Haorui (Harry) Li

Email: hli8@caltech.edu

RESEARCH FOCUS

As a **junior researcher**, my research focuses on **Machine Learning** on non-Euclidean data (e.g. **graphs**), with fundamental understanding in theory and applications to real-world problems in **Life Science (Chemistry/Biology)**.

- **Fields:** AI for Science, Generative Model, Deep Learning for Molecule/Protein, Drug Discovery
- **Methods:** Graph Neural Networks, Generative Models, Organic Chemistry, Physical Chemistry

EXPERIENCES

California Institute of Technology, Pasadena, California *July, 2024 – Present*
- **Undergraduate researcher** in Computing + Mathematical Sciences Department
- Advisors: Dr. **Shengchao Liu** & Prof. **Anima Anandkumar**

Huazhong University of Science and Technology (HUST), Wuhan, Hubei *Sep., 2021 –Present*
- Year 4 Undergraduate
- **Bachelor of Science**, School of Chemistry and Chemical Engineering
- **Bachelor of Engineering**, School of Computer Science & Technology
- **GPA**: **4.0/4.0**, Average Score: **91.8/100**, Ranking: **1**st (Chemistry)
  4.0/4.0, Average Score: **91.6/100**, Ranking: Not available (Computer Science)

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),

SCHOLARSHIPS & AWARDS

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

SKILLS

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

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.

Huazhong University of Science and Technology, Wuhan, Hubei
School of Chemistry and Chemical Engineering *Nov., 2023 – Feb., 2024*
Advisor: Prof. Yuzhou Wu
Project : Predicting the active sites of artificial enzymes through machine learning and generative model

• Contribute to the project ideation by suggesting the use of generative model to address the problem, a proposal that was ultimately adopted by the professor

• Propose the utilization of a diffusion model, inspired by a technique originally employed in the field of conformation prediction for small molecule-protein binding.

Huazhong University of Science and Technology, Wuhan, Hubei
School of Computer Science and Technology *May, 2023 – Sep., 2023*
Advisor: Prof. Yao Wan
Project : An overview of the evolution of NL2VIS——from the era of deep-learning to the era of LLM

• Conduct literature search, review almost all papers related to NL2VIS, and thoroughly read influential articles in the field.

• Summarize the development of the NL2VIS field, particularly during the eras of deep learning and large model, and extract key points and innovations from significant articles in the domain.

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
School 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.