Last Updated: November 25, 2023

Jieyu Zhao

PHE #332, USC, Los Angeles, CA | https://jyzhao.net | jieyuz@usc.edu

RESEARCH INTERESTS

- Detecting potential stereotypes in machine learning models
- Developing computational approaches to enhance the fairness in NLP applications

EDUCATION AND EXPERIENCE

Assistant Professor. University of Southern California, CA, USA 2023.08 - Current

Postdoc. University of Maryland, College Park, MD, USA 2022.01 - 2023.08
UMD Institute for Advanced Computer Studies
CLIP Lab; w/ Hal Daumé III

Ph.D. University of California, Los Angeles, CA, USA 2017.09 - 2021.12
Dept. of Computer Science
UCLANLP Group; w/ Kai-Wei Chang

Ph.D. University of Virginia (transferred), VA, USA 2016.08 - 2017.08

M.S. Beihang University, Beijing, China 2013.09 - 2016.01
Major: Virtualization, Data Mining; School of Computer Science and Engineering

B.S. Beihang University, Beijing, China 2009.09 - 2013.07
Major: Computer Science and Engineering; School of Advanced Engineering

Google Research 2021.06 - 2021.09
Intern, SIR Responsible ML, New York City (Mentor: Xuezhi Wang; Yao Qin)

Allen Institute for Artificial Intelligence 2020.09 - 2020.12
Part-time Intern, Aristo team, Seattle (Mentor: Daniel Khashabi)

Microsoft 2020.06 - 2020.09
Intern, Microsoft Research, Redmond (Mentor: Chris Brockett)

Microsoft 2019.06 - 2019.09
Intern, Microsoft Research, Redmond (Mentor: Ahmed Hassan Awadallah)

DiDi Chuxing 2016.01 - 2016.05
Intern, DiDi Chuxing Research, Beijing (Mentor: Pinghua Gong)

HONORS, AWARDS, AND SCHOLARSHIPS

1. Computing Innovation Fellows 2021
2. Rising stars in EECS 2021
3. Microsoft PhD Fellowship 2020
4. SoCalNLP Symposium 2018 Best Poster Award 2018
5. UCLA Graduate Division Fellowships 2017
6. EMNLP 2017 Best Long Paper Award 2017
7. Outstanding Graduate of Beijing City 2016
8. Master Thesis Award 2016
9. 2nd Prize University Level Scholarship 2013, 2014, 2015
10. Outstanding Graduate, Beihang University. 6/236. 2014
11. National Scholarship. 2/236. 2014
12. The Award of " Outstanding Graduates" of Beihang University 2013
13. Special Scholarship for Freshmen. Top 6 of all freshmen (around 3000). 2009

PUBLICATIONS

Pre-print

1. J. Qu, L. Li, **J. Zhao**, S. Dev, K.-W. Chang. DisinfoMeme: A Multimodal Dataset for Detecting Meme Intentionally Spreading Out Disinformation.

Published

1. Sandra Sandoval, **J. Zhao**, Marine Carpuat, Hal Daume. A Rose by Any Other Name would not Smell as Sweet: Social Bias in Name Mistranslations. EMNLP 2023.
2. Yixin Wan, **J. Zhao**, Aman Chadha, Nanyun Peng, Kai-Wei Chang. Are Personalized Stochastic Parrots More Dangerous? Evaluating Persona Biases in Dialogue Systems. EMNLP 2023 Findings.
3. Ruyuan Zuo, **J. Zhao**. Mind What You Measure For: A Study on Reliability of Prompt-Based Bias Measurement. WiNLP 2023.
4. Ruijie Zheng, Xiyao Wang, Yanchao Sun, Shuang Ma, **J. Zhao**, Huazhe Xu, Hal Daume, Furong Huang. TACO: Temporal Latent Action-Driven Contrastive Loss For Visual Reinforcement Learning. NeurIPS 2023.
5. H. An, Z. Li, **J. Zhao**, R. Rudinger. SODAPOP: Open-Ended Discovery of Social Biases in Social Commonsense Reasoning Models. EACL 2023.
6. A. Ovalle, S. Dev, **J. Zhao**, M. Sarrafzadeh, K.-W. Chang. Auditing Algorithmic Fairness in Machine Learning for Health with Severity-Based LOGAN. AAAI 2023 Health Intelligence Workshop.
7. **J. Zhao**, X. Wang, Y. Qin, J. Chen, K.-W. Chang. Investigating Ensemble Methods for Model Robustness Improvement of Text Classifiers. EMNLP Findings, 2022.
8. S. Dev, E. Sheng, **J. Zhao**, A. Smstutz, J. Sun, Y. Hou, M. Sanseverino, J. Kim, A. Nishi, N. Peng, K.-W. Chang. On Measures of Biases and Harms in NLP. AACL 2022.
9. A. Kwako, E. Wan, **J. Zhao**, K.-W. Chang, L. Cai, and M. Hansen. Using Item Response Theory to Measure Gender and Racial Bias of a BERT-based Automated English Speech Assessment System. BEA-2022: 17th Workshop on Innovative Use of NLP for Building Educational Applications, NAACL 2022.
10. **J. Zhao**, D. Khashabi, T. Khot, A. Sabharwal, K.-W. Chang. Ethical-Advice Taker: Do Language Models Understand Natural Language Interventions? ACL Findings, 2021.
11. C. Zhang, **J. Zhao**, H. Zhang, K.-W. Chang, C.-J. Hsieh. Double Perturbation: On the Robustness of Robustness and Counterfactual Bias Evaluation. NAACL, 2021.
12. **J. Zhao**, and K.-W. Chang. LOGAN: Local Group Bias Detection by Clustering. Conference on Empirical Methods in Natural Language Processing (EMNLP), 2020.
13. **J. Zhao**, S. Mukherjee, S. Hosseini, K.-W. Chang, and A. H. Awadallah. Gender Bias in Multilingual Embeddings and Cross-Lingual Transfer. Association for Computational Linguistics (ACL), 2020.
14. Y. Zhou, J-Y. Jiang, **J. Zhao**, K.-W. Chang, and W. Wang. "The Boating Store Had Its Best Sail Ever": Pronunciation-attentive Contextualized Pun Recognition. Association for Computational Linguistics (ACL), 2020.
15. S. Jia*, T. Meng*, **J. Zhao**, and K.-W. Chang. Mitigating Gender Bias Amplification in Distribution by Posterior Regularization. Association for Computational Linguistics (ACL), 2020.
16. A. Gaut, T. Sun, S. Tang, Y. Huang, J. Qian, M. ElSherief, **J. Zhao**, D. Mirza, E. Belding, K.-W. Chang, and W. Y. Wang. Towards Understanding Gender Bias in Relation Extraction. Association for Computational Linguistics (ACL), 2020.

17. Z. Fu, Y. Xian, R. Gao, **J. Zhao**, Q. Huang, Y. Ge, S. Xu, S. Geng, C. Shah, Y. Zhang, G. de Melo. Fairness-Aware Explainable Recommendation over Knowledge Graphs. Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR), 2020.
18. T. Wang, **J. Zhao**, M. Yatskar, K. Chang, V. Ordonez. Balanced Datasets Are Not Enough: Estimating and Mitigating Gender Bias in Deep Image Representations. International Conference on Computer Vision (ICCV), 2019
19. P. Zhou, W. Shi, **J. Zhao**, K.-H. Huang, M. Chen, K.-W. Chang. Examining Gender Bias in Languages with Grammatical Gender. Conference on Empirical Methods in Natural Language Processing & International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), 2019
20. T. Sun, A. Gaut, S. Tang, Y. Huang, M. ElSherief, **J. Zhao**, M. Diba, B. Elizabeth, K.-W. Chang and W.Y. Wang. Mitigating Gender Bias in Natural Language Processing: Literature Review. Association for Computational Linguistics (ACL), 2019
21. **J. Zhao**, T. Wang, M. Yatskar, R. Cotterell , V. Ordonez, K.-W. Chang. Gender Bias in Contextualized Word Embeddings. North American Chapter of the Association for Computational Linguistics (NAACL), 2019
22. **J. Zhao**, Y. Zhou, Z. Li, W. Wei, K.-W. Chang. Learning Gender Neutral Word Embeddings. Conference on Empirical Methods in Natural Language Processing (EMNLP), 2018
23. **J. Zhao**, T. Wang, M. Yatskar, V. Ordonez, K.-W. Chang. Gender Bias in Coreference Resolution: Evaluation and Debiasing Methods. North American Chapter of the Association for Computational Linguistics (NAACL), 2018
24. **J. Zhao**, T. Wang, M. Yatskar, V. Ordonez, K.-W. Chang. Men Also Like Shopping: Reducing Gender Bias Amplification using Corpus-level Constraints. Conference on Empirical Methods in Natural Language Processing (EMNLP), 2017 **(Best Long Paper Award)**
25. **J. Zhao**, J. Li, B. Zhou, F. Chen, P. Tomchik, W. Ju. Parallel Algorithms for Anomalous Subgraph Detection. Concurrency and Computation: Practice and Experience, 29(3), e3769. 2017
26. J. Li, **J. Zhao**, Y. Li, L. Cui, B. Li, L. Liu, J. Panneerselvam. iMIG: Toward an Adaptive Live Migration Method for KVM Virtual Machines. The Computer Journal, 58(6), 1227-1242. 2015
27. B. Shi, B. Li, L. Cui, **J. Zhao**, J. Li. SyncSnap: Synchronized Live Memory Snapshots of Virtual Machine Networks. IEEE International Conference on High Performance Computing and Communications (HPCC), 2014

TALKS AND PANELS

- Invited talk @CHAI in UChicago. 2023.03
- Trustworthy and Responsible AI: Fairness, Interpretability, Transparency and Their Interactions. AAAI tutorial. 2023.02
- Gender Bias in Natural Language Processing Workshop 2021.08
- Talk @Apple Research. 2020.08
- AKBC KG-BIAS Workshop 2020.06
- Gender Equality and Corporate Social Responsibility, UN-Women China 2019.11
- Grace Hopper Celebration 2018.09
- NLP Highlights Podcast 2018.08
- Mid-Atlantic Student Colloquium on Speech, Language and Learning 2017.05

PROFESSIONAL ACTIVITIES

Organizer:

- Workflow Chair, AAAI 2023
- Women in Machine Learning (WiML), NeurIPS 2021.
- Workshop on Socially Responsible Machine Learning, ICML 2021.
- NLP for Positive Impact Workshop, ACL 2021, EMNLP 2022

Program Committee/Reviewer:

- Area Chair: EMNLP 2023
- NeurIPS 2023
- AACL 2022
- ARR 2021, 2022
- TrustNLP workshop 2021, 2022
- NAACL 2019, 2021, 2022
- ACL 2020, 2021
- AAAI 2020, 2021
- EMNLP 2018, 2021, 2022
- NLPCC-English 2018, 2019, 2020
- NLPCC-Chinese 2018, 2 019
- AKBC 2019

TEACHING EXPERIENCE

USC

- CSCI 699, Ethics in NLP 2023 Fall

Guest Lecture

- CSCI 697, USC 2023.08
- CMSC 396H, UMD 2022.10
- CSCI 544, USC 2022.04
- CS seminar (w/ 200 enrollment, show case of EDI research), UCLA 2021.12
- IST 597, PSU 2021.11
- CSE 290C, UCSC 2021.05
- CS263, UCLA 2021.03
- Ethics and Fairness in AI, UPitt 2021.02

UCLA

- Teaching Assistant, Introduction to Machine Learning, Kai-Wei Chang. 2019 Fall

University of Virginia

- Teaching Assistant, Algorithm, Kong-Cheng Wong. 2017 Spring
- Teaching Assistant, Software Development Methods, Nada Basit and David Edwards. 2017 Spring
- Teaching Assistant, Algorithm, Gabriel Robins. 2016 Fall
- Teaching Assistant, Discrete Math, David Edwards. 2016 Fall