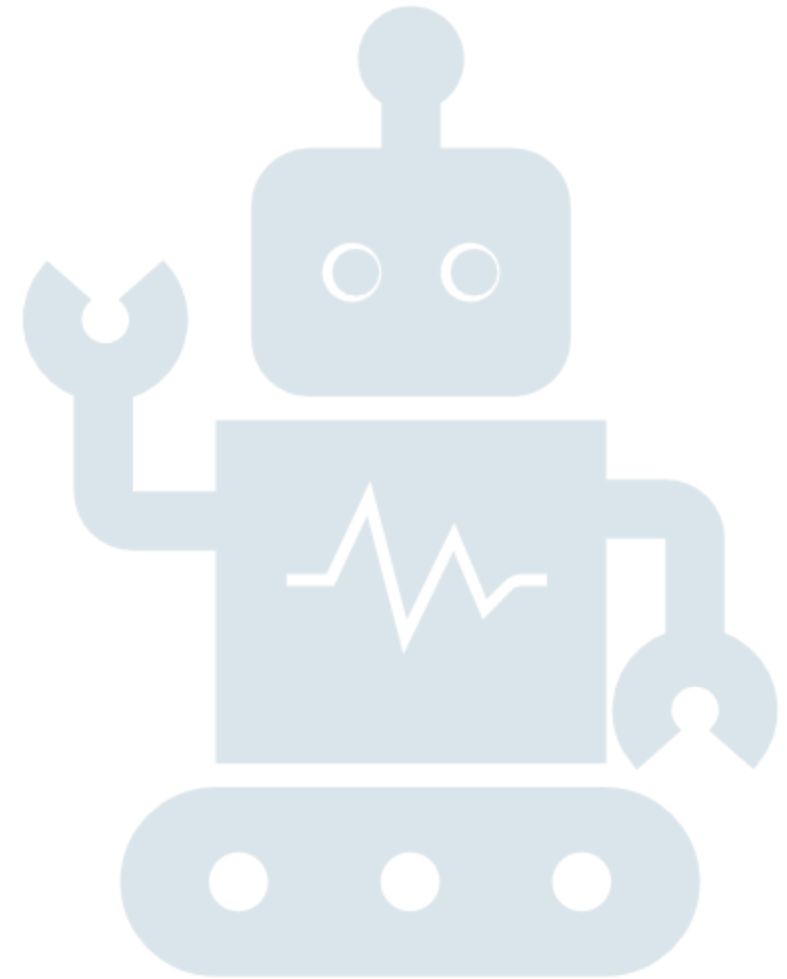
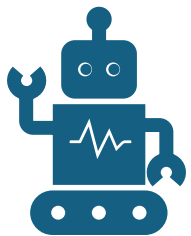




Responsible Artificial Intelligence (AI)



Discussion

Scenario: An institution of higher education is creating an admissions system powered by AI (to determine who is admitted).

Which of the following policies do you think are most fair? Why?

- Ignore gender (don't use it as a variable at all): if 70% of people who get in are women, that's ok if we haven't used gender as a factor.
- Adjust for different groups: women might have had experiences that led to different numbers than men or non-binary people, and biased previous data (we need to adjust)
- Output reflects input: if 30% of applicants are women, 30% of accepted students should be women
- Prediction is correct: percent of admitted women who graduate is equal to the percent of admitted men and non-binary people who graduate.

Algorithmic Bias and Fairness

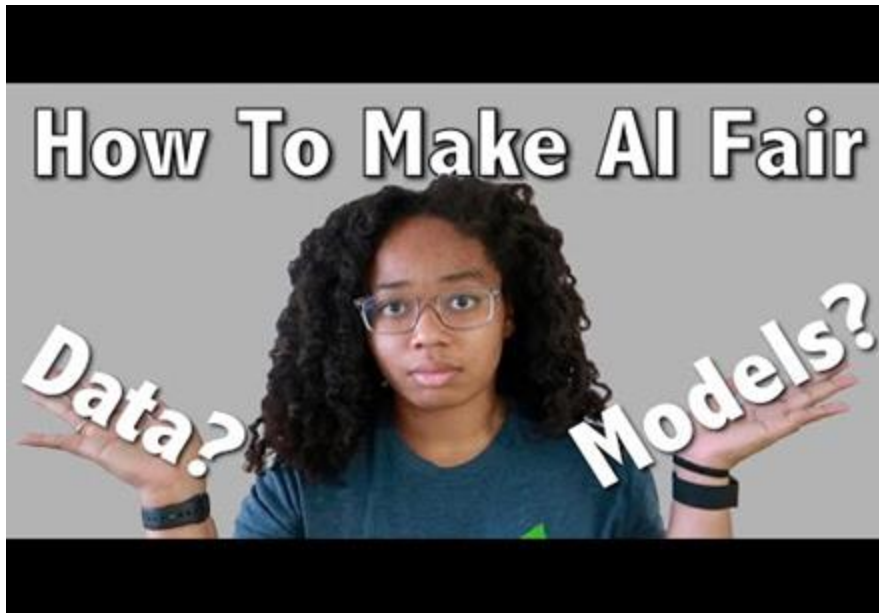
Video Overview

- Common types of algorithmic bias include:
 - Data that reflects existing biases.
 - Unbalanced classes in training data.
 - Data that does not capture the right value.
 - Data that is amplified by feedback loops.
 - Malicious data.
- Bias or preference is often used in the human brain to efficiently find patterns.
- However, when implementing algorithms, if bias is not properly acknowledged as exceptions in patterns it may then yield inadmissible output.
- Establish differences between bias and discrimination.



https://youtu.be/gV0_raKR2UQ?si=GYNLjmr-icnjMI11

Making Algorithms Fairer



https://youtu.be/hZG9tyOcyx0?si=pyBbbbLtgJku5-_F&t=299

Video Overview

- Class imbalance: when some classes (or groups) have substantially more instances than other classes.
- Often race or gender bias originates from having samples that poorly represent certain groups (e.g., women or people of color).
- Class imbalance mitigation: reduce the size of dataset by removing data from categories that are overrepresented.

References for Responsible AI



RESPONSIBLE AI BY AWS
MACHINE LEARNING
UNIVERSITY: [YOUTUBE](#)



AI & JUSTICE SYSTEM:
[\[URL\]](#)



ARTICLE BY
SMITHSONIAN
MAGAZINE: [\[URL\]](#)



PRINCETON DIALOGUES
ON AI AND ETHICS [\[URL\]](#)