

Education

Massachusetts Institute of Technology

Sep. 2025 - Present

PH.D. STUDENT IN ELECTRICAL ENGINEERING AND COMPUTER SCIENCE

- GPA: **5.0/5.0**
- Relevant Coursework: Deep Learning (**A+**), AI for Protein Biology (A)

Tsinghua University

Sep. 2021 - Jul. 2025

B.ENG. IN COMPUTER SCIENCE AND ENGINEERING

- GPA: **3.94/4.0 (19/178)**
- Relevant Coursework: Fundamentals of Engineering Biology (**A+**), Fundamentals of Programming (**A+**), Theory of Computer Network (**A+**), Computational Biology (A), Introduction to Artificial Intelligence (A), Introduction to Machine Learning (A), Artificial Neural Network (A), Data Mining (A)

Publications

- Bo Chen*, **Zhilei Bei***, Xingyi Cheng, Pan Li, Jie Tang, Le Song. *MSAGPT: Neural Prompting for Protein Structure Prediction via MSA Generative Pre-Training*. *NeurIPS* (2024).
- Bo Chen*, Xingyi Cheng*, Pan Li, Yangli-ao Geng, Jing Gong, Shen Li, **Zhilei Bei**, et al. *xTrimoPGLM: Unified 100B-Scale Pre-Trained Transformer for Deciphering the Language of Proteins*. *Nature Methods* (2025).
- Ajinkya Deshpande, **Zhilei Bei**, Jian Ma, Spencer Krieger. *MIMYR: Generative Modeling of Missing Tissue in Spatial Transcriptomics*. *Preprint on bioRxiv* (2025).
- Yujiang Li, Zhenyu Hou, Xiaohan Jia, Zihan Luo, **Zhilei Bei**, Rui Lu, Hong Huang, Jie Tang, Yuxiao Dong. *MLE-RL: Reinforcement Learning for Self-Improvement in Machine Learning Agents*. *Submitted to ICML* (2026).
- Boyan Wang, Yangliao Geng, Xingyi Cheng, Bo Chen, **Zhilei Bei**, et al. *ProtGO: Universal Protein Function Prediction Utilizing Multi-Modal Gene Ontology Knowledge*. *Bioinformatics* (2025).

Research Experiences

Massachusetts Institute of Technology

Rotation Advisor: Prof. Alex Rives

Sep. 2025 - Present

• Conditional Generation from Few-shot Examples for Protein Design

- Developed a masked diffusion language model for few-shot, context-conditioned protein generation.
- Incorporated classifier-free guidance to steer fitness prediction on deep mutational scanning benchmarks.

Carnegie Mellon University

Research Assistant, Advisor: Prof. Jian Ma

Jun. 2024 - Sep. 2025

• Predicting Missing Tissue Regions in Spatial Transcriptomics

- Developed a Variational Autoencoder model to learn cell representations from gene expression profiles.
- Developed a Vision Transformer model to generate spatial transcriptomics data for missing tissue regions.

BioMap Inc. Research Intern, Advisors: Prof. Le Song & Prof. Jie Tang

Aug. 2023 - Jun. 2024

• Protein Language Foundation Models

- Discovered multiple forms of pre-training scaling law tailoring different downstream tasks.
- Enhanced model generative performance by 300% with supervised fine-tuning and reinforced self-training.

• Protein Multiple Sequence Alignment (MSA) Generation

- Designed a simple yet effective MSA generation framework to aid protein structure prediction in low-MSA regime.
- Further enhanced model performance by leveraging feedback from AlphaFold2 for reinforcement learning.

Tsinghua Knowledge Engineering Group

Advisors: Prof. Yuxiao Dong & Prof. Jie Tang

Nov. 2022 - May. 2023

• Molecule Generation with Diffusion Model

- Developed a Graph Neural Network model for pocket-ligand pair affinity prediction.
- Designed a mixed training strategy utilizing unlabeled protein and molecule for data augmentation.
- Enhanced conditional generation performance by 13% using mixed training.

Tsinghua Machine Learning and Computational Biology Group

Advisor: Prof. Jianyang Zeng

Jul. 2022 - Sep. 2022

• Gene Selection and Drug Screening Based on Knowledge Graph and Community Detection

- Aligned embeddings from biological knowledge graphs with orthogonal gene and drug datasets for data integration.
- Tested algorithms for community detection on heterogeneous knowledge graph.