Rika Chan

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

EDUCATION

Barnard College, Columbia University | New York, NY **May 2026**
B.A. in Computational Biology | GPA: 3.7/4.0

TECHNICAL SKILLS

Programming Languages: Python, R, Rust, TypeScript, Bash, HTML/CSS

Machine Learning: PyTorch, TensorFlow, Keras, scikit-learn, Hugging Face Transformers | Deep learning (CNNs, LSTMs, Transformers), variational autoencoders, contrastive learning, ensemble methods, active learning, attention mechanisms

Software Development: React, Tauri, REST APIs, Git, Docker, CI/CD, test-driven development

Data & Infrastructure: High-Performance Computing (SLURM), AWS, Google Cloud Platform, Snakemake, SQL

RESEARCH & ANALYTICAL EXPERIENCES

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

- Developed ML-enhanced single-cell RNA sequencing pipelines to analyze datasets with 20K+ cells and 18K+ genes, partnering with wet-lab biologists to optimize quality control parameters and ensure data integrity for downstream analysis
- Architected custom pipeline for single-cell and spatial transcriptomics combining ML-based data integration method, allowing downstream analyses such as cell-cell communication and trajectory inference, enabling biological insight extraction from multi-modal datasets
- Built analytical tools and custom visualization outputs that enabled non-coding researchers to explore high-dimensional datasets, improving research team productivity by 80%
- Identified novel biomarkers through differential expression and pathway enrichment analysis, with findings subsequently validated experimentally, demonstrating ability to translate complex datasets into actionable biological insights
- Developed interactive HTML-based visualization tools for spatial transcriptomics results, enabling wet-lab collaborators to validate and iterate on analyses without computational expertise

Machine Learning Researcher | Rabadan Lab, Columbia University Irving Medical Center

New York, NY | Sep 2024 - Present

- Engineered protein sequence classification system using ESM-2 large language models, achieving 92% accuracy in predicting viral host adaptation patterns relevant to H5N1 pandemic preparedness
- Designed end-to-end ML pipeline architecture from data preprocessing through model evaluation, collaborating with domain experts to translate biological questions into supervised learning frameworks
- Deployed reproducible ML pipeline on HPC infrastructure using Python and SLURM, implementing version control with Git and automated testing to ensure production-grade reliability
- Implemented attention-based interpretability analysis to identify key amino acid positions driving host specificity predictions, providing biologically meaningful insights into viral adaptation mechanisms

Data Science Consultant | Empirical Reasoning Center, Barnard College

New York, NY | Mar 2023 - Present

- Consulted with 100+ students across disciplines to scope analytical projects and deliver custom data solutions using Python, R, SQL, Stata, and Excel
- Facilitated analytical workshop sessions translating complex statistical concepts for non-technical audiences, enabling data-driven decision-making across research projects
- Created technical documentation and training materials adopted college-wide, improving data literacy and reducing repeated consultation requests

SELECTED PROJECTS

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

- Built full-stack desktop application combining Python/Rust backend with React/TypeScript frontend and Tauri wrapper for end-to-end single-cell dataset analysis, featuring interactive visualizations and high-performance scientific computing for users without coding experience
- Engineered data infrastructure for large-scale biological datasets (10GB+), including support for AnnData/H5AD formats, quality control checks, and intuitive analytics interfaces enabling wet-lab researchers to perform downstream analyses such as clustering, differential expression, and pathway enrichment without coding
- Deployed platform publicly on LinkedIn; actively used by a research lab at Barnard College, increasing productivity by 80%, and by 12 students in an undergraduate computational biology course for final projects analyzing single-cell datasets in brown fat development research

ML System for Genetic Perturbation Prediction | Python, PyTorch | Github

- Architected ensemble deep learning system combining variational autoencoder with baseline methods to predict cellular responses to combinatorial genetic perturbations, achieving 80% reduction in prediction error on held-out samples
- Developed uncertainty quantification framework measuring epistemic disagreement between models, demonstrating strong correlation with prediction error and enabling reliable identification of high-confidence predictions
- Designed active learning pipeline for iterative experiment selection; identified key limitations in model convergence during rapid retraining cycles, informing future improvements including fine-tuning strategies and early stopping protocols