

Wu, Renjie (Roger)

Boston, MA | renjiuwu@hsph.harvard.edu | +1 (857) 209-1097 | [linkedin.com/in/renjiuwu](https://www.linkedin.com/in/renjiuwu) | [ren-jie-wu.github.io](https://github.com/ren-jie-wu)

ABOUT ME

Machine Learning Engineer Internship applicant with a robust, diverse background in machine learning and genomics and 2+ years of hands-on experience developing and applying machine learning models for biomedical applications. Skilled at working both independently and collaboratively. Seeking to leverage academic expertise and practical experience in a dynamic biotech environment to drive innovative solutions in healthcare.

EDUCATION

Harvard University, School of Public Health

Boston, MA

M.S. in Computational Biology

08/2024 – 05/2026 (Expected)

- **GPA:** 4.0/4.0; **Courses:** Architecting & Engineering Software System, Data Science, Statistical Inference
- **Awards:** Harvard Chan Fellowship (2024)

Peking University, School of Life Sciences

Beijing, China

B.Eng. in Bioinformatics

09/2020 – 07/2024

- Major **GPA:** 3.8/4.0 (1st in the major); **Courses:** Data Structures & Algorithms, Probability & Statistics, Bioinformatics
- **Awards:** Award for Contribution in Student Organizations (2022), Qin Wanshun – Jin Yunhui Scholarship (2021)

SKILLS

- **Domains:** Data Science, Machine Learning, Software Development & Engineering, Statistical Analysis, Genomics
- **Programming Languages:** Proficient in Python, R | Experienced in Bash, C/C++, HTML, JavaScript | Basic in SQL
- **Frameworks:** Experienced in Flask, React, Pytorch, TensorFlow, Scikit-learn, Scanpy
- **Languages:** Native in Mandarin Chinese | Proficient in English | Basic in Cantonese Chinese, Japanese

EXPERIENCES

Engineering Application of Generative AI-Powered Knowledge Management System Boston, MA

Course Project in Architecting & Engineering Software System, MIT

11/2024 – 12/2024

- **Developed** independently a full-stack web application for end-to-end deployment using **Flask** (backend) and **React** (frontend) for a knowledge management system with a user-friendly interface.
- **Integrated** Large Language Model (LLM) API and **Graph RAG** techniques to enable effective information retrieval, theoretically improving retrieval **accuracy** by up to **3 times** (dependent on the data and question types) compared with the non-graph-based approach.

Engineering Deep Learning Model for Predicting Gene Expression Dynamics

Beijing, China

Undergraduate Researcher in Dr. Jinzhao Wang's Lab, Peking University

12/2023 – 05/2024

- **Developed** a machine learning model to predict gene expression changes over time using single-cell genomic data.
- **Built** scalable data preprocessing pipelines for feature transformation, selection, and dimensionality reduction.
- **Designed** and **implemented** a hybrid deep learning architecture combining **Graph Neural Networks** and **Transformer** in **Pytorch** framework, **optimized** by regularization, hyper-parameter tuning, and ablation studies, achieving a **35%** prediction **accuracy** improvement compared with the linear model.
- **Applied** the trained model to infer gene regulatory networks, reaching an **AUC** of **0.78**, surpassing prior studies (~0.53)

Designing Deep Learning Model for Predicting Cancer Subtypes from Multi-omics

Beijing, China

Research Assistant in Dr. Jinzhao Wang's Lab, Peking University

03/2023 – 06/2023

- **Collaborated** with computational specialists, contributing to a **publication** in *Briefings in Bioinformatics* as a co-author.
- **Integrated** biological insights to guide both architecture **design** and **validation** of a deep learning model framework for predicting cancer subtypes and markers from multi-model genomic datasets (~9,000 cells).

Applying Machine Learning to Gene Expression Analyses in Colorectal Cancer

Beijing, China

Research Assistant in Dr. Hongkui Deng's and Dr. Cheng Li's Lab, Peking University

11/2022 – 05/2023

- **Collaborated** with biological experimentalists, contributing by **translating** biological problems into computational tasks and effectively **communicating** the analysis results.
- **Built** data pipelines for processing **large-scale** genomic datasets (~60,000 cells), incorporating feature engineering and batch correction.
- **Applied** unsupervised learning and statistical analysis to construct gene expression profiles (*in vivo* and *in vitro*), delivering actionable insights in manipulating colorectal cancer.
- **Explored** and **utilized** a variety of machine learning software and algorithms for inference in therapeutic targets.