

ALGORITHMS & BIAS RISK



2018 SRA Annual Meeting

December 2018, New Orleans, LA

YVONNE STEVENS, LL.M., LL.B. | yshens@asu.edu
Arizona State University - Sandra Day O'Connor College of Law

Outline

Defining algorithms

Defining algorithmic bias

Examples

Bias oversight

Resolving bias: progress & challenges

Defining Algorithms

- Algorithm: a set of instructions (going back hundreds of years) for analyzing data, performing tasks & solving problems.
- Today's digital algorithms have infiltrated most, if not all, industries → legal risk everywhere (will keep lawyers busy!).
- Algorithms are deployed for speed, low-cost, efficiency and, ideally, accuracy.

Defining Algorithmic Bias

- Algorithmic bias most typically occurs when human values and static historical data are applied to the development of algorithms → impact their problem-solving capabilities.
- Bias may target race, gender, ethnicity, social status, geographical location and so forth → affects privacy, health, safety and security of those sectors (among other repercussions).
- “Myth of neutrality and objectivity in algorithms”.
- Automation bias: human habit to rely on decisions made by automated systems based on above noted myth across industries (military, healthcare, education, law, etc.) → moderate to serious implications.

Getty Images: “woman” search 2007 vs 2017



Examples – 2018 Reveals

- June 2018 reveal that ICE algorithm was modified to produce only one result: detention of 100% immigrants in custody.
- October 2018 reveal that “Amazon’s machine learning system for resume scanning shown to discriminate against women, even downranking CVs simply for containing the word ‘women.’”
- July 2018 reveal that Amazon’s new facial recognition service was incorrectly identifying 28 members of congress as criminals with racial implications.
- February 2018 reveal that MIT “found that three of the latest gender-recognition AIs could correctly identify a person’s gender from a photograph 99 per cent of the time – but only for white men. For dark-skinned women, accuracy dropped to just 35 per cent.”

Source: AI Now Institute - <https://ainowinstitute.org/> (AI Now Institute 2018 Report & October 16, 2018 Symposium); and New Scientist <https://www.newscientist.com/article/2166207-discriminating-algorithms-5-times-ai-showed-prejudice/>

Amazon Rekognition **FALSE MATCHES**



28 current members of Congress

Bias Oversight: Players

- **Legislation** (New York City’s “Automated decision systems used by agencies” law / task force; EU GDPR).
- **Common law** (case law – to be discussed on next slide).
- **Public policy** (*RAND Corporation* <https://www.rand.org/blog/2018/08/keeping-artificial-intelligence-accountable-to-humans.html>).
- **Standards** (*IEEE Standards Association* – P7003 Algorithmic Bias Considerations <https://standards.ieee.org/project/7003.html>).
- **AI industry self-policing** *IBM, Facebook, Microsoft* “bias busting”.
- **Internal corporate policy** (*Deloitte* - <https://www2.deloitte.com/content/dam/Deloitte/us/Documents/center-for-board-effectiveness/us-cbe-nov-board-oversight-algorithmic-risk.pdf>).
- **Media** (*ProPublica* COMPAS investigation - www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing).
- **Think-tanks** (*Fairness Accountability and Transparency in Machine Learning* – fatml.org)
- **Universities** (*AI Now Institute* NYU – “Algorithmic Impact Assessment Framework - ainowinstitute.org; [ASU Law’s CLSI!](#)).
- **Bias Consultants** (Cathy O’Neil – mathematician, author (“Weapons of Math Destruction”) & entrepreneur).

Recent case law, due process and other challenges

- “COMPAS”: algorithm used to assess recidivism risk among criminal offenders. WI Supreme Court conclusion: **“proprietary nature of COMPAS prevents disclosure of how risk is calculated”**. [State v. Loomis, 881 N.W. 2d 749 (WI 2016); WI Supreme Court; On certification from C.A.; petition for certiorari denied by U.S. Supreme Court]. (OUTLIER)
- “Budget Tool”: algorithm used to assess budgets for developmentally disabled persons. U.S. District Court conclusion: **“patients and the public have a right to transparency of the algorithmic process”**. [K.W. v. Armstrong, 180 F. Supp. 3d 703 (D. Idaho 2016); on remand from U.S. C.A. Ninth Circuit].
- “EVAAS”: algorithm used to make decisions about teacher performance & outcomes. U.S. District Court conclusion: **“teachers have no meaningful way to ensure correct calculation of their EVAAS scores, and as a result are unfairly subject to mistaken deprivation of constitutionally protected property interests in their jobs”**. [Houston Federation of Teachers Local 2415, et al. v. Houston Independent School District, 251 F.Supp.3d 1168 (D. Texas 2017)].
- “Resource Utilization Groups System”: algorithm used to make decisions on hours of home-care eligibility. AK Supreme Court conclusion: **“beneficiaries have suffered and will continue to suffer irreparable harm”** without an order(TRO) barring DHS from applying its new reassessment system. [Arkansas Department of Human Services v. Bradley Ledgerwood, et al., 530 S.W.3d 336 (AK 2017)].

Current Legislation



Examples: hr5, sres9, "health care"



MORE OPTIONS

[Home](#) > [Legislation](#) > [115th Congress](#) > H.R.4625[Print](#) [Subscribe](#) [Share/Save](#) [Give Feedback](#)[BACK TO RESULTS](#)

2 OF 3 RESULTS



H.R.4625 - FUTURE of Artificial Intelligence Act of 2017

115t

(H) Matters relating to machine learning **bias** through core cultural and societal norms.

(I) Matters relating to how artificial intelligence can serve or enhance opportunities in rural communities.

(J) Government efficiency, including matters relating to how to promote cost saving and streamline operations.

(2) STUDY.—The Advisory Committee shall study and assess the following:

(A) How to create a climate for public and private sector investment and innovation in artificial intelligence.

(B) The possible benefits and effects that the development of artificial intelligence may have on the economy, workforce, and competitiveness of the United States.

(C) Whether and how networked, automated, artificial intelligence applications and robotic devices will displace or create jobs and how any job related gains relating to artificial intelligence can be maximized.

(D) How **bias** can be identified and eliminated in the development of artificial intelligence and in the **algorithms** that support them, including with respect to the following:

(C) to promote and support the **unbiased** development and application of artificial intelligence; and

Su

(Text)

Resolving Bias: progress

- Awareness.
- Recognition of transparency & fairness (re: fairness → predictive parity, equal false-positive error rates, and equal false-negative error rates involve a *whole* other presentation).
- Public trust increase as a result of oversight.
- Accountability (through oversight).
- Grandfathering.
- Others?

Resolving Bias: challenges

- Privacy (re: transparency → can't have it both ways).
- Security (re: transparency → e.g. cyber).
- Industry push-back (disclosure & bottom line \$).
- Potential conflicts between oversight models/needs and IP protections /lack of.
- Insufficient bias studies due to industry push-back.
- Algorithm complexity from input to output (e.g. understanding).
- Financially burdensome (e.g. government / private resources).
- “Anchoring” (too much reliance on digital systems even when inconsistent).
- Monitoring efforts.
- Others?

COURT ROOM



OF COURSE YOU'LL GET A LONGER SENTENCE THAN HIM, ITS YOUR THIRD CONVICTION.



Shoiki