



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 11:00 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 11:00 AM

Are Facebook Ads Discriminatory?

Silicon Valley Mar 20
Facebook is going to stop letting advertisers target by race, gender, or age — 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 11:00 AM

Are Facebook Ads Discriminatory?

Silicon Valley Mar 20
Facebook is going to stop letting advertisers target by race, gender, or age — 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.

suomi ⇄ englanti

Hän on fiksu. Hän on kaunis. Hän on insinööri. Hän on siivoja. Hän urheilee. Hän tiskaa. Hän ajaa autoa. Hän hoitaa lapsia. Hän on johtaja. X

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 11:00 AM

Are Facebook Ads Discriminatory?

Silicon Valley Mar 20
Facebook is going to stop letting advertisers target by race, gender, or age — 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.

suomi



englanti

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

July 7, 2021 / Lydia X. Z. Brown

kaunis. Hän on insinööri. Urheilee. Hän tiskaa. Hän on lapsia. Hän on johtaja. X

beautiful. He is an engineer. 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 11:00 AM

Are Facebook Ads Discriminatory?

Silicon Valley Mar 20

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

suomi



englanti

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

July 7, 2021 / Lydia X. Z. Brown

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

The accuracy, fairness, and limits of predicting recidivism

Julia Dressel and Hany Farid*



Intersectional Accuracy on PPB

94.0%



**DARKER
MALES**

79.2%



**DARKER
FEMALES**

100%



**LIGHTER
MALES**

98.3%



**LIGHTER
FEMALES**



www.gendershades.org



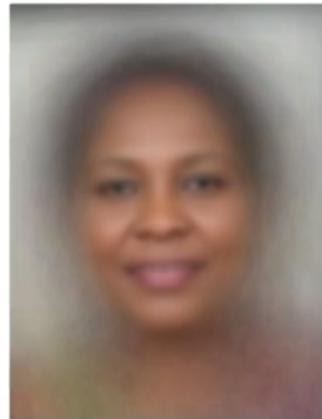
Intersectional Accuracy

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?

POTENTIAL HARMS FROM ALGORITHMIC DECISION-MAKING	
INDIVIDUAL HARMS	COLLECTIVE SOCIAL HARMS
ILLEGAL DISCRIMINATION	UNFAIR PRACTICES
HIRING	LOSS OF OPPORTUNITY
EMPLOYMENT	
INSURANCE & SOCIAL BENEFITS	
HOUSING	
EDUCATION	
CREDIT	ECONOMIC LOSS
DIFFERENTIAL PRICES OF GOODS	
LOSS OF LIBERTY	SOCIAL STIGMATIZATION
INCREASED SURVEILLANCE	
STEREOTYPE REINFORCEMENT	
DIGNATORY HARMS	

Chart Contents Courtesy of Megan Smith, Former CTO of the United States

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

Equal opportunity

Each demographic class is offered the same opportunity

Equal outcome

Each demographic class gets the same results

Group fairness

Two classes are treated similarly

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.

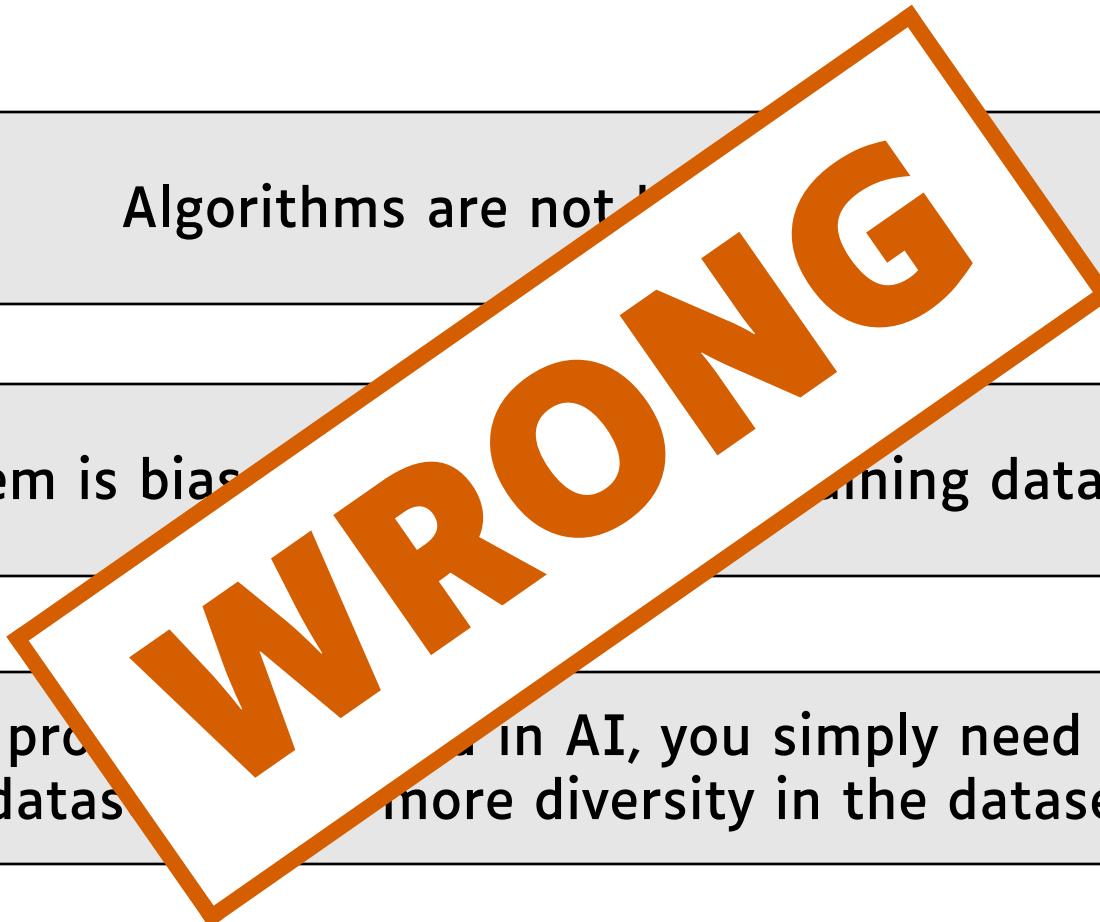
To solve the problem of biased in AI, you simply need to unbiase the datasets / add more diversity in the datasets

Quick Myth-busting

Algorithms are not biased.

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

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



Two important takes

1

Biased in AI systems is not limited to ML

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

2

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

“Bias” only exists related to a societal context.

How can systems become biased?

Biased data

The data analysed is biased, leading to a biased analysis

Biased measure

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

Conflicting costs

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

How can systems become biased?

Biased data

The data analysed is biased, leading to a biased analysis

Biased measure

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

Conflicting costs

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

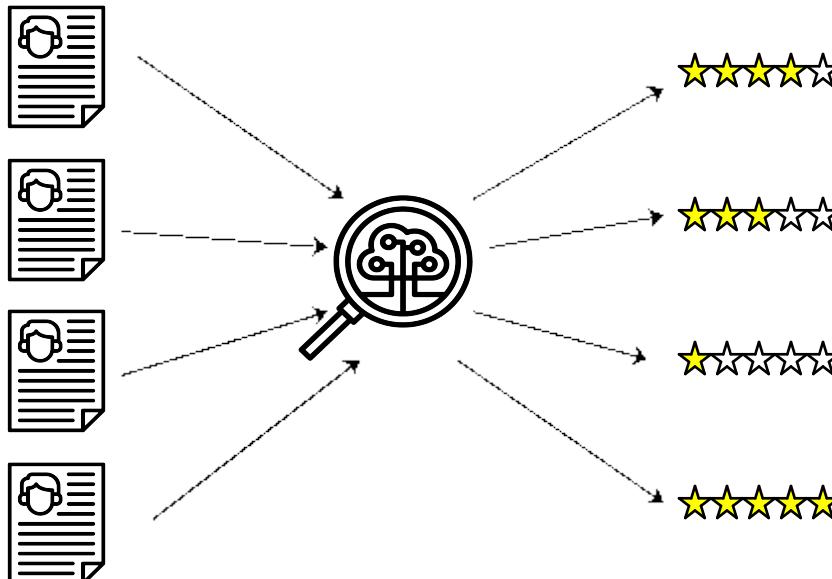
Biased data: Amazon's recruiting tool

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

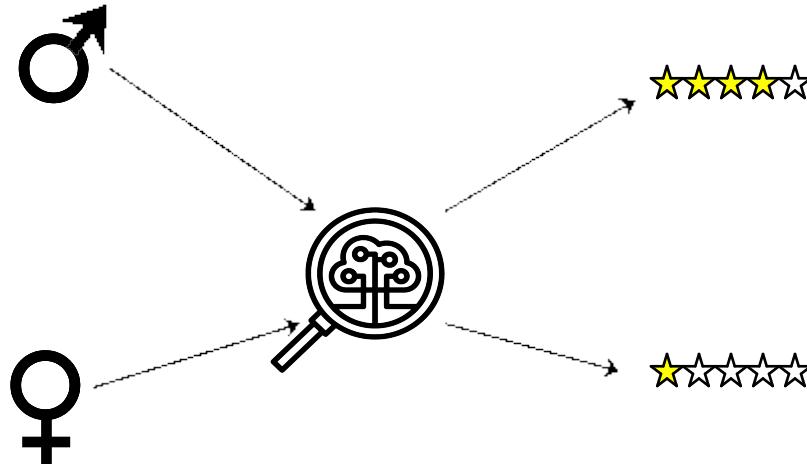
Amazon scraps secret AI recruiting tool that showed bias against women

By Jeffrey Dastin

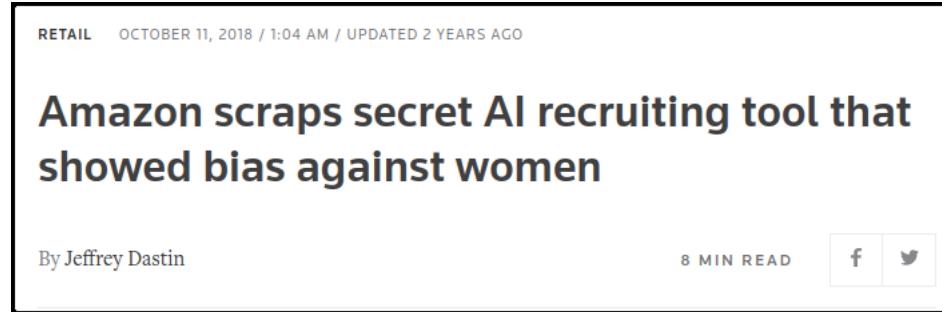
8 MIN READ [f](#) [t](#)



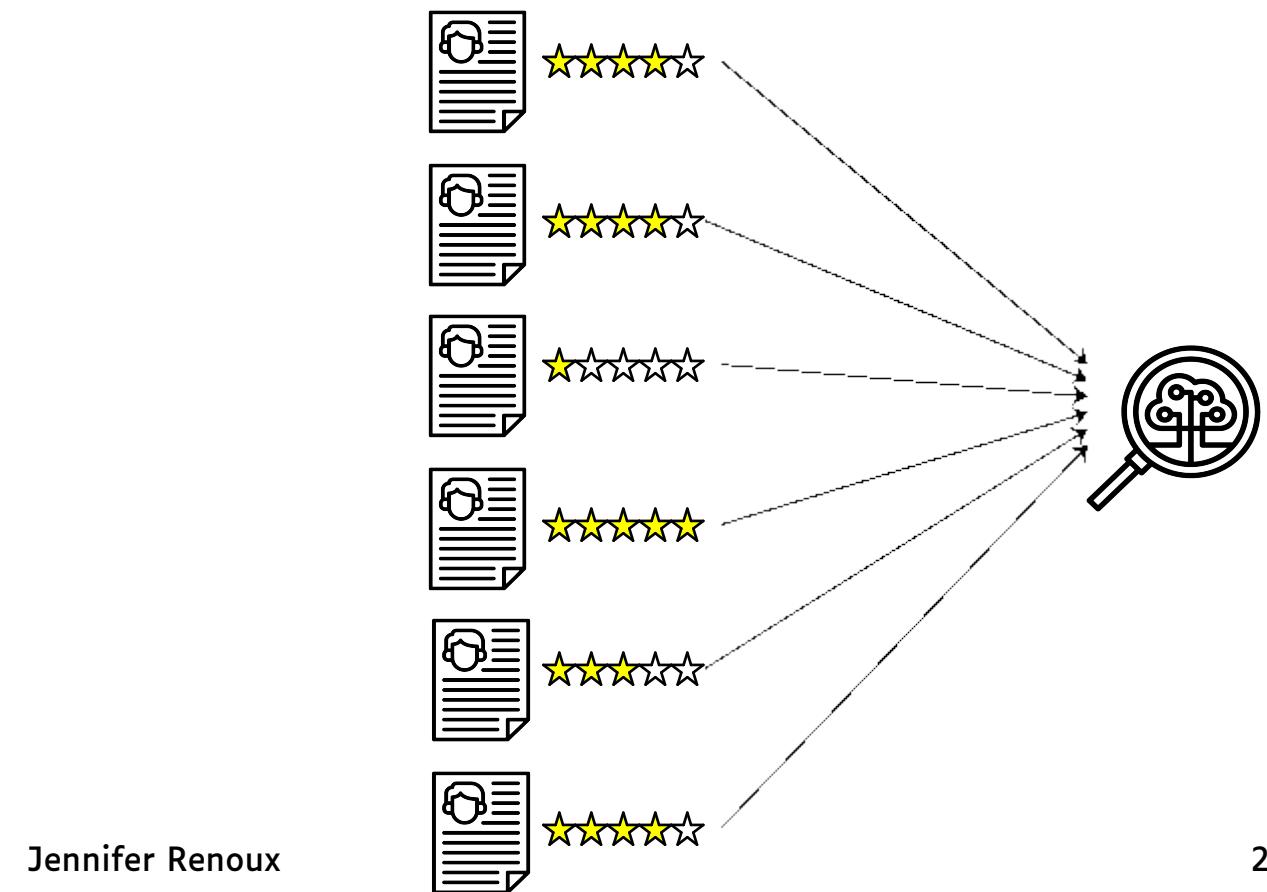
Biased data: Amazon's recruiting tool



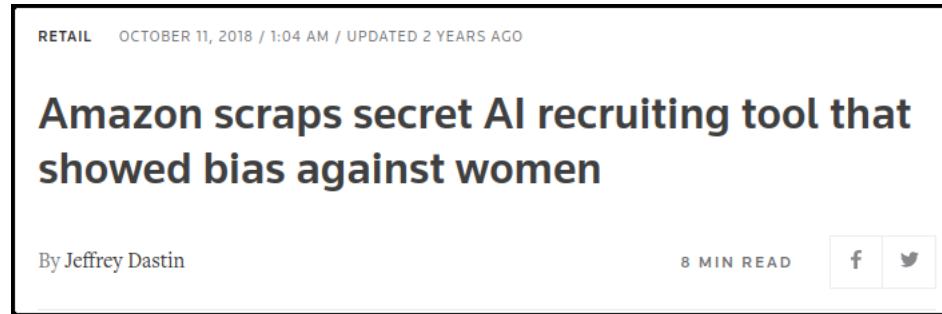
Biased data: Amazon's recruiting tool



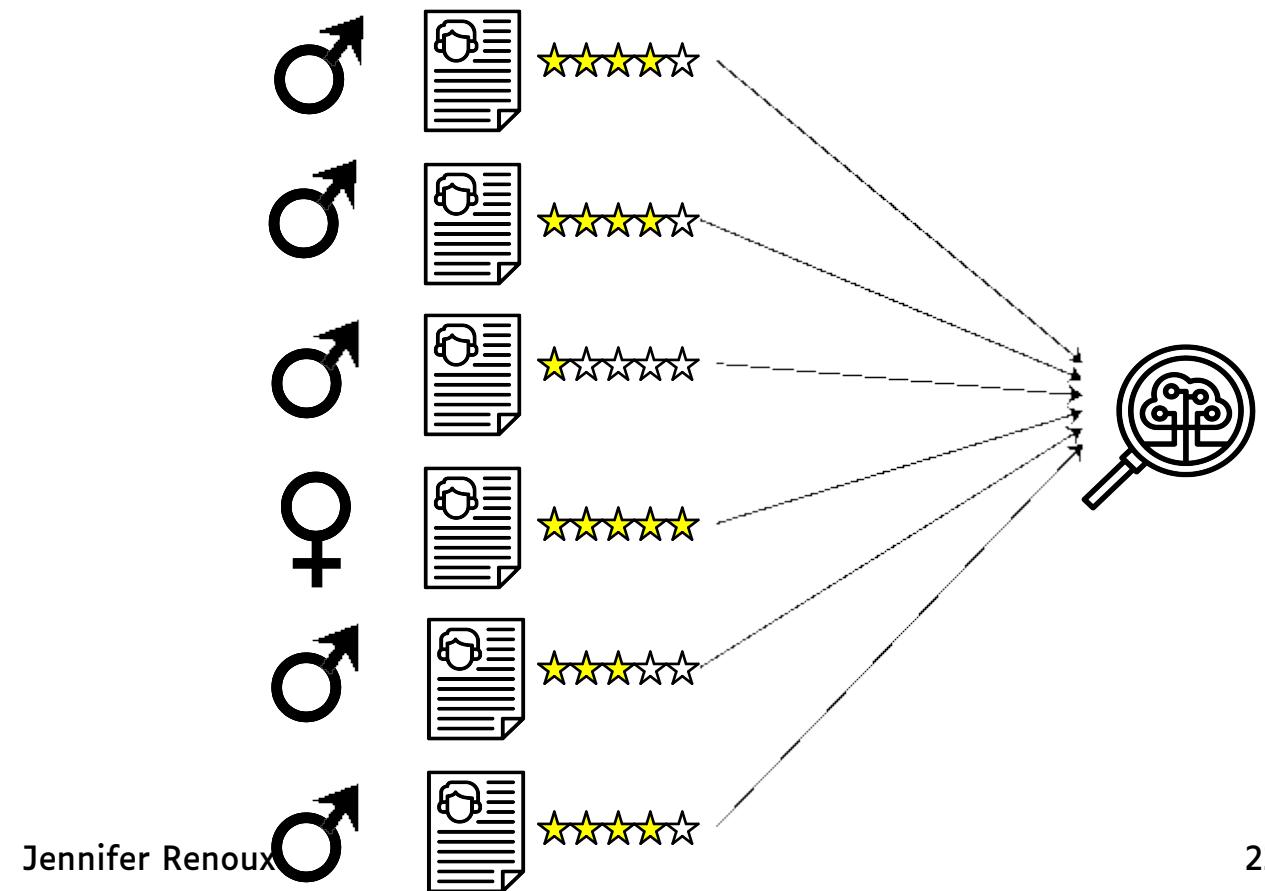
Previous applicants and results (10 years)



Biased data: Amazon's recruiting tool



Previous applicants and results (10 years)



How can systems become biased?

Biased data

The data analysed is biased, leading to a biased analysis

Biased measure

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

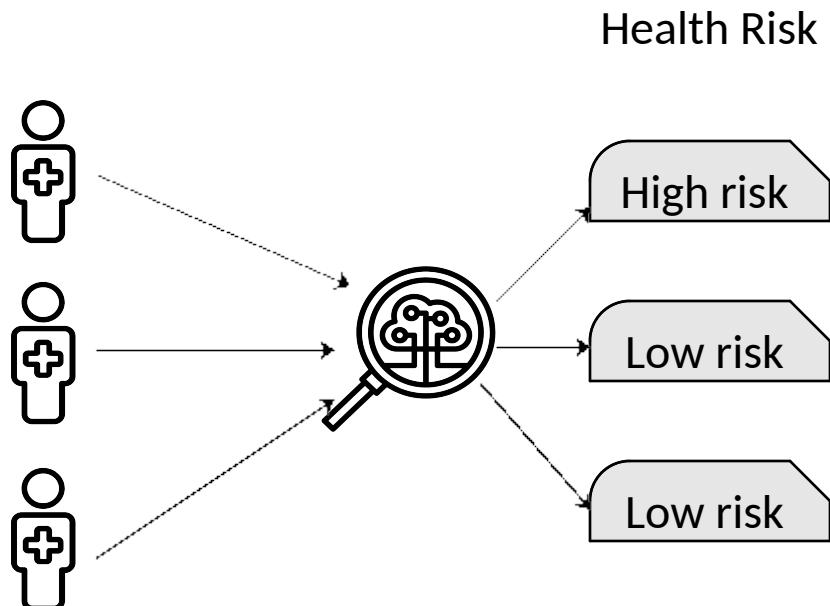
Conflicting costs

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

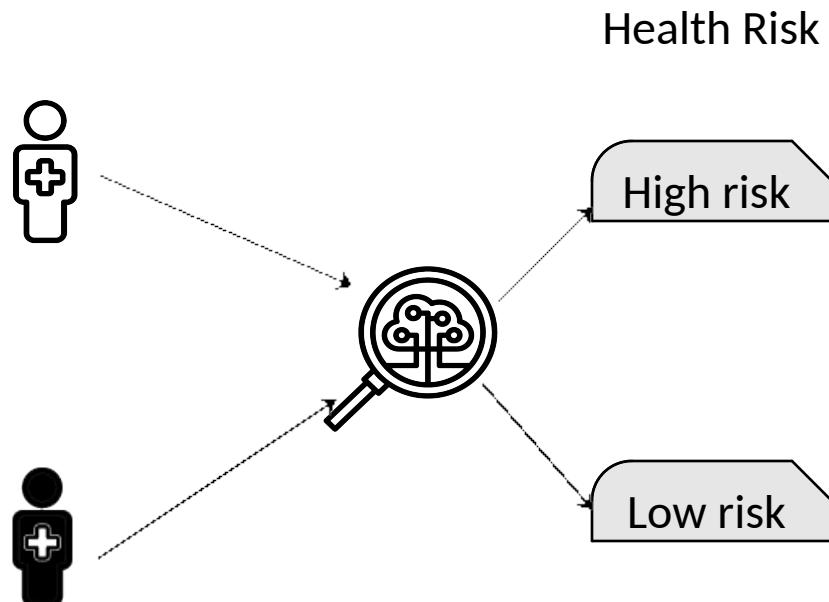
Ziad Obermeyer^{1,2*}, Brian Powers³, Christine Vogeli⁴, Sendhil Mullainathan^{5*†}



Biased measure: Framing the problem

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

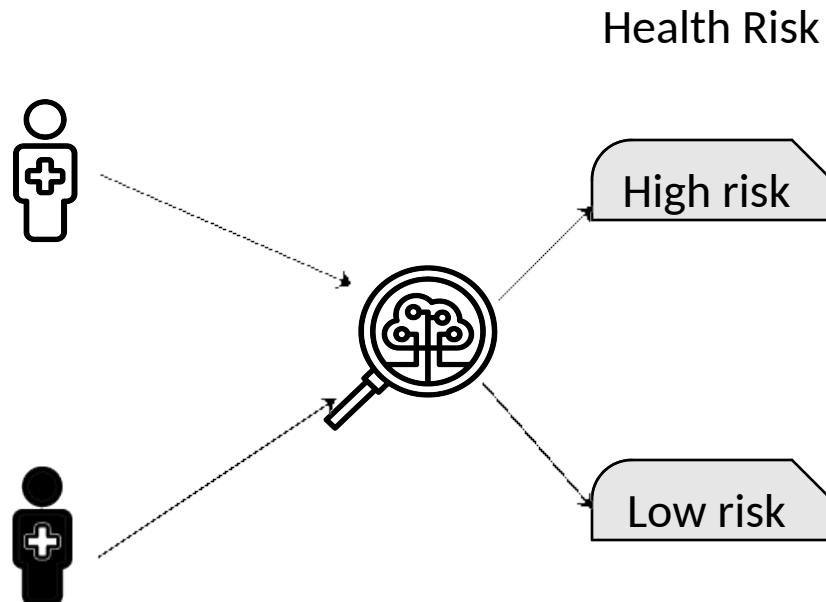
Ziad Obermeyer^{1,2*}, Brian Powers³, Christine Vogeli⁴, Sendhil Mullainathan^{5*†}



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*†}

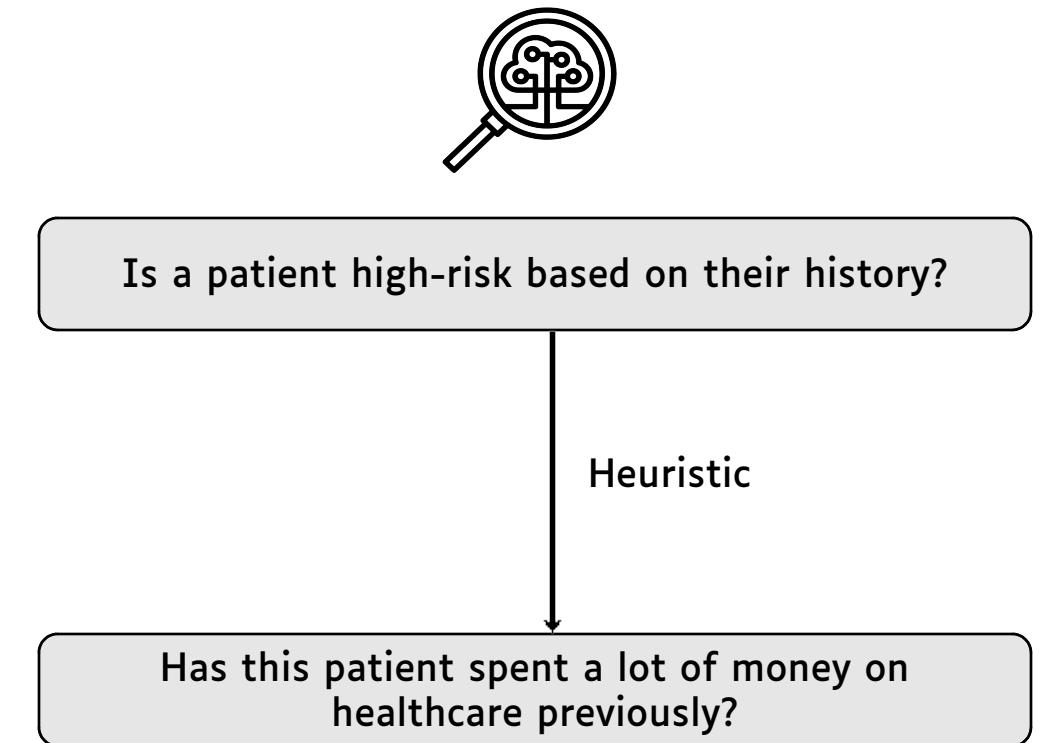
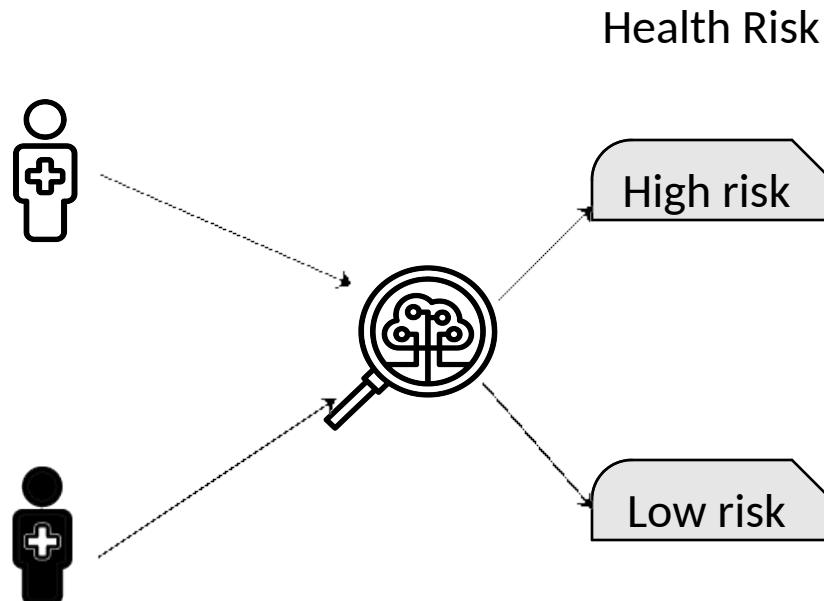


Is a patient high-risk based on their history?

Biased measure: Framing the problem

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

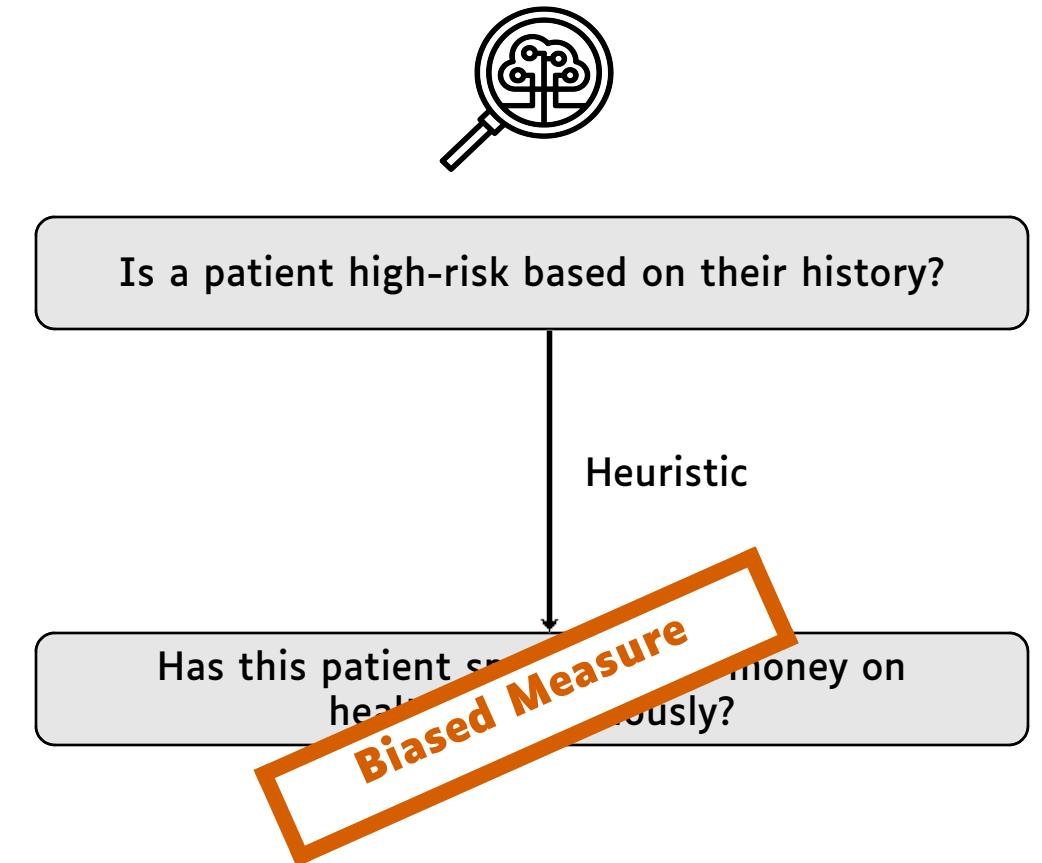
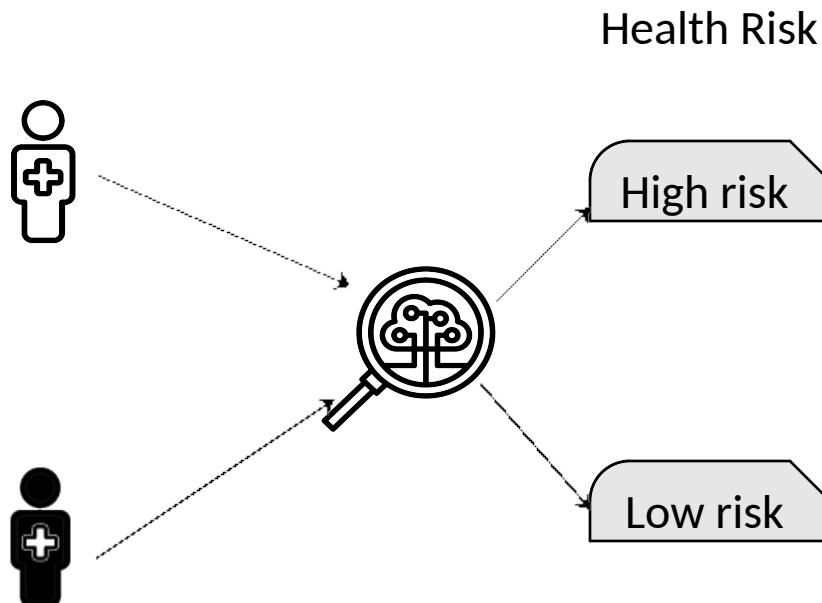
Ziad Obermeyer^{1,2*}, Brian Powers³, Christine Vogeli⁴, Sendhil Mullainathan^{5*†}



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 data

The data analysed is biased, leading to a biased analysis

Biased measure

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

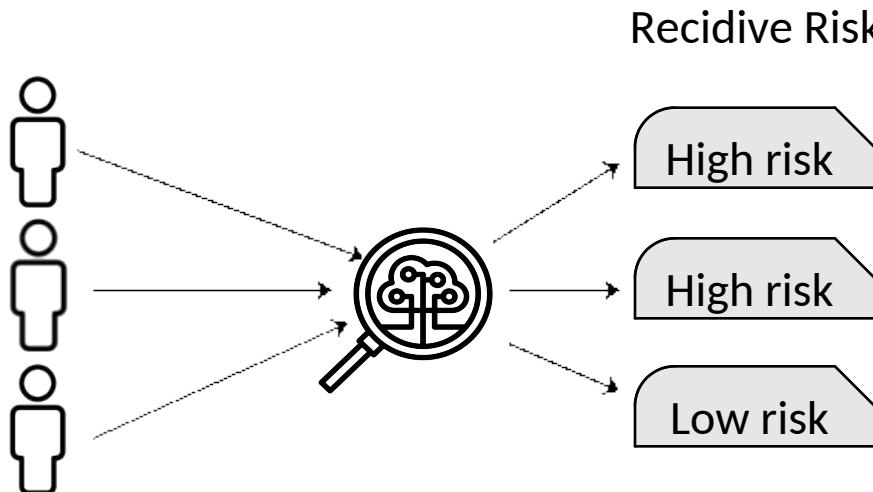
Conflicting costs

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

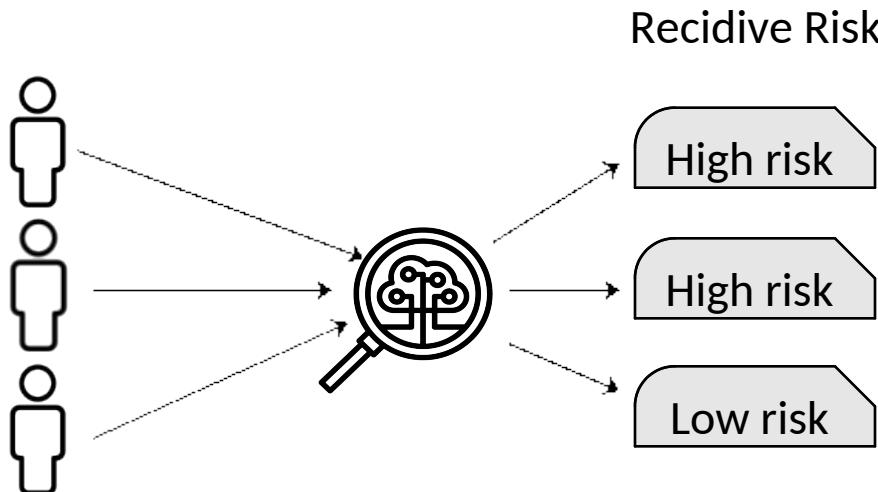
Julia Dressel and Hany Farid*



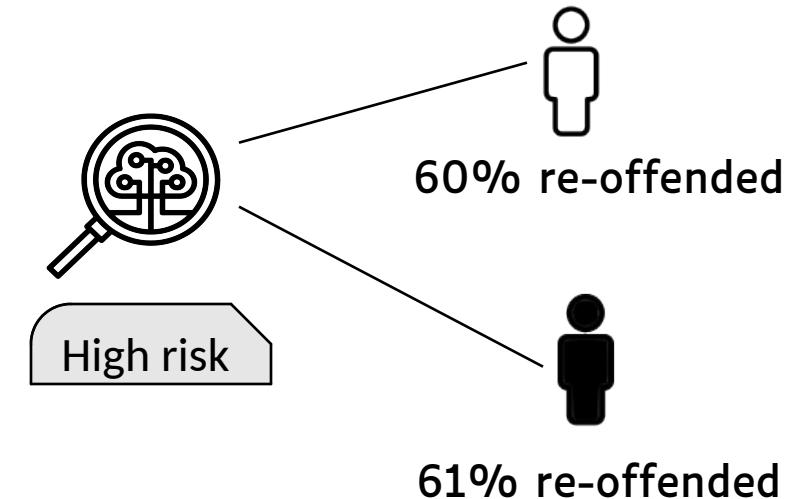
Conflicting costs: The example of COMPAS

The accuracy, fairness, and limits of predicting recidivism

Julia Dressel and Hany Farid*

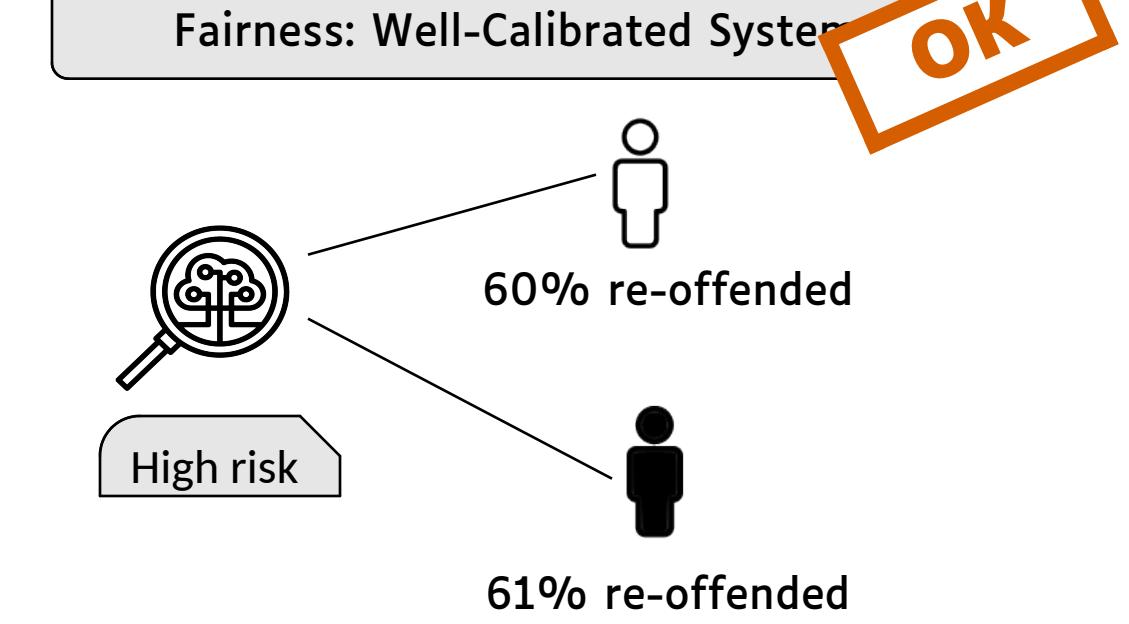
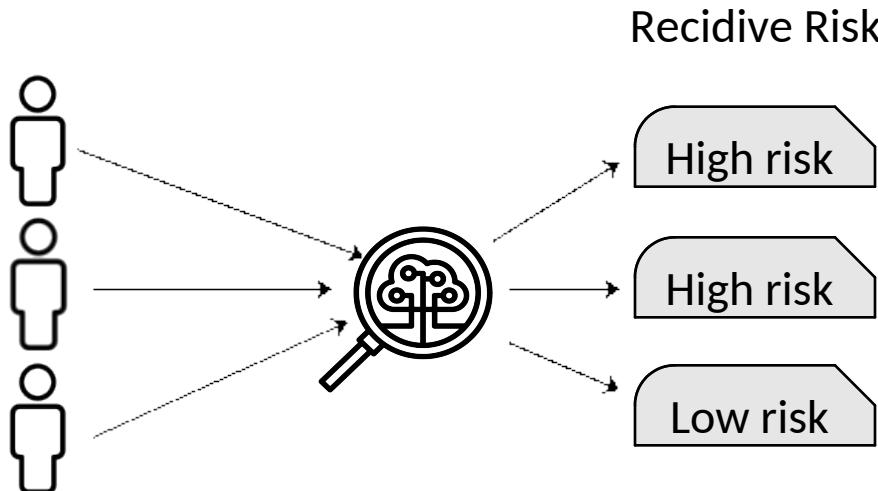


Fairness: Well-Calibrated System



Conflicting costs: The example of COMPAS

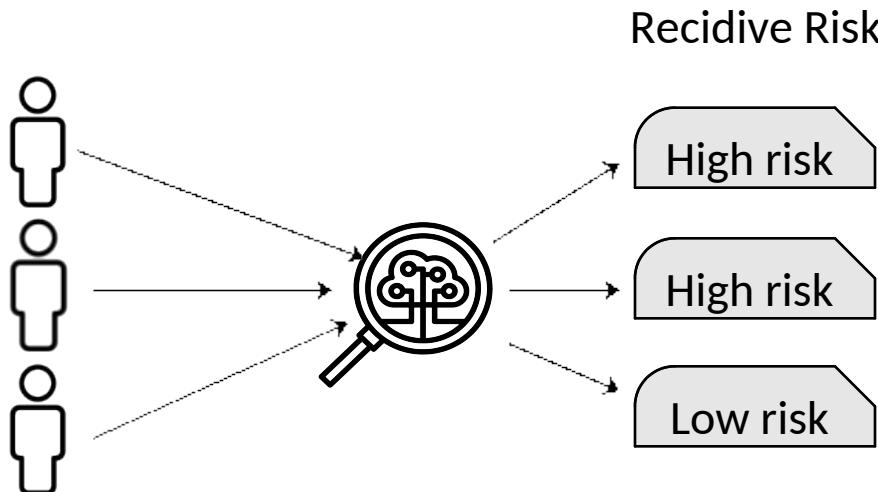
The accuracy, fairness, and limits
of predicting recidivism
Julia Dressel and Hany Farid*



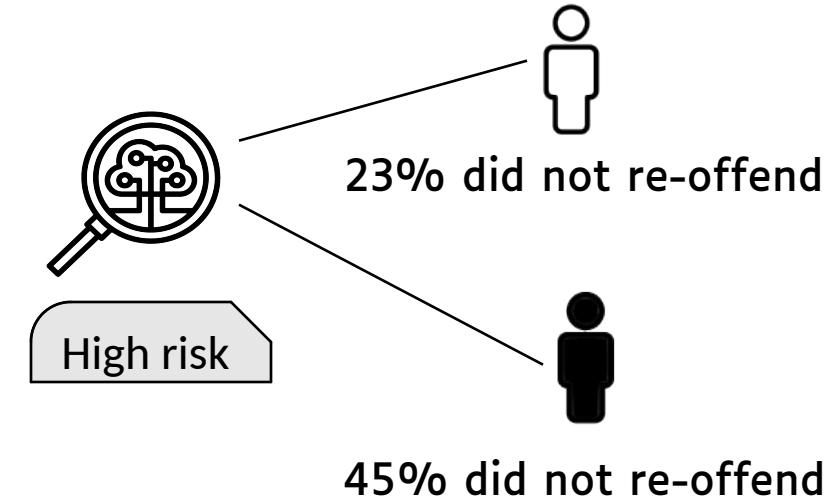
Conflicting costs: The example of COMPAS

The accuracy, fairness, and limits of predicting recidivism

Julia Dressel and Hany Farid*



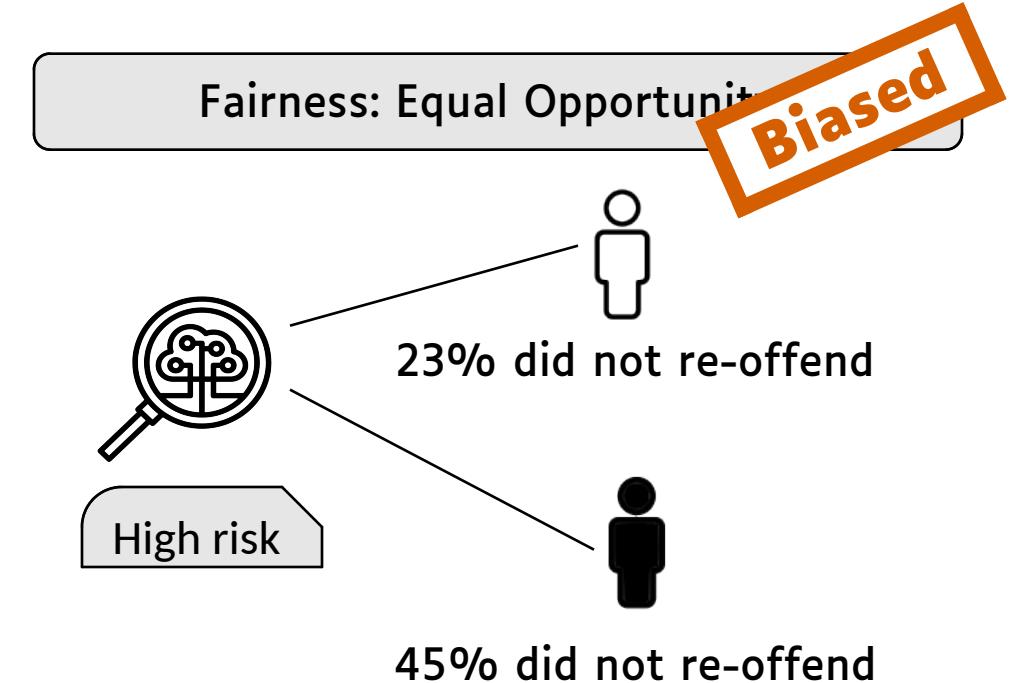
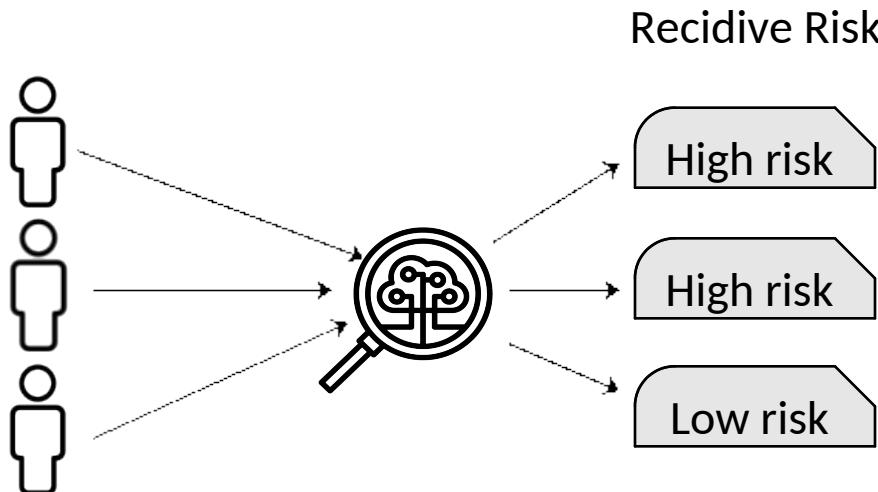
Fairness: Equal Opportunity?



Conflicting costs: The example of COMPAS

The accuracy, fairness, and limits of predicting recidivism

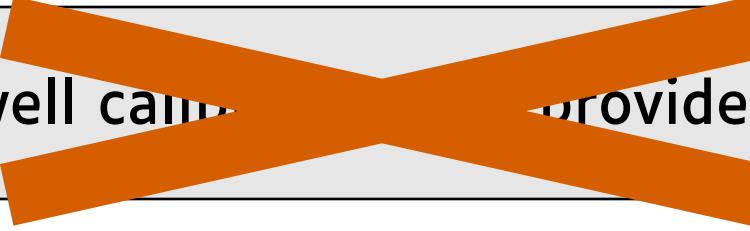
Julia Dressel and Hany Farid*



Conflicting costs: the conundrum

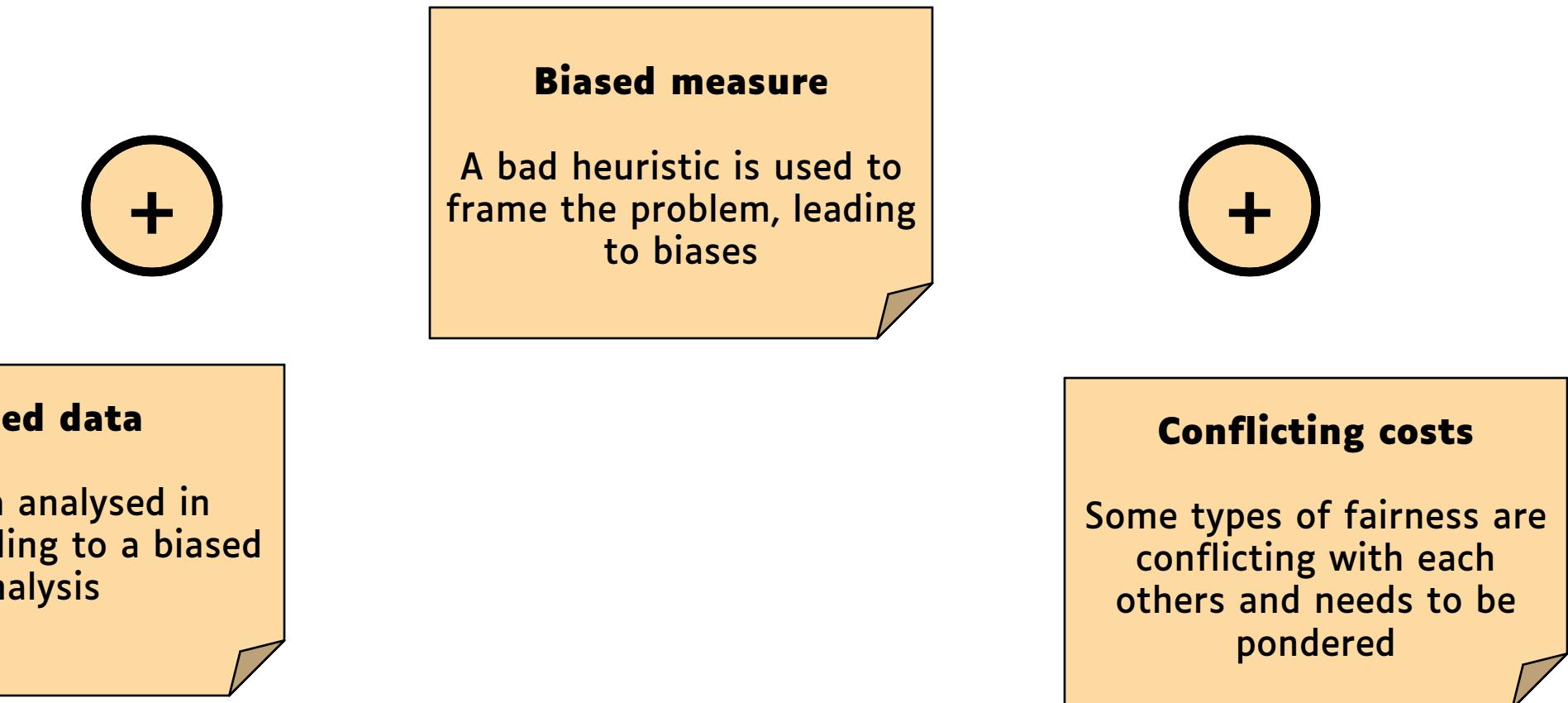
Can a system be well calibrated **and** provide equal opportunity?

Conflicting costs: the conundrum

Can a system be well calibrated?  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?



What can we do?

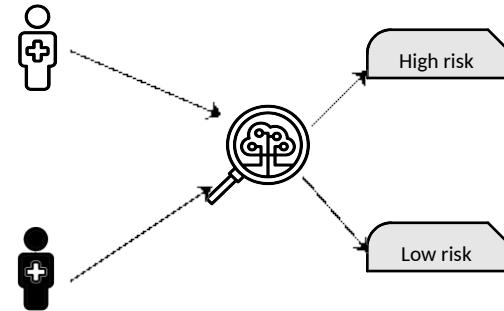
Involving the right stakeholders

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 concered

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.

2

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

"Bias" only exists related to a societal context.

3

Consider fairness from the start

Fairness is not something we add "when we have time".