

## STATEMENT OF PURPOSE

I bring statistics and machine learning together with critical perspectives from social science to consider when, how, and why data and modeling succeed in their aims—and when, how, and why they can fail. I am passionate about improving practice towards more responsible, robust, effective, and just uses of data and modeling, as well as engaging in outreach to help practitioners in policy, public health, medicine, law, government, journalism, social science, industry, civil society, and elsewhere understand and adopt machine learning and data science.

## EDUCATION

### School of Computer Science, Carnegie Mellon University, Aug 2013–Aug 2018

PhD in Societal Computing (Institute for Software Research) and MS in Machine Learning (Machine Learning Department). Research on social media and sensor data. ARCS Foundation award. Dissertation: “Bias and beyond in digital trace data.” Committee: Jürgen Pfeffer and Anind K. Dey (Coadvisors), Cosma R. Shalizi (Department of Statistics), and David Lazer (Northeastern University).

### Oxford Internet Institute, University of Oxford, Oct 2011–Sep 2012

MSc with distinction in Social Science of the Internet. Master’s thesis on locating the emergence of Internet studies in the 20-year evolution of a large co-authorship network. Advisor: Eric T. Meyer.

### Department of the History of Science, Harvard University, Sep 2004–Mar 2009

AB cum laude in History and Science with Music minor. Senior thesis on narratives of the early 20th century South Indian mathematician Srinivasa Ramanujan. Advisor: Lukas Rieppel.

## EXPERIENCE

### Director of Data Science, Avant Garde Health, Boston, MA, Nov 2020–present

Leading a small team of data scientists at a healthcare data startup, born out of Harvard Business School’s value-based healthcare research and serving about a dozen client hospitals. Research into value-based healthcare; predictive modeling to support bundling payments for Medicare and Medicaid patients; integrating analysis techniques into the product platform.

### Data Analyst / Scientist, MoveOn, Remote, Aug 2020–Nov 2020

Analytics support for progressive organizing efforts around the US 2020 General Election. Tracking disinformation and digital voter suppression targeting minority voters; counter-messaging campaign analytics; designing and implementing voter mobilization experiments.

### Data Science Postdoctoral Fellow, Berkman Klein Center for Internet & Society at Harvard University, Cambridge, MA, Sep 2018–Aug 2020

Statistical learning and network modeling for data in Media Cloud and other projects; technical advising and tutoring for scholars and practitioners in social science, journalism, law, and policy; building in-house data science capacity; forging stronger university-wide ties between social and data science; and research into conceptual, ethical, theoretical, and practical challenges of deploying data science.


### Data Science for Social Good Fellow, Lisbon, Portugal, Summer 2017

Project for Tuscan agencies applying machine learning and network analysis to urban data for sustainable tourism in Florence. Run by the Center for Data Science and Public Policy, University of Chicago, and Nova School of Business and Economics, Universidade NOVA de Lisboa.

## SELECTED PUBLICATIONS

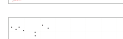
 Nicole C. Nelson, Kelsey Ichikawa, Julie Chung, and **Momin M. Malik**. 2020 (in submission). Mapping the discursive dimensions of the reproducibility crisis: A mixed methods analysis. [MetaArXiv:10.31222/osf.io/sbv3q](https://arxiv.org/abs/10.31222/osf.io/sbv3q).


 Eugene T. Richardson, **Momin M. Malik**, William A. Darity, Jr., A. Kirsten Mullen, Michelle E. Morse, Maya Malik, Aletha Maybank, Mary T. Bassett, Paul E. Farmer, Lee Worden, and James Holland Jones. 2021. Reparations for American descendants of persons enslaved in the U.S. and their potential impact on SARS-CoV-2 transmission. *Social Science & Medicine*.

 **Momin M. Malik**, Afsaneh Doryab, Mike A. Merrill, and Anind K. Dey. 2020 (in submission). Can co-location data detect friendship? It depends how you model it.

 **Momin M. Malik**. 2020 (in submission). A hierarchy of limitations in machine learning. [arXiv: 2002.05193](https://arxiv.org/abs/2002.05193).

 **Momin M. Malik** and Jürgen Pfeffer. 2016. Identifying platform effects in social media data. In *Proceedings of the Tenth International AAAI Conference on Web and Social Media (ICWSM-16)*, 241-249.

 **Momin M. Malik** and Jürgen Pfeffer. 2016. A macroscopic analysis of news content on Twitter. *Digital Journalism* 4 (8): 955-979.

 **Momin M. Malik**, Hemank Lamba, Constantine Nakos, and Jürgen Pfeffer. 2015. Population bias in geotagged tweets. In *Papers from the 2015 ICWSM Workshop on Standards and Practices in Large-Scale Social Media Research (ICWSM-15 SPSM)*, 18-27.

## SKILLS

