Chujie Zheng 郑楚杰

Contact Information

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

Research Overview and Highlights

- I have a broad research interest in building **efficient, scalable, and trustworthy** AI systems, with the current focus on **LLM alignment and reasoning**. My research goal is to **advance and oversee AI systems with minimal human intervention and ensure they work responsibly and transparently**.

- Previously, I have conducted extensive research on **LLMs for social good**, with a main focus on building **LLMs for emotional support**.

- I also maintain the GitHub repository of **chat templates for LLMs**, which has received **600+** stars.

- Google Scholar citations **1800+**, h-index **21**, i10-index **23**.

Education

- **Ph.D. candidate** in Computer Science and Technology, Tsinghua University   Sep 2020 – Present
  Advisor: Minlie Huang

- **Visiting Scholar**, UCLA   Nov 2023 – Jun 2024
  Host: Nanyun (Violet) Peng

- **B.S.** in Foundational Mathematics and Physics, Tsinghua University   Aug 2016 – Jun 2020

Work Experiences

- Research Intern. Qwen Post-training Team, **Alibaba Cloud**   Oct 2024 – Present
  - Contributed to the **QwQ** series reasoning models
  - Built the **ProcessBench** benchmark for process error identification in mathematical reasoning
  - Contributed to the **Qwen2.5-Math-PRM-7B/72B** process reward models for mathematical reasoning
- Research Intern. AI Alignment Team, **01.AI**   Jul 2024 – Oct 2024
  - Contributed to **Yi-Lightning**, which **ranks #6 on Chatbot Arena and #3 in Math Category (as of 10/14/2024)**
- Research Intern. General Dialogue Team, **Baidu**   Feb 2022 – Jun 2022
  - Improved PLATO's dialogue consistency via building contradiction detectors

Selected Papers

1. **ProcessBench: Identifying Process Errors in Mathematical Reasoning**
   <u>**Chujie Zheng**</u>, Zhenru Zhang, Beichen Zhang, Runji Lin, Keming Lu, Bowen Yu, Dayiheng Liu, Jingren Zhou, Junyang Lin
   [paper] [repo] [🤗 data]

2. **The Lessons of Developing Process Reward Models in Mathematical Reasoning**
   Zhenru Zhang, <u>**Chujie Zheng**</u>, Yangzhen Wu, Beichen Zhang, Runji Lin, Bowen Yu, Dayiheng Liu, Jingren Zhou, Junyang Lin
   [paper] [🤗 model]

3. **Yi-Lightning Technical Report**
   01.AI
   [paper]

4. **Weak-to-Strong Extrapolation Expedites Alignment**
   <u>**Chujie Zheng**</u>, Ziqi Wang, Heng Ji, Minlie Huang, Nanyun Peng
   arXiv:2404.16792 *(120K+ downloads)*
   [paper] [repo] [🤗 model]

5. **On Prompt-Driven Safeguarding for Large Language Models**
   <u>**Chujie Zheng**</u>, Fan Yin, Hao Zhou, Fandong Meng, Jie Zhou, Kai-Wei Chang, Minlie Huang, Nanyun Peng
   ICML 2024 || SeT LLM Workshop @ ICLR 2024 *(Oral: 5%)*
   [paper] [repo]

6. **Large Language Models Are Not Robust Multiple Choice Selectors**
   <u>**Chujie Zheng**</u>, Hao Zhou, Fandong Meng, Jie Zhou, Minlie Huang
   ICLR 2024 *(Spotlight: 5%)*
   [paper] [repo]

Other Papers

7. **Click: Controllable Text Generation with Sequence Likelihood Contrastive Learning**
   <u>**Chujie Zheng**</u>, Pei Ke, Zheng Zhang, Minlie Huang
   Findings of ACL 2023

8. **AugESC: Dialogue Augmentation with Large Language Models for Emotional Support Conversation**
   <u>**Chujie Zheng**</u>, Sahand Sabour, Jiaxin Wen, Zheng Zhang, Minlie Huang
   Findings of ACL 2023

9. **CASE: Aligning Coarse-to-Fine Cognition and Affection for Empathetic Response Generation**
   Jinfeng Zhou*, <u>**Chujie Zheng***</u>, Bo Wang, Zheng Zhang, Minlie Huang
   ACL 2023 *(Oral)*

10. **EVA2.0: Investigating Open-domain Chinese Dialogue Systems with Large-scale Pre-training**
    Yuxian Gu*, Jiaxin Wen*, Hao Sun*, Yi Song, Pei Ke, <u>**Chujie Zheng**</u>, Zheng Zhang, Jianzhu Yao, Lei Liu, Xiaoyan Zhu, Minlie Huang

Machine Intelligence Research 2023

11. **CDConv: A Benchmark for Contradiction Detection in Chinese Conversations**
<u>Chujie Zheng*</u>, Jinfeng Zhou*, Yinhe Zheng, Libiao Peng, Zhen Guo, Wenquan Wu, Zhengyu Niu, Hua Wu, Minlie Huang
EMNLP 2022 (Oral)

12. **COLD: A Benchmark for Chinese Offensive Language Detection**
Jiawen Deng*, Jingyan Zhou*, Hao Sun, <u>Chujie Zheng</u>, Fei Mi, Helen Meng, Minlie Huang
EMNLP 2022 (Oral)

13. **On the Safety of Conversational Models: Taxonomy, Dataset, and Benchmark**
Hao Sun*, Guangxuan Xu*, Jiawen Deng, Jiale Cheng, <u>Chujie Zheng</u>, Hao Zhou, Nanyun Peng, Xiaoyan Zhu, Minlie Huang
Findings of ACL 2022

14. **CEM: Commonsense-aware Empathetic Response Generation**
Sahand Sabour, <u>Chujie Zheng</u>, Minlie Huang
AAAI 2022 (Oral)

15. **Exploring Prompt-based Few-shot Learning for Grounded Dialog Generation**
<u>Chujie Zheng</u>, Minlie Huang
arXiv:2109.06513

16. **EVA: An Open-Domain Chinese Dialogue System with Large-Scale Generative Pre-Training**
Hao Zhou*, Pei Ke*, Zheng Zhang*, Yuxian Gu, Yinhe Zheng, <u>Chujie Zheng</u>, Yida Wang, Chen Henry Wu, Hao Sun, Xiaocong Yang, Bosi Wen, Xiaoyan Zhu, Minlie Huang, Jie Tang
arXiv:2108.01547

17. **Towards Emotional Support Dialog Systems**
Siyang Liu*, <u>Chujie Zheng*</u>, Orianna Demasi, Sahand Sabour, Yu Li, Zhou Yu, Yong Jiang, Minlie Huang (*: Equal contribution)
ACL-IJCNLP 2021 (Oral)

18. **CoMAE: A Multi-factor Hierarchical Framework for Empathetic Response Generation**
<u>Chujie Zheng</u>, Yong Liu, Wei Chen, Yongcai Leng, Minlie Huang
Findings of ACL-IJCNLP 2021

19. **PsyQA: A Chinese Dataset for Generating Long Counseling Text for Mental Health Support**
Hao Sun*, Zhenru Lin*, <u>Chujie Zheng</u>, Siyang Liu, Minlie Huang
Findings of ACL-IJCNLP 2021

20. **Difference-aware Knowledge Selection for Knowledge-grounded Conversation Generation**
<u>Chujie Zheng</u>, Yunbo Cao, Daxin Jiang, Minlie Huang
Findings of EMNLP 2020

21. **KdConv: A Chinese Multi-domain Dialogue Dataset Towards Multi-turn Knowledge-driven Conversation**
Hao Zhou*, <u>Chujie Zheng*</u>, Kaili Huang, Minlie Huang, Xiaoyan Zhu
ACL 2020

22. **ChID: A Large-scale Chinese IDiom Dataset for Cloze Test**
<u>Chujie Zheng</u>, Minlie Huang, Aixin Sun

ACL 2019

Selected Awards and Honors

- Comprehensive Merit Scholarship, Tsinghua University   2021 – 2024
- Outstanding Undergraduate, Tsinghua University   2020
- China National Scholarship (Top 2/100)   2019
- Comprehensive Merit Scholarship, Tsinghua University   2018

Academic Services

- **Area Chair:** ACL (24/25), EMNLP (24), NAACL (25), ACL Rolling Review (24/25)

- **Reviewer:** ICLR (25), NeurIPS (24/25), ICML (24), COLM (24/25), ACL (22/23), EMNLP (21/22), NAACL (24), EACL (23), ACL Rolling Review (21/22/23), CogSci (24), AAAI (22/23)