Katy Felkner

felkner@isi.edu | http://www.katyfelkner.com | (405) 808-2703

EDUCATION

University of Southern California — Los Angeles, CA
Ph.D. in Computer Science — Aug. 2020 – Present
- Expected May 2027; GPA: 3.88/4.00; Supported by **NSF Graduate Research Fellowship**
- Selected Coursework: Adv. NLP, Fairness in AI, Probability for Data Sci., Representation Learning in NLP

University of Oklahoma — Norman, OK
B.S. Computer Science (summa cum laude), B.A. Letters (summa cum laude); GPA: 4.0/4.0 — Aug. 2016 – May 2020

RESEARCH INTERESTS

Research agenda focuses on developing high-quality, **well-grounded benchmark datasets** for robust evaluation of language models, with particular emphases on community-engaged methods and on rigorous measurement of social bias.

TECHNICAL SKILLS

Programming Languages and Skills: Python, C, C++, R, SQL, Bash Scripting, HPC, Slurm

ML/NLP/Data Science Tools: HuggingFace, PyTorch, TensorFlow, SpaCy, Pandas, Numpy, Scikit-Learn, NLTK

RESEARCH EXPERIENCE

LLM Fairness and Bias, USC Information Sciences Institute — Los Angeles, CA
Graduate Research Assistant — Jan. 2022 – Present
- Develop two community-sourced benchmark datasets for detecting anti-LGBTQ+ and antisemitic bias in LLMs.
- Design and distribute two community surveys and analyze 500 responses to inform LLM bias benchmarks.
- Lead team of 3 to collect 8M sentences of community data, then train 32 debiased models via continued pretraining.
- Assess LLM capability for fairness benchmark construction and conduct detailed human analysis of model failures.
- **Mentoring**: Mentor undergrad on project analyzing behavior of textual entailment (NLI) as an LLM bias metric.
- **Publications**: Lead interdisciplinary teams to publish four papers (two ACL, one workshop, one under review).

Border Translation Project, USC Information Sciences Institute — Los Angeles, CA
Graduate Research Assistant — Aug. 2020 – Dec. 2021
- Built and evaluated NMT systems for extremely low-resource languages.
- Developed NMT system for K'iche', with intended application of providing legal assistance to asylum seekers.

Ultrascale Systems Research Center, Los Alamos National Laboratory — Remote
Post-Baccalaureate Research Intern — May 2020 – Aug. 2020
- Tested 16 combinations of word embedding and clustering algorithms to predict failures from supercomputer syslog.
- **Publication**: Paper accepted to Second Workshop on ML for Computing Systems, Supercomputing 2020.

SELECTED PUBLICATIONS

Felkner, Virginia K., Allison Lim, and Jonathan May (Oct. 2025). *Textual Entailment and Token Probability as Bias Evaluation Metrics*. Under review. arXiv: 2510.07662 [cs.CL]. URL: https://arxiv.org/abs/2510.07662.

Felkner, Virginia K., Jennifer Thompson, and Jonathan May (Aug. 2024). "GPT is Not an Annotator: The Necessity of Human Annotation in Fairness Benchmark Construction". In: *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics*. DOI: 10.18653/v1/2024.acl-long.760.

Felkner, Virginia K., Ho-Chun Herbert Chang, et al. (July 2023). "WinoQueer: A Community-in-the-Loop Benchmark for Anti-LGBTQ+ Bias in Large Language Models". In: *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics*. DOI: 10.18653/v1/2023.acl-long.507.

TEACHING EXPERIENCE

USC CSCI 662: Advanced NLP: *Teaching Assistant* — Aug. 2025 – Dec. 2025
USC CSCI 350: Intro to Operating Systems: *Teaching Assistant* — May 2025 – Aug. 2025

PROFESSIONAL SERVICE

Seminar Organizing: USC ISI Natural Language Seminar — Aug. 2024 – May 2025
Reviewing: EMNLP 2025 Ethics Review, NAACL SRW 2025, EMNLP 2023, ACL 2023, EMNLP 2022 Theme Track