

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	https://chujiezheng.github.io/
Google Scholar	https://scholar.google.com/citations?user=55zBNgUAAAAJ
Semantic Scholar	https://www.semanticscholar.org/author/Chujie-Zheng/146452866

Education

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

Research Interests

- **Robust and Aligned** Large Language Models (LLMs)
- LLMs for **Social Good**
- Dialogue System, Natural Language Generation

Research Highlights and Overview

- Published **10+** papers at the top-tier conferences (ACL, EMNLP, AAAI).
- Google Scholar citations **620+**, h-index **12**, i10-index **13**.
- Conducted a series of research on **LLMs' robustness and alignment**, covering the following topics:
 - *Perturbation robustness.* [Preprint](#) investigates LLMs' bias and robustness in multiple choice questions.
 - *Out-of-distribution generalization.* [AugESC](#) improves small models' generalization with LLM-synthesized data.
 - *Adversarial attack and defense.* [CDConv](#) constructs queries at scale to trigger chatbots' self-contradictions; [Click](#) trains LLMs via contrastive learning to align generated texts with reward functions.
 - *Dialogue safety.* [DiaSafety](#) identifies typical categories of dialogue systems' safety failures. [COLD](#) 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.* [ESC](#) proposes a practical emotional support framework for modern neural dialogue systems and makes the first step towards data-driven emotional support systems; [AugESC](#) 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.* [CoMAE](#) formulates a fine-grained conceptual framework to characterize empathy expression; [CEM](#) and [CASE](#) further enhance empathy expression with commonsense acquisition.
- Built a series of **popular NLP datasets**, including [ChID](#), [KDConv](#), [ESConv](#), [CDConv](#), [DiaSafety](#), and [COLD](#).

Publications (*: Equal Contribution)

1. **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](#).
2. **Chujie Zheng**, Pei Ke, Zheng Zhang, Minlie Huang. *Click: Controllable Text Generation with Sequence Likelihood Contrastive Learning*. [Findings of ACL 2023](#).
3. Jinfeng Zhou*, **Chujie Zheng***, Bo Wang, Zheng Zhang, Minlie Huang. *CASE: Aligning Coarse-to-Fine Cognition and Affection for Empathetic Response Generation*. [ACL 2023](#).
4. **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](#).
5. Siyang Liu*, **Chujie Zheng***, Orianna Demasi, Sahand Sabour, Yu Li, Zhou Yu, Yong Jiang, Minlie Huang. *Towards Emotional Support Dialog Systems*. [ACL 2021](#).

6. **Chujie Zheng**, Yong Liu, Wei Chen, Yongcai Leng, Minlie Huang. *CoMAE: A Multi-factor Hierarchical Framework for Empathetic Response Generation*. Findings of ACL 2021.
7. **Chujie Zheng**, Yunbo Cao, Daxin Jiang, Minlie Huang. *Difference-aware Knowledge Selection for Knowledge-grounded Conversation Generation*. Findings of EMNLP 2020.
8. 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.
9. **Chujie Zheng**, Minlie Huang, Aixin Sun. *ChID: A Large-scale Chinese IDiom Dataset for Cloze Test*. ACL 2019.
10. Yuxian Gu*, Jiabin 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.
11. Jiawen Deng*, Jingyan Zhou*, Hao Sun, **Chujie Zheng**, Fei Mi, Helen Meng, Minlie Huang. *COLD: A Benchmark for Chinese Offensive Language Detection*. EMNLP 2022.
12. 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.
13. Sahand Sabour, **Chujie Zheng**, Minlie Huang. *CEM: Commonsense-aware Empathetic Response Generation*. AAAI 2022.
14. 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

15. **Chujie Zheng**, Hao Zhou, Fandong Meng, Jie Zhou, Minlie Huang. *Large Language Models Are Not Robust Multiple Choice Selectors*. arXiv:2309.03882.
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.

Awards and Honors

- | | |
|--|------|
| • Comprehensive Scholarship (2nd Price), Tsinghua University | 2022 |
| • Chi-Sun YEH Scholarship (Top 5/100), Dept. of Physics, Tsinghua University | 2020 |
| • Outstanding Undergraduate, Tsinghua University | 2020 |
| • China National Scholarship (Top 2/100) | 2019 |
| • Overall Excellence Scholarship, Tsinghua University | 2018 |

Services

- Reviewer: ACL (22/23), EMNLP (21/22), AAAI (22/23), EACL (23), KNOSYS, TIST
- Review Assistant: EMNLP (20), AAAI (21), COLING (20)