Contents

1. Predictive Policing
    a. PredPol
    b. COMPAS
    c. Criminal Risk Assessment
    d. Steps in the Right Direction
2. General Resources
    a. ACM Teach LA Introduction to AI and ML
    b. Stanford CS224N Lecture on Bias in AI
    c. How AI Bias Happens
    d. Google Fairness Tutorial
    e. Book on Fairness in Machine Learning
    f. Explainable AI
    g. Book on Interpretable Machine Learning
    h. Mathwashing
    i. Course on Fairness in AI and Algorithms
3. Biases in AI and ML
    a. Gender Bias
    b. Chatbots and Bias
    c. Bias in Speech Recognition
    d. Racial Bias
    e. Other Industries affected
    f. Steps in the Right Direction
4. Questions to Think About
    a. General AI Bias
    b. Making AI Explainable
    c. Voice Assistants
    d. Voice Recognition
    e. Facial Recognition
    f. Industries affected by AI

(Use document outline to navigate)
(If you do not see an outline go to View → Show document outline)

Predictive Policing

PredPol

- How does predpol work?
  https://www.predpol.com/how-predictive-policing-works/
- Falsified Data being used for Predictive Policing
  https://www.technologyreview.com/2019/02/13/137444/predictive-policing-algorithms-ai-crime-dirty-data/
- Predictive Policing Algorithms are Racist
  https://www.technologyreview.com/2020/07/17/1005396/predictive-policing-algorithms-racist-dismantled-machine-learning-bias-criminal-justice/
- A case study in predictive policing:
  https://rss.onlinelibrary.wiley.com/doi/full/10.1111/j.1740-9713.2016.00960.x
- Paper on ML and Policing:
  https://academic.oup.com/policing/advance-article/doi/10.1093/police/paz035/5518992
- A fight against racially biased policing algorithms:

  https://www.technologyreview.com/2020/06/05/1002709/the-activist-dismantling-racist-police-algorithms/
- Is AI used to predict crime biased?:
  https://www.smithsonianmag.com/innovation/artificial-intelligence-is-now-used-predict-crime-is-it-biased-180968337/

COMPAS

System used to predict recidivism (the likelihood that a criminal will re-offend)
Used for bail, sentencing and parole

- https://www.nytimes.com/2017/06/13/opinion/how-computers-are-harming-criminal-justice.html
- https://hdsr.mitpress.mit.edu/pub/7z10o269/release/4

Criminal Risk Assessment

https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

Steps in the Right Direction

- Interpretable Machine Learning for Recidivism Prediction
  https://arxiv.org/abs/2005.04176
- Model Cards for Model Reporting - to make ML more transparent
  https://arxiv.org/pdf/1810.03993.pdf

General Resources

ACM Teach LA Introduction to AI and ML + Resources

https://teachla.uclaacm.com/classes/ml

An amazing resource for a complete introduction to AI and ML, also has a bunch of useful resources linked at the bottom!

Stanford CS224N Lecture 19: Bias in AI

Slides on Bias in the Vision and Language of AI
https://web.stanford.edu/class/cs224n/slides/cs224n-2019-lecture19-bias.pdf
Corresponding Lecture Video:
https://www.youtube.com/watch?v=XR8YSRcuVLE&list=PLoROMvodv4rOhcuXMZkNm7j3fVwBBY42z&index=19

Slightly high-level, but very interesting to watch nevertheless. Provides a good overview of the types of possible biases.

How AI Bias Happens

- MIT Tech Review
  https://www.technologyreview.com/2019/02/04/137602/this-is-how-ai-bias-really-happensand-why-its-so-hard-to-fix/
- McKinsey Article on tackling bias in AI
  https://www.mckinsey.com/featured-insights/artificial-intelligence/tackling-bias-in-artificial-intelligence-and-in-humans
- Harvard Business Review Article on what can be done about biases in AI
  https://hbr.org/2019/10/what-do-we-do-about-the-biases-in-ai

Google Fairness Tutorial

https://sites.google.com/view/fairness-tutorial

Book on Fairness in Machine Learning

https://fairmlbook.org/

This book is extremely detailed and provides a mathematical understanding of different perspectives of fairness and discusses approaches to mitigate bias

Explainable AI

Aiming to make AI less of a black-box and make its decisions more explainable

- https://www.darpa.mil/program/explainable-artificial-intelligence
- https://www.zdnet.com/article/explainable-ai-artificial-intelligence-a-guide-for-making-black-box-machine-learning-models-explainable/

Book on Interpretable Machine Learning

https://christophm.github.io/interpretable-ml-book/

Mathwashing

https://www.mathwashing.com/

A course on Fairness in AI and Algorithms

http://catherineyeo.tech/ai_fairness_course

Biases in AI and ML

Gender Bias

1. **Female-voiced Voice Assistants**
    a. https://medium.com/inclusive-conversational-ai/inclusive-conversational-ai-the-case-of-female-voice-assistants-2212d45742be

2. **Amazon's biased recruiting algorithm**
   Amazon stopped the use of an applicant sorting algorithm from trials after it realized that the algorithm was severely biased towards men:
    a. https://www.reuters.com/article/us-amazon-com-jobs-automation-insight-idUSKCN1MK08G
    b. https://becominghuman.ai/amazons-sexist-ai-recruiting-tool-how-did-it-go-so-wrong-e3d14816d98e
    c. https://www.theverge.com/2018/10/10/17958784/ai-recruiting-tool-bias-amazon-report

3. **Gender Bias in Word Embeddings**
    a. http://wordbias.umiacs.umd.edu/

Chatbots - GPT-2 and GPT-3

- https://www.analyticssteps.com/blogs/what-openai-gpt-3
- https://www.technologyreview.com/2020/07/20/1005454/openai-machine-learning-language-generator-gpt-3-nlp/
- https://blog.exxactcorp.com/what-can-you-do-with-the-openai-gpt-3-language-model/
- https://www.theverge.com/21346343/gpt-3-explainer-openai-examples-errors-agi-potential

Bias

- https://towardsdatascience.com/gender-bias-in-gpt-2-acf65dc84bd8
- https://medium.com/fair-bytes/how-biased-is-gpt-3-5b2b91f1177
- https://www.technologyreview.com/2020/10/23/1011116/chatbot-gpt3-openai-facebook-google-safety-fix-racist-sexist-language-ai/

Bias in Speech Recognition

Usually happens with respect to accent and gender

1. **Alzheimer's Prediction** (Healthcare)

    a. Example 1 in https://www.quantib.com/blog/understanding-the-role-of-ai-bias-in-healthcare
    b. https://www.forum-wbp.com/artificial-intelligence-in-health-new-technology-old-biases/
    c. https://medium.com/@swethaxnarayanan/an-inconvenient-truth-the-problem-with-data-quality-in-medtech-ai-476d7486dae0
2. **Gender and Racial Biases in Voice** https://hbr.org/2019/05/voice-recognition-still-has-significant-race-and-gender-biases
3. **Car voice recognition did not work for women** (outdated) https://techland.time.com/2011/06/01/its-not-you-its-it-voice-recognition-doesnt-recognize-women/

Racial Bias

1. **Healthcare** https://www.scientificamerican.com/article/racial-bias-found-in-a-major-health-care-risk-algorithm/
   They fixed this by working on it later and bringing it up to 84% accuracy.
2. **Facial Recognition**
    a. https://www.nist.gov/news-events/news/2019/12/nist-study-evaluates-effects-race-age-sex-face-recognition-software
    b. https://time.com/5520558/artificial-intelligence-racial-gender-bias/
3. **Predictive Policing** https://www.technologyreview.com/2020/07/17/1005396/predictive-policing-algorithms-racist-dismantled-machine-learning-bias-criminal-justice/

Other Industries Affected

https://dzone.com/articles/aiml-bias-explained-with-examples
Employment, Housing, Banking and Education, among others are affected

Steps in the right direction

Gender Neutral Voice Assistant:
https://www.npr.org/2019/03/21/705395100/meet-q-the-gender-neutral-voice-assistant

Questions to think about

General AI Bias

- As the articles show, various factors can cause bias to happen. Can you think of one way to try and reduce the influence of each factor?

Making AI Explainable

- A large part of AI working well happens because of its black-box nature. What do you think is more important? Accuracy or understanding how it works?
- How do you think the black-box nature of AI and ML can be overcome?

Voice Assistant

- Why do you think the AI voice assistant behaves this way?
- Would you expect this to happen in real life? How does your answer reflect on the way the AI works?

Voice Recognition Bias

- How would you suggest this bias be overcome? What do you think is the reason this bias happens?
- Do you think speech recognition bias is important to prevent? Why / Why not?

Facial Recognition bias

- What do you think causes these inaccuracies when it comes to recognizing people with darker skin?
- What kinds of effects do you think this type of bias can have on society?

Industries Affected

- Where else do you think AI bias could have an unseen but relevant effect? How?

Even if AI did somehow manage to be *perfect*, would you say it should be used all the time? If no, where do you think it should / should not be used?