Rika Chan

+1 (646) 255-9183 | rc3517@barnard.edu | Linkedin | Portfolio

EDUCATION

Barnard College, Columbia University | New York, NY
B.A., Computational Biology | GPA: 3.7 | Expected May 2026

TECHNICAL SKILLS

Programming Languages: Python, R, Bash, HTML/CSS, Rust, Typescript
ML & DL: Attention, Ensemble methods, Active learning, Variational autoencoders, Autoencoders, Contrastive learning, Deep learning (CNNs, LSTMs, TCNs)
ML Frameworks & Tools: PyTorch, TensorFlow, Keras, scikit-learn, Hugging Face, CUDA
Infrastructure: High-Performance Computing (HPC), Google Cloud Platform, Docker, Git, Snakemake

RESEARCH EXPERIENCE

Machine Learning Researcher | Rabadan Lab, Columbia University Irving Medical Center
New York, NY | Sep 2024 - Present

- Applied large language models (ESM-2) to protein sequence data, building classification system achieving 92% accuracy for predicting viral host adaptation, directly relevant to pandemic preparedness and clinical decision-making
- Collaborated with researchers to understand domain requirements, translate biological questions into ML frameworks, and communicate model capabilities and limitations to non-technical stakeholders
- Developed data validation pipelines using Python on HPC infrastructure, ensuring data quality and reproducibility in production environment with Git version control

Computational Biology Researcher | Mansfield Lab, Barnard College
New York, NY | Sep 2024 - Present

- Analyzed biological datasets with 20K+ cells and 18K+ genes using ML-enhanced analysis pipelines, collaborating with wet-lab biologists to ensure data quality met requirements for downstream scientific analyses
- Created analytical support tools and custom data outputs enabling researchers without coding expertise to explore high-dimensional datasets, improving research team productivity and data accessibility
- Identified novel biomarkers through data analysis that was subsequently validated experimentally, demonstrating ability to generate actionable scientific insights from complex datasets

Data Science Consultant | Empirical Reasoning Center, Barnard College
New York, NY | Mar 2023 - Present

- Mentored 100+ students across disciplines to understand analytical needs, scope projects, and deliver custom data solutions using Python, R, SQL, Stata, and Excel
- Led analytical workshop sessions, breaking down complex statistical concepts for non-technical audiences and influencing decision-making through clear communication of data-driven insights
- Developed technical documentation and training materials improving data literacy across the college, demonstrating ability to translate technical knowledge into accessible formats

Data Scientist | Uhlemann Lab, Columbia University Irving Medical Center
New York, NY | Jan 2024 - Aug 2024

- Analyzed clinical genomic data from 10 immunocompromised transplant patients, applying statistical methods and ML to extract clinically relevant patterns in bacterial evolution and antibiotic resistance
- Developed automated ETL pipeline on cloud infrastructure for processing whole-genome sequencing data, implementing quality control checks and validation steps to ensure data quality
- Communicated analytical findings to clinical researchers through clear data visualizations and written reports, translating complex genomic analyses into actionable clinical insights

FEATURED PROJECTS

ML Framework for Genetic Perturbation Prediction | *Python, scikit-learn, PyTorch*

- Built ensemble ML framework with systematic validation pipeline to assess model quality, measure prediction uncertainty, and identify areas needing improvement, achieving 18% performance gain
- Developed automated evaluation datasets with version-controlled Python pipelines, ensuring reproducible validation of ML systems across multiple iterations

Research Data Platform Development (scViewer) | *Python, Rust, React, Typescript*

- Designed production data system for biological datasets (10GB+), building ETL pipelines, implementing data quality checks, and creating user-friendly analytics tools for non-coding researchers
- Created comprehensive documentation enabling researchers to validate data quality and understand analysis capabilities and limitations