



CPSC 2910 – Lesson 7

AI Bias

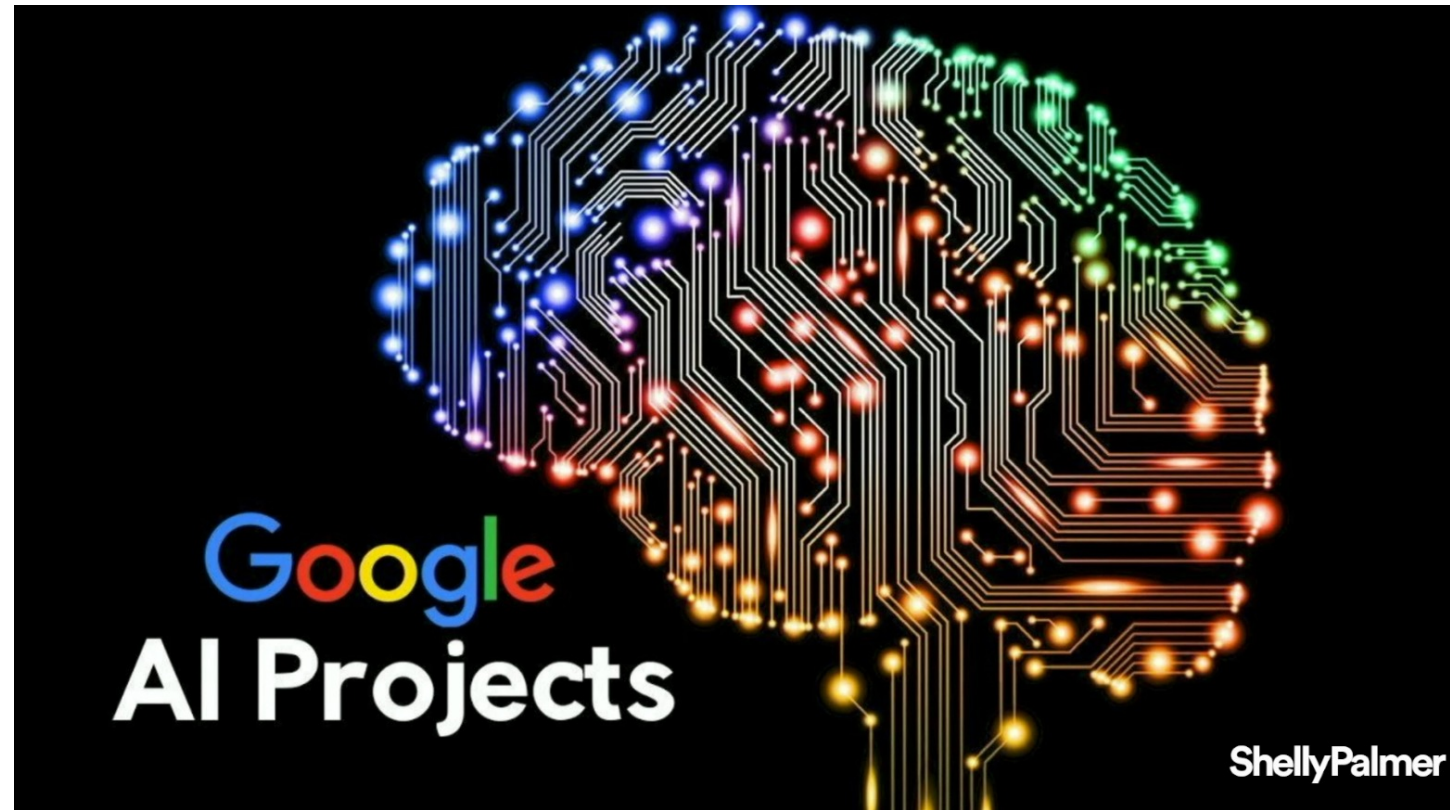
Carrie L Russell | Clemson University, Spring 2023

Agenda

- **Quiz 3**
- **What is AI Bias and Why is there Bias?**
- **Examples of Bias**
- **How to Hold Algorithms Responsible – Minimizing Bias**

AI Bias with Google Translate

Barak Turovsky at 2020 Shelly Palmer Innovation Series Summit:



<https://www.shellypalmer.com/2020/02/shelly-palmer-innovation-series-summit-ces-2020-highlights-ai-bias/>

What is “AI Bias” and Why does it happen?

An anomaly -

- AI bias is an anomaly in the output of machine learning algorithms.
- Could be due to the prejudiced assumptions made during algorithm development process or prejudices in training data.

Main reasons for bias -

- **Cognitive biases** – effective feelings towards a person or a group based on their perceived group membership.
- **Lack of complete data** – if data is not complete, it may not be representative and therefore it may include bias.

AI Bias Examples

Amazon's biased recruiting tool -

- Did not rate candidates fairly and were biased against women.

Racial bias in healthcare risk algorithm -

- Algorithm used on more than 2000 million US citizens produced faulty results that favored white patients or black patients.

Bias in Facebook ads -

- In 2019, Facebook was allowing it's advertisers to intentionally target adverts according to gender, race, and religion.

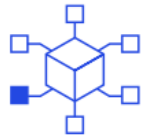
Minimizing Bias: Processes

Minimizing bias will be critical if artificial intelligence is to reach its potential and increase people's trust in the systems.

Minimizing bias will be critical if artificial intelligence is to reach its potential and increase people's trust in the systems.

Six potential ways forward for artificial-intelligence (AI) practitioners and business and policy leaders to consider

1



Be aware of contexts in which AI can help correct for bias and those in which there is high risk for AI to exacerbate bias

2



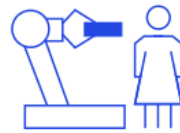
Establish processes and practices to test for and mitigate bias in AI systems

3



Engage in fact-based conversations about potential biases in human decisions

4



Fully explore how humans and machines can best work together

5



Invest more in bias research, make more data available for research (while respecting privacy), and adopt a multidisciplinary approach

6



Invest more in diversifying the AI field itself

Minimizing Bias: Tools

AI Fairness 360 - IBM released an open-source library to detect and mitigate biases in unsupervised learning algorithms that has 34 contributors (as of 9/20) on [Github](#). It enables AI programmers to –

- Test biases in models and datasets with a comprehensive set of metrics
- Mitigate biases with the help of 12 packaged algorithms such as Learning Fair Representations, Reject Option Classification, Disparate Impact Remover.
- NOTE: AI Fairness 360's bias detection and mitigation algorithms are designed for binary classification problems – it needs to be extended to multiclass and regression problems if the problem is more complex.

IBM Watson OpenScale - IBM's [Watson OpenScale](#) performs bias checking and mitigation in real time when AI is making its decisions.

Google's What-If Tool - Using the [What-If Tool](#), you can test performance in hypothetical situations, analyze the importance of different data features, and visualize model behavior across multiple models and subsets of input data, and for different ML fairness metrics.

Minimizing Bias: Organizations

Algorithmic Justice League - mission is to raise awareness about the impacts of AI, equip advocates with empirical research, build the voice and choice of the most impacted communities, and galvanize researchers, policy makers, and industry practitioners to mitigate AI harms and biases. <https://www.ajl.org>

Google's Responsible AI Practices - Google is committed to making progress in the responsible development of AI and to sharing knowledge, research, tools, datasets, and other resources with the larger community. <https://ai.google/responsibilities/responsible-ai-practices/?category=fairness>

Discussion Topic 1

Using the articles mentioned in this lecture as a guideline, answer the following –

An online course in artificial intelligence at a major university had 300 students and nine teaching assistants who answered questions and conversed with students online. The students were not told, and most did not guess, that one of the teaching assistants was an AI program. If you were in the class, would you want to be told that you were conversing with a program? Give your reasons. Explain whether you think there is an ethical issue involved.

Discussion Topic 2

Using the articles mentioned in this lecture as a guideline, answer the following –

Should we leverage AI to deploy fully-automated and autonomous weapons of war? Why or why not?

In-Class Teamwork

As a team – answer this week’s discussion post in Canvas

AI and Bias –

- You will work together in class to complete the online discussion post - due end of class.
- Individual responses to other posts due midnight before next week’s class.

Reading and Discussion Rubric-Sp22		
Criteria	Ratings	Pts
Provided thoughtful answer to the specified question. Showed thought and reflection in their answer. Referenced key elements of the article to back up their answer.	This area will be used by the assessor to leave comments related to this criterion.	12.5 pts
Commented on at least 2 classmate answers. Provided a thoughtful response either supporting or challenging an answer provided by two of their classmates answers.	This area will be used by the assessor to leave comments related to this criterion.	5 pts
No major grammar or spelling issues in the submission. Shows an attention to detail by ensuring that the grammar and spelling are correct in their submission.	This area will be used by the assessor to leave comments related to this criterion.	2.5 pts
Total Points: 20		

Looking Ahead

Next week -

- Reading and Online Discussion –

Read <https://techcrunch.com/2019/09/26/privacy-queen-of-human-rights-in-a-digital-world/> and <https://www.npr.org/sections/alltechconsidered/2014/03/03/285334820/if-theres-privacy-in-the-digital-age-it-has-a-new-definition>

- Be prepared to discuss as a class and in small groups next week. You will work together in class to complete the online discussion post, so it is important that you **READ AHEAD**.
- In-class presentation on Privacy in the Digital Age – Team 12 (benefits) and Team 9 (harms)
- In-class quiz on this week's topics

Questions?

Use our Teams channel first – this is preferred for general questions!

- Prof Russell - crusse4@clemson.edu
- Matt Bernard - mberna2@clemson.edu