Wenhao Gao

whgao@mit.edu | Website | Google Scholar | Github | ORCID

Education

Massachusetts Institute of Technology — Cambridge, MA
Doctoral student in Chemical Engineering — *Sep 2020 - now*

Johns Hopkins University — Baltimore, MD
Master of Science and Engineering in Chemical and Biomolecular Engineering — *Aug 2018 - May 2020*

Peking University — Beijing, China
Bachelor of Science in Chemistry — *Sep 2014 - Jun 2018*

Research Experience

Connor W. Coley Research Group — Cambridge, MA
Graduate Student, Massachusetts Institute of Technology — *Jan 2021 - now*

- Designed machine learning algorithms to constrain molecular generation within the synthesizable chemical space.
- Developed machine learning algorithms to improve data efficiency in discriminative and generative molecular design.

Jefferey J. Gray Research Group — Baltimore, MD
Graduate Student, Johns Hopkins University — *Sep 2019 - Jun 2020*

- Designed a reinforcement learning algorithm to enhance protein conformation sampling with the Rosetta score function.

Klavs F. Jensen Research Group — Cambridge, MA
Visiting Student, Massachusetts Institute of Technology — *May 2019 - Aug 2019*

- Investigate the synthetic accessibility issue of molecule generative models using data-driven synthesis planning software.

Luhua Lai Research Group — Beijing, China
Undergraduate Researcher, Peking University — *Sep 2017 - Jun 2018*

- Developed machine learning algorithms to predict and design class A GPCR antagonists and agonists.

Jian Liu Research Group — Beijing, China
Undergraduate Researcher, Peking University — *Sep 2015 - Aug 2017*

- Developed a computational workflow to model the peak intensity and location of fluorescence for near-infrared fluorescent proteins, leveraging QM/MM molecular dynamics and TD-DFT.

Professional Experience

Los Alamos National Laboratory — Los Alamos, NM
Machine Learning Scientist Intern at Computer, Computational, and Statistical Sciences Division — *Jun 2020 - Aug 2020*

- Developed machine learning methods to aid the recycling of radioactive elements for sustainable nuclear power usage.

Honors & Awards

- **2023:** Google PhD fellowship (Link)
- **2022:** MIT-Takeda fellowship (Link)
- **2016:** Supported by National University Student Innovation Program (Ministry of Education, China)
- **2016:** May Fourth Scholarship (Peking University, China)

Publications

* indicates equal contribution.

- Wang, H.*, Fu, T.*, Du, Y.*, Gao, W., ... , & Zitnik, M. (2023) "Scientific discovery in the age of artificial intelligence." *Nature, 620*(7972), 47-60. Link
- Huang, K.*, Fu, T.*, Gao, W.*, Zhao, Y., Roohani, Y., Leskovec, J., Coley, C.W., Xiao, C., Sun, J., & Zitnik, M. (2022) "Artificial Intelligence Foundation for Therapeutic Science." *Nature Chemical Biology, 18*(10), 1033-1036. Link
- Gao, W.*, Fu, T.*, Sun, J. & Coley, C. W. (2022). "Sample Efficiency Matters: A Benchmark for Practical Molecular Optimization." *Proceeding of Neural Information Processing Systems, Datasets and Benchmarks Track* Link
- Fu, T.*, Gao, W.*, Coley, C. W. & Sun, J. (2022). "Reinforced Genetic Algorithm for Structure-based Drug Design." *Proceeding of Neural Information Processing Systems* Link

- Gao, W., Raghavan, P. & Coley, C. W. (2022). "Autonomous Platforms for Data-driven Organic Synthesis." *Nature Communications 13*(1), 1075. Link
- Gao, W., Mercado, R. & Coley, C. W. (2021). "Amortized Tree Generation for Bottom-up Synthesis Planning and Synthesizable Molecular Design." *Proceeding of International Conference on Learning Representations* (Spotlight) Link
- Fu, T.*, Gao, W.*, Xiao, C., Yasonik, J., Coley, C. W., & Sun, J. (2021). "Differentiable Scaffolding Tree for Molecular Optimization." *Proceeding of International Conference on Learning Representations* Link
- Tynes, M., Gao, W., Burrill, D.J., Batista, E.R., Perez D., Yang, P., & Lubbers, N. (2021) "Pairwise Difference Regression: A Machine Learning Meta-algorithm for Improved Prediction and Uncertainty Quantification in Chemical Search." *Journal of Chemical Information and Modeling, 61*(8), 3846-3857 Link
- Huang, K.*, Fu, T.*, Gao, W.*, Zhao, Y., Roohani, Y., Leskovec, J., Coley, C.W., Xiao, C., Sun, J., & Zitnik, M. (2021) "Therapeutics Data Commons: Machine Learning Datasets and Tasks for Therapeutics." *Proceeding of Neural Information Processing Systems, Datasets and Benchmarks Track* Link
- Gao, W., Mahajan, S. P., Sulam, J., & Gray, J. J. (2020) "Deep Learning in Protein Structural Modeling and Design." *Patterns, 1*(9). Link
- Gao, W., & Coley, C. W. (2020) "Synthesizability of Molecules Proposed by Generative Models." *Journal of Chemical Information and Modeling, 60*(12), 5714-5723 Link

Presentation (invited)

- "Chemistry-tailored AI algorithms for molecular discovery: harnessing synthesis to bridge the gap." *The MIT Corporation*. (Feb. 2024)
- "AI in small molecule drug discovery–design and synthesis." *Johns Hopkins University (JHU)*. (Apr. 2023)
- "Toward autonomous molecular discovery and data-driven therapeutic design." *Memorial Sloan Kettering Cancer Center (MSKCC)*. (Dec. 2022)
- "Sample efficiency matters: A benchmark for practical molecular optimization." *Molecular Modeling and Drug Discovery Seminar, Valence and Montreal Institute for Learning Algorithms (Mila) [remote].* (Sep. 2022) Link to Video
- "Toward autonomous molecular discovery." *Nutshell Therapeutics [remote].* (Aug. 2022)
- "Amortized tree generation for bottom-up synthesis planning and synthesizable molecular design." *AI TIME [remote].* (Jun. 2022) Link to Video
- "Molecular synthesizability and synthetic tree generation for molecular design." *Molecular Modeling and Drug Discovery Seminar, Valence and Montreal Institute for Learning Algorithms (Mila) [remote].* (Jan. 2022) Link to Video
- "Synthetic accessibility and synthesizable molecular design." *Deep Potential Technology (DP) [remote].* (Jan. 2022)
- "Amortized tree generation for bottom-up synthesis planning and synthesizable molecular design." *Nanyang Technological University (NTU) [remote].* (Dec. 2021)
- "Amortized tree generation for bottom-up synthesis planning and synthesizable molecular design." *Korea Advanced Institute of Science and Technology (KAIST) [remote].* (Oct. 2021)
- "The synthesizability of molecules proposed by generative models." *A.I. Socratic Circles (AISC)[remote].* (Jun. 2020) Link to Video

Presentation (contributed)

- "Substrate scope contrastive loss: Repurposing human bias to learn representations of reactive atoms" Oral and poster presentation at the spring conference of *American Chemistry Society (ACS)*, New Orleans, LA. (Mar. 2024)
- "Substrate scope contrastive loss: Repurposing human bias to learn atomic representation" Poster presentation at the conference of *Molecular Machine Learning Conference*, Cambridge, MA. (Nov. 2023)
- "Sample efficiency matters: A benchmark for practical molecular optimization" Oral presentation at the fall conference of *American Chemistry Society (ACS)*, San Francisco, CA. (Aug. 2023)
- "Reinforced genetic algorithm for structure-based drug design" Poster presentation at the *Thirty-sixth Conference on Neural Information Processing Systems (NeurIPS)*, New Orleans, LA. (Dec. 2022)
- "Sample efficiency matters: A benchmark for practical molecular optimization" Poster presentation at the *Thirty-sixth Conference on Neural Information Processing Systems (NeurIPS)*, New Orleans, LA. (Dec. 2022)
- "Sample efficiency matters: A benchmark for practical molecular optimization" Poster presentation at the conference of *Molecular Machine Learning Conference*, Cambridge, MA. (Oct. 2022)
- "Therapeutics data commons: Artificial intelligence foundation for therapeutic science" Oral presentation at the fall conference of *American Chemistry Society (ACS)*, Chicago, IL. (Aug. 2022)

- "Sample efficiency matters: A benchmark for practical molecular optimization" Poster presentation at the fall conference of *American Chemistry Society (ACS)*, Chicago, IL. (Aug. 2022)
- "Reinforced genetic algorithm for structure-based drug design" Poster presentation at the *AI for Science Workshop, Thirty-ninth International Conference on Machine Learning (ICML)*, Baltimore, MD. (Jul. 2022)
- "Sample efficiency matters: A benchmark for practical molecular optimization" Poster presentation at the *AI for Science Workshop, Thirty-ninth International Conference on Machine Learning (ICML)*, Baltimore, MD. (Jul. 2022)
- "Synthetic tree generation for synthesis planning and synthesizable molecular design" Oral presentation at the spring conference of *American Chemistry Society (ACS)*, San Diego, CA. (Mar. 2022)
- "Amortized synthetic tree generation for synthesis planning and synthesizable molecular design" Poster presentation at the conference of *Machine Learning for Pharmaceutical Discovery and Synthesis Consortium*, Remote (Oct. 2021)
- "Can we synthesize the molecules proposed by generative models?" Poster presentation at the conference of *3rd RSC-BMCS / RSC-CICAG Artificial Intelligence in Chemistry Symposium*, Remote (Sep. 2020)
- "Can we synthesize the molecules proposed by generative models?" Poster presentation at the conference of *AI Powered Drug Discovery and Manufacturing*, Cambridge, MA. (Feb. 2020)

Teaching Experience

- **Instructor**, *Massachusetts Institute of Technology*
  Machine Learning for Molecular Design (6.S085; IAP 2024) This class is one that I designed, prepared, and taught, focusing on machine learning algorithms for molecular design purposes. The syllabus and course material are open to the public at our course website.
- **Instructor**, *University of Maryland*
  Scientific Computing From Scratch (Summer 2023) An NSF and MOSSI sponsored summer bootcamp on scientific computing for beginners organized by Prof. Pratyush Tiwary, University of Maryland (Link).
- **Teaching Assistant**, *Massachusetts Institute of Technology*
  Introduction to Chemical Engineering (10.10; Fall 2023)
- **Training**, *Massachusetts Institute of Technology*
  Graduate Teaching Certificate from MIT Teaching+Learning Lab (Fall 2023)

Mentorship

Research Supervisor (UG: undergraduate, MS: master's)
Jacob Yasonik (UG, 2021, MIT), Chanwoo Yoon (UG, 2022, MIT), Christian Ulmer (MS, 2021, TU Berlin and KTH Stockholm), Sabrina Cai (UG, 2023, MIT), Ziang Li (UG, 2023, Tsinghua University), Ron Shprints (UG, 2022-2024, MIT)

Professional Service

Journal/Conference Peer Reviewing

ACS Omega, Chem, Chemical Science, Communications Chemistry, Digital Discovery, Frontiers in Bioengineering and Biotechnology, International Conference on Learning Representations, International Conference on Learning Representations Workshop on Deep Generative Models for Highly Structured Data, International Conference on Machine Learning Workshop on AI for Science, International Conference on Machine Learning Workshop on Structured Probabilistic Inference & Generative Modeling, Journal of Chemical Informatics and Modeling, Special Interest Group on Knowledge Discovery and Data Mining, Learning on Graphs Conference, Nature Communications, Conference on Neural Information Processing Systems, Conference on Neural Information Processing Systems Datasets and Benchmarks Track, Conference on Neural Information Processing Systems Workshop on AI for Science, Nature Machine Intelligence

Conference and Symposium Organization

- Organizing committee of AI for Science Workshops at NeurIPS 2021, ICML 2022, NeurIPS 2023
- Organizing committee of Machine Learning and AI for Organic Chemistry Symposium at ACS 2023 Fall, ACS 2024 Spring, ACS 2024 Fall

Professional Memberships

- 2022-now American Chemical Society
- 2024-now American Institute of Chemical Engineers