

| | | |
|-----------|--|-------------------------|
| Education | The University of Texas at Dallas | 2024 - present |
| | - Ph.D. in Computer Science | Dallas, TX, USA |
| | - Advisor: Zhiyu Zoey Chen | |
| | Virginia Tech (transferred out) | 2023 - 2024 |
| | - Ph.D. in Computer Science | Blacksburg, VA, USA |
| | Soochow University | 2020 - 2023 |
| | - M.S. in Computer Science | Suzhou, Jiangsu, China |
| | - Advisors: Wenliang Chen , Xiabing Zhou | |
| | - Outstanding Graduate; Outstanding Thesis | |
| | Nanjing University of Posts and Telecommunications | 2016 - 2020 |
| | - B.Eng. in Computer Science | Nanjing, Jiangsu, China |
| | - GPA: 3.96 / 5 (Top 3% in the Department) | |

| | | |
|------------------|---|--|
| Research Summary | My current research focuses on enhancing (large) language models for social good, with a particular emphasis on <i>health domains</i> [1, 2, 4]. I am actively working on post-training techniques (<i>reinforcement learning from human feedback</i>) [1, 6], and synthetic data generation [1, 7] to equip models with complex and reliable capabilities, such as <i>instruction following</i> and <i>personalization</i> . I have experience in evaluating and improving language models, particularly in their dialogue abilities , such as keeping <i>self-consistent</i> [5] and <i>safe</i> [6] in conversation or understanding speaker's <i>core semantic meaning</i> [7] and <i>emotion</i> [8]. Additionally, I have a strong interest in vision-language models, retrieval-augmented methods, and language agents. Totally, I have 9 papers along these directions including 7 first-authored papers and 7 peer-reviewed publications in top-tier AI conferences such as ACL, NAACL, NeurIPS, etc. | |
|------------------|---|--|

| | | |
|----------------------|---|-----------------------|
| Research Experiences | Human Language Technology Research Institute at UTD | Sep. 2024 - present |
| | <i>Student Researcher</i> (Advisor: Zhiyu Zoey Chen) | Dallas, TX |
| | Topics: Preference Learning, Personalization | |
| | Tencent AI Lab | May. 2024 - Jul. 2024 |
| | <i>Research Intern</i> (Mentors: Haitao Mi , Linfeng Song) | Seattle, WA |
| | Topics: Reinforcement Learning from Human Feedback | |
| | Tencent AI Lab | Dec. 2021 - Dec. 2022 |
| | <i>Research Intern</i> (Mentor: Linfeng Jin) | Shenzhen, China |
| | Topics: Dialogue Understanding, Semi-supervised Learning | |
| | Institute of Human Language Technology at Soochow University | Sep. 2020 - Jul. 2021 |
| | <i>Student Researcher</i> (Advisors: Wenliang Chen , Xiabing Zhou) | Suzhou, China |
| | Topics: Dialogue State Tracking, Conversational Emotion Recognition | |

| | | |
|--------|--|--|
| Papers | (*denotes equal contribution) | |
| | 1. Aligning Large Language Models to Assist in Psychotherapy (<i>Preprint</i>) | |
| | Mian Zhang , Shaun M. Eack, Zhiyu Chen. | |
| | 2. CBT-Bench: Evaluating Large Language Models on Assisting Cognitive Behavior Therapy | |
| | (NAACL 2025) | |

Mian Zhang*, Xianjun Yang*, Xinlu Zhang, Travis Labrum, Jamie C. Chiu, Shaun M. Eack, Fei Fang, William Yang Wang, Zhiyu Chen.

3. [IDEA: Enhancing the Rule Learning Ability of Large Language Model Agent through Induction, Deduction, and Abduction](#) (*NeurIPS 2024 Workshop on Open-World Agents, ACL under submission*)
Kaiyu He, **Mian Zhang**, Shuo Yan, Peilin Wu, Zhiyu Zoey Chen.
4. [Large Language Models for Disease Diagnosis: A Scoping Review](#) (*Preprint*)
Shuang Zhou*, Zidu Xu*, **Mian Zhang***, Chunpu Xu*, Yawen Guo, Zaifu Zhan, Sirui Ding, Jiashuo Wang, Kaishuai Xu, Yi Fang, Liqiao Xia, Jeremy Yeung, Daochen Zha, Mingquan Lin, Rui Zhang.
5. [Inconsistent dialogue responses and how to recover from them](#) (*EACL 2024*)
Mian Zhang, Lifeng Jin, Linfeng Song, Haitao Mi, Dong Yu.
6. [SafeConv: Explaining and Correcting Conversational Unsafe Behavior](#) (*ACL 2023 Oral*)
Mian Zhang, Lifeng Jin, Linfeng Song, Haitao Mi, Wenliang Chen, Dong Yu.
7. [Friend-training: Learning from Different but Related Tasks](#) (*EACL 2023*)
Mian Zhang, Lifeng Jin, Linfeng Song, Haitao Mi, Xiabing Zhou, Dong Yu.
8. [Emotion Recognition in Conversation from Variable-Length Context](#) (*ICASSP 2023*)
Mian Zhang, Xiabing Zhou, Wenliang Chen, Min Zhang.
9. [A Pairing Enhancement Approach for Aspect Sentiment Triplet Extraction](#) (*KSEM 2023*)
Fang Yang, **Mian Zhang**, Gongzhen Hu, Xiabing Zhou.

Skills

Programming Language: Python, C/C++, Shell, \LaTeX , Matlab

Machine Learning Framework: PyTorch, NumPy, Transformers, Scikit-learn

Tool & Software: Vim, Git, pdb, Matplotlib, Pandas

Natural Language: Mandarin (native), English (advanced)

Selected Awards

| | |
|---|-------------|
| Outstanding Master's Thesis | 2024 |
| Outstanding Graduate Award at Soochow University | 2023 |
| First-class Scholarship at Soochow University | 2020 - 2021 |
| Principal's Award | 2019 |
| CUMCM National First Prize (top 1%) | 2018 |
| First-class Scholarship at NJUPT | 2017 - 2019 |