Zhehao Zhang

(+1) 640-240-4027 • zhehao.zhang.gr@dartmouth.edu • LinkedIn Profile
GitHub Profile • Google Scholar • Home Page • @Zhehao_Zhang123

Education

Dartmouth College **Hanover, NH**
Master of Science in Computer Science, *Sep 2023-Jun 2025 (Expected)*

Shanghai Jiao Tong University (SJTU) **Shanghai, China**
Bachelor of Engineering in Artificial Intelligence Honor Class, *Sep 2019-Jun 2023*

Related Coursework: Natural language processing, Data mining, Computer Vision, Deep Learning, Machine Learning, Reinforcement learning, Data structure, Knowledge representation and reasoning, Intelligent speech recognition

Publications

[1] Ziems Caleb, William Held, Omar Shaikh, Jiaao Chen, **Zhehao Zhang**, and Diyi Yang. "Can Large Language Models Transform Computational Social Science?." arXiv preprint arXiv:2305.03514 (2023). PDF

Industry Experience

Microsoft Research Asia **Beijing, China**
Research Intern, Data, Knowledge, and Intelligence Lab, Mentor: Dr. Yan Gao *Dec 2022 - Aug 2023*

• Explored Large Language Models'(*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. (Submitted to EMNLP 2023)

• Built a table analysis system for large hierarchical tables in a zero-shot manner 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 state-of-the-art performances.

Research Experience

Stanford University **Stanford, CA**
Visiting Research Intern, Social and Language Technologies (SALT) Lab, Advisor: Prof.Diyi Yang *Jun 2022 - Present*

• Searched for biased grammar patterns on hate speech detection datasets. Analyzed the spuriousness of different biases using causal interference, and then proposed a method to mitigate such biases based on several confounders. Validated the effectiveness of the method by running experiments across nine hate speech detection datasets with an out-of-domain challenge set to reach positive conclusions on its use for reducing hate speech bias. (Submitted to EMNLP 2023)

• 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. PDF

Open-source Projects

Chinese Medical Named Entity Recognition (NER)

• Located and classified medical-related entities (e.g., symptoms, organs, etc.) on a large-scale Chinese biomedical dataset (CBLUE). Implement BERT and Roberta with conditional random field (or Long short-term memory) baseline model for both vanilla and nested settings. Improve the baseline model by further introducing the Flat-Lattice Transformer model (FLAT) and other techniques, including adversarial training and layer-wise learning rate decay.

Class-incremental Learning with Large Distribution Shift.

• Investigated continual learning (CL) with large domain shift which requires machine learning models to learn from a continuous stream of different data over time. Systematically analyzed existing CL methods and proposed a new distillation approach which has an 8.6% accuracy improvement over other replay-free methods.

Skills

Programming Languages: Python, C/C++, MATLAB
Tools and Frameworks: LangChain, Git, GitHub,LaTeX, PyTorch, Huggingface transformers, Numpy, Scikit-learn, pandas
Spoken Language: English, Mandarin

Service

NeurIPS 2023: Reviewer, Thirty-seventh Conference on Neural Information Processing Systems Reviewer **2023**
EMNLP 2023: Reviewer, The 2023 Conference on Empirical Methods in Natural Language Processing **2023**