ZHIYU WANG

zhiyu wang work@outlook.com
+44 (0) 7763688280
0009-0001-8938-3663

🏶 wzy-zhiyu-wang.github.io 🞓 gRGHW9wAAAAJ 🛅 zhiyu-wang-243335252

Geometric Deep Learning

Large Language Model

AI4Biology/Healthcare



11 EDUCATION

University of Cambridge

Advanced Computer Science, Master of Philosophy (Distinction, expected)

✓ Cambridge, UK **d** Oct 2024 — (Jun 2025)

- Courses: Geometric Deep Learning, Affective AI, Mobile & Wearable ML, NLP, Multi-agent RL

University College London

✓ London, UK

Computer Science, Bachelor of Science (First Class Honors)

苗 Sep 2021 — Jun 2024

- Main Courses: Machine Learning, Computer Vision, Reinforcement Learning, Mathematics & Statistics, Software Engineering

Research Experience

Protein Representation Learning With TNNs | UNIVERSITY OF CAMBRIDGE

Cambridge, UK 苗 Oct 2024 — Present

Research Assistant | Supervisor: Prof. Pietro Liò

- Developed TCPNet, an SE(3)-equivariant topological neural network that simultaneously learns multiple levels of protein structures (amino acids, secondary structures, proteins), achieving a 5% improvement in antibody developability and protein folding classification tasks.

- Built a complete and extensible protein structure learning framework, including downloading raw PDB data, preprocessing into combinatorial complexes data structure, and computing features for various structural levels.
- Refactored TopoX's neighborhood matrix calculation to pure PyTorch, preventing GPU → CPU → GPU overhead and improving GPU utlization from 15% to 50% and shorten training time from 380 min to 90 min.
- Performed remote development on a Linux HPC server (with 4 A100 GPUs) using SLURM for job queuing and WandB for monitoring.

Multi-Omics Integration Data with GNNs | UNIVERSITY OF CAMBRIDGE

Cambridge, UK

Research Assistant | Supervisor: Prof. Pietro Liò

Feb 2025 — Present

- Using heterogeneous graph network for embedding single cell multi-omics (e.g. RNA, protein, and ATAC modalities) data to a shared latent space and conduct cell clustering.
- Training time on BM-CITE and LUNG-CITE datasets takes only a few minutes (RTX 4070 GPU), with 3% lower ARI and 2% lower NMI compared to the statistical method MOJITOO.
- Developed a pipeline with Python and R for downloading scRNA/ADT/ATAC-seq datasets stored in SeuratObject format and converting them to Python-readable AnnData.

Graph-based Facial Expresission Analysis | University of Cambridge

✓ Cambridge, UK

Research Assistant | Supervisor: Prof. Hatice Gunes

- Accepted for presentation at IJCAI25 MiGA Workshop. (https://arxiv.org/abs/2505.19802v2)
- Proposed the GraphAU-Pain model that fuses a ResNet-50 backbone with graph-based facial action unit (AU) representation learning for pain intensity estimation, allowing better clinical interpretability and generalization compared to purely image-driven methods.
- Presented a novel transfer learning strategy to address domain differences and missing annotations, improving GraphAU-Pain's AU occurrence prediction F1-score by 30% on the UNBC dataset.

Healthcare LLM Instruction Dataset Generation | XUNFEI HEALTHCARE TECHNOLOGY

◀ Hefei, China

Machine Learning Engineer | Supervisor: Jianyu Cai

- **i** Jul 2024 Sep 2024
- Designed a pipeline using 10 human-written examples to generate 1000+ medical report instruction, and output examples with an instruction-tuned LLM; the generated examples improve instruction-following ability of an untuned LLM from 34% to 82%.
- Enhanced instruction generation methods, extending short general instructions (10-50 words) to domain-specific ones averaging 500 words for advanced healthcare-specific LLM training.

RPM Surgery Complication Prediction Model | University College London

Research Assistant | Supervisor: Prof. Ivana Drobnjak

✓ London, UKi May 2023 — Jun 2024

- Built a PostgreSQL pipeline and XGBoost model (ROC-AUC 0.79±0.03) on RPM data, enhancing clinical interpretability with SHAP.
- Developed a patient-independent stratified bootstrapping framework to mitigate data leakage and address small, imbalanced dataset challenges, ensuring credible evaluation and robust model generalizability.

♣ Industrial Experience

Low-cost AI Chatbot Generation System for Hospitals | National Health Service Intern Software Engineer

◀ London, UK

Example 2022 — May 2023

- Delivered a live demonstration at Great Ormond Street Hospital and presented at UCL to IBM, Microsoft, and Intel, earning media recognition from IBM Master Inventor John McNamara. (bit.ly/ixn-nhs-ucl-and-ibm).
- Led a team of 3 to create an automated chatbot generation service for hospitals using React.js and Django, reducing development time from weeks to 30 minutes.
- Deployed a Scrapy web crawler to extract key hospital information and formatted it into Q&A pairs.
- Engineered a multi-layer Q&A architecture that reduced service costs by 50% and improved answer accuracy by 20%.

LLM Knowledge Injection Pipeline | Luojin Data Information

Shanghai, China

Machine Learning Engineer

May 2023 — Aug 2023

- Designed an LLM-based pipeline to translate natural language queries into database queries, achieving 92% accuracy in selecting data for knowledge injection.
- Created a LangChain Agent to label financial reports with 10 category tags, reducing a 2-minute task to 15 seconds per article.
- Implemented semantic search via a MongoDB vector database for investment category tag embedding, ensuring a 97% match rate for user-entered company names.

Fresher Coding Sessions | University College London

◀ London, UK

Programming Tutor

Sep 2022 — May 2023

- Organized biweekly group discussions for 12 first-year computer science students, offering academic guidance and peer support.