Making Algorithms Less Biased

MADDY WESTERGAARD

This presentation

1. How algorithms are biased and how to combat bias
2. How algorithms are used in the government
3. The toolkit developed by GovX
4. Other changes being made in the industry

How are algorithms biased?

- Propagate the bias of the data
- Can't detect the bias themselves
- People are inherently biased, which is passed on to the algorithms they are producing

Some of the worst cases of Algorithm Bias

- Sexist Recruiting Tactics
- Microsoft AI ChatBot "TayTweets"
- Gender bias in AI
- Racist Predictions in Criminal Justice

the bias of Northpointe's Algorithm

	WHITE	AFRICAN AMERICAN
Labeled Higher Risk, But Didn't Re-Offend	23.5%	44.9%
Labeled Lower Risk, Yet Did Re-Offend	47.7%	28.0%

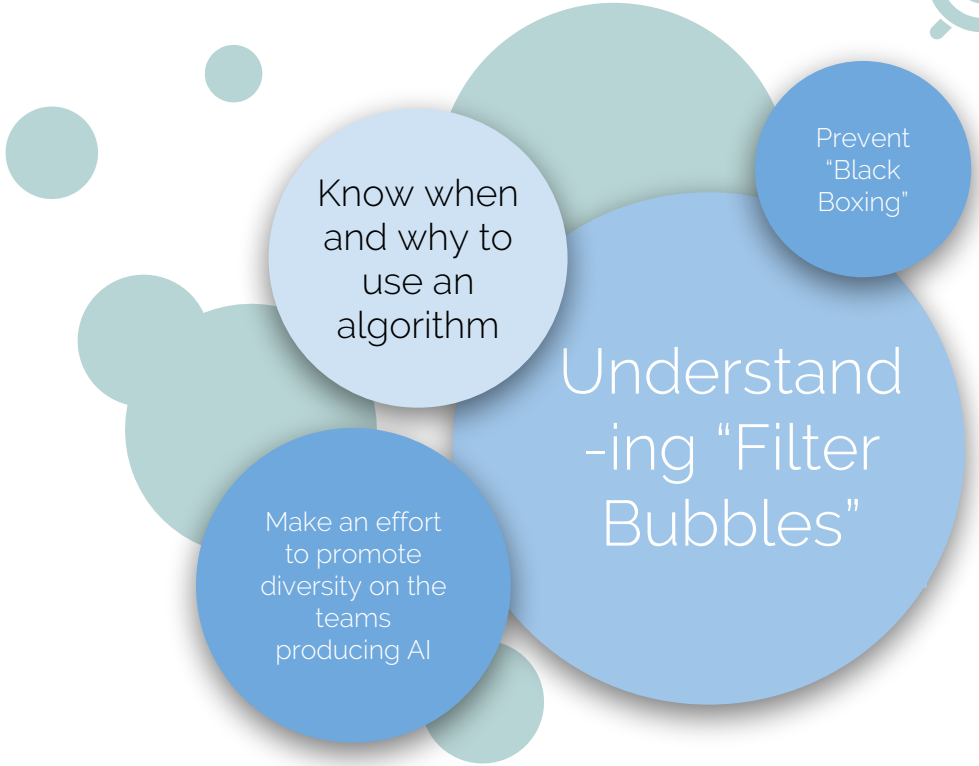
Overall, Northpointe's assessment tool correctly predicts recidivism 61 percent of the time. But blacks are almost twice as likely as whites to be labeled a higher risk but not actually re-offend. It makes the opposite mistake among whites: They are much more likely than blacks to be labeled lower risk but go on to commit other crimes. (Source: ProPublica analysis of data from Broward County, Fla.)

- Every convicted criminal is given a risk score
- Intended for use in combating crime, yet used in prosecution

“

The data scientist knows that in machine learning the answers can be useful only if we ask the right questions.

how to combat algorithm bias



Where are algorithms used in the Government?

1. Segmenting and targeting residents for individual-level results

2. Optimizing civic processes

3. Personal qualification and performance optimization

4. Improving healthcare and public health.

5. Improving science and research

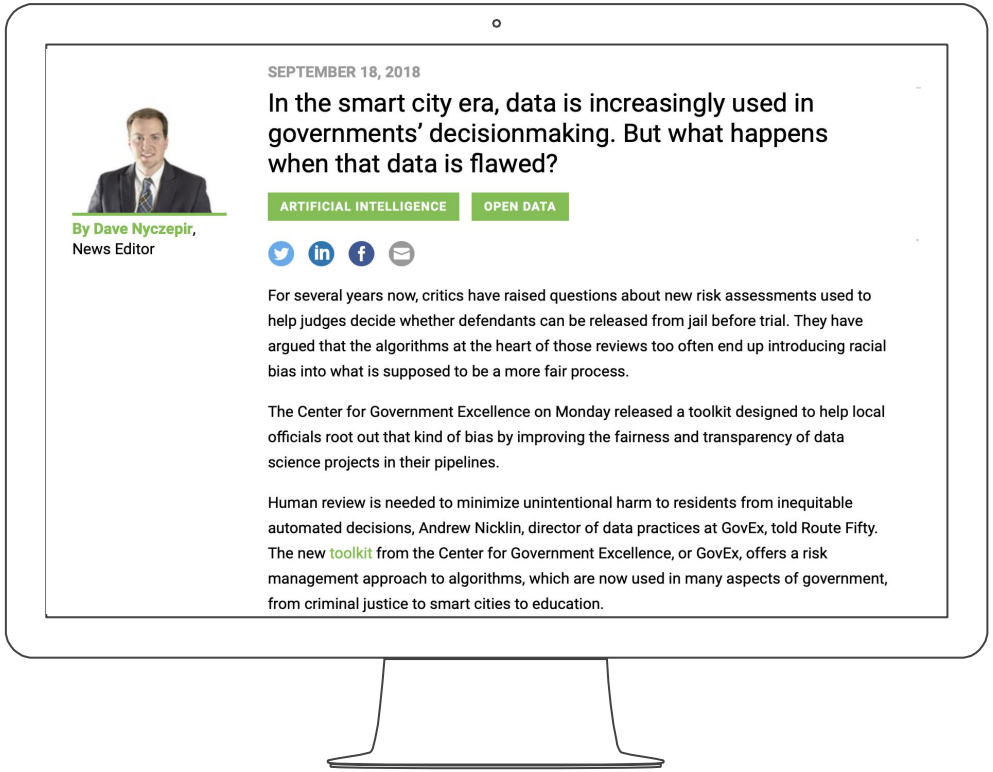
6. Optimizing machine performance

7. Improving security and law enforcement

The Article

Written by Dave Nyczepir

- A toolkit has been developed to mitigate bias with government used AI algorithms



Part 1: Assess Algorithm Risk

1. **Impact -** type, degree, scale, direction
2. **Appropriate Use** – consistency, compatibility, risk score
3. **Accountability** – role of humans in employing the algorithm
4. **Bias** – underlying influence of the data

Part 2: Manage Algorithm Risk

1. Risk to Mitigation Matching worksheet
2. List of 20 Mitigations to aid in steps needed to manage risk

Mitigation 11. Evaluate human-intervened decisions periodically to control for unintended rater bias.

Mitigation 13. Shifting the monitoring of an algorithm to an internal or additional third party adds one more level of subjectivity to an algorithm's methodology.

Other Changes being made

- New York passed the “Automated decision systems used by agencies” bill
- AI Now Institute
  - understanding AI’s implications for society
- New field called Explainable AI

Bibliography

https://www.nextgov.com/analytics-data/2018/09/making-algorithms-less-biased/151350/

https://ainowinstitute.org/about.html

https://www.mckinsey.com/business-functions/risk/our-insights/controlling-machine-learning-algorithms-and-their-biases

https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

https://www.fastcompany.com/40536485/now-is-the-time-to-act-to-stop-bias-in-ai

http://www.govtech.com/policy/Could-New-York-Citys-AI-Transparency-Bill-Be-a-Model-for-the-Country.html

https://www.theverge.com/2016/3/24/11297050/tay-microsoft-chatbot-racist

https://www.forbes.com/sites/theyec/2018/06/27/how-can-we-eliminate-bias-in-our-algorithms/#1e031bcd337e

Thanks!

Any questions?