Zhilei **Bei**

Undergraduate Student · Tsinghua University

(+86) 173-1022-4006 | bzl21@mails.tsinghua.edu.cn | https://zhileibei.github.io

Education

Tsinghua University, B.Sc. Computer Science and Engineering *Sep. 2021 - Present*

- GPA: **3.93/4.0 (19/178)**; TOFEL: **119**; GRE: **336+5**
- 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)

Research Experience

Carnegie Mellon University *Research Assistant, Advisor: Prof. Jian Ma* *Jun. 2024 - present*

- **Tissue Generator: Predicting Missing Tissue Regions in Spatial Transcriptomics**
  - Developed a Variational Autoencoder (VAE) model to embed single cell 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 scaling law for 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, achieving state-of-the-art performance.
  - 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**
  - Designed a mixed training strategy utilizing unlabeled protein and molecule for data augmentation.
  - Developed a Graph Neural Network (GNN) model in python for pocket-ligand pair affinity prediction.
  - Enhanced conditional generation performance by 13% using classifier guidance.

Tsinghua Machine Learning and Computational Biology Group *Advisor: Prof. Jianyang Zeng* *Jun. 2022 - Dec. 2022*

- **Gene Selection and Drug Screening Based on Knowledge Map and Community Detection**
  - Aligned embeddings from biological knowledge graphs with orthogonal gene and drug datasets for data integration.
  - Designed algorithms for community detection on heterogeneous knowledge graph.

Institute of Biophysics, Chinese Academy of Sciences *Advisor: Prof. Pingsheng Liu* *Jan. 2019 - Sep. 2020*

- **Blocking the Binding of Bacterial Lipid Droplets to Phage DNA**
  - Discovered an effective blocker for MLDS protein and viral DNA binding through *in vitro* experiments.
  - Acquired expertise in laboratory techniques, including confocal microscopy imaging, spectrophotometry, cell culturing and Western blotting.

Publication

- Bo Chen*, **Zhilei Bei***, Xingyi Cheng, Pan Li, Jie Tang, Le Song. *"MSAGPT: Neural Prompting for Protein Structure Prediction via MSA Generative Pre-Training."* In *Proceedings of the 38th Annual Conference on Neural Information Processing Systems (NeurIPS)*, preprint at arXiv:2406.05347, 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."* Acceptance in Principle at *Nature Methods*, preprint at arXiv:2401.06199, 2024.

Honors & Awards

2023 **Comprehensive Scholarship**, Department of Computer Science and Technology, Tsinghua University *Beijing, China*

2022 **Comprehensive Scholarship**, Department of Computer Science and Technology, Tsinghua University *Beijing, China*

2020 **Finalist**, International Science and Engineering Fair (no specific awards announced due to COVID-19) *Virtual*

Skills

Programming Languages Python (Pytorch), C/C++, MATLAB, Rust, Javascript, HTML, CSS, SQL, x86 assembly

Tools & Frameworks Git, Django, MySQL, React