

# Hongli Zhan

Ph.D. Candidate, UT Austin

Department of Linguistics  
The University of Texas at Austin  
✉ [honglizhan@utexas.edu](mailto:honglizhan@utexas.edu)  
🌐 Website: [honglizhan.github.io](https://honglizhan.github.io)  
🐙 Github: [github.com/honglizhan](https://github.com/honglizhan)  
in LinkedIn: [hongli-zhan-761291198](https://www.linkedin.com/in/hongli-zhan-761291198)

## Education



### Ph.D. in Computational Linguistics

Aug 2021 – Present

Department of Linguistics, The University of Texas at Austin, Austin, TX

- Advisor: Dr. [Junyi Jessy Li](#)
- Committee: Dr. [Desmond C. Ong](#), Dr. [David I. Beaver](#), Dr. [Kyle Mahowald](#)



### B.A. in English Linguistics

Sep 2017 – June 2021

School of Foreign Languages, Shanghai Jiao Tong University, Shanghai, China

## Industry Experience



### Research Scientist Intern (Part-Time)

Jan 2026 – May 2026

MBZUAI Institute of Foundation Models, Silicon Valley Lab, Sunnyvale, CA

- Post-training



### Research Scientist Intern

May 2025 – Aug 2025

IBM Research, IBM Thomas J. Watson Research Center, Yorktown Heights, NY

- Manager: Dr. [Raya Horesh](#); Mentors: Dr. [Muneeza Azmat](#) & Dr. [Pin-Yu Chen](#)
- Developed a scalable method for continuous safety governance of LLMs; the pipeline directly supports the safety and trustworthiness of IBM's [Granite](#) models
- Work in preparation for a first-authored paper & a first-authored U.S. patent



### Research Scientist Intern

May 2024 – Aug 2024

IBM Research, IBM Thomas J. Watson Research Center, Yorktown Heights, NY

- Manager: Dr. [Raya Horesh](#); Mentors: Dr. [Muneeza Azmat](#) & Dr. [Mikhail Yurochkin](#)
- Designed and implemented SPRI, a context-aware, principle-guided framework that enables LLMs to achieve human-level alignment with minimal manual oversight; SPRI forms the core synthetic data generation pipeline for IBM's [Granite Guardian](#) models
- Work led to a first-authored paper at ICML 2025 [C6] & a first-authored U.S. patent [P1]





## Refereed Publications

\* denotes equal contributions

(C stands for conference publications, J for journals, and P for patents)

### In Conference Proceedings

- [C6] 2025 **Hongli Zhan**, Muneeza Azmat, Raya Horesh, Junyi Jessy Li, and Mikhail Yurochkin. SPRI: Aligning large language models with context-situated principles. In *Proceedings of the 42nd International Conference on Machine Learning (ICML)*, 2025. [26.9% acceptance rate (3,260 out of 12,107 submissions)]
- [C5] 2024 **Hongli Zhan**, Allen Zheng, Yoon Kyung Lee, Jina Suh, Junyi Jessy Li, and Desmond C. Ong. Large language models are capable of offering cognitive reappraisal, if guided. In *Proceedings of the 1st Conference on Language Modeling (COLM)*, 2024. [28.8% acceptance rate (299 out of 1,036 submissions)]

- [C4] 2024 Yoon Kyung Lee, Jina Suh, **Hongli Zhan**, Junyi Jessy Li, and Desmond C. Ong. Large language models produce responses perceived to be empathic. In *Proceedings of the 12th IEEE International Conference on Affective Computing and Intelligent Interaction (ACII)*, 2024. [36.7% acceptance rate (44 out of 120 submissions)] .
- [C3] 2023 **Hongli Zhan**, Desmond C. Ong, and Junyi Jessy Li. Evaluating subjective cognitive appraisals of emotions from large language models. In Houda Bouamor, Juan Pino, and Kalika Bali, editors, *Findings of the Association for Computational Linguistics: EMNLP 2023 (EMNLP Findings)*, pages 14418–14446, Singapore, December 2023. Association for Computational Linguistics. [45.4% acceptance rate (1,758 out of 3,868 submissions)] .
- [C2] 2023 Tiberiu Sosea\*, **Hongli Zhan\***, Junyi Jessy Li, and Cornelia Caragea. Unsupervised extractive summarization of emotion triggers. In Anna Rogers, Jordan Boyd-Graber, and Naoaki Okazaki, editors, *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers) (ACL)*, pages 9550–9569, Toronto, Canada, July 2023. Association for Computational Linguistics. [23.5% acceptance rate (910 out of 3,872 submissions)] .
- [C1] 2022 **Hongli Zhan\***, Tiberiu Sosea\*, Cornelia Caragea, and Junyi Jessy Li. Why do you feel this way? summarizing triggers of emotions in social media posts. In Yoav Goldberg, Zornitsa Kozareva, and Yue Zhang, editors, *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 9436–9453, Abu Dhabi, United Arab Emirates, December 2022. Association for Computational Linguistics. [22.1% acceptance rate (715 out of 3,242 submissions)] .

## Patents

- [P1] **Hongli Zhan**, Mikhail Yurochkin, Raya Horesh, Adriana Alvarado Garcia, Heloisa Caroline de Souza Pereira Candello, and Muneeza Azmat. Context-aware principle-guided (framework and system) for synthetic data generation. **U.S. Patent Application** [Filed and Pending].

## Professional Activities

### Organization and Service

- 2023 – 2024 **Organizer**, *Natural Language Learning Reading Group*, UT Austin
- Student Volunteer**, *ACL 2023*

### Reviewing

- 2025 **Conference on Language Modeling**, (*COLM*)
- 2025 **Annual Meeting of the Association for Computational Linguistics**, (*ACL*)
- 2024 **IEEE International Conference on Robotics and Automation**, (*ICRA*)
- 2023 **Conference on Empirical Methods in Natural Language Processing**, (*EMNLP*)

## Teaching Experience

### The University of Texas at Austin

- Fall 2024 **LIN 371: Machine Learning for Text Analysis**, *Guest Lecturer*  
Instructor: Prof. [Junyi Jessy Li](#)
- Spring 2024 **LIN 371: Machine Learning for Text Analysis**, *Graduate Teaching Assistant*  
Instructor: Prof. [Junyi Jessy Li](#)
- Spring 2023 **LIN 306: Intro to the Study of Language**, *Guest Lecturer*  
Instructor: Seyeon Park
- Fall 2022 **LIN 373N: Machine Learning Toolbox for Text Analysis**, *Graduate Teaching Assistant*  
Instructor: Prof. [Junyi Jessy Li](#)
- Fall 2021 **LIN 350: Computational Semantics**, *Graduate Teaching Assistant*  
Instructor: Prof. [Katrin Erk](#)

## Mentoring

2023 – 2024 **Allen Zheng**, *B.S. in Computer Science (Turing Honors) & Mathematics*, UT Austin

## Fellowships & Awards

### IBM Internal Awards

May 2025 **Patent File Award**, *IBM Research*, 867 USD

### Awards from The University of Texas at Austin

Fall 2024 **Professional Development Award for Attending COLM 2024**, *UT Austin*, 1,200 USD

Fall 2023 **Professional Development Award for Attending EMNLP 2023**, *UT Austin*, 1,160 USD

Fall 2022 **Professional Development Award for Attending EMNLP 2022**, *UT Austin*, 1,200 USD

Spring 2022 **COLA Supplemental Graduate School Fellowship**, *UT Austin*, 5,000 USD

### Awards from Shanghai Jiao Tong University

2021 **Outstanding Graduate**, *Shanghai Jiao Tong University*

2021 **Outstanding Undergraduate Thesis Award**, *Shanghai Jiao Tong University*

## Invited Talks

Oct 11, 2023 **Can I Cheer You Up? Psychologically-Grounded Emotional Support: Leveraging LLMs for Targeted Re-Appraisals**, *Computational Affective and Social Cognition Lab (CASCogLab)*, Department of Psychology, The University of Texas at Austin, Austin, TX

Feb 22, 2023 **An Introduction to Computational Linguistics**, *Westlake High School*, Austin, TX