10+ million cells, applying Bonferroni correction to ensure statistical significance.

Build Team Project Manager Jun 2023 – Jul 2024
UCLA Biomedical Engineering Society (BMES) *Los Angeles, CA*

- Directed a team of 35 undergraduate students to develop a full-fledged pulse oximeter project through hands-on experience in Arduino (C++) and Processing (Java) programming, circuitry, and computer-aided design over the course of a year.
- Assisted with developing content for weekly modules and hands-on workshops, created starter code for students' to build upon, and spearheaded project funding applications to facilitate procurement of necessary supplies.