



AI Accountability

DR. JOSEPH WILLIAMS
SEATTLE DIRECTOR
PACIFIC NORTHWEST
NATIONAL LABORATORY

PNNL-SA-159866

JULY 2021

What is AI accountability?

A framework for examining discriminatory and inequitable outcomes through the application of artificial intelligence. At its heart is **algorithmic accountability**.

The outcomes can arise from problematic...

- Data
- Models
- Applications

Issues related to algorithmic accountability

Bias

Bias that results in unfairness

Transparency

Standards

Trust

Two parts in algorithmic accountability

Bias

How and why artificial intelligence tends to be biased



Accountability

What scientists and programmers are trying to do to manage and correct for bias



Examples of Data and model bias

Data Bias

- Apple credit cards (NY Times, 11/10/19)
- How might this get used?
 - Credit limits
- Couples filing jointly applied
- Men were approved for credit limits up to 20 times larger than women when they filed income taxes together
- Cause: Historical gender bias in the data with respect to participation by women in the workforce

Model Bias

- Predicting who survived the Titanic Disaster
 - Women & Children first
 - First class versus 3rd class
- Predict whether Molly Brown survived
 - Financial status bias
 - Gender bias
- How might this get used?
 - Insurance premiums

Application bias

- COMPAS (Correctional Offender Management Profiling for Alternative Sanctions)
 - Used to score the risk of recidivism for defendants in pretrial hearings
 - Built by a private firm, validated by Florida State University researchers
 - COMPAS was found in 2016 to systematically mistakenly score white defendants as low-risk, black defendants as high-risk
 - Think about it this way: suppose Whites were +5 underscored on risk, Blacks -5 overscored on risk
 - Model builders claimed no statistical bias as the model was, on average, correct on overall recidivism predictions
- On January 9th, 2020, a black man (Robert Williams) was arrested for theft in Detroit on the basis of facial recognition alone despite well-known weaknesses in the model as applied to black and brown populations. He was subsequently released after police verified his alibi.

Facial recognition, Data, Bias, AI

- What you may not know:
 - At last count, there are 641 million images in the FBI's facial recognition unit (GAO + DHS, 6/14/19)
 - Black and Asian people are 100 times more likely to be misidentified by facial recognition than whites (NIST, 12/17/20)
 - More than 600 law enforcement agencies use Clearview AI (NY Times, 1/31/21)
- Clearview AI says it is working hard to fix misidentification of black and brown people.
- What is the moral conundrum here if Clearview AI gets to racial parity in facial recognition?

Accountability

- **Fix the data**

- Remove bias in the data
 - Hard to do – many AI models rely on humans to train them
 - Many organizations use Mechanical Turk to label images / data (not an unbiased group, tend to be Indo-European males)
 - Then use crowdsourcing to identify biased outcomes
- Ultimately, people are biased. What does it mean to have unbiased data?

- **Fix the models**

- Use of statistics to better validate model integrity
- Construct and use some kind of “fairness” metric
 - The technology is there for this
 - However, people are sharply divided on what is “fair”
- Human in the loop systems

Fix the application: The Algorithmic accountability act of 2019

- Introduced in Congress by Senators Booker and Wyden and Representative Clark
- Would have directed the FTC to develop regulations requiring large firms to conduct impact assessments for “high-risk automated decision systems”
- Review and analyses highlighted how challenging it will be to regulate algorithmic accountability
- Was attacked from all sides on practicality, market realities, technology realities, and questions regarding the government’s overall competence to be able to enforce its provisions
- Did not provide for a private right of action
- Legislation didn’t go anywhere. Also failed in New Jersey
- Updated version has just been introduced in California (AB 13 – stuck in committee) and in Washington (SB 5116 – died in committee)



OPEN DISCUSSION

Algorithmic Accountability

Dr. Joseph Williams
joseph.williams@pnnl.gov