# Nic Fishman

CAUSAL INFERENCE · MACHINE LEARNING · MULTISCALE MODELING

## Education

**Harvard University** — *Cambridge, MA*
Anticipated PhD in Statistics — *aug 2023 - present*

**Oxford University** — *Oxford, UK*
MSc in Statistics, advised by Prof. Yee Whye Teh — *oct 2021 - jun 2023*

**Stanford University** — *Stanford, CA*
B.S. in Computer Science and B.A. in Sociology with Honors – GPA: 3.9 — *sept 2017 - jun 2021*

## Honors & Awards

**2023** — **James Mill Peirce Fellowship**, Harvard University — *Cambridge, MA*
The James Mill Peirce Fellowship recognizes excellent graduate students in the natural sciences, mathematics, and engineering at the Harvard Graduate School of Arts and Sciences.

**2023** — **Graduate Research Fellowship**, National Science Foundation — *Cambridge, MA*
The NSF GRFP recognizes and supports outstanding graduate students in science, technology, engineering, and mathematics. Awarded for work on machine learning and causal inference.

**2021** — **Rhodes Scholarship**, Rhodes Trust — *Oxford, UK*
The Rhodes Scholarship is a postgraduate award for students to study at the University of Oxford.

**2021** — **Outstanding Senior Thesis Award and Firestone Medal**, Stanford University — *Stanford, CA*
Best thesis award in sociology and medal for top Stanford-wide theses. Received for work on the history of trial avoidance mechanisms and how colonialism and imperialism drove their diffusion.

**2020** — **Spotling Talk**, Computational Biology Workshop, ICML — *Vienna, AT*
Exceptional submissions to the workshop are invited to give a presentation on their work.

**2017** — **G.R.E.A.T. Award**, NHGRI, National Institutes of Health — *Bethesda, MD*
Genome Recognition of Employee Accomplishments and Talents, for work in the Ostrander Lab.

## Publications

### Published Papers

Alexander C Furnas, **Nic Fishman**, Leah Rosenstiel, and Dashun Wang, 2025. Partisan disparities in the funding of science in the United States. *Science*.

**Nic Fishman** and Evan Rosenman, 2024. Estimating Vote Choice in US Elections with Approximate Poisson-Binomial Logistic Regression. *OPT 2024*.

Jake Fawkes, **Nic Fishman**, Mel Andrews, and Zachary Lipton, 2024. The fragility of fairness: Causal sensitivity analysis for fair machine learning. *Advances in Neural Information Processing Systems*.

Hamed Nilforoshan, Wenli Looi, Emma Pierson, Blanca Villanueva, **Nic Fishman**, Yiling Chen, John Sholar, Beth Redbird, David Grusky, and Jure Leskovec, 2023. Human mobility networks reveal increased segregation in large cities. *Nature*.

**Nic Fishman**, Leo Klarner, Valentin De Bortoli, Emile Mathieu, and Michael Hutchinson, 2023. Diffusion models for constrained domains. *Transactions on Machine Learning Research*.

**Nic Fishman**, Leo Klarner, Emile Mathieu, Michael Hutchinson, and Valentin De Bortoli, 2023. Metropolis sampling for constrained diffusion models. *Advances in Neural Information Processing Systems*.

**Nic Fishman** and Nicholas T Davis, 2022. Change we can believe in: Structural and content dynamics within belief networks. *American Journal of Political Science*.

**Nic Fishman** and Leif Hancox-Li, 2022. Should attention be all we need? The epistemic and ethical implications of unification in machine learning. *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency*.

**Fishman**, N., Shrikumar, A., Marinov, G., Kundaje, and A., . Systematic characterization of generative models for de novo design of regulatory DNA. *(Spotlight) International Conference on Machine Learning, Computational Biology Workshop, 2020*.

### Working Papers

**Nic Fishman**, Gokul Gowri, Peng Yin, Jonathan Gootenberg, and Omar Abudayyeh, 2025. Generative Distribution Embeddings. *arXiv preprint arXiv:2505.18150*.

**Nic Fishman**, 2021. Making (Global) Criminal Procedure: Empire, the End of Justice, and the Rise of Efficiency. *Stanford University*.

Ben-Menachem, J., **\* Fishman**, N., \* Galper, and G.\*, . The False Promise of Controlling False Positives: Ephemerality and the "Replication Crisis" in Observational Studies.

**Fishman** and N., . Principled Models for Analyzing Max-Diff Surveys.

# Presentations

| | | |
|---|---|---|
| 2025 | **Generative modeling on the space of empirical measures**, Harvard Statistics Seminar | *Cambridge, MA* |

Outlining the history of the field of fair machine learning and arguing that "theoretical fairness" is a doomed project.

| | | |
|---|---|---|
| 2025 | **False Promises and False Premises of Fair Machine Learning**, Columbia Algorithms and Society Seminar | *New York, NY* |

Outlining the history of the field of fair machine learning and arguing that "theoretical fairness" is a doomed project.

| | | |
|---|---|---|
| 2025 | **Predicting perturbations at single-cell resolution**, Cell Circuits and Epigenomics Seminar | *Cambridge, MA* |

An overview of the problem of single-cell perturbation prediction, presentation of results from benchmarking existing methods, and our novel approach for *in silico* experimentation.

| | | |
|---|---|---|
| 2022 | **Invited talk, Should attention be all we need?**, Stanford Human Centered Artificial Intelligence | *Stanford, CA* |
| 2022 | **Invited talk, Should attention be all we need?**, DeepMind Sociotechnical Systems | *London, England* |
| 2022 | **Conference talk**, Law and Society Association Annual Meeting | *Lisbon, Portugal* |

On 'Making (Global) Criminal Procedure'. Outlining the history of two trial avoidance mechanisms, plea bargaining and penal orders, and how colonialism and imperialism drove their diffusion.

| | | |
|---|---|---|
| 2021 | **Conference talk**, American Political Science Association Annual Meeting | *Seattle, WA* |

Developed sparse-covariance Gaussian Mixture Models for studying political belief networks and the connections to Brandom's inferentialist account of belief formation. With J. Green and N. T. Davis.

# Research Experience

**Harvard Medical School, Harvard University**
*Cambridge, MA*
RESEARCH WITH PROF. MARINKA ZITNIK
*apr 2024 - present*

- Building a causal foundation model synthesizing 280 million patient medical records and medical knowledge graphs integrating information across scales (molecules, proteins, genetics, pathways, and disease information) to predict patient counterfactuals.
- Developing new methods for conditional dynamic treatment effect estimation from a foundation model, with applications to 1) optimizing treatment policies and 2) generating clinically informed drug representations.

**Harvard Medical School, Harvard University**
*Cambridge, MA*
RESEARCH WITH PROFS. JONATHAN GOOTENBERG AND OMAR ABUDAYYEH
*may 2024 - present*

- Developed novel approach for approximate optimal transport and flow matching for predicting effects of genetic perturbations, enabling generalization across unseen cell types.
- Designed and implemented a transformer-based models to develop embeddings for sets of cells to generate fine-grained cell-type embeddings, enabling better generalization across cell types.
- Led extensive benchmarking and validation efforts, systematically evaluating and unifying existing models for out-of-distribution prediction in single-cell sequencing using novel performance metrics.

**Institute for Quantitative Social Science, Harvard University**
*Cambridge, MA*
RESEARCH WITH PROF. KOSUKE IMAI
*june 2020 - sept 2024*

- Constructed a non-parametric framework for causal inference under general interference using regression.
- Developed efficient algorithms and theory for non-parametric monotone regression for estimation and inference in the general interference framework, with applications in large-scale network and spatial experiments.
- Worked on adaptive methods for pure exploration in combinatorial designs in generalized linear models.

**Statistics Department, University of Oxford**
*Oxford, England*
RESEARCH IN OxCSML GROUP UNDER YEE WHYE TEH
*sept 2021 - june 2023*

- Built novel models of cyclic peptides and antibody loops using constrained diffusion and flow-matching models.
- Developed techniques for extending diffusion models to constrained domains in general Riemannian geometries.

**Center for Science of Science & Innovation, Northwestern University**
*Evanston, IL*
RESEARCH FELLOW
*june 2021 - sept 2023*

- Synthesized huge international datasets on scientific grantmaking and publication to generate a picture of the landscape and political implications of scientific funding.
- Leveraged LLMs to develop fine-grained topic classifications across grants and paper abstracts to understand when grant-making institutions lead scientific investigation as opposed to following what is already being studied.

**Economics Department, Massachusetts Institute of Technology**
*Cambridge, MA*
RESEARCH ASSISTANT FOR PROF. VICTOR CHERNOZHUKOV
*june 2021 - june 2022*

- Developed methods for machine learning dynamic treatment effects for improved policy learning in healthcare.
- Specifically extended g-estimation to non-linear blip functions using tools from minimax optimization.

**Stanford Center on Poverty and Inequality and Stanford Network Analysis Project**    *Stanford, CA*
RESEARCH INTERN FOR PROF. JURE LESKOVEC AND PROF. DAVID GRUSKY    *jan 2020 - june 2021*
- Used path crossings to study socioeconomic stratification in the United States, developing insights about heterogeneity in segregation across different social contexts and across different income deciles.
- Worked on algorithms to identify when individuals cross paths in time and space using massive GPS data.

**Kundaje Lab, Stanford AI Lab**    *Stanford, CA*
RESEARCH INTERN FOR PROF. ANSHUL KUNDAJE    *jan 2018 - june 2021*
- Created modular system for generating DNA sequences using GANs/VAEs/Transformers and optimizing these models to produce sequences with specified properties (level of gene expression or chromatin accessibility).
- Developed a k-NN algorithm to evaluate (1) the fidelity of samples from generative models and (2) the robustness of neural network predictions on regression outputs (extending existing work on classification).

# Work Experience

**Knowledge Graph Team, Google**    *San Fransisco, CA*
RESEARCH INTERN    *summer 2022*
- Research integrating LLMs and knowledge graphs (KGs) for search, particularly combining LLM representations with graph neural networks over a knowledge graph to produce KG-informed representations.
- Developed methods to identify author expertise using LLM/KG representations for multimedia topic modeling.

**Data for Progress**    *New York, NY*
DATA SCIENTIST    *june 2018 - may 2021*
- Led development of polling infrastructure that produced the most accurate poll results in the Democratic Primary. Built out MySQL database for storage and analysis of survey responses. Automated chart and report generation from this database. Developed search engine and website for internal use to assist in research.
- Designed, conducted, and analyzed polls used to guide policy change for the Green New Deal, Medicare for All, several HR-1 issues, and criminal justice reform, among other progressive issues.
- Developed novel poll weighting scheme achieving state-of-the-art accuracy.
- Automated argument detection so that non-technical colleagues to easily interpret open-ended survey responses in policy briefs by building hierarchical Dirichlet process models for non-parametric topic modeling.
- Created an ecological loss function and corresponding neural networks, extending ecological inference.
- Led team creating word2vec models to analyze gender/racial bias in news articles around the 2016 election.

**Sunrise Movement**    *Washington, DC*
DATA SCIENCE CONSULTANT    *fall 2020*
- Used LASSO and random forest algorithms to build interpretable heuristics for identifying low-propensity swing state voters for a large-scale get-out-to-vote campaign in advance of the 2020 presidential election.

**Star Lab Corporation**    *Washington, DC*
MACHINE LEARNING SPECIALIST    *summer 2018*
- Worked on a Red Hat kernel module to log system activity and leveraged it for neural net anomaly detection.