

Working with a diverse set of students and collaborators confirms my belief that everyone has the potential to become a successful researcher, just not necessarily receiving the same resources and mentorship. During my PhD, I have put substantial efforts into promoting diversity and inclusion, and I will continue these efforts as a professor.

Broadening Participation through Inclusive Advising and Democratizing Education Resources. I care deeply about diversity in my advising, having mentored students with diverse demographics, including eight women, 20 students of color, and two LGBTQ+ students, with diverse cultural backgrounds from 12 different origin countries, including Bangladesh, China, Germany, Hungary, India, Italy, Mexico, Russia, Sweden, Switzerland, Taiwan, and the US. As a senior graduate student, I have also mentored two junior PhD students, who are the first-generation PhD students in their family, whom I helped with their first papers at top NLP conferences like ACL, thus embarking on a good start to their PhD.

Alongside diversifying my personal mentorship, I also support open-source educational resources and mentorship insights, so that students with all socioeconomic backgrounds across the world can benefit. For my research-related resources, I provided open access to my Causal NLP tutorial (2,500 views), paper reading lists for [Causal NLP](#) (400+ stars) and [NLP for Social Good](#) (250+ stars), as well as all my research talk videos at top conferences such as [NeurIPS 2022](#), [EMNLP 2022](#), [NAACL 2022](#), [EMNLP 2021 \(1\)](#), [EMNLP 2021 \(2\)](#), [ACL 2021](#), and [ACL 2020](#).

To expand mentorship outreach, I established a [mentorship GitHub](#) with advising suggestions for junior researchers in NLP, which received over 700 stars on GitHub. In the era of large language models (LLMs), I also helped bring together 20 researchers to collect research direction suggestions given the drastic change that LLMs bring to the traditional NLP community [6]. In addition to the written suggestions, I also organize dynamic mentorship sessions as a leader and one of the initiators of the [ACL Year-Round Mentorship](#). Inheriting the same diversity and inclusion spirit, we opened access to all our mentorship videos online, with the [most popular session](#) receiving up to 2,500 views. We receive positive responses from students in our mentorship program such as *“I’d like to express my gratitude for the mentorship program; it has been immensely beneficial to me! The sessions on research topics and the PhD application process have been particularly enlightening.”*

Ensuring Inclusion in Leadership Roles. In addition to broadening the outreach to junior students, I also take leadership roles in various initiatives, and encourage representation diversity. Among the five conference workshops I have co-organized, we ensure that the organizer committee includes 50% or more women. We also ensure balanced gender and demographic representations among the invited speakers, so underrepresented attendees are more likely to find the role models in their own identities. Similarly for the ACL Year-Round Mentorship, we invited speakers from more than 20 different countries, and further diversified the representation in our mentorship sessions by rotating the time zones, diversifying the speakers’ gender and demographics, as well as covering both a diverse set of universities or organizations of the speakers.

Casting Positive Impact through Research. As NLP becomes more and more powerful as a technology, its various applications and social implications also start to flourish. I actively use NLP as a tool to help the diversity, equity, and inclusion (DEI) agenda, where I have papers reporting the large gender difference in the AI publica-

tion world [3], revealing that editors’ opinions are biased towards the gender of authors in certain publication domains [2], and measuring the inherent bias and stereotypes within LLMs [8]. Moreover, in our EMNLP 2022 and ACL 2023 papers [9, 10], we also identify ways to protect the user privacy in LLMs, so that underrepresented groups are shielded from discrimination and exploitation due to their identity when using this technology. I also work towards uniting the community efforts towards socially-minded NLP by leading the three **NLP for Positive Impact Workshops** [1, 4], publishing papers to provide sites of the current progress [5] and future directions [7] of NLP for social good research.

Moving forward, I plan to continue these efforts by incorporating the spirit of DEI throughout my mentorship and advising principles, course design, community organization, and research for positive social impact. Through these concerted efforts, I aspire to create an inclusive academic environment that not only fosters a richly diverse community, but also champions equitable opportunities for learning, growth, and innovation.

References

- [1] Laura Biester, Dorottya Demszky, **Zhijing Jin**, Mrinmaya Sachan, Joel Tetreault, Steven Wilson, Lu Xiao, and Jieyu Zhao. *Proceedings of the Second Workshop on NLP for Positive Impact (NLP4PI)*, Abu Dhabi, United Arab Emirates (Hybrid), December 2022. Association for Computational Linguistics.
- [2] Anna Costello, Ekaterina Fedorova, **Zhijing Jin**, and Rada Mihalcea. Editing a woman’s voice. In *ICSSl*, 2023.
- [3] Yiwen Ding*, Jiarui Liu*, Zhiheng Lyu*, Kun Zhang, Bernhard Schölkopf, **Zhijing Jin**†, and Rada Mihalcea†. Voices of her: Analyzing gender differences in the AI publication world. 2022.
- [4] Anjalie Field, Shrimai Prabhumoye, Maarten Sap, **Zhijing Jin**, Jieyu Zhao, and Chris Brockett. *Proceedings of the 1st Workshop on NLP for Positive Impact*, Online, August 2021. Association for Computational Linguistics.
- [5] Fernando Gonzalez*, **Zhijing Jin***, Jad Beydoun, Bernhard Schölkopf, Tom Hope, Mrinmaya Sachan, and Rada Mihalcea. Beyond good intentions: Reporting the research landscape of NLP for social good. *CoRR*, abs/2305.05471, 2023.
- [6] Oana Ignat*, **Zhijing Jin***, Artem Abzaliev, Laura Biester, Santiago Castro, Naihao Deng, Xinyi Gao, Aylin Gunal, Jacky He, Ashkan Kazemi, Muhammad Khalifa, Namho Koh, Andrew Lee, Siyang Liu, Do June Min, Shinka Mori, Joan Nwatu, Verónica Pérez-Rosas, Siqi Shen, Zekun Wang, Winston Wu, and Rada Mihalcea. A PhD student’s perspective on research in NLP in the era of very large language models. *CoRR*, abs/2305.12544, 2023.
- [7] **Zhijing Jin**, Geeticka Chauhan, Brian Tse, Mrinmaya Sachan, and Rada Mihalcea. How good is NLP? A sober look at NLP tasks through the lens of social impact. In *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*, pages 3099–3113, Online, August 2021. Association for Computational Linguistics.
- [8] Justus Mattern, **Zhijing Jin**, Mrinmaya Sachan, Rada Mihalcea, and Bernhard Schölkopf. Understanding stereotypes in language models: Towards robust measurement and zero-shot debiasing. *CoRR*, abs/2212.10678, 2022.
- [9] Justus Mattern, **Zhijing Jin**, Benjamin Weggenmann, Bernhard Schölkopf, and Mrinmaya Sachan. Differentially private language models for secure data sharing. In *Proceedings*

of the 2022 Conference on Empirical Methods in Natural Language Processing, pages 4860–4873, Abu Dhabi, United Arab Emirates, December 2022. Association for Computational Linguistics.

- [10] Justus Mattern, Fatemehsadat Mireshghallah, **Zhijing Jin**, Bernhard Schölkopf, Mrinmaya Sachan, and Taylor Berg-Kirkpatrick. Membership inference attacks against language models via neighbourhood comparison. In *Findings of the Association for Computational Linguistics: ACL 2023*, Toronto, Canada, July 2023. Association for Computational Linguistics.