

Zhehao Zhang

Phone: (+1) 640-240-4027 ◊ Email: zhehao.zhang.gr@dartmouth.edu

Homepage: <https://zzh-sjtu.github.io/zhehaozhang.github.io/>

Google Scholar: <https://scholar.google.com/citations?user=QG-BAGwAAAAJ&hl=en>

EDUCATION

Dartmouth College

M.S. in Computer Science

Hanover, NH

Sep 2023-Jun 2025 (Expected)

Shanghai Jiao Tong University (SJTU)

B.E. in Artificial Intelligence (Honor Class), GPA: 3.70/4.0

Shanghai, China

Sep 2019-Jun 2023

PUBLICATIONS

- [1] **Zhang, Zhehao**, R. A. Rossi, B. Kveton, *et al.*, “Personalization of large language models: A survey,” *arXiv preprint arXiv:2411.00027*, 2024.
- [2] **Zhang, Zhehao**, R. A. Rossi, T. Yu, *et al.*, “Vipact: Visual-perception enhancement via specialized vlm agent collaboration and tool-use,” *ICLR (Under Review)*, 2025.
- [3] J. Wu, **Zhang, Zhehao**, Y. Xia, *et al.*, “Visual prompting in multimodal large language models: A survey,” *arXiv preprint arXiv:2409.15310*, 2024.
- [4] **Zhang, Zhehao**, J. Chen, and D. Yang, “DARG: Dynamic evaluation of large language models via adaptive reasoning graph,” *Advances in Neural Information Processing Systems (NeurIPS)*, 2024.
- [5] **Zhang, Zhehao**, W. Ma, and S. Vosoughi, “Is gpt-4v (ision) all you need for automating academic data visualization? exploring vision-language models’ capability in reproducing academic charts,” *Findings of Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2024.
- [6] **Zhehao Zhang**, Y. Gao, and J. Lou, “E⁵: 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.
- [7] C. Ziems, W. Held, O. Shaikh, J. Chen, **Zhehao Zhang**, and D. Yang, “Can large language models transform computational social science?” *Computational Linguistics*, 2024.
- [8] **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.
- [9] **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.
- [10] **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.

RESEARCH EXPERIENCE

Research Intern

Stanford University, Supervisor: *Prof. Diyi Yang*

Jun 2022 - May 2024

Stanford, CA

· **Dynamic Evaluation of Large Language Models (LLMs) [4]:**

- Built the **DARG** framework, introducing adaptive reasoning graphs to dynamically generate test data with controlled complexity, enhancing evaluation robustness for 15 SOTA LLMs across diverse reasoning tasks.
- Benchmarked LLMs on DARG-generated data, revealing performance drops under increased complexity and demonstrating the framework’s utility in improving LLMs through **fine-tuning with dynamic datasets**.

· **Bias Mitigation in Hate Speech Detection [10]:**

- Conducted **causal analysis** to identify confounding factors in hate speech detection, introducing the **Relative Spuriousness** metric for evaluating spurious features and guiding effective bias mitigation.
- Proposed **Multi-Task Intervention** and **Data-Specific Intervention** to mitigate spurious correlations in hate speech detection, achieving robust improvements across 9 datasets and enhanced OOD generalization.

· **LLMs for Computational Social Science (CSS) [7]:** Developed a roadmap for integrating LLMs into CSS, implemented prompting practices, and built fine-tuned models (e.g., T5, RoBERTa) to benchmark 13 LLMs on 24 tasks, showcasing their potential in augmenting human annotation and generative tasks.

Research Intern

Oct 2023 – Feb 2024

Dartmouth College, Supervisor: Prof. Soroush Vosoughi

Hanover, NH

· Vision-Language Models (VLMs) for Automatic Data Visualization [5]:

- Developed **AcademiaChart**, a dataset of 2525 **high-resolution** academic charts extracted directly from LaTeX source code on arXiv, showcasing diverse visualization types from top AI conferences and establishing a foundation for advancing automated data visualization and reproducibility.
- Benchmarked six VLMs for **code generation** to replicate academic charts, employing fine-grained human evaluations and automated metrics to demonstrate the potential of SOTA closed-source VLMs (e.g., GPT-4-V) in **significantly reducing researchers' effort** in creating accurate and reusable visualizations.

INDUSTRY EXPERIENCE

Research Scientist/Engineer Intern

Jun 2024 - Sep 2024

Adobe Research - Data Science Lab, Mentor: Dr. Ryan A. Rossi

San Jose, CA

· Multi-Agent Framework for VLMs [2]:

- Developed VIPACT, an agent framework that enhances VLMs by integrating **multi-agent collaboration** and **vision expert models**, enabling more precise visual perception and comprehensive System-2 reasoning.
- Outperformed SOTA prompt engineering techniques and agentic frameworks in **visual perception and reasoning tasks**, excelling in **visual prompt** [3] comprehension and **multi-image** inference.

- **A Survey on Personalized LLMs [1]:** Conducted a comprehensive survey on the personalization of LLMs, bridging the gap between **text generation** and **downstream applications**. Introduced novel taxonomies for personalization techniques, granularity, evaluation metrics, and datasets. Formalized foundational concepts and identified critical challenges and open research directions.

Research Scientist Intern

Dec 2022 - Aug 2023

Microsoft Research Asia - Data, Knowledge, Intelligence Lab, Mentor: Dr. Jian-Guang Lou Beijing, China

· Hierarchical Table Analysis with Code-Augmented LLM-based Agent [6]:

- Built E^5 , a tool-augmented framework for **hierarchical table QA** using GPT-4, eliminating the need for hand-crafted exemplars and achieving **SOTA performance** with a **44.38 Exact Match improvement**.
- Developed F^3 , an adaptive algorithm built on E^5 , reducing token usage by **93%** to enable **efficient large-scale table analysis** with **limited-context LLMs** while preserving accuracy.

· Complex Reasoning QA over Tabular Data (CRT-QA) [9]:

- Explored the reasoning abilities of LLMs on structured data by developing **CRT-QA**, the **first** large-scale table QA dataset requiring multi-step complex reasoning and introducing a **detailed reasoning taxonomy**.
- Proposed **ARC (Auto-exemplar-guided Reasoning with Code)**, a tool-augmented language agent framework leveraging Python (Pandas), achieving **SOTA results** without handcrafted exemplars.

ACHIEVEMENTS

Merit Scholarship, awarded by Dartmouth College

2023-2025

Zhiyuan Honor Scholarship and Merit Scholarship, awarded by SJTU [8]

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

SERVICE

Reviewer

EMNLP 2023, 2024; NeurIPS 2023, 2024; NAACL 2024; ACL 2024; COLM 2024

CIKM 2024; ICLR 2025; COLING 2025; IEEE TNNLS Journal

Volunteer

EMNLP 2023; NAACL 2024