Danny Collinson

dannycollinson12@gmail.com | 610-290-3410 | github.com/dannycollinson | linkedin.com/in/danny-collinson

Education

California Institute of Technology September 2019 – June 2024

Bachelor of Science (B.S.), Computation and Neural Systems ***GPA: 4.0***

Relevant Coursework: Large Language & Vision Models, Deep Learning, Relational Databases, LLMs as Agents, Data Analysis & Statistical Inference, Machine Learning & Data Mining, Computational Biology & Bioinformatics.

Experience

Data Science Intern June 2023 – September 2023

Recursion Pharmaceuticals - Salt Lake City, UT

- Developed 2 novel metrics to analyze large experimental datasets, improving accuracy of future modeling decisions.
- Implemented recent ML techniques in PyTorch, enhancing power and efficiency in the data transformation pipeline.
- Collaborated with large cross-functional teams to deploy monitoring tools, upgrading data science and QA workflows.

Computational Biology Research Fellow February 2022 – September 2022

Parker Lab, Caltech Biology and Bioengineering - Pasadena, CA

- Increased measurement accuracy by 11% and sped up processing by a factor of 100 for an upcoming publication.
- Created pipeline using deep learning and statistics for automatic microscopy analysis, improving accuracy and speed.
- Oversaw 3 projects combining bioinformatics, ML video analysis, and wet lab, fostering deeper research insights.

Computational Astrophysics Research Fellow February 2020 – February 2021

Harrison Lab, NuSTAR Group, Caltech Physics and Astronomy - Pasadena, CA

- Performed statistical analysis to classify ultraluminous x-ray sources, improving the accuracy of source identification.
- Delivered actionable insights from NuSTAR telescope data using Python and other software, utilizing large datasets.
- Initiated development of an auto-classification pipeline for raw NuSTAR data, streamlining processing and analysis.

Teaching Assistant September 2023 – December 2023

Justin Bois, Caltech Bioengineering - Pasadena, CA

- Instructed 85 students in the graduate-level course Introduction to Data Analysis in the Biological Sciences.
- Taught statistical modeling, numerical optimization, data visualization, and exploratory data analysis in Python.

Relevant Skills

Python: PyTorch, NumPy, pandas, scikit-learn, Jupyter, Matplotlib, SciPy, PyTorch Lightning, Bokeh, and more.
Tools: SQL, git, cloud computing (AWS, GCP, HPC), Docker, Bash/shell, Linux/Unix, GitHub, MCMC, API usage.
Topics: deep learning, large datasets, statistics, EDA, computer vision, LLMs, Bayesian models, prompt engineering.

Selected Projects

Temperature Prediction from GIS Spectra September 2023 – December 2023

- Led team of 3 in ML project to predict surface temperatures from spectral data, achieving error of less than 1 ℃.
- Created a new ML dataset from NASA data, built training and testing pipelines, and designed model architectures.
- Collaborated with Prof. Katie Bouman and JPL scientists to study novel research problems and solve technical issues.

CheXpert Machine Learning Competition April 2023 – June 2023

- Orchestrated training of large multi-GPU models on HPC clusters using Slurm scripts, enabling powerful classifiers.
- Processed over 224k chest x-ray images, adding augmentations to improve model diagnosis accuracy by 8%.
- Placed 3rd as a solo team competing against teams of up to 5 in Caltech's annual ML competition.