

# BOWEN GAO

Beijing, China ◊ +86 13810620833

billgao0111@gmail.com ◊ LinkedIn ◊ Google Scholar

## EDUCATION

### Tsinghua University

Ph.D. in Computer Science and Technology

Beijing, China

Aug 2024 - Dec 2027 (Expected)

- Supervised by Professor Ya-Qin Zhang and Professor Yanyan Lan

### California Institute of Technology

Master of Electrical Engineering

Pasadena, CA

Oct 2019 - Jun 2021

- GPA: 4.2 / 4.3
- Advised by Professor Yaser Abu-Mostafa and Professor Yisong Yue

### University of Toronto

Bachelor of Computer Science

Toronto, Canada

Sep 2014 - Jun 2019

- GPA: 3.85 / 4.0
- Dean's Honour List for all academic years
- Graduated with Highest Honors

## RESEARCH INTEREST

My research focuses on leveraging artificial intelligence for drug discovery (AIDD), with a particular emphasis on developing and applying deep learning models for the representation and generation of small molecules and proteins. I aim to build **data-centric and knowledge-centric** methods to address the data scarcity problem in the AIDD domain, and build **agent systems** that can effectively accelerate the drug discovery process.

## WORK EXPERIENCE

### Institute for AI Industry Research, Tsinghua University (AIR)

Sep 2022 - Aug 2024

Full Time Research Engineer

### Applied Machine Learning (AML) at ByteDance

Jul 2021 - Sep 2022

Full Time Machine Learning Engineer

### Uber ATG

Jun 2020 - Sep 2020

Autonomous Driving Algorithm Intern

## PUBLICATIONS

1. Bowei He, Bowen Gao, Yankai Chen, Yanyan Lan, Chen Ma, Philip S. Yu, Ya-Qin Zhang, and Wei-Ying Ma. **S<sup>2</sup>Drug: Bridging Protein Sequence and 3D Structure in Contrastive Representation Learning for Virtual Screening**. In *AAAI Conference on Artificial Intelligence (AAAI)*, 2026.
2. Jianhui Wang\*, Wenyu Zhu\*, Bowen Gao\*, Xin Hong, Ya-Qin Zhang, Wei-Ying Ma, and Yanyan Lan. **Learning Protein–Ligand Binding in Hyperbolic Space**. In *AAAI Conference on Artificial Intelligence (AAAI)*, 2026.
3. Bowen Gao\*, Yanwen Huang\*, Yiqiao Liu, Wenxuan Xie, Bowei He, Haichuan Tan, Wei-Ying Ma, Ya-Qin Zhang, and Yanyan Lan. **CIDD: Collaborative Intelligence for Structure-Based Drug Design Empowered by LLMs**. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2025.
4. Wenyu Zhu\*, Jianhui Wang\*, Bowen Gao\*, Yinjun Jia, Haichuan Tan, Ya-Qin Zhang, Wei-Ying Ma, and Yanyan Lan. **AANet: Virtual Screening under Structural Uncertainty via Alignment and Aggregation**. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2025.

5. Bowen Gao\*, Haichuan Tan\*, Yanwen Huang, Minsi Ren, Xiao Huang, Wei-Ying Ma, Ya-Qin Zhang, and Yanyan Lan. **Reframing Structure-Based Drug Design Model Evaluation via Metrics Correlated to Practical Needs**. In *The Thirteenth International Conference on Learning Representations*, 2025.
6. Yanwen Huang\*, Bowen Gao\*, Yinjun Jia, Hongbo Ma, Wei-Ying Ma, Ya-Qin Zhang, and Yanyan Lan. **SIU: A Million-Scale Structural Small Molecule-Protein Interaction Dataset for Unbiased Bioactivity Prediction**. In *International Conference on Learning Representations (ICLR)*, 2025.
7. Bowen Gao\*, Minsi Ren\*, Yuyan Ni, Yanwen Huang, Bo Qiang, Zhi-Ming Ma, Wei-Ying Ma, and Yanyan Lan. **Rethinking Specificity in SBDD: Leveraging Delta Score and Energy-Guided Diffusion**. In *International Conference on Machine Learning (ICML)*, 2024.
8. Bowen Gao\*, Yinjun Jia\*, YuanLe Mo, Yuyan Ni, Wei-Ying Ma, Zhi-Ming Ma, and Yanyan Lan. **Self-supervised Pocket Pretraining via Protein Fragment-Surroundings Alignment**. In *International Conference on Learning Representations (ICLR)*, 2024.
9. Bowen Gao\*, Bo Qiang\*, Haichuan Tan, Yinjun Jia, Minsi Ren, Minsi Lu, Jingjing Liu, Wei-Ying Ma, and Yanyan Lan. **Drugclip: Contrastive protein-molecule representation learning for virtual screening**. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2023.
10. Bo Qiang, Yuxuan Song, Minkai Xu, Jingjing Gong, Bowen Gao, Hao Zhou, Wei-Ying Ma, and Yanyan Lan. **Coarse-to-fine: a hierarchical diffusion model for molecule generation in 3d**. In *International Conference on Machine Learning (ICML)*, 2023.

## PREPRINTS

---

1. Yinjun jia\*, Bowen Gao\*, Jiaxin Tan\*, Jiqing Zheng\*, Xin Hong\*, Wenyu Zhu, Haichuan Tan, Yuan Xiao, Yanwen Huang, Yue Jin, Yafei Yuan, et al. **Deep contrastive learning enables genome-wide virtual screening**. *Under Review*, 2025
2. Wenyu Zhu, Chengzhu Li, Xiaohe Tian, Yifan Wang, Yinjun Jia, Jianhui Wang, Bowen Gao, Ya-Qin Zhang, Wei-Ying Ma, and Yanyan Lan. **Coder as Editor: Code-driven Interpretable Molecular Optimization**. *arXiv preprint arXiv:2510.14455*, 2025
3. Bowen Gao\*, Yanwen Huang\*, Yiqiao Liu\*, Wenxuan Xie\*, Wei-Ying Ma, Ya-Qin Zhang, and Yanyan Lan. **Pharmagents: Building a virtual pharma with large language model agents**. *arXiv:2503.22164*, 2025
4. Yuanle Mo, Xin Hong, Bowen Gao, Yinjun Jia, and Yanyan Lan. **Multi-level Interaction Modeling for Protein Mutational Effect Prediction**. *arXiv:2405.17802*, 2024.

## ACADEMIC SERVICES

---

- Reviewer for International Conference on Learning Representations (ICLR) 2025, 2026
- Reviewer for AAAI Conference on Artificial Intelligence (AAAI) 2026
- Reviewer for Neural Information Processing Systems (NeurIPS) 2024, 2025
- Reviewer for International Conference on Machine Learning (ICML) 2025
- Reviewer for International Conference on Artificial Intelligence and Statistics (AISTATS) 2025
- Reviewer for IEEE Transactions on Neural Networks and Learning Systems (TNNLS)