HAOWEI ZHANG

✉ haowei.zhang123@gmail.com · ☎ (+86) 18516760840 · % haowei-freesky

⚑ Research Interests: Multimodal Reasoning, Video Understanding, Efficient Inference

🎓 EDUCATION

Fudan University, Shanghai, China   2025 – Present

Ph.D. student in FudanNLP, supervised by Prof. Xipeng Qiu and Jinlan Fu

Technical University of Munich (TUM), Munich, Germnay   2021 – 2024

M.Sc. in Robotics, Cognition, Intelligence
Final Grade: 1.7 / 1.0 (German system, equivalent to 91 / 100)

Tongji University, Shanghai, China   2017 – 2021

B.Eng. in Mechatronics
Final Grade: 4.54 / 5.0 (equivalent to 91 / 100)

👥 EXPERIENCE

Yale NLP Lab remote   2024.10 – 2025.02

% https://yale-nlp.github.io/
Research Assistant   Leader: Yilun Zhao, Prof. Arman Cohan

- video understanding and reasoning of multimodal LLMs
- one paper accepted at CVPR as co-first author

LMU Research Group Munich, Germany   2023.11 – 2024.03

% https://tresp-lab.github.io/
Research Assistant   Leader: Zhen Han, Jindong Gu, Prof. Volker Tresp

- improving reasoning and calibration capabilities of multimodal LLMs
- one paper accepted at EMNLP 2024 Findings as first author

Yale NLP Lab remote   2023.03 – 2023.10

% https://yale-nlp.github.io/
Research Assistant   Leader: Yilun Zhao, Prof. Arman Cohan

- table-to-text generation capabilities of LLMs
- one paper accepted at EMNLP 2023 Industry Track as co-first author
- one paper accepted at EMNLP 2024 Demo Track as first author

🔍 PROJECTS

**GroundedPRM: Tree-Guided and Fidelity-Aware Process Reward Modeling
for Step-Level Reasoning**   2024.12 – 2025.03
- proposing a tree-guided and fidelity-aware framework for automatic process supervision
- under review at AAAI 2026 (second phase) as second author

**MMVU: A Comprehensive Benchmark for Expert-Level Multi-Disciplinary
Video Understanding and Reasoning**   2024.10 – 2025.02
% Paper Link | % Project Link
- comprehensive expert-level, multi-disciplinary benchmark for evaluating foundation models in video understanding
- attracting great attention from the community (download >20k on HF), used by Qwen2.5-VL, Kimi-VL, Seed1.5-VL testing, etc.

- accepted at CVPR 2025 as co-first author

Visual Question Decomposition on Multimodal Large Language Models   2023.10 – 2024.03

% Paper Link | % Project Link

- development of a novel evaluation framework to assess the quality of generated sub-questions
- proposing a finetuning pipeline to enhance the capabilities of multimodal LLMs for visual question decomposition
- instructing multimodal LLMs to perform selective question decomposition
- accpeted at EMNLP 2024 Findings as first author

OpenT2T: An Open-Source Toolkit for Table-to-Text Generation   2023.07 – 2023.09

% Paper Link | % Project Link

- first open-source toolkit for table-to-text generation, designed to reproduce performance comparison and to expedite the development of new models
- accepted at EMNLP 2024 Demo Track as first author

Investigating Table-to-Text Generation Capabilities of Large Language Models in Real-World Information Seeking Scenarios   2023.03 – 2023.07

% Paper Link | % Project Link

- investigation of the table-to-text capabilities of different LLMs within real-world information seeking scenarios
- demonstrating that high-performing LLMs can effectively serve as table-to-text generators, evaluators, and feedback generators
- accepted at EMNLP 2023 Industry Track as co-first author

♡ COMMUNITY SERVICE

Organizer of Knowledge-Intensive Multimodal Reasoning @ ICCV 2025
Reviewer of Reasoning and Planning for LLMs @ ICLR 2025
Reviewer of ACL Rolling Review 2025

♛ HONORS AND AWARDS

Outstanding Graduate of Tongji University	2021.04
DAAD-Scholarship (German national level scholarship)	2020.11

⚙ SKILLS

- Research: Python / Natural Language Processing / (Multimodal) Large Language Models
- Languages: English, German - Fluent / Chinese - Native speaker