Steve Herrin

CONTACT INFORMATION

jobs@steveherrin.com
650-814-8865
San Jose, CA

www.github.com/steveherrin
www.linkedin.com/in/herrinsteve

SUMMARY

Engineering leader with over a decade of experience building and leading teams that use software, data, and machine learning to solve novel problems. Scientific (physics PhD) background with demonstrated adaptability to other fields like biotech.

EXPERIENCE

Pathos AI, Chicago, IL (Remote) June 2022 – present

Vice President of Engineering

- Grew an engineering team from zero to four engineers of varying seniority
- Built a data catalog and management system to search, utilize, understand, and audit access to thousands of datasets and analysis results
- Prototyped LLM RAG generative AI pipeline for scientific literature review
- Used GCP, Terraform, and Nextflow to automate data analysis and processing

D2G Oncology, Mountain View, CA April 2021 – May 2022

Staff Software Engineer

- Created *in silico* simulations of PCR and DNA sequencing, used for oligo design, QC, and automated processing of several multiplexed sequencing runs per month
- Built a pipeline using pydantic to automate ETL and validation of Benchling LIMS data from an HTTP API to a PostgreSQL warehouse
- Developed a Python library exposing GraphQL and REST APIs for accessing and traversing a large knowledge graph of lab, sequencing, public, and analysis data

23andMe, Sunnyvale, CA January 2014 – March 2021

Engineering Management (∼2 years; title: engineering manager)

- Created and grew 3 machine learning and data -focused engineering teams totaling 16 engineers, including a mix of leads and individual contributors
- Led committee of engineering leads to standardize interviewing guidelines and open source release processes, subsequently adopted organization-wide

Engineering Individual Contributor (∼5 years; final title: sr. tech lead engineer)

- Architected containerized (Docker, AWS ECS) machine learning systems to ensure quality and reproducibility of models in a regulated medical device setting
- Designed and implemented a library with a unified API for accessing data & metadata across application-specific data stores, eventually used for all customer content
- Created web portal for external researchers to recruit for genomic studies and receive data back, increasing sales by 2% and producing strategic data-sharing agreements
- Developed and performed a maximum likelihood analysis combining private and public datasets to replace thousands of ineffective genotyping probes
- Built 3 generations of distributed data pipelines with Celery, Luigi, and AWS to run Python, C++, and R algorithms operating on petabytes of genetic data

SLAC National Accelerator Lab, Menlo Park, CA May 2008 – August 2013

Research Associate

- Applied machine learning & statistics to improve detector energy resolution by 25%
- Used computer vision algorithms to repurpose the detector for 3D cosmic ray muon reconstruction, yielding a 10x reduction in cosmogenic background uncertainty
- Built, networked, and programmed PLC control systems with over 600 channels of heterogeneous sensor data at a site with unreliable internet connectivity, successfully protecting $10M of liquid xenon
- Created a PHP logbook webapp with a MySQL backend for tracking lab work
- Developed batch data pipelines using Python, C++, and shell scripts to routinely measure detector characteristics by processing TB of calibration data.

Skills

Languages: Python, Rust, C, C++, SQL, Shell Scripting, Elm, JavaScript/TypeScript, R

Tools: AWS, GCP, NumPy, SciPy, Scikit-Learn, Mypy, Pydantic, FastAPI, Flask, Django, React, MySQL, PostgreSQL, Git, HBase, Spark, LaTeX

Other: Machine Learning, Data Analysis, Bayesian/Frequentist Statistics, Simulation, CI/CD, Sensors, Analog & Digital Electronics, Neutrino & Particle Physics

Education

Insight Data Science, Mountain View, CA December 2013

- Postdoctoral Fellowship

Stanford University, Stanford, CA June 2013

- Ph.D. (Physics)

Rice University, Houston, TX May 2007

- B.S. (Physics)