Jinglin (Ollie) Jian

(+1) 217-819-7884   jj50@illinois.edu   jianjinglin.github.io   https://www.linkedin.com/in/olliejian1999/

EDUCATION

The Scripps Research Institute, Skaggs Graduate School — California, US
Doctoral Program in Chemical and Biological Sciences — *Aug 2025 - May 2029*

University of Illinois Urbana-Champaign, School of Information Sciences — Illinois, US
M.S. in Information Sciences, GPA: 3.92/4.0 — *Aug 2023 - May 2025*
Course: ML for Bioinformatics, Text Information Systems2, Large Language Models, Data Mining

Peking University, National School of Development — Beijing, China
B.Econ. Minor — *Sep 2021 - Jul 2023*
Course: Statistics, Applied Econometrics, Causal Inference Models

Beijing Normal University, School of Educational Technology — Beijing, China
B.S. in Educational Technology, GPA: 3.76/4.0 — *Sep 2017 - Jul 2021*
Course: Data Structure, Database, WebDev, OOP, Information Retrieval, Intelligent System, Data Mining

PUBLICATIONS AND CONFERENCE

[1] Hou, R., **Jian, J.**, & Zhou, D. (2024). **GeoCM: Exploring Consistency Models and EGNNs for Molecular 3D Structure Prediction**. In *CS582 ML for Bioinformatics Workshop*.

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

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

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

[5] Xiao, Y. and **Jian, J**. (2024). **Which Animal Would You Like to See on Your Flashcards? Designing Visual Aids Together with Kids Using GIMs**. In *The 25th International Conference on Artificial Intelligence in Education* [Website]

RESEARCH EXPERIENCE

[arXiv'24] GeoCM: Exploring Consistency Models and EGNNs for Molecular 3D Structure Prediction — *Oct 2024 – Present*
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 (Beijing Normal University)
– Developed a **semi-automatic** 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 ($0.54 \rightarrow$ **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.

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 and Teacher Volunteer *May 2019 – Aug 2022*
China Starry Night (non-profit organization) [Web]
– SPresented 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 (Retrieval-Augmented Generation) #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

– **National Innovation and Entrepreneurship Training Award (1%)**, Ministry of Education *2021*
– **Outstanding Teaching Practice (1%)**, Beijing Normal University *2021*
– **Jianghaiziqiang Scholarship (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, 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