Haorui (Harry) Li

Email: haoruili46@gmail.com

RESEARCH FOCUS

My research focuses on utilizing deep learning—such as graph neural networks (GNNs), generative models, multi-modal learning, and large language models (LLMs)—to advance the research of molecule/protein.

- **Fields:** AI for Science, Representation Learning, Geometric Deep Learning, Drug Discovery
- **Methods:** Graph Neural Networks, Generative Models, Organic Chemistry, Physical Chemistry

EXPERIENCES

Huazhong University of Science and Technology, Wuhan, Hubei *Sep., 2021 –Present*

- **Year 4 Undergraduate**
- **Bachelor of Science**, School of Chemistry and Chemical Engineering
- **Bachelor of Engineering**, School of Computer Science and Technology

GPA: 3.9/4.0, Average Score: 90.9/100, Ranking: 1st (Chemistry)
4.0/4.0, Average Score: 91.6/100, Ranking: Not available (Computer Science)

California Institute of Technology, Pasadena, California *July, 2024 – Present*

- **Undergraduate researcher** in Computing + Mathematical Sciences Department
- Advisor: Dr. **Shengchao Liu** & Prof. **Anima Anandkumar**

Vanderbilt University, Nashville, Tennessee *May, 2024 – Aug., 2024*

- **Intern** in Department of Computer Science
- Advisors: Dr. **Yu Wang** & Prof. **Tyler Derr**

National University of Singapore, Singapore *July, 2023 – Aug., 2023*

- **Intern** in School of Computing
- Advisor: Prof. **LEK HSIANG HUI**

SCHOLARSHIPS & AWARDS

- **2024 National Scholarship** (Highest scholarship awarded by the Chinese government,$<$ **0.1%**)
- **2023 National Scholarship**
- **2022 National Scholarship**
- **2023 Merit Student Scholarship** (given to students that excel in academics, athletics, arts of special interest,$<$ **1%**)
- **2022 Merit Student Scholarship**
- **2022 Outstanding Undergraduates in Term of Academic Performance** (Greatest honor in HUST,$<$ **0.1%**)

PUBLICATIONS

Haorui Li*, Shengchao Liu*, Hongyu Guo, Anima Anandkumar. **Geometry-text Multi-modal Foundation Model for Reactivity-oriented Molecule Editing**. *AI for New Drug Modalities Workshop*, ***38th Conference on Neural Information Processing Systems (NeurIPS 2024)***

CORE COURSES

- **Chemistry**: Quantum Chemistry (4.0/4.0), Physical Chemistry (4.0/4.0), Inorganic Chemistry(Element) (4.0/4.0), Organic Chemistry (4.0/4.0), Structure Chemistry (4.0/4.0), Chemoinformatics (4.0/4.0), Organic Structure Analysis (4.0/4.0), Crystal Chemistry (4.0/4.0), Catalytic Chemistry (4.0/4.0)
- **CS**: Operating Systems Principles (4.0/4.0), Object-Oriented Programming (4.0/4.0), Data Structure and Algorithms (4.0/4.0), Computer Network(4.0/4.0), Digital Circuit and Logic Design (4.0/4.0), Computer Architecture (4.0/4.0), Algorithm Analysis and Design (4.0/4.0), Software Engineering (4.0/4.0)

- **Mathematics**: Calculus A (4.0/4.0), Discrete Mathematics (4.0/4.0), Linear Algebra (4.0/4.0), Probability and Statistics (4.0/4.0)

RESEARCH PROJECTS

California Institute of Technology, Pasadena, California
Computing + Mathematical Sciences Department *July, 2024 – Present*
Advisors: Dr. Shengchao Liu & Prof. Anima Anandkumar
Project : Geometry-text Multi-modal Foundation Model for Molecule Discovery

• Construct a novel large-scale 3D structure-text dataset containing approximately 163K molecules with 202K text-structure pairs.

• Apply contrastive learning to align latent representations between 3D molecular structure (processed by 3D GNNs) and textual descriptions (handled by LLMs).

• Design a range of novel and challenging downstream tasks, such as reactivity-oriented molecule editing, to demonstrate the superiority of the 3D structure-text joint molecular representation.

Vanderbilt University, Nashville, Tennessee
Department of Computer Science *May, 2024 – Aug., 2024*
Advisors: Dr. Yu Wang & Prof. Tyler Derr
Project : On Domain Transferability of Diffusion-augmented Graph Classification

• Perform graph classification on various datasets within two major categories: chemistry and social networks, using GIN and GCN models.

• Train a diffusion model (DiGRESS) on individual datasets and use it to generate new graphs through diffusion-augment, followed by graph classification.

• Train DiGRESS on the entire chemistry or social networks category, apply it to individual datasets to generate new graphs via diffusion-augment, and analyze the transferability across different datasets during graph classification.

National University of Singapore, Singapore
Department of Information Systems and Analytics, School of Computing *July, 2023 – Aug, 2023*
Advisor: Prof. LEK HSIANG HUI
Project : A player recommendation system for NBA team managers—utilizing web mining techniques to collect player and team data

• Lead a team of four undergraduate students in the successful development of a recommendation system.

• Propose and conceptualize the design of an advanced NBA player recommendation system to assist team managers in making informed decisions during player recruitment processes.

• Leverage web crawler, machine learning, and data analytics to provide personalized recommendations based on team requirements and playing strategies.

Huazhong University of Science and Technology, Wuhan, Hubei
College of Chemistry and Chemical Engineering *March, 2022 – July, 2022*
Advisor: Prof. Deli Wang
Project : High-nickel ternary layered cathode materials for lithium-ion batteries

• Focus on addressing the challenges of microcrack formation and poor structural stability in high-nickel cathodes, which lead to reduced rate performance and cycling life in lithium-ion batteries.

• Propose a dual modification strategy, combining primary particle structure design and tungsten and fluorine co-doping (W-F-NCM95) to modify the $Li[Ni_{0.95}Co_{0.025}Mn_{0.025}]O_2$ cathode.

SKILLS

Computer Science: Proficient in Python, PyTorch, LaTeX, Gaussian09
Chemistry: Proficient in Organic Chemistry, and Physical Chemistry
Language: English, Mandarin(Native)