

# Yifei Wang

+1-781-354-5091 | [yifeiwang2025@gmail.com](mailto:yifeiwang2025@gmail.com) | [Homepage](#) | [in LinkedIn](#) | [GitHub](#) | [Google Scholar](#) | Waltham, MA, USA

## RESEARCH TOPICS

Machine Learning, Deep Learning, Efficiency, Optimization, Generative AI, Interpretable AI, AI4Science, AI For Drug Discovery, AI For Biomedical Applications.

## EDUCATION

- Brandeis University** Jan 2020 - Feb 2025  
*Ph.D. student in Computer Science* Massachusetts, USA
  - GPA: 4.0/4.0. GSAS Fellowship, Travel and Research Grant.
  - Research Focus: Deep Learning, Large Language Model, Generative AI, AI4Science, AI for Drug Discovery.
- Tsinghua University** Aug 2015 - Jul 2019  
*Bachelor of Pure and Applied Mathematics* Beijing, China
  - GPA: 3.6/4.0. Top 10% Undergraduate Thesis.
  - Major Focus: Statistics, Algorithm Design, Optimization, Machine Learning.

## WORK EXPERIENCE

- Biogen** Feb 2025 - Present  
*Machine Learning Research Fellow* Massachusetts, USA
  - Developing AI tools to accelerate drug discovery
  - Designing Generative AI approaches for synthesizable small molecule design
- Brigham and Women's Hospital** Aug 2022 - Jan 2025  
*Research Trainee part-time advised by Li Zhou, Ph.D., M.D. and Liqin Wang, Ph.D.* Massachusetts, USA
  - Improved reliability and interpretability of healthcare AI models using EHR data across ML and LLMs.
  - Detected and evaluated fairness and bias in healthcare AI decision systems.
- XtalPi Inc.** July 2019 - Dec 2019  
*Algorithm Intern for AI-aided drug discovery* Shenzhen, China
  - Implemented a transformer-like architecture for retrosynthetic reaction path prediction.
  - Improved model architectures and training strategies. Refined the pipeline of reaction prediction.

## SELECTED PUBLICATIONS

- Yifei Wang\***, Yunrui Li\*, Pengyu Hong, Hao Xu. "Advancing Drug Discovery with Enhanced Chemical Understanding via Asymmetric Contrastive Multimodal Learning." *Journal of Chemical Information and Modeling*, 2025.
- Xinsong Du\*, Zhengyang Zhou\*, **Yifei Wang\***, Ya-Wen Chuang, Yiming Li, Richard Yang, Pengyu Hong, David W. Bates, and Li Zhou. "Performance and improvement strategies for adapting generative large language models for electronic health record applications: A systematic review." *International Journal of Medical Informatics*, 2025.
- Yifei Wang**, Liqin Wang, Zhengyang Zhou, John Laurentiev, Joshua R. Lakin, Li Zhou, and Pengyu Hong. "Assessing fairness in machine learning models: A study of racial bias using matched counterparts in mortality prediction for patients with chronic diseases." *Journal of Biomedical Informatics*, 2024.
- Yifei Wang\***, Zhengyang Zhou\*, Liqin Wang, John Laurentiev, Peter Hou, Li Zhou, Pengyu Hong. "Counterpart Fairness – Addressing Systematic Between-group Differences in Fairness Evaluation" In *NeurIPS Workshop*, 2024.
- Yifei Wang**, Shiyang Chen, Guobin Chen, Ethan Shurberg, Hang Liu, and Pengyu Hong. "Motif-Based Graph Representation Learning with Application to Chemical Molecules." *Informatics*, vol. 10, no. 1, p. 8. MDPI, 2023.
- Tong Yang\*, **Yifei Wang\***, Long Sha\*, Jan Engelbrecht, and Pengyu Hong. "Knowledgebra: An Algebraic Learning Framework for Knowledge Graph." *Machine Learning and Knowledge Extraction* 4, no. 2 (2022): 432-445.
- Li, Peizhao, **Yifei Wang**, Han Zhao, Pengyu Hong, and Hongfu Liu. "On dyadic fairness: Exploring and mitigating bias in graph connections." In *International Conference on Learning Representations (ICLR)*, 2021.

\*Equal contribution.

## PROJECTS

---

- **Project 1: An Algebraic Learning Framework for Knowledge Graph** [\[Paper\]](#) [\[Code\]](#)
  - Developed an algebraic framework for learning consistent relation embeddings in knowledge graphs and proposed an algebraic-based instantiation for a knowledge graph embedding model.
  - Proposed simplified variants for learning low-dimensional relation embeddings using shared block settings and regularization-based method to integrate human logic rules.
- **Project 2: Motif-based Graph Representation Learning on Molecules** [\[Paper\]](#) [\[Code\]](#)
  - Designed a novel convolution module for graph representational learning on molecules with an efficient pretraining strategy, enabling the capture of local structural and semantic information from graph motifs.
  - Demonstrated the superiority of the proposed model in molecular learning across multiple dimensions: (1) better results in various molecular property prediction tasks, (2) more stable training, and (3) improved interpretability.
- **Project 3: Explainable Contrastive Multimodal Learning on Molecules** [\[Paper\]](#) [\[Code\]](#)
  - Proposed Asymmetric Contrastive Multimodal Learning, an effective and training-efficient framework tailored for molecules, promoting cross-modality understanding between the molecular graph and other chemical modalities.
  - Demonstrated the effectiveness and interpretability of the proposed framework in various key tasks: (1) large-scale cross-modality retrieval, (2) isomer discrimination, and (3) molecular property prediction.
- **Project 4: Exploring and Mitigating Bias in Graph Connections** [\[Paper\]](#)
  - Investigated the disparity of graph embedding for link prediction bridging on intra- and inter-sensitive groups. Proposed a chain of theoretical analyses on how graph connections could affect dyadic fairness.
  - Accordingly introduced an algorithm for fair link prediction by adjusting the adjacency weight matrix to address the fairness-utility trade-off.
- **Project 5: Unveiling Systematic Differences in Group Fairness Evaluation in Healthcare Contexts** [\[Paper\]](#) [\[Code\]](#)
  - Explored the influence of systematic differences and confounding effects on group fairness estimation and proposed a counterpart-based fairness (CFair) index to improve fairness estimation by mitigating these influences.
  - Studied long-term mortality prediction of patients with chronic diseases, utilizing CFair to assess racial disparities in model predictions and offering insights into the influence of racial bias on model performance.

## SKILLS

---

- **Programming Languages:** Python (familiar with most deep learning packages and platforms such as Sklearn, Pytorch, Pytorch Geometric, Transformers, Peft, HuggingFace, etc.), MATLAB, R.
- **Languages:** English (proficient), Chinese (native).
- **Research Skills:** Algorithm Development, Theoretical Proof, Quantitative Analysis, Experiment Design, Interdisciplinary collaboration, Presentation, Writing and Communication.
- **Soft Skills:** Time Management, Problem Solving, Project Management, Collaboration, Mentorship.

## SERVICES

---

- **Reviewer:** NeurIPS, ICLR, ICML, COLING, KDD Workshop, AMIA Annual Symposium, International Journal of Medical Informatics, BMC Medical Informatics and Decision Making
- **Teaching Assistant:** Computation Theory, Spring 2024 & Statistical Machine Learning, Fall 2020, 2021, 2022.
- **Undergraduate Thesis Advisor:** Supervised a thesis that earned the Kukin Moskowitz Prize for Excellence 2024.