James Fu

jamesfup@gmail.com | linkedin.com/in/jamespfu | jamespfu.com | (925) 875-8886

EDUCATION

The University of Texas at Austin Expected December 2025

M.S. in Data Science, GPA 3.90

- **Relevant Coursework:** Data Science Statistics, Natural Language Processing, Probability and Simulation

University of California, Los Angeles June 2024

B.S. in Computational Biology, Data Science, GPA 3.74

- **Relevant Coursework:** Data Structures and Algorithms, Machine Learning, Probability and Discrete Math

SKILLS

- **Languages:** Python (pandas, TensorFlow, scikit-learn), SQL, R (Tidyverse), C++, JavaScript, HTML, CSS, LaTeX
- **Frameworks and Tools:** Jupyter Notebook (Anaconda), PyCharm, Git, React, APIs, ETL, Data Science Pipeline (cleansing, wrangling, visualization, modeling, interpretation)

EXPERIENCE

Zipursky Lab, Los Angeles, CA September 2023 - June 2024

UCLA Undergraduate Research Thesis

- Developed a ML pipeline using K-means to cluster 9M+ astrocytes from MERFISH data based on CCF coordinates and gene expression. Identified three astrocyte subtypes and used SVM (87% accuracy) to validate them against known neuronal layer markers, assessing whether astrocytes exhibited similar layer specificity.
- Conducted differential gene expression analysis on 1,122 genes, identifying 104 astrocyte-specific markers with Bonferroni correction for statistical significance and visualizing results in a volcano plot.

UCLA Semel Institute, Los Angeles, CA November 2022 - February 2024

Student IT Technician

- Spearheaded procurement and rental of refurbished desktops and peripherals, maintaining inventory, upgrading components, and processing RMA requests to reduce e-waste and costs by $10K+ annually.
- Imaged, encrypted, and deployed 100+ HIPAA-compliant PCs. Optimized network infrastructure for 150+ offices by utilizing batch scripts and Excel to track inventory and identify non-UCLA Health compliant hardware.

Silicon Valley Education Foundation, San Leandro, CA June 2022 - August 2022

Computer Science Innovate Teaching Assistant

- Facilitated student learning in introductory Python concepts taught via Adafruit's Circuit Playground Express.
- Led 10+ Python lessons on recursion, control flow, data structures, and GitHub deployment via live demos. Debugged student code and prepared supplementary coding tutorials and challenges to reinforce OOP concepts.

PROJECTS

Enhancing Robustness in Natural Language Inference (NLI) Models October 2024 - December 2024

- Generated 5+ contrast sets using OpenAI's GPT-4o mini API to create syntactic distractors modifying the SNLI dataset. Evaluated robustness against dataset artifacts on NLI tasks before and after dataset cartography.
- Extended dataset cartography codebase by Swayamdipta et al. to support ELECTRA-small, overcoming the original codebase's BERT-base model limitation. Classified training data by difficulty, optimizing fine-tuning on challenging examples, resulting in an 8% F1 increase in NLI robustness evaluated on novel contrast sets.

Equitable Healthcare Cost Modeling through Deep Learning April 2024 - Present

- Built Faircare, a Python-Flask ML-powered healthcare cost modeling tool using React, surpassed 300+ visits in the first month. Integrated synthetic data generation using MEPS dataset, increasing data volume by over 250%.

LEADERSHIP

UCLA Biomedical Engineering Society, Los Angeles, CA June 2023 - July 2024

Build Team Project Manager

- Led a team of 35 undergraduates in a year-long project, teaching Arduino (C++), Processing (Java), circuitry, and CAD through weekly modules and hands-on workshops to help teams develop a full-fledged pulse oximeter.
- Designed instructional content and demos for weekly modules and workshops and created starter code for students to build upon.