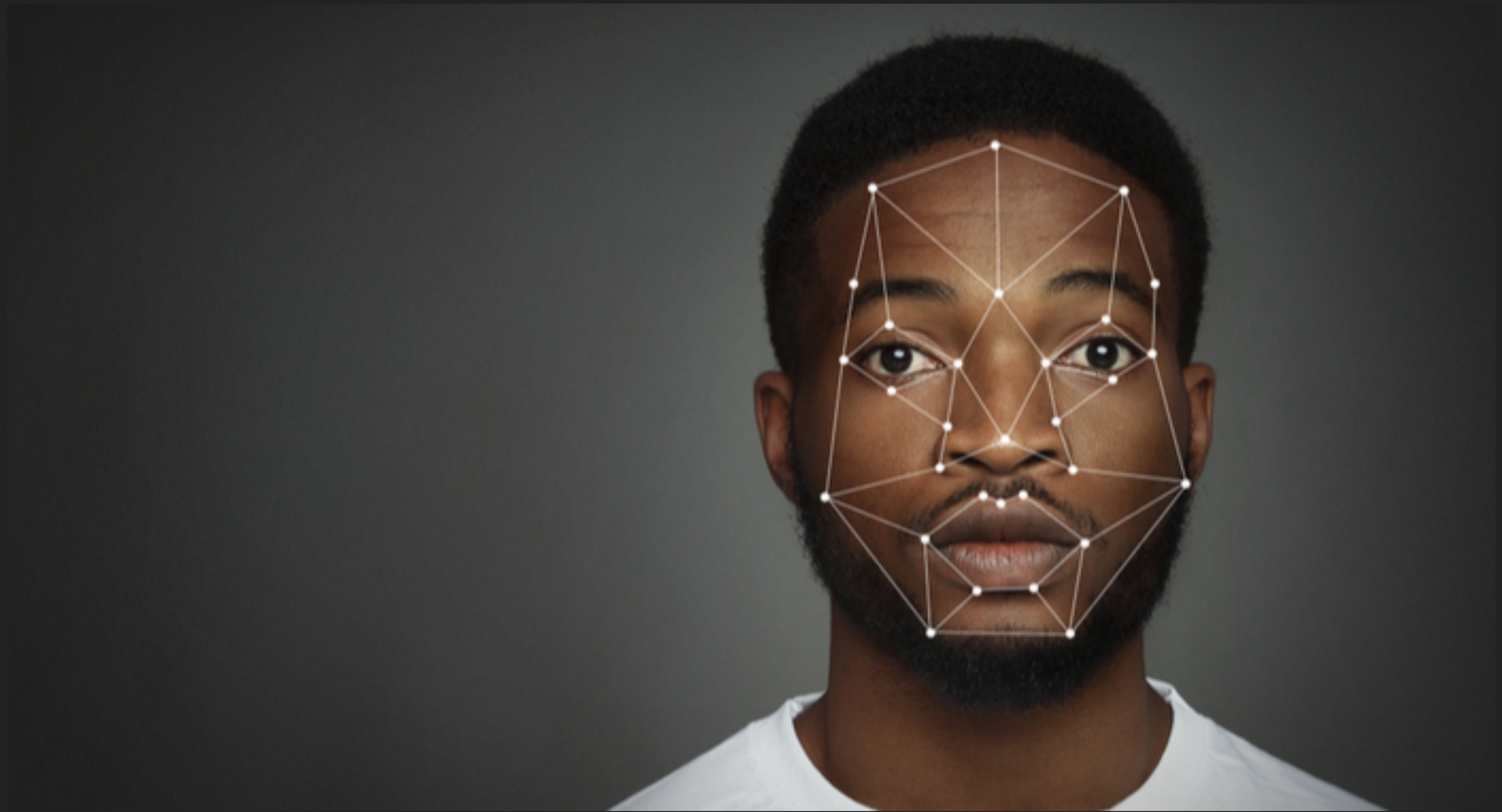
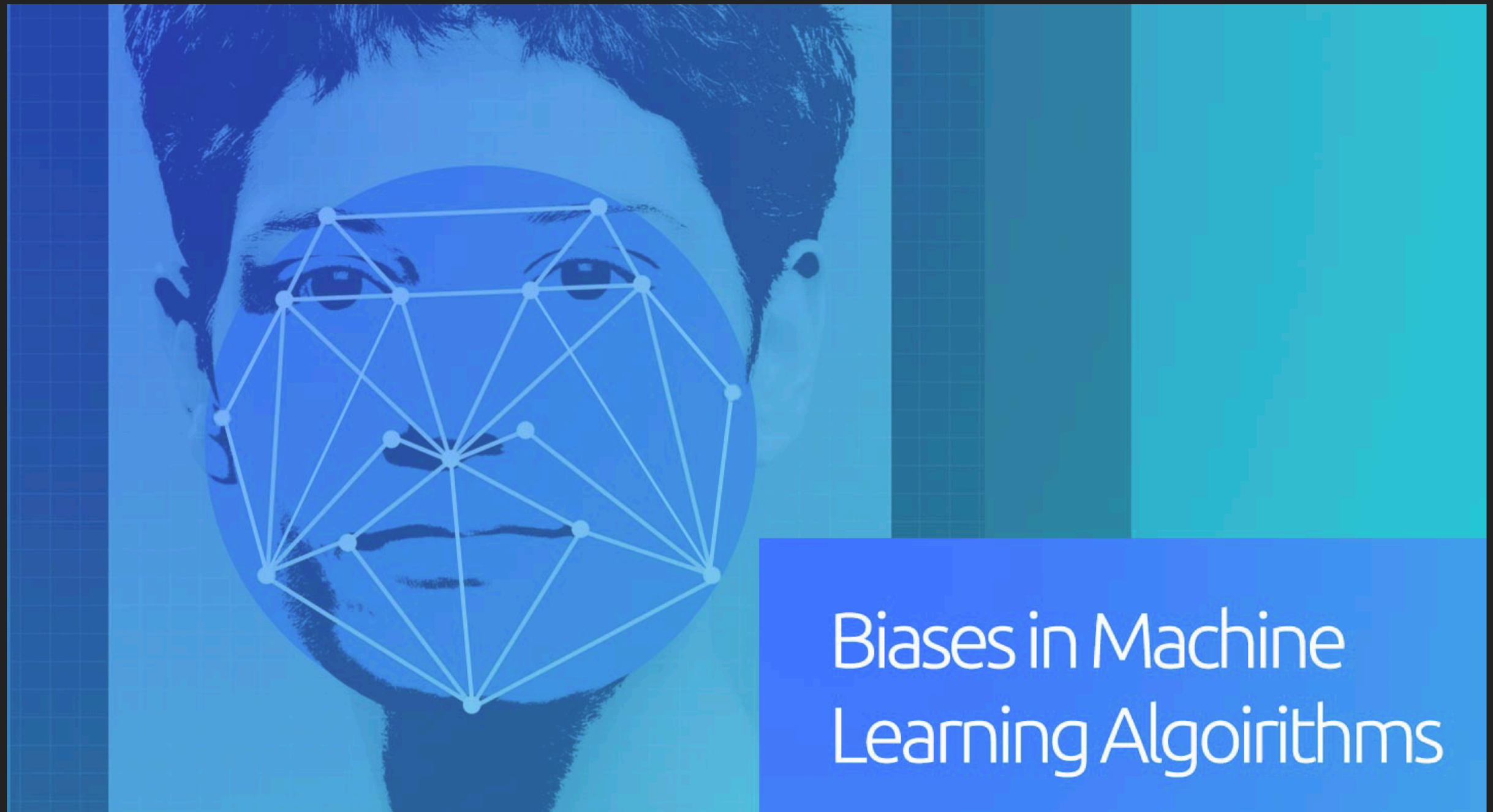
(Un)Ethical Computing

face recognition



predicting recidivism

human resources

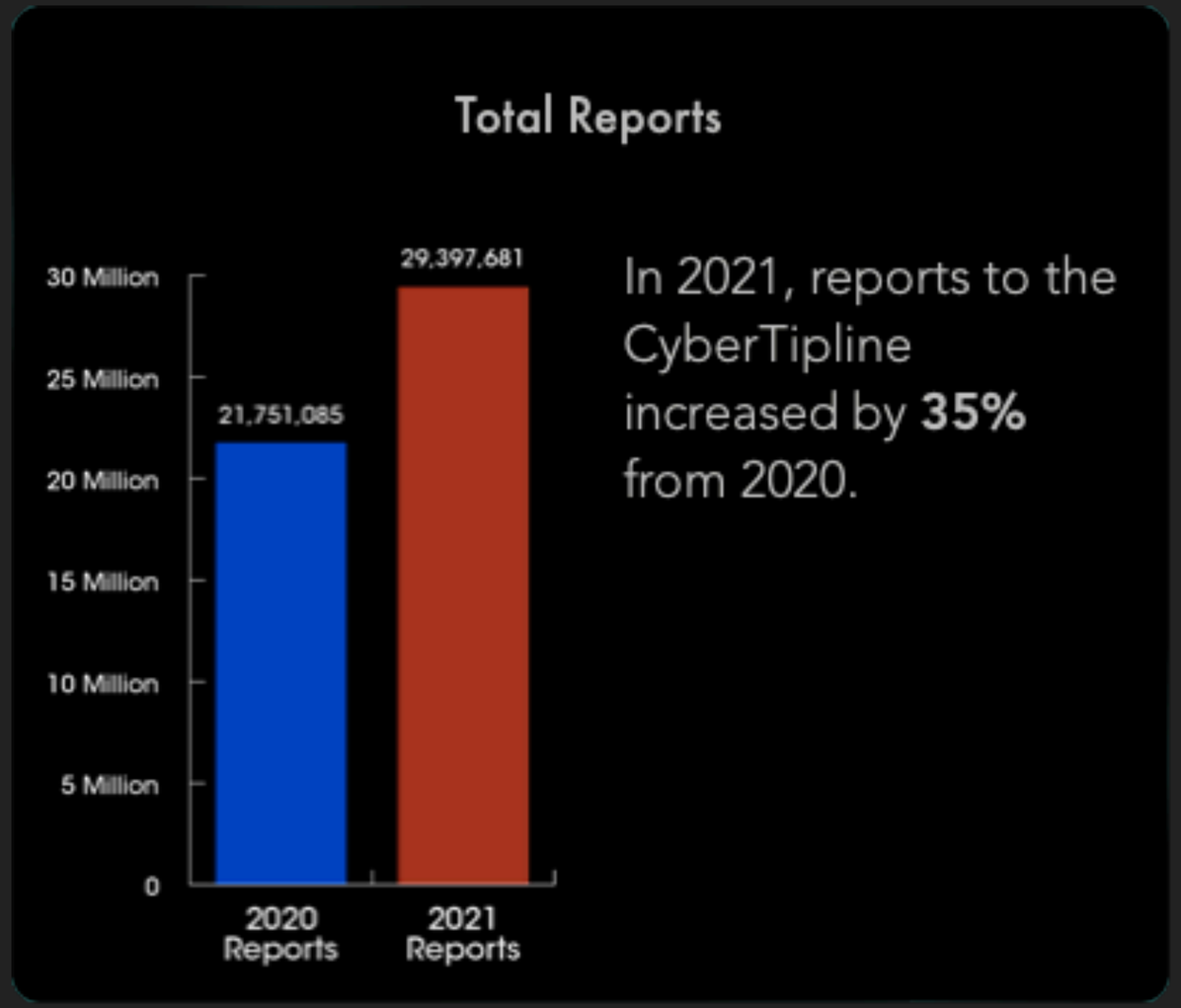


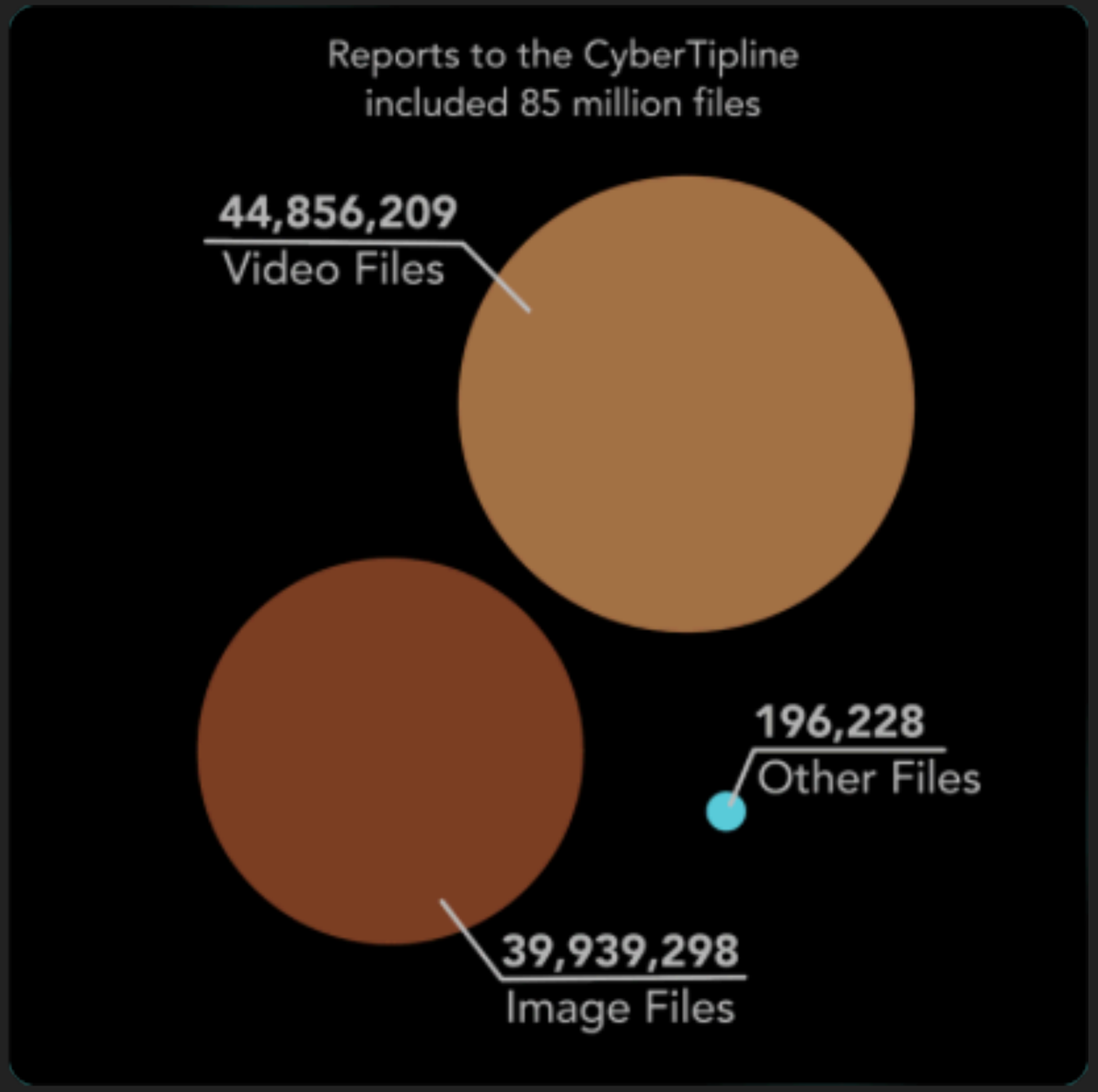
financial lending

non-consensual imagery

addiction (body image, self harm, bullying,…)

child sexual abuse

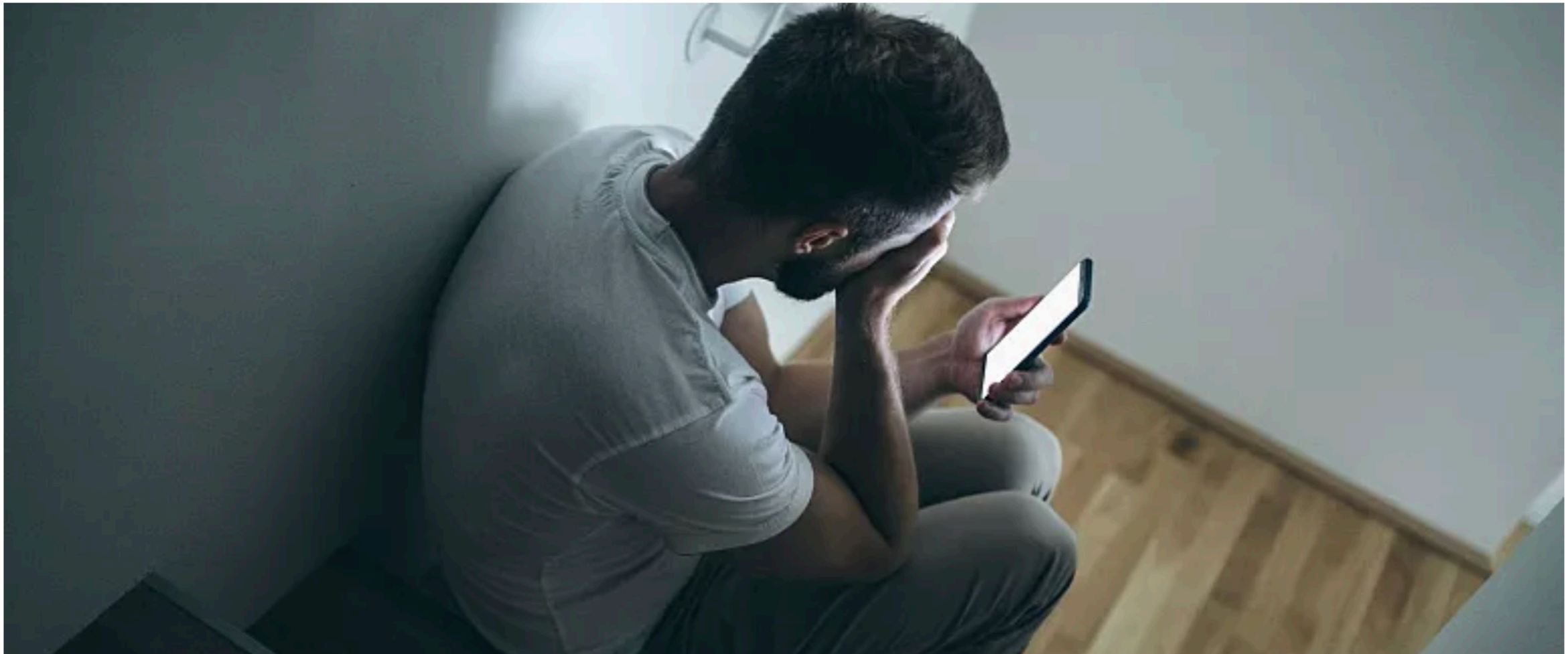




narcotics

generative AI

Man ends his life after an AI chatbot 'encouraged' him to sacrifice himself to stop climate change



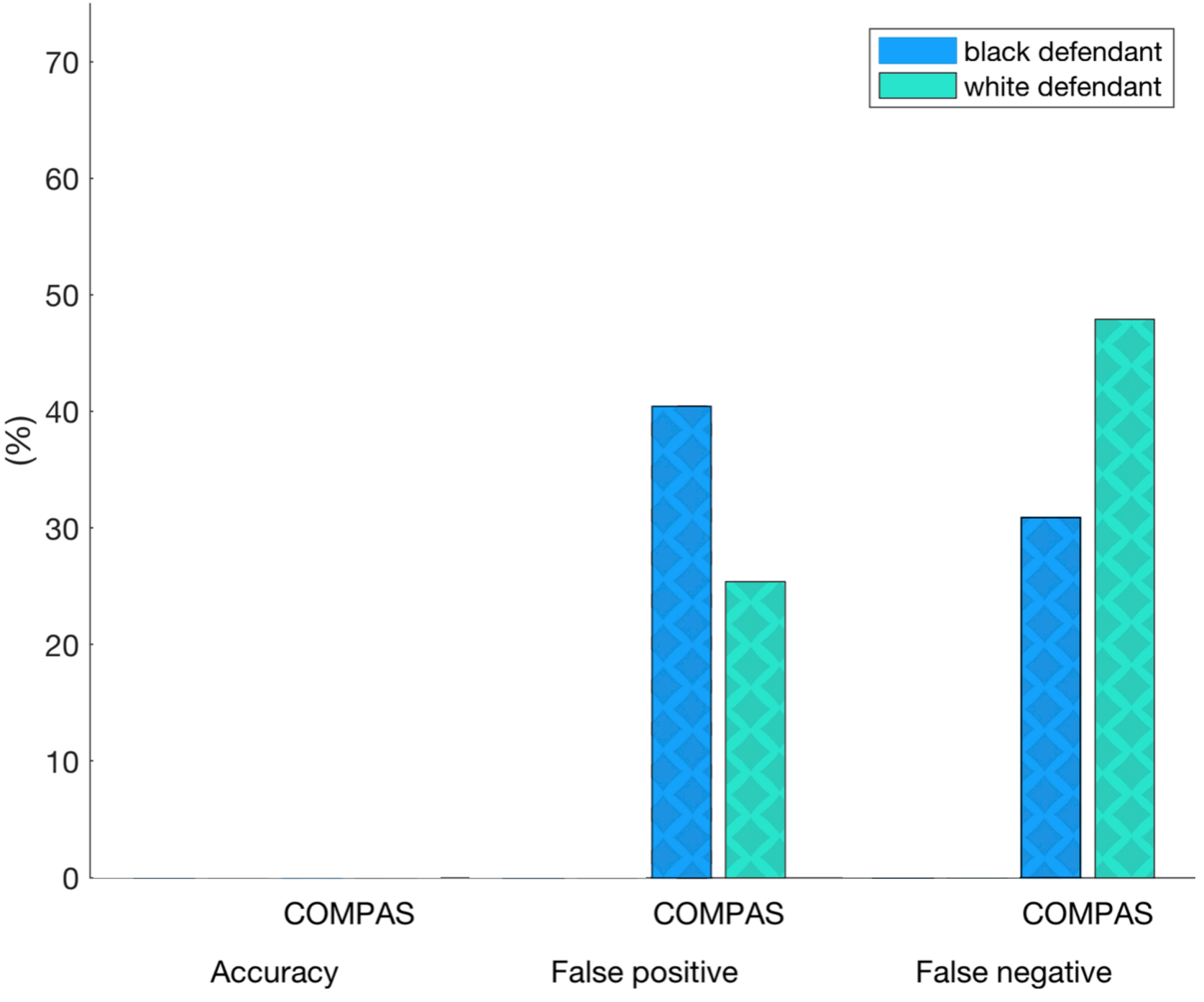
predicting recidivism

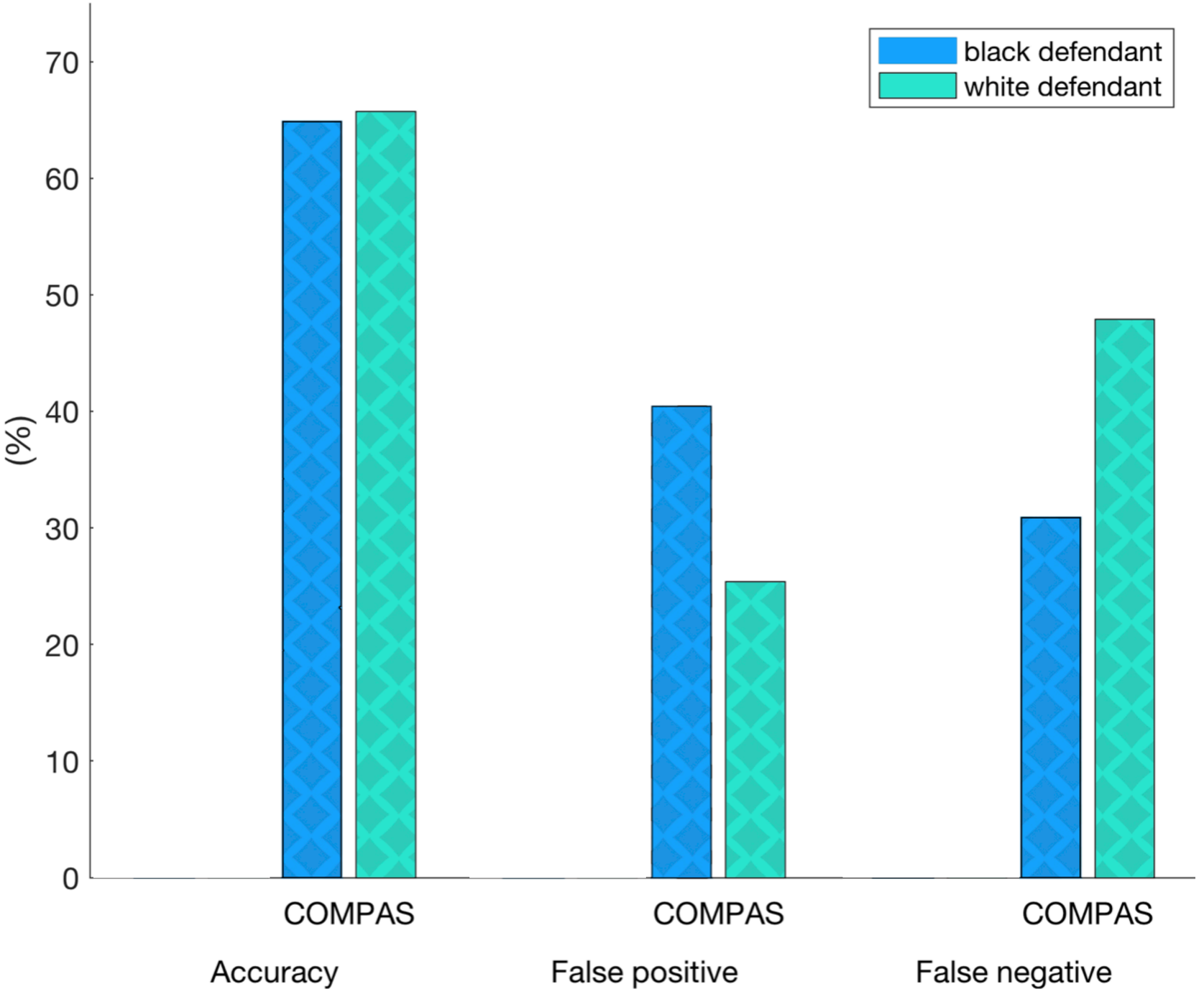
Machine Bias

There's software used across the country to predict future criminals. And it's biased against blacks.

by Julia Angwin, Jeff Larson, Surya Mattu and Lauren Kirchner, ProPublica

May 23, 2016


black defendant
white defendant
70
60
50
40
30
20
10
0
(%)
COMPAS
COMPAS
COMPAS
Accuracy
False positive
False negative


black defendant
white defendant
70
60
50
40
30
20
10
0
(%)
COMPAS
COMPAS
COMPAS
Accuracy
False positive
False negative

Your current accuracy is: 63%

The defendant is a male aged 43. They have been charged with: Disorderly Intoxication. This crime is classified as a misdemeanor. They have been convicted of 2 prior crimes. They have 0 juvenile felony charges and 0 juvenile misdemeanor charges on their record. Do you think this person will commit another crime within 2 years?

Yes

No

Disorderly Intoxication: When an intoxicated person endangers the safety of another person or property, or causes a disturbance in public

>>

Your current accuracy is: 63%

The defendant is a male aged 43. They have been charged with: Disorderly Intoxication. This crime is classified as a misdemeanor. They have been convicted of 2 prior crimes. They have 0 juvenile felony charges and 0 juvenile misdemeanor charges on their record. Do you think this person will commit another crime within 2 years?

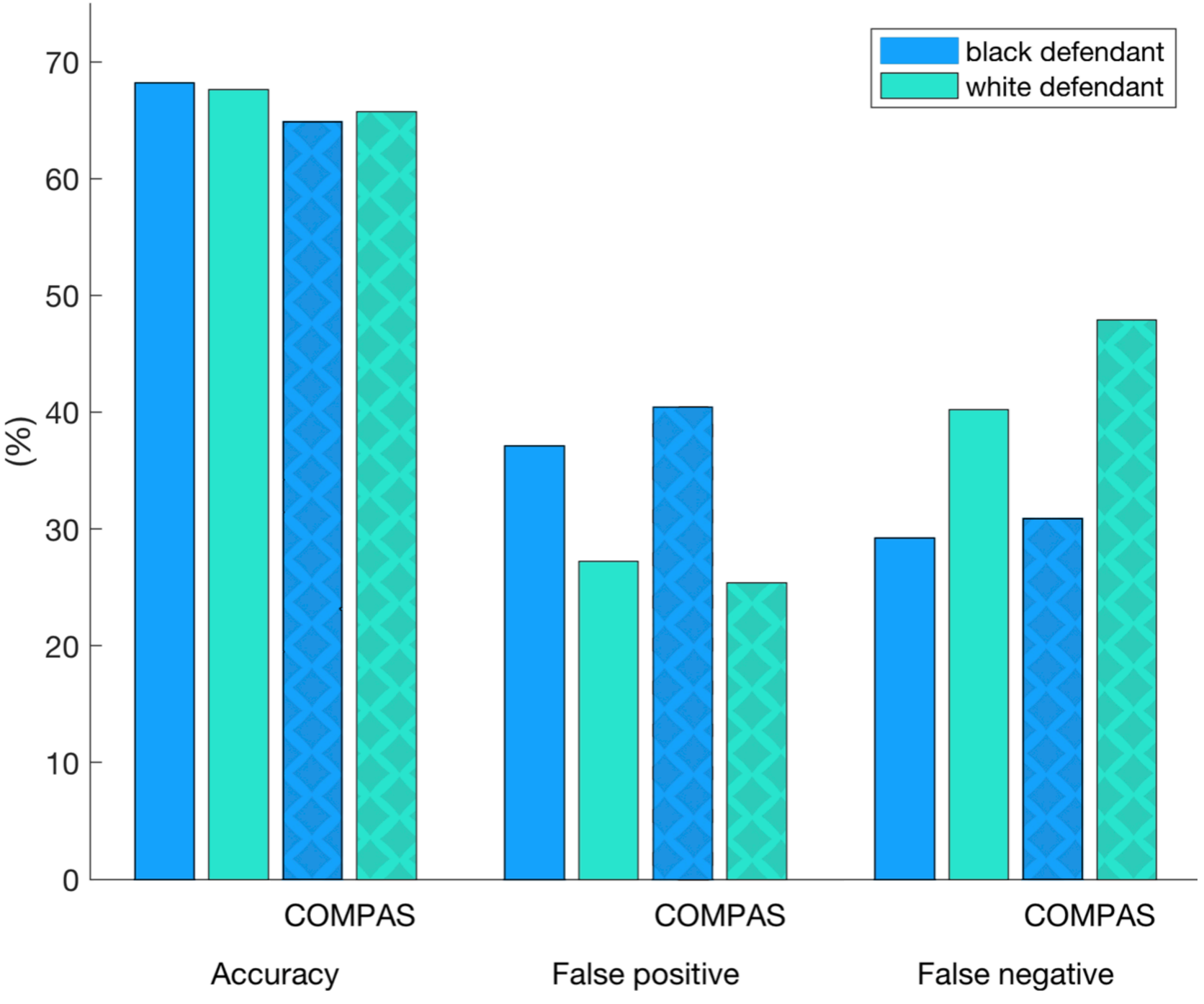
Yes

No

Disorderly Intoxication: When an intoxicated person endangers the safety of another person or property, or causes a disturbance in public

>>

- N = 400
- 50 questions/participant
- feedback
- catch trials
- $1 payment
- $5 bonus for > 65%


black defendant
white defendant
70
60
50
40
30
20
10
0
(%)
COMPAS
COMPAS
COMPAS
Accuracy
False positive
False negative

Your current accuracy is: 63%

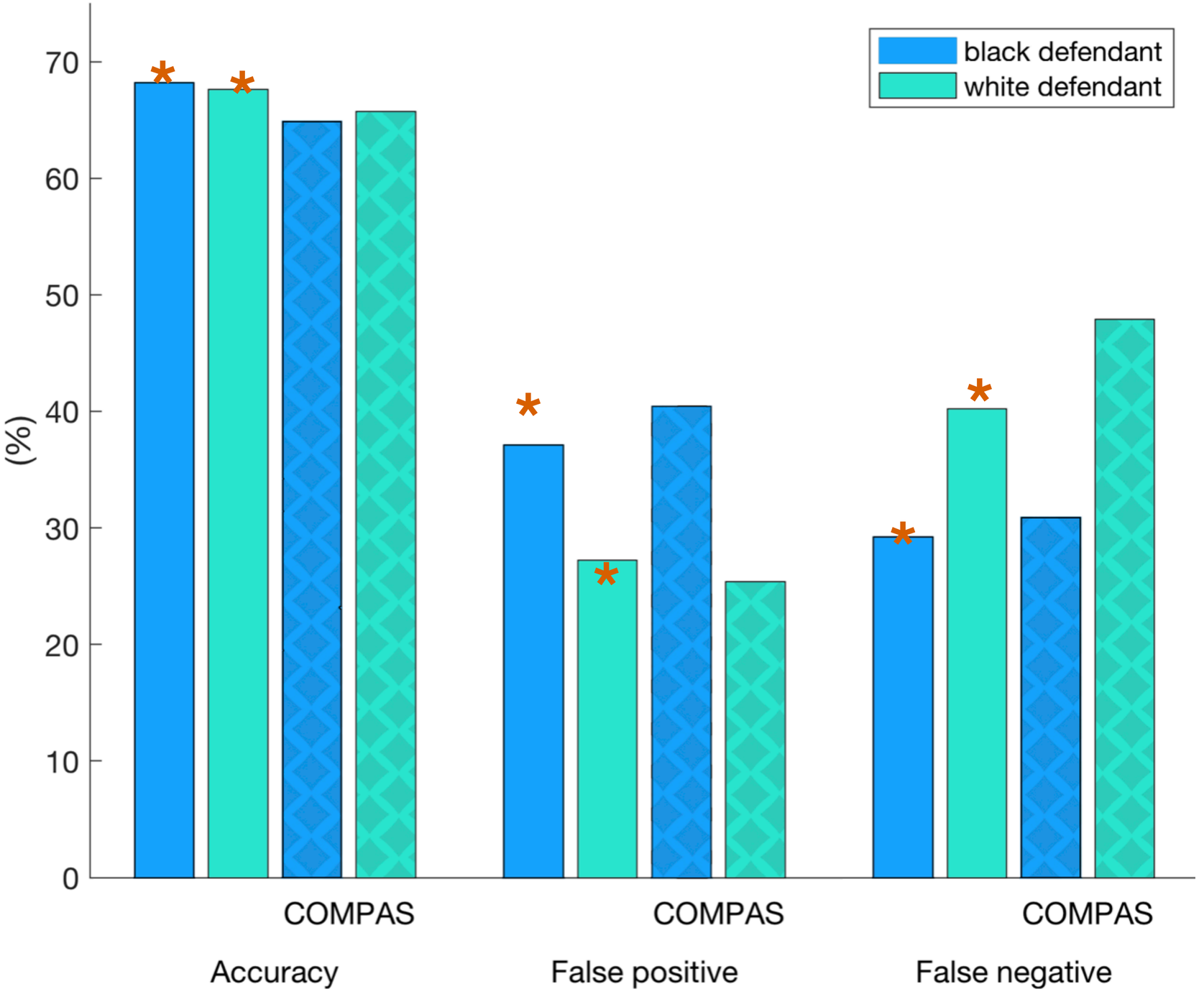
The defendant is a white male aged 43. They have been charged with: Disorderly Intoxication. This crime is classified as a misdemeanor. They have been convicted of 2 prior crimes. They have 0 juvenile felony charges and 0 juvenile misdemeanor charges on their record. Do you think this person will commit another crime within 2 years?

Yes

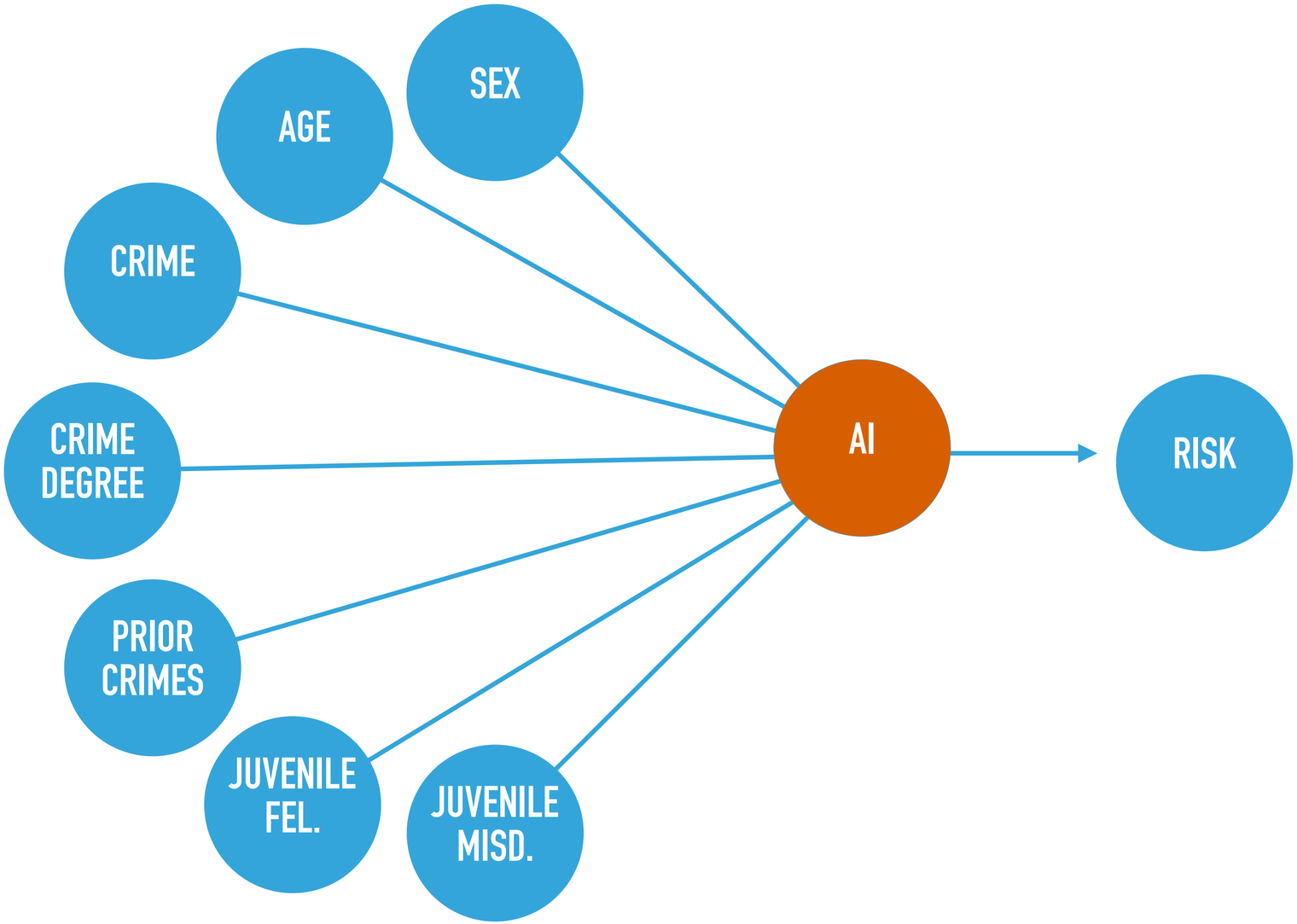
No

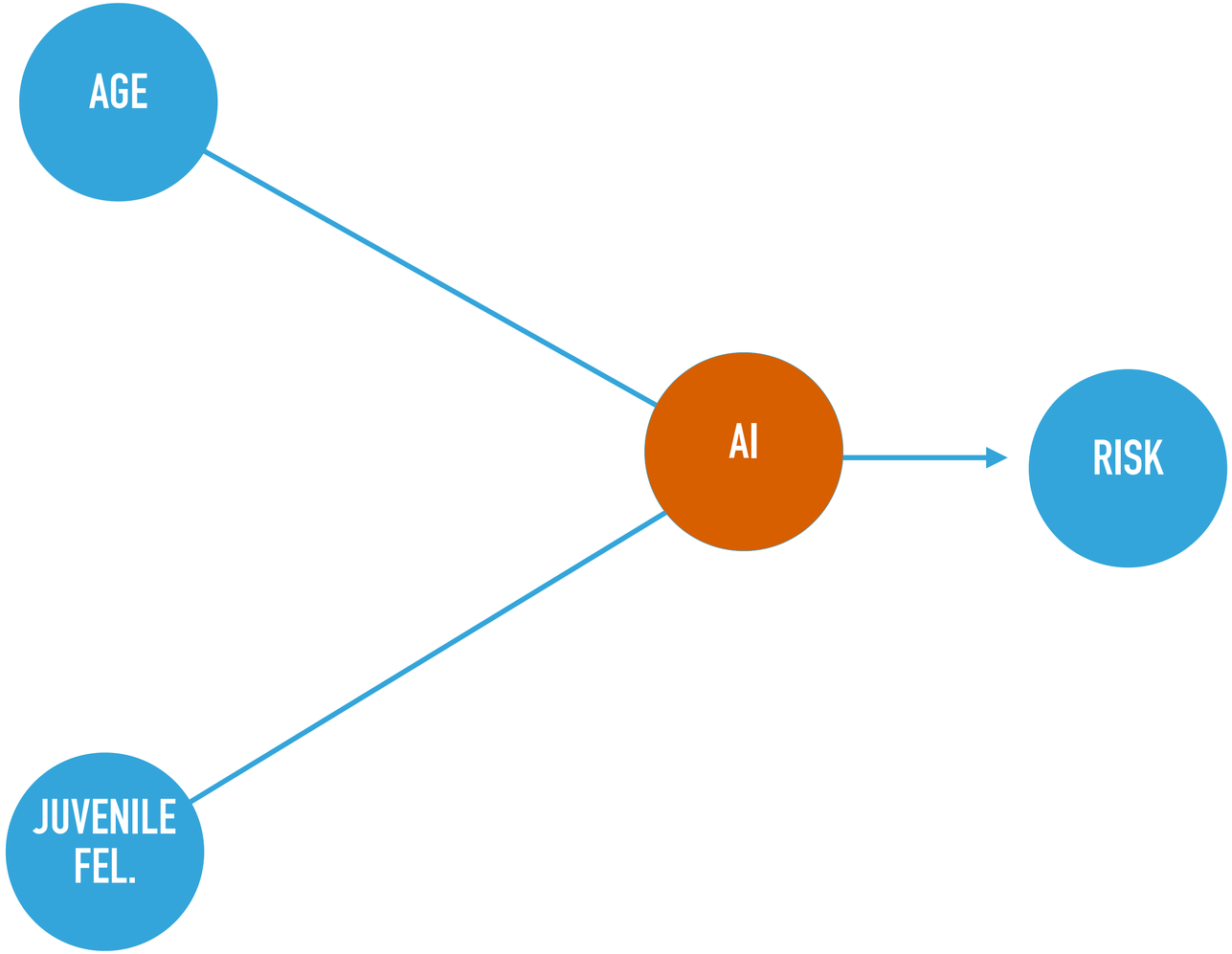
Disorderly Intoxication: When an intoxicated person endangers the safety of another person or property, or causes a disturbance in public

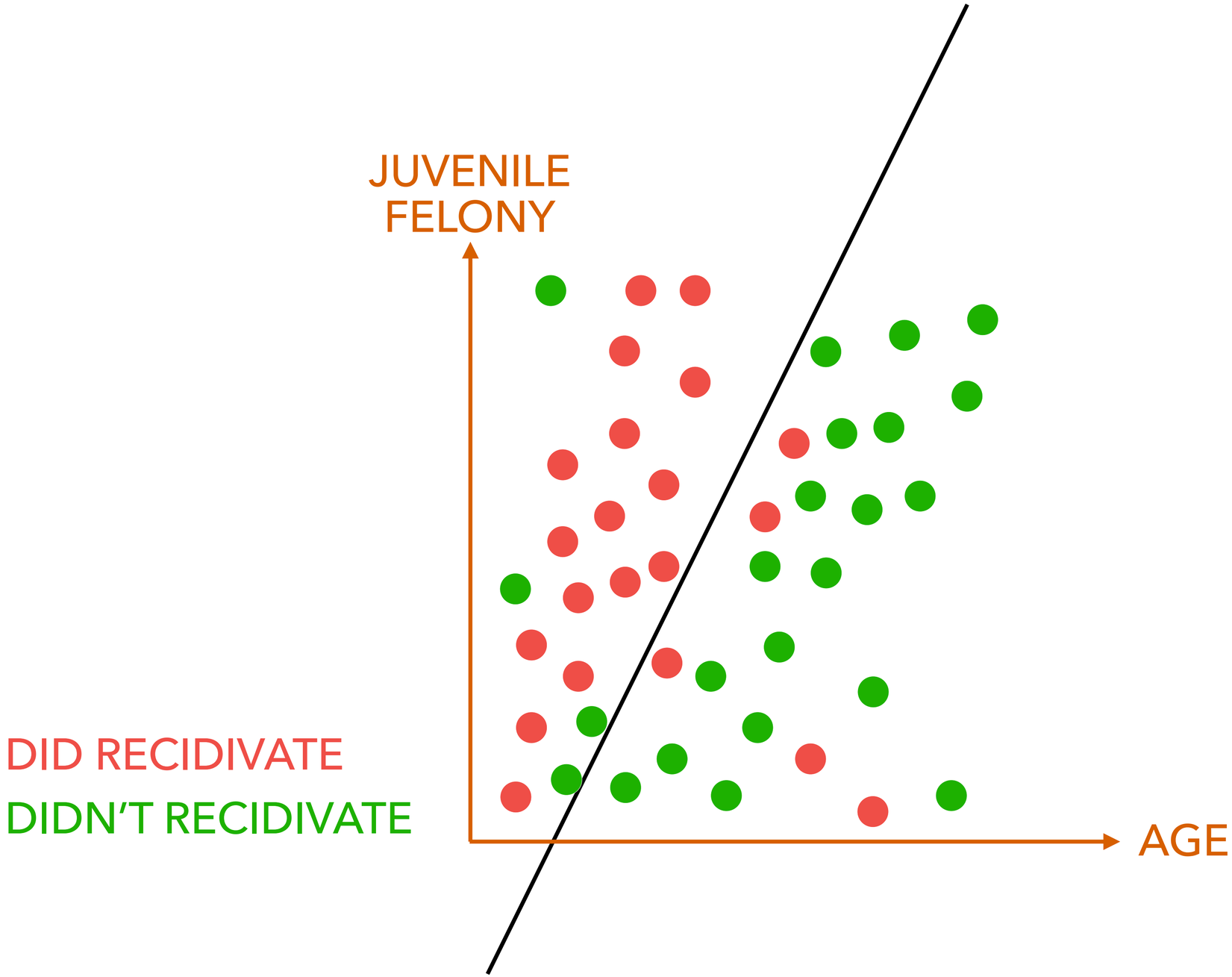
>>

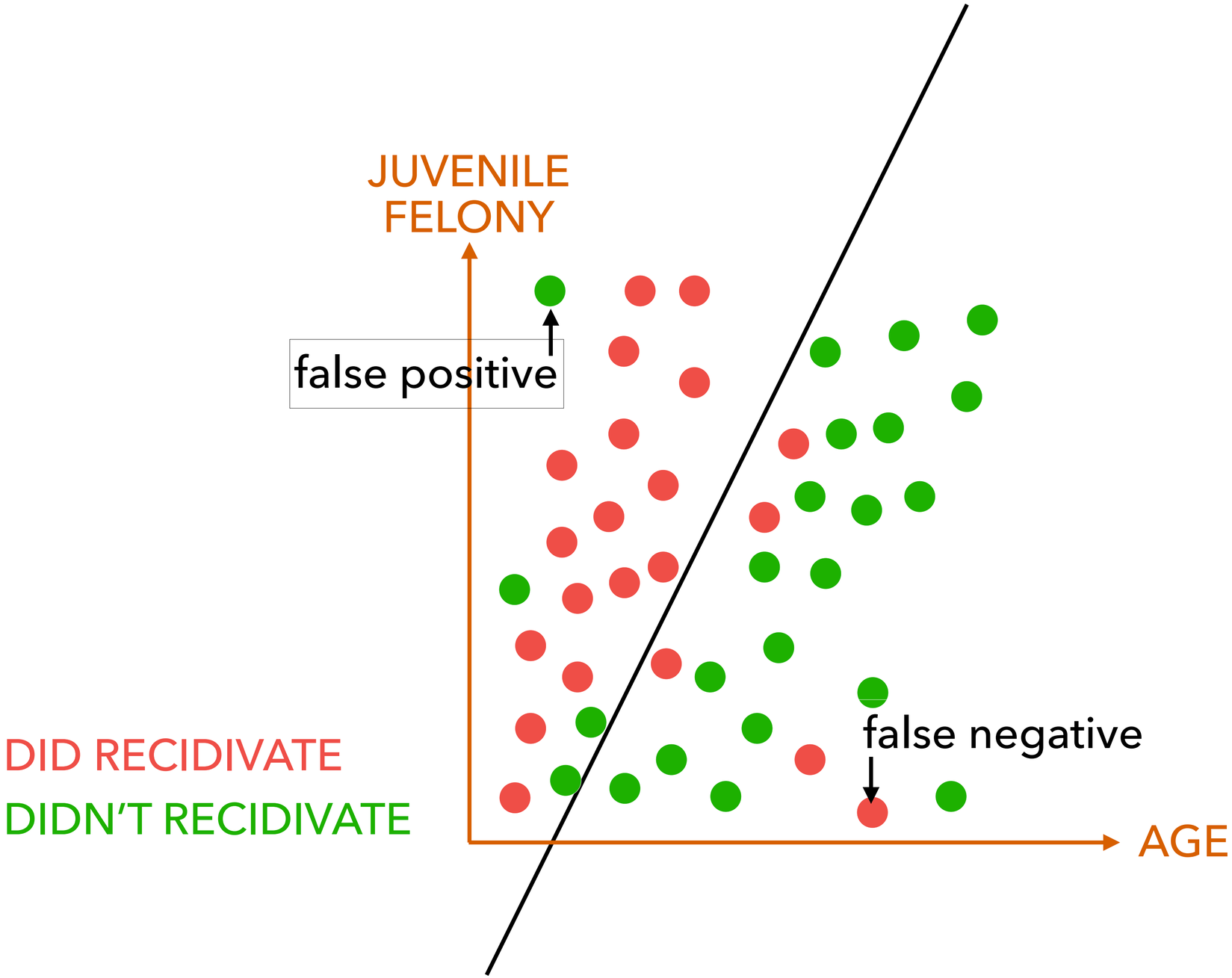

black defendant
white defendant
(%)
70
60
50
40
30
20
10
0
COMPAS
COMPAS
COMPAS
Accuracy
False positive
False negative

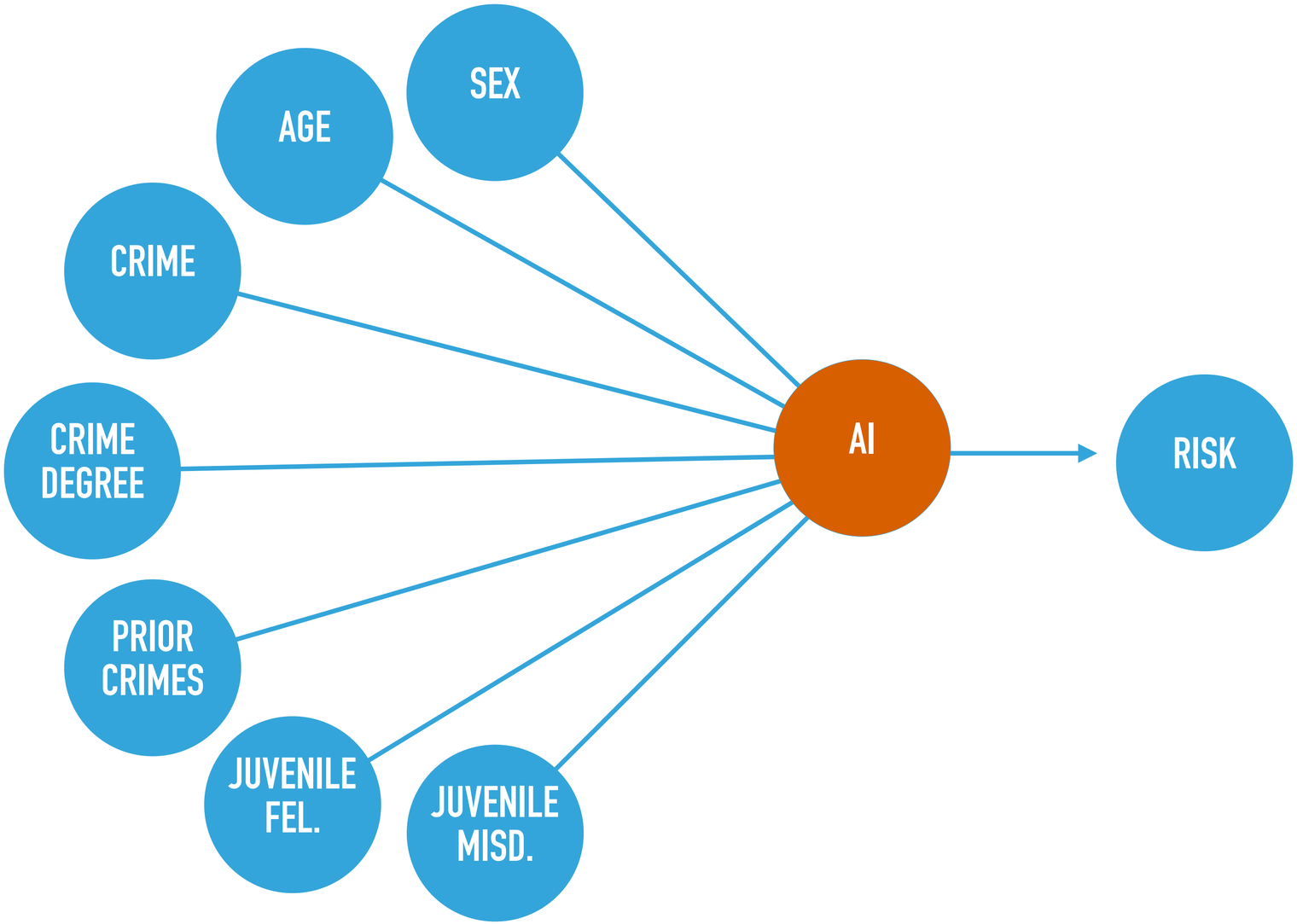
1. non-experts = AI?
2. biased without race?

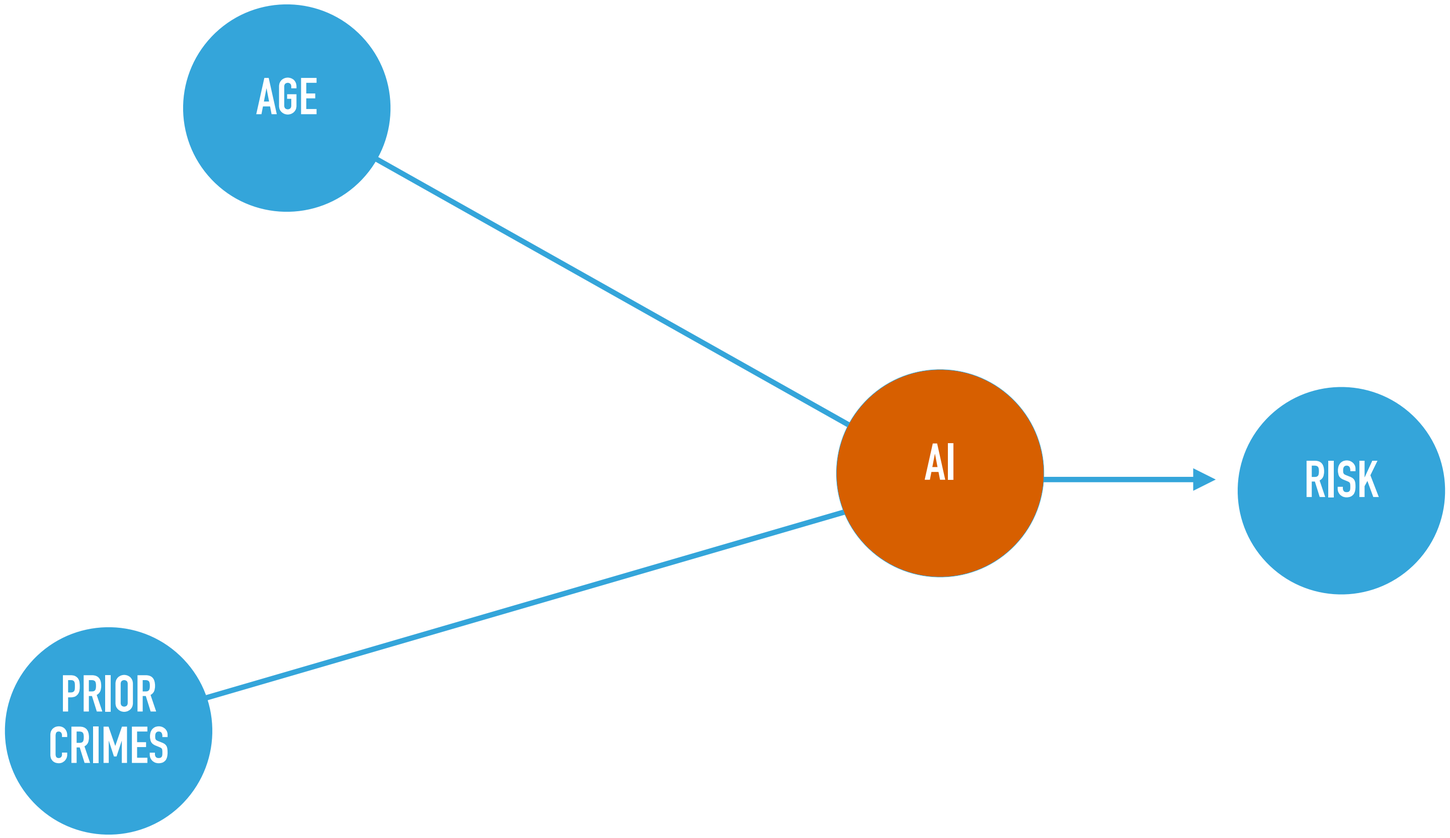

SEX
AGE
CRIME
CRIME DEGREE
PRIOR CRIMES
JUVENILE FEL.
JUVENILE MISD.
AI
RISK

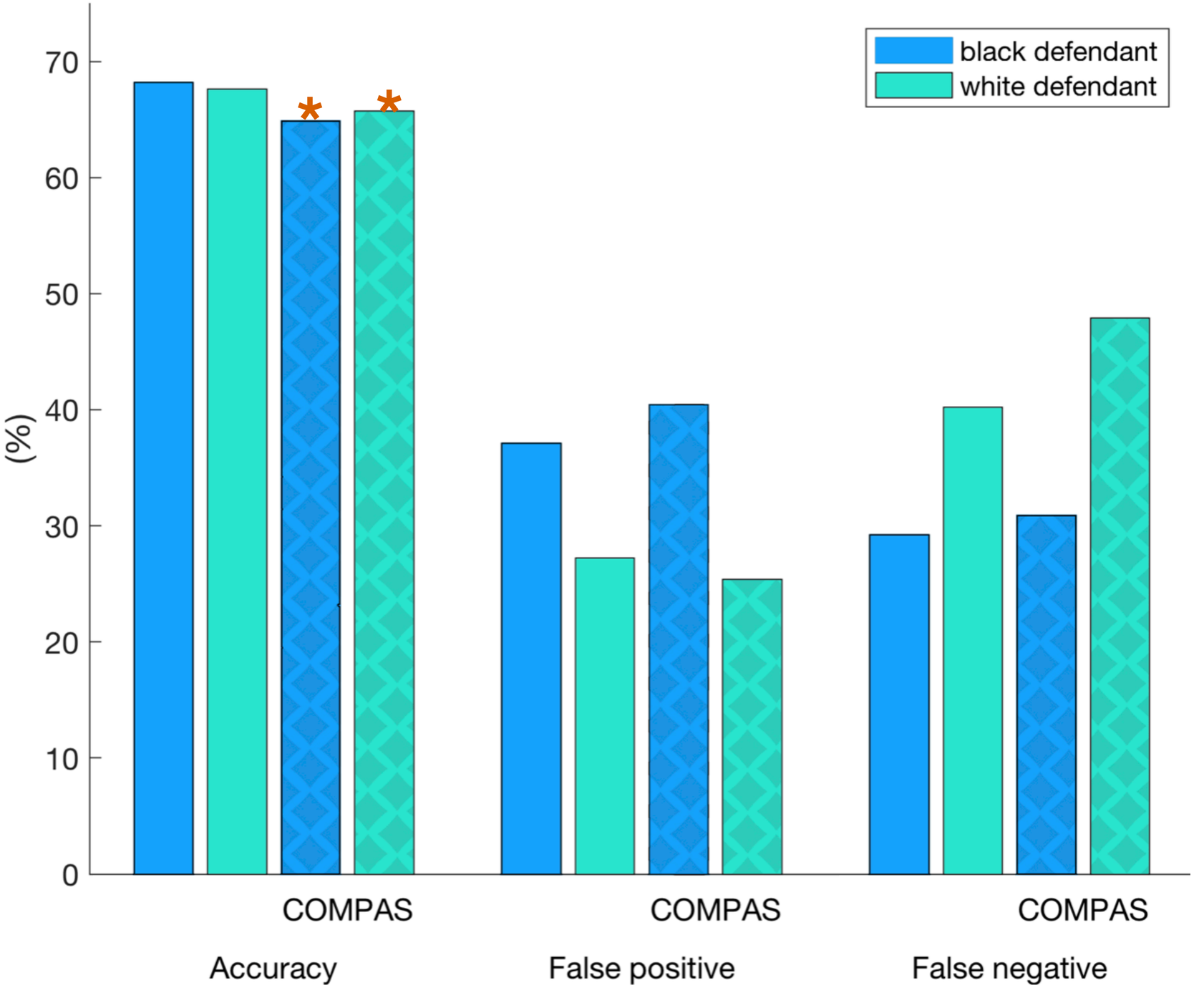

AGE
AI
RISK
JUVENILE
FEL.

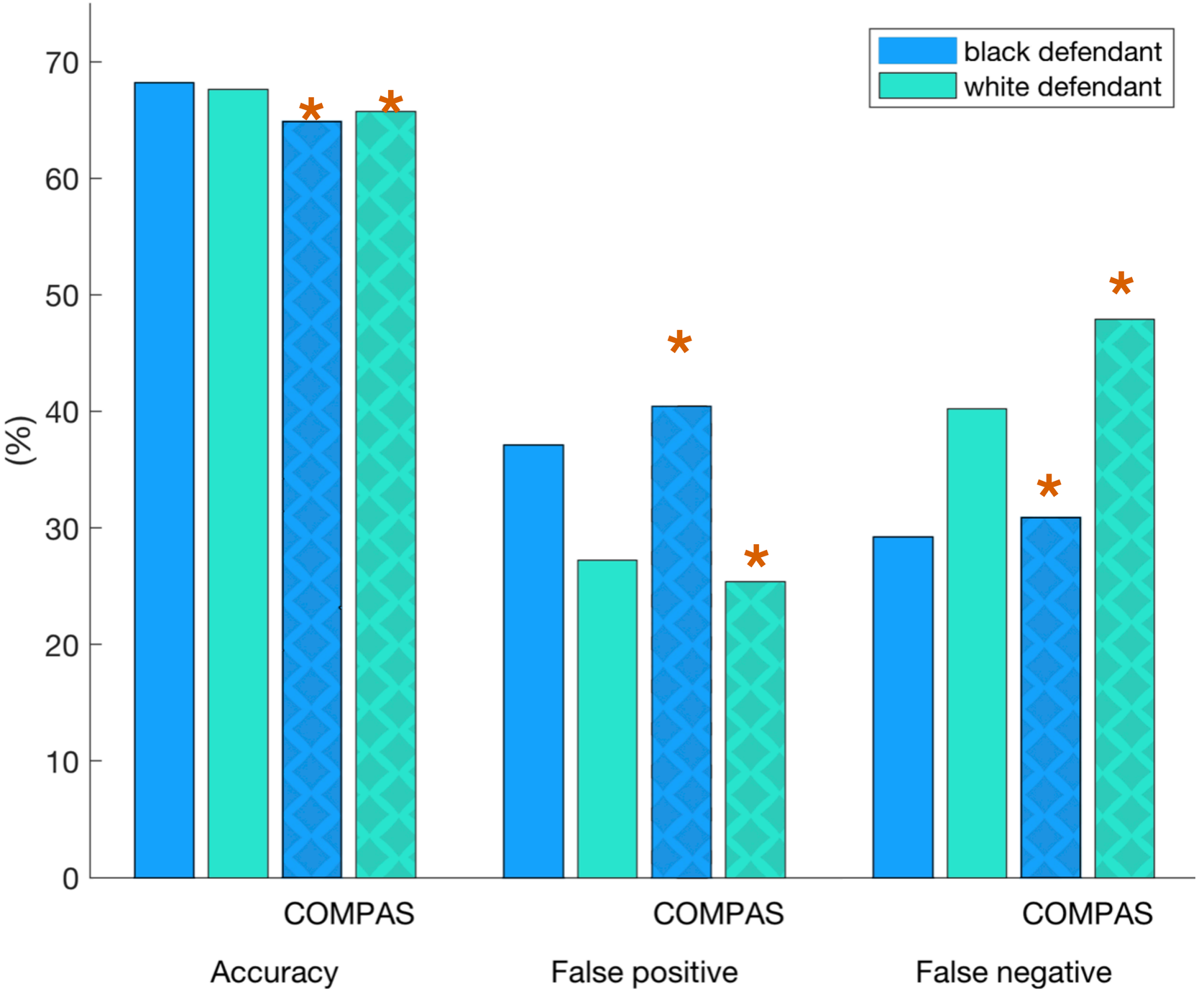

JUVENILE
FELONY
DID RECIDIVATE
DIDN'T RECIDIVATE
AGE


JUVENILE
FELONY
false positive
false negative
DID RECIDIVATE
DIDN'T RECIDIVATE
AGE

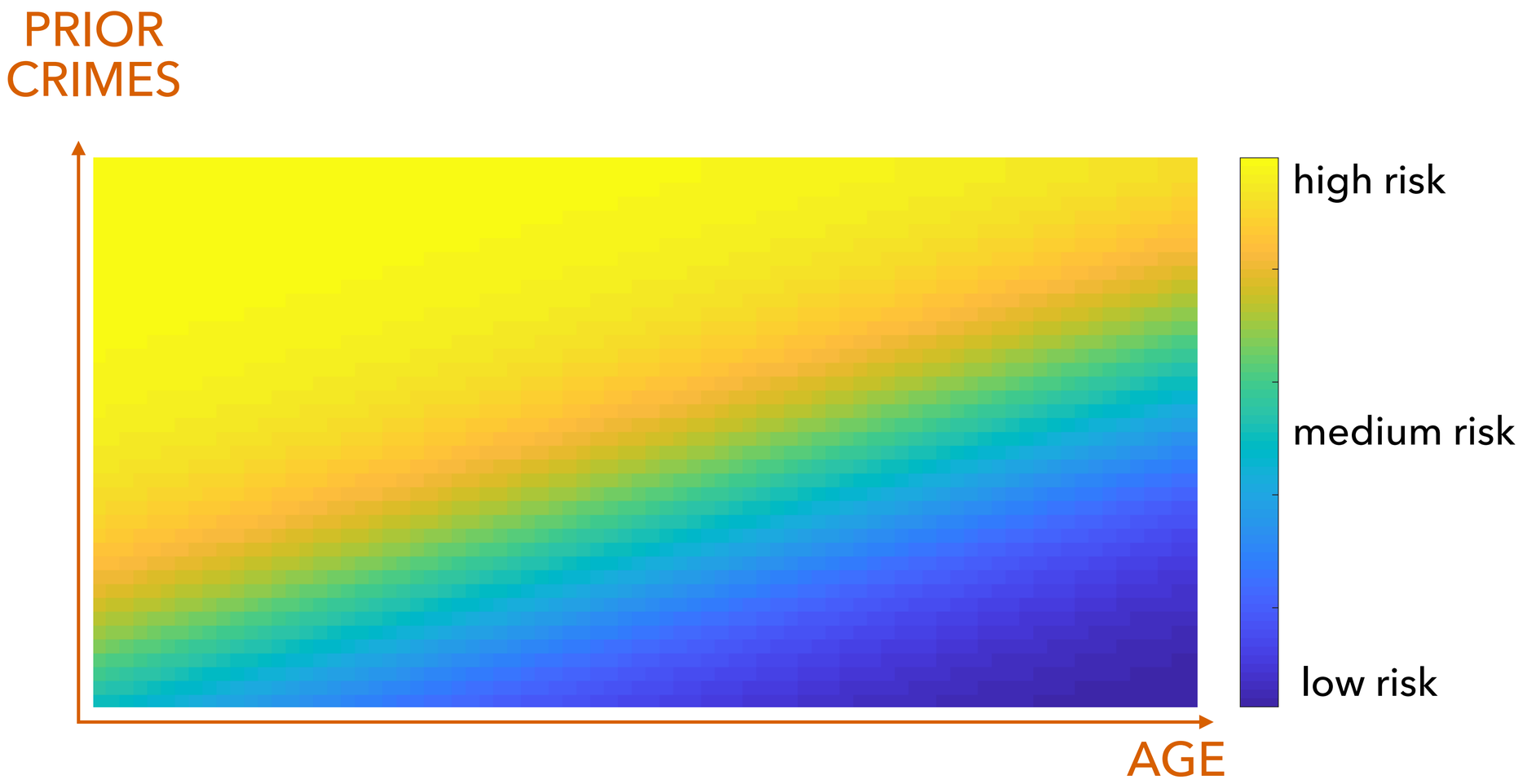

SEX
AGE
CRIME
CRIME DEGREE
PRIOR CRIMES
JUVENILE FEL.
JUVENILE MISD.
AI
RISK


AGE
AI
RISK
PRIOR
CRIMES


black defendant
white defendant
70
60
50
40
30
20
10
0
(%)
*
*
COMPAS
COMPAS
COMPAS
Accuracy
False positive
False negative


black defendant
white defendant
(%)
70
60
50
40
30
20
10
0
COMPAS
COMPAS
COMPAS
Accuracy
False positive
False negative

1. non-experts = AI?
2. biased without race?


PRIOR
CRIMES
AGE
high risk
medium risk
low risk

2. Prior Crimes is a proxy for Race

1. classification is easy (but not accurate)

so what?

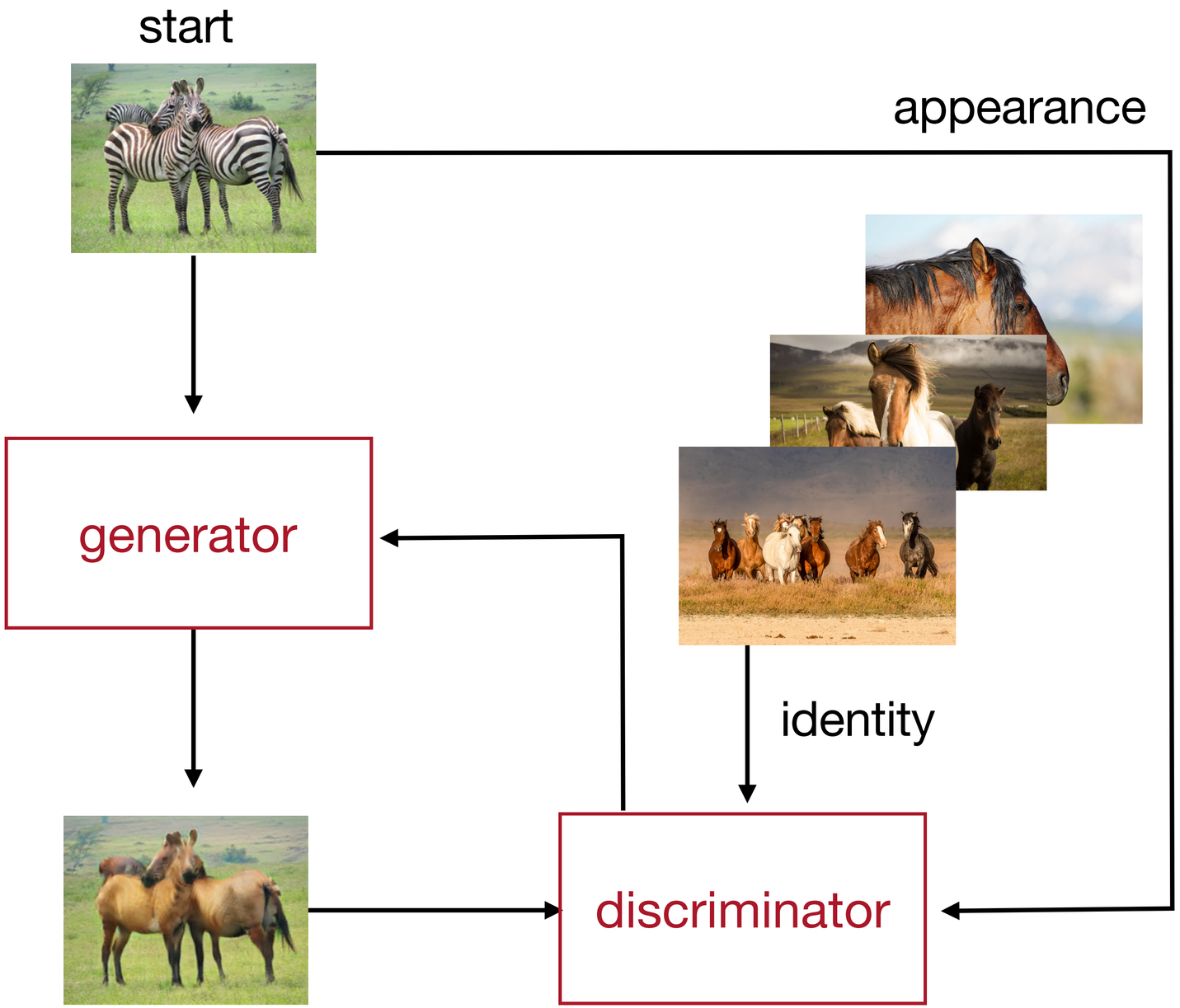
Explainable AI/ML is important

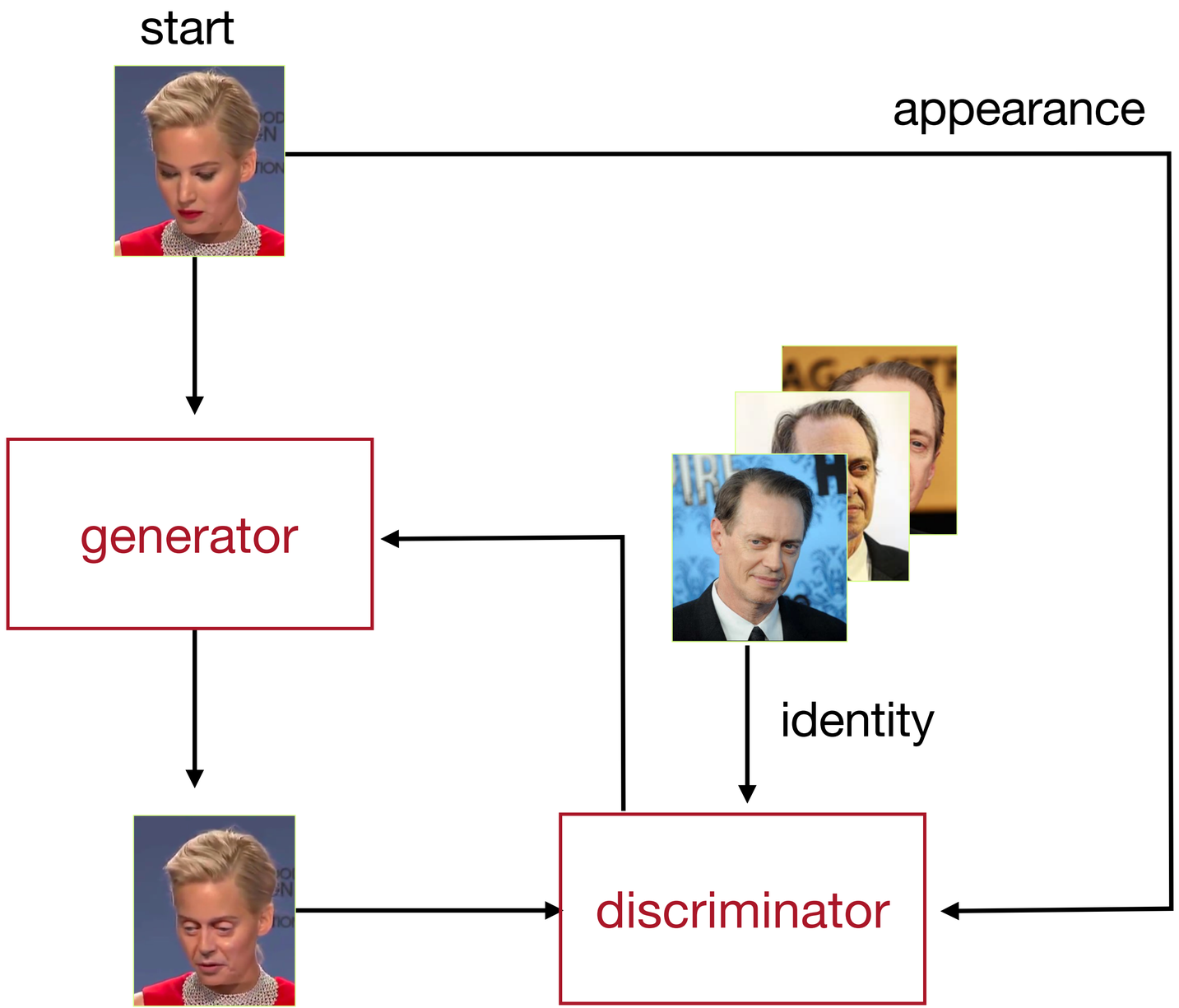
The best way to repeat history is to train AI/ML on historical data

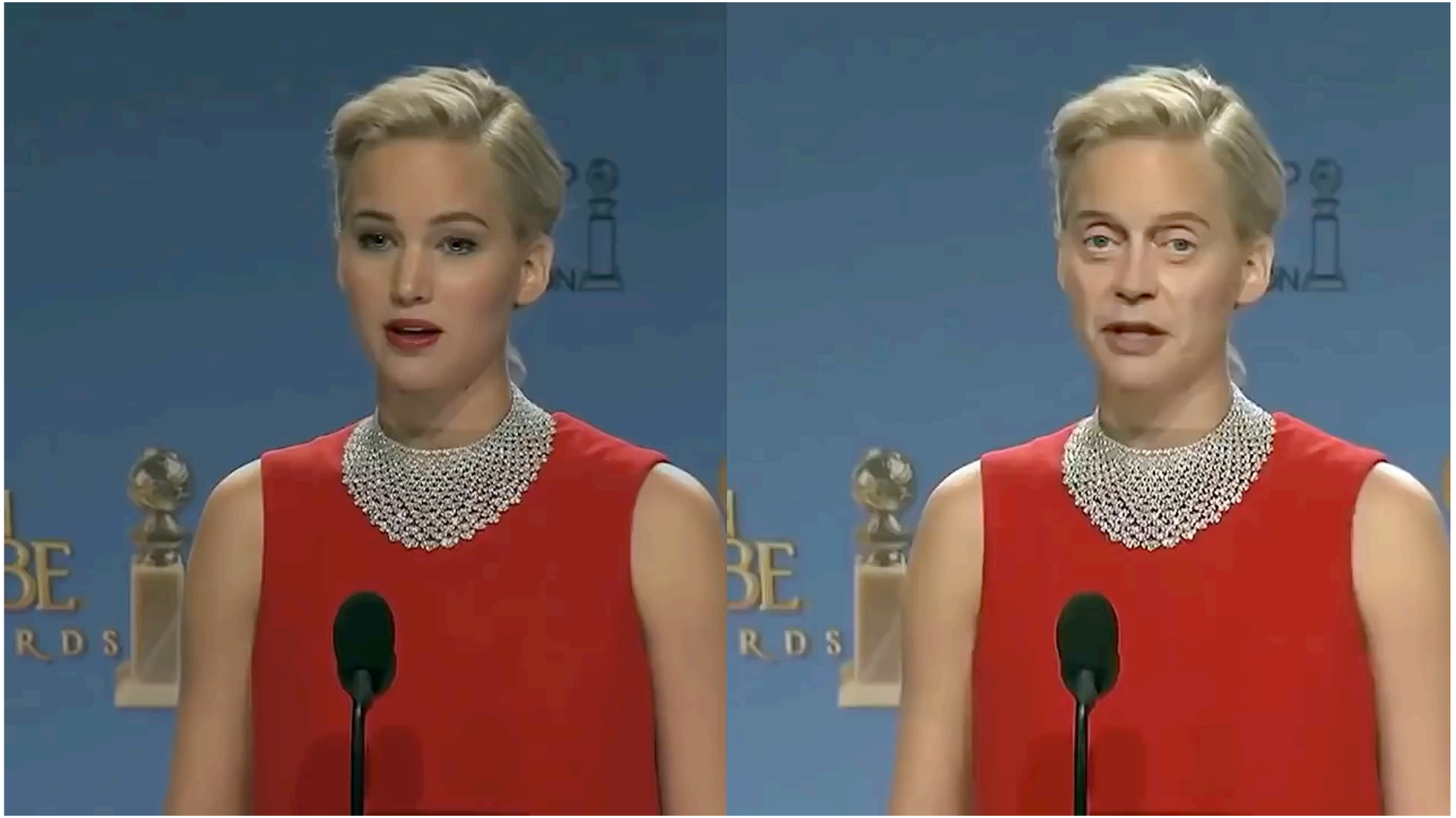
J. Dressel and H. Farid. The Dangers of Risk Prediction in the Criminal Justice System. MIT Case Studies in Social and Ethical Responsibilities of Computing, 2021.

non-consensual imagery

image-to-image translation
(deep fakes)


start
appearance
generator
identity
discriminator


start
appearance
generator
identity
discriminator




TikTok
@deeptomcruise
DeepTomCruise

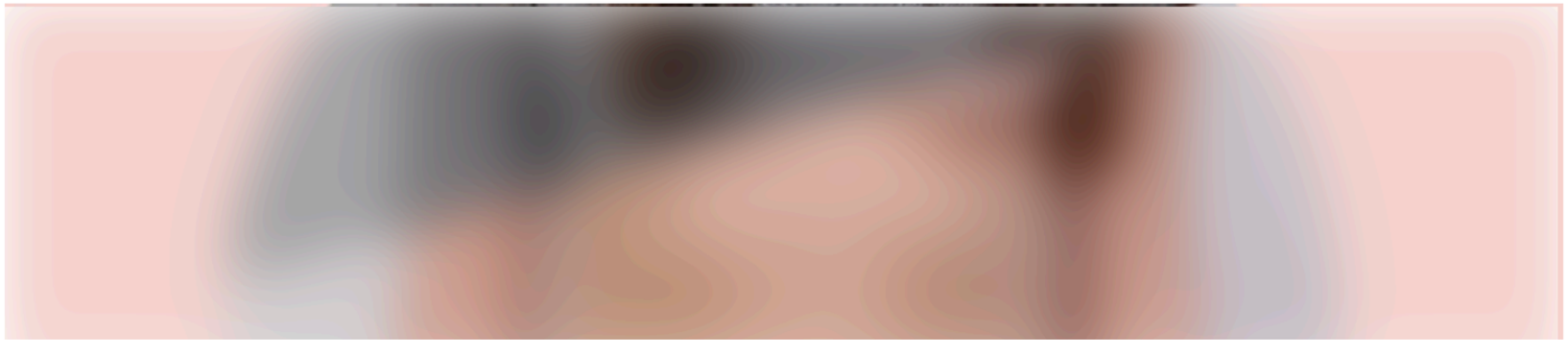
MOTHERBOARD
TECH BY VICE

This Horrifying App Undresses a Photo of Any Woman With a Single Click

The $50 DeepNude app dispenses with the idea that deepfakes were about anything besides claiming ownership over women's bodies.

By Samantha Cole

June 26, 2019, 2:48pm Share Tweet Snap



technologists build because they can,

not necessarily because they should

“move fast and break things”

was, is, and always will be a dumb motto

(Un)Ethical Computing