

Shawn Schulz

shawnschulz@gmail.com • Cell: (510)-334-8000 • <https://github.com/shawnschulz>

EDUCATION

University of California, Berkeley, Berkeley, CA 2024—

Masters of Science, Molecular Science and Software Engineering

University of California, Berkeley, Berkeley, CA 2018 — 2022

Bachelor of Arts, Major in Molecular and Cell Biology

GPA: 3.9

WORK EXPERIENCE

Data Engineer, *Stanford Medicine, Kuo and Gawad Labs* June 2022–Nov. 2023

- Developed enterprise user focused HPC pipeline in **python**, **R**, **C++**, obtaining **\$1M+**
- Improved machine learning analysis reducing errors by **500%** using a **gaussian process** model.
- Developed and maintained a database serving **800TB** of genomic data, saving over **400 days** of user runtime via improved concurrency with **bash**, and **linux server environment shell**.
- Automated building and testing, improved error handling and compatibility serving **30+** enterprise users using **R**, **docker**, **nextflow**, **github actions**, **git cli**.
- Performed high impact data analysis on large raw datasets for **\$100,000+** grant using **PCA** and **k-means** clustering, preventing overfitting and improving performance via **numpy** and **pytorch**.
- Solved data science challenges in cancer detection and presented to audiences of 50+ peers.

Machine Learning Researcher, *Berkeley Artificial Intelligence Research (BAIR)* Nov. 2021—May 2022

- Developed, tested, evaluated and benchmarked SCVI **variational autoencoder deep learning** model serving thousands of users. Solved sophisticated machine learning problems.
- Produced high impact results and figures for deep learning models in a \$100,000+ grant.
- Improved model performance by **200%** using **sci-kit learn**, **numpy** and **pytorch** libraries to improve processing throughput and streamline batch correction.
- Modeled using **PCA** and visualized using **matplotlib** to find SARS-CoV2 infection pathology.
- Used **gaussian mixture model** infection level classifier to find novel mechanisms of COVID-19

PROJECTS

gpt-flow

- Created AI workflow visualization tool for text **LLM** data inputs and outputs using **javascript/typescript**, **nodejs**, and **React** framework.
- Developed web server backend serving **100** users using **flask**.

transcriptome-transformer

- Developing multi-ome classifier and protein solver using **C++** and **CUDA**.

Robotic liquid handler

- Embedded systems programming of robotic pipette attachment for 3d printers using **C++**.

OTHER EXPERIENCE

Undergraduate Researcher, *UCSF Department of Surgery, Wang Lab* Nov. 2020–Aug. 2021

- Used a computer vision package in **Matlab** to analyze immunofluorescence microscopy data to find diameter size of blood vessels and flow rate.

Ashland Free Medical Clinic, *Clinic Coordinator* Dec. 2020—

- Managed a team of **nine** volunteers and managed web EHR for over **200 patients**.

PUBLICATIONS

- NF- κ B inhibitor alpha has a cross-variant role during SARS-CoV-2 infection in ACE2-overexpressing human airway organoid
- Cellular and molecular characterization of peripheral glia in the lung and other organs

SKILLS and COURSEWORK

- **C++**, **C#**, **Python**, **Javascript/Typescript**, **R**, **Rust**
- **git**, **AWS**, **OpenAI API**, **react**, **SQL**, **Excel**, **sci-kit learn**, **jupyter**, **conda**, **flask**, **html**, **css**
- Biotech Software, Machine Learning, High Performance Computing, Full Stack Development