





YIJIA XIAO

 [yijia-xiao.github.io](https://github.com/yijia-xiao) |  [linkedin.com/in/yijia-xiao](https://www.linkedin.com/in/yijia-xiao) |  Google Scholar |  yijia.xiao@cs.ucla.edu
Los Angeles, CA, USA

OBJECTIVE

Seeking to apply my expertise in **Multimodal Large Language Models** and **LLM Agents** to:

- I. Multi-Modal Alignment:** Fuse domain-specific data (e.g., science, finance) with text/image modalities using LLMs.
- II. First Principles:** Derive complex properties from primary data like financial time series and biological sequences.

EDUCATION

• University of California, Los Angeles

Ph.D. Student in Computer Science

- Advisor: Professor [Wei Wang](#)

2022 – June 2027

Los Angeles, CA, USA

• Tsinghua University

Bachelor of Computer Science and Technology

- Advisor: Professor [Jie Tang](#)

2018 – 2022

Beijing, China

EXPERIENCE

• Amazon Web Services

Applied Scientist Intern

- Researched the applications of Large Language Model Agents in automated code deployment.
- Developed an LLM multi-agents framework to automate GitHub repositories deployment by enabling instruction understanding, shell script drafting, tool execution (e.g., Linux Shell, Perplexity Search), system log analysis, and issue database retrieval using Retrieval-Augmented Generation (RAG).
- First-Authored a paper based on internship work, accepted at NAACL Main Track (Publication [C.2]).
- Received excellent feedback rating from mentor and manager, resulting in a **return offer** for next year's internship.

June 2024 – Sep. 2024

Seattle, WA, USA

• NEC Labs America

Research Intern

- Conducted research on privacy-preserving fine-tuning of Large Language Models, proposing a lightweight tuning scheme that balances utility and privacy.
- Published the work at the EMNLP 2024 main conference (Publication [C.3]).
- Applied the method to medical scenarios, leading to a U.S. patent application filed on behalf of the company.

July 2023 – Sep. 2023

Princeton, NJ, USA

• Tsinghua University

Project Leader

- Led and Pre-trained the world's 2nd-largest Protein LLM **ProteinLM** (*Wen Su*), significantly improving contact prediction accuracy from 36% to 75%. [Wikipedia Page](#) for my work, part of WuDao LLM.
- Developed ProteinLM, a protein language model pretraining framework based on NVIDIA's Megatron-LM.
- Managed a cluster of 504 NVIDIA Tesla V100 (32G) GPUs to pre-train *Wen Su*.
- Model request form statistics show that ProteinLM has been used by researchers from over 20 institutions, including Harvard, University of Toronto, and NVIDIA, etc.

Dec. 2020 – Aug. 2021

Beijing, China

• Sequoia Capital China

Seed Group Intern

- Focused on investments in computer infrastructure, renewable energy (photovoltaic, hydrogen), and AI4Science.
- Contributed to over 20 investment deals by assisting in market analysis, technical roadmap reviews, due diligence, team roadshows, term sheet drafting, and more.

July 2022 – Sep. 2022

Beijing, China

• Toyota Technological Institute at Chicago (TTIC)

Research Assistant

- Led the project *Unsupervised Contact Prediction with Large MSA Language Model*.
- Proposed *fragment training* to balance the width and depth of Multiple Sequence Alignments (MSA), allowing pre-training of a 1-billion-parameter MSA model (the world's largest) using only 10% of the data required by previous state-of-the-art methods.
- Achieved a 3.5% improvement over state-of-the-art results and exceeded Meta AI's performance using significantly less training data.
- Demonstrated that increasing model scale enhances the capture of long-range dependencies in sequences.

June 2021 – June 2022

Chicago, IL, USA

• Tencent Inc.

Research Intern

- Developed models for protein multiple sequence alignments and localized protein database services.
- Published research on deep-learning-powered protein-ligand docking affinity prediction under data-scarce scenarios in *Briefings in Bioinformatics*.

Sep. 2021 – May 2022

Shenzhen, China

I. Large Language Model Agents

- [C.1] Yijia Xiao, Edward Sun, Di Luo, Wei Wang. (2025). **TradingAgents: Multi-Agents LLM Financial Trading Framework**. Accepted at *Multi-Agent AI in the Real World, AAAI 2025*.
 - TradingAgents has received interest from venture capital firms and financial startups.
 - See a financial expert's [Summary on LinkedIn](#). Visit the [TradingAgents Homepage](#) for more details.
- [C.2] Yijia Xiao, Runhui Wang, Luyang Kong, Davor Golac, Wei Wang. (2024). **GSR-Bench: Benchmarking Deployment in GitHub Science Repositories with Large Language Model Agents**. Accepted at *NAACL 2025*.
 - GSR-Agents is a multi-agents LLM framework for automated code repository deployment.
 - My internship performance received excellent feedback from my mentor & manager, leading to a return offer. The paper and framework will be released soon.
- [C.3] Yiqiao Jin, Qinlin Zhao, Yiyang Wang, Hao Chen, Kaijie Zhu, Yijia Xiao, Jindong Wang. (2024). **AgentReview: Exploring Peer Review Dynamics with LLM Agents**. In *Empirical Methods in Natural Language Processing (EMNLP) 2024*.

II. Multimodal Large Language Model

- [S.1] Yijia Xiao, Edward Sun, Yiqiao Jin, Qifan Wang, Wei Wang. (2024). **ProteinGPT: Multimodal LLM for Protein Property Prediction and Structure Understanding**. Submitted to *KDD 2025*.
 - Collaboration with *Meta AI*: ProteinGPT is a multi-modal Protein ChatGPT-like system that integrates sequence and structure encoders with an LLM for precise analysis and responses.
- [C.4] Yijia Xiao, Edward Sun, Yiqiao Jin, Wei Wang. (2024). **RNA-GPT: Multimodal Generative System for RNA Sequence Understanding**. In *Machine Learning for Structural Biology Workshop, NeurIPS 2024*.
- [S.2] Yijia Xiao, Edward Sun, Wei Wang. (2024). **LogicVista: Multimodal LLM Logical Reasoning Benchmark in Visual Contexts**.
- [C.5] Tianyu Liu, Yijia Xiao, Xiao Luo, Hua Xu, W. Jim Zheng, Hongyu Zhao. (2023). **Geneverse: A Collection of Open-source Multimodal Large Language Models for Genomic and Proteomic Research**. In *Empirical Methods in Natural Language Processing (EMNLP) 2024*.
- [C.6] Fan Zhang, Changhu Wang, et al., Dongjie Wang, Yijia Xiao, Chong Chen, Xian-Sheng Hua, Xiao Luo. (2025). **DREAM: Decoupled Discriminative Learning with Bigraph-aware Alignment for Semi-supervised 2D-3D Cross-modal Retrieval**. In *AAAI 2025*.

III. AI for Biology & Healthcare

- [C.7] Yijia Xiao, Jiezhong Qiu, Ziang Li, Chang-Yu Hsieh, Jie Tang. (2021). **Modeling Protein Using Large-scale Pretrain Language Model**. In *Pretrain Workshop, KDD 2021*.
 - The 1st and **largest** Protein LLM in China as of 2021-2023, ProteinLM (Wen Su), is part of the Wu Dao project.
 - Wikipedia page covering Protein LLM Wen Su: [Wu Dao](#).
- [J.1] Zi-Yi Yang, Zhao-Feng Ye, Yi-Jia Xiao, Chang-Yu Hsieh, Sheng-Yu Zhang. (2022). **SPLDExtraTrees: Robust Machine Learning Approach for Predicting Kinase Inhibitor Resistance**. In *Briefings in Bioinformatics*.
- [S.3] Yijia Xiao, Dylan Steinecke, Alexander Russell Pelletier, Yushi Bai, Peipei Ping, Wei Wang. (2023). **Know2BIO: A Comprehensive Dual-View Benchmark for Evolving Biomedical Knowledge Graphs**.
- [C.8] Mingyu Ma, Yijia Xiao, Anthony Cuturrufo, Xiaoxuan Wang, Vijay Nori, Wei Wang. (2024). **Memorize and Rank: Elevating Large Language Models for Clinical Diagnosis Prediction**. In *AAAI 2025*.
- [C.9] Rakesh Bal, Yijia Xiao, Wei Wang. (2023). **PGraphDTA: Improving Drug Target Interaction Prediction using Protein Language Models and Contact Maps**. In *NeurIPS AI for Science Workshop 2023*.
- [S.4] Yijia Xiao, Fang Sun, Xiao Luo, Yizhou Sun, Wei Wang. (2024). **RLMC: RNA Language Model with Contrastive Learning**.

IV. LLM Security & Evaluation

- [C.10] Yijia Xiao, Yiqiao Jin, Yushi Bai, Yue Wu, Xianjun Yang, Xiao Luo, Wenchao Yu, Xujiang Zhao, Yanchi Liu, Quanquan Gu, Haifeng Chen, Wei Wang, Wei Cheng. (2023). **PrivacyMind: Large Language Models Can Be Good Privacy Protection Learners**. Accepted at *Empirical Methods in Natural Language Processing (EMNLP) 2024*.
- [P.1] Wei Cheng, Wenchao Yu, Yanchi Liu, Xujiang Zhao, Haifeng Chen, and Yijia Xiao. (2023). **Privacy Protection Tuning for LLMs in Medical Decision Making**. Patent Application No. Docket 23060.
- [C.11] Yushi Bai, Jiahao Ying, Yixin Cao, et al., Yijia Xiao, Haozhe Lyu, Jiayin Zhang, Juanzi Li, Lei Hou. (2023). **Benchmarking Foundation Models with Language-Model-as-an-Examiner**. In *NeurIPS 2023*.

HONORS AND AWARDS

- **Research Excellence Scholarship**, Tsinghua University, 2021. Ranked top 2%, 3rd out of 230 students.
- **Science and Technology Innovation Scholarship**, Tsinghua University, 2021. Ranked top 2%, 3rd out of 230 students.
- **Technology Innovation Scholarship**, Tsinghua University, 2020. Ranked top 5%, 6th out of 230 students.
- **Silver Medal**, International Collegiate Programming Contest (ICPC) Asia East Continent Final, 2020.
- **Gold Medal**, International Collegiate Programming Contest (ICPC) Asia Regional Contest, 2020.
- **First Prize**, Chinese Collegiate Physics Olympiad, 2019.
- **National Bronze Medal**, Chinese Physics Olympiad, 2017.
- **National Bronze Medal**, China Western Mathematics Invitational, 2017.

LEADERSHIP EXPERIENCE

- **Business Development Team** | *Chinese Entrepreneurs Organization* 2022 – 2024
Engaged in business development activities at Stanford, CA
- **Organizer** | *Tsinghua Future Internet Computing Club* 2020 – 2022
Organized events and workshops related to future internet computing
- **Member** | *Tsinghua Entrepreneurship Training Program* 2020 – 2022
Selected as one of 30 members per year in Tsinghua University

PROFESSIONAL MEMBERSHIPS

- **Reviewer**, NeurIPS 2023/2024, ICML 2023/2024, ICLR 2023, AISTATS 2024, ACM Multimedia, PLOS ONE, etc.