



AI Ethics for Engineers

Jennifer Renoux
Örebro University

jennifer.renoux@oru.se

RETAIL OCTOBER 11, 2018 / 1:04 AM / UPDATED 2 YEARS AGO

Amazon scraps secret AI recruiting tool that showed bias against women

By Jeffrey Dastin

8 MIN READ



RETAIL OCTOBER 11, 2018 / 1:04 AM / UPDATED 2 YEARS AGO

Amazon scraps secret AI recruiting tool that showed bias against women

By Jeffrey Dastin

8 MIN READ



ANTONIO GARCÍA MARTÍNEZ

IDEAS 06.11.2019 11:08 AM

Are Facebook Ads Discriminatory? It's Complicated

The company's system for targeting ads is under fire for gender and ethnic bias. In some cases, the cure could be worse than the disease.

RETAIL OCTOBER 11, 2018 / 1:04 AM / UPDATED 2 YEARS AGO

Amazon scraps secret AI recruiting tool that showed bias against women

By Jeffrey Dastin

8 MIN READ



ANTONIO GARCÍA MARTÍNEZ

IDEAS 06.11.2019 11:08 AM

Are Facebook Ads Discriminating?

Silicon Valley Mar 20

Facebook is going to stop letting advertisers target by race, gender, or age

is complicated

The company's system for targeting ads is under fire for gender and ethnic bias. In some cases, the cure could be worse than the disease.

RETAIL OCTOBER 11, 2018 / 1:04 AM / UPDATED 2 YEARS AGO

Amazon scraps secret AI recruiting tool that showed bias against women

By Jeffrey Dastin

8 MIN READ



ANTONIO GARCÍA MARTÍNEZ

IDEAS 06.11.2019 11:08 AM

Are Facebook Ads Discriminating?

Silicon Valley Mar 20

Facebook is going to stop letting advertisers target by race, gender, or age

The company's system for targeting ads is under fire for gender and ethnic bias. In some cases, the cure could be worse than the disease.

suomi ↔ englanti

Hän on fiksu. Hän on kaunis. Hän on insinööri. Hän on siivooja. Hän urheilee. Hän tiskaa. Hän ajaa autoa. Hän hoitaa lapsia. Hän on johtaja. ✕

He's smart. She is beautiful. He is an engineer. She is a cleaner. He's playing sports. She was doing the dishes. He drives a car. She takes care of the children. He is a leader. ➔

Jennifer Renoux

RETAIL OCTOBER 11, 2018 / 1:04 AM / UPDATED 2 YEARS AGO

Amazon scraps secret AI recruiting tool that showed bias against women

By Jeffrey Dastin

8 MIN READ



ANTONIO GARCÍA MARTÍNEZ

IDEAS 06.11.2019 11:08 AM

Silicon Valley Mar 20

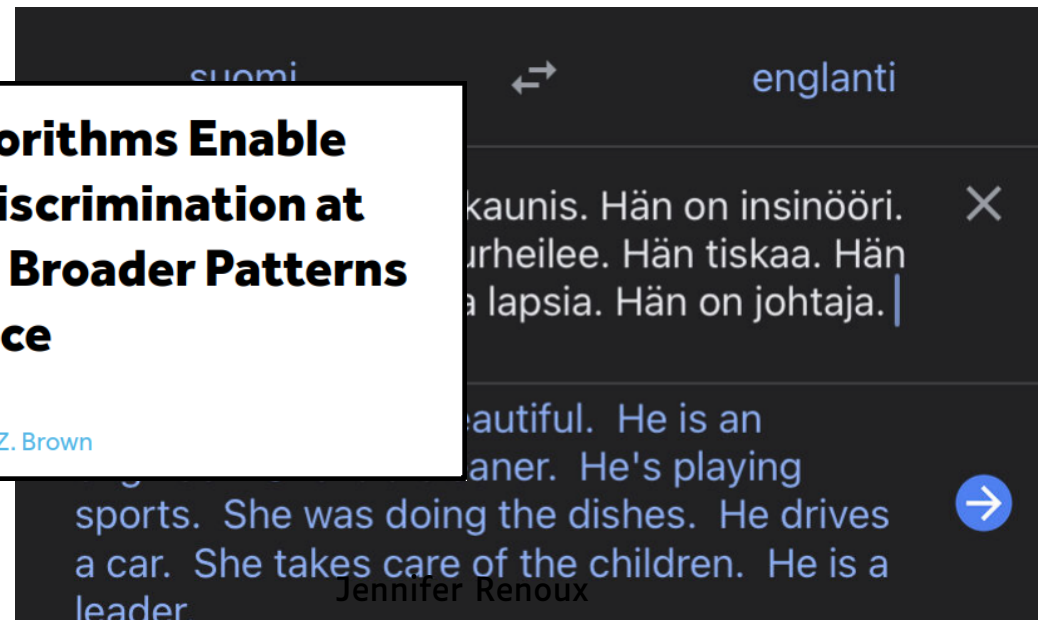
Facebook is going to stop letting advertisers target by race, gender, or age

Are Facebook Ads Discriminating? It's Complicated

The company's system for targeting ads is under fire for gender and ethnic bias. In some cases, the cure could be worse than the disease.

Tenant Screening Algorithms Enable Racial and Disability Discrimination at Scale, and Contribute to Broader Patterns of Injustice

July 7, 2021 / Lydia X. Z. Brown



RETAIL OCTOBER 11, 2018 / 1:04 AM / UPDATED 2 YEARS AGO

Amazon scraps secret AI recruiting tool that showed bias against women

By Jeffrey Dastin

8 MIN READ



ANTONIO GARCÍA MARTÍNEZ

IDEAS 06.11.2019 11:08 AM

Silicon Valley Mar 20

Facebook is going to stop letting advertisers target by race, gender, or age

Are Facebook Ads Discriminating? It's Complicated

The company's system for targeting ads is under fire for gender and ethnic bias. In some cases, the cure could be worse than the disease.

Tenant Screening Algorithms Enable Racial and Disability Discrimination at Scale, and Contribute to Broader Patterns of Injustice

July 7, 2021 / Lydia X. Z. Brown

The accuracy, fairness, and limits of predicting recidivism

Julia Dressel and Hany Farid*

sports. She was doing the dishes. He drives a car. She takes care of the children. He is a leader.

Jennifer Renoux

Intersectional Accuracy on PPB

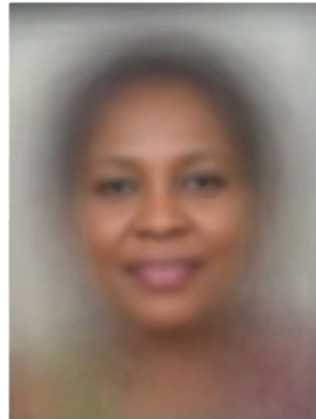


94.0%



**DARKER
MALES**

79.2%



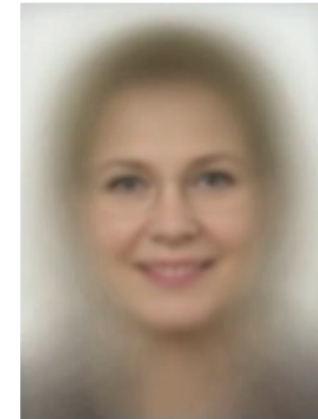
**DARKER
FEMALES**

100%



**LIGHTER
MALES**

98.3%



**LIGHTER
FEMALES**

www.gendershades.org

Intersectional Accuracy

amazon

Aug '18

98.7%

68.6%

100%

92.9%



2017 94.0%

79.2%

100%

98.3%

2018 99.7%

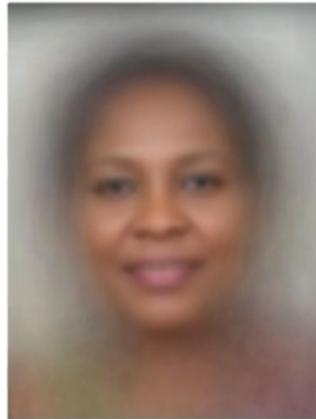
98.5%

100%

99.7%



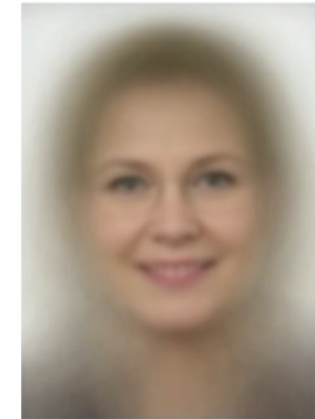
**DARKER
MALES**



**DARKER
FEMALES**

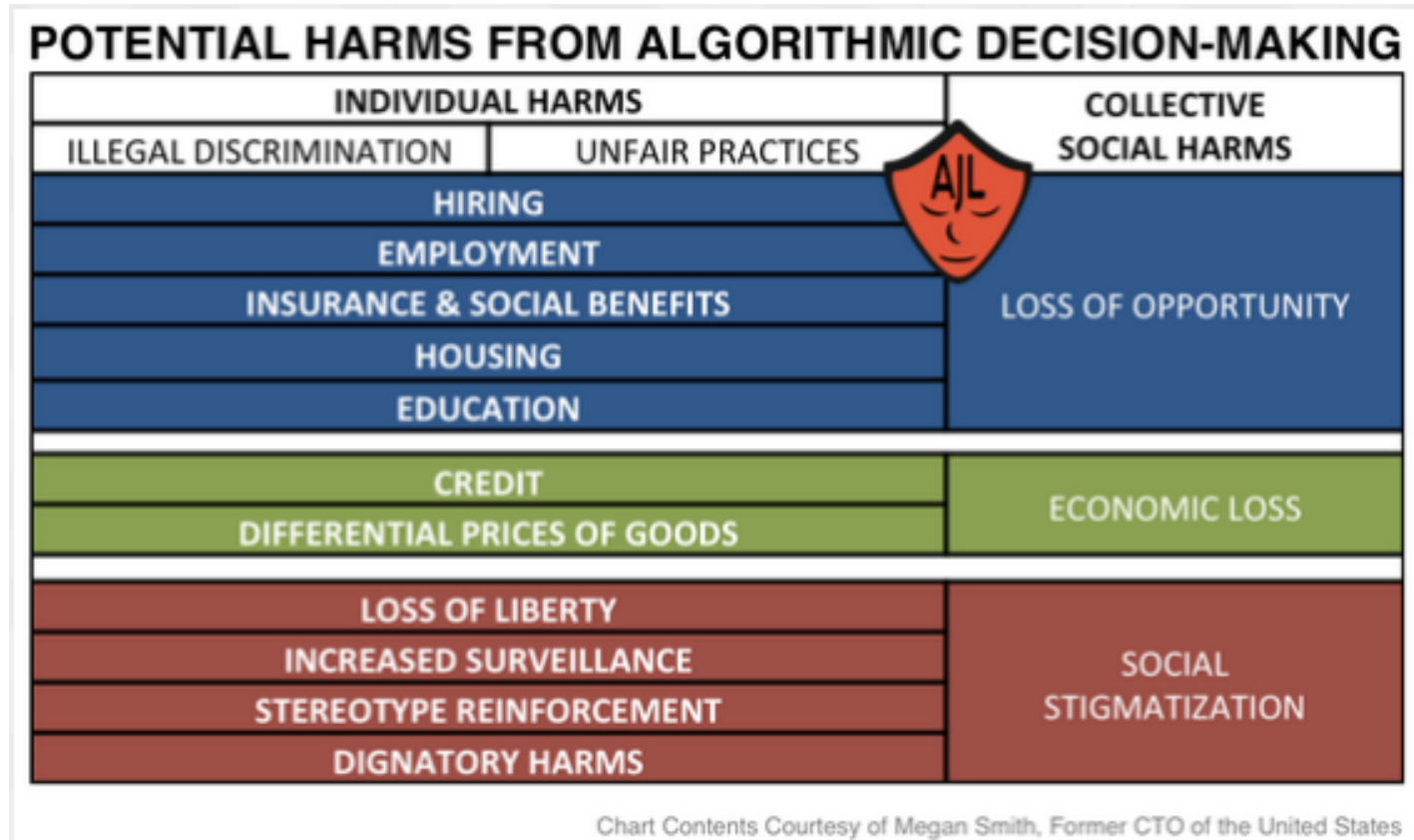


**LIGHTER
MALES**



**LIGHTER
FEMALES**

Why does it matter?



Why does it matter?

Developers of AI systems have moral responsibility to ensure that their systems are fair.

Embedded biases in systems impact the system's fairness

Some definitions

AI System:

Any system that does or assist in decision-making processes

Algorithmic Bias:

A **systematic** and **repeatable** pattern in a computer system that creates unfair outcome

What is fairness?

Individual Fairness

Individuals are treated similarly, regardless of the class they are in

Group fairness

Two classes are treated similarly

Equal opportunity

Each demographic class is offered the same opportunity

Equal outcome

Each demographic class gets the same results

Fairness through unawareness

If race and gender are deleted, system cannot discriminate

Quick Myth-busting

Algorithms are not biased. Data is.

If the system is biased, it's because the training dataset is biased.

To solve the problem of biased in AI, you simply need to unbiase the datasets / add more diversity in the datasets

Quick Myth-busting

Algorithms are not biased.

If the system is biased, the training dataset is biased.

To solve the problem of bias in AI, you simply need to unbiase the datasets with more diversity in the datasets

WRONG

Two important takes

1

Biased in AI systems is not limited to ML

Model-based AI can be biased too. Algorithms can be biased.

Creating "non-biased system" is a socio-technical challenge

"Bias" only exists related to a societal context.

2

How can systems become biased?

Biased measure

A bad heuristic is used to frame the problem, leading to biases

Biased data

The data analysed is biased, leading to a biased analysis

Conflicting costs

Some types of fairness are conflicting with each others and needs to be pondered

How can systems become biased?

Biased measure

A bad heuristic is used to frame the problem, leading to biases

Biased data

The data analysed in biased, leading to a biased analysis

Conflicting costs

Some types of fairness are conflicting with each others and needs to be pondered

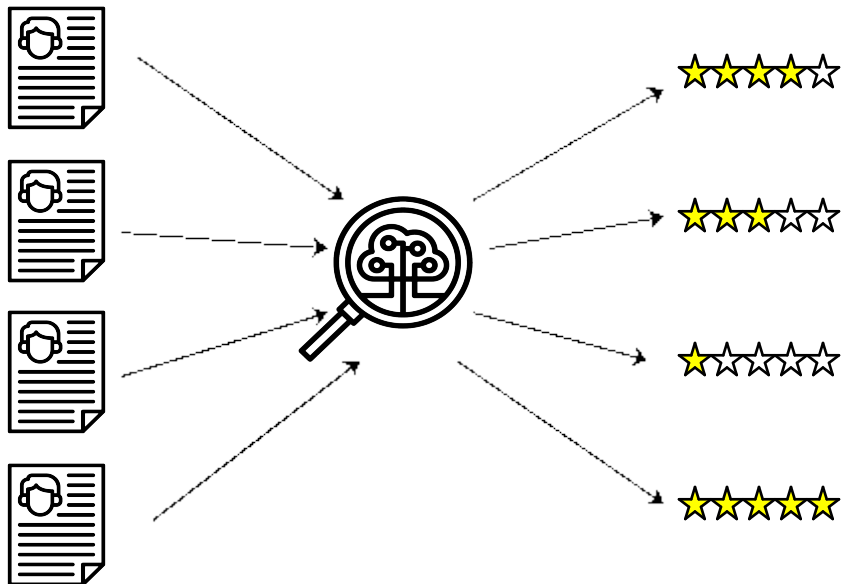

Biased data: Amazon's recruiting tool

RETAIL OCTOBER 11, 2018 / 1:04 AM / UPDATED 2 YEARS AGO

Amazon scraps secret AI recruiting tool that showed bias against women

By Jeffrey Dastin

8 MIN READ



Biased data: Amazon's recruiting tool

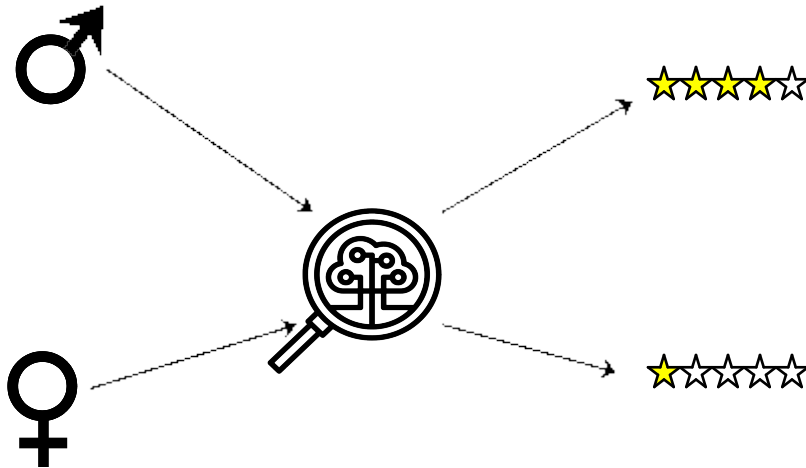
RETAIL OCTOBER 11, 2018 / 1:04 AM / UPDATED 2 YEARS AGO

Amazon scraps secret AI recruiting tool that showed bias against women

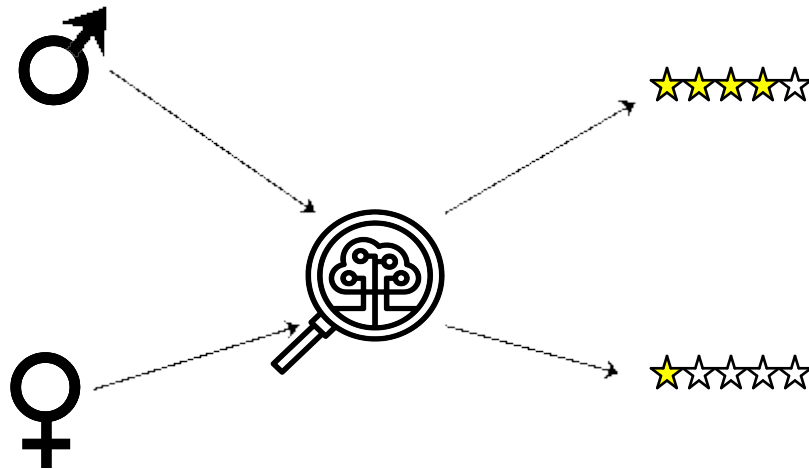
By Jeffrey Dastin

8 MIN READ

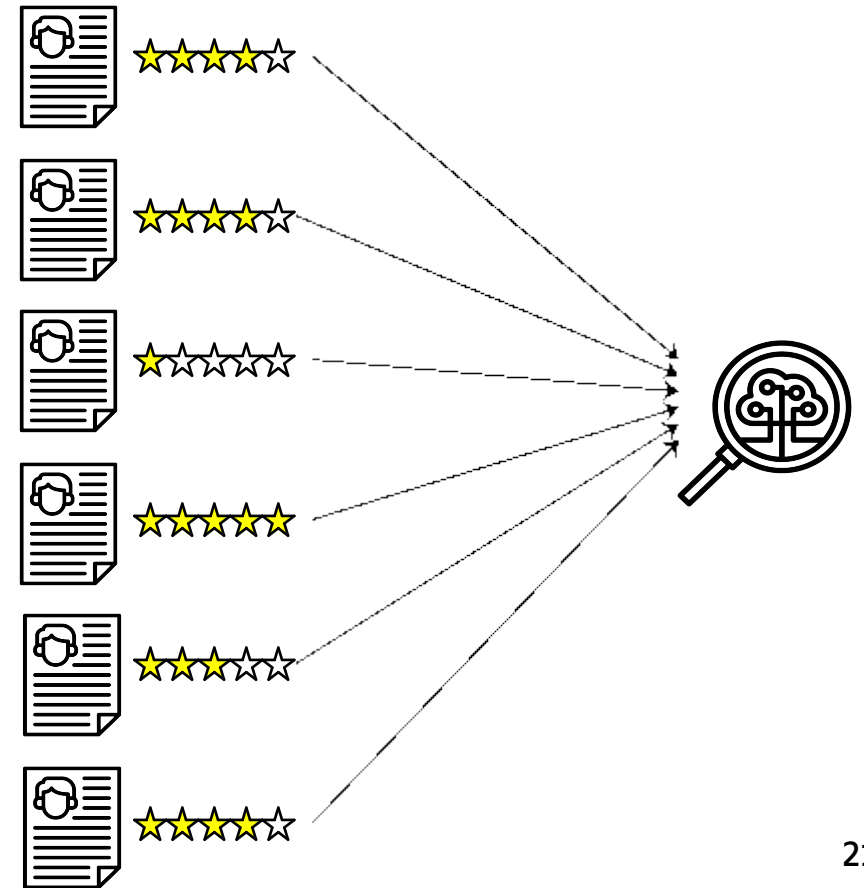
[f](#) [t](#)



Biased data: Amazon's recruiting tool



Previous applicants and results (10 years)




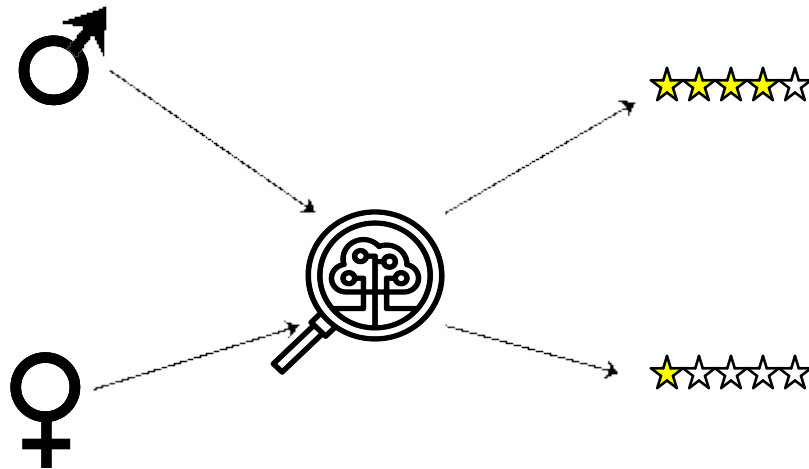
Biased data: Amazon's recruiting tool

RETAIL OCTOBER 11, 2018 / 1:04 AM / UPDATED 2 YEARS AGO

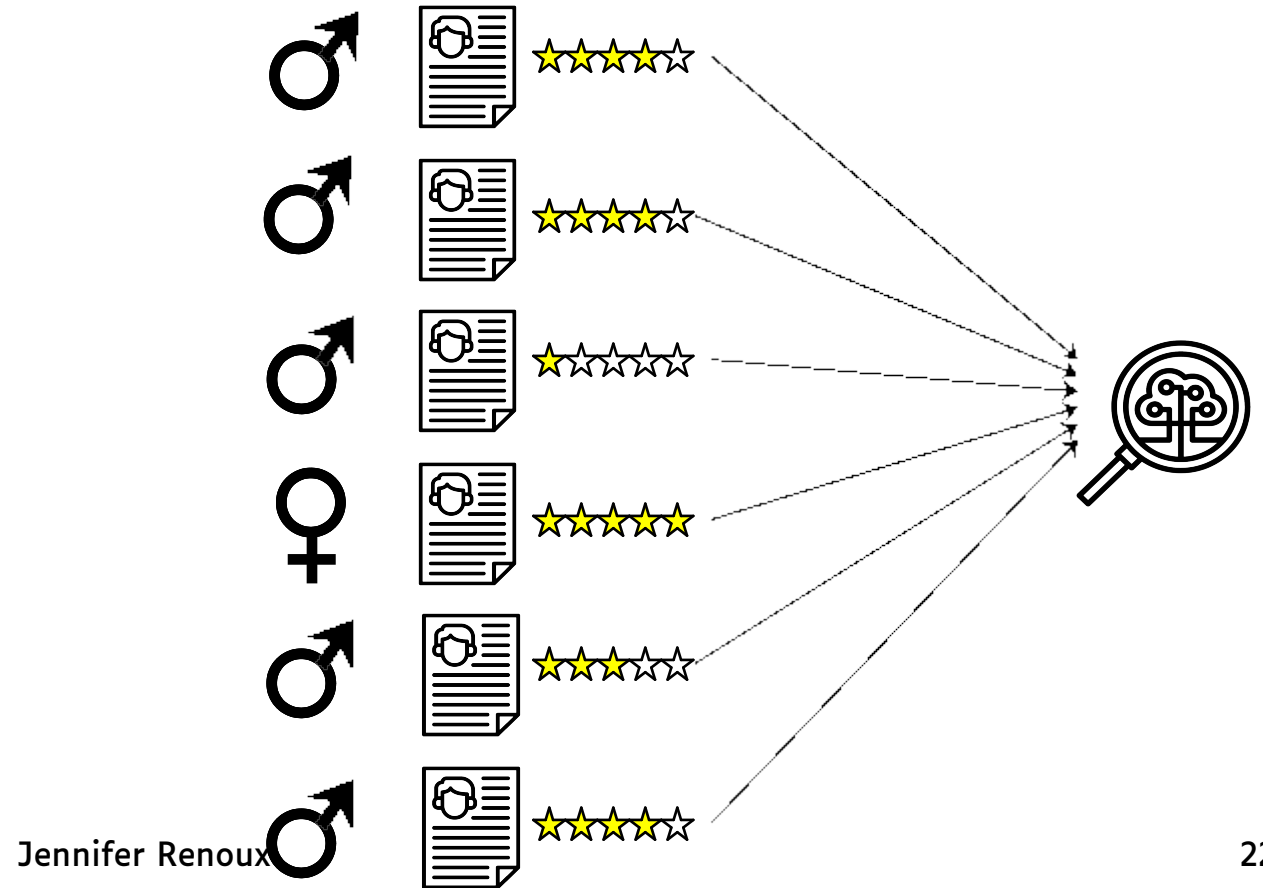
Amazon scraps secret AI recruiting tool that showed bias against women

By Jeffrey Dastin

8 MIN READ  



Previous applicants and results (10 years)



How can systems become biased?

Biased measure

A bad heuristic is used to frame the problem, leading to biases

Biased data

The data analysed in biased, leading to a biased analysis

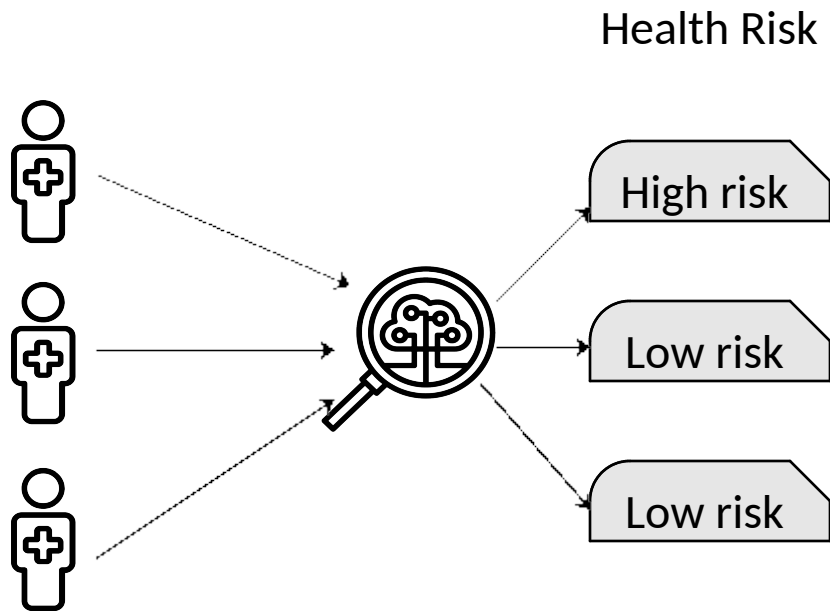
Conflicting costs

Some types of fairness are conflicting with each others and needs to be pondered

Biased measure: Framing the problem

Dissecting racial bias in an algorithm used to manage the health of populations

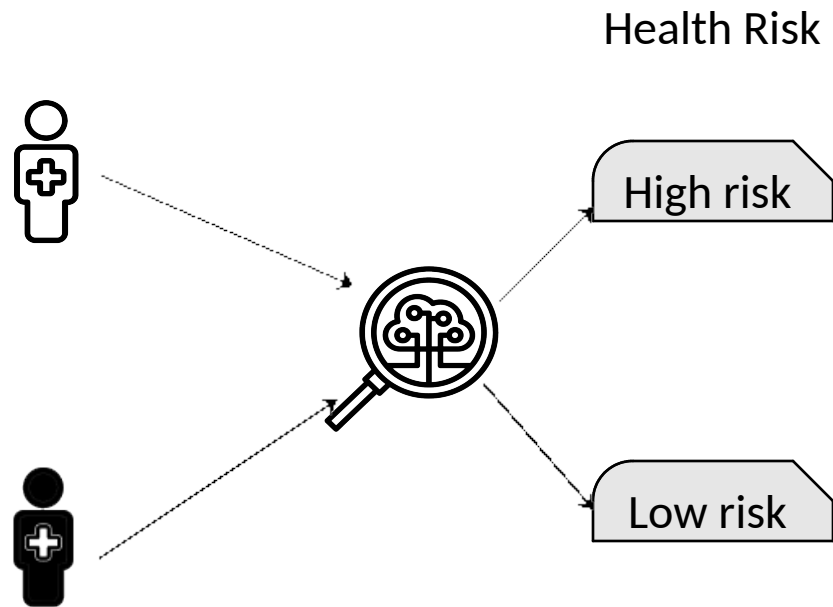
Ziad Obermeyer^{1,2*}, Brian Powers³, Christine Vogeli⁴, Sendhil Mullainathan^{5*†}



Biased measure: Framing the problem

Dissecting racial bias in an algorithm used to manage the health of populations

Ziad Obermeyer^{1,2*}, Brian Powers³, Christine Vogeli⁴, Sendhil Mullainathan^{5*†}



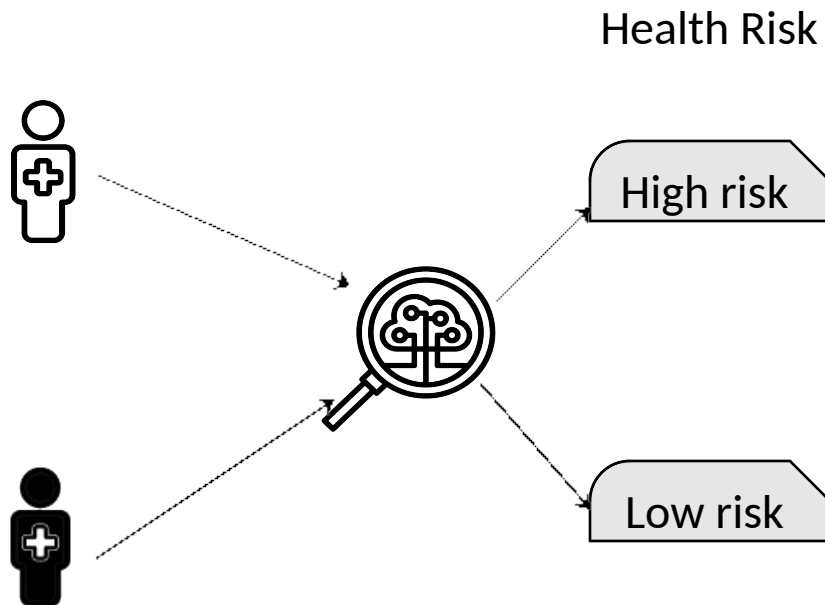
Biased measure: Framing the problem

Dissecting racial bias in an algorithm used to manage the health of populations

Ziad Obermeyer^{1,2*}, Brian Powers³, Christine Vogeli⁴, Sendhil Mullainathan^{5*†}



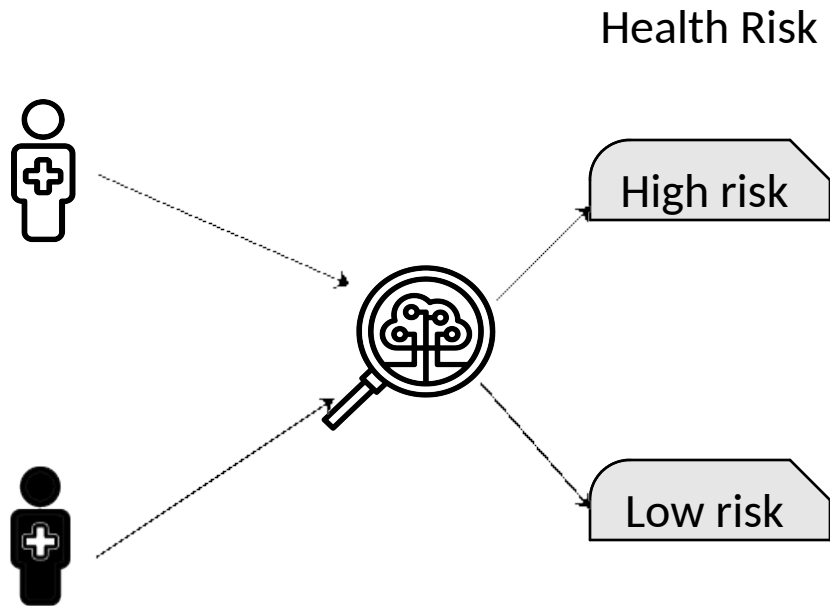
Is a patient high-risk based on their history?



Biased measure: Framing the problem

Dissecting racial bias in an algorithm used to manage the health of populations

Ziad Obermeyer^{1,2*}, Brian Powers³, Christine Vogeli⁴, Sendhil Mullainathan^{5*†}



Is a patient high-risk based on their history?

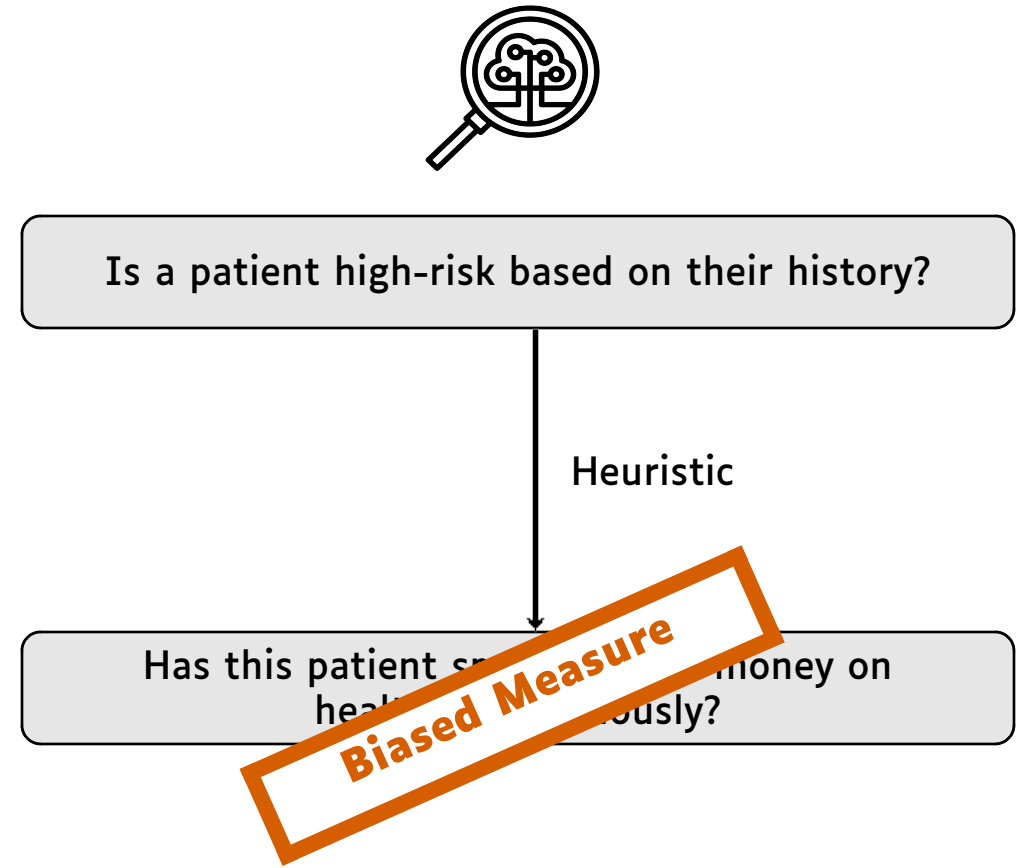
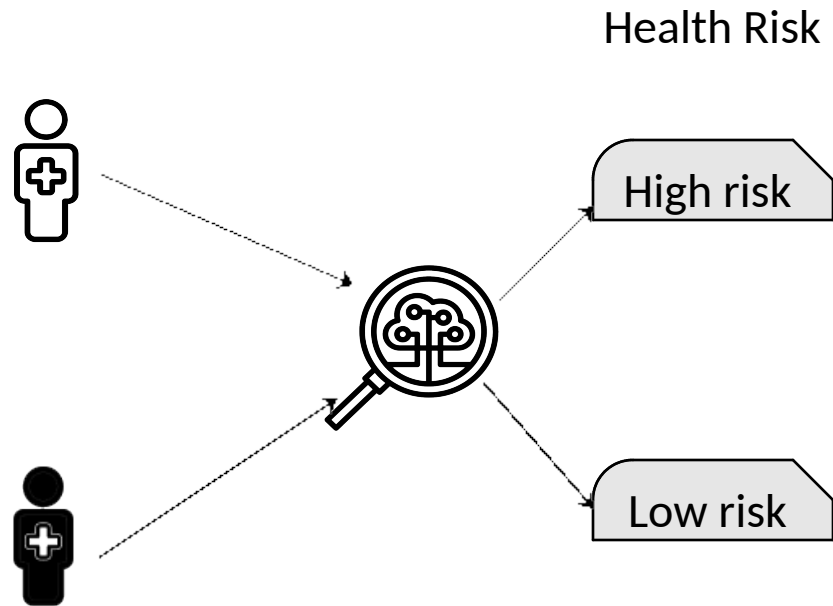
Heuristic

Has this patient spent a lot of money on healthcare previously?

Biased measure: Framing the problem

Dissecting racial bias in an algorithm used to manage the health of populations

Ziad Obermeyer^{1,2*}, Brian Powers³, Christine Vogeli⁴, Sendhil Mullainathan^{5*†}



How can systems become biased?

Biased measure

A bad heuristic is used to frame the problem, leading to biases

Biased data

The data analysed in biased, leading to a biased analysis

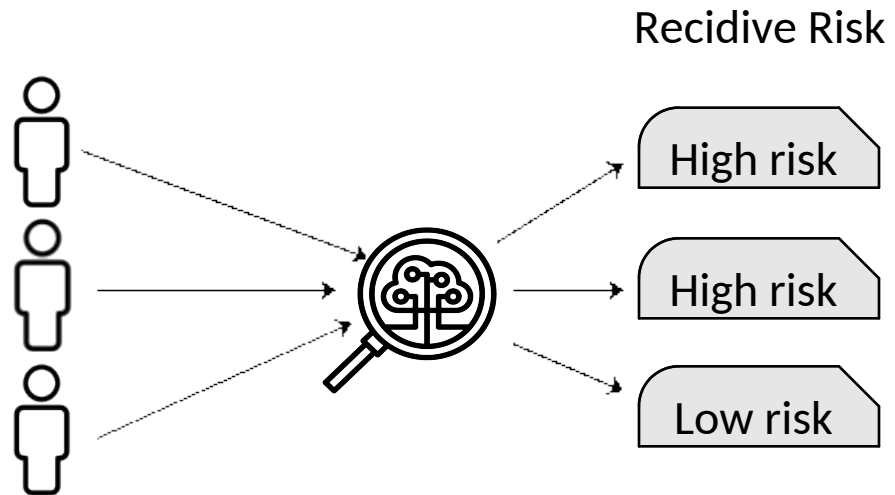
Conflicting costs

Some types of fairness are conflicting with each others and needs to be pondered

Conflicting costs: The example of COMPAS

The accuracy, fairness, and limits of predicting recidivism

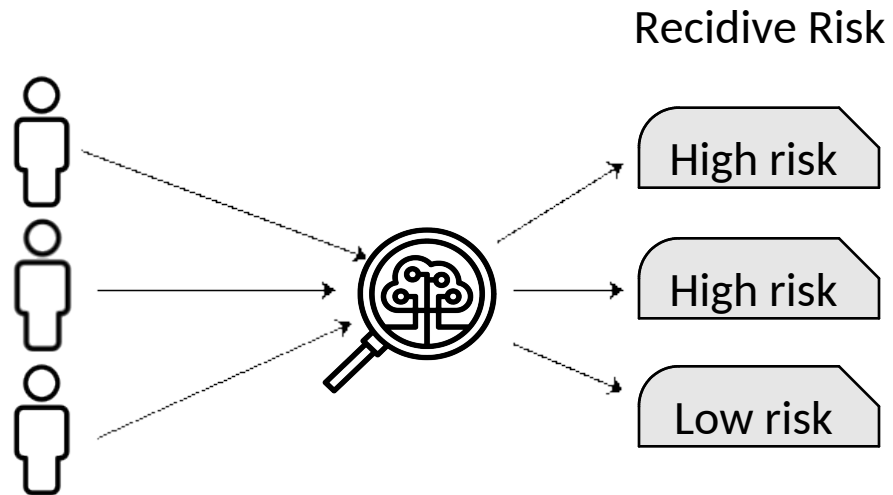
Julia Dressel and Hany Farid*



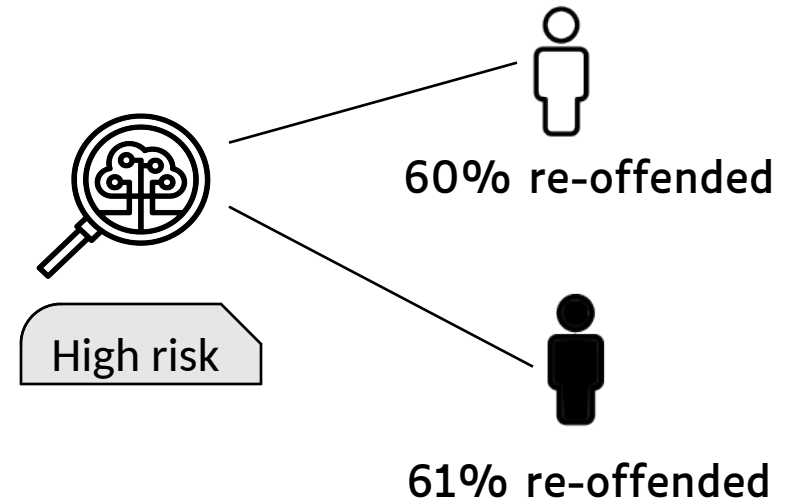
Conflicting costs: The example of COMPAS

The accuracy, fairness, and limits of predicting recidivism

Julia Dressel and Hany Farid*



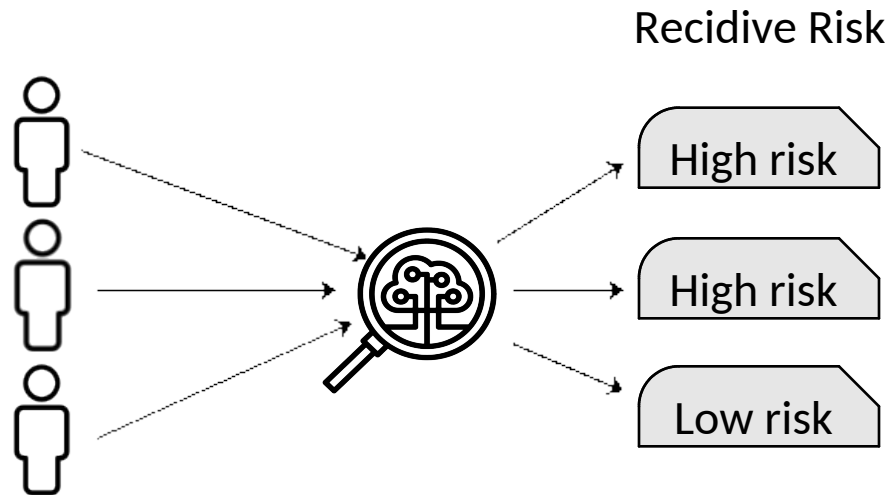
Fairness: Well-Calibrated System



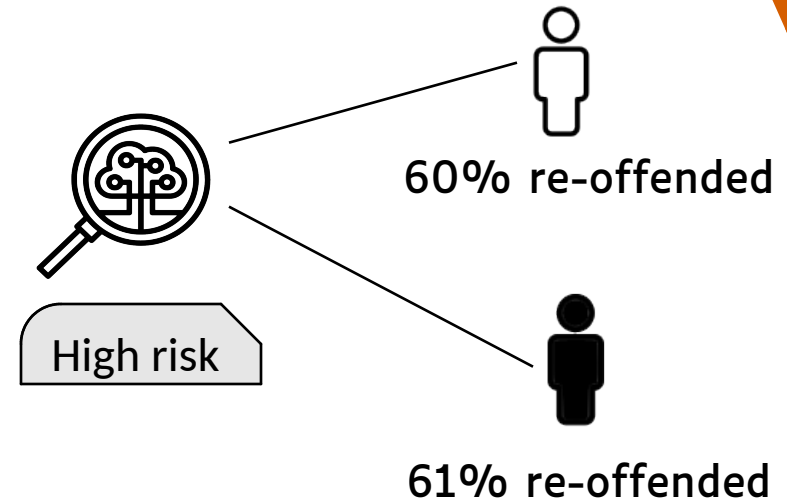
Conflicting costs: The example of COMPAS

The accuracy, fairness, and limits of predicting recidivism

Julia Dressel and Hany Farid*



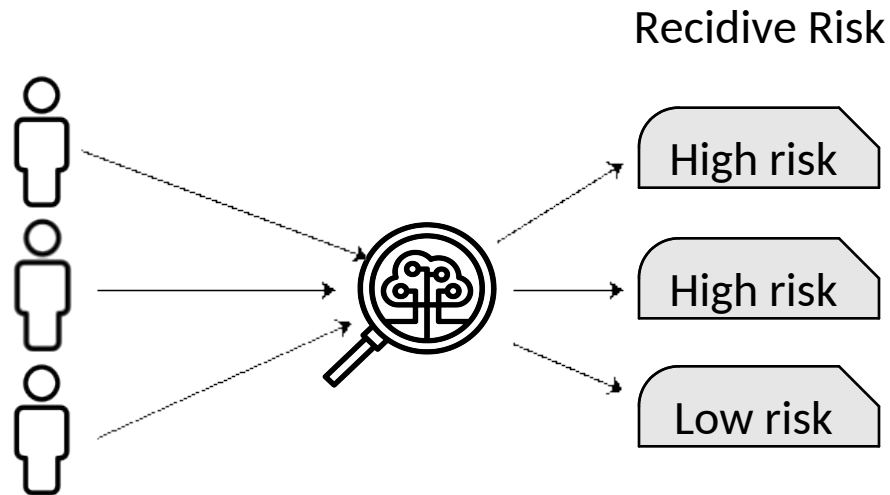
Fairness: Well-Calibrated System **OK**



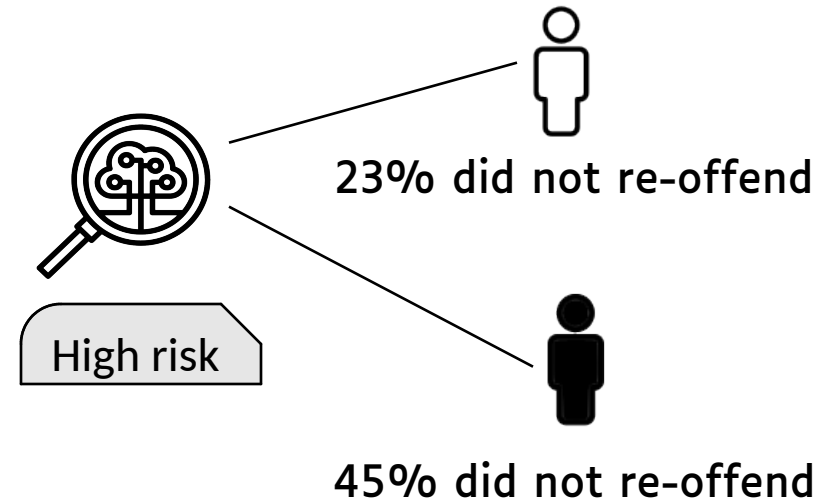
Conflicting costs: The example of COMPAS

The accuracy, fairness, and limits of predicting recidivism

Julia Dressel and Hany Farid*



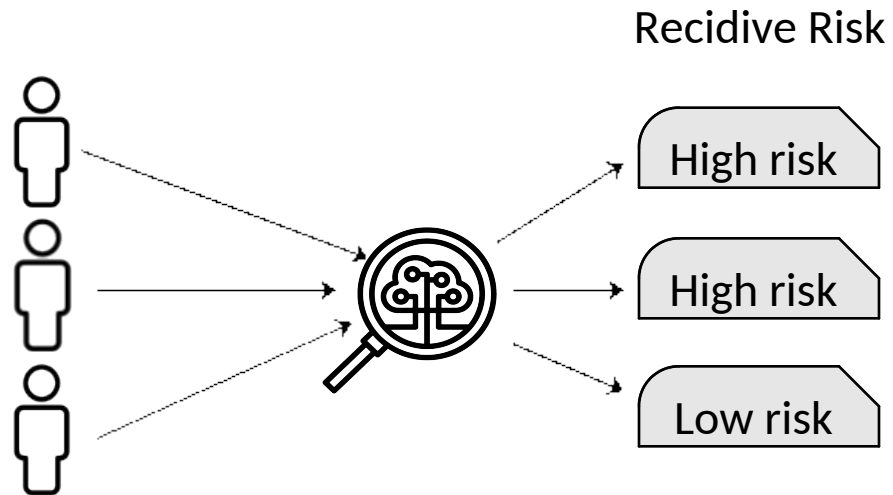
Fairness: Equal Opportunity?



Conflicting costs: The example of COMPAS

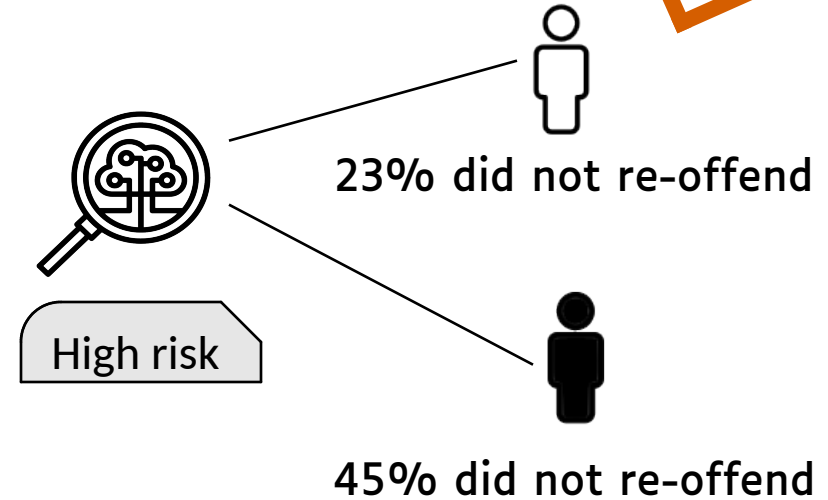
The accuracy, fairness, and limits of predicting recidivism

Julia Dressel and Hany Farid*



Fairness: Equal Opportunity

Biased



Conflicting costs: the conundrum

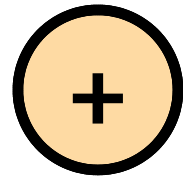
Can a system be well calibrated **and** provide equal opportunity?

Conflicting costs: the conundrum

Can a system be well calibrated to provide equal opportunity?

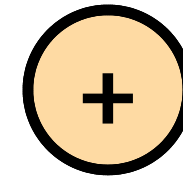
Kleinberg, J. et al. "Inherent Trade-Offs in the Fair Determination of Risk Scores."
ArXiv abs/1609.05807 (2017): n. pag.

How can systems become biased?



Biased measure

A bad heuristic is used to frame the problem, leading to biases



Biased data

The data analysed in biased, leading to a biased analysis

Conflicting costs

Some types of fairness are conflicting with each others and needs to be pondered

What can we do?

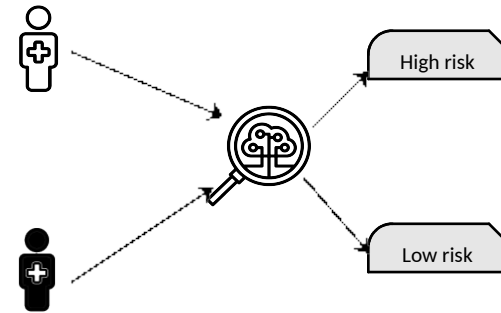
Involving the right stakeholders

Who are you going to talk to when eliciting the requirements?

Involving the right stakeholders

Dissecting racial bias in an algorithm used to manage the health of populations

Ziad Obermeyer^{1,2*}, Brian Powers³, Christine Vogeli⁴, Sendhil Mullainathan^{5*†}



Requirement: create an automated system helping nurses to decide which patient to send to the emergency room.

Who would you involve if you were tasked to develop such a system?

Involving the right stakeholders

Hospital
administration

Medical experts

Nurses

Involving the right stakeholders

Hospital
administration

Medical experts

Nurses

Representative for
patients

Others?

Social scientists

Reconsidering the objective function



Reconsidering the objective function



May reward those with advantageous educational opportunities, enforcing class boundaries

Reconsidering the objective function



May reward those with advantageous educational opportunities, enforcing class boundaries

Better chance to cut across class boundaries and choose from a broader pool, but more difficult to evaluate

Reconsidering the objective function



May reward those with advantageous educational opportunities, enforcing class boundaries

Better chance to cut across class boundaries and choose from a broader pool, but more difficult to evaluate

Whatever the choice, it has to be an **explicit** reasoning and **conscious** decision

The power of a diverse team

You are more likely to spot problems if you are directly concerned

The power of a diverse team

You are more likely to spot problems if you are directly
concerned

Did you know that some people are motion sick when playing video games?
(Digital Motion Sickness)

Is your GUI understandable for color-blind people?

Is your human-machine dialog model appropriate for non-binary or transgender
people?

Key Takeaways

1

Software systems can cause involuntary harm

Software systems are used in many critical life situations. A biased system can cause “real-world” harm.

Creating “non-biased system” is a socio-technical challenge

“Bias” only exists related to a societal context.

2

3

Consider fairness from the start

Fairness is not something we add “when we have time”.