Section Three

Class COGS 9

Teaching Assistant: Matthew Feigelis

A Mulching Proposal

- General thoughts?

Background

- Algorithmic systems are ubiquitous in everyday life
- These systems may create unjust or unfair outcomes
- Standards / Guidelines have been developed to ensure good values are being used

FAT Framework

- A set of guidelines that may minimize unfair or unjust outcomes from your algorithm
- **Fair**: lacking biases which create unfair and discriminatory outcomes
- **Accountable**: answerable to the people subject to them
- **Transparent**: open about how and why particular decisions were made

- This paper: Applying FAT guidelines to an algorithmic solution to ensure it does not create unfair and unjust outcomes

The Problem

- The problem:
  - World-hunger due to low amounts of fertile soil to grow food.
  - Growing elderly populations

The Algorithmic Solution

- **Proposal**: “mulching elderly people into high-nutrient slurry, which could be used as a source of fertilizer to solve food scarcity issues”
- **Algorithm**:
  - ID low social connectivity people on social media
  - Extract face image from profile
  - Computer vision algorithm to predict age >60. Filter for elderly
  - Send database to UAV system which locates and collects elderly

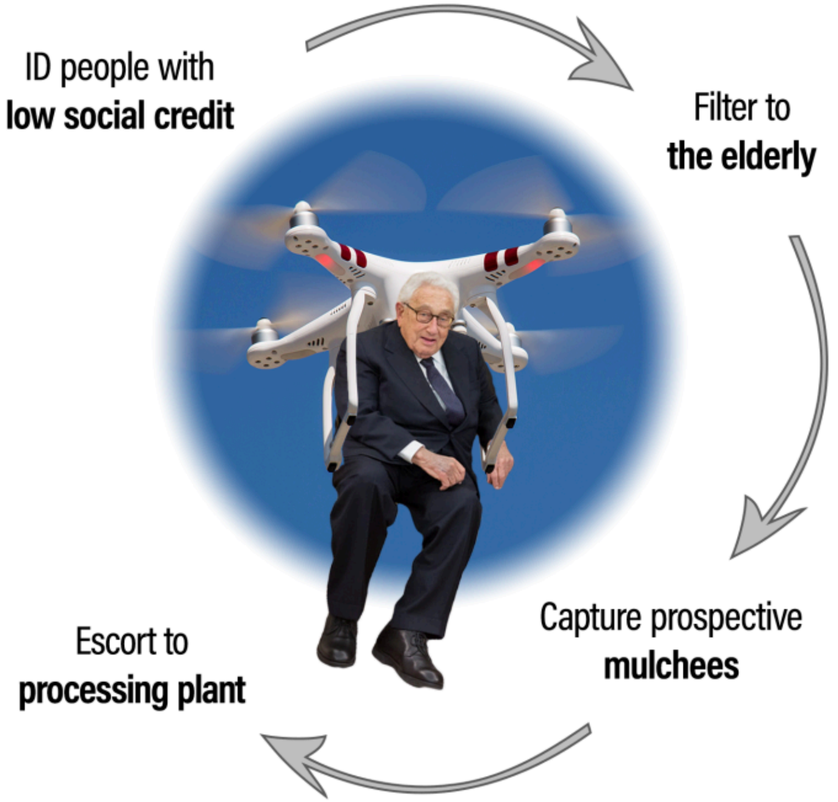

Figure 1: A publicity image for the project, produced by Logan-Nolan Industries

Algorithmic Systems Design

- Application of the FAT framework

- Demographic fairness: Did the system (either in determining social connectivity or age) evenly distribute false positives and negatives across racial and gender groups?

Algorithmic Systems Design

- Sampled 900 images of individuals to determine demographics have largest % of being mulched.

Table 1: Percentage of individuals tagged as worthy of mulching, by demographic.

	Mulching Probability				
Race	*Cis Man*	*Cis Woman*	*Trans Man*	*Trans Woman*	*Non-Binary Person*
White	44.6%	33.3%	2.2%	3.2%	1.1%
Asian-American	22.2%	16.3%	2.8%	1.2%	1.8%
African-American	26.9%	11.2%	2.3%	1.9%	3.4%
Latino	16.9%	18.7%	3.3%	1.2%	1.7%
Native American	14.4%	12.4%	1.0%	0.8%	1.5%
Hawaiian & Pacific Islander	11.6%	7.8%	2.4%	1.1%	0.7%

- We see that white cis males disproportionately did, which is unfair. The algorithm team decided to collects more data on other demograhpics and this fixed the fairness issue.

Algorithmic Systems Design

Table 1: Percentage of individuals tagged as worthy of mulching, by demographic.

	Mulching Probability				
Race	*Cis Man*	*Cis Woman*	*Trans Man*	*Trans Woman*	*Non-Binary Person*
White	44.6%	33.3%	2.2%	3.2%	1.1%
Asian-American	22.2%	16.3%	2.8%	1.2%	1.8%
African-American	26.9%	11.2%	2.3%	1.9%	3.4%
Latino	16.9%	18.7%	3.3%	1.2%	1.7%
Native American	14.4%	12.4%	1.0%	0.8%	1.5%
Hawaiian & Pacific Islander	11.6%	7.8%	2.4%	1.1%	0.7%

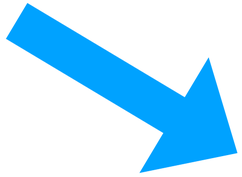

Table 2: Post-audit mulching probabilities.

	Mulching Probability				
Race	*Cis Man*	*Cis Woman*	*Trans Man*	*Trans Woman*	*Non-Binary Person*
White	44.6%	43.3%	44.2%	46.3%	41.2%
Asian-American	52.2%	51.3%	55.8%	49.6%	52.3%
African-American	46.9%	51.1%	53.2%	49.1%	53.3%
Latino	56.9%	48.2%	47.3%	51.1%	47.4%
Native American	54.4%	54.2%	51.5%	48.8%	51.2%
Hawaiian & Pacific Islander	51.6%	48.6%	44.9%	51.1%	47.0%

Algorithmic Systems Design

- Accountability: Refers to the ability of people to address any failure or inaccuracy produced by an algorithmic system.
- Transparency: are users aware of how the algorithm reaches decisions?

- What are the two points of failure in the algorithm?
  - The computer vision algorithm could fail, incorrectly classifying someone as elderly
  - The analysis of social connections might be inaccurate due to a person's limited presence on social media sites

Algorithmic Systems Design

- To make the algorithm more accountable and transparent the authors suggest an pre-mulching accountability mechanisms:
    - On location of an elderly person the drone explains why they are being collected
    - The elderly person has 10 seconds to present reasons why they should not be collected
    - A customer service member is then contacted to continue bargaining with the elderly person
    - The customer service member ultimately decides whether to proceed with the collection or not

Conclusion

- This paper highlight the dangers of blindly trusting algorithms and guidelines
- Highlights the need for ethical considerations in technological advancements during all stages of development
- Applying frameworks such as FAT is a minimum. You could satisfy a framework and still be unethical, as demonstrated by the satire here

Discussion Questions

- Discuss amongst groups of 2-5 people the following. Take 10 minutes. Then we'll reconvene and make a list together.

1. What algorithms (or **products**) do you interact with in your day to day life do you consider may be unethical?

2. What could we do as data scientists or engineers to avoid unethical algorithms being deployed to the public?

1. What algorithms do you interact with in your day to day life do you consider may be unethical?

- Recommender systems: use your information —-> Reinforcement learning algorithm —> gives you recommendations (product for you to buy, or video for you to watch)
- Privacy concerns: Face recognition, Satellite imagery

- Examples:
  - FaceBook feed recommender system
  - Credit scores
  - Face recognition
  - Satellite imagery

What could we do as data scientists or engineers to avoid unethical algorithms being deployed to the public?

- Oversight, committees, regulatory agencies
- Andon cord. Toyota. Employees can pull a cord to stop production if they see a problem. Management then would discuss and implement changes before resuming production
- Adding Ethical challenges to hiring process

❑ Have we listed how this technology can be attacked or abused?

❑ Have we tested our training data to ensure it is fair and representative?

❑ Have we studied and understood possible sources of bias in our data?

❑ Does our team reflect diversity of opinions, backgrounds, and kinds of thought?

❑ What kind of user consent do we need to collect and use the data?

❑ Do we have a mechanism for gathering consent from users?

❑ Have we explained clearly what users are consenting to?

❑ Do we have a mechanism for redress if people are harmed by the results?

❑ Can we shut down this software in production if it is behaving badly?

❑ Have we tested for fairness with respect to different user groups?

❑ Have we tested for disparate error rates among different user groups?

❑ Do we test and monitor for model drift to ensure our software remains fair over time?

❑ Do we have a plan to protect and secure user data?

Attendance

- Enter your number and today's word into the attendance form

- Todays word: Luke
- Form: https://forms.gle/tx9GcpANHEMwj8Jv7

- https://tinyurl.com/cog9-spring-23