Menghua (Rachel) Wu

rmwu{at}mit.edu · github.com/rmwu/ · people.csail.mit.edu/rmwu

Education

2020–
Cambridge, MA

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

PhD Candidate in Computer Science. Advised by Profs. Regina Barzilay and Tommi Jaakkola.

2019–2020
Cambridge, MA

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

MEng in Computer Science (GPA 5/5)

2015–2019
Cambridge, MA

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

BS in Computer Science, BS in Mathematics (GPA 4.9/5)

Research Experience

MIT CSAIL

2018–present
Cambridge, MA

Research assistant
deep learning for biology, molecules, and drug discovery

EMERALD INNOVATIONS, WITH PROF. DINA KATABI

06-08.2020
Cambridge, MA

Software engineering intern
deep learning on RF signals for person re-identification; robustness to domain shift

MIT-IBM WATSON AI LAB

06-08.2019
Cambridge, MA

Graduate research intern
semi-supervised approaches for information extraction, transfer learning for NLP

STANFORD SCHOOL OF MEDICINE, WITH PROF. ATUL J BUTTE

06-08.2014
Stanford, CA

Summer intern in medical research (SIMR)
analyzed statistical trends in clinical data for drug combinations; oral presentation at 2015 PSB

Teaching

MACHINE LEARNING (G)

09-12.2018 teaching assistant

INTRODUCTION TO MACHINE LEARNING (U)

01-06.2019 teaching assistant

09-12.2017 teaching assistant

01-06.2017 grader

EECS ASSOCIATE ADVISING

2018-2019 associate academic advisor for Prof. Suvrit Sra

Selected Publications

Menghua Wu, Yujia Bao, Regina Barzilay, Tommi Jaakkola. Sample, estimate, aggregate: A recipe for causal discovery foundation models. *Preprint* (2024).

Bo Qiang, Wenxian Shi, Yuxuan Song, **Menghua Wu**. PROflow: An iterative refinement model for PROT-AC-induced structure prediction. *Preprint* (2024).

Wenxian Shi, Jeremy Wohlwend, **Menghua Wu**, Regina Barzilay. Improving influenza A vaccine strain selection through deep evolutionary models. *Preprint* (2023).

Mohamed Amine Ketata, Cedrik Laue, Ruslan Mammadov, Hannes Stärk, **Menghua Wu**, Gabriele Corso, Céline Marquet, Regina Barzilay, Tommi S. Jaakkola. DiffDock-PP: Rigid Protein-Protein Docking with Diffusion Models. *ICLR Machine Learning for Drug Discovery Workshop* (2023).

Yujia Bao*, **Menghua Wu***, Shiyu Chang, Regina Barzilay. Few-shot Text Classification with Distributional Signatures. *International Conference on Learning Representations* (2020).

Menghua Wu, Marina Sirota, Atul Butte, Bin Chen. Characteristics of drug combination therapy in oncology by analyzing clinical trial data on ClinicalTrials.gov. In *Pacific Symposium on Biocomputing* 20:68-79. 10.1142/9789814644730_0008 (2015).

Service

- Organizer for Generative AI and Biology workshop at NeurIPS 2023.
- Organizer for Molecular Machine Learning (MoML) conference 2022-2023.
- Reviewer for Machine Learning in Structural Biology workshop at NeurIPS 2023.
- Reviewer for *Scientific Reports Bioinformatics*.

Honors

2020 NSF Graduate Research Fellowship
IBM Outstanding Poster Award

2015 Intel Science Talent Search Semifinalist
USA Biology Olympiad Semifinalist
National Merit Finalist
Presidential Scholar Candidate

Languages

native english, chinese
intermediate spanish, japanese

Other Experience

Bloomberg LP

06-08.2018
New York, NY

Software engineering intern
performed statistical analyses of large-scale
mortgage bonds data for anomaly detection

IBM Watson

01-02.2018
Cambridge, MA

Artificial intelligence intern
improved out-of-domain detection in conversation
agents with VAEs

Amazon Lab 126

05-08.2017
Cupertino, CA

Software development intern
built IoT cloud services with AWS serverless architecture (Python),
designed and created React app for API testing

Amazon Lab 126

06-08.2016
Cupertino, CA

Software development intern
developed end-to-end prototype for IoT management service with
Java backend framework and d3.js GUI

MIT Media Lab, Camera Culture Group

2015-2016
Cambridge, MA

Research assistant
developed client-facing web application for diabetic retinopathy
detection with neural models

National Institutes of Health

06-08.2013
Bethesda, MD

Research intern
developed a genetic network to memorize intercellular
signaling events via self-coupling receptors that trigger
fluorescent proteins