Machine Learning System Design Challenges

Ashis Kumer Biswas, Ph.D.

Potential Challenges in ML System Design

- AI Systems are only as good as the data we put into them.
- Bad data can contain implicit racial, gender biases.
- A crucial principle, for both humans and machines, is to avoid bias and therefore prevent discrimination.

An analogy

- AI has been used for good purposes and bad purposes.
- No matter how much of a good intend to put behind using something for good, it's going to be used and misconstrued in some way to be used for bad.
- "Fire".
  - It's great for cooking food, getting warmth, and all.
  - But it's also notorious for burning people's houses down.
  - At the same time, you can use it to chase away the people who were burning people's houses.

Fairness in Machine Learning

TheUpshot

ROBO RECRUITING

Can an Algorithm Hire

B

By (

June

Hir

be

computers lack, like making conversation and reading social cues.

But people have biases and predilections. They make hiring decisions, often unconsciously, based on similarities that have nothing to do with the job requirements — like whether an applicant has a friend in common, went to the same school or likes the same sports.

That is one reason researchers say traditional job searches are broken. The question is how to make them better.

A new wave of start-ups — including Gild, Entelo, Textio, Doxa and GapJumpers — is trying various ways to automate hiring. They say

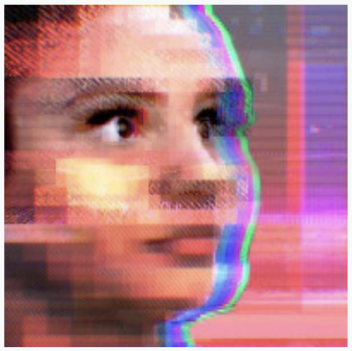
- (2014) Amazon's hiring tool, based on AI, was biased towards hiring men[1]
  - The company decided to scrap the project early 2017.

Am I being watched 24/7?

- The services are gaining popularity in catching the porch pirates
- Recent partnership with the service providers and the Law enforcement also raised concerns on the Privacy rights and civil liberties.
- What if you put camera pointing outside your property line?
- What if you lose control over the camera(s) and recordings?

Microsoft's TAY: Thinking About You

Tay



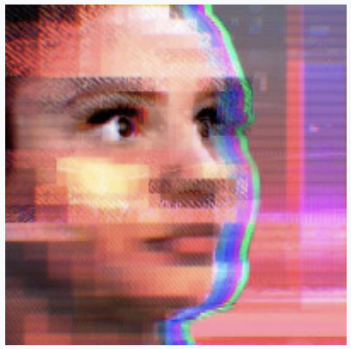
The Twitter profile picture of Tay

Developer(s)	Microsoft Research, Bing
Available in	English
Type	Artificial intelligence chatterbot
License	Proprietary
Website	tay.ai

- Started tweeting on March 23, 2016.
- Tay was designed to mimic the language patterns of a 19-year-old American girl, and to learn from interacting with human users of Twitter.
- It started to follow the Godwin's law: "As an online discussion grows longer, the probability of a comparison involving Nazis or Hitler approaches to 1".

Microsoft's TAY: Thinking About You

Tay



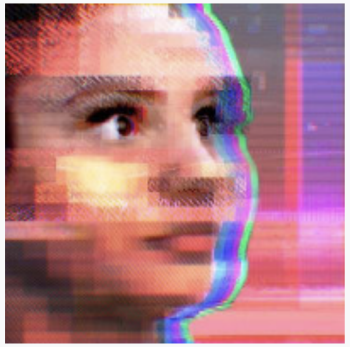
The Twitter profile picture of Tay

Developer(s)	Microsoft Research, Bing
Available in	English
Type	Artificial intelligence chatterbot
License	Proprietary
Website	tay.ai

- Some users on Twitter began tweeting politically incorrect phrases, teaching it inflammatory messages revolving around common themes on the internet, such as "redpilling", GamerGate, and "cuckservatism". As a result, the robot began releasing racist and sexually-charged messages in response to other Twitter users.

Microsoft's TAY: Thinking About You

Tay



The Twitter profile picture of Tay

Developer(s)	Microsoft Research, Bing
Available in	English
Type	Artificial intelligence chatterbot
License	Proprietary
Website	tay.ai

- It caused subsequent controversy with a huge number of inflammatory and offensive tweets through its Twitter account.
- It was taken down by Microsoft just 16 hours after its launch.

Blame game

- Should we blame the technology?
  - Absolutely not!
  - Ethics is not a technological problem; it is pretty much human problem.

How to teach AI ethics?

- General Data Protection Regulation (GDPR) rules that you can't allow a computer to make a decision that you can't explain as a human.
- To avoid bias, can you just remove that data – for instance, gender and race data?
  - Yes. However, how do you get that data to look fair is also a challenge.

Transparency in Algorithms

- The creators of an algorithm may not know why the algorithm is making a certain type of prediction.
- Companies won't allow their algorithms to be publicly scrutinized.
- Example: few algorithms have been used to fire teachers, without being able to give them an explanation of why the model indicated they should be fired [2].
- Let's work hard for **XAI** = Explainable Artificial Intelligence.

[2] https://www.bloomberg.com/opinion/articles/2017-05-15/don-t-grade-teachers-with-a-bad-algorithm

Few (un)fairness examples

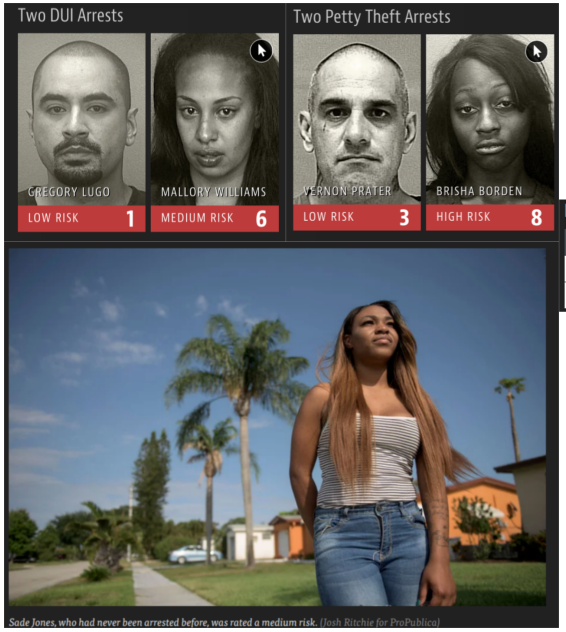
- Not all doctors or CEOs are men. Not all nurses or receptionists are women.
- But you may think otherwise if you search these professions in Google images. (*A study by Univ of Washington, 2015*)

Who will make the final decision?

- Human or AI?

Machine Bias – another level

- There's software, COMPAS, used across the country to predict future criminals (recidivism prediction). And it's biased against blacks. (ProPublica.org, 2016)
- COMPAS, Correctional Offender Management Profiling for Alternative Sanctions (NorthPointe, now Equivant)



Prediction Fails Differently for Black Defendants

	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%

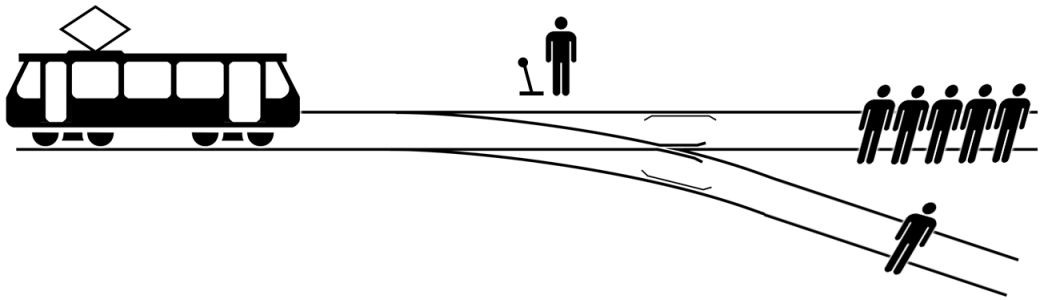
Fake news and Fake videos

- "Great power comes with great responsibility" – Peter Parker (Spiderman by Stan Lee)
- "False news were 70% more likely to be retweeted than real news" [3].
  - Impacts in political campaigns
- If we know that videos can be faked, what will we be acceptable as evidence in a courtroom?

[3] Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. Science, 359(6380), 1146-1151.

The Trolley Problem (by Philippa Foot, 1967) : Should we let AI kill us?

Should you pull the lever to divert the runaway trolley onto the side track?



- Problems analogous to the Trolley problem arise in the design of software controls in autonomous vehicles.
- Car owners should determine their car's ethical values, such as favoring safety of the owner or the owner's family over the safety of others.

Introducing Morality into machines

- Choices are –
  - Utilitarianism – the doctrine that actions are right if they are useful for a majority
  - Rule-based doctrine – e.g., Don’t tell a lie.
  - Nurture based strategy
    - Watch and learn vs. Act and learn
    - Apprenticeship based learning vs. Reinforcement learning