Beyond AI Hype

Algorithmic and Dataset Bias

Chinasa T. Okolo
www.chinasatokolo.com

Session III

Session III Schedule

Group Discussion

Bias Lecture

Bias Case Studies Activity

Case Studies Presentations

Case Studies Debrief

Group Discussion

- What do you know about bias?

- What are some examples of bias?

- Do you know how bias impacts ML models?

Bias Lecture

What is bias?

Bias refers to a tendency or inclination to favor or oppose something or someone, often without considering all the relevant facts or being objective. Bias can be conscious or unconscious, and it can affect how we perceive and interpret information. It can lead to unfair judgments or unequal treatment based on factors like race, gender, or personal beliefs.
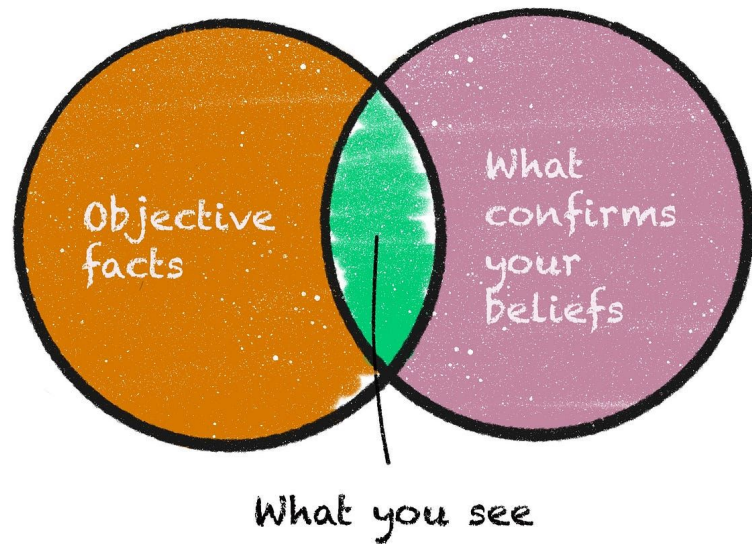
…bias affects many aspects of our daily lives!

Confirmation Bias

When we seek out information that confirms our existing beliefs or opinions.

Objective facts

What confirms your beliefs

What you see

Stereotyping

Making assumptions or generalizations about individuals or groups based on characteristics such as their race, gender, nationality, or profession.
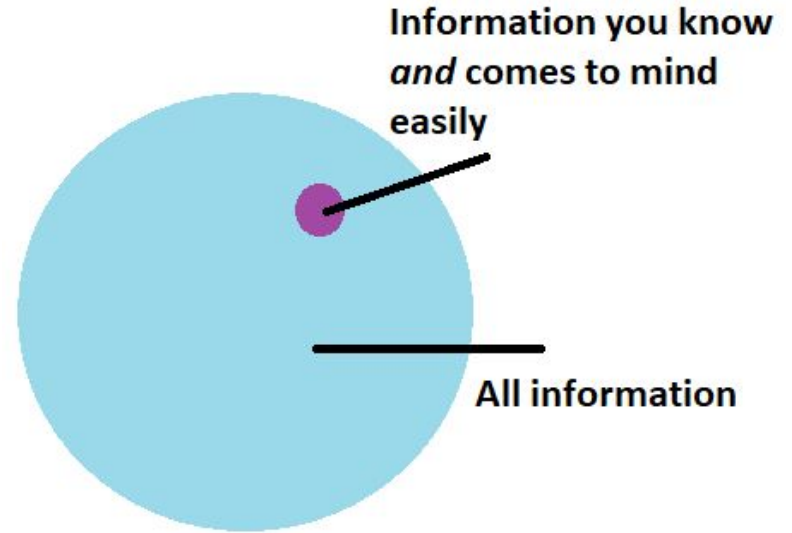
Availability Bias

When we rely heavily on information or examples that come readily to mind, rather than considering a broader range of data.

Information you know *and* comes to mind easily

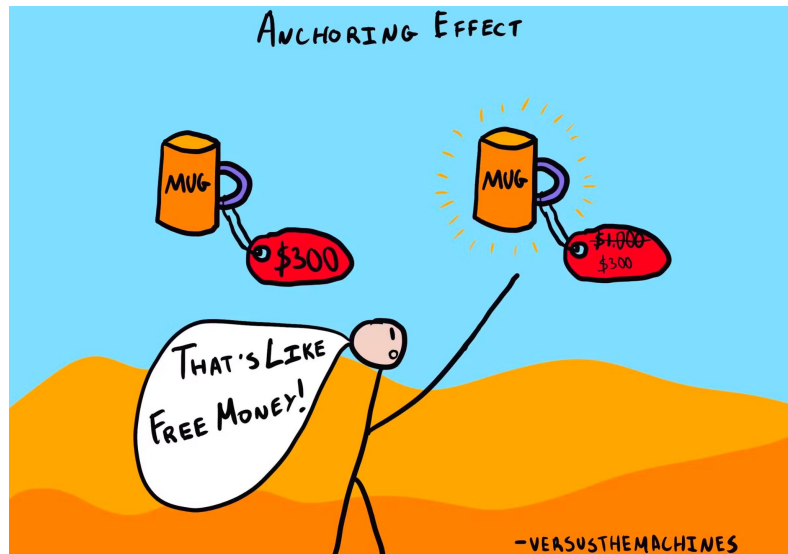All information

Implicit Bias

Unconscious attitudes or stereotypes that influence our behavior and decisions without us realizing it.
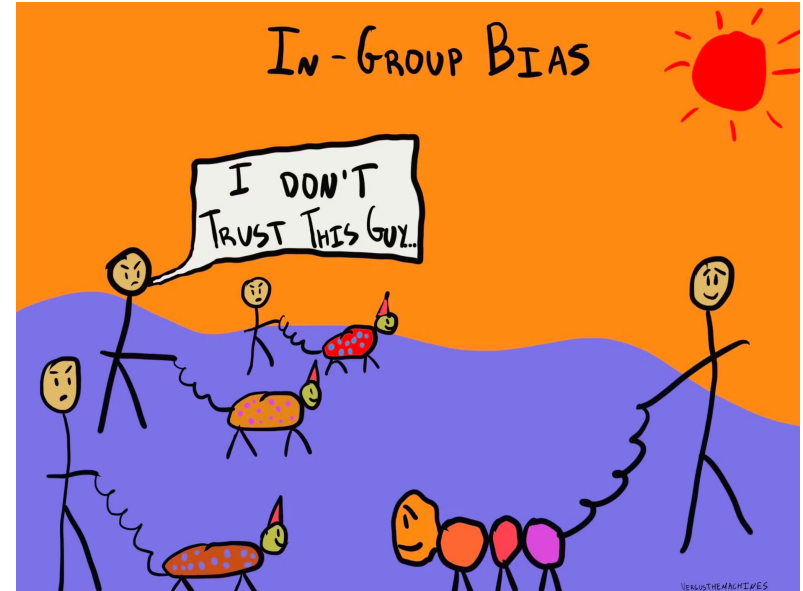
Anchoring Bias

When we rely too heavily on the first piece of information we encounter when making decisions, even if it may not be the most accurate or relevant information.

In-group Bias

The tendency to favor individuals who belong to the same group as us (in-group) over those from different groups (out-group).

What is Bias?

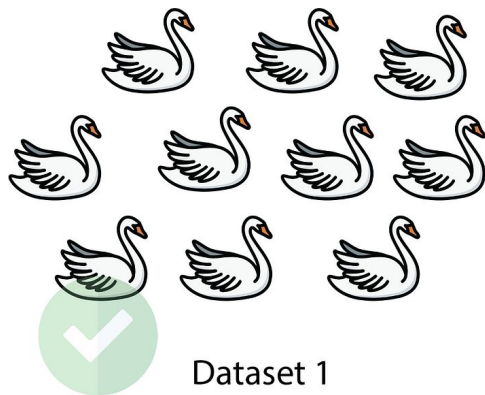**What is Bias? - Definition & Types**
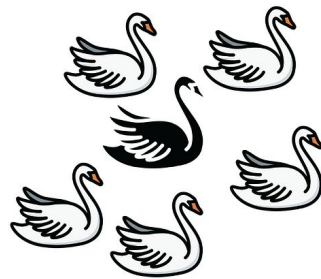
Bias in Machine Learning

Dataset Bias

- People have biases that affect how they perceive and value things

- Datasets can be biased in how they are collected, organized, and annotated
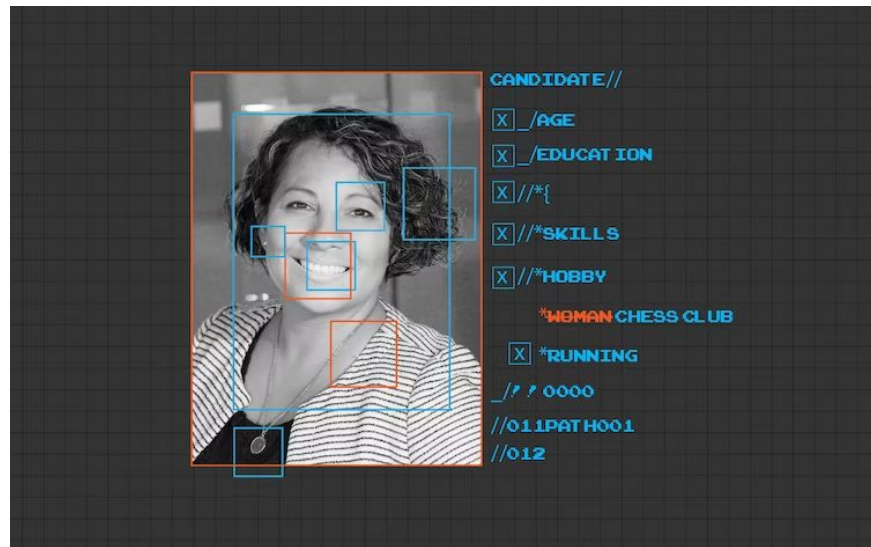

Dataset 1

All swans are white

Dataset 2

https://towardsdatascience.com/types-of-biases-in-data-cafc4f2634fb

How does bias impact ML models?

Algorithmic Bias

- Algorithmic bias refers to the presence of unfair or discriminatory outcomes in automated decision-making systems or algorithms

- Biased datasets can also lead to biased algorithms

Algorithmic Bias

Bias Case Studies

Bias Case Study

Bias Case Study Instructions

- Find an example of AI bias that has been featured in the media or covered in research articles

- Read through a few articles or watch a few videos to learn more
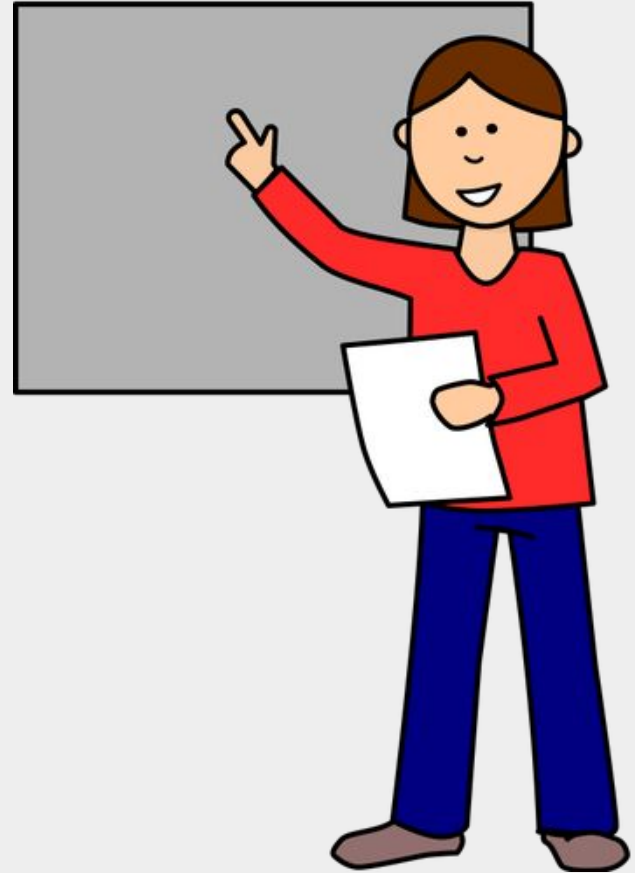
- Write up your findings

AI Bias Ideas

- Amazon's biased hiring algorithm

- Microsoft's Tay AI chatbot

- Epic (healthcare company) biased medical algorithm

- Biased remote test-taking software

- …use your favorite search engine to find more!

Bias Case Study Write-Up

- **What** is the bias about and what field does it impact (medicine, finance, hiring, etc.)?

- **Who** (company, researchers, etc.) developed the biased tool or algorithm?

- **When** did this happen?

- **How** did it affect users?

- **Why** do you think this is an important issue?

Case Studies Presentations

Ethical Issues of Data Annotation Work

Data Labeling Work

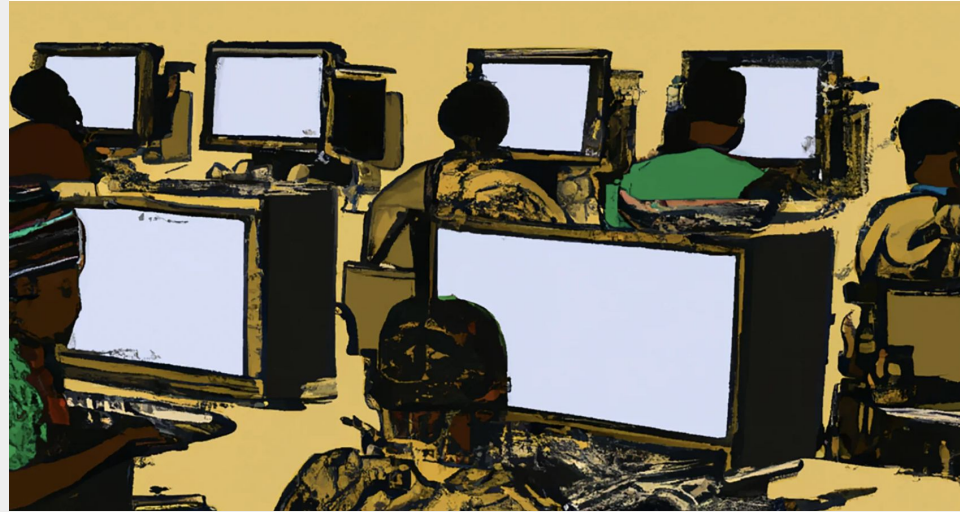# OpenAI Used Kenyan Workers on Less Than $2 Per Hour to Make ChatGPT Less Toxic

— Billy Perrigo, TIME

Case Studies
Debrief

Case Studies Debrief

- How did you like the activity?

- What did you learn?

- What questions do you still have?

The danger of AI is weirder than you think

Questions?

End of Session III

Thank you!
Chinasa T. Okolo
www.chinasatokolo.com

This presentation is copyrighted by Chinasa T. Okolo.

It is available under a
Creative Commons Attribution 4.0 International License
(CC BY-NC-SA 4.0)