

JULIUS ADEBAYO

juliusad@mit.edu

EDUCATION

PhD Student, Electrical Engineering and Computer Science

Massachusetts Institute of Technology,

Research Focus: developing tools and methods to assess and verify properties relating to fairness (or bias), robustness, & privacy of machine learning systems.

SM, Electrical Engineering and Computer Science

SM, Technology and Policy

Massachusetts Institute of Technology, 2013 - 2016.

THESIS: *FairML: Toolbox for Investigating Bias in Predictive Models.*

Bachelor of Science, Mechanical Engineering

Minors: Computer Science & Mathematics

Brigham Young University, 2012.

RESEARCH/WORK EXPERIENCE

Internet Policy Research Initiative (CSAIL, MIT)

September 2018 - Present

(Research Assistant/PhD Student)

- Developing tools and methods to assess and verify properties relating to fairness (or bias), robustness, & privacy of machine learning systems.

Google Brain.

July 2017 - July 2018

(AI Resident)

- Research on improving interpretability, security, and privacy of deep neural networks (DNNs). Published work, at NeurIPS 2018, showing that several commonly used methods for interpreting the predictions of DNNs are invariant to network weight randomizations (See publication section).
- Implemented/extended methods for training DNNs with differential privacy.

Fast Forward Labs Inc.

September 2016 - June 2017

(Research Engineer)

- Wrote and open-sourced python Cuckoo filter data structure for efficient streaming data processing.
- Continued work on FairML, a python package for auditing machine learning models for bias.
- Developed (w/ team) an interpretability API for understanding ensemble and complicated ML methods.

Decentralized Information Group (CSAIL)

August 2014 - January 2016

MIT

(Research Assistant)

- Masters thesis focused on developing, FairML, a system for reverse engineering black-box predictive models to detect bias/discrimination in access to employment, housing, and banking services.
- Worked on maintaining a mobile development platform (in Java), which enables humanitarian organizations and practitioners to easily develop and deploy mobile applications.
- Research here resulted in 4 refereed publications & 3 conference presentations.

Apple Inc.

June 2016 - September 2016

(Data Science Intern)

January 2015 - February 2015

- Wrote and deployed a time series anomaly detection package for monitoring internal metrics across all devices. Anomaly detection methodology was an ensemble of three different methods: generalized extreme studentized deviates test, thresholded median absolute deviation, and Robust PCA.
- Selected as one of the best intern projects, and presented the to SVP of Software engineering.

New England Complex Systems Institute

August 2012 - July 2013

Cambridge, MA

(Research Analyst)

- Built a python application for automatically collecting and clustering Twitter data.
- Collected Twitter data in order to assess theories relating to contagion and spread of information in social networks, which resulted in publication in the journal on complexity.

Data Science Fellow at Data Science for Social Good

June 2014 - August 2014

Harris School of Public Policy, University of Chicago

(Fellow)

- Worked (in a team of 4) with the office of president of Mexico to identify key factors that can be implemented to reduce maternal mortality on a municipality level in Mexico.
- Used various classification and regression techniques to help model maternal mortality at the municipality level leading to identification of key factors: prenatal care, cesarean section, and health insurance.
- The output of this project resulted in a large scale randomized control trial by the Mexican government to assess the influence of these factors on maternal mortality.

IDeA Labs, CS Department

August 2010 - August 2012

Brigham Young University, Provo, UT

(Research Assistant)

- Worked on applying techniques from control theory to synthetic biology. In particular, worked on understanding and extending the use of the method of dynamical structure functions for inferring the pas kinase system.
- A paper detailing the results of this research was published in IEEE CDC December 2011 (See publication section).
- Awarded BYU Orca grant (2010, 2011) worth \$3000 to undertake research.

Papers

- **J. Adebayo**, J. Gilmer, I. Goodfellow, M. Muelly, M. Hardt, B. Kim. Sanity Checks for Saliency Maps. Spotlight Paper at Advances in Neural Information Processing Systems (NIPS), 2018.
- **J. Adebayo**, J. Gilmer, I. Goodfellow, M. Hardt, B. Kim. Local Explanation Methods for Deep Neural Networks are not Sensitive to Parameter Values. ICLR Workshop, April 2018.
- S. Tan, **J. Adebayo**, K. Inkpen, E. Kamar. Investigating Human + Machine Complementarity for Recidivism Predictions. Arxiv Preprint, 2018.
- P. Kindermans, S. Hooker, **J. Adebayo**, M. Alber, K.T. Schutt, S. Dahne, D. Erhan, B. Kim. The (Un)reliability of saliency methods. NIPS Interpretability Workshop, December 2017.
- **J. Adebayo**, L. Kagal. Iterative Orthogonal Feature Projection for Diagnosing Bias in Black-box Models. FATML Workshop, 2016.
- Hurley, M., and **J. Adebayo**. "CREDIT SCORING IN THE ERA OF BIG DATA." Yale Journal of Law & Technology. 18 (2016): 148-275. <http://yjolt.org/credit-scoring-era-big-data>
- **J. Adebayo**, L. Kagal. A Privacy Protection Procedure for Large Scale Individual Level Data. Proceedings of the Conference Intelligence and Security Informatics(ISI) IEEE, May 2015.
- A. Jain, **J. Adebayo**, E. D. Leon, W. Li, L. Kagal, P. Meier, and C. Castillo. Mobile Application Development for Crises Data. Proceedings of Humanitarian Technology: Science, Systems and Global Impact (HumTech), 2015.
- W. Li, **J. Adebayo**, F. Shih, L. Kagal. The Role of Mobile Application in Humanitarian Relief. 12th International Conference on Information Systems for Crisis Response and Management, 2015.
- **J. Adebayo**, T. Southwick, V. Chetty, E. Yeung, Y. Yuan, J. Goncalves, J. Grose, G.B. Stan, S. Warnick. Dynamical Structure Function Identifiability Conditions Enabling Signal Structure Reconstruction. Proceedings of the Conference on Decision and Control IEEE, December 2011.
- **J. Adebayo**, Tiziana Musso, Kawandeeep Virdee, Casey Friedman, Yaneer Bar-Yam, An Exploration of Social Identity: The Structure of the BBC News-Sharing Community on Twitter. Complexity 19 (5) pp.55-63 (2014)