

# Wenhao Gao

[whgao@mit.edu](mailto:whgao@mit.edu) | [Website](#) | [Google Scholar](#) | [Github](#) | [ORCID](#)

## EDUCATION

### Massachusetts Institute of Technology

Doctoral student in Chemical Engineering

Cambridge, MA

Sep 2020 - now

**Advisor:** Connor W. Coley

### Johns Hopkins University

Master of Science and Engineering in Chemical and Biomolecular Engineering

Baltimore, MD

Aug 2018 - May 2020

**Thesis:** [Development of a Protein Folding Environment for Reinforcement Learning](#)

**Advisor:** Jeffrey J. Gray

### Peking University

Bachelor of Science in Chemistry

Beijing, China

Sep 2014 - Jun 2018

**Thesis:** Classification and Generation of Class A GPCRs Modulators based on ChemVAE

**Advisor:** Jianfeng Pei, Luhua Lai

## PROFESSIONAL

### Los Alamos National Laboratory

Machine Learning Scientist Intern at Computer, Computational, and Statistical Sciences Division

Los Alamos, NM

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:** Academic Excellence Award (Peking University, 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), SELECTED

---

- "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)
- "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)
- "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)

## PROFESSIONAL SERVICE

---

### Journal/Conference Peer Reviewing

Chem, Chemical Science, Nature Machine Intelligence, Nature Communications, Journal of Chemical Informatics and Modeling, Digital Discovery, ACS Omega, Communications Chemistry, Frontiers in Bioengineering and Biotechnology, NeurIPS, NeurIPS Datasets and Benchmarks Track, ICLR, KDD, Learning on Graphs, NeurIPS AI for Science Workshop, ICML AI for Science Workshop

### Conference and Symposium Organization

Organizing committee and program chair of NeurIPS 2021 [AI for Science Workshop](#), ICML 2022 [AI for Science Workshop](#), NeurIPS 2023 [AI for Scientific Discovery Workshop](#). Organizing committee of [ACS Machine Learning](#) and [AI for Organic Chemistry Symposium](#).

### Open-sourced Software Contribution

Founding member and core developer of [Therapeutics Data Commons](#)