# Nic **Fishman**
STATISTICS · MACHINE LEARNING · SOCIOLOGY

## **Edu**cation

**Rhodes Scholar, Oxford University**                              *Oxford, UK*
Anticipated MSc in Statistics, MSc in History of Science, Medicine and Technology    *oct 2021 - jun 2023*

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

## **Wor**k Experience

**Pritzker School of Law, Northwestern University**                  *Chicago, IL*
Research Assistant for Prof. Josh Kleinfeld                         *fall 2020*
- Used network models to explain plea bargaining's diffusion and its impacts on global criminal justice systems.
- Conducted research on various questions of legal/intellectual history.

**Sunrise Movement**                                            *Chicago, IL*
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.

**Institute for Quantitative Social Science, Harvard University**        *Boston, MA*
Research Intern for Prof. Kosuke Imai                            *summer 2020*
- Constructed a non-parametric framework for causal inference with continuous treatments, focused on distance-based treatments.
- Leveraged kernel symmetries to build efficient algorithms for enforcing permutation invariances in Gaussian processes, with applications in ecological inference, spillover estimation, and multiple treatments settings.
- Developed Variational Gaussian Processes approach to estimating monotonic treatment functions.

**Stanford Center on Poverty and Inequality and Stanford Network Analysis Project**    *Stanford, CA*
Interdisciplinary Research Intern for Prof. Jure Leskovec and Prof. David Grusky    *since jan 2020*
- 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.

**English Department, Stanford University**                        *Stanford, CA*
Research assistant for Prof. Mark Grief                           *apr - dec 2019*
- Conducted literature reviews on topics including social theory, psychology, and gender and sexuality studies.

**Kundaje Lab, Stanford AI Lab**                                 *Stanford, CA*
Research Intern for Prof. Anshul Kundaje                          *since jan 2018*
- Created modular system for generating DNA sequence using GANs/VAEs/Transformers and optimizing these models to produce sequences with specified properties, e.x. 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).
- Constructed neural net for predicting protein expression from whole genome ChIP-exo data. Used causal inference methods to extract and build a synthetic DNA regulatory logic for simulations in generative genomics.
- Contributed to SimDNA, a Python library for simulating DNA datasets to evaluate genetic machine learning methods in simulations studies.
- Built a novel method for generating regulatory DNA to achieve targeted levels of protein expression through augmenting conditional GAN architectures. (CS229 project)

**Data for Progress**                                           *New York, NY*
Data Scientist                                                 *since jun 2018*
- 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.

**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.

**Ostrander Lab, National Human Genome Research Institute**
Research Fellow

*Bethesda, MD*
*dec 2015 - sept 2017*

- Responsible for building a DNA database to facilitate access and analysis of structural variants.
- Used random forest approach to identify candidate cancer risk genes from SNP arrays.

## Honors & Awards

| | | |
|---|---|---|
| 2020 | **Rhodes Scholar**, Rhodes Trust | *Oxford, UK* |

The Rhodes Scholarship is an international postgraduate award for students to study at the University of Oxford. One of 32 US scholars elected in 2020.

| | | |
|---|---|---|
| 2019 | **Lunsford Finalist**, Program in Writing and Rhetoric | *Stanford, CA* |

The Lunsford Oral Presentation Award recognizes outstanding oral presentations by second-year students. Presented on the computerization of the IRS and its impact on auditing and inequality.

| | | |
|---|---|---|
| 2018 | **Best Project**, CS229 | *Stanford, CA* |

Best project of 400 completed for Stanford's graduate-level machine learning course. Outlined a novel architecture for conditional sampling from GANs.

| | | |
|---|---|---|
| 2018 | **Boothe Prize Finalist**, Program in Writing and Rhetoric | *Stanford, CA* |

The Boothe Prize recognizes outstanding expository and argumentative writing by first-year students. Paper was on the "right to privacy" in the context of digital surveillance and advertising.

| | | |
|---|---|---|
| 2018 | **Boothe Prize Finalist**, Thinking Matters | *Stanford, CA* |

See above. Paper was on the ethics of allocation of Spinraza to SMA patients in hospitals.

| | | |
|---|---|---|
| 2017 | **Winner**, CS109 Competition | *Stanford, CA* |

Won the competition in the introductory statistics course for computer science for developing a framework for simulating how people would evacuate a given room from the blueprint.

| | | |
|---|---|---|
| 2017 | **G.R.E.A.T. Award**, NHGRI, National Institutes of Health | *Bethesda, MD* |

The Genome Recognition of Employee Accomplishments and Talents (G.R.E.A.T) Award, given for work in the Ostrander Lab on the DOGSV system for storing and analyzing structural variants.

## Presentations

| | | |
|---|---|---|
| 2020 | **Spotlight talk**, International Conference on Machine Learning, Computational Biology Workshop | *Vienna, Austria* |

Introduced a taxonomy of DNA sequence generation algorithms and applications to regulatory sequence. Presented a framework for evaluating the quality of generated sequences.

| | | |
|---|---|---|
| 2016 | **Poster presentation**, Ostrander Lab Quadrennial Review, National Institutes of Health | *Bethesda, MD* |

Presented DOGSV: A Relational Database for the Storage and Analysis of Structural Variants.

## Publications

### Published Papers

Fishman, N. and Davis, N. T. 'Change We Can Believe In: Structural and Content Dynamics within Belief Networks', *American Journal of Political Science*, 2020.

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

Fishman, N. 'Fleeing from Terror: Considering Safety When Designing Public Spaces in the Age of Mass Murder', *Stanford Journal of Public Health* 7(1), 2018.

### Working Papers

(In preparation for submission to Nature) Looi, W.,* Pierson, E.,* Redbird, B.,* Villanueva, B., Fishman, N., Chen, Y., Kleinberg, J., Grusky, D., Lescovec, J. 'The Segregation of Interaction', 2020.

(Under Review at ICLR) Fishman, N. and McAuliffe, C. 'Deep Ecological Inference', 2020.

(Honors Thesis) Fishman, N.. 'Imperialism and Plea Bargaining: How Efficiency Killed Justice in Criminal Procedure'.

Fishman, N.. 'The Political Economy of the Computerizing IRS'.

Shrikumar, A.,* Fishman, N.,* Kundaje, A. 'SimDNA: a library for simulating regulatory genomic sequences'.

## Extracurricular Activity

| | | | |
|---|---|---|---|
| Democratic Socialists of America | *since sept 2019* | Stanford Academic Accessibility Directory | *nov 2019* |
| Queer Dungeons and Dragons | *apr 2019 - march 2020* | JStreet University | *jan 2018 - jun 2019* |
| Biomedical Engineering Society | *sept 2017 - jun 2019* | Stanford Debate Society | *sept 2017 - jun 2018* |