Yining (Sam) Huang

+1 (847) 894 – 0068 | huangsam@seas.upenn.edu

EDUCATION

University of Pennsylvania , PA   Aug. 2025 - Present
PhD in **Computer Science**

Harvard University, MA   Aug. 2023 – May. 2025
Master's in **Biomedical Informatics**   GPA: 3.96/4.00

Northwestern University, IL   Sep. 2019 – Jun. 2023
Bachelor of Science in **Computer Science** & **Statistics**, Minor in **Mathematics** & **Finance**   GPA: 3.99/4.00

Honors: Summa Cum Laude, McCormick School of Engineering High Honor, Tau Beta Pi - The Engineering Honor Society

PUBLICATIONS

[1] Natalie Maus, Yimeng Zeng, Haydn Thomas Jones, **Yining Huang**, Gaurav Ng Goel, Alden Rose, Kyurae Kim, Hyun-Su Lee, Marcelo Der Torossian Torres, Fangping Wan, Cesar de la Fuente-Nunez, Mark Yatskar, Osbert Bastani, Jacob R. Gardner. **Purely Agentic Black-Box Optimization for Biological Design**. [paper]

[2] Felix Teufel, Aaron W Kollasch, **Yining Huang**, Ole Winther, Kevin K Yang, Pascal Notin, Debora Marks. **Few-shot Protein Fitness Prediction via In-context Learning and Test-time Training**. NeurIPS 2025: AI for Science Workshop. [paper]

[3] Zuobai Zhang*, Pascal Notin*, **Yining Huang**, Aurelie Lozano, Vijil Chenthamarakshan, Debora Marks, Payel Das, Jian Tang. **Multi-Scale Representation Learning for Protein Fitness Prediction**. NeurIPS 2024: Conference on Neural Information Processing Systems. [paper]

[4] Guanghao Wei*, **Yining Huang***, Chenru Duan, Yue Song, Yuanqi Du. **Navigating Chemical Space with Latent Flows**. NeurIPS 2024: Conference on Neural Information Processing Systems. [paper]

[5] **Yining Huang***, Zuobai Zhang*, Jian Tang, Debora Marks, Pascal Notin. **Augmenting Evolutionary Models with Structure-Based Retrieval**. ICML 2024: ML4MLS Workshop. [paper]

[6] Dina Sharon*, **Yining Huang***, Motolani Oyewole, Sammy Mustafa. **How to Go with The Flow: An Analysis of Flow Matching Molecular Docking Performance with Priors of Varying Information Content**. ICLR 2024: GEM Workshop. [paper]

[7] **Yining Huang***, Steffanie Paul*, Debora Marks. **An Energy Based Model for Incorporating Sequence Priors for Target-Specific Antibody Design**. NeurIPS 2023: GenBio Workshop. [paper]

[8] Zaixi Zhang*, Jiaxian Yan*, **Yining Huang***, Qi Liu, Enhong Chen, Mengdi Wang, Marinka Zitnik. **Geometric Deep Learning for Structure-Based Drug Design: A Survey**. ACM Computing Survey. [paper]

[9] Yuanqi Du*, Yingheng Wang*, **Yining Huang**, Jianan Canal Li, Yanqiao Zhu, Tian Xie, Chenru Duan, John Gregoire, Carla P Gomes. **M^{2}Hub: Unlocking the Potential of Machine Learning for Materials Discovery**. NeurIPS 2023: Conference on Neural Information Processing Systems (Datasets and Benchmarks Track). [paper]

[10] Qiwen Zhang, Xueke Tian, Guang Chen, Ze Yu, Xiaojian Zhang, Jingli Lu, Jinyuan Zhang, Peile Wang, Xin Hao, **Yining Huang**, Zeyuan Wang, Fei Gao, Jing Yang. **A Prediction Model for Tacrolimus Daily Dose in Kidney Transplant Recipients With Machine Learning and Deep Learning Techniques**. Frontiers in Medicine. [paper]

(* Equal Contribution)

RESEARCH EXPERIENCES

Evolutionary Model with Sequence, Structure, and Domain-Informed Retrieval   Apr. 2024 – Present
Graduate Student – Advised by Prof. Debora Marks
- Proposed a systematic MSA construction method leveraging protein sequence, structure [5], and functional domain information
- Introduced a filtering strategy to select protein homologous using structure-aware PLMs and informed by downstream tasks
- Developed an alignment-based model that learns evolutionary information from MSA based on protein functional domains
- Demonstrated method efficacy on fitness prediction and observed significant improvement on previously underperformed assays

Quantum Inspired Many-Body GNN   Jun. 2024 – Present
Researcher – Advised by Prof. Di Luo & Yuanqi Du
- Introduced a computational framework inspired by quantum tensor network allowing design of alternative neural architectures, that are not rooted in human intuition and domain-specific knowledge, using different tensor operations and arithmetic
- Described existing neural architectures and proposed over 50 alternative GNN designs under this computational framework
- Demonstrated superior performance of alternative GNN designs across diverse tasks, including molecular property prediction

Multi-Scale Model for Protein Fitness Prediction [3]   Jan. 2024 – Aug. 2024
Graduate Student – Advised by Prof. Debora Marks
- Proposed a multi-modal representation learning framework for proteins that integrates sequence, structure, and surface features
- Introduced instances of this framework S3F for protein fitness prediction that combines sequence information from pre-trained protein language model embeddings with protein backbone structure and detailed surface topology encoded by GVP networks
- Evaluated S3F on ProteinGym benchmark and demonstrated state-of-the-art performance on protein fitness prediction task

Chemical Space Exploration with Potential Flow [4]   Jul. 2023 – Jun. 2024
Researcher - Mentored by Yuanqi Du
- Developed a novel framework based on potential flows to efficiently explore the latent space of molecule generative models
- Unified previous molecule latent space traversal methods and PDE-inspired nonlinear methods under the realm of flow
- Proposed a variety of regularization on the flow dynamics and examined their unique properties on solving different tasks
- Validated the method efficacy on single- and multi-objective molecule optimization under supervised and unsupervised settings

Analysis of Flow Matching Docking Performance [6]   Oct. 2023 – Feb. 2024
Course Project - Advised by Prof. Manolis Kellis
- Conducted a thorough evaluation of flow-matching based ligand docking performance from biology and chemistry perspectives
- Investigated effects of different flow-matching priors with varying molecular conformation details on docking performances

De Novo Antibody Design [7]   Aug. 2023 – Nov. 2023
Graduate Student – Advised by Prof. Debora Marks
- Investigated the sitewise performance of different antigen-conditioned antibody generative models on antibody CDR3 design
- Proposed energy based model for de novo antibody design that combines the structure information from an antigen-conditioned graph-based antibody generative model and the sequence information from a pre-trained antibody language model

TEACHING EXPERIENCES

CIS 6200: Advanced Machine Learning (Graduate-level)   Jan. 2026 – May. 2026
Teaching Assistant - Instructor: Prof. Jacob Gardner
- Assisted in teaching of PhD-level machine learn

STAT415 Introduction to Machine Learning (Graduate-level)   Mar. 2023 – Jun. 2023

Teaching Assistant - Instructor: Prof. Bradly Stadie
- Assisted in the teaching of a Ph.D.-level seminar on advanced machine learning and generative AI

CS336 Design & Analysis of Algorithms   Mar. 2022 – Jun. 2022
Teaching Assistant - Instructor: Prof. Donald Stull
- Led weekly office hours to recap and expand on class materials; graded weekly proof-based assignments with detailed feedback

RELEVANT COURSEWORK

Computation: Natural Language Processing, Reinforcement Learning, Graph Neural Network, CUDA Programming, MLOps, Quantum Computing, Optimization, Stochastic Processes, Time Series Modeling, Bayesian Statistics
Biology: Computational Biology, Clinical Informatics, Cell Biology, Physiology, Genetics, Single Cells, Rare Disease Analysis

TECHNICAL SKILLS

Programming Languages: Python, C, C++, Java, SQL, R, Rust, LaTeX, HTML/CSS, JavaScript, Racket
Libraries & Frameworks: PyTorch, TensorFlow, Lightning, PyG, DGL, RDKit