Joyce Yu

530-908-3451
joyceyu579@berkeley.edu
https://www.linkedin.com/in/joyceyu579/
https://github.com/joyceyu579

Education

University of California, Berkeley — **Berkeley, CA**
Bachelor of Arts in Integrative Biology (GPA: 3.60) — *Aug 2015 – May 2019*
College of San Mateo — **San Mateo, CA**
Post-baccalaureate education — *June 2023 - Aug 2024*
Relevant Coursework: Machine Learning Algorithms, Principles and Techniques of Data Science, From Data Warehousing to Big Data, Data Structures and Algorithms, Probability and Statistics in Biology using R, Programming languages for Molecular Sciences: Python and C++, Software Engineering Fundamentals for Molecular Sciences, Deep-Learning with convolution and graph neural networks

Languages and Technologies

- **Languages:** Python, Bash, Groovy/Nextflow, C++, SQL, HTML/CSS, R
- **Machine Learning & Libraries:** scikit-learn, PyTorch, TensorFlow, Keras, BioPython, pysam, NumPy, SciPy, Pandas, Matplotlib, Plotly, RDKit, SQLite, pytest, argparse, re, OpenBLAS, Eigen, tidyverse, dplyr, ggplot2, tibble
- **Data Analysis & Bioinformatics Tools:** Nextflow, Samtools, wgsim, bedtools, seqtk, BWA, BAMSurgeon, picard, bcftools, IGV, FlowJo, PRISM
- **Development Tools:** MySQL, Git/GitHub/GitLab, VSCode, Jupyter Notebook, RStudio
- **DevOps & Cloud Computing:** HPC clusters, Amazon Web Services (CLI, EC2, and S3)

Work Experience

Exact Sciences — **200 Van Buren St., Phoenix, AZ**
NGS Application Developer (contract) — *June 7, 2025 – August 29, 2025*

- Developed a suite of software tools in Python, Bash, and Nextflow for genomic data simulation and pipeline support.
- Leveraged cloud computing services (AWS CLI, EC2, S3), HPC clusters, and SLURM to enable scalable computation.
- Implemented CI/CD pipelines with GitLab and contributed to automated testing frameworks, improving software reliability and deployment speed.
- Integrated bioinformatics tools (e.g., samtools, wgsim, bwa-mem, picard, bedtools, seqtk, IGV, etc.) with Python libraries (argparse, BioPython, numpy, pandas, re, matplotlib, pysam) for computational modeling and simulations.
- Facilitated the development of variant calling algorithms, validation studies, and diagnostic pipelines
- Collaborated with cross-functional teams using Jira and Confluence for project management.
- Presented project results to stakeholders of the BBDS (Bioinformatics and Biological Data Science) team highlighting technical contributions to precision oncology workflows.

Curia Global — **201 Industrial Road, San Carlos, CA**
Research Scientist III — *March 2020 – May 2025*

- Served as technical lead for 27 R&D projects resulting in over 100+ uniquely discovered and characterized antibodies
- Built an interactive client-facing tool using Python to visualize and summarize 6 months of lab data
- Analyzed lab data, generated client-facing reports and presented results at weekly team meetings
- Selected by executive leadership committee to represent 3000+ employees at an international scale and assisted with global marketing content https://www.linkedin.com/feed/update/urn:li:activity:7139090251876077569/
- Trained and mentored 4 scientists and 1 college intern on workflows, wet and dry lab techniques, data collection, software for data analysis, report generation, and provided feedback for improvement
- Helped develop a new R&D service focused on anti-drug antibodies and pharmacokinetic (PK) studies.
- Led a team of 8 scientists and 3 project managers to execute operational duties, projected expenses with data analytical tools to minimize costs, delegated tasks during company acquisition and transitionary phases

Stanford University (School of Medicine) **300 Pasteur Dr., Stanford, CA 94305**

Life Science Research Professional *June 2019 – Feb 2020*

- Assisted with data collection and analysis using R programming and provided additional lab support
- Conducted hypoxia experiments to study genetic mechanisms underlying different lung diseases
- Maintained detailed records of experiments, sample inventory, and the development of new protocols

Projects

MACHINE LEARNING FOR PRECISION ONCOLOGY **Spring 2025**

- Developed a machine learning pipeline to predict personalized anti-cancer therapeutics and treatment responses using a dataset of 204,026 patients with 86 clinical, experimental, and drug-related features.
- Performed extensive data cleaning, normalization, and integration from heterogeneous sources to produce a high-quality analytical dataset.
- Built and validated predictive models using PCA, UMAP, Gaussian Naive Bayes, Nearest Neighbors, feed-forward neural networks, binary classifiers, and k-fold cross-validation to improve precision oncology outcomes.

NATURAL LANGUAGE PROCESSING FOR CHATBOT ANALYSIS **Spring 2025**

- Applied NLP and ML techniques to evaluate 28 large language models across 12,000+ diverse prompts and responses.
- Conducted exploratory data analysis on embedded response vectors and enhanced features using one-hot encoding, K-means clustering, and ELO ratings.
- Utilized PCA and statistical feature extraction for dimensionality reduction and interpretability.
- Created predictive models with linear, logistic, lasso, and ridge regression, validated by k-fold cross-validation and confusion matrices.
- Produced visual insights through histograms, KDE plots, box plots, and waterfall diagrams.

MOLECULAR SUBSTRUCTURE SEARCH VIA GRAPH BASED MODELING AND ALGORITHMS **Fall 2024**

- Developed a python library and molecular fingerprinting algorithm that enables users to perform substructure searches of molecules and functional groups, perform graph-based modeling and visualizations for user-chosen molecules, and can identify aromatic structures and different functional groups using graph traversal algorithms.

SYSTEMIC EVALUATION AND RUN TIME ANALYSIS OF MOLECULAR SIMULATIONS **Fall 2024**

- Refactored on-going python scripts into C++ to improve the space and run time complexity of different methods used to compute the energy of particles in motion and make accurate predictions of larger systems through molecular dynamic simulations

LIGAND-BASED VIRTUAL SCREENING USING MACHINE LEARNING AND ALGORITHMS **Fall 2024**

- Helped develop a virtual screening pipeline to process >100,000 small molecules, identifying novel drug candidates and filtering out toxic compounds.
- Employed machine learning, ROC analysis, and clustering techniques (Taylor-Butina/agglomerative), along with kernel density estimation and retrosynthesis, to prioritize diverse and pharmacologically relevant compounds based on molecular descriptors and structural properties .

INTERACTIVE VISUAL TOOL FOR ANTIBODY WORKFLOWS AND DATA SUMMARIES **Fall 2023**

- Built an interactive client facing visualization tool used to summarize 6 months' worth of lab data and product workflow.
- Ensured compliancy of client proprietary information in collaboration with IT and cybersecurity personnel.
- Taught BASH and Python basics to colleagues to help advance technical skills of others.

Publications, Patents, and Awards

Wnt7a deficit is associated with dysfunctional angiogenesis in pulmonary arterial hypertension
European Respiratory Journal | June 08, 2023 | https://doi.org/10.1183/13993003.01625-2022

Field of invention: Antibodies for treating, preventing, and/or detecting SARS-COV-2 infection
Patent Publication #20230115257 | May 17, 2022 | https://patents.justia.com/patent/20230115257

Mural cell SDF1 signaling is associated with the pathogenesis of pulmonary arterial hypertension
AJRCMB | Feb 20, 2020 | https://doi.org/10.1165/rcmb.2019-0401OC

Received 3 awards by team and company vote at Curia for Employee's Choice (2022), Spot Award (2021), and Employee Appreciation (2020)