

YUANYUAN ZHANG

✉ zhang038@purdue.edu · ☎ (+1) 7657758429 · 🌐 Website · 📄 Google Scholar ·

- Expected graduation date: 2025.12

🎓 EDUCATION

Purdue University , West Lafayette, IN, US <i>Ph.D student</i> in Computer Science, GPA: 3.8/4.0	2021.9 – 2025.12
University of Chinese Academy of Sciences , Beijing, China <i>Master</i> in Computer Science	2018.8 – 2021.6
Sichuan University , Sichuan, China <i>Bachelor</i> in Computer Science and Technology	2014.9 – 2018.6

👤 EXPERIENCE

Samsung America , Seattle WA, US <i>Research intern</i> <i>Large Language Model post-training, Chain-Of-Thought reasoning on scaled model</i> <ul style="list-style-type: none">• Developed novel LLM-based log anomaly detection method that provides root cause analysis beyond detection, achieving 97% F1 score with 8B model (beating 235B baselines).• Proposed novel hierarchical Chain-of-Thought knowledge distillation from teacher model to student model significantly reducing training cost while maintaining performance.• Achieved 3.2x improvement in reasoning quality and 16x improvement in root cause diagnosis compared to larger LLMs, reducing debugging time by 60% in production tests	2025.5 – 2025.8
Noble Lab and Wang Lab, University of Washington Seattle, WA, US <i>Research intern</i> Advisor: Prof. William Stafford Noble and Sheng Wang <i>Self-supervised foundation model on Hi-C image data</i> <ul style="list-style-type: none">• developed a contrastive based masked autoencoder for Hi-C data.• integrated vision transformer as network architecture. <i>Neural radiance field(NeRF) on cryo-EM 3D image reconstruction</i> <ul style="list-style-type: none">• worked on NeRF application on cryo-EM 3D image• designed heterogeneity-aware encoder for cryo-EM heterogeneity problem• utilized hash encoding to effectively encode the query coordinates.	2024.6 – 2024.8
KiharaLab, Purdue University West Lafayette, IN, US <i>Research assistant</i> Advisor: Prof. Daisuke Kihara <i>Flow-matching on cryo-EM map segmentation</i> <ul style="list-style-type: none">• developed a conditioned rectified flow matching algorithm for cryo-EM 3D density image segmentation• used Unet based transformer architecture to effectively extract features <i>3D detection on Cryo-EM map</i> <ul style="list-style-type: none">• proposed a 3D-MUNet to accurately detect structural class from 3D density input. <i>Protein structure prediction based on AlphaFold2</i> <ul style="list-style-type: none">• proposed a fine-tuning pipeline to optimize structure module in AF2 under constraints.• designed multi-task loss to optimize structure prediction	2022.5 – Present
Key Laboratory of Network Data Science and Technology, CAS Beijing, China <i>Research assistant</i> Sentiment analysis and knowledge graph network by deep learning	2019.8 – 2021.6

SELECTED PUBLICATIONS

- **Zhang, Y.**, Bai, Y., Zhu, J., Hao, H. & Min, S. (2025) LLMAD-mini: Efficient Distilling Chain-of-Thought for Interpretable Log Anomaly Reasoning and Detection using Large Language Model. (*Submitted to ICLR 2026*)
- Jiang, Y.*, Li, X.*, **Zhang, Y.***, Han, J., Xu, Y., Pandit, A., ... & Liu, J. (2025). PoseX: AI Defeats Physics Approaches on Protein-Ligand Cross Docking. (*Submitted to ICLR 2026*) <https://arxiv.org/abs/2505.01700>
- **Zhang, Y.**, Wang X. , Terashi G. , & Kihara D. (2025) Flow-All: A generalized flow-matching model for 3D structure modeling on low resolution cryo-EM maps. (*In submission*)
- Liu, X.*, **Zhang, Y.***, Lu, Y., Yin, C., Hu, X., Liu, X., ... & Wang, X. (2025). Biomedical Foundation Model: A Survey. <https://arxiv.org/abs/2503.02104>
- Wang, X.*, **Zhang, Y.***, Ray, S., Jha, A., Fang, T., Hang, S., ... & Wang, S. (2024). A generalizable Hi-C foundation model for chromatin architecture, single-cell and multi-omics analysis across species. *bioRxiv. Nature Methods* (Under 2nd review). <https://www.biorxiv.org/content/10.1101/2024.12.16.628821v1.abstract>
- **Zhang, Y.**, Zhang, Z., Kagaya, Y., Terashi, G., Zhao, B., Xiong, Y., & Kihara, D. (2023). Distance-AF: Modifying Predicted Protein Structure Models by Alphafold2 with User-Specified Distance Constraints. *Accepted by Nature Communications Biology*. <https://www.biorxiv.org/content/10.1101/2023.12.01.569498v1>
- Hu, M., Ma, C., ..., **Zhang, Y.**, ... & Zhou, B. (2025). A Survey of Scientific Large Language Models: From Data Foundations to Agent Frontiers. *arXiv preprint arXiv:2508.21148*. <https://arxiv.org/abs/2508.21148>
- Xu, Z., Wu, F., **Zhang, Y.**, & Zhao, Y. (2024). Retrieval-reasoning large language model-based synthetic clinical trial generation. (*Accepted by 2025 SIGKDD*) <https://arxiv.org/abs/2410.12476>
- Qu, H., Wang, X., **Zhang, Y.**, Wang, S., Noble, W. S., & Chen, T. (2025). CryoNeRF: reconstruction of homogeneous and heterogeneous cryo-EM structures using neural radiance field. *Nature Biotechnology*. (Under 1st review) <https://www.biorxiv.org/content/10.1101/2025.01.10.632460v1>
- **Zhang, Y.**, Wang, X., Li, S., Terashi, G., Nakamura, T. & Kihara, D. (2024). DAQ-ATOM score for protein models evaluation from high-resolution Cryo-EM maps. *In submission to Nature Methods*
- Xu, Z., Wu, F., **Zhang, Y.**, & Zhao, Y. (2024). Retrieval-reasoning large language model-based synthetic clinical trial generation. *arXiv preprint arXiv:2410.12476*. <https://arxiv.org/abs/2410.12476>
- **Zhang, Y.**, Wang, X., Zhang, Z., Huang, Y., & Kihara, D. (2024). Assessment of Protein-Protein Docking Models Using Deep Learning. *Protein-Protein Docking: Methods and Protocols*, 149-162.
- Wang, X., **Zhang, Y.**, Yu, S., Liu, X., & Wang, F. Y. (2018). Computerized adaptive English ability assessment based on deep learning. In *Image and Video Technology: PSIVT 2017 International Workshops*, Wuhan, China, November 20-24, 2017, Revised Selected Papers 8 (pp. 158-171). Springer International Publishing.
- Wang, X., **Zhang, Y.**, Yu, S., Liu, X., Yuan, Y., & Wang, F. Y. (2017, October). E-learning recommendation framework based on deep learning. In *2017 IEEE international conference on systems, man, and cybernetics (SMC)* (pp. 455-460). IEEE.
- Farheen, F., Broyles, B. K., **Zhang, Y.**, Ibtihaz, N., Erkin, A. M., & Kihara, D. (2024). Predicting transcriptional activation domain function using Graph Neural Networks. *bioRxiv*, 2024-05.
- Bou Abdallah, F., Fish, J., Terashi, G., **Zhang, Y.**, Kihara, D., & Arosio, P. (2024). Unveiling the stochastic nature of human heteropolymer ferritin self assembly mechanism. *Protein Science*, 33(8), e5104.
- Gagliardi, L., Raffo, A., Fugacci, U., Biasotti, S., Rocchia, W., Huang, H., Amor, B.B., Fang, Y., **Zhang, Y.**, Wang, X. and Christoffer, C., 2022. SHREC 2022: Protein-ligand binding site recognition. *Computers & Graphics*, 107, pp.20-31.

- Lensink, M. F., Brysbaert, G., Raouraoua, N., Bates, P. A., Giulini, M., Honorato, R. V., ..., **Zhang, Y.**, ...& Wodak, S. J. (2023). Impact of AlphaFold on structure prediction of protein complexes: The CASP15 CAPRI experiment. *Proteins: Structure, Function, and Bioinformatics*, 91(12), 1658-1683.

🏆 HONORS AND AWARDS

D.E. Shaw DESRES Doctoral Fellowship, New York, US	2024.3
Graduate Student Scholarship of ICT CAS (Top 1%), Beijing, China	2018.9-2021.6
Outstanding Graduate of Sichuan University (Top 1%) Chengdu, China	2018.6
Outstanding Student of Sichuan University (Top 3%), Chengdu, China	2015.9-2018.6
National Endeavor Fellowship (Top 1%), Sichuan University, Chengdu, China	2016.9-2017.6

⚙️ SKILLS

- **Expertise:** Python, Deep Learning, Machine Learning, Pytorch, TensorFlow
- **Capable:** Hadoop, Spark, C, C++, Java, Matlab