Maojia Song

Homepage: maojiasong.github.io
Email: maojia_song@mymail.sutd.edu.sg

Education

Singapore University of Technology and Design 01/2024 – Present
Ph.D., Information Systems Technology and Design **Supervisor: Soujanya Poria**

University of Leeds 09/2019 – 08/2023
B.Eng., Electronic and Electrical Engineering **Rank: 1st in the department**
First-Class Honours

Work Experience

Research Assistant 09/2023 – 12/2023
DeCLaRe Lab, Singapore University of Technology and Design
Supervisor: Soujanya Poria

Undergraduate Intern 03/2022 – 09/2022
Business AI Lab, Nanyang Technological University
Supervisor: Teoh Teik Toe

Undergraduate Intern 10/2021 – 04/2022
Computer Laboratory, University of Cambridge
Supervisor: Pietro Liò

Honors & Awards

OpenAI Researcher Access Program Grant 2024
Outstanding Student in InternLM Camp, Shanghai AI Laboratory 2024
Doctoral Research Scholarship granted by MOE, Singapore 2024
Best Undergraduate Thesis Finalist by EEE, University of Leeds 2023
Winner of U.S.-China Young Maker Competition 2021

Publications (* *denotes equal contributions*)

[2] Yew Ken Chia, Liying Cheng, **Maojia Song**, Hou Pong Chan, Chaoqun Liu, Mahani Aljunied, Soujanya Poria, Lidong Bing. "M-Longdoc: A Benchmark for Multimodal Super-Long Document Understanding and a Retrieval-Aware Tuning Framework." *arXiv preprint.* 2024.

Submitted to ICLR 2025. Reviewer rating: 8,5,3

[1] **Maojia Song**, Shang Hong Sim, Rishabh Bhardwaj, Hai Leong Chieu, Navonil Majumder, Soujanya Poria. "Measuring and Enhancing Trustworthiness of LLMs in RAG through Grounded Attributions and Learning to Refuse." *arXiv preprint.* 2024.

Submitted to ICLR 2025. Reviewer rating: 8,8,8,8

Projects

Enhancing scalable MCQA by Synthesizing Negative Answers Using LLMs 2024

- Investigated the feasibility of massively synthesizing high-quality negative answers in the MCQA task, reducing human labour while preserving expert-labelled positive answers.
- Developed a chain-of-thought synthesis process where GPT-4 generated reasoning steps towards positive answers, and a secondary GPT-4 model modified the correct reasoning inconsistencies to improve the difficulty of synthesized negative annotations.

- Achieved performance improvements on ANLI, CommonsenseQA, and QASC datasets, with minimal performance drops on complex datasets like HellaSWAG and SIQA (within 2%), showcasing the scalability and cost-efficiency of the approach.

Counterfactual Evaluation of LLMs: Disentangling Reasoning from Knowledge 2023

- Designed an evaluation framework targeting LLM reasoning and knowledge abilities using counterfactual facts to isolate memory-based responses from genuine understanding.
- Developed and synthesized counterfactual test datasets using GPT-4 to assess LLM responses to logical and conceptual shifts.
- Evaluated GPT-4 and Llama2 models, revealing GPT-4's strengths in reasoning but weaknesses in knowledge comprehension, while Llama2 showed weaker overall performance but comparatively better knowledge understanding.

Invited Talks

SSNLP 2024 11/2024
Singapore Symposium on Natural Language Processing

Teaching

Teaching assistant

- 50.045 Information Retrieval Fall 2024
  Instructor: Soujanya Poria
- 50.006 User Interface Design and Implementation Fall 2024
  Instructor: Simon Perrault

Professional Services

Reviewer

- ICLR 2025