

Lang Cao

Tel: (+1) 217-621-0532 | Email: langcao0820@gmail.com
Homepage: zclang-c.github.io | GitHub: github.com/zclang-c

Education

University of Illinois at Urbana-Champaign (UIUC)

Urbana, US

Doctor of Philosophy in Computer Science

Aug. 2025 (Expected Start)

- Research Area: Machine Learning; Machine Reasoning; AI for Health
- Advisor: Prof. Yue Guo

Master of Science in Computer Science

Aug. 2022 - May 2024

- Research Area: AI for Healthcare; Natural Language Processing

Wuhan University of Technology (WUT)

Wuhan, China

Bachelor of Engineering in Software Engineering

Sept. 2018 - June 2022

- Rank 1st/79, GPA 93.51/100 (3.94/4.0), National Scholarship

Selected Publications (Full List on [\[Google Scholar\]](#))

Learn to Refuse: Making Large Language Models More Controllable and Reliable through Knowledge Scope Limitation and Refusal Mechanism [\[Paper\]](#)[\[Code\]](#)

- **Lang Cao.**
- **EMNLP 2024 Main Conference**, 2024 Conference on Empirical Methods in Natural Language Processing.

PILOT: Legal Case Outcome Prediction with Case Law [\[Paper\]](#) [\[Code\]](#)

- **Lang Cao**, Zifeng Wang, Cao Xiao, Jimeng Sun.
- **NAACL 2024 Main Conference**, 2024 Conference of the North American Chapter of the Association for Computational Linguistics.

KG-FIT: Knowledge Graph Fine-Tuning Upon Open-World Knowledge [\[Paper\]](#) [\[Code\]](#)

- Pengcheng Jiang, **Lang Cao**, Cao Xiao, Parminder Bhatia, Jimeng Sun, Jiawei Han
- **NeurIPS 2024**, The Thirty-Eighth Annual Conference on Neural Information Processing Systems.

GraphReason: Enhancing Reasoning Capabilities of Large Language Models through A Graph-Based Verification Approach [\[Paper\]](#)[\[Code\]](#)

- **Lang Cao.**
- ACL 2024, Natural Language Reasoning and Structured Explanations Workshop.

AutoRD: An Automatic and End-to-end Rare Disease Knowledge Graph Construction System Based on Ontologies-enhanced Large Language Models [\[Paper\]](#)[\[Code\]](#)

- **Lang Cao**, Adam Cross, Jimeng Sun.
- JMIR Medical Informatics, 2024.

TablePilot: Recommending Human-Preferred Tabular Data Analysis with Large Language Models [\[Paper\]](#)

- Deyin Yi, Yihao Liu, **Lang Cao**, Mengyu Zhou, Haoyu Dong, Shi Han, Dongmei Zhang
- ACL 2025 Industry Track (Oral).

Fortune: Formula-Driven Reinforcement Learning for Symbolic Table Reasoning in Language Models

- **Lang Cao**, Jingxian Xu, Hanbing Liu, Jinyu Wang, Mengyu Zhou, Haoyu Dong, Shi Han, Dongmei Zhang.
- Under Review (NeurIPS 2025).

Bingo: Boosting Efficient Reasoning of LLMs via Dynamic and Significance-based Reinforcement Learning

- Hanbing Liu, **Lang Cao**, Yuanyi Ren, Mengyu Zhou, Haoyu Dong, Xiaojun Ma, Shi Han, Dongmei Zhang.
- Under Review (NeurIPS 2025).

DeepRetrieval: Hacking Real Search Engines and Retrievers with Large Language Models via Reinforcement Learning [\[Paper\]](#) [\[Code\]](#)

- Pengcheng Jiang, Jiacheng Lin, **Lang Cao**, Runchu Tian, SeongKu Kang, Zifeng Wang, Jimeng Sun, Jiawei Han.
- Under Review (COLM 2025).

SuperRL: Reinforcement Learning with Supervision to Boost Language Model Reasoning

- Yihao Liu, Shuo Cheng Li, **Lang Cao**, Yuhang Xie, Mengyu Zhou, Haoyu Dong, Xiaojun Ma, Shi Han, Dongmei Zhang.
- Under Review (NeurIPS 2025).

TableMaster: A Recipe to Advance Table Understanding with Language Models [\[Paper\]](#)[\[Code\]](#)

- **Lang Cao**, Hanbing Liu.
- Under Review (NeurIPS 2025).

RAS: Enhanced Knowledge-Intensive LLM Generation with Iterative Retrieval-And-Structure [\[Paper\]](#)

- Pengcheng Jiang, **Lang Cao**, Ruike Zhu, Minhao Jiang, Yunyi Zhang, Jimeng Sun, and Jiawei Han.
- Under Review (NeurIPS 2025).

Accelerating Clinical Evidence Synthesis with Large Language Models [\[Paper\]](#)

- Zifeng Wang, **Lang Cao**, Benjamin Danek, Qiao Jin, Zhiyong Lu, Jimeng Sun.
- npj Digital Medicine, 2025.

A foundation model for human-AI collaboration in medical literature mining [\[Paper\]](#)

- Zifeng Wang, Lang Cao, Qiao Jin, et al. (23 authors total), Jimeng Sun.
- Nature Communications, 2025.

Experiences

Ubiquant (Top Quant Fund in China)

AI and Quant Intern

- Research Focus: AI for Quantitative Trading.
- Mentor: Ziwei Yang

Shanghai, China

May 2025 – Sept. 2025

Microsoft Research

Research Intern at Data Knowledge and Intelligence Group

- Research Focus: Spreadsheet Intelligence, Table LLMs, General Machine Reasoning.
- Mentor: Haoyu Dong, Mengyu Zhou

Beijing, China

Aug. 2024 – June 2025

Tsinghua University

Research Assistant at THU-NLP Lab

- Research Focus: Multi-modal Learning and Reasoning.

Beijing, China

Nov. 2024 – Feb. 2025

University of Illinois at Urbana-Champaign

Research Assistant at Sunlab

- Research Focus: NLP / LLMs Applications for Healthcare and Legal.

Urbana, US

Jan. 2023 – May 2024

iFLYTEK

AI Algorithm Intern at Smart Car Technology R&D Division

- Focus: AI Applications on Smart Cars.
- Mentor: Shen'an Li

Hefei, China

June 2021 - Aug. 2021

Selected Honors & Awards

- Silver Medal, top 5% in Kaggle Common Lit Readability Prize (2021.8)
- Top 2% in Alibaba Tianchi NLP Chinese Pre-training Model Generalization Ability Challenge (2021.1)
- National Scholarship (1%), WUT (2020); Merit Student Model Honor (5%), WUT (2020)
- Outstanding Student Leader of the School (2%), WUT (2019)
- Outstanding Graduate (2%), WUT (2022.6); Outstanding Thesis (1%), WUT (2022.6)
- **The National Champion** of the 2014 FIRST LEGO League Challenge in China (core) (2014.6); **Gold Award** at the 2016 Asia-Pacific Championship of the FIRST LEGO League Challenge (leader) (2%) (2016.7)

Invited Talks

- “Thoughts of DeepSeek R1 & Reasoning in Language Models”, invited by Haoyu Dong at Microsoft Research, (2025.02).
- “Language Models’ Table Understanding with Formula”, invited by Mengyu Zhou at Microsoft Research, (2025.04).

Services

- Paper Review: ACL Rolling Review (ACL, EMNLP, NAACL), ICLR, ICML, NeurIPS, KDD