# Algorithmic Bias & High-Stakes Gambling

## Shrey Gupta

# Introduction

- Duke University '20: Computer Science & Statistics
- Software engineering at Applied Intuition, working on self-driving simulation
- Previously software engineering at Google & quantitative research at a hedge fund

*Disclaimer: all opinions are my own and not the views of my employer, current or past.*

# Agenda

- What is algorithmic bias?

- What are the implications?

- What can I do?

- **Algorithm:** set of step-by-step instructions that computers follow to perform tasks to automate their decision-making
  - **Ex:** determining the ads you see on Facebook
  - **Ex:** picking the individuals who qualify for a loan

# Are algorithms objective?

- We often assume these decisions are **objective** because they're done by computers

- **But they're not...** These algorithms inherit biases from:
  - Unrepresentative data used to "teach" the algorithms
  - Subconscious decisions of programmers

- **Algorithmic bias:** racial, gender, and socioeconomic bias in algorithmic decision-making

# Example: Recidivism

- **Black box models** are being deployed today to determine the chance a convicted criminal will re-offend in the future
  - Individuals in **low-income neighborhoods** are flagged for greater risk of recidivism, *regardless of the steps the individual has taken to correct themself*
  - These communities skew towards neighborhoods containing predominantly **people of color** due to decades of structural racial inequality

# What are the implications?

- Historical biases like these can become **embedded** in models that we rely on to inform or make decisions
- Models like this can **widen the structural inequality** seen today
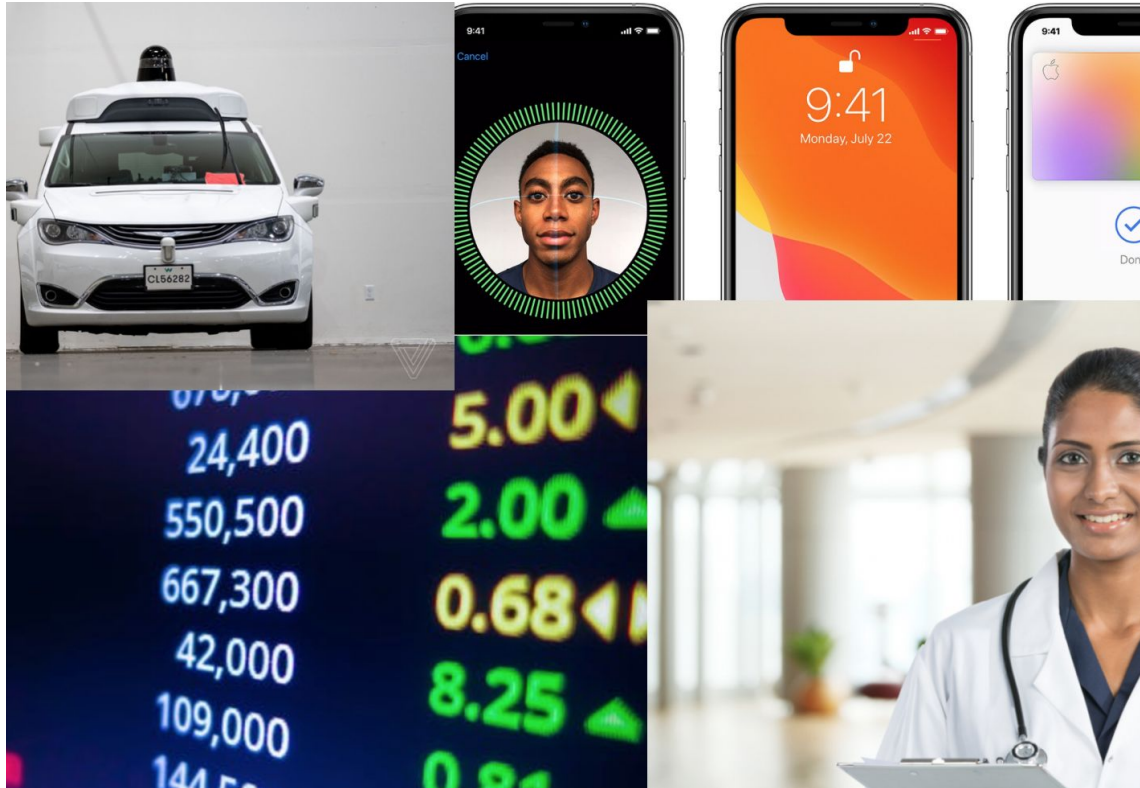
- *(5-10 min)* Discuss potential **instances** of algorithmic bias, and the **implications**
- Consider the following areas:
  - Finance
  - Healthcare
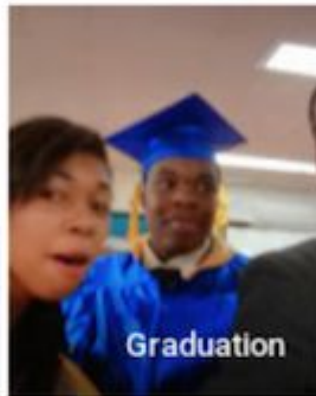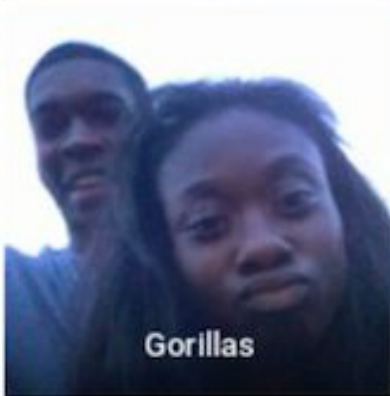  - Criminology & defense
  - Autonomous driving

# Unrepresentative Data
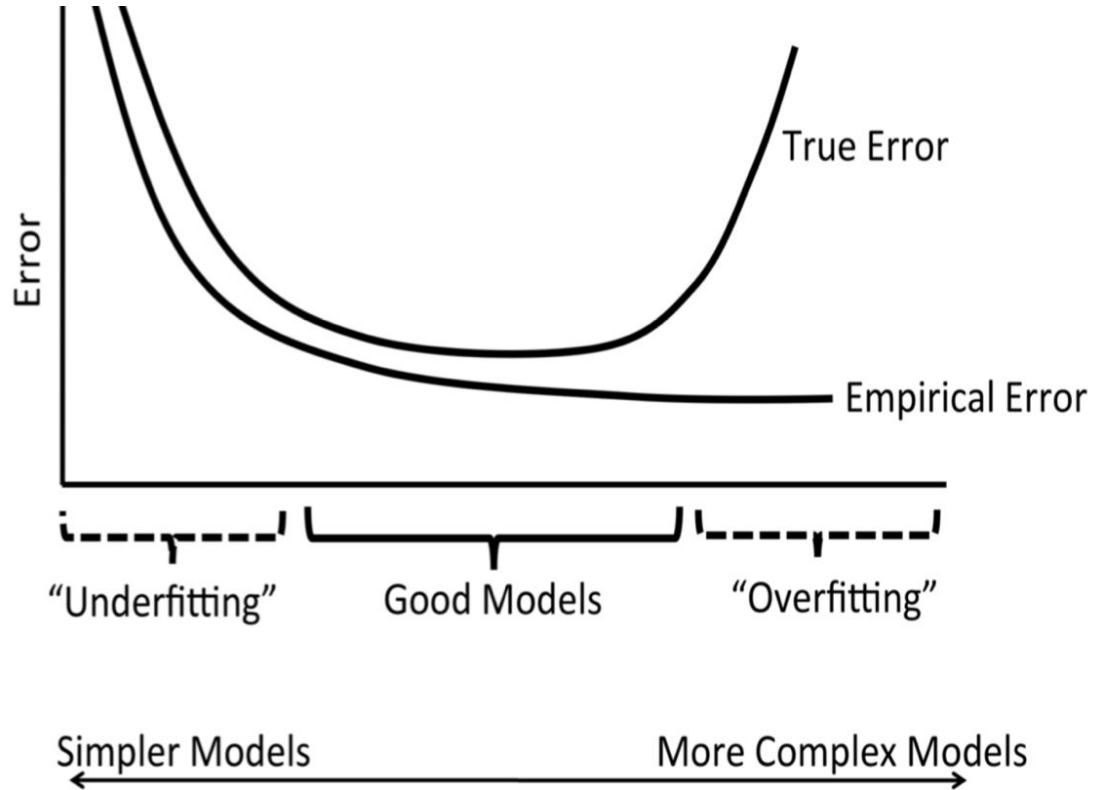
- Pattern recognition and machine learning involve **training models** using examples in datasets
  - But the **datasets** used can be biased, for example, by under- or over-representing certain groups
  - This bias is then transferred to the models, causing them to be systematically inaccurate for **subsets** of the data

# Unrepresentative Data

- We use **random sampling** to create training data: every member of the target population has equal odds of being a part of the training data
  - True random sampling can be **difficult** to achieve due to various factors surrounding how these datasets are collected

- Downstream from actual data collection, there are some

  **strategies** researchers can use to try and **detect bias**

  - Checking for anomalies

  - Experimenting with model settings

  - Running simulations

- What **assumptions** are you making with your model?

- Need to take into account how those assumptions and algorithms are:

  - Propagating **existing societal biases** into new areas

  - Creating **positive feedback loops** where groups receive less opportunity

- Discrimination by algorithms is **difficult to litigate**

  - Anti-discrimination laws are **outdated** and often plaintiffs lack adequate **access to technology**

  - In court, one must be able to show the algorithm disfavors a protected group at a **statistically significant** level

  - Without information on an algorithm's processes or output patterns, an individual is **unlikely to even know** they've been discriminated against

- In areas with high-stakes decision making, future policy should:

  - Require more **transparency**

  - Implement a **review process** and accuracy **standard**

  - …

- *Weapons of Math Destruction* by Cathy O'Neil

- TED Talk: *How I'm fighting bias in algorithms* by Joy Buolamwini

# Questions?

shreygupta.me/ssi-algorithmic-bias