# Zhehao Zhang

Phone: (+1) 640-240-4027 ◇ Email: zhehao.zhang.gr@dartmouth.edu
Homepage Google Scholar ◇ Github ◇ LinkedIn ◇ Twitter

## EDUCATION

**Dartmouth College** *Hanover, NH*
M.S. in Computer Science, GPA: 4.0/4.0 *Sep 2023-Jun 2025 (Expected)*

**Shanghai Jiao Tong University (SJTU)** *Shanghai, China*
B.E. in Artificial Intelligence (Honor Class), GPA: 3.74/4.0 Sep 2019-Jun 2023

## RESEARCH INTERESTS

My research interests span a broad range of topics within Natural Language Processing (NLP), with a focus on large language models (LLMs) and multi-modal large language models (MLLMs). I am particularly passionate about synthetic data for robust LLM evaluation and enhancement [1], tool-augmented LLMs (Language Agents) especially for tabular data [2], [3], human-centered NLP, and NLP for social good [4], [5]. Additionally, I am excited about topics such as grounded MLLMs, retrieval-augmented generation (RAG), long-context LLMs, multi-agent interaction, and LLM's core abilities like reasoning, planning, and knowledge extraction.

## PUBLICATIONS

[1] **Zhang, Zhehao**, J. Chen, and D. Yang, "DARG: Dynamic evaluation of large language models via adaptive reasoning graph," *arxiv*, 2024.

[2] **Zhang, Zhehao**, X. Li, Y. Gao, and J. Lou, "CRT-QA: A dataset of complex reasoning question answering over tabular data," *Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2023.

[3] **Zhehao Zhang**, Y. Gao, and J. Lou, "$E^5$: Zero-shot hierarchical table analysis using augmented LLMs via explain, extract, execute, exhibit and extrapolate," *Conference of the North American Chapter of the Association for Computational Linguistics (NAACL)*, 2024.

[4] **Zhang, Zhehao**, J. Chen, and D. Yang, "Mitigating biases in hate speech detection from a causal perspective," *Findings of Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2023.

[5] C. Ziems, W. Held, O. Shaikh, J. Chen, **Zhehao Zhang**, and D. Yang, "Can large language models transform computational social science?" *Computational Linguistics*, 2024.

[6] **Zhang, Zhehao**, T. Yu, H. Zhao, K. Xie, L. Yao, and S. Li, "Exploring soft prompt initialization strategy for few-shot continual text classification," *2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2024.

## RESEARCH EXPERIENCE

**Research Intern** Jun 2022 - May 2024
*Stanford University, Supervisors: Prof. Diyi Yang* Stanford, CA

- Developed DARG (Dynamic Evaluation of LLMs via Adaptive Reasoning Graph), a novel framework for dynamically evaluating LLMs by extending existing benchmarks with controlled complexity and diversity. Found significant performance decreases across 15 SOTA LLMs on reasoning tasks as complexity increased, revealing limitations of static benchmarks and highlighting the need for adaptive evaluation approaches. [1].
- Developed biased mitigation methods from the perspective of causal inference in hate speech detection models. Demonstrated significant performance improvements across 9 datasets, with an average 8.61 F1 score increase over baseline models and enhanced out-of-domain generalization [4].
- Participated in constructing a road map for using LLMs as computational social science (CSS) tools and contributed a set of prompting best practices and an extensive evaluation pipeline to measure the zero-shot performance of 13 language models on 24 representative CSS benchmarks. Responsible for building and analyzing various baseline models (e.g., T-5, Roberta, etc.) on all CSS datasets. [5]

## INDUSTRY EXPERIENCE

**Research Scientist/Engineer Intern** Jun 2024 - Sep 2024
*Adobe Research - Data Science Lab* San Jose, CA

- Proposed a multi-agent framework for multi-modal LLMs (e.g., GPT-4o) augmented by external tools such as code and task-specific vision models to conduct grounded fine-grained visual perception and reasoning tasks, especially on visual prompt perceptions and multi-image inference. Our proposal generally outperforms previous prompt engineering methods, in-context learning approaches, and agentic frameworks.
- Conducted a comprehensive survey on the personalization of LLMs, bridging the gap between text generation and downstream applications like recommendations; introduced new taxonomies for techniques, granularity, evaluation, and datasets; formalized foundational concepts; and highlighted key challenges and open problems

**Research Scientist Intern** Dec 2022 - Aug 2023
*Microsoft Research Asia - Data, Knowledge, and Intelligence Lab* Beijing, China

- Built a tool-augmented LLM-based agent table analysis system for large hierarchical table analysis using LangChain, which avoided hand-crafted in-context exemplars and considerably decreased the token usage in calling LLMs. This approach makes it possible for models with limited context length to analyze large-scale tabular data and achieve SOTA performances [3].
- Explored LLMs'(*e.g.,* GPT-4 etc.) reasoning ability on structured data. Constructed the first large-scale table question-answering dataset which requires the model to have multi-step complex reasoning capability with a detailed reasoning taxonomy. Comprehensively investigate LLMs' ability on different reasoning types [2].

## ACHIEVEMENTS

Merit Scholarship, awarded by Dartmouth College *2023-2025*
Zhiyuan Honor Scholarship, awarded by SJTU *2019-2023*
Merit Scholarship, awarded by SJTU [6] *2019-2023*

## SKILLS

| | |
|---|---|
| **Programming Languages** | Python, C/C++, MATLAB, JavaScript |
| **Machine Learning Tools** | PyTorch, Huggingface, Numpy, Scikit-learn, Pandas |
| **LLM-related Tools** | LangChain, LlamaIndex, Gradio, Ollama, Groq, LitGPT |
| **Spoken Language:** | English, Mandarin |

## SERVICE

| | |
|---|---|
| **EMNLP 2023** | Reviewer, Volunteer |
| **NeurIPS 2023, 2024** | Reviewer |
| **NAACL 2024** | Reviewer, Volunteer |
| **ACL 2024** | Reviewer |
| **COLM 2024** | Reviewer |
| **CIKM 2024** | Reviewer |