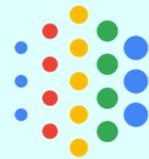




# Timnit Gebru



Google AI

Friday, October 26, 2018

3:00 PM

Luddy Hall, Rm 1106

## Understanding the limitations of AI: When Algorithms Fail

**Abstract:** Automated decision making tools are currently used in high stakes scenarios. From natural language processing tools used to automatically determine one's suitability for a job, to health diagnostic systems trained to determine a patient's outcome, machine learning models are used to make decisions that can have serious consequences on people's lives. In spite of the consequential nature of these use cases, vendors of such models are not required to perform specific tests showing the suitability of their models for a given task. Nor are they required to provide documentation describing the characteristics of their models, or disclose the results of algorithmic audits to ensure that certain groups are not unfairly treated. I will show some examples to examine the dire consequences of basing decisions entirely on machine learning based systems, and discuss recent work on auditing and exposing the gender and skin tone bias found in commercial gender classification systems. I will end with the concept of an AI datasheet to standardize information for datasets and pre-trained models, in order to push the field as a whole towards transparency and accountability.

**Biography:** Timnit Gebru is currently a research scientist in the Ethical AI team at Google, and just finished her postdoc at Microsoft Research, New York City in the FATE (Fairness Transparency Accountability and Ethics in AI) group, where she studied algorithmic bias and the ethical implications underlying any data mining project. She received her Ph.D. from the Stanford Artificial Intelligence Laboratory, studying computer vision under Fei-Fei Li. Her thesis pertains to data mining large-scale publicly available images to gain sociological insight and working on computer vision problems that arise as a result. The Economist and others have recently covered part of this work. Some of the computer vision areas she is interested in include fine-grained image recognition, scalable annotation of images, and domain adaptation. Prior to joining Fei-Fei's lab, Gebru worked at Apple designing circuits and signal processing algorithms for various Apple products including the first iPad. She also spent an obligatory year as an entrepreneur (as all Stanford undergrads seem to do). Her research was supported by the NSF foundation GRFP fellowship and the Stanford DARE fellowship.

