Heng Wang

Address: Room 1111, Siebel Center for Computer Science 201 N. Goodwin Avenue, Urbana, IL 61801, USA
Email: hengwang0301@gmail.com, heng6@illinois.edu,
Homepage: arthur-heng.github.io

Research Interests

I am interested in understanding and controlling LMs, reasoning for structured data, and AI for scientific discovery.

Education

University of Illinois Urbana-Champaign, Champaign, IL, USA 2025.08 - present
Ph.D. student in Computer Science

Xi'an Jiaotong University, Xi'an, Shaanxi, China 2021.09 - 2025.06
B.E. in Computer Science and Technology
Thesis Advisor: Prof. Minnan Luo

University of California, Berkeley, Berkeley, California 2023.08 - 2023.12
Visiting Student

Publications (* indicates equal contribution)

Generalizable LLM Learning of Graph Synthetic Data with Reinforcement Learning
Yizhuo Zhang*, Heng Wang*, Shangbin Feng*, Zhaoxuan Tan, Xinyun Liu, Yulia Tsvetkov
arXiv preprint 2025

Continuously Steering LLMs Sensitivity to Contextual Knowledge with Proxy Models
Yilin Wang, Heng Wang, Yuyang Bai, Minnan Luo
In *Proceedings of EMNLP 2025 (Oral)*

Can Language Models Solve Graph Problems in Natural Language?
Heng Wang*, Shangbin Feng*, Tianxing He, Zhaoxuan Tan, Xiaochuang Han, Yulia Tsvetkov.
In *Proceedings of the NeurIPS 2023* **(spotlight)**

SmartBackdoor: Malicious Language Model Agents that Avoid Being Caught
Heng Wang*, Ruiqi Zhong*, Jiaxin Wen, Jacob Steinhardt.
In *ICML NextGenAISafety Workshop 2024*

Can LLM Graph Reasoning Generalize beyond Pattern Memorization?
Yizhuo Zhang*, Heng Wang*, Shangbin Feng*, Zhaoxuan Tan, Xiaochuang Han, Tianxing He, Yulia Tsvetkov.
In *Proceedings of EMNLP 2024, findings*

Detecting Spoilers in Movie Reviews with External Movie Knowledge and User Networks
Heng Wang, Wenqian Zhang, Yuyang Bai, Zhaoxuan Tan, Shangbin Feng, Qinghua Zheng, Minnan Luo.
In *Proceedings of EMNLP 2023*

Explaining Datasets in Words: Statistical Models with Natural Language Parameters
Ruiqi Zhong, Heng Wang, Dan Klein, Jacob Steinhardt.
In *Proceedings of NeurIPS 2024*

Resolving Knowledge Conflicts in Large Language Models
Yike Wang*, Shangbin Feng*, Heng Wang, Weijia Shi, Vidhisha Balachandran, Tianxing He, Yulia Tsvetkov.
In *Proceedings of COLM 2024*

DELL: Generating Reactions and Explanations for LLM-Based Misinformation Detection
Herun Wan*, Shangbin Feng*, Zhaoxuan Tan, Heng Wang, Yulia Tsvetkov, Minnan Luo In *Proceedings of ACL 2024, findings*

Unveiling the Hidden: Movie Genre and User Bias in Spoiler Detection
Haokai Zhang*, Shengtao Zhang*, Zijian Cai, Heng Wang, Ruixuan Zhu, Zinan Zeng, Minnan Luo
In *Proceedings of ECML-PKDD 2025*

From predictions to analyses: Rationale-augmented fake news detection with large vision-language models Xiaofan Zheng, Zinan Zeng, Heng Wang, Yuyang Bai, Yuhan Liu, Minnan Luo In *Proceedings of WWW 2025*

BotMoE: Twitter Bot Detection with Community-Aware Mixtures of Modal-Specific Experts
Yuhan Liu, Zhaoxuan Tan, Heng Wang, Shangbin Feng, Qinghua Zheng, Minnan Luo.
In *Proceedings of SIGIR 2023*

Research Experience

TsvetShop @ University of Washington 2023.02 - 2024.6

Advisor: Prof. Yulia Tsvetkov Mentor: Shangbin Feng

Can Language Models Solve Graph Problems in Natural Language?

- Proposed one of the first graph reasoning benchmarks for language models.
- Conducted extensive experiments to evaluate LLMs and prompting approaches on the benchmark.
- Provided insights into how LLMs can be applied to graphs.
- Provided two prompting methods to improve LLMs' graph reasoning: Build-a-Graph and Algorithmic prompting, which are widely used in later papers on graph reasoning.
- Published a paper as the first author at NeurIPS 2023 (spotlight).

Can LLM Graph Reasoning Generalize beyond Pattern Memorization?

- Designed experiments to quantify the success of "generalization" of LLM graph reasoning.
- Showed that on simpler patterns (e.g. semantic) it is ok-ish, but on harder patterns (e.g. transferring from synthetic training data to real-world tasks), tuning barely helped and could even be counterproductive.
- The results cast doubt on the benefit of fine-tuning with synthetic graph problems.
- Further analysis showed that code-mixing, alignment, and better data mixtures could help generalization to some extent.
- Published a paper as a co-first author at EMNLP 2024, findings.

University of California, Berkeley 2023.07 - 2024.5

Advisor: Prof. Jacob Steinhardt Mentor: Ruiqi Zhong

Explaining Datasets in Words: Statistical Models with Natural Language Parameters

- Explored expanding the Statistical Models with Natural Language to vision data beyond text data.
- Published a paper as the second author at NeurIPS 2024.

SmartBackdoor: Malicious Language Model Agents that Avoid Being Caught

- Speculated a new family of cyber attacks in which malicious actors provide a backdoored LLM agent; when the victim uses the agent, the agent uses information from its environment to detect whether it is overseen by the victim user; if not, the agent acts maliciously against the victim.
- Used AutoGPT as a case study and provide a proof-of-concept: to exfiltrate a private key without being caught, a backdoored LLM agent can analyze the command running itself or infer the skill level of the human user, thus predicting whether it will get caught.

- Published a paper as the first author at ICML NextGenAISafety Workshop 2024 and submitted it to a conference.

Luo lab Undergraduate Division (LUD) @ Xi'an Jiaotong University

Advisor: Prof. Minnan Luo

- **Director:** Mentored a sophomore to work on knowledge conflicts of language models and published a paper at EMNLP 2025. Mentored a junior to work on explainable fake news detection and published a paper at WWW 2025. 2024.06 - 2025.6

Honors and Awards

SenseTime Scholarship (awarded to 30 students in China)	2023
National Scholarship (0.2% nationwide)	2022
Golden Award (team 3rd place), ACM ICPC Shaanxi Provincial Programming Contest	2023
Silver Award (ranking 26/115), ACM ICPC Asia Hong Kong Regional	2023
NeurIPS Scholar Award	2023
Silver Award, CCF Collegiate Computer System and Programming Contest	2022
Academic Research Award, Xi'an Jiaotong University	2022
Dean's List, Xi'an Jiaotong University	2022

Services

Reviewer for ICLR	2025
Reviewer for ACL Rolling Review	2024, 2025
Reviewer for NeurIPS	2024, 2025
Reviewer for COLM	2024, 2025
Reviewer for AAAI	2025
Reviewer for TIST	2024
Reviewer for AGI Workshop @ ICLR	2024
Reviewer for SeT LLM Workshop @ ICLR	2024
Virtual Volunteer for EMNLP	2023
Reviewer for EMNLP	2023
Reviewer for TOIS	2023
Reviewer for NeurIPS, Datasets and Benchmarks Track	2022, 2023, 2024

Skills

- Programming Skills: C/C++, Python, PyTorch, Pascal
- Language Skills: Mandarin (native), English (TOFEL 108: R 30, L 29, S 24, W 25)