Steven (Hanqi) Su

| | | hanqisu@sas.upenn.edu | Portfolio

EDUCATION

University of Pennsylvania, *Accelerated M.S.E. in Robotics* | Philadelphia, US — Expected May 2026
University of Pennsylvania, *B.A. Computer Science + Computational Biology* | Philadelphia, US — Expected May 2026

SKILLS

Programming	C, Python, Java, JavaScript, Swift, R, SQL, Linux/Unix
DevTools	Git, Docker, Docker Compose, Kubernetes, Node.js, Kafka, Flask, FastAPI, Django, IaC, React
Data	Spark, Apache Airflow, PyTorch, Scikit-Learn, Pandas, Databricks, GCP, AWS, PostgreSQL, MySQL, MongoDB, Firebase, Redis, LangChain, Vector Databases

EXPERIENCE

Grove AI, *Software Engineering Intern* | San Francisco, CA (Remote) — July 2024 - Present

- Implemented secure, anonymized patient data migration from production to staging database using **SQLAlchemy**
- Leveraged **GitHub Actions** to automate CI/CD and **cron job** configurations for company's database utility tools
- Created complex **PostgreSQL** query scripts to develop embedded dashboards that monitor company's LLM agents
- Conducted extensive EDA, feature engineering, and modeling using **Pandas** and **Scikit-Learn** to generate data-driven insights from clinical trial statistics

Ancient Tech, *Software Engineering Intern* | Palo Alto, CA (Remote) — May 2024 - August 2024

- Received UPenn's Startup Internship Award, doubled the size of training data for company's Reinforcement Learning for Finance projects by connecting 3 new ticker screeners to **MySQL** in the ETL pipeline of the production codebase
- Utilized **Docker**, **Selenium, and X virtual framebuffer** to deploy Docker images that enable automated browser interactions in Linux-based production servers without graphical user interfaces
- Leveraged **Redis** to enable robust market data caching and persistence and reduce model training time by 30%

SciSketch, *Technical Founder and CEO* | Philadelphia, PA — Feb 2024 - Present

- Designed and implemented a backend framework utilizing **Apache Airflow**, **AWS RDS**, and **AWS S3** to automate biomedical image and text scraping, periodic LLM finetuning, and model deployment
- Engineered a text-icon mapping framework through semantic search with **vector databases** and **LangChain**
- Established detached, distributed, and GPU-enabled computing infrastructures that reduce computation time by 70% utilizing **AWS EC2**, **Databricks**, **Spark** and **GCP**
- Used **PyTorch** and Huggingface APIs to fine-tune Google Flan-T5 and produce two custom LLMs, BioPhraser and BioRegressor

Penn Epigenetics Institute, *Bioinformatics Researcher* | Philadelphia, PA — Sep 2023 - Sep 2024

- Spearheaded development and maintenance of MEAnalysis, an open-source **R**-package for institute internal use in streamlined analysis of neuron multielectrode array data. Leveraged **Objected-Oriented Programming** paradigm in program design and created 5 novel data visualization methods, reducing manual analysis time by 60%

PROJECTS

CompGrid | *Node.js, React, Redis, PostgreSQL, Docker Compose, Kubernetes, CI/CD, IaC* — BioHacks NYC, Nov 2024

- Placed **2nd Place** out of 30 teams. Developed solo a unified platform for local compute resource management
- Led full-stack dev with **Node.js, Express.js, React, PostgreSQL**. Deployed frontend via **Nginx**. Added **Redis** caching for compute instance usage logs and **JWT** for secure user authentication
- Simulated ArchLinux instances with **UTM**. Deployed VM control middleware using **Flask**, **AppleScript**, and **Ngrok**
- Deployed containerized frontend and backend services on **AWS Kubernetes Service** with **AWS Load Balancer** for horizontal scalability, high availability, and self-healing capabilities
- Optimized development process with **Docker Compose**, CI/CD via **GitHub Actions**, and IaC using **Terraform**

Underflow | *FastAPI, Uvicorn, MySQL, Next.js, Github API, OpenAI API* — PennApps, Sep 2024

- PennApps **Best Use of Code Generation Challenge Winner**. A command-line tool that scans a codebase and recommends improved tech stack choices for cost savings and performance enhancements
- Set up and integrated MySQL instance and backend **FastAPI** endpoints with command-line and Next.js dashboard
- Integrated **Github API** into the project architecture to enable code scanning of entire repositories

Extracurricular

CIS 5450 Big Data Analytics, *Teaching Assistant* | Philadelphia, PA — August 2024 - Present

- Leads recitations to teach data science concepts including **lazy execution**, SQL, and **hypothesis testing**
- Host office hours to answer questions on challenging **relational database design** concepts and debug students' code
- Serve as **Project TA Lead**. Manage logistics for course final project and mentor 7 student groups on their final project's dataset selection, feature engineering, and choice of machine learning/deep learning models