

# AI: Strategy + Marketing (MGT 853)

## AI in Healthcare (Session 6)

Vineet Kumar

Yale School of Management  
Spring 2024

# Project

- Everyone has chosen projects now, right?
- Each group meets with me next week
- Choose a 10 minute slot that works for \*ALL\* group members
- Read the paper and have a preliminary plan before we meet
  - introduction and results fully, skim other parts
- In class: Each group gets 12 minutes to present + 3 minutes of Q&A and Discussion
- Asking questions of other groups also counts in participation

# Beyond Transparency

- Researchers were comparing models for predicting likelihood of death for pneumonia patients in hospitals

# Beyond Transparency

- Researchers were comparing models for predicting likelihood of death for pneumonia patients in hospitals
- Rule based learning indicated that patients with **Asthma** were less likely to die

# Beyond Transparency

- Researchers were comparing models for predicting likelihood of death for pneumonia patients in hospitals
- Rule based learning indicated that patients with **Asthma** were less likely to die
- Puzzling from a medical perspective!

# Beyond Transparency

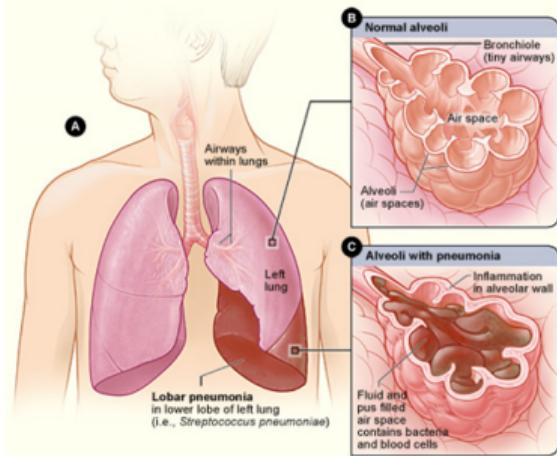
- Researchers were comparing models for predicting likelihood of death for pneumonia patients in hospitals
- Rule based learning indicated that patients with **Asthma** were less likely to die
- Puzzling from a medical perspective!
- Why does this happen?

# Beyond Transparency

- Researchers were comparing models for predicting likelihood of death for pneumonia patients in hospitals
- Rule based learning indicated that patients with **Asthma** were less likely to die
- Puzzling from a medical perspective!
- Why does this happen?

# Beyond Transparency

- Researchers were comparing models for predicting likelihood of death for pneumonia patients in hospitals
  - Rule based learning indicated that patients with **Asthma** were less likely to die
  - Puzzling from a medical perspective!
  - Why does this happen?



# Gaming in Learning (Adversarial Learning)

# Gaming in Learning

- User browsing a retailer's website

# Gaming in Learning

- User browsing a retailer's website
- Retailer observes a number of different behaviors  $X$

# Gaming in Learning

- User browsing a retailer's website
  - Retailer observes a number of different behaviors  $X$
  - Wants to figure out which users to give a coupon to

# Gaming in Learning

- User browsing a retailer's website
- Retailer observes a number of different behaviors  $X$
- Wants to figure out which users to give a coupon to
- Specify this as a prediction problem  $y = f(X)$

# Gaming in Learning

- User browsing a retailer's website
- Retailer observes a number of different behaviors  $X$
- Wants to figure out which users to give a coupon to
- Specify this as a prediction problem  $y = f(X)$

# Gaming in Learning

- User browsing a retailer's website
  - Retailer observes a number of different behaviors  $X$
  - Wants to figure out which users to give a coupon to
  - Specify this as a prediction problem  $y = f(X)$

# Customer Journeys Are Complex



# Gaming in Learning

- User browsing a retailer's website
  - Retailer observes a number of different behaviors  $X$
  - Wants to figure out which users to give a coupon to
  - Specify this as a prediction problem  $y = f(X)$

## Customer Journeys Are Complex



So, what's the problem?

# Ethics in AI

# Algorithmic Decisions

- Where do Algorithms make decisions in business + society?

# Algorithmic Decisions

- Where do Algorithms make decisions in business + society?
    - Human Capital: Resume screening, University admissions / enrollment

# Algorithmic Decisions

- Where do Algorithms make decisions in business + society?
  - Human Capital: Resume screening, University admissions / enrollment
  - Medical Care: Which patients to monitor more intensively or even escalate care?

# Algorithmic Decisions

- Where do Algorithms make decisions in business + society?
  - Human Capital: Resume screening, University admissions / enrollment
  - Medical Care: Which patients to monitor more intensively or even escalate care?
  - Loans: Which customers to approve for a home / auto / personal loan?

# Algorithmic Decisions

- Where do Algorithms make decisions in business + society?
  - Human Capital: Resume screening, University admissions / enrollment
  - Medical Care: Which patients to monitor more intensively or even escalate care?
  - Loans: Which customers to approve for a home / auto / personal loan?
  - Criminal Justice: Pre-trial bail

# Algorithmic Decisions

- Where do Algorithms make decisions in business + society?
  - Human Capital: Resume screening, University admissions / enrollment
  - Medical Care: Which patients to monitor more intensively or even escalate care?
  - Loans: Which customers to approve for a home / auto / personal loan?
  - Criminal Justice: Pre-trial bail

# Algorithmic Decisions

- Where do Algorithms make decisions in business + society?
  - Human Capital: Resume screening, University admissions / enrollment
  - Medical Care: Which patients to monitor more intensively or even escalate care?
  - Loans: Which customers to approve for a home / auto / personal loan?
  - Criminal Justice: Pre-trial bail

Major Challenge:

**Algorithmic Bias leads to differential performance across groups  
(think race, gender, age, income)**

Humans may be able to override decisions, but...

- often have limited time / energy / attention
- power of defaults from marketing and psychology

# Impact of Bias: Human Capital

RETAIL OCTOBER 10, 2018 / 7:04 PM / UPDATED 4 YEARS AGO

## Amazon scraps secret AI recruiting tool that showed bias against women

By Jeffrey Dastin

8 MIN READ



SAN FRANCISCO (Reuters) - Amazon.com Inc's [AMZN.O](#) machine-learning specialists uncovered a big problem: their new recruiting engine did not like women.

# Impact of Bias: Human Capital

RETAIL OCTOBER 10, 2018 / 7:04 PM / UPDATED 4 YEARS AGO

## Amazon scraps secret AI recruiting tool that showed bias against women

By Jeffrey Dastin

8 MIN READ



SAN FRANCISCO (Reuters) - Amazon.com Inc's [AMZN.O](#) machine-learning specialists uncovered a big problem: their new recruiting engine did not like women.

### ML Problem:

Prediction Problem: Predict quality or fit of applicant (one to five stars)

Input to algorithm: Resume

Decision: Interview or Not

Bias: Women-related words decreased stars

# Impact of Bias: Human Capital



REPORT

**Enrollment algorithms are contributing to the crises of higher education**

Alex Engler · Tuesday, September 14, 2021

# Impact of Bias: Human Capital



REPORT

**Enrollment algorithms are contributing to the crises of higher education**

Alex Engler · Tuesday, September 14, 2021

## ML Problem:

Prediction problem (Y): Likelihood of accepting

Input to algorithm (X): Student information

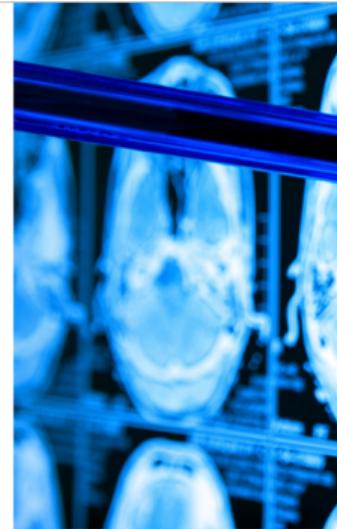
Decision: How much financial aid to offer

Bias: More accurate for higher income

# Impact of Bias: Medical Care

## Algorithms Are Making Decisions About Health Care, Which May Only Worsen Medical Racism

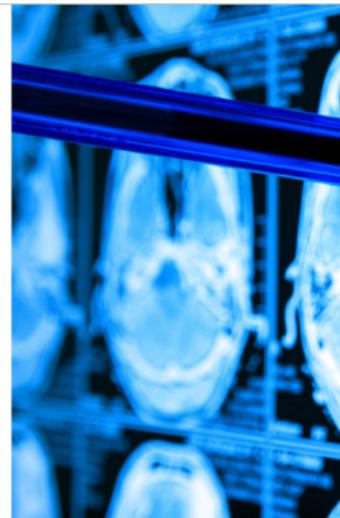
Unclear regulation and a lack of transparency increase the risk that AI and algorithmic tools that exacerbate racial biases will be used in medical settings.



# Impact of Bias: Medical Care

## Algorithms Are Making Decisions About Health Care, Which May Only Worsen Medical Racism

Unclear regulation and a lack of transparency increase the risk that AI and algorithmic tools that exacerbate racial biases will be used in medical settings.



### ML Problem:

Prediction problem (Y): who is likely to have a serious condition

Input to algorithm (X): insurance claims, diagnosis codes, etc.

Decision: extra medical attention and care

Bias: For same risk assessment, Black patients sicker than White patients

# Impact of Bias: Criminal Justice



Machine Bias

# Impact of Bias: Criminal Justice



## ML Problem:

Prediction problem (Y): likelihood of re-offending

Input to algorithm (X): 137 question survey

Decision: Offer Bail or Not

Bias: Higher FPR among Blacks

- Is Prediction Accuracy a good metric?
- Accuracy cannot distinguish between FP and FN (Type 1 and Type 2)

# Where does Algorithmic Bias come in?

# Concepts of Fairness in ML

# Fairness through Unawareness – 1

## Orchestrating Impartiality: The Impact of “Blind” Auditions on Female Musicians

By CLAUDIA GOLDIN AND CECILIA ROUSE\*

*A change in the audition procedures of symphony orchestras—adoption of “blind” auditions with a “screen” to conceal the candidate’s identity from the jury—provides a test for sex-biased hiring. Using data from actual auditions, in an individual fixed-effects framework, we find that the screen increases the probability a woman will be advanced and hired. Although some of our estimates have large standard errors and there is one persistent effect in the opposite direction, the weight of the evidence suggests that the blind audition procedure fostered impartiality in hiring and increased the proportion women in symphony orchestras.* (JEL J7, J16)

Sex-biased hiring has been alleged for many occupations but is extremely difficult to prove. The empirical literature on discrimination, de-

riving from the seminal contributions of Gary Becker (1971) and Kenneth Arrow (1973), has focused mainly on disparities in earnings between groups (e.g., males and females), given differences in observable productivity-altering characteristics. With the exception of various audit studies (e.g., Genevieve Kenney and Douglas A. Wissoker, 1994; David Neumark et al. 1996) and others, few researchers have been

# Fairness through Unawareness – 2

The New York Times

CRITIC'S NOTEBOOK

## To Make Orchestras More Diverse, End Blind Auditions

If ensembles are to reflect the communities they serve, the audition process should take into account race, gender and other factors.

Share full article



# Can we satisfy all concepts of fairness?

## The (Im)possibility of Fairness: Different Value Systems Require Different Mechanisms For Fair Decision Making

By Sorelle A. Friedler, Carlos Scheidegger, Suresh Venkatasubramanian

Communications of the ACM, April 2021, Vol. 64 No. 4, Pages 136-143

10.1145/3433949

Comments



Automated decision-making systems (often machine learning-based) now commonly determine criminal sentences, hiring choices, and loan applications. This widespread deployment is concerning, since these systems have the potential to discriminate against people based on their demographic characteristics. Current sentencing risk assessments are racially biased,<sup>4</sup> and job advertisements discriminate on gender.<sup>8</sup> These concerns have led to an explosive growth in fairness-aware machine learning, a field that aims to enable algorithmic systems that are fair by design.

[Back to Top](#)

Kev Insights

**SIGN IN** for Full Access

User Name

Password

[» Forgot Password?](#)

[» Create an ACM Web Account](#)

**SIGN IN**

**ARTICLE CONTENTS:**

Introduction
Key Insights
An Example
Spaces: Construct vs. Observed and Features vs. Decisions
Fairness and Non-Discrimination
Worldviews and Assumptions

It is impossible to simultaneously satisfy all the notions of fairness

# How can fairness impact profitability?

- You're a bank making loans to 2 groups of people: (O)range and (B)lue

# How can fairness impact profitability?

- You're a bank making loans to 2 groups of people: (O)range and (B)lue
- A successful loan makes \$300

# How can fairness impact profitability?

- You're a bank making loans to 2 groups of people: (O)range and (B)lue
- A successful loan makes \$300
- An unsuccessful loan costs \$700

# How can fairness impact profitability?

- You're a bank making loans to 2 groups of people: (O)range and (B)lue
  - A successful loan makes \$300
  - An unsuccessful loan costs \$700
  - **What is the bank's profit?**

# How can fairness impact profitability?

- You're a bank making loans to 2 groups of people: (O)range and (B)lue
- A successful loan makes \$300
- An unsuccessful loan costs \$700
- **What is the bank's profit?**
- Everyone has a credit score between 0 and 100

# How can fairness impact profitability?

- You're a bank making loans to 2 groups of people: (O)range and (B)lue
- A successful loan makes \$300
- An unsuccessful loan costs \$700
- **What is the bank's profit?**
- Everyone has a credit score between 0 and 100
- Threshold classifier: above / below the bar?

# How can fairness impact profitability?

- You're a bank making loans to 2 groups of people: (O)range and (B)lue
- A successful loan makes \$300
- An unsuccessful loan costs \$700
- **What is the bank's profit?**
- Everyone has a credit score between 0 and 100
- Threshold classifier: above / below the bar?

# How can fairness impact profitability?

- You're a bank making loans to 2 groups of people: (O)range and (B)lue
- A successful loan makes \$300
- An unsuccessful loan costs \$700
- **What is the bank's profit?**
- Everyone has a credit score between 0 and 100
- Threshold classifier: above / below the bar?

How to trade-off profitability with fairness?

## Simulation

See how different notions of fairness lead to different levels of profitability in this simulation:

[http:](http://)

# Confusion Matrix for Bank Loans

# Pareto Frontier

# Intersectionality of Race and Gender

Proceedings of Machine Learning Research 81:1–15, 2018

Conference on Fairness, Accountability, and Transparency

## Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification\*

Joy Buolamwini

*MIT Media Lab 75 Amherst St. Cambridge, MA 02139*

JOYAB@MIT.EDU

Timnit Gebru

*Microsoft Research 641 Avenue of the Americas, New York, NY 10011*

TIMNIT.GEBRU@MICROSOFT.COM

**Editors:** Sorelle A. Friedler and Christo Wilson

### Abstract

Recent studies demonstrate that machine learning algorithms can discriminate based on classes like race and gender. In this work, we present an approach to evaluate bias present in automated facial analysis algorithms and datasets with respect to phenotypic subgroups. Using the dermatologist approved Fitzpatrick Skin Type classification system, we characterize the gender and skin type distribution of two facial analysis benchmarks, IJB-A and Adience. We find that these datasets are overwhelmingly composed of lighter-skinned subjects

who is hired, fired, granted a loan, or how long an individual spends in prison, decisions that have traditionally been performed by humans are rapidly made by algorithms (O’Neil, 2017; Citron and Pasquale, 2014). Even AI-based technologies that are not specifically trained to perform high-stakes tasks (such as determining how long someone spends in prison) can be used in a pipeline that performs such tasks. For example, while face recognition software by itself should not be trained to determine the fate of an individual in the criminal justice system, it is very likely that such software is used to identify suspects. Thus,

# Intersectionality of Race and Gender

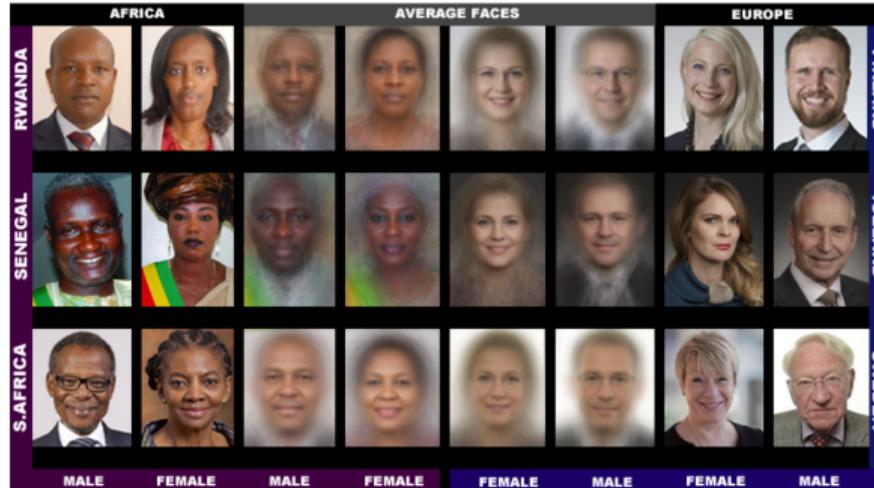


Figure 1: Example images and average faces from the new Pilot Parliaments Benchmark (PPB). As the examples show, the images are constrained with relatively little variation in pose. The subjects are composed of male and female parliamentarians from 6 countries. On average, Senegalese subjects are the darkest skinned while those from Finland and Iceland are the lightest skinned.

# Intersectionality in ML

Classifier	Metric	All	F	M	Darker	Lighter	DF	DM	LF	LM
MSFT	PPV(%)	93.7	89.3	97.4	87.1	99.3	79.2	94.0	98.3	<b>100</b>
	Error Rate(%)	6.3	10.7	2.6	12.9	0.7	<b>20.8</b>	6.0	1.7	0.0
	TPR (%)	93.7	96.5	91.7	87.1	99.3	92.1	83.7	<b>100</b>	98.7
	FPR (%)	6.3	8.3	3.5	12.9	0.7	<b>16.3</b>	7.9	1.3	0.0
Face++	PPV(%)	90.0	78.7	99.3	83.5	95.3	65.5	<b>99.3</b>	94.0	99.2
	Error Rate(%)	10.0	21.3	0.7	16.5	4.7	<b>34.5</b>	0.7	6.0	0.8
	TPR (%)	90.0	98.9	85.1	83.5	95.3	98.8	76.6	<b>98.9</b>	92.9
	FPR (%)	10.0	14.9	1.1	16.5	4.7	<b>23.4</b>	1.2	7.1	1.1
IBM	PPV(%)	87.9	79.7	94.4	77.6	96.8	65.3	88.0	92.9	<b>99.7</b>
	Error Rate(%)	12.1	20.3	5.6	22.4	3.2	<b>34.7</b>	12.0	7.1	0.3
	TPR (%)	87.9	92.1	85.2	77.6	96.8	82.3	74.8	<b>99.6</b>	94.8
	FPR (%)	12.1	14.8	7.9	22.4	3.2	<b>25.2</b>	17.7	5.20	0.4

# Fairness & Bias in ML – Takeaways 1

- ➊ Fairness can be a problem in ML if we only focus on accuracy. We can have multiple goals beyond accuracy.

# Fairness & Bias in ML – Takeaways 1

- ➊ Fairness can be a problem in ML if we only focus on accuracy. We can have multiple goals beyond accuracy.
- ➋ A lot of factors contributing to bias happens before our ML process even begins

# Fairness & Bias in ML – Takeaways 1

- ➊ Fairness can be a problem in ML if we only focus on accuracy. We can have multiple goals beyond accuracy.
- ➋ A lot of factors contributing to bias happens before our ML process even begins
- ➌ Many definitions of fairness: Statistical Parity, Equalized odds,

# Fairness & Bias in ML – Takeaways 1

- ➊ Fairness can be a problem in ML if we only focus on accuracy. We can have multiple goals beyond accuracy.
- ➋ A lot of factors contributing to bias happens before our ML process even begins
- ➌ Many definitions of fairness: Statistical Parity, Equalized odds,
  - Concepts of fairness are not equivalent

# Fairness & Bias in ML – Takeaways 1

- ➊ Fairness can be a problem in ML if we only focus on accuracy. We can have multiple goals beyond accuracy.
- ➋ A lot of factors contributing to bias happens before our ML process even begins
- ➌ Many definitions of fairness: Statistical Parity, Equalized odds,
  - Concepts of fairness are not equivalent
  - may be opposed to one another (impossible to satisfy \*all\*)

# Fairness & Bias in ML – Takeaways 1

- ① Fairness can be a problem in ML if we only focus on accuracy. We can have multiple goals beyond accuracy.
- ② A lot of factors contributing to bias happens before our ML process even begins
- ③ Many definitions of fairness: Statistical Parity, Equalized odds,
  - Concepts of fairness are not equivalent
  - may be opposed to one another (impossible to satisfy \*all\*)
- ④ Pareto frontier characterizes the tradeoff between accuracy and fairness

# Fairness & Bias in ML – Takeaways 1

- ➊ Fairness can be a problem in ML if we only focus on accuracy. We can have multiple goals beyond accuracy.
- ➋ A lot of factors contributing to bias happens before our ML process even begins
- ➌ Many definitions of fairness: Statistical Parity, Equalized odds,
  - Concepts of fairness are not equivalent
  - may be opposed to one another (impossible to satisfy \*all\*)
- ➍ Pareto frontier characterizes the tradeoff between accuracy and fairness
  - If within frontier, may not be a tradeoff. At frontier, there is a tradeoff.

# Fairness & Bias in ML – Takeaways 1

- ➊ Fairness can be a problem in ML if we only focus on accuracy. We can have multiple goals beyond accuracy.
- ➋ A lot of factors contributing to bias happens before our ML process even begins
- ➌ Many definitions of fairness: Statistical Parity, Equalized odds,
  - Concepts of fairness are not equivalent
  - may be opposed to one another (impossible to satisfy \*all\*)
- ➍ Pareto frontier characterizes the tradeoff between accuracy and fairness
  - If within frontier, may not be a tradeoff. At frontier, there is a tradeoff.
- ➎ Where can bias enter into an ML system?

# Fairness & Bias in ML – Takeaways 1

- ➊ Fairness can be a problem in ML if we only focus on accuracy. We can have multiple goals beyond accuracy.
- ➋ A lot of factors contributing to bias happens before our ML process even begins
- ➌ Many definitions of fairness: Statistical Parity, Equalized odds,
  - Concepts of fairness are not equivalent
  - may be opposed to one another (impossible to satisfy \*all\*)
- ➍ Pareto frontier characterizes the tradeoff between accuracy and fairness
  - If within frontier, may not be a tradeoff. At frontier, there is a tradeoff.
- ➎ Where can bias enter into an ML system?
  - Data (Training, Validation and Test)

# Fairness & Bias in ML – Takeaways 1

- ➊ Fairness can be a problem in ML if we only focus on accuracy. We can have multiple goals beyond accuracy.
- ➋ A lot of factors contributing to bias happens before our ML process even begins
- ➌ Many definitions of fairness: Statistical Parity, Equalized odds,
  - Concepts of fairness are not equivalent
  - may be opposed to one another (impossible to satisfy \*all\*)
- ➍ Pareto frontier characterizes the tradeoff between accuracy and fairness
  - If within frontier, may not be a tradeoff. At frontier, there is a tradeoff.
- ➎ Where can bias enter into an ML system?
  - Data (Training, Validation and Test)
  - Algorithm

# Fairness & Bias in ML – Takeaways 2

- ① Intersectionality: Combinations of groups (intersections) may face challenges.

# Fairness & Bias in ML – Takeaways 2

- ① Intersectionality: Combinations of groups (intersections) may face challenges.
    - Can be a challenge to look at all possible intersections (think gender, race, age, income,...)

# Fairness & Bias in ML – Takeaways 2

- ➊ Intersectionality: Combinations of groups (intersections) may face challenges.
  - Can be a challenge to look at all possible intersections (think gender, race, age, income,...)
- ➋ Interventions to reduce bias:

# Fairness & Bias in ML – Takeaways 2

- ➊ Intersectionality: Combinations of groups (intersections) may face challenges.
  - Can be a challenge to look at all possible intersections (think gender, race, age, income,...)
- ➋ Interventions to reduce bias:
  - Pre-processing

# Fairness & Bias in ML – Takeaways 2

- ➊ Intersectionality: Combinations of groups (intersections) may face challenges.
  - Can be a challenge to look at all possible intersections (think gender, race, age, income,...)
- ➋ Interventions to reduce bias:
  - Pre-processing
  - In-processing

# Fairness & Bias in ML – Takeaways 2

- ➊ Intersectionality: Combinations of groups (intersections) may face challenges.
  - Can be a challenge to look at all possible intersections (think gender, race, age, income,...)
- ➋ Interventions to reduce bias:
  - Pre-processing
  - In-processing
  - Post-processing

# Fairness & Bias in ML – Takeaways 2

- ➊ Intersectionality: Combinations of groups (intersections) may face challenges.
  - Can be a challenge to look at all possible intersections (think gender, race, age, income,...)
- ➋ Interventions to reduce bias:
  - Pre-processing
  - In-processing
  - Post-processing
- ➌ Fairness is not the only issue wrt AI and society  $\implies$  Responsible AI

# Pop Quiz (0 Points)

**Multiple choice: More than one option  
may be true**

# Q1

- You're working with an online retailer who has browsing and purchase data for users. They would like you to identify which consumers have been most responsive to their messaging promotions over the past year. Which of the following approaches could be helpful here?
  - ① Unsupervised
  - ② Supervised
  - ③ Reinforcement

## Q2

- Which of these is an example of reducing model complexity?
  - ➊ Deciding not to use certain features  $X$  in a prediction problem
  - ➋ Including interaction terms 
$$y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1 X_2$$
  - ➌ Dropping out neurons in a neural network
  - ➍ Penalizing models that have too many parameters

# Q3

- Which of the following weather related problems would Unsupervised Learning **\*not\*** be suitable for?
  - ① Identify features for inclusion in prediction problems
  - ② Predict tomorrow's weather
  - ③ Determine which cities have similar weather patterns
  - ④ Discover a rule like "If it has rained today, and it is sunny now, we're likely to see a rainbow"

## Q4

- Suppose we're predicting prices for cars based on visual features using images, in addition to characteristics like mpg, hp etc. You have annual data from 2010-2014. You split the data into training and test samples to predict prices and use Deep Learning to predict sales prices. Which options below are correct?
  - ① Data splitting is not helpful because we are not leveraging all the data
  - ② Data splitting is a good practice because of model complexity
  - ③ Data splitting is helpful because prediction on test data provides the best estimate
  - ④ There is data leakage when we split the data by years

## Q5

- Observe the following reinforcement learning example.  
Can you specify the following?

- State  $s$
- Action  $a$
- Rewards  $R$

## Q5

- Observe the following reinforcement learning example. Can you specify the following?

- State  $s$
- Action  $a$
- Rewards  $R$



- States  $s$  could include the type of chess piece at each square and whom it belongs to (white or black)
- Action  $a$  involves the set of all possible moves that can be played by a particular player
- Rewards  $R$ : This could either be defined as being based on every move (dense rewards) or only for eventually winning the game (sparse rewards)