

Shiyu Jiang

Department of Computer and Information Science, Herbert Wertheim College of Engineering

1345 CENTER DR, Gainesville, FL 32610

☎ (667) 379-4478

✉ jiang.shiyu@ufl.edu

🌐 jasonjiangs.github.io

GitHub: github.com/JasonJiangs

ORCID: orcid.org/0000-0001-9450-7379

Last updated: September 21, 2026

RESEARCH INTERESTS

AI-driven drug discovery: generative modeling and representation learning for molecular prediction, design, and optimization; drug-induced cellular phenotype prediction; and machine learning-guided protein evolution.

Language models for life sciences: single-cell foundation models and language models for protein and nucleic acid sequences, spanning model development, evaluation, and applications.

Agent memory systems and continual learning: parametric and non-parametric memory, long-horizon adaptation, self-evolving agents, reinforcement learning, and agentic systems for scientific discovery.

EDUCATION AND TRAINING

Ph.D. Student in Computer Science

Aug. 2026 – Present

Dept. of Computer and Information Science, University of Florida, Gainesville, FL, USA

Selected coursework: PHA 6195 - AI in Drug Discovery;

Advisor: Prof. Yanjun Li

Ph.D. Student in Computational Biology and Bioinformatics

Aug. 2025 – May. 2026

Dept. of Quantitative and Computational Biology, University of Southern California, Los Angeles, CA, USA

Selected coursework: QBIO 570 - Intro to Data Structures & Algorithms for Science; QBIO 540 - Prob & Stats for Computational Sciences; QBIO 502 - Molecular Biology for Quantitative Scientists;

M.S.E. in Computer Science, Bioinformatics Track

Aug. 2022 – Dec. 2023

Dept. of Computer Science, Johns Hopkins University, Baltimore, MD, USA

Selected coursework: CSCI 601.649 - Computational Genomics: Applied Comparative Genomics; CSCI 601.646: Sketching & Indexing for Sequences;

Thesis Advisor: Prof. Kimia Ghobadi; Course Advisor: Prof. Ben Langmead

Committee: Kimia Ghobadi, Anton Dahbura, Claus Aranha (Univ. of Tsukuba)

Thesis: Analyzing epidemic spread and emotion contagion with agent-based simulation and system dynamic modeling

B.S. in Computer Science; Minor in Mathematical Sciences

Aug. 2018 – May. 2022

Dept. of Computer Science, Wenzhou-Kean University, Wenzhou, China

Advisor: Prof. Aloysius Wong

PUBLICATIONS & PREPRINTS

([†] [‡] = equal contribution; * = co-corresponding author [Google Scholar](#): Citations 373; h-index 9; i10-index 9)

Under Review

- **Jiang S**, Wei J, Taghavi A, Yang Z, Seabra G, Childs-Disney JL, Li C, Disney MD, Li Y^{*}. SMARTFlexDB: a paired apo-holo RNA structure database and analysis platform for RNA conformational flexibility and small molecule recognition. *bioRxiv*. Under review.
- **Jiang S[†]**, Fang Z[†], Zhang Y, Zhang X, Kalfon J, Wang W, Fu X, Patel A, Rizvi S, Lu YR, He S, Wang Y, Lu Y, Xu W, Wang M, Miao Z, Lin J, Ding J, Wang ZJ, Ouyang W, Chen T, Li M, Ma J, Wang F, Xie Y, Tang J, Rabadan R, van Dijk D, Xie P, He P, Fox EB, Song L, Theis FJ, Xing E, Theodoris CV^{*}, Qiu X^{*}, Ding J^{*}. The landscape of single-cell foundation models: design principles, applications, and open challenges. *bioRxiv*. Under review.
- **Jiang S[†]**, Taghavi A[†], Wang T, Sung K, Meyer SM, Springer NA, Wei J, Childs-Disney JL, Li C, Disney MD^{*}, Li Y^{*}. Small Molecule Approach to RNA Targeting Binder Discovery (SMARTBind) Using Deep Learning Without Structural Input. *bioRxiv*, Under review.
- Zheng X[†], Zhou Z[†], **Jiang S**, Li K, Qiu P, Cheng L, Wei J, Mu Y, Seabra G, Li C, Bian Y, Li Y^{*}. Beyond the Rigid Receptor: Coupled De Novo Ligand Generation with Protein Adaptation for Flexible Structure-Based Drug Design. Under review.
- Ding J^{†*}, Lin J[†], Mao Z[†], Mechtel N[†], **Jiang S**, Wang Y, Fang Z, Martin-Rufino JD, Weng C, Saunders R, Weissman JS, Ouyang W, Li M^{*}, Tang J^{*}, Lu R.Y.^{*}, Qiu X^{*}. Predictive single cell foundation model for gene regulation and aging with privacy-preserving tabular learning. *arXiv*. Under review.
- Zhou X[†], Han C[†], Zhang Y[†], Du H[†], Tian J[†], Su J[†], Liu R[†], Zhuang K[†], **Jiang S[†]**, Gitter A, Liu L, Li H, Wu M, You S, Yuan Z, Ju F, Zhang H, Zhang W, Dai F, Zhou Y, Tao Y, Wu D, Zhao Z, Liu Y^{*}, Lu H^{*}, Yuan F^{*}. Decoding the Molecular Language of Proteins with Evolla. *bioRxiv*. Under review.

Selected Publications

- **Jiang S***, Liu X, Wang JZ*. Evaluating DNA function understanding in genomic language models using evolutionarily implausible sequences. *ACS Synthetic Biology*, 2026.
- Cheng L[†], Zheng X[†], **Jiang S[†]**, Hu Y, Liu Y, Yang K, Rui J, Ding H, Zhang M, Yuan T, Ye H, Li C, Yang KK, Huang X*, Xiao H*. Sequence Display enables large-scale sequence-activity datasets for rapid protein evolution. *Nature Biotechnology*, 2026. [Cover article] [*Research Highlight in Nature Methods*]
- Ding J[†], Lin J[†], **Jiang S[†]**, Wang Y, Mao Z, Fang Z, Tang J*, Li M*, Qiu X*. Tabula: A Tabular Self-Supervised Foundation Model for Single-Cell Transcriptomics. In *NeurIPS*, 2025.
- Su J[†], He Y[†], You S[†], **Jiang S**, Zhou X, Zhang X, Wang Y, Su X, Tolstoy I, Chang X*, Lu H*, Yuan F*. A tri-modal protein language model enables advanced protein searches. *Nature Biotechnology*, 2025: 1–7.
- **Jiang S***, Kim H, Tanaka F, Aranha C*, Bogdanova A, Ghobadi K, Dahbura A. Simulating Disease Spread during Disaster Scenarios. In *Conference on Artificial Life (ALIFE)*, 2023.
- **Jiang S**, Abdalla H, Bi C, Zhu Y, Tian X, Yang Y*, Wong A*. HNOXPred: a web tool for the prediction of gas-sensing H-NOX proteins from amino acid sequences. *Bioinformatics*, 38(19): 4643–4644, 2022.

Journal Articles

- Cheng L[†], Zheng X[†], Ding H[†], **Jiang S**, Hu Y, Wang C, Li C, Rui J, Ye H, Liu Y, Tian Z, Yuan T, Huang X*, Xiao H*. Sequence Display generates large-scale sequence-activity datasets for reprogramming SlugCas9 toward difficult-to-access PAMs. *Nature Communications*, Accepted in principle.
- Yuan T[†], Zhang M[†], Cheng L, Zheng X, **Jiang S**, Huang X, Xiao H*. Biocatalytic Synthesis of N-protected α -Amino Acids through 1,3-Nitrogen Migration by Nonheme Iron Enzymes. *Journal of the American Chemical Society*, 2025.
- Su J, Li Z, Han C, Zhou Y, He Y, Shan J, Zhou X, Chang X, **Jiang S**, Ma D, OPMC, Steinegger M, Ovchinnikov S, Yuan F*. Democratizing protein language model training, sharing and collaboration. *Nature Biotechnology*, 2025: 1–7.
- Hu Y, Wang Y, Cheng L, Wang C, Liu Y, Wang Y, Chen Y, Yang S, Guo Y, **Jiang S**, Yang K, Xiao H*. Engineering unnatural cells with a 21st amino acid as a living epigenetic sensor. *Nature Communications*, 16(1): 9388, 2025.
- Hu Y[†], Cheng L[†], Liu Y, Liu R, **Jiang S**, Yuan T, Wang Y, Ye H, Xiao H*. Biosynthesis of Unnatural Cyclodipeptides through Genetic Code Expansion and Cyclodipeptide Synthase Evolution. *Journal of the American Chemical Society*, 147(38): 34517–34526, 2025.
- Ye H, **Jiang S**, Yan Y, Zhao B, Grant ER, Kitts DD, Yada RY, Singh AP, Baldelli A, Yang T*. Integrating Metal Phenolic Networks-Mediated Separation and Machine Learning-Aided SERS for High-Precision Quantification and Classification of Nanoplastics. *ACS Nano*, 18(38): 26281–26296, 2024. [Cover article].
- Guo Y[†], Cheng L[†], Hu Y, Zhang M, Liu R, Wang Y, **Jiang S**, Xiao H*. Biosynthesis of halogenated tryptophans for protein engineering using genetic code expansion. *ChemBioChem*, 25(20): e202400366, 2024. [VIP paper].

Conference Articles

- Zheng X, **Jiang S**, Seabra G, Li C, Li Y*. Apo2Mol: 3D Molecule Generation via Dynamic Pocket-Aware Diffusion Models. In *AAAI*, 2026.
- Qi X[†], Zhang Z[†], Handoko AB[†], Zheng H, Chen M, Huy TD, Ohan VMH, Zhang L, Cheng L, **Jiang S**, Zhang Z, Liao Z, Zhao Y*, To MS*. ProjectedEx: Enhancing Generation in Explainable AI for Prostate Cancer. In *IEEE CBMS*, 2025.
- Song Z[†], Ouyang G[†], Fang M^{†*}, Na H, Shi Z, Chen Z, Fu Y, Zhang Z, **Jiang S**, Fang M, Chen L, Chen X*. Hazards in Daily Life? Enabling Robots to Proactively Detect and Resolve Anomalies. In *NAACL*, 2025.
- Zheng S, Wu Y, **Jiang S**, Lu C, Gupta G*. Deblur-yolo: Real-time object detection with efficient blind motion deblurring. In *IJCNN*, 2021.

Workshop Papers

- **Jiang S***, Liu X, Wang JZ*. Predicting function of evolutionarily implausible DNA sequences. In *ICML 2025 Workshop on Generative AI and Biology*.

RESEARCH EXPERIENCE

Graduate Research Assistant, University of Florida

Aug. 2026 – Present

Advisors: Prof. Yanjun Li

- Developed SMARTBind for sequence-based RNA–ligand binding prediction and virtual screening.
- Built SMARTFlexDB, a paired apo–holo RNA structure database for ligand-induced conformational changes.

Remote Research Assistant, Dept. of Chemistry, Rice University

2024 – 2026

Advisors: Prof. Han Xiao & Dr. Kevin K. Yang (Microsoft Research)

- Assisted protein evolution by fine-tuning protein language models on sequence display assay data.

- Analyzed multiple assay datasets, including next-generation sequencing and nanopore sequencing data.

Remote Research Assistant, College of Pharmacy, University of Florida

2023 – 2026

Advisors: Prof. Yanjun Li & Prof. Matthew D. Disney (Scripps Research)

- Developed SMARTBind for RNA–ligand interaction prediction, enabling the identification of novel small molecules targeting the precursor of oncogenic microRNA-21, with hits validated by *in vitro* and cellular assays.
- Developed Apo2Mol, a diffusion-based molecular generation model for dynamic drug discovery.

Visiting Master Student, Dept. of Genetics, Stanford University

2023 – 2026

Advisors: Prof. Xiaojie Qiu

- Pretrained a single-cell foundation model, Tabula, using tabular learning and federated learning.
- Designed an *in silico* perturbation experiment to predict gene regulation across developmental biological systems including Hematopoiesis, Cardiogenesis, Neurogenesis, and Pancreatic Endogenesis.
- Benchmarked with scGPT, Geneformer, and scFoundation on multiple cell-level and gene-level tasks.

Research Associate, Westlake University

Aug. 2024 – Jul. 2025

Advisors: Prof. Fajie Yuan & Prof. Zitong Jerry Wang

- Developed and evaluated protein language models for downstream applications and wet lab discovery.
- Evaluated the evolutionary understanding of genomic language model by proposing a dataset of mutated synthetic expression cassettes.

Lab Specialist, School of Medicine, University of Virginia

Jan. 2024 – Jul. 2024

Advisors: Prof. Chongzhi Zang & Prof. Sheng'en Shawn Hu

- Developed and evaluated a ChIP-seq broad peak-calling tool (SICER).

Postgraduate Research Assistant, Dept. of Computer Science, Johns Hopkins

Oct. 2022 – Jul. 2023

Advisors: Prof. Kimia Ghobadi, Prof. Anton Dahbura & Prof. Claus Aranha (Univ. of Tsukuba)

- Developed an agent-based modular simulator for small-community disease spread.

Undergrad Research Assistant, Dept. of Biology, Wenzhou-Kean University

Oct. 2021 – May. 2022

Advisor: Prof. Aloysius Wong

- Developed a bioinformatics webserver for protein motif recognition.

PROFESSIONAL EXPERIENCE

Software Engineer Intern, Alibaba Cloud, Hangzhou, China

Jun. 2022 – Aug. 2022

- PolarDB database development (Java and MySQL).

POSTERS & PRESENTATIONS

- “Small Molecule Approach to RNA Targeting Binder Discovery (SMARTBind) Using Deep Learning Without Structural Input.” **Poster presentation**, UF Health Cancer Institute Annual Research Showcase 2026, University of Florida, Gainesville, FL, USA. Jan. 29, 2026.
- “Tabula: A Tabular Self-Supervised Foundation Model for Single-Cell Transcriptomics.” **Poster presentation**, *NeurIPS* 2025, Hilton Reforma and San Diego Convention Center, San Diego, CA, USA. Dec. 2, 2025.
- “Predicting function of evolutionarily implausible DNA sequences.” **Poster presentation**, Q-BIO 2025 Conference: Emergent Orders in Living Systems Across Scales, Center of Quantitative Biology, Peking University, Beijing, China. Jul. 23, 2025.
- “Predicting function of evolutionarily implausible DNA sequences.” **Poster presentation**, *ICML* 2025 Workshop for Generative AI and Biology, Vancouver, Canada. Jul. 18, 2025.
- “Sequence Display-Enable Machine Learning for Protein Evolution.” **Poster presentation**, Synthetic Biology: Engineering, Evolution, & Design, Houston, TX. Jun. 23, 2025.
- “Simulating Disease Spread during Disaster Scenarios.” **Talk presentation**, The 2023 Conference on Artificial Life, Sapporo, Japan. Jul. 23, 2023.

HONORS & AWARDS

USC Dana and David Dornsife College Graduate School Fellowship

2025

PROFESSIONAL SERVICE

Journal Reviewer

- *IEEE Transactions on Computational Biology and Bioinformatics*
- *PLoS Computational Biology*

Conference Reviewer

- *Association for the Advancement of Artificial Intelligence (AAAI) 2026, 2027*
- *The Annual Conference on Neural Information Processing Systems (NeurIPS) 2026*
- *The International Conference on Learning Representations (ICLR) 2027*