

# Pei Zhou

*Ph.D. candidate in NLP and ML*

✉ [peiz@usc.edu](mailto:peiz@usc.edu)

📄 <https://shaoxia57.github.io/>

*Large language models, language agents,  
and complex reasoning*

## Education

- 2019 – 2024 **University of Southern California**, Los Angeles, CA, United States  
(expected) **PhD in Computer Science**, Advisor: *Xiang Ren & Jay Pujara*
- 2015 – 2019 **University of California, Los Angeles**, Los Angeles, CA, United States  
**Bachelor of Science in Mathematics of Computation**, Latin Honor

## Research Experience

- 09/2023 – **Google DeepMind**, Mountain View, CA, United States.  
Now Research Collaborator with Swaroop Mishra, Huaixiu Steven Zheng, and Quoc Le
- 05/2023 – **Google-Bard**, Mountain View, CA, United States.  
08/2023 Research Intern with Shyam Upadhyay and Manaal Faruqui
- 05/2022 – **Allen Institute for AI**, Seattle, WA, United States.  
12/2022 Research Intern with Prithviraj Ammanabrolu and Yejin Choi
- 5-8/2021 **Amazon Alexa AI**, Sunnyvale, CA, United States.  
5-8/2020 Applied Scientist Intern with Yang Liu and Dilek Hakkani-Tur
- 09/2017 – **UCLA-NLP Group**, Los Angeles, CA, United States.  
06/2019 Undergraduate Researcher with Kai-Wei Chang

## Selected Publications

Full list see Google Scholar.

- ArXiv 2023 **Pei Zhou**, Aman Madaan, Srividya Pranavi Potharaju, Aditya Gupta, Kevin R McKee, Ari Holtzman, Jay Pujara, Xiang Ren, Swaroop Mishra, Aida Nematzadeh, Shyam Upadhyay, Manaal Faruqui. How FaR Are Large Language Models From Agents with Theory-of-Mind? (Under Submission)
- ACL 2023 **Pei Zhou**, Andrew Zhu, Jennifer Hu, Jay Pujara, Xiang Ren, Chris Callison-Burch, Yejin Choi, Prithviraj Ammanabrolu. I Cast Detect Thoughts: Learning to Converse and Guide with Intents and Theory-of-Mind in Dungeons and Dragons.

- EMNLP 2022 **Pei Zhou**, Hyundong Cho, Pegah Jandaghi, Dong-Ho Lee, Bill Yuchen Lin, Jay Pujara, Xiang Ren. Reflect, Not Reflex: Inference-Based Common Ground Improves Dialogue Response Quality.
- ACL 2022 **Pei Zhou**, Karthik Gopalakrishnan, Behnam Hedayatnia, Seokhwan Kim, Jay Pujara, Xiang Ren, Yang Liu, Dilek Hakkani-Tur. Think Before You Speak: Explicitly Generating Implicit Commonsense Knowledge for Response Generation.
- EMNLP 2021 **Pei Zhou**, Rahul Khanna, Seyeon Lee, Bill Yuchen Lin, Daniel Ho, Jay Pujara, Xiang Ren. RICA: Evaluating Robust Inference Capabilities Based on Commonsense Axioms.
- EMNLP 2021 Ninareh Mehrabi\*, **Pei Zhou\***, Fred Morstatter, Jay Pujara, Xiang Ren, Aram Galstyan. Lawyers are Dishonest? Quantifying Representational Harms in Commonsense Knowledge Resources. (\* denotes equal contribution)
- Findings of EMNLP 2021 **Pei Zhou**, Pegah Jandaghi, Bill Yuchen Lin, Justin Cho, Jay Pujara, Xiang Ren. Probing Commonsense Explanation in Dialogue Response Generation.
- SIGDIAL 2021 **Pei Zhou**, Karthik Gopalakrishnan, Behnam Hedayatnia, Seokhwan Kim, Jay Pujara, Xiang Ren, Yang Liu, Dilek Hakkani-Tur. Commonsense-Focused Dialogues for Response Generation: An Empirical Study.
- Findings of EMNLP 2020 Bill Yuchen Lin, Wangchunshu Zhou, Ming Shen, **Pei Zhou**, Chandra Bhagavatula, Yejin Choi, Xiang Ren. CommonGen: A constrained text generation challenge for generative commonsense reasoning.
- EMNLP 2019 **Pei Zhou**, Weijia Shi, Jieyu Zhao, Kuan-Hao Huang, Muhao Chen, Ryan Cotterell, Kai-Wei Chang. Examining Gender Bias in Languages with Grammatical Gender.
- EMNLP 2019 Weijia Shi, Muhao Chen, **Pei Zhou**, Kai-Wei Chang. Retrofitting contextualized word embeddings with paraphrases.
- EMNLP 2023 Hyunwoo Kim, Jack Hessel, Liwei Jiang, Peter West, Ximing Lu, Youngjae Yu, **Pei Zhou**, Ronan Le Bras, Malihe Alikhani, Gunhee Kim, Maarten Sap, Yejin Choi. SODA: Million-scale Dialogue Distillation with Social Commonsense Contextualization. Outstanding Paper Award
- ArXiv 2023 Aman Madaan\*, Pranjal Aggarwal\*, Ankit Anand, Srividya Pranavi Potharaju, Swaroop Mishra, **Pei Zhou**, Aditya Gupta, Dheeraj Rajagopal, Karthik Kappaganthu, Yiming Yang, Shyam Upadhyay, Mausam, Manaal Faruqui. AutoMix: Automatically Mixing Language Models.
- EMNLP 2020 Workshop Ting-Yun Chang, Yang Liu, Karthik Gopalakrishnan, Behnam Hedayatnia, **Pei Zhou**, Dilek Hakkani-Tur. Incorporating commonsense knowledge graph in pretrained models for social commonsense tasks.

---

## Awards and Honors

- 2019-2023 **Annenberg Fellowship** (approximately \$45,000 per year)
- 2021 **USC Graduate School Travel Award**
- 2019 **ACL Student Research Travel Grant**
- 2018-2019 **UCLA Internet Research Initiative Prize** (<5%)
- 2018 **ASA Datafest Judge's Choice Award**

---

## Invited Talks

- Aug 2023 **Situated Reasoning for Large Language Model-Based Agents**  
Google, Mountain View, CA, United States
- Aug 2023 **Common Ground Reasoning for Communicative Agents**  
USC-NLP Group Talk Series, Los Angeles, CA, United States
- Aug 2022 **Learning to Guide and Converse with Theory-of-Mind and Reinforcement Learning in Grounded Dialogues**  
AI2, Seattle, WA, United States
- Sep 2021 **Smooth and Robust Commonsense Reasoning for Communication Models**  
Information Sciences Institute (ISI)-NLP Seminar, Marina Del Rey, CA, United States

---

## Professional Services

- Organizer **Workshop on Theory-of-Mind at ICML 2023**  
**AI Rising Star Symposium 2019**
- Mentoring **AACL-IJCNLP Student Research Workshop 2021**
- Program 100+ papers from **ACL, EMNLP, NAACL** (2020-present)
- Committee **ACL Rolling Review** (2021-present), **ICML and AAI** (2020-present)

---

## Skills

- Programing Python, Pytorch, Bash, JAX, C++, Colab, R, Matlab
- Language English (bilingual proficiency), Chinese (native), and Japanese (Intermediate)

## Teaching and Mentoring

Teaching Teaching Assistant for **USC-CSCI699: Ethics in Natural Language Processing** (PhD-level seminar)

Lecture at **USC-CSCI588 on Commonsense Reasoning for Language Models**

Mentoring Twice-elected mentor for **USC-Center for Undergraduate Research in Viterbi Engineering (CURVE) Fellowship**

Mentored two undergraduate students. Both received poster presentation on research projects.