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

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

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

👥 Experience

Institute for AI Industry Research (AIR)   Jul. 2023 – Present
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. (2022). Zero-shot 3d drug design by sketching and generating. Advances in Neural Information Processing Systems, 35, 23894-23907. (**NeurIPS 2022, CCF-A**)
- **Long, S.**, Zheng, K., Lu, T., Yang, J., Dai, X., Zhang, M., ... & Zhou, H. (2024). Multi-Scale Protein Language Model for Unified Molecular Modeling. bioRxiv, 2024-03. (**ICML 2024, CCF-A**)
- **Long, S.**, Wang, R., Tao, K., Zeng, J., & Dai, X. (2020, December). Synonym Knowledge Enhanced Reader for Chinese Idiom Reading Comprehension. In Proceedings of the 28th International Conference on Computational Linguistics (pp. 3684-3695). (**COLING 2020, CCF-B**)
- **Long, S.**, Wu, J., Zheng, K., Zhang, J., Dai, X., Zhou, H., & Ma, W. (2024). Drug Property Prediction by Integrating Local and Global Information.
- Wang, R., **Long, S.**, Dai, X., Huang, S., & Chen, J. (2021, November). Meta-LMTC: meta-learning for large-scale multi-label text classification. In Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing (pp. 8633-8646). (**EMNLP 2021, CCF-B**)
- Zheng, N., **Long, S.**, & Dai, X. (2023). BED: Bi-Encoder-Decoder Model for Canonical Relation Extraction. arXiv preprint arXiv:2312.07088.
- Yang, J., Zheng, K., **Long, S.**, Nie, Z., Zhang, M., Dai, X., ... & Zhou, H. (2024). MOL-AE: Auto-Encoder Based Molecular Representation Learning With 3D Cloze Test Objective. bioRxiv, 2024-04. (**ICML 2024, CCF-A**)
- Zhao, F., Wu, Z., **Long, S.**, Dai, X., Huang, S., & Chen, J. (2022, October). Learning from adjective-noun pairs: A knowledge-enhanced framework for target-oriented multimodal sentiment classification. In Proceedings of the 29th International Conference on Computational Linguistics (pp. 6784-6794). (**COLING 2022, CCF-B**)