

Xiaofei Wen

Davis, CA | xfw@ucdavis.edu | (+1) 279-766-9275 | [Homepage](#) | [Linkedin](#) | [Github](#)

Bio

Xiaofei Wen is a CS Ph.D. student at University of California, Davis, advised by Prof. Muhao Chen.

His research lies at the intersection of Video Audio Alignment. He focuses on leveraging RL & post-training to enhance LLM reasoning and video understanding capabilities.

Publication

- [a] **When Vision Speaks for Sound** May 2026
[Xiaofei Wen](#), Wenjie Jacky Mo, Xingyu Fu, Rui Cai, Tinghui Zhu, Wendi Li, Yanan Xie, Muhao Chen, Peng Qi
[Preprint](#) [paper] [webpage]
- [b] **Video Models Can Reason with Verifiable Rewards** May 2026
Tinghui Zhu, Sheng Zhang, James Y. Huang, Selena Song, [Xiaofei Wen](#), Yuankai Li, Hoifung Poon, Muhao Chen
[Preprint](#) [paper] [webpage]
- [c] **ModelLens: Finding the Best for Your Task from Myriads of Models** May 2026
Rui Cai, Wenjie Jacky Mo, [Xiaofei Wen](#), Qiyao Ma, Wenhui Zhu, Xiwen Chen, Muhao Chen, Zhe Zhao
[Preprint](#) [paper] [webpage]
- [d] **DebugLM: Learning Traceable Training Data Provenance for LLMs** Mar 2026
Wenjie Jacky Mo, Qin Liu, [Xiaofei Wen](#), Wenxuan Zhou, Zhe Zhao, Muhao Chen
[Preprint](#) [paper]
- [e] **OmniGuard: Unified Omni-Modal Guardrails with Deliberate Reasoning** Dec 2025
Boyu Zhu, [Xiaofei Wen](#), Wenjie Jacky Mo, Tinghui Zhu, Yanan Xie, Peng Qi, Muhao Chen
[Preprint](#) [paper] [webpage]
- [f] **Learning Efficient Guardrails for Compliance** Oct 2025
[Xiaofei Wen](#), Wenjie Jacky Mo, Yanan Xie, Peng Qi, Muhao Chen
ICML 2026 [paper] [webpage]
- [g] **REDCODER: Automated Multi-Turn Red Teaming for Code LLMs** July 2025
Wenjie Jacky Mo, Qin Liu, [Xiaofei Wen](#), Dongwon Jung, Hadi Askari, Wenxuan Zhou, Zhe Zhao, Muhao Chen
ACL 2026 [paper] [code]
- [h] **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]
- [i] **ThinkGuard: Deliberative Slow Thinking Leads to Cautious Guardrails** Feb 2025
[Xiaofei Wen](#), Wenxuan Zhou, Wenjie Jacky Mo, Muhao Chen

ACL 2025 Findings [paper] [code]

[j] **Red Teaming Language Models for Processing Contradictory Dialogues** Nov 2024
Xiaofei Wen, Bangzheng Li, Tenghao Huang, Muhao Chen
EMNLP 2024 [paper] [code]

[k] **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]

[l] **Sequential Topic Selection Model with Latent Variable for Topic-Grounded Dialogue** Dec 2022
Xiaofei Wen, Wei Wei, Xianling Mao
EMNLP 2022 Findings [paper]

Experience

Research Intern, Codedroid – Milpitas, CA June 2025 – Sept 2025

- Advisor: Huan Ren
- Worked on a mobile LLM-based coding assistant (Cursor-like) for on-device development workflows.
- Built a pipeline to interpret user queries and localize file(s) in a code repository requiring changes.

Research Intern, NetEase – Hangzhou, China June 2023 – Sept 2023

- Curated and standardized dialogue-hallucination data; verified via fine-tuning.
- Fine-tuned LLaMA2 for persona dialogue 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 in Alipay.

Education

University of California, Davis, Ph.D. in Computer Science Sept 2024 - Present

- **GPA:** 3.95/4.00

Huazhong University of Science and Technology, M.S. in Computer Science Sept 2021 – June 2024

- **GPA:** 3.90/4.00
- **Awards:** First-Class Graduate Scholarship

Dalian University of Technology, B.Eng in Software Engineering Sept 2017 – June 2021

- **GPA:** 3.89/4.00
- **Awards:** Mitsubishi & Outstanding Scholarship, Meritorious Winner of 2019 ICM

Service

Reviewer: ACL, EMNLP, NAACL, since 2023.

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