

Shuxian Zou

Abu Dhabi, UAE | +971 (0) 585 906 319 | shuxian.zou@mbzuai.ac.ae | [LinkedIn](#) | [GitHub](#) | [Google Scholar](#)

RESEARCH PROFILE

>> **Machine learning researcher specializing in biomolecular foundation models, structure prediction, gene expression prediction, and multimodal learning for biology.** Experienced in self-supervised learning and large-scale distributed training with PyTorch, Megatron-LM, and Hugging Face, including pretraining on 64 NVIDIA A100 GPUs.

>> **Highlights:**

- Developed a 1.6B-parameter RNA foundation model ([GB.RNA](#)) and contributed to a 16B-parameter protein Mixture-of-Experts model ([GB.Protein](#));
- [Gold Medalist \(6th/1,516 teams\)](#) in the 2025 Stanford RNA 3D Folding Kaggle Competition.

EDUCATION

PhD in Machine Learning, Mohamed bin Zayed University of Artificial Intelligence | Abu Dhabi, UAE

2022 – Dec 2026 (expected) | GPA: 3.95/4.00, Full Sponsorship

Thesis: Representation Learning with Foundation Models for Biomolecular Structure and Function Prediction (*in progress*)

Advisor: Prof. Le Song (MBZUAI & GenBio AI)

Thesis committee: Eric P. Xing (MBZUAI & CMU), Kun Zhang (CMU & MBZUAI), Christoph Feinauer (GenBio AI)

MSc in Computer Science, University of Chinese Academy of Sciences | Beijing, China

2019 - 2022 | GPA: 3.86/4.00, Academic Scholarship

Thesis: Research on Decoding Natural Language from Functional Brain Images | **Advisor:** Prof. Jiajun Zhang

BSc in Statistics, Sun Yat-sen University | Guangzhou, China

2010 - 2014 | GPA: 4.0/5.0, Top 10%

RESEARCH EXPERIENCE

Research Intern | GenBio AI | Dec 2024 - Dec 2025 & June - Dec, 2026 | Abu Dhabi, UAE

- Developed [GB.RNA](#) and [GB.RNA-CDS](#), RNA foundation models for noncoding RNA and coding sequence, and evaluated them on 100+ downstream tasks spanning RNA structure, expression, and function prediction.
- Benchmarked [GB.Protein](#), [GB.Protein-RAG](#), and [GB.Protein2StructureToken](#) on ProteinGym DMS and xTrimopGLM benchmarks for protein understanding and fitness prediction.
- Curated a [transcript isoform expression dataset](#) from GTEx and developed multimodal models for tissue-specific RNA isoform expression prediction integrating DNA, RNA and protein data.
- Implemented sequence-level tasks and multimodal fusion in [GB.ModelGenerator](#), a framework for adapting pretrained biological models to downstream scientific tasks.
- Collaborated with the RNA design team and an external CRO to experimentally validate designed RNA aptamers through wet-lab assays.

Visiting PhD Student | Sailing Lab, Carnegie Mellon University | Jun-Sep 2024 | Pittsburgh, USA

Host: Prof. Eric P. Xing

- Investigated pretraining data and strategies to scale RNA foundation models to billions of parameters and enable unified modeling of non-coding and protein-coding RNA sequences.
- Explored efficient pretraining and downstream evaluation strategies for a 16B-parameter protein language model.

Algorithm Intern | BioMap | Jul – Sep 2023 | Beijing, China

- Pretrained cross-modal protein models on large-scale STRING protein–protein interaction data using Megatron-LM for antibody–antigen binding affinity prediction.

Research Assistant | Institute of Automation, Chinese Academy of Sciences | Jul 2019–May 2022 | Beijing, China

- Proposed new tasks and methods for fMRI-based brain-to-word and brain-to-text decoding using pretrained encoder–decoder language models.

SELECTED PUBLICATIONS

A large-scale foundation model for RNA function and structure prediction.

Shuxian Zou, Tianhua Tao, Sazan Mahbub, Caleb Ellington, Robin Algayres, Dian Li, Yonghao Zhuang, Hongyi Wang, Le Song, Eric P. Xing

Under review at *Nature Methods*. NeurIPS 2024 AIDrugX Workshop (**Spotlight**).

Mixture of experts enable efficient and effective protein understanding and design

Ning Sun, **Shuxian Zou**, Tianhua Tao, Sazan Mahbub, Dian Li, Yonghao Zhuang, Hongyi Wang, Xingyi Cheng, Le Song, Eric P. Xing

Under review at *Nature Methods*. NeurIPS 2024 AIDrugX Workshop (**Spotlight**).

Improving RNA 3D Structure Prediction via Language Model-Augmented AlphaFold 3

Shuxian Zou, Jiayou Zhang, Bingkang Zhao, Hui Li, Eric P. Xing, Le Song

NeurIPS 2025 MLSB Workshop.

Template-based RNA structure prediction advanced through a blind code competition

Youhan Lee, Shujun He, Toshiyuki Oda, ..., **Shuxian Zou**, ..., David Baker, ..., Rhiju Das

Accepted in principle at *Nature Methods*.

Rapid and Reproducible Multimodal Biological Foundation Model Development with AIDO.ModelGenerator

Caleb Ellington, Dian Li, **Shuxian Zou**, Elijah Cole, Ning Sun, Sohan Addagudi, Le Song, Eric P. Xing

ICML 2025 GenBio Workshop (**Spotlight**).

Linker-Tuning: Optimizing Continuous Prompts for Heterodimeric Protein Prediction

Shuxian Zou, Hui Li, Shentong Mo, Xingyi Cheng, Eric P. Xing, Le Song

arXiv preprint, 2023.

Cross-Modal Cloze Task: A New Task to Brain-to-Word Decoding

Shuxian Zou, Shaonan Wang, Jiajun Zhang, Chengqing Zong

Findings of ACL 2022.

TEACHING & ADDITIONAL EXPERIENCE

Teaching Assistant, MBZUAI | Advanced Machine Learning (Spring 2025); Machine Learning with Python (Fall 2025).

Data Analyst, China Mobile | 2014 – 2018 | Guangzhou, China | Customer analysis, and KPI project management.

SELECTED AWARDS

Gold Medal, Stanford RNA 3D Folding Kaggle Competition, 2025 - Ranked 6th / 1,516 teams

Outstanding Teaching Assistant Award, MBZUAI, Spring 2025

SKILLS

Languages: English (professional proficiency), Chinese (native), Cantonese (fluent), Hakka (native)

Programming: Python, Shell, SQL

Machine Learning: PyTorch, Megatron-LM, Hugging Face, Lightning, TensorFlow, scikit-learn

HPC & Infrastructure: SLURM, Weights & Biases (W&B), distributed multi-GPU training

Developer Tools: Git, Claude Code, Codex