

Rebecca Dorn

(301) 832-2668 | rdorn@usc.edu | rebedorn.github.io

I am a Ph.D. candidate at the University of Southern California specializing in generative AI fairness, modeling, and assessment research. I am particularly passionate about evaluating and improving how LLM systems treat dialects of historically marginalized groups. I have 6 peer-reviewed papers, including an *Outstanding Paper Award* at EMNLP 2024.

EDUCATION

- Pursuing Ph.D., Computer Science** Aug. 2021-Present
University of Southern California
Advisors: Kristina Lerman and Fred Morstatter
- B.S., Computer Science** Sep. 2016-Jun. 2020
University of California, Santa Cruz
Advisor: Lise Getoor
Thesis: Bias Exploration in Face Verification Systems

WORK EXPERIENCE

- Applied Science PhD Intern**, Amazon Summer 2025
Computer Science Education Coordinator, USC Stimulating STEM Summer Program Summer 2024
Data Science Intern, Families USA Winter 2021

INVITED TALKS

- Microsoft**, GLEAM Speaker Series Apr. 2025
Mila, FATE Seminar Apr. 2025
USC SWE, PhD's & Platters Mar. 2025
USC Information Science Institute, Natural Language Processing Seminar Dec. 2024
UCLA, Natural Language Processing Fairness Group Nov. 2024
USC ISI, Fairness and Bias Seminar Nov. 2024
USC, Center for AI in Society Mar. 2023
USC, Women in Science and Engineering (WiSE) STEM Bytes Oct. 2022

PUBLICATIONS

- Chu, D., He, Z., **Dorn, R.** & Lerman, K. Improving and Assessing the Fidelity of Large Language Model Alignment to Online Communities. *NAACL 2025*.
- Leitner, M., **Dorn, R.** & Morstatter, F. Characterizing Network Structure of Anti-Trans Actors on TikTok. *ACM WebSci 2025*.
- Ranjit, J., Joshi, B., **Dorn, R.**, Petry, L., Koumoundouros, O., Bottarini, J., Liu, P., Rice, E. & Swayamdipta, S. OATH-Frames: Characterizing Online Attitudes Towards Homelessness via LLM. *ACM EMNLP 2024*. ***Outstanding Paper Award**
- He, Z., **Dorn, R.**, Guo, S., Chu, D. & Lerman, K. COMMUNITY-CROSS-INSTRUCT: Unsupervised Instruction Generation for Aligning Large Language Models to Online Communities. *ACM EMNLP 2024*.
- Dorn, R.**, Kezar, L., Morstatter, F. & Lerman, K. Harmful Speech Detection by Language Models Exhibit Gender-Queer Dialect Bias. *ACM EAAMO 2024*. **(11.5% Acceptance Rate)**
- Sanchez, C., Chu, D., He, Z., **Dorn, R.**, Murray, S. & Lerman, K. Feelings about Bodies: Emotions on Diet and Fitness Forums Reveal Gendered Stereotypes and Body Image Concerns. *Under Review*.
- Dorn, R.**, Mokhberian, N., Jiang, J. Abramson, J., Morstatter, F. & Lerman, K. Non-Binary Gender Expression in Online Interactions. *IEEE ASONAM 2024*.

POSTER PRESENTATIONS

- Dorn, R.**, Chance, C., & Rusti, C. Dialect Bias in Affective Computing: AAVE and Text-Based Emotion Classification. *SoCalNLP Symposium 2024*.
- Lee, E, Baird, A., Young, L. & **Dorn, R.** The Role of Public Facing Organizations in the Transgender Rights Debate: An Issue Management Framework. *Organizational Communication Research Escalator at ICA 2024*.
- Ranjit, J., Joshi, B., **Dorn, R.**, Petry, L., Koumoundouros, O., Bottarini, J., Liu, P., Rice, E. & Swayamdipta, S. OATH-Frames: Characterizing Online Attitudes Towards Homelessness via LLM. *USC ShowCAIS 2024*. ***Best Poster Award**

Dorn, R., Kezar, L., Morstatter, F. & Lerman, K. Harmful Speech Detection by Language Models Exhibit Gender-Queer Dialect Bias. *SoCalNLP 2023*.

Ranjit, J., Joshi, B., **Dorn, R.**, Petry, L., Koumoundouros, O., Bottarini, J., Liu, P., Rice, E. & Swayamdipta, S. OATH-Frames: Characterizing Online Attitudes Towards Homelessness via LLM. *SoCalNLP 2023*.

Dorn, R., Mokhberian, N., Jiang, J. Abramson, J., Morstatter, F. & Lerman, K. Non-Binary Gender Expression in Online Interactions. *CMU SBP-BRiMS 2023*.

TEACHING EXPERIENCE

Graduate Teaching Assistant , Machine Learning for Data Science	Summer 2024, Spring 2025
Graduate Teaching Assistant , Data Structures	Fall 2024, Fall 2023
Graduate Teaching Assistant , Fairness in Artificial Intelligence	Spring 2023
Undergraduate Teaching Assistant , Ethics & Algorithms	Winter 2020, Spring 2020
Undergraduate Peer Tutor	Winter 2019, Spring 2019

HONORS AND AWARDS

EMNLP Outstanding Paper Award, 2024
EAAMO Conference Travel Award, 2024
WiSE Travel Award, 2024
USC ShowCAIS Best Poster Award, 2024
SBP-BRiMS Scholarship, 2023
Inclusive STEM Educator Award, 2019

INVOLVEMENT

Memberships: EAAMO (2024-Present), QueerInAI (2023-Present), Women in Science and Engineering (2021-Present)
External Reviewer: CLARe6, LREC, IJHCI, CSCW, WebSci, JMIR
Organizer: Fairness and Bias ISI Seminar Series (2024-Present)

MEDIA COVERAGE

Zaner, Amanda. "Intersectional Disparities within Automated Hate-speech Detection Across US Centered Social Media Content". Guest Post in *Center for Democracy and Technology*. December 2024.

Lee, Stephanie. "USC ISI at the 2024 EMNLP Conference". Article in *USC Information Science Institute Research Newsletter*. November 2024.

Cohen, Julia. "Flagged for Being Queer". Article in *USC Viterbi School of Engineering Newsletter*. October 2024.

Soetirto, Rania. "AI Solutions for Social Good: Ph.D. Student Enlists LLM Assistants on Project Addressing Homelessness". Article in *USC Viterbi School of Engineering Newsletter*. September 2024.

Russell, Adam (Host). "There is no end goal for AI ethics, there will always be something new to mitigate". In *AI/insiders Podcast*, June 2024.

TECHNICAL SKILLS

Programming	Python, SQL, R, C, C++, Git, Docker, CUDA, HPC
Machine Learning	PyTorch, HuggingFace, TensorFlow, Scikit-Learn, NumPy, Pandas, Matplotlib, Seaborn
Large Language Models	Fine Tuning (RAG, Instruction Tuning), Value Alignment, Prompt Engineering, Dataset Curation