Moxin Li

Homepage: https://li-moxin.github.io/HelloFromMoxin/   Mobile: +65 80381637 / +86 13683398638
Github: https://github.com/li-moxin   Email: limoxin@u.nus.edu

Education

- National University of Singapore   Aug. 2021 – Now
  Ph.D. Candidate in Computer Science, supervised by Professor Tat-Seng Chua.

- Peking University   Sep. 2016 – Jun. 2020
  Bachelor in Intelligence Science and Technology, Yuanpei College.

Experience

- National University of Singapore   Sep. 2020 – Jul. 2021
  Research Intern, supervised by Professor Tat-Seng Chua and Professor Fuli Feng.

- Wangxuan Institute of Computer Technology, Peking University   Jun. 2019 – Jun. 2020
  Research Intern, supervised by Professor Rui Yan.

Research Interest

My research interest lies in trustworthy language models. I have focused on topics including:
- Robustness: robust question answering by counterfactual reasoning, robust prompt optimization
- Confidence calibration for mitigating hallucination
- Multi-objective alignment towards thorough and unbiased trustworthiness
I am also interested in LLM evaluation, LLM knowledge boundary, prompt optimization, and table-and-text QA.

Selected Publications & Preprints

16 papers published in top-tier conferences.
6 papers as (co)first author.
Google Scholar citation -> 207. h-index -> 9.

- <u>Moxin Li</u>*, Yuantao Zhang*, Wenjie Wang, Wentao Shi, Zhuo Liu, Fuli Feng, Tat-Seng Chua. Self-improvement towards pareto optimality: Mitigating preference conflicts in multi-objective alignment. ACL (findings) 2025.

- <u>Moxin Li</u>*, Yong Zhao*, Yang Deng, Wenxuan Zhang, Shuaiyi Li, Wenya Xie, See-Kiong Ng, Tat-Seng Chua. Knowledge Boundary of Large Language Models: A Survey. ACL 2025 (oral presentation).

- <u>Moxin Li</u>, Wenjie Wang, Fuli Feng, Fengbin Zhu, Qifan Wang, Tat-Seng Chua. Think Twice Before Trusting: Self-Detection for Large Language Models through Comprehensive Answer Reflection. EMNLP (findings) 2024.

- <u>Moxin Li</u>, Wenjie Wang, Fuli Feng, Yixin Cao, Jizhi Zhang, Tat-Seng Chua. Robust prompt optimization for large language models against distribution shifts. EMNLP 2023 (oral presentation).

- <u>Moxin Li</u>, Wenjie Wang, Fuli Feng, Hanwang Zhang, Qifan Wang, Tat-Seng Chua. Hypothetical training for robust machine reading comprehension of tabular context. ACL (findings) 2023.

- <u>Moxin Li</u>, Fuli Feng, Hanwang Zhang, Xiangnan He, Fengbin Zhu, Tat-Seng Chua. Learning to imagine: Integrating counterfactual thinking in neural discrete reasoning. ACL 2022.

  (* denotes Equal Contribution or Core Contributor.)

Professional Service and Presentation

- Tutorial Organizer: "Unveiling Knowledge Boundary of Large Language Models for Trustworthy Information Access" SIGIR 2025

- Conference and Journal Reviewer: ACL ARR, EMNLP ARR, ICLR, AAAI, KDD, CL

- Conference Volunteer: ICAIF 2025, EMNLP 2024, WWW 2024, WSDM 2023

- Session Chair and Presenter: The 2022, 2023, 2024 Singapore Symposium on Natural Language Processing

- Presenter: Singapore ACM SIGKDD Symposium 2024

- Test Designer and Organizer: CCIR Cup 2022

Honors and Awards

- NUS Research Achievement Award   National University of Singapore, 2023