# Ethics and Fairness

School of Information Studies
Syracuse University

# Ethics?

***What is ethics?***

- Ethics is concerned with ***"protecting*** and advancing central ***human values,*** such as life, health, security, happiness, freedom, knowledge, resources, power, and opportunity."

***Example issue***

- Linking data sets that identifies people from multiple anonymous data sets

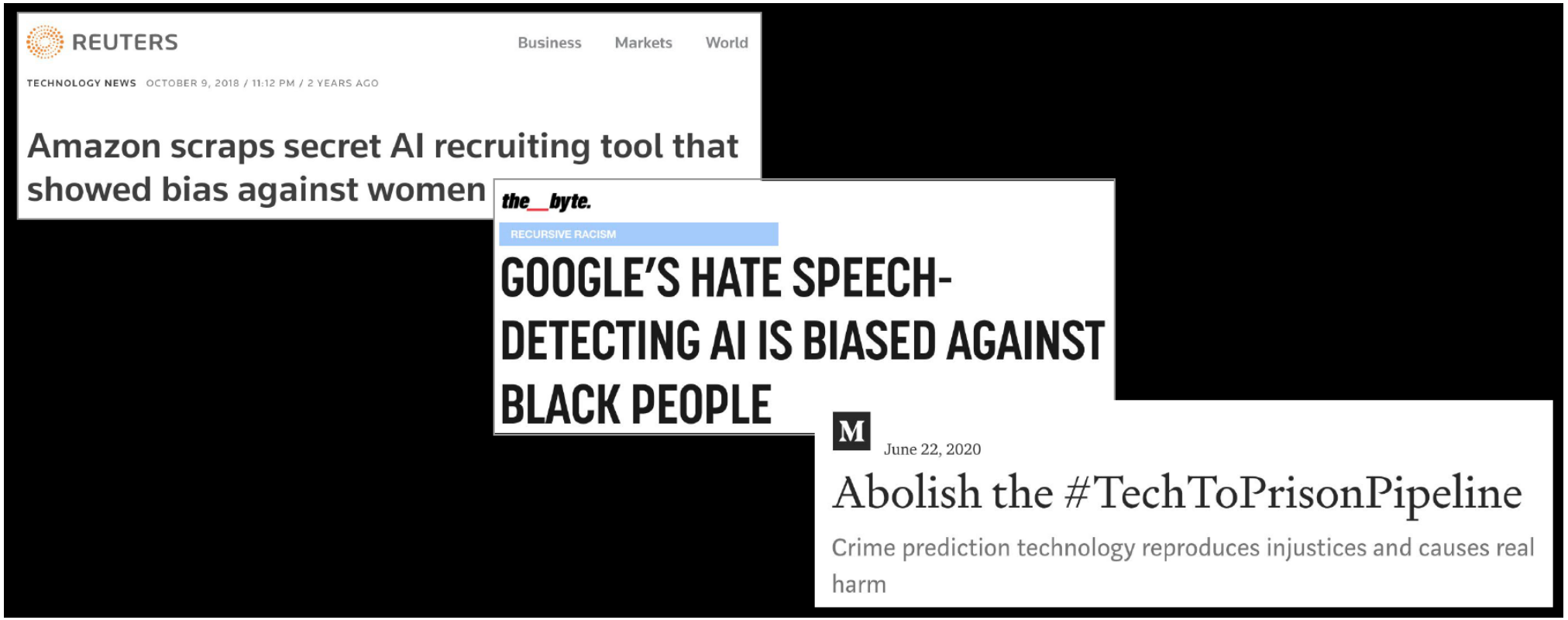
# Question

**Why focus/think about ethics?**

# Ethics and Fairness (cont.)

School of Information Studies
Syracuse University

# Examples of Models Gone Wrong



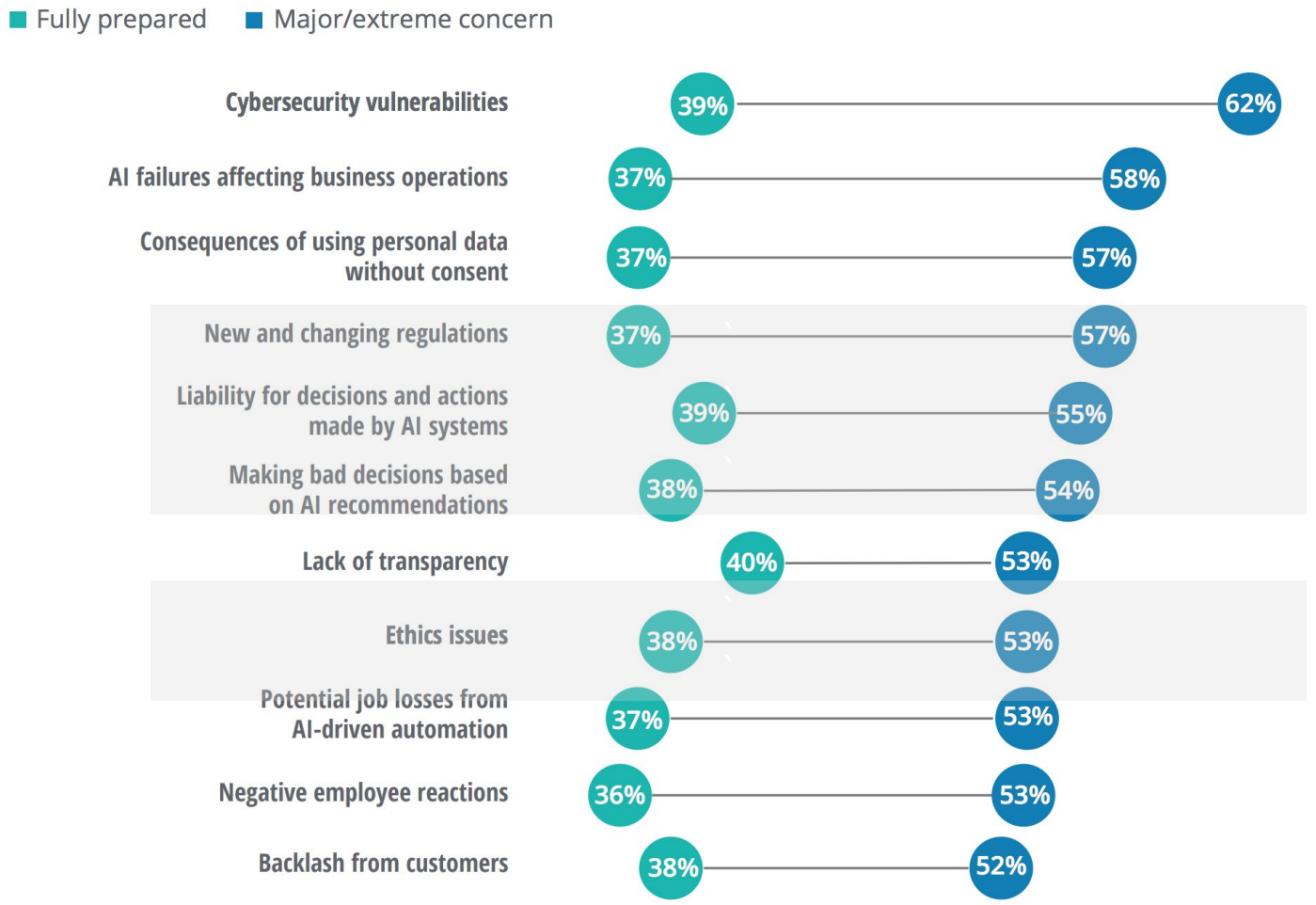
# The Need for Data Science Ethics

**Some examples of "things to think about"**

- Just because data is available does not mean it's ethical to use the data.
- Is there subjectivity in building/using models?
- How handling missing data can bias results.
- Are some combination of attributes causing bias in our model?

76% of data science (DS) professionals think that DS education should include ethics.

# Ethics/Risks: Industry Perceptions




Source: Deloitte, *State of AI in the Enterprise, 3rd Edition*, 2020.

# Focus of Ethics Discussion

**Avoid potential issues in the creation and use of the model**

- Help ensure an ethical analysis (e.g., no bias)
- Key questions during the life of the project
- Not focused on societal issues…
  - But on what the team should consider

# General Standards of Conduct

Professional conduct

Duty to client

Duty to colleagues/industry

Nothing really new but **do not forget**

# Data-Related Challenges

## Privacy and anonymity

- Ensure that personal information remains private
- Data donor information remains anonymous
- Aggregation and linking bring new challenges

## Data misuse

- Use data in the spirit of how the data provider intended
  - Just because data is available doesn't mean it is ethical to use it.

## Data accuracy and validity

- How to ensure accurate/valid data?

# Model-Related Challenges

## Personal and group harm

- Models can perpetuate and amplify bias
  (e.g., model built using data that records a bias)
  - Can lead to a group of people being disadvantaged

## Subjective model design

- Subjective decisions
  - What algorithm to use
  - What data to use and how to treat missing data

## Model misuse/misinterpretation

- Ensure everyone understands the quality of the model (e.g., prediction quality)

# How to Integrate Into a Project?

Key ethical considerations by phase of a project

| Project Phase | Key Ethical Themes | Ethical Considerations |
|---|---|---|
| Business Understanding | Project Initiation / Management Challenges | Personal and Group Harm |
| | | Team Accountability |
| Data Understanding | Data Challenges | Data Misuse |
| Data Preparation | | Data Privacy & Anonymity |
| | | Data Accuracy |
| Modeling | Model Challenges | Personal and Group Harm |
| Evaluation | | Subjective Model Design |
| Deployment | | Misuse / Misinterpretation |

# Example Ethics Questions

***Project initiation and management-related challenges***

**Q1: Which laws and regulations might be applicable to our project?**

It is important to consider which laws and regulations might be relevant.

**Q2: How are we achieving ethical accountability?**

It should be clear who will be accountable to minimize the potential harm.

# Example Ethics Questions (Cont.)

*Data-related challenges*

**Q3: How might the legal rights of an individual be impinged by our use of data?**

The project must have the right to use the data for that purpose.

**Q4: How might individuals' privacy and anonymity be impinged?**

How to ensure anonymity must be reexamined due to aggregations and linking.

**Q5: How do we know that the data is ethically available for its intended use?**

Being able to access data does not mean that it is ethical to use that data.

**Q6: How do we know that the data is valid for its intended use?**

This includes data accuracy and imputing missing values or excluding records.

# Example Ethics Questions (cont.)

*Model-related challenges*

**Q7: How have we identified and minimized any bias in the data or in the model?**
Models built using biased data can also be biased.

**Q8: How was any potential modeler bias identified, and if appropriate, mitigated?**
There could be subjectivity within the model building process.

**Q9: How transparent does the model need to be and how is transparency achieved?**
How important it is that the model can be explained and understood.

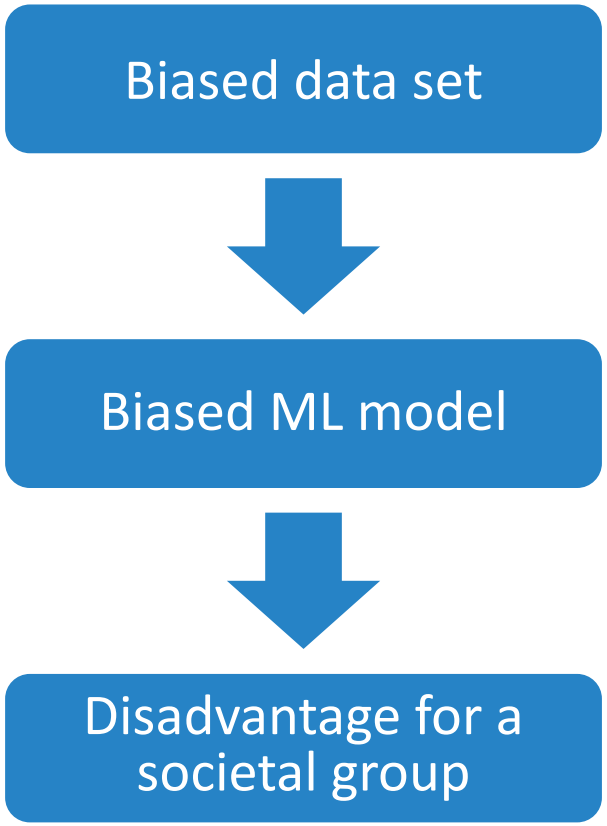
**Q10: What are likely misinterpretations of the results and what can be done to prevent those misinterpretations?**
The decisions made via an ML model should reflect the scale, accuracy, and precision of the data that was used and the resulting model.

# Biases in Data Sets

**Historical bias:**
Existing bias in the data and socio-technical issues in the world

**Representation bias:**
Comes from the way we define and sample a population (or data set)

**Measurement bias:**
Comes from the way we choose, utilize, and measure a particular feature



# Fairness

There is no standard definition of what is fair *(whether decisions are made by humans or machines).*

## For example: Which is most fair to give a loan?

- Ensure loans are made at the same rate to two different groups?
  or
- Focus on each person's expected payback rate and group attribute?
  or
- Focus on each person's expected payback rate, ignoring the group attribute?
  or
- Is neither of these the most fair?

***Even for situations that seem simple, people may disagree about what is fair.***

## Questions include:

- When is it fair to define a group (vs. better factoring on individual differences)?
- At what level of granularity should groups be defined?

# Four Types of Fairness

1. **Maximum profit**
   - The most profitable, since there are no constraints. But the two groups have different thresholds, meaning they are held to different standards.

2. **Group unaware**
   - Both groups have the same threshold (i.e., ignore the grouping).
     - One group will have fewer actions.
     - There might be bias in the training data
       (e.g., this might put one group at a disadvantage).

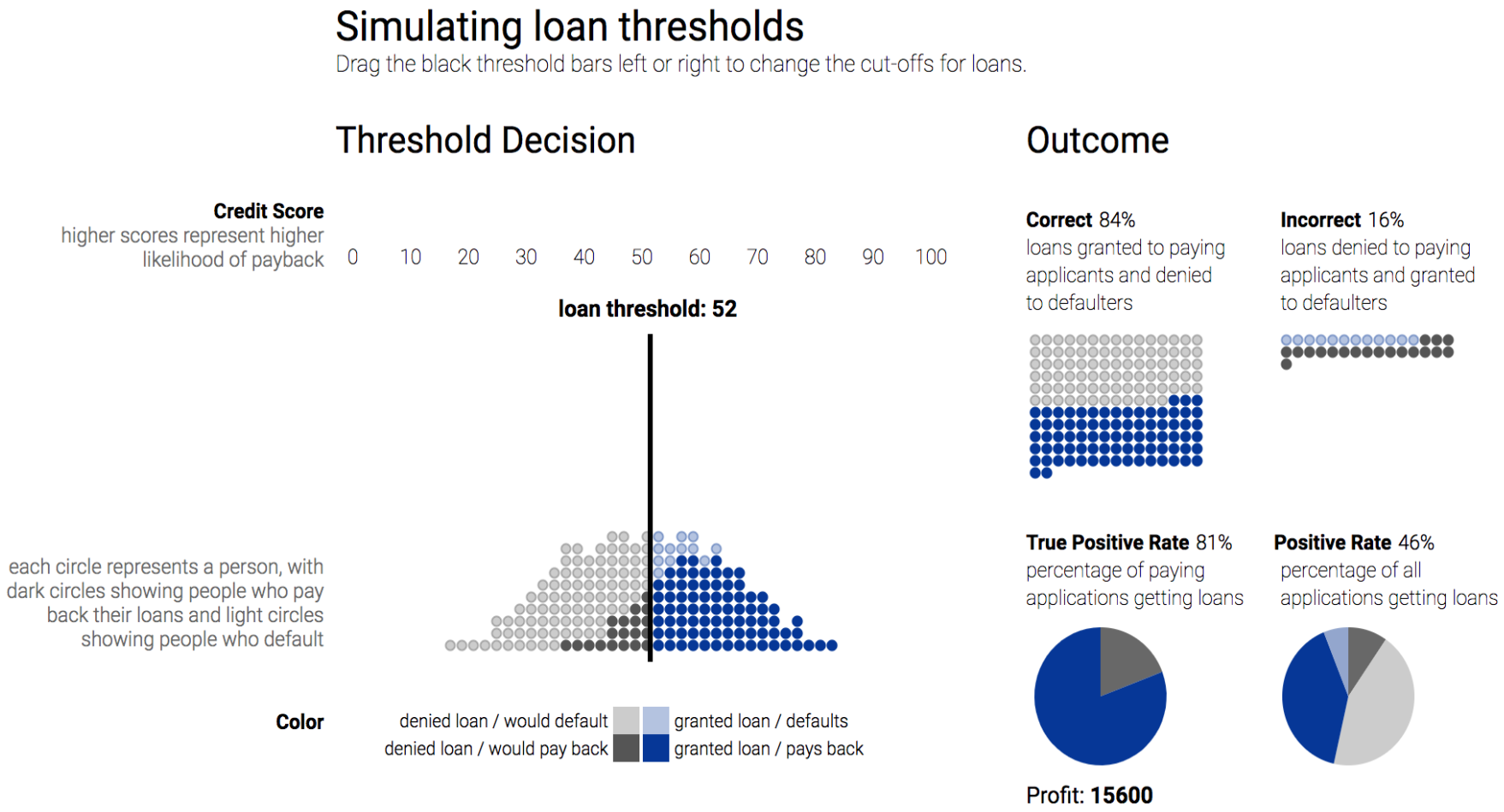
> "Individuals who qualify for a desirable outcome should have an equal chance of being correctly classified for this outcome."
> —Moritz Hardt (https:/ai.googleblog.com/2016/10/equality-of-opportunity-in-machine.html)

# Four Types of Fairness (cont.)

3. **Demographic parity**
   - The number of actions for each group is the same, but among people who qualify, one group is at a disadvantage (group with more qualified candidates).
     - Imagine group A represents 70% of the population (and group B represents 30%). If the company decides to accept 10 applicants and use demographic parity, seven will be from group A while three will be selected from group B (i.e., 10% of each group of candidates get selected).
     - What if in group A there are 60 qualified candidates (out of 100 candidates), while group B has two qualified applicants (out of the 50 candidates)? One unqualified candidate will be selected.
4. **Equal opportunity**
   - Among people who would meet the threshold, both groups do equally well (the percentage of actions for people who meet the threshold are the same).
     - Imagine group A represents 70% of the population (and group B represents 30% of the population). If group A has 60 qualified candidates (out of 100 candidates), while group B has two qualified applicants (out of the 50 candidates). If the

**Which is:** "Individuals who qualify for a desirable outcome should have an equal chance of being correctly classified for this outcome."

# Example of Fairness Evaluation

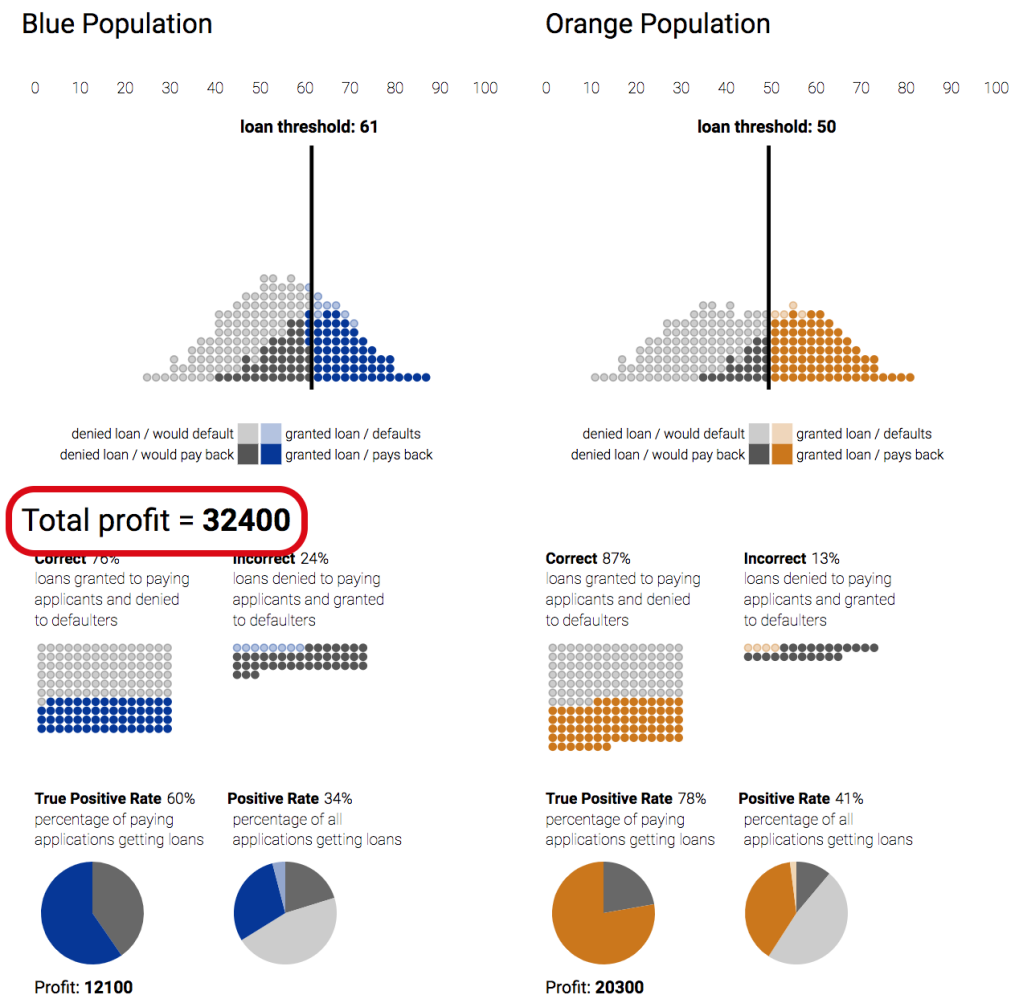



https://research.google.com/bigpicture/attacking-discrimination-in-ml/

# Maximum Profit

The most profitable, since there are no constraints.

But the two groups have different thresholds, meaning they are held to different standards.

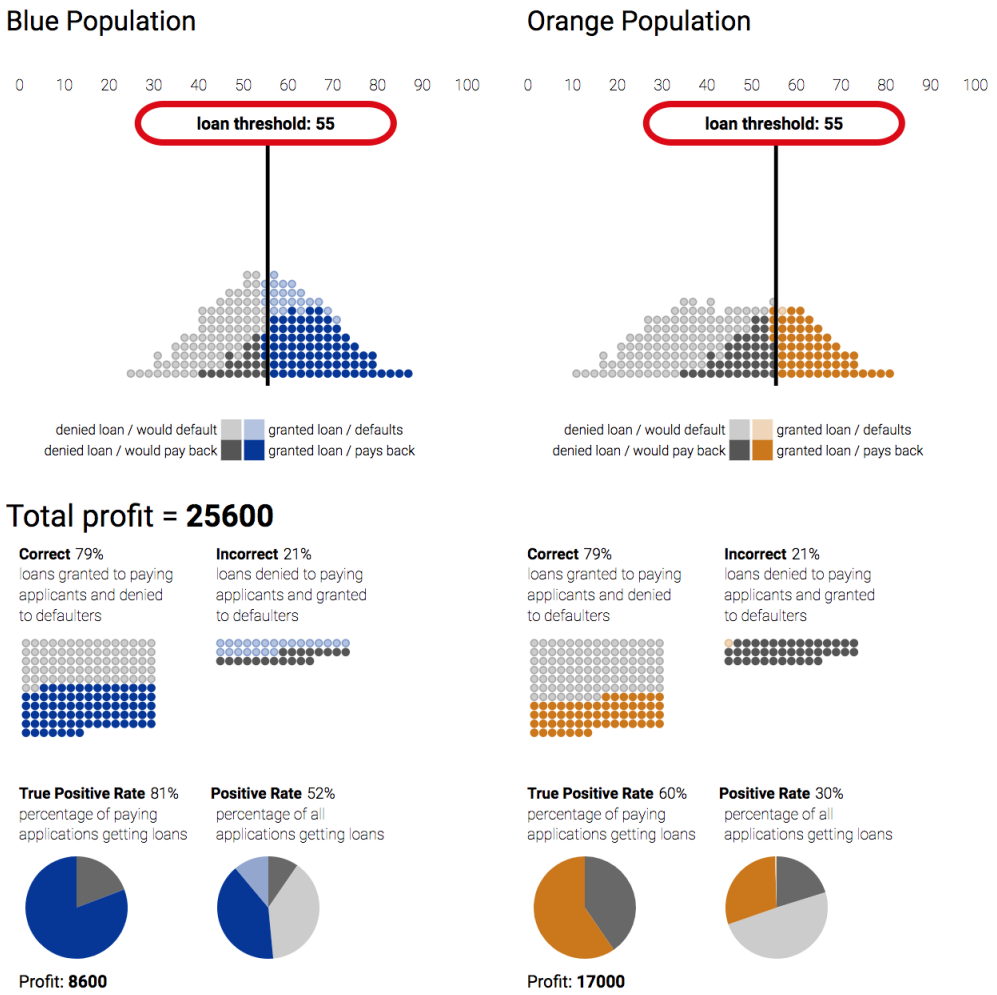


# Group Unaware

Both groups have the same threshold.

But the orange group has been given fewer loans overall.

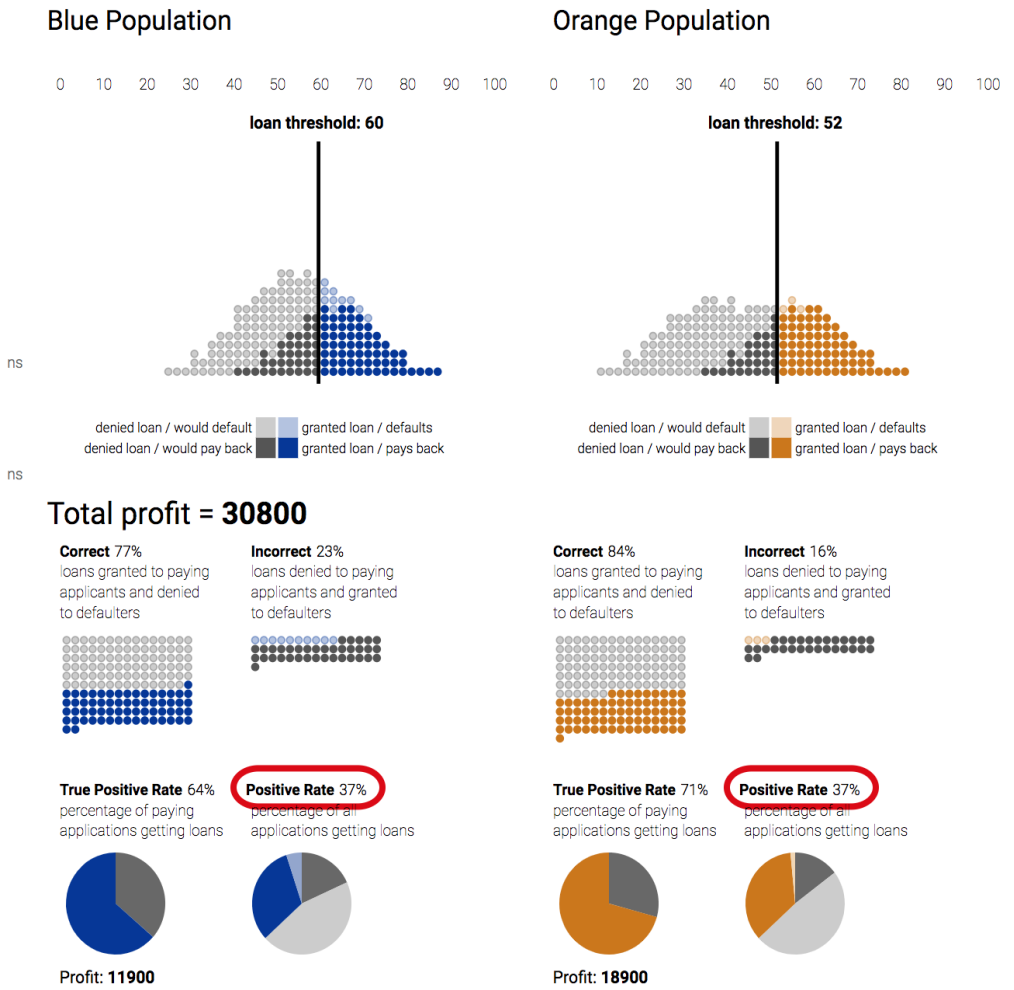
Among people who would pay back a loan, the orange group is also at a disadvantage.



# Demographic Parity

The percentage of loans given to each group is the same.

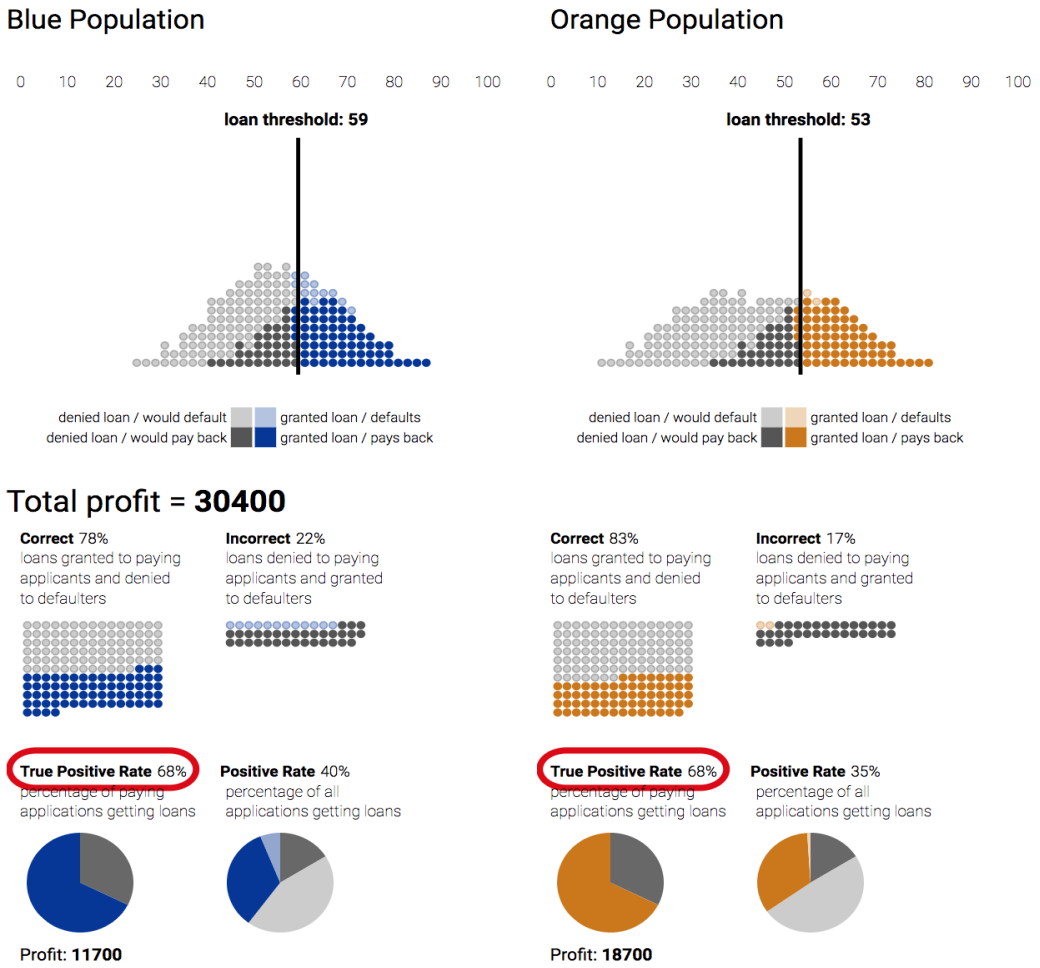
But among people who would pay back a loan, the blue group is at a disadvantage.



# Equal Opportunity

Among people who would pay back a loan, blue and orange groups do equally well.

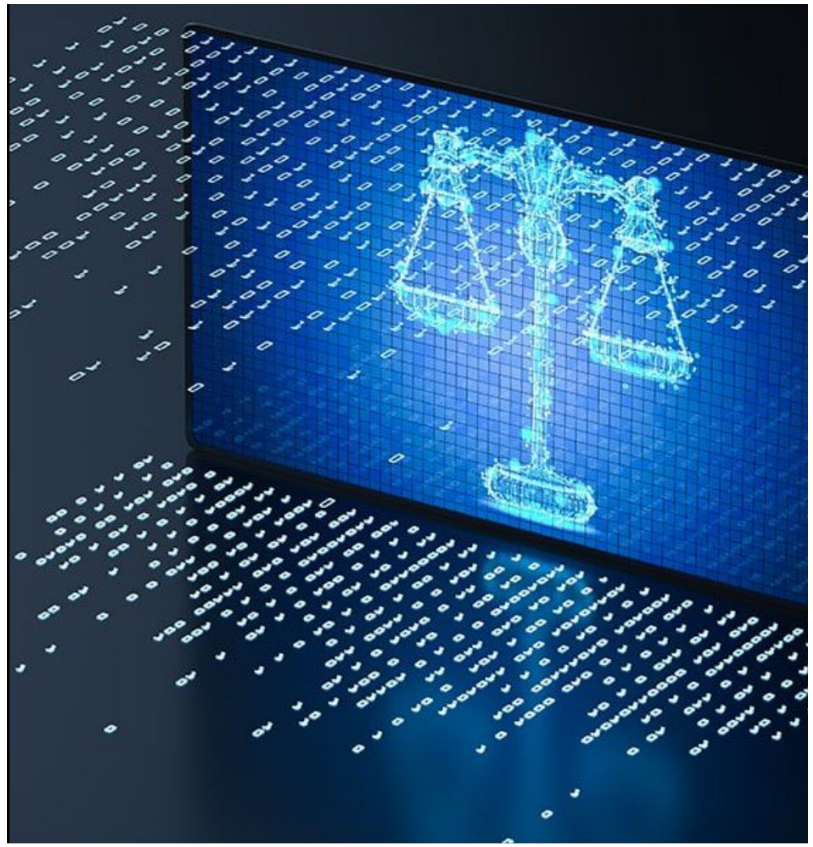
This choice is almost as profitable as demographic parity, and about as many people get loans overall.



# Fairness Is Hard

Lack consensus about which fairness to apply.

Each type of fairness requires both technical and nontechnical decisions and trade-offs.



# Ethics Questions for Discussion

1. What should I do if my manager asks me to do something—not use data incorrectly, or not check for bias in a model?

2. How can one ensure that the data sources used in predictions isn't?

3. Will having stronger data privacy laws and regulations prevent misuse of data?