Jiawei Zhang

Phone: (+86)18920816307 | Email: jiawei_zhang1@163.com

EDUCATION

CIVIL AVIATION UNIVERSITY OF CHINA *Tianjin, China*

Faculty of Economics and Management, Bachelor of Business Administration *Sep. 2016 – Jul. 2020*

GPA: 3.67/4.0 (85.4/100)

Key courses: Advanced Mathematics/92 (89+95); Probability-and-Statistics/80; Business Statistics and Forecasting/89

PUBLICATION

Jiawei Zhang, Wang Ma, Hui Yao. "Accurate TCR-pMHC interaction prediction using a BERT-based transfer learning method", Briefings in Bioinformatics, 2024

Wang Ma, **Jiawei Zhang**, Hui Yao. "NeoMUST: an accurate and efficient multi-task learning model for neoantigen presentation", Life Science Alliance, 2024

Kai Zhuang†, **Jiawei Zhang†**, Yumou Liu, Hanqun Cao, et al. "Lost in Tokenization: Context as the Key to Unlocking Biomolecular Understanding in Scientific LLMs", arXiv, 2025
In submission to International Conference on Learning Representations (ICLR 2026 Original Score: 8,6,6,6).

Yuyang Zhou†, Jin Su†, **Jiawei Zhang**, Wangyang Hu, Tianli Tao. "PRIME: A Multi-Agent Environment for Orchestrating Dynamic Computational Workflows in Protein Engineering", bioRxiv, 2025.

Jin Su, Zhikai Li, ..., **The OPMC**, ..., Fajie Yuan. "Democratizing protein language model training, sharing and collaboration", Nature Biotechnology, 2025.

RESEARCH EXPERIENCES

MULTI-TOOL INTERGRATED ENZYME MINING PIPELINE

Westlake University, Hangzhou China

Research Assistant Jun. *2025 – Present*

- Project content: Enzyme mining workflow based on vector search, structural features, and other characteristics
- Main responsibilities: Tool Setup, Construction and processing of datasets, Construction of model structures

LLM-BASED FUNCTIONAL ANNOTATION OF PROTEINS

Westlake University, Hangzhou China

Research Assistant Jan. *2025 – Present*

- Project content: Based on motifs, GO terms, and other features, mapping from protein sequence and structure to function through test-time scaling.
- Main responsibilities: Idea, Construction and processing of datasets, Construction of model structures
- Open-source address: https://github.com/opendatalab-raiser/CoKE

ENERGY-BASED TCR-PMHC BINDING PREDICTION

Northeastern University, Remote

Research Assistant (Remote) Oct. *2024 – May. 2025*

- Project content: Using multimer structure prediction models (Boltz, Chai, TCRmodel2, etc.) to predict TCR-pMHC binding based on energy functions derived from its output metrics (pAE, pTMe, etc.).
- Main responsibilities: Construction and processing of datasets, Conducting benchmark on different models, Performing statistical analysis of experimental datasets

DEVELOP AN LLM-BASED BIOLOGICAL AGENT

Westlake University, Hangzhou China

Research Assistant *May. 2024 – June. 2025*

- Project content: Design and develop an LLM-based agent system that automatically executes corresponding AI models and bioinformatic tools based on the user's needs.
- Main responsibilities: Construct the agent framework; Deploy some of the deep learning models

DEVELOP COLABSAPROT AND SAPROTHUB

Westlake University, Hangzhou China

Research Assistant *Apr. 2024 – Aug. 2024*

- Project content: Develop ColabSaprot and SaprotHub to support scientific research, allowing biologists to easily train and use Protein Language Models. SaprotHub is widely used for protein-related tasks, with wet lab experiments validating its results.
- Main responsibilities: Test and optimize the function of ColabSaprot; Assist in building ColabSaprot and SaprotHub
- Completed a paper as the OPMC author: “SaprotHub: Making Protein Modeling Accessible to All Biologists”. (Published in ***Nature Biotechnology***)
- Open-source address: https://github.com/westlake-repl/SaprotHub

RESEARCH AN LLM-BASED WEB3 AGENT

The Hong Kong University of Science and Technology, China

Research Assistant (Remote) *Apr. 2024 – Sep. 2024*

- Project content: Design an agent that can decompose user requirements and then select and execute the appropriate tool for each step of the task.
- Main responsibilities: Design the agent framework; Investigate appropriate fine-tuning algorithms

WORK EXPERIENCES

PREDICT TCR-PMHC BINDING USING DEEP LEARNING

Fresh Wind Biotechnologies Inc. Tianjin, China

Assistant Bioinformatics R&D Engineer *Apr. 2022 – Apr. 2024*

- Project content: Predict the binding of TCR peptide-MHC-1 complexes by BERT-based deep learning model with transfer learning.
- Main responsibilities: Data collection and organization; Modification of model structure; Code writing and GitHub repository creation; Model training and testing; Writing of the paper
- Results: Completed paper as the first author: “Accurate TCR-pMHC Interaction Prediction Using a BERT-based Transfer Learning Method”. (Published in ***Briefings in Bioinformatics***)
- Open-source address: https://github.com/Freshwind-Bioinformatics/TABR-BERT

IDENTIFY NEOANTIGENS USING DEEP LEARNING

Fresh Wind Biotechnologies Inc. Tianjin, China

Assistant Bioinformatics R&D Engineer *Apr. 2022 – Apr. 2024*

- Project content: Accurate identification of neoantigens by multi-task learning architecture with LSTM as the feature extractor.
- Main responsibilities: Data collection and organization; Assist in GitHub repository creation; Assist in model testing; Assist in writing of the paper
- Results: Completed paper as the second author: “Neo-MUST: an Accurate and Efficient Multi-Task Learning Model for Neoantigen Presentation”. (Published in ***Life Science Alliance***)
- Open-source address: https://github.com/Freshwind-Bioinformatics/NeoMUST

AWARD & ACTIVITY

2017-2018 Renmin Third-Class Scholarship
National Innovation and Entrepreneurship Project / Team member
Director of the Communication Club of the Student Union of the University
Undergraduate class monitor
School Library, Kindergarten Outstanding Volunteer
Summer Social Practice Advanced Individual

SKILLS AND OTHERS

Skills: Python, Linux, Docker, R

Interest: Photography📷, Reading🌊