Racial Bias in ML Models

Models with Bias Used Today

Recruitment Tools

Facial Recognition

Word Associations

Criminal Justice

Models with Bias Used Today

Recruitment Tools

Data from Amazon applicants trained to recognize the current employees as good candidates. The issue was that the current employees are predominantly men, leading to a gender bias in the model.

Models with Bias Used Today

A study of a large language model showed that European names were recognized as more "pleasant" and words like "women" and "girl" were linked to stereotypical jobs and interests, such as the arts rather than science and math

Word Associations

Models with Bias Used Today

Facial Recognition

Most facial data comes from white males, leading to the recognition algorithms we develop to have lower accuracy rates for other groups, such as darker-skinned women.

Models with Bias Used Today

An algorithm that is used for sentencing inmates based on likelihood of committing another crime was shown to be biased against African-Americans by recommending a longer sentence for a similar likelihood of recidivating.

Criminal Justice

Causes of Bias

It is important to be weary of the causes of biases so you can recognize it too.

A typical path to a biased model:

Historical Human Biases (sexism, racism, etc.) → Inaccurate or Misrepresentative Data → Biased Model
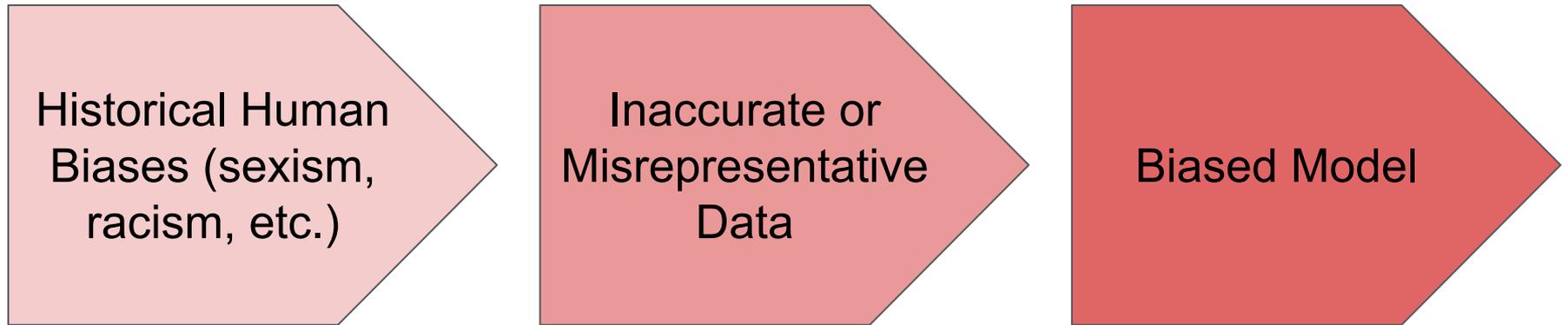
Reflection Question

Try to think of an example of where we might have biased or incomplete data that would impact some ML models?

Hint: Think of different industries and work from there (housing, healthcare, etc.)

Practical Example

Complete this notebook activity:

https://www.kaggle.com/code/anshulshah99/algorithmic-bias-activity

Takeaways from the Example

1. Were you surprised by the results you saw on the Kaggle activity?

2. How will this shape how you interact with and interpret certain ML models?