

# Yiyang (Steven) Yu

https://stevenyuyy.com

Email: [y.yiyang@columbia.edu](mailto:y.yiyang@columbia.edu)

LinkedIn: [in/yuyiy](https://www.linkedin.com/in/yuyiy) | Github: [@54yuyu](https://github.com/54yuyu) | Kaggle: [@yuyu54](https://www.kaggle.com/yuyu54)

## Education

---

- **Columbia University in the City of New York** New York, NY  
*B.S. Biomedical Engineering, Minor in Computer Science. GPA: 3.95/4.0* *Expected 2026*
  - **Relevant Coursework:** High Performance Machine Learning, Theoretical Foundations for LLMs, High Dimensional Stats for Biomedical Data, Advanced Programming, Data Structure, Calculus III, Linear Algebra, Machine Learning - Stanford University (Coursera), Deep Learning Specialization - DeepLearning.AI (Coursera)

## Work Experience

---

- **Leash Bio** Salt Lake City, UT (Remote)  
*Machine Learning Intern* *Aug 2024 - Present*
  - Developing a multimodal transformer model to predict binding affinities of protein-molecule bindings.
  - Optimized the training and inference pipeline of protein language models by up to 30% by reimplementing models using flash attention and provided a validation method to evaluate data from wet-lab experiments.
- **Columbia University Irving Medical Center (AlQuraishi Lab)** New York, NY  
*SURF Fellow / Undergraduate Researcher* *December 2023 - Present*
  - Designing algorithms to select a maximally diversified set of proteins for molecular dynamics simulations, generating data to support subsequent model development for predicting protein conformation trajectories.
  - Developing methods to extract protein conformational ensembles from AlphaFold2 through latent space exploration and systematically created a benchmark library to compare with existing methods.
- **Cold Spring Harbor Laboratory (Koo Lab)** Cold Spring Harbor, NY  
*Research Intern* *March 2022 - December 2023*
  - Developed and implemented evolution-inspired data augmentations ([EvoAug-TE](#)) in **TensorFlow** for genomic deep neural networks and demonstrated its improvement in generalization and interpretability. ([Paper](#)).
  - Designed and evaluated more than 100 deep-learning models for predicting DNA promoters' expression rates using **Python**, **TensorFlow**, and **WandB** in the 2022 [DREAM Challenge](#). Placed 7th in the [final leaderboard](#).

## Project

---

- **Gold, NeurIPS 2024 - Predict New Medicines with BELKA, Kaggle (13/1946)** *April 2024 - July 2024*
  - Developed deep learning models to predict small molecule-protein interactions using the Big Encoded Library for Chemical Assessment. Implemented over 40 types of DL models, including CNNs, GNNs, Transformers, RNNs, and GDBT Models, and finally developed a robust solution after using up all 480 submissions.

## Leadership Experience

---

- **Organizing Committee Leader, Columbia Organizing of Rising Entrepreneurs (CORE)** *January 2024 - Present*
  - Facilitating mentor connections and providing programmatic support for cohorts of 6-8 early-stage startups during Columbia's intensive 8-week Almaxworks accelerator, helping teams prepare for the culminating investor pitch Demo Day to drive fundraising success.

## Selected Honors

---

- **2024 Kaggle Competitions Master (3xGold, 4x Silver)** Current Global Ranking: 133 / 203,363 (Top 0.065%)
- **2024 HackMIT Intersystems Challenge 1st Place** \$2000 prize
- **2024 Gilbert Family Scholarship from Columbia Engineering**

## Technical Skills

---

- **Languages:** Python, Javascript, Typescript, Java, C, C++, HTML/CSS, LaTeX, Bash, MATLAB
- **Libraries:** TensorFlow, PyTorch, Numpy, Pandas, WandB, OpenCV, HuggingFace, React, Node, Matplotlib, Nginx
- **Tools:** Git, GCP, HPC, Slurm, Jupyter, Fusion 360, Photoshop, Premiere Pro, Unity 3D, Blender