

What's Fair?

Cynthia Dwork
Microsoft Research / Harvard University
cdwork@gmail.com

ABSTRACT

Data, algorithms, and systems have biases embedded within them reflecting designers' explicit and implicit choices, historical biases, and societal priorities. They form, literally and inexorably, a codification of values. "Unfairness" of algorithms – for tasks ranging from advertising to recidivism prediction – has attracted considerable attention in the popular press. The talk will discuss the nascent mathematically rigorous study of fairness in classification and scoring.

Author Keywords

Algorithmic bias; data bias; system bias; fairness; classification fairness; scoring fairness; data mining.

BIOGRAPHY

Cynthia Dwork, Distinguished Scientist at Microsoft Research, is renowned for placing privacy-preserving data analysis on a mathematically rigorous foundation. A cornerstone of this work is differential privacy, a strong privacy guarantee frequently permitting highly accurate data analysis. Dwork has also made seminal contributions in cryptography and distributed computing. She is a member of the US National Academy of Sciences and the US National Academy of Engineering, and is a Fellow of the American Academy of Arts and Sciences and the American Philosophical Society. Beginning January, 2017, Dwork will be the Gordon McKay Professor of Computer Science at the Harvard Paulson School of Engineering, the Radcliffe Alumnae Professor at the Radcliffe Institute for Advanced Study, and Professor by Affiliation at Harvard Law School.

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

KDD'17, August 13–17, 2017, Halifax, NS, Canada.

© 2017 Copyright is held by the owner/author(s).

ACM ISBN 978-1-4503-4887-4/17/08.

DOI: <http://dx.doi.org/10.1145/3097983.3105807>