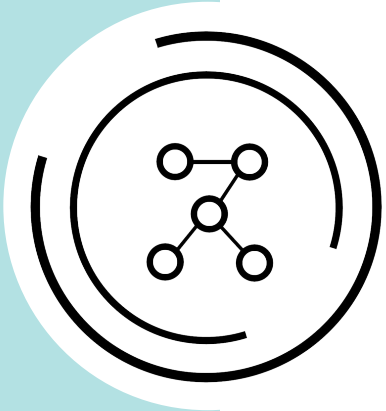


Data Science Society

Data Ethics

Fall 2020 - Bootcamp

By:
Pooja Eega & Kevin Miao

First, go ahead and do the test on the following page:

https://www.moralmachine.net/

Breakout rooms: **2-3 minutes**

Let’s discuss:

Moral Machine.

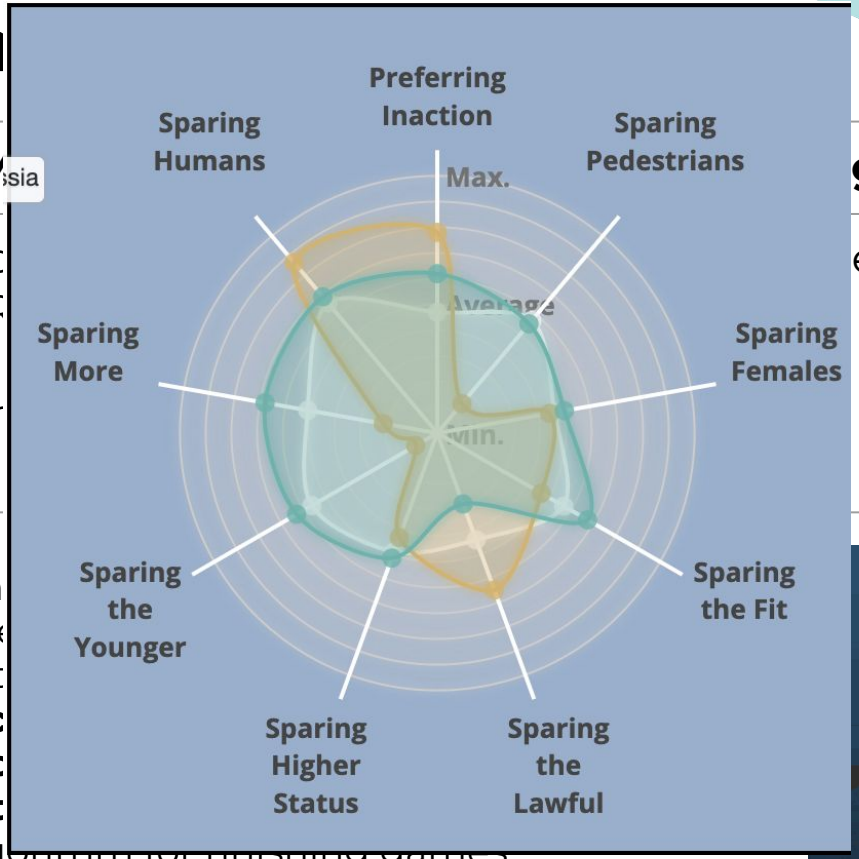
Moral Machine
Discussion

- Who is responsible for harm caused by AI?
- What is the role of lawmakers in regulating emerging technologies?
- How do we decide what is fair?
- **In what ways does human decision-making in the context of rational agents (i.e. AV, Algorithms) influence society?**
  - **Today:** Bias/Algorithmic Inequality

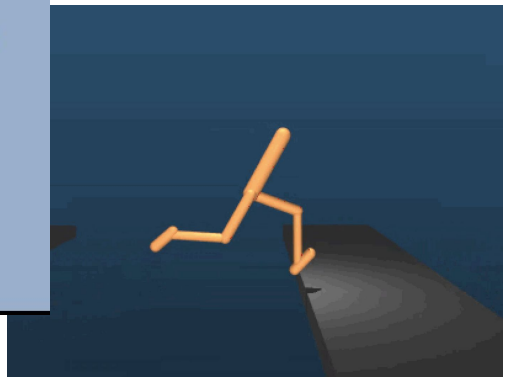
To understa

Human Decision M

- Context Depenc
- (Rarely) Purely F
- Performative
- Culturally deper
- Biased



- **In context of Mora**
  save old people du
  western countries t
  - **North Americ**
  - **South Americ**
- **Out of the context**
  - Computer Alg



Are rational agents **objective**?

Bias

Definition:

> “a tendency to believe that some people, ideas, etc., are better than others that usually results in treating some people unfairly”

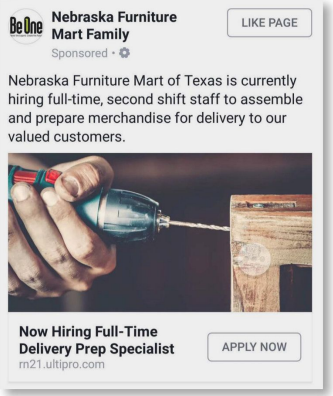
- Development of Algorithms depend on humans
  - **Data can contain bias**
  - **Choices in programming can contain bias**

Algorithmic Inequality

- Difference between **algorithmic inequality** and **human bias**
  - **Reach of algorithms** is extremely **large**
    - **As a consequence, important to assess**

Examples:

- **Black Americans** and **precision healthcare**
- **Job listing** and **women**
- **Cases: Hiring Algorithms, COMPAS**



How to fix **bias?**

Bias is part of being human

- Understanding where during the process there is room for **bias**
  - Data Collection
  - Feature Engineering
  - Team (Diversity of thought)

- Always understand the **complete problem statement** and **context**
- **Transparency** allows for catching errors
- **Evaluate**, **discuss** and be **critical**

Important cases:

1. **hiring algorithms.**
2. **predicting recidivism.**

“AI Systems are only as good as the data they’re trained on and the humans that build them.”

- Adina Sterling, Organizational Behavior Professor at Stanford

Let’s consider:
hiring algorithms.

Types of Hiring Algorithms

- **Sourcing Algorithms**
  - “Configured to search the web for ideal candidates for any given job”
  - Company lists skills they are looking for → the algorithm then looks for candidates with these skills
  - LinkedIn, Ziprecruiter, etc

- **Filtering Algorithms**
  - Filtering the candidates who applied to a job (example: resume screening)
  - Will look at different factors (experience, education, age, etc) to filter candidates for a role

The Problem With Sourcing Algorithms

- **Harvard Business Review Study**
  - In broadly targeted Facebook ads
    - Facebook supermarket cashier positions were shown to an audience of 85% women
    - Jobs with taxi companies went to an audience that was 75% black
  - Even when factors like race of sex are not introduced to the model, it may still find proxies
    - Example: After an audit of an algorithm developed by a resume screening company, they found that the algorithm found two factors to be most indicative of job performance: being named Jared and playing lacrosse

Hiring Algorithms

Filtering Algorithms

- Filtering the candidates who applied to a job (example: resume screening)
- Will look at different factors (experience, education, age, etc) to filter candidates for a role

The Problem with Filtering Algorithms

The Imitation Effect:

- HR department inputs current employee information **so algorithm can construct ideal-candidate profiles**
- As a result: algorithm will favor applications who are closer to what it considers the **“best profiles”**
- This bias continues to get reinforced by the model
- This has an **impact on diversity of candidates**
    - Those with different experiences or backgrounds are at a disadvantage

The Problem with Filtering Algorithms

Amazon:

- In 2014, Amazon started looking into whether AI could help filter candidates for certain job positions
- They found that the algorithm was **only approving resumes that were submitted by men**
- After further investigation, they found that the model was being fed data from the past 10 years, a period when men were dominating the tech industry

The Problem with Filtering Algorithms

HireVue:

- Sells a job interview video platform that claims they can use AI to assess candidates and predict their likelihood to succeed in a position
- The algorithm analyzes how individual candidates answered preselected questions in a recorded video interview, **grading their verbal response and facial movements/gestures**

The Problem with Filtering Algorithms

HireVue:

- Electronic Privacy Information Center (EPIC), a privacy rights nonprofit, filed a complaint with the Federal Trade Commission, pushing the agency to investigate the company for potential bias
- They found that the system **could discriminate against certain groups of people**
- This lawsuit follows the introduction of the Algorithmic Accountability Act in Congress
    - granting the FTC authority to create regulations to check “automated decision systems” for bias

Let’s consider:
predicting recidivism

Background - COMPAS

COMPAS: Correctional Offender Management Profiling for Alternative Sanctions

- Developed by a company called Equivant, which claims **it can predict a defendant's risk of committing another crime**
- Algorithm considers the answers to a 137 item questionnaire
  - https://www.documentcloud.org/documents/2702103-Sample-Risk-Assessment-COMPAS-CORE.html
- Algorithm generates scores on a scale of 1-10 that predicts "Risk of Recidivism" and "Risk of Violent Recidivism"
- Courts take into consideration COMPAS metric when determining criminal sentences

Is the COMPAS algorithm accurate?

Propublica worked to determine **whether the algorithm correctly predicted recidivism**. They found that:

- When controlling for prior crimes, future recidivism, age, and gender, black defendants were **45% more likely** to be assigned higher risk scores than white defenders
- Black defendants were **twice as likely** as white defendants to be misclassified as being a higher risk of violent recidivism
- White violent recidivists were **63% more likely** to have been misclassified as a low risk of violent recidivism

CONCLUSION

- We all have bias; it is just part of being human
- We need to be aware of our bias and how it influences our models
- AI Systems and Machine Learning models can reinforce existing human bias
    - This can have a profound societal impact, as we have seen with hiring algorithms and COMPAS

As you continue your projects this semester, remember to be aware of how human bias may influence the models you create

Classes to take: DATA104, LEGALST123, INFO188, BIOE100, CS195, ISF100J

Additional Information

- **COMPAS**
  - Propublica Analysis Github: https://github.com/propublica/compas-analysis
  - https://www.washingtonpost.com/news/wonk/wp/2016/08/18/why-a-computer-program-that-judges-rely-on-around-the-country-was-accused-of-racism/
  - https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm
  - https://www.theatlantic.com/technology/archive/2018/01/equivant-compas-algorithm/550646/

- **Hiring Algorithms**
  - https://hbr.org/2019/05/all-the-ways-hiring-algorithms-can-introduce-bias
  - https://www.vox.com/recode/2019/12/12/20993665/artificial-intelligence-ai-job-screen
  - https://www.welcometothejungle.com/en/articles/what-if-your-recruiter-were-an-algorithm