

JEFFREY WANG

jdww2@andrew.cmu.edu | (781) 827-1582 | linkedin.com/in/jeffrey-d-wang/ | github.com/jeff-d-wang

EDUCATION

Carnegie Mellon University
MS in Computational Biology

Expected Dec 2026
Pittsburgh, PA

Northeastern University
BS in Computer Science, Minor in Biology, 3.67/4.00

Sep 2021 - May 2025
Boston, MA

Coursework: Machine Learning, Natural Language Processing, Applied Data Science, Bayesian Statistics, Genetics, Molecular Biology, Genome Editing, Cancer Immunotherapies

EXPERIENCE

The Broad Institute, Massachusetts General Hospital - Vazquez-Garcia Lab
Student Researcher

Cambridge, MA
Apr 2025 - Jul 2025

- Authored HPC cluster (SLURM) documentation, improving job submission consistency for a team of 5 researchers.
- **Analysed 392,000+ cancer cells** via a single-cell RNA-seq workflow to perform cell-type clustering, annotation, and differential gene expression. Presented key findings to PI to refine analyses and identify novel gene signatures.
- Fronted the generation of preliminary figures (UMAPs, heatmaps) that were **incorporated into 3 grant applications**.

Blueprint Medicines
Bioinformatics Co-op

Cambridge, MA
Jul 2024 – Dec 2024

- Engineered and deployed a production-grade proteomics analysis pipeline on **AWS** using **Nextflow** and **Docker**, automating QC, dose-response modeling, and differential abundance testing.
- Developed a GSEA reporting workflow for mass spectrometry data, that **slashed analysis and visualization turnaround time by 99%** (from 1 week to 15 minutes).
- Took the initiative to create an **RShiny** app to calculate drug synergy scores, aiding lead compound prioritization.
- Performed differential expression and pathway analysis (**DESeq2, GSEA**) on **RNA-seq data** from 6 cell lines to elucidate mechanisms of drug resistance, presenting findings to project stakeholders.

Verve Therapeutics
Computational Biology Co-op

Boston, MA
Jul 2023 – Dec 2023

- Developed a validation framework in **Python and R** to benchmark and improve CRISPR gRNA on-target efficiency models against applied **transfer learning on published LLMs** and a prototyped two-encoder model with **PyTorch**.
- Trained and optimized a sequence-feature-based **XGBoost** model on over 40,000+ gRNA samples to **predict CRISPR on-target efficiency with an R^2 of 0.91. Reduced the experimental search space for researchers by 30%**.
- Engineered a genomics data pipeline (**BLAST, Ensembl APIs**) to automate the extraction of sequence and structural features, a critical prerequisite for the efficiency model.
- Drove project from model development to final validation, presenting at a company-wide research symposium.

PROJECTS

Machine Learning for Chronic Kidney Disease Prediction | Artificial Intelligence Project Jan 2025 - Apr 2025

- Devised a novel data synthesis method using **Kernel Density Estimation (KDE)** and **rejection sampling** to augment training clinical data and improve model generalization on out-of-distribution samples.
- Made **Random Forest** and **Multilayer Perceptron (MLP)** classifiers from scratch in Python; the MLP, using L2 regularization and dropout, achieved **97% accuracy** on the external validation set.
- Identified clinically relevant biomarkers through permutation feature importance and presented the complete methodology, validation strategy, and findings in a 8-page formal report and research poster.

Full Stack Q and A Web Forum | Fundamentals of SWE Project

Jan 2025 - Apr 2025

- Collaborated in an Agile team of 4 to build a full-stack Q and A web app with two-week sprints and user stories.
- Led the backend development, building the **Node.js/Express** server and **MongoDB** schema.
- Implemented automated background jobs and **RESTful APIs** to update, retrieve, and validate user data.
- Wrote unit and integration tests with **Jest** covering API routes, scheduled jobs, and database interactions.

SKILLS

Languages: Python, R, Bash, SQL, JavaScript, Java, C++, Assembly

ML and Data Science: PyTorch, Tensorflow, NumPy, Pandas, Scikit-Learn, Matplotlib, Seaborn, Tidyverse

Bioinformatics: Seurat, Scanpy, Bioconductor, Nextflow, SLURM, Samtools, BLAST, Docker, Conda, Git

Web and Software Dev: Flask, React.js, AWS (EC2, S3), MongoDB, RESTful APIs, HTML/CSS, GitHub Actions, Jira

Awards: USACO Gold Division, Dean's List (Multiple Semesters), Best Environmental Hack @ AngelHacks