Rebecca Dorn

rdorn@usc.edu +1-301-832-2668 https://rebedorn.github.io/

Research Interests

I am a Ph.D. candidate at the University of Southern California specializing in generative AI fairness, modeling, and assessment research. I have 6 peer-reviewed papers, including an Outstanding Paper Award at EMNLP 2024.

Education

Pursuing Ph.D., Computer Science Aug. 2021 – May 2026
University of Southern California
Advisors: Kristina Lerman and Fred Morstatter

B.S., Computer Science Sep. 2016 – Jun. 2020
University of California, Santa Cruz
Thesis: Bias Exploration in Face Verification Systems, advised by Lise Getoor.

Select Publications and Pre-Prints

Societal Impact of LLMs

Dorn, R., Chance, C., Rusti, C., Bickham Jr., C., Chang, K., Morstatter, F., & Lerman, K. Emotion Detection on African American Vernacular English Minimizes Black Joy. *Under Review.*

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**

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)

Alignment & Safety

Chu, D., He, Z., **Dorn, R.** & Lerman, K. Improving and Assessing the Fidelity of Large Language Model Alignment to Online Communities. *NAACL 2025.*

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.*

Computational Social Science

Leitner, M., **Dorn, R.** & Morstatter, F. Characterizing Network Structure of Anti-Trans Actors on TikTok. *ACM WebSci 2025.*

Dorn, R., Mokhberian, N., Jiang, J. Abramson, J., Morstatter, F. & Lerman, K. Non-Binary Gender Expression in Online Interactions. *IEEE ASONAM 2024.*

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.*

Work Experience

Applied Science Intern, Amazon Summer 2025
Data Science Intern, Families USA Winter 2020

Invited Talks

Amazon, SCINTEX Series Aug. 2025
Microsoft, GLEAM Series Apr. 2025
Mila, FATE Seminar Apr. 2025
USC Information Science Institute, Natural Language Processing Seminar Dec. 2024
UCLA, Natural Language Processing Fairness Group Nov. 2024
USC Information Science Institute, AI Fairness and Bias Seminar Nov. 2024
USC Center for AI Safety, showCAIS Seminar Mar. 2023
USC, Women in Science and Engineering (WiSE) STEM Bytes Oct. 2022

Select 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

Honors & Awards

Finalist, Gabard Award to Advance Understanding of the LGBTQ+ Community	2025
Outstanding Paper Award, EMNLP	2024
Conference Travel Award, EAAMO	2024
Conference Travel Award, WiSE	2024
Best Poster Award, USC ShowCAIS	2024
Conference Travel Award, SBP-BRiMS	2023
Inclusive STEM Educator Award, UC Santa Cruz Learning Support Services	2019

Involvement

Memberships: EAAMO (2024 - Present), QueerinAI (2023 - Present), Women in Science and Engineering (2021 - Present)

Reviewer: AIES, ARR, CSCW, CLARe6, IJHCI, JMIR, LREC, WebSci

Leadership: Seminar Organizer for USC Information Science Institute Fairness & Bias Seminar (2024 – Present), Computer Science Education Coordinator for Stimulating STEM Program (2024)

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.

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/nsiders Podcast*, June 2024.

Technical Skills

Programming: Python, SQL, PySpark, R, C, C++, Git, Docker, CUDA, HPC

Machine Learning: PyTorch, HuggingFace, Scikit-Learn, NumPy, Pandas, Matplotlib, Seaborn

LLM: Fine-Tuning (RAG, Instruction Tuning), Value Alignment, Prompt Engineering, Dataset Curation