

Xiaofei Wen

Davis, CA | xfwe@ucdavis.edu | (279)766-9275 | [Homepage](#) | [Linkedin](#) | [Github](#)

Bio

Xiaofei Wen is a Computer Science Ph.D. student at University of California, Davis, advised by Prof. Muhao Chen. His research interest lies in Natural Language Processing (NLP). Currently, he is working on improving reasoning efficiency and enhancing reasoning generalization.

Publication

- Diagnosing and Mitigating Modality Interference in Multimodal Large Language Models** May 2025
Rui Cai, Bangzheng Li, [Xiaofei Wen](#), Muhao Chen, Zhe Zhao
[Preprint](#) [paper] [code]
- ThinkGuard: Deliberative Slow Thinking Leads to Cautious Guardrails** Feb 2025
[Xiaofei Wen](#), Wenxuan Zhou, Wenjie Jacky Mo, Muhao Chen
[ACL 2025 Findings](#) [paper] [code]
- Red Teaming Language Models for Processing Contradictory Dialogues** Nov 2024
[Xiaofei Wen](#), Bangzheng Li, Tenghao Huang, Muhao Chen
[EMNLP 2024](#) [paper] [code]
- Personalized Topic Selection Model for Topic-Grounded Dialogue** Aug 2024
Shixuan Fan, Wei Wei, [Xiaofei Wen](#), Xianling Mao, Jixiong Chen, Danyang Chen
[ACL 2024 Findings](#) [paper]
- Sequential Topic Selection Model with Latent Variable for Topic-Grounded Dialogue** Dec 2022
[Xiaofei Wen](#), Wei Wei, Xianling Mao
[EMNLP 2022 Findings](#) [paper]

Education

- University of California, Davis**, Ph.D. in Computer Science Sept 2024 - Present
 - Advised by: Prof. Muhao Chen
- Huazhong University of Science and Technology**, M.S. in Computer Science Sept 2021 – June 2024
 - **Research:** Natural Language Processing, Conversational AI
- Dalian University of Technology**, B.Eng in Software Engineering Sept 2017 – June 2021
 - **GPA:** 3.89/4.00
 - **Awards:** Mitsubishi Fellowship, Outstanding Scholarship, Meritorious Winner of 2019 ICM

Experience

- Summer Intern**, NetEase Games – Hangzhou, China June 2023 – Sept 2023
 - Collected and standardised the data used to address the problem of dialogue hallucinations, and validated the validity through fine-tuning methods.
 - Participated in the construction of Chinese persona dialogue model based on LLaMA and LLaMA2. The finetuned model was used for character dialogue fragments generation for the game Treacherous Waters.
- Research Intern**, Ant Group – Hangzhou, China Sept 2022 – Dec 2022
 - Implemented a RL-based conversational(interactive) recommendation system [code].
 - Improved the performance of the interactive recommendation system for the real insurance business in Alipay.

Service

Reviewer: ACL, EMNLP, NAACL, since 2023.

Vice President, Model United Nations of DUT, 2018-2019