

# Chujie Zheng 郑楚杰

## Contact Information

Address	Room 4-504, FIT Building, Tsinghua University, Beijing 100084, China
Tel	86 18800116990
Email	chujiezhengchn@gmail.com
Homepage	<a href="https://chujiezheng.github.io/">https://chujiezheng.github.io/</a>
Google Scholar	<a href="https://scholar.google.com/citations?user=55zBNgUAAAAJ">https://scholar.google.com/citations?user=55zBNgUAAAAJ</a>

## Education

- **Ph.D student**, Department of Computer Science and Technology, Tsinghua University Sep 2020 – Present  
*Advisor: Prof. [Minlie Huang](#)*
- **Visiting Scholar**, Computer Science Department, UCLA Nov 2023 – Present  
*Advisor: Prof. [Nanyun \(Violet\) Peng](#)*
- **B.S.** in Physics, Tsinghua University Aug 2016 – Jun 2020

## Main Research Interests

- **Robustness and Alignment** of Large Language Models (LLMs)
- LLMs for **Social Good**
- Dialogue System, Natural Language Generation

## Research Highlights and Overview

- Published **10+** papers at the top-tier ML/NLP conferences (ICLR, ACL, EMNLP).
- Google Scholar citations **750+**, h-index **13**, i10-index **14**.
- Conducted a series of research on **LLMs' robustness and alignment**, covering the following topics:
  - *Perturbation robustness.* [The preprint](#) investigates LLMs' bias and robustness in multiple choice evaluations.
  - *Out-of-distribution generalization.* [The ACL 2023 paper](#) improves small dialogue models' generalization with LLM-synthesized data.
  - *Adversarial attack and defense.* [The EMNLP 2022 paper](#) constructs queries at scale to trigger chatbots' self-contradictory behaviors; [The ACL 2023 paper](#) trains LLMs via contrastive learning to align generated texts with reward functions.
  - *Dialogue safety.* [The ACL 2022 paper](#) identifies typical categories of dialogue systems' safety failures. [The EMNLP 2022 paper](#) releases practical classifiers to detect Chinese LLMs' offensive language.
- Conducted a series of research on **LLMs for social good**, covering the following topics:
  - *Emotional support dialogue system.* [The ACL 2021 paper](#) proposes a practical emotional support framework for modern neural dialogue systems and makes the first step towards data-driven emotional support systems; [The ACL 2023 paper](#) leverages large-scale LLM-synthesized data to boost open-domain emotional support dialogue; Combining the two research work, we built a Chinese emotional support chatbot [Emohaa](#) (launched by Lingxin AI).
  - *Empathetic dialogue generation.* [The ACL 2021 paper](#) formulates a fine-grained conceptual framework to characterize empathy expression; [The AAAI 2022 paper](#) and [the ACL 2023 paper](#) further enhance empathy expression with commonsense acquisition.
- Built a series of **popular NLP datasets**, including [ChID \(ACL 2019\)](#), [KDCConv \(ACL 2020\)](#), [ESConv \(ACL 2021\)](#), [CDCConv \(EMNLP 2022\)](#), [DiaSafety \(ACL 2022\)](#), and [COLD \(EMNLP 2022\)](#).

## Publications (\*: Equal Contribution)

1. **Chujie Zheng**, Hao Zhou, Fandong Meng, Jie Zhou, Minlie Huang. *Large Language Models Are Not Robust Multiple Choice Selectors*. [ICLR 2024 \(Spotlight: 5%\)](#).
2. **Chujie Zheng**, Pei Ke, Zheng Zhang, Minlie Huang. *Click: Controllable Text Generation with Sequence Likelihood Contrastive Learning*. [Findings of ACL 2023](#).

3. **Chujie Zheng**, Sahand Sabour, Jiaxin Wen, Zheng Zhang, Minlie Huang. *AugESC: Dialogue Augmentation with Large Language Models for Emotional Support Conversation*. [Findings of ACL 2023](#).
4. Jinfeng Zhou\*, **Chujie Zheng\***, Bo Wang, Zheng Zhang, Minlie Huang. *CASE: Aligning Coarse-to-Fine Cognition and Affection for Empathetic Response Generation*. [ACL 2023](#).
5. **Chujie Zheng\***, Jinfeng Zhou\*, Yinhe Zheng, Libiao Peng, Zhen Guo, Wenquan Wu, Zhengyu Niu, Hua Wu, Minlie Huang. *CDConv: A Benchmark for Contradiction Detection in Chinese Conversations*. [EMNLP 2022](#).
6. Siyang Liu\*, **Chujie Zheng\***, Orianna Demasi, Sahand Sabour, Yu Li, Zhou Yu, Yong Jiang, Minlie Huang. *Towards Emotional Support Dialog Systems*. [ACL 2021](#).
7. **Chujie Zheng**, Yong Liu, Wei Chen, Yongcai Leng, Minlie Huang. *CoMAE: A Multi-factor Hierarchical Framework for Empathetic Response Generation*. [Findings of ACL 2021](#).
8. **Chujie Zheng**, Yunbo Cao, Daxin Jiang, Minlie Huang. *Difference-aware Knowledge Selection for Knowledge-grounded Conversation Generation*. [Findings of EMNLP 2020](#).
9. Hao Zhou\*, **Chujie Zheng\***, Kaili Huang, Minlie Huang, Xiaoyan Zhu. *KdConv: A Chinese Multi-domain Dialogue Dataset Towards Multi-turn Knowledge-driven Conversation*. [ACL 2020](#).
10. **Chujie Zheng**, Minlie Huang, Aixin Sun. *ChID: A Large-scale Chinese IDiom Dataset for Cloze Test*. [ACL 2019](#).
11. Yuxian Gu\*, Jiaxin Wen\*, Hao Sun\*, Yi Song, Pei Ke, **Chujie Zheng**, Zheng Zhang, Jianzhu Yao, Lei Liu, Xiaoyan Zhu, Minlie Huang. *EVA2.0: Investigating Open-domain Chinese Dialogue Systems with Large-scale Pre-training*. [Machine Intelligence Research 2023](#).
12. Jiawen Deng\*, Jingyan Zhou\*, Hao Sun, **Chujie Zheng**, Fei Mi, Helen Meng, Minlie Huang. *COLD: A Benchmark for Chinese Offensive Language Detection*. [EMNLP 2022](#).
13. Hao Sun\*, Guangxuan Xu\*, Jiawen Deng, Jiale Cheng, **Chujie Zheng**, Hao Zhou, Nanyun Peng, Xiaoyan Zhu, Minlie Huang. *On the Safety of Conversational Models: Taxonomy, Dataset, and Benchmark*. [Findings of ACL 2022](#).
14. Sahand Sabour, **Chujie Zheng**, Minlie Huang. *CEM: Commonsense-aware Empathetic Response Generation*. [AAAI 2022](#).
15. Hao Sun\*, Zhenru Lin\*, **Chujie Zheng**, Siyang Liu, Minlie Huang. *PsyQA: A Chinese Dataset for Generating Long Counseling Text for Mental Health Support*. [Findings of ACL 2021](#).

## Preprints

---

16. **Chujie Zheng**, Minlie Huang. *Exploring Prompt-based Few-shot Learning for Grounded Dialog Generation*. [arXiv:2109.06513](#).
17. Hao Zhou\*, Pei Ke\*, Zheng Zhang\*, Yuxian Gu, Yinhe Zheng, **Chujie Zheng**, Yida Wang, Chen Henry Wu, Hao Sun, Xiaocong Yang, Bosi Wen, Xiaoyan Zhu, Minlie Huang, Jie Tang. *EVA: An Open-Domain Chinese Dialogue System with Large-Scale Generative Pre-Training*. [arXiv:2108.01547](#).

## Selected Awards and Honors

---

- |  |      |
|--|------|
| • Schlumberger Scholarship, Tsinghua University        | 2023 |
| • Comprehensive Merit Scholarship, Tsinghua University | 2022 |
| • Outstanding Undergraduate, Tsinghua University       | 2020 |
| • China National Scholarship (Top 2/100)               | 2019 |
| • Comprehensive Merit Scholarship, Tsinghua University | 2018 |

## Services

---

- Conference reviewer: ACL (22/23), EMNLP (21/22), NAACL (24), AAAI (22/23), EACL (23), ACL Rolling Review (21/22)
- Journal reviewer: ACM Transactions on the Web (22), ACM Transactions on Intelligent Systems and Technology (22), Knowledge-Based Systems (21)