Jinglin (Ollie) Jian

(+1)217-819-7884 jjian@scripps.edu jianjinglin.github.io linkedin.com/in/olliejian1999

Education

The Scripps Research Institute San Diego, California, US
Ph.D. Program *Aug 2025 - May 2029*
Current Focus: AI for Drug Discovery, advised by Prof. Stefano Forli
Recipient of the **Kellogg Fellowship**, funded by the Kellogg Family and the ALSAM Foundation

University of Illinois Urbana-Champaign (UIUC) Illinois, US
M.S. in Information Sciences *Aug 2023 - May 2025*
Course: ML for Bioinformatics, Text Information Systems, Large Language Models, Data Mining

Peking University (PKU) Beijing, China
B.Econ. in Economics Minor *Sep 2021 - Jul 2023*
Course: Statistics, Applied Econometrics, Causal Inference Models

Beijing Normal University (BNU) Beijing, China
B.S. in Computer Science and Education *Sep 2017 - Jul 2021*
Course: Data Structure, Database, WebDev, OOP, Information Retrieval, Intelligent System, Data Mining

Publications

Google Scholar: https://scholar.google.com/citations?hl=en&user=pXY2xYkAAAAJ

[Ongoing] **Jian, J.**, ..., & Chen, Q. The AI Scientist in Health: Potential, Challenges, and the Road Ahead.

[BPS'26] Torres-Paris, C., ..., **Jian, J.**, Holcomb, M., Forli, S., & Racki, L.R. Allosterically inhibiting *Pseudomonas aeruginosa*'s polyphosphate kinase 2A by disrupting its oligomerization. *Biophysical Society Annual Meeting* (2026).

[ACL'26 Submitted] Wang, Q., ..., **Jian, J.**, Guo, X., & Li, R. Human-AI Co-Discovery. *in submission of ACL 2026*

[TechRxiv'25] **Jian, J.**, ..., Chen, Q., Lu, Z., & Wang, Q. Exploring Agentic Multimodal Large Language Models: A Survey for AIScientists. (2025). *in submission of ACL 2026*

[arXiv'24] Liu, H., Li, Y., **Jian, J.**, Cheng, Y., ... & Wang, H. Toward a Team of AI-Made Scientists for Scientific Discovery from Gene Expression Data. *arXiv preprint arXiv:2402.12391* (2024).

[IEEE Big Data'24] **Jian, J.**, ..., & Chen, J. Big Data-Driven Computational Aptamer Design Framework via Parallel Monte Carlo Tree Search. *IEEE International Conference on Big Data* (2024).

[Accepted] Li, Z., **Jian, J.**, ..., & Zhang, Y. Patient Outcome Predictions via a Multimodal Language Model for Electronic Health Records. *IEEE International Conference on Big Data* (2024).

[Published] Hou, R., Zhou, D., & **Jian, J.** GeoCM: Exploring Consistency Models and EGNNs for Molecular 3D Structure Prediction. *UIUC Machine Learning for Bioinformatics Workshop* (2024).

[AIED'24] Xiao, Y. & **Jian, J.** Which Animal Would You Like to See on Your Flashcards? Designing Visual Aids Together with Kids Using GIMs. *The 25th International Conference on Artificial Intelligence in Education.*

Research Experience

[Ongoing] Agentic Multimodal AI System for Drug Discovery *Aug 2024 – Present*
Advisor: Dr. Stefano Forli (Scripps)

- Fine-tuning a **large model** for **chemistry reasoning** by extracting small molecular data from literature and databases, with human reasoning demonstrations for benchmarking.
- Developing an **agentic AIScientist** system for virtual screening, providing an explainable platform to lower the barrier for scientists without computational expertise.

[BPS'26] Virtual Screening Pipeline for Antibiotic Discovery *Aug 2024 – Present*
Advisor: Prof. Lisa Racki, Prof. Stefano Forli (Scripps)

- Established a virtual screening pipeline targeting Ppk2A in *P. aeruginosa* (WHO Critical Priority Pathogen), utilizing **AutoDock-GPU** and **AlphaFold3** to screen **2.4 million** compounds, identifying **23 candidates** currently undergoing wet-lab experimental validation.

[TechRxiv'25] Exploring Agentic Multimodal Large Language Models: A Survey for AIScientists *Jun 2025 – Nov 2025*
Advisor: Prof. Qingyu Chen (Yale), Prof. Zhiyong Lu (NIH), Prof. Qingyun Wang (William & Mary)

- Lead author of a survey framing AIScientists from a machine learning centric and agentic multimodal LLM perspective.

[arXiv'24] GeoCM: Exploring Consistency Models and EGNNs for Molecular 3D Structure Prediction *Oct 2024 – Dec 2024*
Advisor: Prof. Ge Liu (UIUC)

- Utilized **Equivariant Graph Neural Networks (EGNN)** and **Consistency Models (CM)** to train a self-supervised model to predict molecular 3D structures.
- Established two metrics **Coverage Rate** and **Matching Error** to compare GeoCM models against other models, demonstrating GeoCM **claimed a new SOTA**.

[Big Data'24] Fast and Accurate Drug Discovery Framework *Jan 2024 – Oct 2024*
Advisor: Prof. Yang Zhang (UIUC) and Dr. Jin Chen (Cleveland Clinic)

- Developed an enhanced parallel **Monte Carlo Tree Search** framework, considering aptamers' high-affinity and specificity for target proteins, achieving **98-fold** computational efficiency and **7.59-fold** improved sequence quality.

[arXiv'24] Team of AI-made Scientists (TAIS) *Aug 2023 – Feb 2024*
Advisor: Prof. Haohan Wang (UIUC)

- ML can discover disease-predictive genes from gene expression data. We introduced **TAIS**, a LLM-based framework for automatic streamlining ML analysis, outperforming **GPT4/MetaGPT/AutoGPT**.
- Fetched data from the **GEO/TAGC** database. Created the **GenoTEX**, a NEW benchmark for evaluating the exploration of genomics data, with aligning gene symbols, logging, and statistical corrections.
- Created several **agents** as scientists, via autonomously creating codes (**template-based prompting**), execution (**subprocess**), outputs/errors capture (**logger**), and built communication within (**Data Engineer**, **Code Reviewer**, and **Domain Expert** agents).

[Bachelor's Thesis] Semi-automatic Knowledge Graph Construction *Sep 2020 - Jul 2021*
Advisor: Prof. Qinhua Zheng (BNU)

- Developed a **semi-automatic** ML paradigm for **knowledge graph creation** for addressing time-consuming issues by combining **supervised ML** with **human-in-the-loop** incorporation.
- Developed a benchmark dataset for educational entities by annotating transcriptions using BIO tagging.
- Iterated a supervised **BiLSTM-CRF** model for **entity recognition** and dynamic term re-ranking (**mutual information** and **human feedback**), improving F1-score (0.54 → **0.76**).

Professional Experience

Software Developer Intern *May 2024 – Aug 2024*
Supervisor: David Bachtler and Ian Cowen, ReDirect
– Engineered subscription functionality using Flutter framework and implemented unit testing.

Teaching Assistant, IS 589 Independent Study *Aug 2023 – Dec 2023*
Supervisor: Prof. Haohan Wang, University of Illinois Urbana-Champaign
– Developed and published a comprehensive ML tutorial, the "Gold-Standard-Pipeline-Guide", teaching students to analyze gene expression data for disease-related gene discovery. [GitHub]
– The resource has been forked by 9 students who applied it to their own research projects.

Research Assistant *Aug 2023 – Aug 2024*
Supervisor: Prof. Mackenzie Alston, University of Illinois Urbana-Champaign
– Conducted literature review (randomized controlled trials) using Zotero and scraped 2000+ emails.

Head on Online Learning Department & Teacher Volunteer *May 2019 – Aug 2022*
China Starry Night (NGO) [Web]
– Designed online educational videos to help intangible cultural heritage artisans in rural regions share traditional crafts, enabling younger generations to preserve cultural treasures.
– Courses published on Bilibili (China's leading video platform) garnered 10,000+ views.
– Presented at the 5th China Education Innovation Expo (National Award - Top 1%).

Selected Projects

– **Demo of ChemTutor: AI Q&A system with Chemistry Textbooks** [Code] *2024*
#LLM #Q&A system #RAG #LangChain
– **HMM-DRL Model for Data-driven Auto-Trading** [Paper] *2022*
#Reinforcement Learning #Hidden Markov Model #Time Series Data #Financial Index
– **Evolution of Key Themes in Learning Sciences** [Web] *2020*
#Text Mining #LDA-Topic Model #TF-IDF #Data Visualization

Honors and Awards

– **Kellogg Fellowship**, The Scripps Research Institute *Aug 2025 – Aug 2028*
– **National Innovation and Entrepreneurship Training Award (1%)**, Ministry of Education *2021*
– **Outstanding Teaching Practice (1%)**, Beijing Normal University *2021*
– **Jianghaiziqiang Fellowship (1%)**, Beijing Normal University *2020*
– **First-class Scholarship for Competition Excellence (1%)**, Beijing Normal University *2019*
– **First-class Scholarship for Academic Excellence (10%)**, Beijing Normal University *2017 - 2021*

Skill Set

Machine Learning & NLP	TensorFlow, PyTorch, LangGraph, LangChain, sklearn, NLTK
Programming Language	Python, Java, C, JavaScript, HTML, Stata
Framework & Database	React, Node.js, RESTful API, MySQL, MongoDB, Neo4j
Cloud	AWS, EC2, API Gateway
Code Management & Others	Git/Github, Docker, Tableau, Unit Testing, LaTeX