Jillian Fisher

(214)-893-0342 | jrfish@uw.edu | jfisher52.github.io
Seattle, Washington 98122

EDUCATION

University of Washington – Doctorate, Statistics (NLP + HCI) AUGUST 2020 - OCTOBER 2025
Texas A&M University – Master of Science, Statistics JANUARY 2018 - DECEMBER 2019
University of Texas at Austin – Bachelor of Arts, Mathematics and Psychology AUGUST 2011 - MAY 2015

CURRENT RESEARCH

My research focuses on *AI alignment, controllable generation, and the societal impact of AI*. Advised by Yejin Choi (Computer Science & Engineering) and Thomas Richardson (Statistics) at the University of Washington, I take an interdisciplinary approach that integrates diverse methodologies.

PUBLICATIONS

- **Jillian Fisher**, Chan Young Park, and Jennifer Neville. Beyond Similarity for Personalization: User Memory Selection via Response-Utility Optimization. *In review.*
- **Jillian Fisher**, Shangbin Feng, Robert Aron, Thomas Richardson, Yejin Choi, Daniel Fisher, Jennifer Pan, Yulia Tsvetkov, and Katharina Reinecke. Bias AI can Influence Political Decision-Making. *ACL 2025 (Oral).*
- **Jillian Fisher,** Ruth E. Appel, Chan Young Park, Yujin Potter, Liwei Jiang, Taylor Sorensen, Shangbin Feng, Yulia Tsvetkov, Margaret E. Roberts, Jennifer Pan, Dawn Song, and Yejin Choi. Political Neutrality in AI is Impossible- But Here is How to Approximate it. *ICML 2025 (Oral).*
- Abhilasha Ravichander, **Jillian Fisher**, Taylor Sorensen, Ximing Lu, Yuchen Lin, Maria Antoniak, Niloofar Mireshghallah, Chandra Bhagavatula, Yejin Choi. Information-Guided Identification of Training Data Imprint in (Proprietary) Large Language Models. *NAACL 2025.*
- **Jillian Fisher**, Skyler Hallinan, Ximing Lu, Mitchell Gordon, Zaid Harchaoui, Yejin Choi. StyleRemix: Interpretable Authorship Obfuscation via Distillation and Perturbation of Style Elements. *EMNLP 2024 (0ral).*
- Shangbin Feng, Taylor Sorensen, Yuhan Liu, **Jillian Fisher**, Chan Young Park, Yejin Choi, Yulia Tsvetkov. Modular Pluralism: Pluralistic Alignment via Multi-LLM Collaboration. *EMNLP 2024.*
- **Jillian Fisher**, Ximing Lu, Jaehun Jung, Liwei Jiang, Zaid Harchaoui, Yejin Choi. JAMDEC: Unsupervised Authorship Obfuscation using Constrained Decoding over Small Language Models. *NAACL 2024 (Oral).*
- Taylor Sorensen, Jared Moore, **Jillian Fisher**, Mitchell Gordon, Christopher Michael Rytting, Andre Ye, Liwei Jiang, Ximing Lu, Yejin Choi. Position: A Roadmap to Pluralistic Alignment. *ICML 2024.*
- Peter Wes, Ximing Lu, Nouha Dziri, Faeze Brahman, Linjie Li, Jena D. Hwang, Liwei Jiang, **Jillian Fisher**, Abhilasha Ravichander, Khyathi Raghavi Chandu, Benjamin Newman, Pang Wei Koh, Allyson Ettinger, Yejin Choi. THE GENERATIVE AI PARADOX: "What It Can Create, It May Not Understand". *ICLR 2024.*
- **Jillian Fisher**, Lang Liu, Krishna Pillutla, Yejin Choi, Zaid Harchaoui. Statistical and Computational Guarantees for Influence Diagnostics. *AISTAT 2023*.
  **Awarded Honorable Mention* : ASA Statistical Learning and Data Science 2023 Student Paper Award Competition

WORK EXPERIENCE

Microsoft Research, Redmond, WA – Research Science Intern. MAY 2025 – AUGUST 2025

- Designed and led a research project on developing a Bayesian optimal method for memory selection to improve LLM personalization.
- Produced a research paper submitted to ICLR 2026.

Meta, New York, NY – Data Science Intern JUNE 2024 – SEPTEMBER 2024

- Developed a metric to assess the "contextuality" of ads in Instagram's new Multi-ads format.
- Delivered and presented four analyses on metric impact, leading to its integration in the ads pipeline.
- Collaborated on the design and experimental setup for a large (>2K) human-based study to evaluate Multi-ads, contributing to insights on ad contextuality.

Allen Institute for Artificial Intelligence (AI2), Seattle, WA – Research Intern JUNE 2022 – SEPTEMBER 2022

- Led a project which aimed to enhance models' ability to construct more human-aligned advice
- Constructed and conducted 500+ user studies using Amazon Mechanical Turk to analyze human-alignment

Amazon AWS, Seattle, WA – Data Science Intern JUNE 2021 – JUNE 2022

- Doubled accuracy of AWS hiring forecast, improving hiring funnel to accurately direct talent acquisition team
- Coded, optimized, and deployed Python and SQL scripts to correctly integrate cycle time into the hiring forecast
- Conducted analysis on 1.5M data points of demographic diversity and market factors, incorporating multiple content sources (including BLS) to drive 2022 hiring goals

• Python (Pytorch) • Machine Learning • Natural Language Processing • Statistics • Psychology • Human Study Design