Shiyu Zhao

☎ +1 (650)422-8429 ◇ ✉ shiyuz@stanford.edu◇ https://shiyu-zhao.netlify.app/

EDUCATIONAL BACKGROUND

Stanford University, School of Engineering Stanford, United States
Computer Science (Master's degree) Sep 2023 - Mar 2025(expected)

Tsinghua University, Yao Class Beijing, China
Computer Science and Technology (Bachelor's degree) Sep 2019 - Jun 2023

- **GPA:** 3.84/4.0 **Coursework:** Alogorithm design, Machine learning, Artificial intelligence, Mathematics for computer science, Operating system, Distributed system, Data structure, Numerical analysis, Parallel computing
- **Honors and Awards:** Rank 1/70,000+ students in Chinese College Entrance Examination in Ningxia (2019), Dean's List (2020, 2021, 2022), Outstanding graduate of Yao Class (Highest honor in the department, 2023)

PUBLICATIONS

Xiao Liu*, **Shiyu Zhao***, Kai Su*, Yukuo Cen, Jiezhong Qiu, Mengdi Zhang, Wei Wu, Yuxiao Dong, Jie Tang et al. "Mask and Reason: Pre-Training Knowledge Graph Transformers for Complex Logical Queries" In: *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining.* KDD '22. **(∗indicates equal contribution)**

EXPERIENCES

Machine Learning Engineer Intern Beijing, China
miHoYo (HoYoverse) Jul 2023 - Aug 2023

- Collaborated on the development of **foundation model** that works both in Chinese and English in a five-people agile team from scratch with PyTorch under gpt-neox framework.
- Implemented the **supervised fine-tuning** (SFT) of the foundation model, including the online streaming data collection/processing, prompt designing/engineering, making the model's reply longer($> 20\%$), more fluent, and more cohesive to the streamer's characteristic descriptions.
- Developed **personalized** virtual streamers, integrating LLM to support virtual streamers with different characteristic, signature phrases and perferences. Enabled LLM-supported virtual streamers to interact with audiences with words, movements, and expressions.
- Created a fully automated **evaluation toolkit** for the performance of the model, including foundation model benchmarks and llm-judge for downstream tasks, added into the company's toolkits to expedite the future development.

Machine Learning Engineer Intern Montreal, Canada
Montreal Institute for Learning Alogorithms (MILA) Mar 2022 - Aug 2022

- Built the first systematically generalizable and scalable **reasoning model** on knowledge graph with PyTorch, inpired by forward and backward chaining, creating a new logical perspective with regard to natural language inference.
- Proposed to model one-step logic inference in NLP as triangle update on graph inspired by logic programming, and formalized it under **graph neural network** framework.
- Identified reasonable patterns and conduct partial reasoning with the help of auxiliary edges on graphs, improving **performance, efficiency** as well as **expressiveness**.
- Outperformed **SOTA** by 9.25% on FB15K-237 dataset, paper under review.

Research Assistant Beijing, China
Tsinghua University, ZhipuAI Mentor: Prof. Jie Tang, Prof. Yuxiao Dong Jan 2021 - Dec 2021

- Introduced a well-functioning **pretrain-finetune** large model into knowledge graph area with great generalizability.
- Designed a knowledge graph triple transformation method to apply transformer on knowledge graph with ease and a mechanism to **unify** different downstream tasks of knowledge graph problems.
- Achieved **SOTA** on both in-domain and out-of-domain reasoning task, significantly outperformed previous SOTA CQD(ICLR 2021 best paper) by over 12.1% relatively on FB15k-237 and over 6.4% relatively on NELL995.
- Accepted by **SIGKDD 2022** (research track) as the first co-author.

PROJECTS

Random Matrix Factorization of Large-scale Network Embedding: Improve large-scale network embedding by single-view SVD and speeded up the factorization by freigs algorithm. Expanded the network scale and boosted speed.
GLUE+: Distinguishable graph-linked embedding for multi-omics single-cell data integration: Solved the indistinguishability of aggregating multi-omics data on the graph by using multiple GNN aggregators.

SKILLS

Programming Languages: Python, C/C++, SQL, bash, Java, JavaScript, MATLAB, LaTeX, Verilog, Go, VB
Tools: PyTorch, Pandas, Linux, TensorFlow, Git, NumPy, Unix, MySQL, Azure, Django, Jetbrains, Redis