Making Algorithms Accountable

Dr. Nakeema Stefflbauer

Digital Government Conference

October 22, 2019 - Helsinki

Making Algorithms Accountable: Critical Factors

1- Explainability

2- Privacy-preserving techniques

3- Ethical training

What is needed:

- a framework for tracking problems
- methods to reverse erroneous decisions
- Decision-makers' understanding of potential risks + impact of automation *pre- implementation*.

Product Failure due to **lack of explainability**

Facial recognition wrongly identifies public as potential criminals 96% of time, figures reveal

14-year-old black schoolboy among those wrongly fingerprinted after being misidentified

[United Kingdom]

Eight trials carried in London between 2016 and 2018 resulted in a 96 per cent rate of “false positives” – where software wrongly alerts police that a person passing through the scanning area matches a photo on the database.

Product Failure due to **lack of privacy preservation**

Smart Cities Are Creating a Mass Surveillance Nightmare

+ Hacking Risks

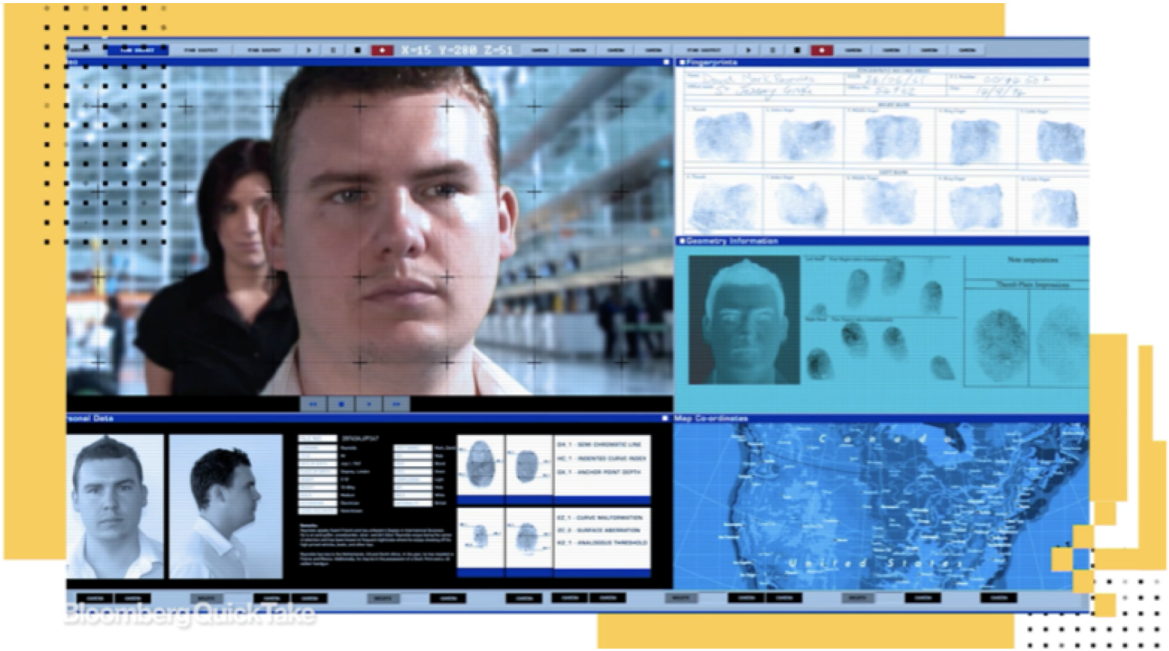
[United States]

Baltimore + Atlanta government functions ground to a halt when ransomware was used successfully

Residents lost access to online bill payments, property deed transfers and court scheduling. **In Baltimore, the city was out of action for weeks & crucial data was permanently lost.**

(Likely) Failure due to **lack of privacy preservation**

France Set to Roll Out Nationwide Facial Recognition ID Program



[FRANCE]

The country's data regulator says the program breaches the European rule of consent & a privacy group is challenging it in France's highest administrative court.

It took a hacker just over one hour to break into a "secure" government messaging app this year, raising concerns about the state's security standards.

Product Failure due to **lack of ethical training**

[United Kingdom]

The British government has knowingly rolled out facial recognition software for passport photos which does not recognise darker skin-colour faces.

How might **an accountable AI system look?**

1. **Run impact tests** before public algorithm rollout, possibly to a (Food-and-Drug-Administration) FDA-type board to assess the risk of violation of existing laws, whether civil rights, human rights, or privacy laws.
2. **Maintain a civil society register for public algorithms** that contains anonymized facts about the raw training data, the algorithms that analyze it, and the decision-making models that emerge.
3. **Make privacy the default for use of personally identifiable data** such that there are very clear guidelines about how you can use that data - and how you cannot.