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**
Master of Science in Molecular Science and Software Engineering (Online, Part-time) — *Aug 2024 – May 2027*
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, Neural Networks, 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, Object-oriented programming, Linear Algebra, Calculus
Organizations*:* Machine Learning at Berkeley (ML@B)

Languages and Technologies

- **Languages:** Python, Bash, C++, SQL, HTML/CSS, R
- **DevOps & Cloud Computing:** Nextflow, High-Performance Computing clusters, Amazon Web Services (EC2 and S3
- **Machine Learning & Libraries:** scikit-learn, PyTorch, TensorFlow, Huggingface, Keras, BioPython, NumPy, SciPy, Pandas, Matplotlib, Plotly, Seaborn, RDKit, duckDB, pytest, argparse, yfinance, OpenBLAS, Eigen, Seurat, ggplot2, tibble
- **Development Tools:** MySQL, Git/GitHub/GitLab, VSCode, Jupyter Notebook, RStudio

Work Experience

Exact Sciences — **200 Van Buren St., Phoenix, AZ**
NGS Application Developer — *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 EC2 and 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.
- 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.
- Explored novel AI tools (e.g. MCPs, Perplexity, Biomni, etc.) and helped team stay current with emerging technologies

Curia Global — **201 Industrial Road, San Carlos, CA**
Research Scientist III — *March 09, 2020 – May 28, 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
- 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., Palo Alto, CA 94305**
Life science research professional I — *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

University of California, Berkeley (Student Technology Services) **2610 Channing Way, Berkeley, CA**

Student Information Technology (IT) Consultant *May 2018 – May 2019*

- Served as part of campus technical support troubleshooting network connectivity issues, assisting campus associates with hardware and software problems, and providing documentation for training staff
- Resolved over 200+ tickets in the ServiceNow ticketing system
- Provided customer service over the phone, online, and in-person for a variety of devices and platforms

Gump Research Station **G55F+RF2, Moorea-Maiao, French Polynesia**

Overseas field researcher *July 2018 – Dec 2018*

- Designed & executed 2 independent field research projects overseas and presented findings at symposium
- Conducted interviews to collect information upon traditional ethnobotanical medicine
- Performed in-vitro assays with plant extracts on embryonic cells of Echinometra mathaei ("The Burrowing Urchin")
- Collaborated with peers and assisted with data collection on individual projects

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.

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

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.

BANKING SYSTEM APPLICATION **Fall 2023**

- Collaborated with 3 engineers to create a banking application using Python and optimized data structures that allows users to simulate a banking system capable of creating accounts, withdrawals, deposits, and money transfers. The application keeps a historical record of transactions and prevents users from creating invalid or dangerous transactions (i.e. overdrafts and duplicate withdrawals).

Publications, Patents, and Awards

- **MSSE Excellence Scholarship**
  Recognized by faculty and admissions committee of UC Berkeley's Master of Molecular Science and Software Engineering program for academic and professional accomplishments. Awarded March 13, 2024.

- **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

- **"The Curia Way – Curiosity"**
  Video and marketing content modeling company core value with global recognition from executive leadership team
  LinkedIn | Dec 08, 2023 | https://www.linkedin.com/feed/update/urn:li:activity:7139090251876077569/

- **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)**

Certifications

- **Agentic AI Fundamentals**: Architecture, Framework, and Applications (LinkedIn Learning, Sept. 2025)

- **Agentic AI for Developers**: Concepts and Application for Enterprises (LinkedIn Learning, Sept. 2025)