Wenhao Gao

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

Education

Massachusetts Institute of Technology — Cambridge, MA
Doctor of Philosophy in Chemical Engineering, Advisor: Connor W. Coley — *Sep 2020 - May 2025 (Expected)*

Thesis: Toward Efficient and Synthesizable In-silico Molecular Design

- Developed deep generative models and combinatorial optimization methods for efficient navigation in synthesizable chemical space.

Johns Hopkins University — Baltimore, MD
Master of Science and Engineering in Chemical and Biomolecular Engineering, Advisor: Jeffrey J. Gray — *Aug 2018 - May 2020*

Thesis: Development of a Protein Folding Environment for Reinforcement Learning

- Designed a reinforcement learning algorithm to enhance protein folding sampling with the Rosetta score function as an oracle.

Peking University — Beijing, China
Bachelor of Science in Chemistry, with a minor in Physics, Advisor: Luhua Lai, Jianfeng Pei — *Sep 2014 - Jun 2018*

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

- Developed discriminative and generative algorithms to predict and design class A GPCR antagonists and agonists.

Additional Research Experience

Los Alamos National Laboratory — Los Alamos, NM
Machine Learning Scientist 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.

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 with a data-driven synthesis planning algorithm.

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

- Developed computational workflow to simulate fluorescence of near-infrared fluorescence proteins based on time-dependent density-functional theory and molecular dynamics simulations.

Honors & Awards

- **2025:** CAS Future Leaders
- **2024:** ACS CINF The Scientific Excellence Award
- **2024:** D. E. Shaw Research Doctoral Fellowship
- **2023:** Google PhD Fellowship (Link)
- **2022:** Takeda Fellowship (Link)
- **2020:** Fellowship of National Security Education Center, Los Alamos National Laboratory
- **2016:** National University Student Innovation Program Grant awarded by the Ministry of Education, China
- **2016:** May Fourth Scholarship awarded by Peking University, China

Book chapters

- Gao, W., & Coley, C. W. "Synthesis-based design – A Practical and Generalizable Approach to de novo Molecular Discovery." in *An Introduction to Generative Drug Discovery*, CRC Press 69-79. Link

PUBLICATIONS

* indicates equal contribution.

- Gao, W., Raghavan, P., Shprints, R., & Coley, C. W. (2025). "Revealing the Relationship Between Publication Bias and Chemical Reactivity with Contrastive Learning." *Journal of the American Chemical Society*, 147(10), 8959-8968. Link
- Gao, W.*, Luo, S.*, & Coley, C. W. (2024). "Generative Artificial Intelligence for Navigating Synthesizable Chemical Space." *arXiv preprint arXiv:2410.03494*. Link (under review)
- Wang, H., Skreta, M., Ser, C. T., Gao, W., Kong, L., Streith-Kalthoff, F., ... & Zhang, C. (2024). "Efficient Evolutionary Search Over Chemical Space with Large Language Models." *Proceeding of International Conference on Learning Representations* Link
- Velez-Arce, A., Huang, K., Li, M. M., Lin, X., Gao, W., Fu, T., ... & Zitnik, M. (2024). "TDC-2: Multimodal Foundation for Therapeutic Science." *bioRxiv preprint bioRxiv:2024-06*. Link (under review)
- Li, X., Wang, P., Fu, T., Gao, W., Li, C., Shi, L., & Liu, J. (2024). "AUTODIFF: Autoregressive Diffusion Modeling for Structure-based Drug Design." *arXiv preprint arXiv:2404.02003*. Link (under review)
- Luo, S.*, Gao, W.*, Wu, Z.*, Peng, J., Coley, C. W. & Ma, J. (2024). "Projecting Molecules into Synthesizable Chemical Spaces." *Proceeding of International Conference on Machine Learning* Link
- Yu, K., Roh, J., Li, Z., Gao, W., Wang, R., & Coley, C. W. (2024). "Double-Ended Synthesis Planning with Goal-Constrained Bidirectional Search." *Proceeding of Neural Information Processing Systems* (Spotlight) Link
- Subramanian, A., Gao, W., ... , & Gomez-Bombarelli, R. (2024) "Closing the Execution Gap in Generative AI for Chemicals and Materials: Freeways or Safeguards." *An MIT Exploration of Generative AI*. Link
- 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* 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* 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

Invited Presentations

- "Toward Efficient and Synthesizable In-silico Molecular Design." *the University of Pennsylvania, Department of Chemical and Biomolecular Engineering*. (Feb. 2025)
- "Toward Synthesizable and Efficient Molecular Design." *Peking University, the College of Chemistry and Molecular Engineering*. (Dec. 2024)
- "Generative AI for navigating synthesizable chemical space." *Machine Learning for Pharmaceutical Discovery and Synthesis Consortium*. (Oct. 2024)
- "Generative AI for navigating synthesizable chemical space." *Machine Learning and Informatics for Chemistry and Materials Workshop at Telluride Science & Innovation Center*. (Oct. 2024)
- "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
- "Therapeutics data commons: Artificial intelligence foundation for therapeutic science" Oral presentation at the fall national conference of *American Chemistry Society (ACS)*, Chicago, IL. (Aug. 2022)
- "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

Contributed Presentations (selected)

- "Substrate scope contrastive loss: Repurposing human bias to learn representations of reactive atoms." Oral and poster presentation at the spring national conference of *American Chemistry Society (ACS)*, New Orleans, LA. (Mar. 2024)
- "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 fall national conference of *American Chemistry Society (ACS)*, Chicago, IL. (Aug. 2022)
- "Differentiable Scaffolding Tree for Molecule Optimization." Poster presentation at the *Tenth International Conference on Learning Representations (ICLR)*, Remote. (Apr. 2022)
- "Amortized Tree Generation for Bottom-up Synthesis Planning and Synthesizable Molecular Design." Poster presentation at the *Tenth International Conference on Learning Representations (ICLR)*, Remote. (Apr. 2022)
- "Synthetic tree generation for synthesis planning and synthesizable molecular design" Oral presentation at the spring national conference of *American Chemistry Society (ACS)*, San Diego, CA. (Mar. 2022)
- "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)

Teaching Experience

- **Instructor**, *Massachusetts Institute of Technology*
  Machine Learning for Molecular Design (6.S085; IAP 2024) A course focused on machine learning algorithms for molecular design, covering foundational architectures and their application in drug and material discovery. I was responsible for designing, preparing, and teaching. The syllabus and course material are open to the public at our course website.
- **Instructor**, *University of Maryland*
  Scientific Computing From Scratch (2023, 2024) An NSF and MOSSI sponsored summer bootcamp on scientific computing for beginners organized by Prof. Pratyush Tiwary, University of Maryland (Course Website).
- **Guest Lecturer**, *Northeastern University*
  Seminar in Artificial Intelligence: Frontier in AI for Science (CS7170, 2025) A seminar series about how AI and ML are transforming science. (Course Website).
- **Teaching Assistant**, *Massachusetts Institute of Technology*
  Introduction to Chemical Engineering (10.10; Fall 2023) An introductory course in chemical engineering.
- **Training**, *Massachusetts Institute of Technology*
  Graduate Teaching Certificate (Fall 2023) (link) from MIT Teaching+Learning Lab.

Mentoring Experience

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, Recipient of MIT EECS Takeda Undergraduate Research and Innovation Scholar)

Proposal Writing Experience

- **Contributing Author**, *National Science Foundation* (2024)
  Contributed to a proposal establishing a multi-institutional collaborative center for foundational materials intelligence. Responsible for drafting the methodology and detailing potential applications.
- **Sole Author**, *Google Ph.D. Fellowship* (2022)
  Solely developed a successful proposal for the Google Ph.D. Fellowship on generative design methods for small molecule therapeutics, securing $300,000 to cover three years of tuition, stipend, and insurance. Took full responsibility for the definition of research objectives, methodology, and broader impacts.
- **Sole Author**, *Takeda Fellowship* (2021)
  Solely developed a successful research proposal for the Takeda Fellowship on developing sample-efficient design methods for small molecule drugs, securing approximately $100,000 to cover one year of tuition, stipend, and insurance. Took full responsibility for the definition of research objectives, methodology, and broader impacts.
- **Contributing Author**, *Office of Naval Research* (2020)
  Contributed to a successful proposal on generative models for autonomous molecular design, securing $510,000 in funding. Responsible for drafting the methodology and detailing potential applications.

Professional Service

Conference and Symposium Organization

- **2025**: Organizing committee of Generative Modeling for Chemistry, Biology, & Material Discovery Symposium at ACS.
- **2024**: Organizing committee of Broad Institute Machine Learning in Drug Discovery Symposium.
- **2024**: Organizing committee of AI for Chemistry and Materials Science Symposium.
- **2023-2024**: Organizing committee of Machine Learning and AI for Organic Chemistry Symposium at ACS.
- **2021-2023**: Organizing committee and program chair of AI for Science Workshop Series at NeurIPS and ICML.

Policy and Ethics

OpenAI Red Teaming Network: assess the risks & safety of OpenAI models.

Journal/Conference Peer Reviewing

ACS Omega, Chem, Chemical Science, Communications Chemistry, Digital Discovery, Frontiers in Bioengineering and Biotechnology, ICLR, Journal of Chemical Informatics and Modeling, Journal of Cheminformatics, KDD, LoG, Nature Communications, NeurIPS, NeurIPS Datasets and Benchmarks Track, Nature Machine Intelligence