Zihao He

zihaoh@usc.edu Homepage

RESEARCH INTERESTS

NLP, Societal Impacts of LLMs, Alignment and Safety, Computational Social Science

EDUCATION

University of Southern California — **Los Angeles, CA, U.S.**
Ph.D. in Computer Science — *08/2019 - 02/2025*

- Advisor: Prof. Kristina Lerman

Tsinghua University — **Beijing, China**
M.Eng. in Computer Technologies (Transferred) — *09/2018 - 06/2019*

Beijing University of Posts and Telecommunications — **Beijing, China**
B.Eng. in Communications Engineering — *09/2014 - 06/2018*

SELECTED PUBLICATIONS AND PREPRINTS

Societal Impacts of LLMs

- COMMUNITY-CROSS-INSTRUCT: Unsupervised Instruction Generation for Aligning Large Language Models to Online Communities
  Zihao He, Rebecca Dorn, Siyi Guo, Minh Duc Chu, Kristina Lerman
  In *Proceedings of EMNLP'24* [Paper]

- Improving and Assessing the Fidelity of Large Language Models Alignment to Online Communities
  Minh Duc Chu, **Zihao He**, Rebecca Dorn, Kristina Lerman
  To appear in *Proceedings of NAACL'25* [Preprint]

- Reading Between the Tweets: Reading Between the Tweets: Deciphering Ideological Stances of Interconnected Mixed-ideology Communities
  Zihao He, Ashwin Rao, Siyi Guo, Negar Mokhberian, Kristina Lerman
  In *Findings of EACL'23* [Paper]

Alignment and Safety

- How Susceptible are Large Language Models to Ideological Manipulation?
  Kai Chen, **Zihao He**, Jun Yan, Taiwei Shi, Kristina Lerman
  In *Proceedings of EMNLP'24* [Paper] Best Paper Runner-up at SeT LLM @ ICLR 2024

- Whose Emotions and Moral Sentiments Do Language Models Reflect?
  Zihao He, Siyi Guo, Ashwin Rao, Kristina Lerman
  In *Findings of ACL'24* [Paper] [Media Coverage]

Multimodal Machine Learning

- ALCAP: Alignment-Augmented Music Captioner
  Zihao He, Weituo Hao, Wei-Tsung Lu, Changyou Chen, Kristina Lerman, Xuchen Song
  In *Proceedings of EMNLP'23* [Paper]

Representation Learning

- Infusing Knowledge from Wikipedia to Enhance Stance Detection
  Zihao He, Negar Mokhberian, Kristina Lerman
  In *Proceedings of WASSA'22* [Paper]

- Speaker Turn Modeling for Dialogue Act Classification
  Zihao He, Leili Tavabi, Kristina Lerman, Mohammad Soleymani
  In *Findings of EMNLP'21* [Paper]

- Detecting Polarized Topics Using Partisanship-aware Contextualized Topic Embeddings
  Zihao He, Negar Mokhberian, António Câmara, Andrés Abeliuk, Kristina Lerman
  In *Findings of EMNLP'21* [Paper] [Media Coverage]

- Graph Embedding with Personalized Context Distribution
  Zihao He, Di Huang, Yuzhong Huang, Kexuan Sun, Sami Abu-El-Haija, Bryan Perozzi, Kristina Lerman, Fred Morstatter, Aram Galstyan
  In *Companion Proceedings of the Web Conference'20* [Paper]

INDUSTRIAL EXPERIENCE

Research Scientist Intern @ ByteDance **San Jose, CA, U.S.**
Speech, Audio, and Music Intelligence Team *05/2023 – 08/2023, 05/2022 - 08/2022*

- Mentors: Weituo Hao, Xuchen Song
- Projects: Improving Music Interpretation Generation by Music-Text Alignment; Pretraining of Audio-Language Models

Applied Scientist Intern @ Amazon **Seattle, WA, U.S. (Remote)**
Amazon Kids Team *05/2021 – 08/2021*

- Mentors: Matthew Butler
- Project: Text Embedding-based Book Personalization

Algorithm Development Intern @ DiDi Global **Beijing, China**
Research Team *04/2019 – 07/2019*

- Mentors: Chao Feng
- Project: Multi-task Learning for Multi-label Driver Misbehavior Classification

HONORS AND AWARDS

- Graduate School Travel Award, University of Southern California. 2024.
- Annenberg Fellowship, University of Southern California. 2019.
- National Scholarship for Excellent Academic Performance (top 1.2%). 2018.

SERVICES

- PC/Reviewer: ACL Rolling Review, EMNLP, ACL, NAACL, SIGKDD, WSDM, WWW, CSS

TEACHING EXPERIENCE

- DSCI-531: Fairness in Artificial Intelligence. Instructor: Kristina Lerman. Spring 2022
- CSCI-566: Deep Learning and its Applications. Instructor: Yue Zhao. Spring 2024

KEY SKILLS

Programming Languages	Python, C++, Java, HTML
Machine Learning Libraries	PyTorch, Transformers, Scikit-Learn, NumPy, Pandas