Zixuan Weng

zxweng0701@gmail.com · Google Scholar · Homepage

EDUCATION

Georgia Institute of Technology Aug. 2026 – Present

- *Ph.D. in Computer Science and Engineering*
- *Advisor: Ling Liu*

Beijing Jiaotong University Sept. 2020 – Jun. 2024

- *B.Eng in Computer Science and Technology*

PUBLICATIONS&PREPRINTS

1. **Zixuan Weng**, Jinghuai Zhang, Kunlin Cai, Ying Li, Peiran Wang, Yuan Tian. FineSteer: A Unified Framework for Fine-grained Inference-time Steering in Large Language Models. *ACL 2026 (**Oral**, **Top4%**)*. Link
2. **Zixuan Weng**, Xiaolong Jin, Jinyuan Jia, Xiangyu Zhang. Foot-In-The-Door: A Multi-turn Jailbreak for LLMs. *EMNLP 2025 (**Oral**, **Top4%**)*. Link
3. **Zixuan Weng**, Jindong Han, Wenzhao Jiang, Hao Liu. SDE: A Simplified and Disentangled Dependency Encoding Framework for State Space Models in Time Series Forecasting. *KDD 2025*. Link
4. Xiaolong Jin, **Zixuan Weng**, Hanxi Guo, Siyuan Cheng, Guangyu Shen, Chenlong Yin, Xiangyu Zhang. JailbreakDiffBench: A Comprehensive Benchmark for Jailbreaking Diffusion Models. *ICCV 2025*. Link
5. Xingyi Zhang, **Zixuan Weng**, Sibo Wang. Towards Deeper Understanding of PPR-based Embedding Approaches: A Topological Perspective. *WWW 2024 (**Oral**, **Top9%**)*. Link
6. Jingyu Xiao, Qingsong Zou, Qing Li, Dan Zhao, Kang Li, **Zixuan Weng**, Ruoyu Li, Yong Jiang. I Know Your Intent: Graph-Enhanced Intent-Aware User Device Interaction Prediction via Contrastive Learning. *Ubicomp 2023*. Link
7. Hao Wang*, **Zixuan Weng***, Hao Liu. DAMBench: A Multi-Modal Benchmark for Deep Learning-based Atmospheric Data Assimilation. *In submission*

RESEARCH EXPERIENCE

VLM Red-teaming (Talent Program) Sep. 2025 – May 2026

Talent Program (English: Jindouyun; Chinese: 筋斗云) Intern (ByteDance) Mentor: Yang Bai

- Conducted red-teaming for Doubao (Cizi) —currently China's AI application with the largest traffic and downloads —and Volcano Ark —currently China's AI cloud service platform with the highest usage.
- Built a security assessment platform for the technology middle platform.
- Addressed real problems based on real data and oriented to real needs.

LLM Steering Jun. 2025 – Jan. 2026

Research Assistant, Advisor:Prof.Yuan Tian (UCLA)

- Proposed a method for efficiently and effectively constructing query-specific steering vectors
- Achieved SOTA (State-of-the-Art) performance on both tasks of defending against jailbreaks and reducing hallucinations, with little to no impact on the model's general performance.
- Submitted a paper as the first author to ACL.

LLM Jailbreak Dec. 2024 – May. 2025

Research Assistant, Advisors:Prof.Jinyuan Jia (PSU) and Prof.Xiangyu Zhang (Purdue)

- Motivated by the Foot-In-The-Door phenomenon, a novel multi-turn jailbreak strategy was introduced.
- Achieving a jailbreak success rate of more than 94% in mainstream LLMs.
- Contributed a paper as the first author to EMNLP.

Time Series Forecasting Feb. 2024 – May.2024

Research Assistant, Advisor:Prof.Hao Liu (HKUST)

- Identified and formally define three dependencies in time seires
- Analyzed the strength of Mamba in time series forecasting and the challenge involved in adapting Mamba for this purpose
- Proposed a simplified Mamba by removing nonlinearities and incorporating a theoretically sound disentangled encoding strategy

Analyzing and explaining the PPR embedding of graph Jul. 2023 – Oct. 2023

Research Assistant, Advisor:Prof.Sibo Wang (CUHK)

- Conducted a survey on graph embedding methods based on random walks
- Proposed and implemented an optimization method for analyzing the information encoded by the PPR-embedding method
- Found the proposed approach results in a 10% improvement in performance recovery over the baseline average path length and a 5% improvement in community connectivity recovery

User Behavior Prediction Using Graph Contrastive Learning Oct. 2022 – Jun. 2023

Research Assistant, Advisor:Prof.Yong Jiang (THU)

- Proposed and implemented the Cluster-based Historical Attention Mechanism, which utilizes the k-means method to cluster behavior sequences and then employ an attention mechanism to capture multi-level periodicity in user behavior
- Discovered the proposed model has an average improvement of 7% compared to the SOTA baseline model and better interpretability
- Contributed to a paper published at the Ubicomp/ISWC 2023 conference as the co-author

HONROS&AWARDS

- KDD Outstanding Reviewer **(Top 10%)** 2025
- National College Students' Innovative Entrepreneurial Training Plan Program **(Top 5%)** 2023
- National Scholarship (**Top 1%**) 2023
- First Class Scholarship for Academic Performance **(Top 2%)** 2023
- Second Class Scholarship for Social Work Performance **(Top 5%)** 2023

ACADEMIC SERVICES

- Session Chair: KDD 2025
- Reviewer: AAAI 2027; NeurIPS 2026; CVPR 2026; ARR 2025 May; WWW 2026; KDD 2024,2025,2026,2027
- Volunteer: KDD 2025; EMNLP 2025