Xingyue Liu

aroma@mail.dlut.edu.cn | 15842668147 | recynie.github.io
Github | Google Scholar

Profile

Dalian University of Technology Dalian, China

Artificial Intelligence, School of Future Technology

2023-2027 Undergraduate

Academic Score: 87.0/100
Language Test: CET-4 618, CET-6 591

Publications

1. **Xingyue Liu**†, Zijie. Xing†, Runze. Wang, Luoming. Hu, and Yanming. Shen*, "Language-guided expertise evolution for protein optimization," *ICLR 2026 Workshop RSI. (Accepted)* [Link]
2. R. Wang, Y. Ma, **Xingyue Liu**, Z. Xing, and Y. Shen*, "Motif-driven molecular graph representation learning," *Expert Systems with Applications*, vol. 269, p. 126484, 2025. *(Published)* [Link]
3. Runze Wang, Zijie Xing, **Xingyue Liu**, Mingqi Yang, Che He, Yanming Shen*, "From Graphs to Tokens: Substructure-Aware Molecular Representation for Large Language Models," *Information Processing and Management*. *(Published)* [Link]
4. Zijie Xing†, **Xingyue Liu**†, Runze Wang, Luoming Hu, Yanming Shen*, "KnowEvo: Knowledge Evolution for Protein Optimization," *NeurIPS 2026*. *(Under Review)* [PDF]

Research Experience

04/2025 – 05/2026

Language-Guided Expertise Evolution for Protein Optimization

Advisor: Prof. Yanming Shen
Overview: Developed a self-evolving agent system for protein optimization that improves accuracy through iterative "explore–verify–refine" cycles, maintaining a population of code blocks as a knowledge base to accumulate executable prior knowledge.
Core Contributions:
(1) Code as Knowledge: Structured code populations serve as the knowledge base; each code block integrates a gating mechanism to match and optimize target protein sub-distributions.
(2) Self-Evolving System: Built a high-concurrency self-evolving agent system from scratch, decoupling forward exploration from knowledge updates to improve efficiency, with multi-server architecture for better resource utilization.
(3) Search Algorithm: Thompson sampling during training and LCB scoring at inference to balance exploration–exploitation in selecting the final knowledge base.
Outcome: An early version won the Best Innovation Award at the Bohr+SciMaster AI4S Agent Development Competition (only 2 teams awarded) and was shortlisted for the Qizhi Startup Camp interview (top 10%). Published as co-first author at ICLR 2026 RSI Workshop [1] currently under review at NeurIPS 2026 [4].

10/2024 – 12/2024

Molecular Representation Learning

Advisor: Prof. Yanming Shen

Overview: Integrated high-level molecular motif information into both graph neural networks (GNNs) and large language models (LLMs) to enhance molecular representation learning, overcoming the expressive limits of 1-WL testing and 1D SMILES representations.

Core Contributions:

(1) Uni-Motif Framework (GNNs): Designed and evaluated three plug-and-play fusion mechanisms for incorporating molecular substructure information; conducted ablation studies on different motif decomposition strategies.

(2) S^2 Token Strategy (LLMs): Proposed and systematically evaluated a novel substructure-aware tokenization method; reproduced multiple state-of-the-art baselines to validate the advantage of motif-augmented approaches.

Outcome: Published as co-author in two peer-reviewed journals: *Expert Systems with Applications* [2] and *Information Processing and Management* [3].

12/2023 – Present

Prediction of Protein Crystallization Conditions

Advisor: Prof. Song Xue

Overview: Predicted protein crystallization conditions by modeling the complex relationships among protein sequences, 3D structures, and chemical environments.

Core Contributions:

- Dataset Construction: Led the full project pipeline to build a crystallization condition database (~7k proteins), extracting, filtering, and integrating crystallization records from the PDB.
- pLM Fine-tuning: Fine-tuned SaProt (a pretrained structure-aware protein language model) to capture latent relationships among protein sequence, structure, and crystallization conditions.
- Crystallization Condition Similarity Analysis: Clustered proteins by structural similarity using Foldseek, analyzing combinatorial effects of solution components and the distribution of crystallization conditions under similar structures.

Outcome: Experimental results indicate that pLMs struggle to effectively represent protein crystallization conditions, motivating the incorporation of hydration information into existing protein embedding models. Work currently in progress.

Awards & Honors

1. AY 2023–2024, AY 2024–2025 **Academic Excellence Scholarship**
2. AY 2024–2025 **National Scholarship**

Research Interests

- **Auto Scientific Research**: Improve the long-horizon scientific research ability of AI, focusing on iterating ideas with experimental feedback and proposing valuable ideas.

Skills

- **AI agent**: Self-evolving agents, agent harness design
- **AI for Science**: Protein representation learning, molecular representation learning, protein optimization