

Monday, December 3, 2018 2:56 PM

#NeurIPS 2018

- Societal Implication
- Technology is neutral — humans are not neutral
- AI: we are handing off decision-making to a computer
 - how decisions affect users? (if decisions are wrong)
 - Are certain groups of people disadvantaged or treated differently
 - Do people understand how decisions are made?
 - Is technology used to avoid accountability
 - Do users have access to remedy?
- Humans fail too — remedy

FATX

Fairness — discrimination — biased data

Accountability — remedy, who is responsible if errors are

Transparency — How do we know how decisions are made?

Ethics

what applications should we consider deepFakes

Baised Data = (prejudicial Bias)

2015 - Gorillas

2017 - AI sentences

2018 - Amazon hiring

Criminal Face Detector

PRO PUBLICA

FEMALE
Nurse

lazy
beautiful

MALE
Doctor

Entrepreneur
hard working
strong

Why?

- ① Data generated by humans — innocent
- ② Pattern matching is lazy

How to prevent:

- ⑥ Know the provenance

- ① Restricted data - Database - blank out columns
- ② De-bias our data
- ③ Rules that override the AI
- ④ Red team

Liability: who is responsible in the eye of the law

- ① AI - No
- ② Engineer
- ③
- ④