Jillian Fisher Curriculum Vitae

(214) 893-0342 · fisherj@stanford.edu · jfisher52.github.io

ACADEMIC APPOINTMENTS

Stanford University – Postdoctoral Scholar, Communication Department October 2026 – Present

University of Washington – Postdoctoral Scholar, Paul G. Allen School of CSE January – October 2026

EDUCATION

University of Washington – Ph.D., Statistics (with CSE) August 2020 – October 2025

Dissertation – *Statistical Methods Towards Trustworthy AI: From Diagnosis to Control and Societal Implications*

Advisors: Yejin Choi (CSE) and Thomas Richardson (Statistics)

Texas A&M University – M.S., Statistics January 2018 – December 2019

University of Texas at Austin – B.A., Mathematics and Psychology August 2011 – May 2015

RESEARCH INTERESTS

AI and the information environment; human–AI interaction and persuasion; AI evaluation methodology and measurement; AI and political communication; pluralistic alignment and personalization.

PUBLICATIONS

Selected Research Publications

- **Jillian Fisher**, Shangbin Feng, Robert Aron, Thomas Richardson, Yejin Choi, Daniel Fisher, Jennifer Pan, Yulia Tsvetkov, Katharina Reinecke. Biased LLMs 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, Yejin Choi. Position: Political Neutrality in AI Is Impossible – But Here Is How to Approximate It. *ICML 2025 (Oral).*
- **Jillian Fisher**, Daniel W. Fisher, Thomas Richardson, Jennifer Pan, Yulia Tsvetkov, Katharina Reinecke. Whose Guardrail? User Identity Shapes the Effectiveness of Model Disclaimers in Mitigating LLM Political Persuasion. *CHI 2027 (under review).*
- Matt Wilde, **Jillian Fisher**, Andrew B. Hall, Kathryn Salam, Robbie Goldfarb. NewsBench: Expert-Grounded Evaluation of Epistemic Quality in AI News Reporting. *Forum AI 2026 (white paper).*
- **Jillian Fisher**, Jennifer Neville, Chan Young Park. Response-Aware User Memory Selection for LLM Personalization. *EMNLP 2026.*

Additional Publications

- Betty Xiong, **Jillian Fisher**, Benjamin Newman, Meng Hu, Shivangi Gupta, Yejin Choi, Lanyan Fang, Russ B. Altman. FDARxBench: Benchmarking Regulatory and Clinical Reasoning on FDA Generic Drug Assessment. *EMNLP 2026.*
- Taylor Sorensen, Benjamin Newman, Jared Moore, Chan Young Park, **Jillian Fisher**, Niloofar Mireshghallah, Liwei Jiang, Yejin Choi. Spectrum Tuning: Post-Training for Distributional Coverage and In-Context Steerability. *ICLR 2026.*

- **Jillian Fisher***, Chan Young Park*, Marius Memmel, Dipika Khullar, A. Yun, Abhishek Gupta, Yejin Choi. Making VLMs More Robot-Friendly: Self-Critical Distillation of Low-Level Procedural Reasoning. *EMNLP 2025.*
- 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 (Oral).*
- 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.*
- Jaehun Jung, Peter West, Liwei Jiang, Faeze Brahman, Ximing Lu, **Jillian Fisher**, Taylor Sorensen, Yejin Choi. Impossible Distillation for Paraphrasing and Summarization: How to Make High-Quality Lemonade out of Small, Low-Quality Models. *NAACL 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, Niloofar Mireshghallah, Christopher Michael Rytting, Andre Ye, Liwei Jiang, Ximing Lu, Nouha Dziri, Tim Althoff, Yejin Choi. Position: A Roadmap to Pluralistic Alignment. *ICML 2024.*
- Peter West, 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.*
- Ximing Lu, Faeze Brahman, Peter West, Jaehun Jung, Khyathi Chandu, Abhilasha Ravichander, Prithviraj Ammanabrolu, Liwei Jiang, Sahana Ramnath, Nouha Dziri, **Jillian Fisher**, Bill Yuchen Lin, Skyler Hallinan, Lianhui Qin, Xiang Ren, Sean Welleck, Yejin Choi. Inference-Time Policy Adapters (IPA): Tailoring Extreme-Scale LMs without Fine-Tuning. *EMNLP 2023.*
- **Jillian Fisher**, Lang Liu, Krishna Pillutla, Yejin Choi, Zaid Harchaoui. Statistical and Computational Guarantees for Influence Diagnostics. *AISTATS 2023*.

  *Equal contribution

MANUSCRIPTS IN PREPARATION AND RESEARCH IN PROGRESS

- Elinor Poole-Dayan, **Jillian Fisher**, Atoosa Kasirzadeh, Jacob Andreas, Mitchell Gordon, Michiel A. Bakker. A Roadmap to Impactful Pluralistic Alignment Research. *2026 (preprint).*

Ongoing Projects

- Search Engine AI Overviews field audit examining demographic disparities in AI-generated search summaries and their implications for information processing.
- Election information audit examining how search engines, social media, and chatbots change the information they provide around the 2026 U.S. election.

AWARDS AND GRANTS

- NSF ACCESS Discover Allocation (Principal Investigator) – approximately $40,000 in CloudBank compute credits for research on political bias in multi-agent LLM systems, 2026.
- Honorable Mention, ASA Statistical Learning and Data Science Student Paper Award, 2023.

MEDIA COVERAGE

- Science News (2025) – "Chatbots spewing facts, and falsehoods, can sway voters".
- NPR / KUOW Soundside (2025) – "Making political decisions with the help of voter guides, pamphlets and... ChatGPT?"
- UW News (2025) – "With just a few messages, biased AI chatbots swayed people's political views" (most-read story of the year).
- GeekWire (2025) – "UW study: Politically persuasive AI chatbots offer potential benefits – and worrying influence".
- KIRO Seattle Morning News (2025) – "AI and Politics".

PRESENTATIONS

Invited Talks

- "Biased, Disclosed, Neutral? How AI Shapes Political Decisions and What We Can Do About It." *AI Ethics Course guest lecture, University of California, Los Angeles, October 2026.*
- "When AI Takes a Side: Measuring and Mitigating the Influence of Political Bias in Language Models." *First Workshop on NLP 4 Democracy, COLM 2025, Montreal, Canada, October 2025.*
- "When AI Takes a Side: Measuring and Mitigating the Influence of Political Bias in Language Models." *Hoover Institution, Stanford University, March 2025.*
- "The Power of Bias: How AI Influences Political Decision-Making and What We Can Do About It." *Cohere For AI, December 2024.*
- "Small but Mighty: Empowering Small Language Models to Outperform Their Larger Counterparts." *Qualcomm, November 2024.*
- "Influence Diagnostics Under Self-Concordance and Application to Natural Language Models." *AI Transfer Lab, September 2023.*

Conference Talks

- "Statistical and Computational Guarantees for Influence Diagnostics." *Joint Statistical Meetings (JSM), Toronto, Canada, August 2023.*
- "Influence Diagnostics Under Self-Concordance." *IMS International Conference on Statistics and Data Science (ICSDS), Florence, Italy, December 2022.*
- "Model Editing in Language Models Using Influence Functions." *Joint Statistical Meetings (JSM), Washington, DC, August 2022.*

TEACHING EXPERIENCE

University of Washington, Seattle, WA **2020 – 2025**

- Teaching Assistant: STAT311 Elements of Statistical Methods; CSSS221 Statistics Concepts and Methods for Social Science; STAT591 Special Topics (Deep Learning).

Concordia University, Austin, TX **Fall 2019**

- Primary Instructor: Math205 Introduction to Statistics (designed and taught course).

University of Texas at Austin, Austin, TX **2015 – 2020**

- Primary Instructor: M408C Differential and Integral Calculus (online); M408D Sequences, Series, and Multivariable Calculus (online).
- Teaching Assistant: M302 Introduction to Mathematics; M408K Differential Calculus; M408L Integral Calculus; M340L Matrices and Matrix Calculations.

INDUSTRY EXPERIENCE

Forum AI, Remote – Research Consultant **January – June 2026**

- Advised on evaluation design for NewsBench, a benchmark of epistemic quality in AI political news.
- Co-authored the NewsBench white paper for an academic audience.

Microsoft Research, Redmond, WA – Research Science Intern **May – August 2025**

- Designed and led a project developing a Bayesian-optimal memory selection method for LLM personalization.
- Produced a first-author paper accepted at EMNLP 2026.

Meta, New York, NY – Data Science Intern **June – 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.
- Co-designed a large (>2K) human study evaluating ad contextuality in the Multi-ads format.

Allen Institute for AI (AI2), Seattle, WA – Research Intern **June – September 2022**

- Led a project that aimed to enhance models' ability to construct more human-aligned advice.
- Created and conducted a 500+ user study using Amazon Mechanical Turk to analyze human alignment.

Amazon AWS, Seattle, WA – Data Science Intern **June 2021 – June 2022**

- Doubled the accuracy of the AWS hiring forecast, improving the hiring funnel to accurately direct the 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.

CONSULTING PUBLICATIONS

- Daniel Fisher, **Jillian Fisher**, S. R. Urfer, N. V. A. Tran, K. F. Kerr, M. Darvas. Higher Burden of Neuropsychiatric Symptom-Like Behaviors Associated with Canine Cognitive Dysfunction Compared to Normal Aging in the Dog Aging Project. *GeroScience, 2025.*
- S. Patel, **Jillian Fisher**, L. S. Chae, A. Soto, B. Kasi, M. Konda, et al. Phase II Basket Trial of Dual Anti-CTLA-4 and Anti-PD-1 Blockade in Rare Tumors (DART) SWOG S1609: Pancreatic Neuroendocrine Neoplasm (PNEN) Cohort. *Journal for ImmunoTherapy of Cancer, 13(6), 2025.*

- S. O. Baird, M. Rinck, D. Rosenfield, M. L. Davis, **Jillian Fisher**, E. S. Becker, M. B. Powers, J. A. J. Smits. Reducing Approach Bias to Achieve Smoking Cessation: A Pilot Randomized Placebo-Controlled Trial. *Cognitive Therapy and Research, 41(4), 2017.*

PROFESSIONAL SERVICE

- Institute for Foundations of Machine Learning (IFML) Workshop, 2023 (lead organizer).
- Conference Reviewer: ICML, ICLR, ACL, EMNLP, NAACL.
- Workshop Reviewer: NLP for Democracies (MultiDems), 2026; Pluralistic Alignment Workshop, 2026.
- Pre-Application Review Service (PARS), UW Statistics, 2023 – 2024 (a free mentorship initiative in which current Ph.D. students review and provide feedback on application materials for prospective graduate applicants).
- Department of Statistics Outreach and Community Committee (SOCC), UW Statistics, 2021 – 2025 (a committee that advises the department on ways to promote an equitable, inclusive, and engaged departmental community).
- Statisticians and Biostatisticians of Underrepresented Genders (SBUG) member, UW Statistics, 2022 – 2025 (a student-run organization for supporting underrepresented genders in these fields).