Housekeeping

- Please mute your mic during the talk
- There will be time at the end for discussion in chat, publically or privately
- This session will be recorded
- Slides and recording will be available after
- Follow on twitter @equal4success

Biases in AI

Izzy Hampson

Definition

- AI (artificial intelligence) is the leading technology in "Fourth Industrial Revolution". It refers to the technological advances – from biotechnology to big data – that are rapidly reshaping the world as we

Bias and discrimination in AI

- AI systems are often biased, particularly along race and gender lines.

- Amazon recruitment algorithm displayed gender biases.

- The algorithm was trained on historical data and the preferential recruitment of males, it ultimately could not be fixed and had to be dropped.

Gender Biases

- Link between development of AI systems with gender biases and the lack of women in design teams.

- AI Now report: clear connection between the male dominated AI industry and its discriminatory systems and products.

- Less recognition of the ways AI products incorporate stereotyped representations of gender in their design.
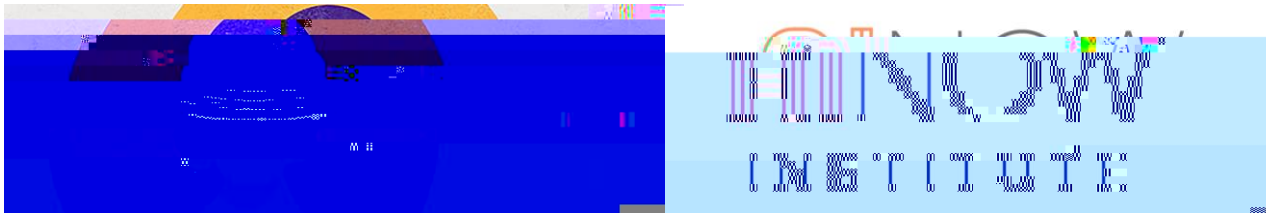
Word-embeddings

- AI fills in the word queen in the sentence

- Words are converted to numerical representations used as inputs models.

- Words are represented as a sequence, or a vector of numbers. If two words have similar meanings, their "embeddings" will be mathematically close to each other.

- The issues arise in cases where AI fills in sentences like

Racial Bias in AI

- Chatbots that become racist in less than a day

- Facial technology that fails to recognize users with darker skin colours

- Ad-serving algorithms that discriminate by gender and race

- An AI hate speech detector that's racially biased itself

Same or different person?
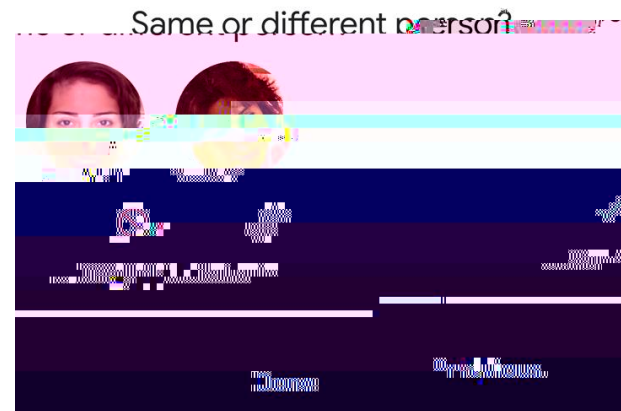
- Racial biases are thoroughly ingrained in society, and have the potential to be exacerbated by algorithms, such as in the criminal justice system.

- Significant problems include the lack of unbiased historical data, an unbalanced workforce, and limited user testing.

What causes AI bias?

- An incomplete or skewed training dataset
- Labels used for training
- Features and modelling techniques

The Bride Problem:
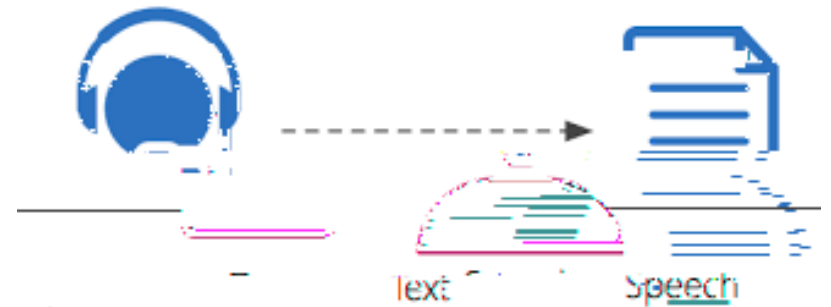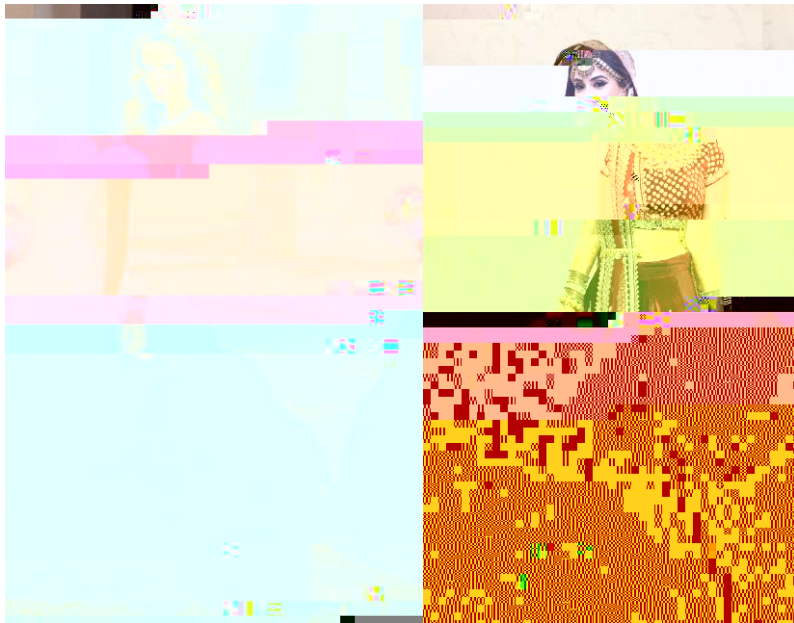
Ways to address bias in AI

- Ensure diversity in the training samples
- Ensure that humans labelling the samples come from diverse backgrounds.

- Encourage machine-learning teams to measure accuracy levels separately for different demographic categories and to identify when one category is being treated unfavourably e.g. Counterfactual Fairness
- Collect more training data associated with underrepresented groups and apply machine learning de

responsibility

-James Manyika, Jake Silberg and Brittany Presten, 2019, Harvard Business Review

Resources

- Man is to Computer Programmer as Woman is toHomemaker? Debiasing Word Embeddings (https://papers.nips.cc/paper/6228-man-is-to-computer-programmer-as-woman-is-to-homemaker-debiasing-word-embeddings.pdf)

- Gender Shades: Intersectional Accuracy Disparities inCommercial Gender Classification (http://proceedings.mlr.press/v81/buolamwini18a/buolamwini18a.pdf)

- AI and Gender Four Proposals for Future Research (