Siyu Long

✉ longsy@smail.nju.edu.cn · ○ https://github.com/longlongman · % https://longlongman.github.io

⚗ My research area

My research focuses on the exploration and development of generative and pre-trained models, aiming to deeply understand their wide applications in the scientific domain. This includes utilizing generative models for targeted drug design, innovative protein structure design, and the construction of pre-trained models for biological structures.

🎓 Education

Tsinghua University (THU), Beijing, China   2025 – Present
Postdoc in Computer Science (CS)

Nanjing University (NJU), Nanjing, China   2019 – 2024
Ph.D in Computer Science (CS)

Xi'an Jiaotong University (XJTU), Shaanxi, China   2015 – 2019
B.S. in Computer Science (CS)

👥 Experience

Tsinghua, Institute for AI Industry Research (AIR)   Jul. 2023 – Dec. 2024
Intern Student   Developing a unified pre-training model for multi-scale biological structures

ByteDance AI Lab   Sep. 2021 – Jul. 2023
Intern Researcher   Zero-shot drug design and protein backbone structure design based on generative models

📖 Publications

- **Long, S.**, Zhou, Y., Dai, X., Zhou, H. Zero-Shot 3D Drug Design by Sketching and Generating. (**NIPS22**)
- **Long, S.**, Zheng, K., Lu, T., Yang, J., Dai, X., Zhang, M., Nie, Z., Ma, W., Zhou, H. ESM All-Atom: Multi-Scale Protein Language Model for Unified Molecular Modeling. (**ICML24**)
- **Long, S.**, Gong, J. Pei, Y., Song, Y., Zhang, Z., Huang, W., Cao, Z., Zhang, S., Zhou, H., Ma, W. Steering Protein Family Design through Profile Bayesian Flow. (**ICLR25**)
- **Long, S.**, Wang, R., Tao, K., Zeng, J., Dai, X. Synonym Knowledge Enhanced Reader for Chinese Idiom Reading Comprehension. (**COLING20**)
- **Long, S.**, Wu, J., Zhou, Y., Sha, F., Dai, X. Deep Neural Networks for Knowledge-Enhanced Molecular Modeling. (**Neurocomputing**)
- Wang, R., **Long, S.**, Dai, X., Huang, S., Chen, J. Meta-LMTC: Meta-Learning for Large-Scale Multi-Label Text Classification. (**EMNLP21**)
- Zheng, N., **Long, S.**, Dai, X. BED: Bi-Encoder-Decoder Model for Canonical Relation Extraction.
- Yang, J., Zheng, K., **Long, S.**, Nie, Z., Zhang, M., Dai, X., Ma, W., Zhou, H. MOL-AE: Auto-Encoder Based Molecular Representation Learning With 3D Cloze Test Objective. (**ICML24**)
- Zhao, F., Wu, Z., **Long, S.**, Dai, X., Huang, S., Chen, J. Learning from Adjective-Noun Pairs: A Knowledge-Enhanced Framework for Target-Oriented Multimodal Sentiment Classification. (**COLING22**)