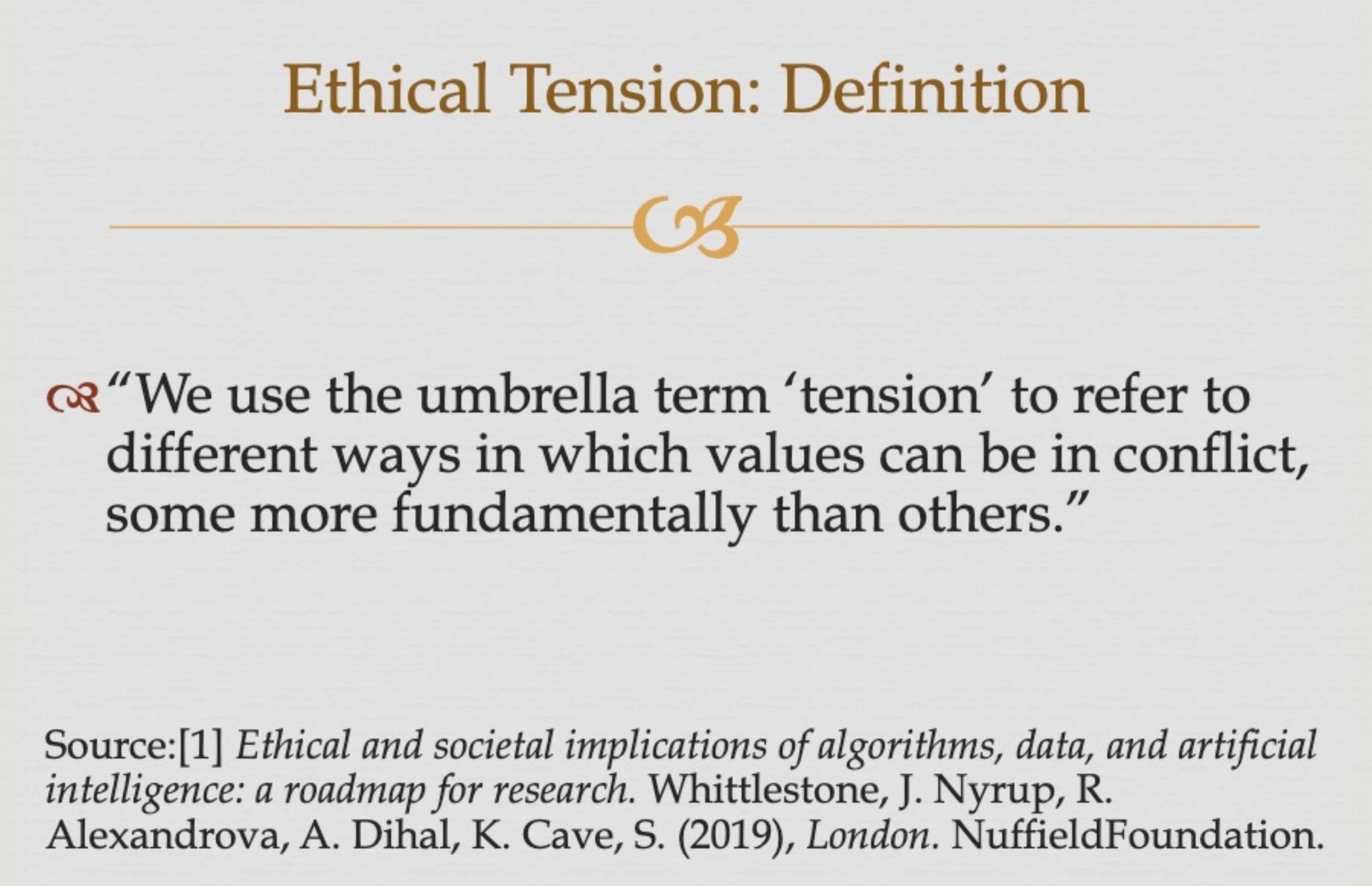
On the ethics of algorithmic decision-making in healthcare

Sune Holm, Associate Professor

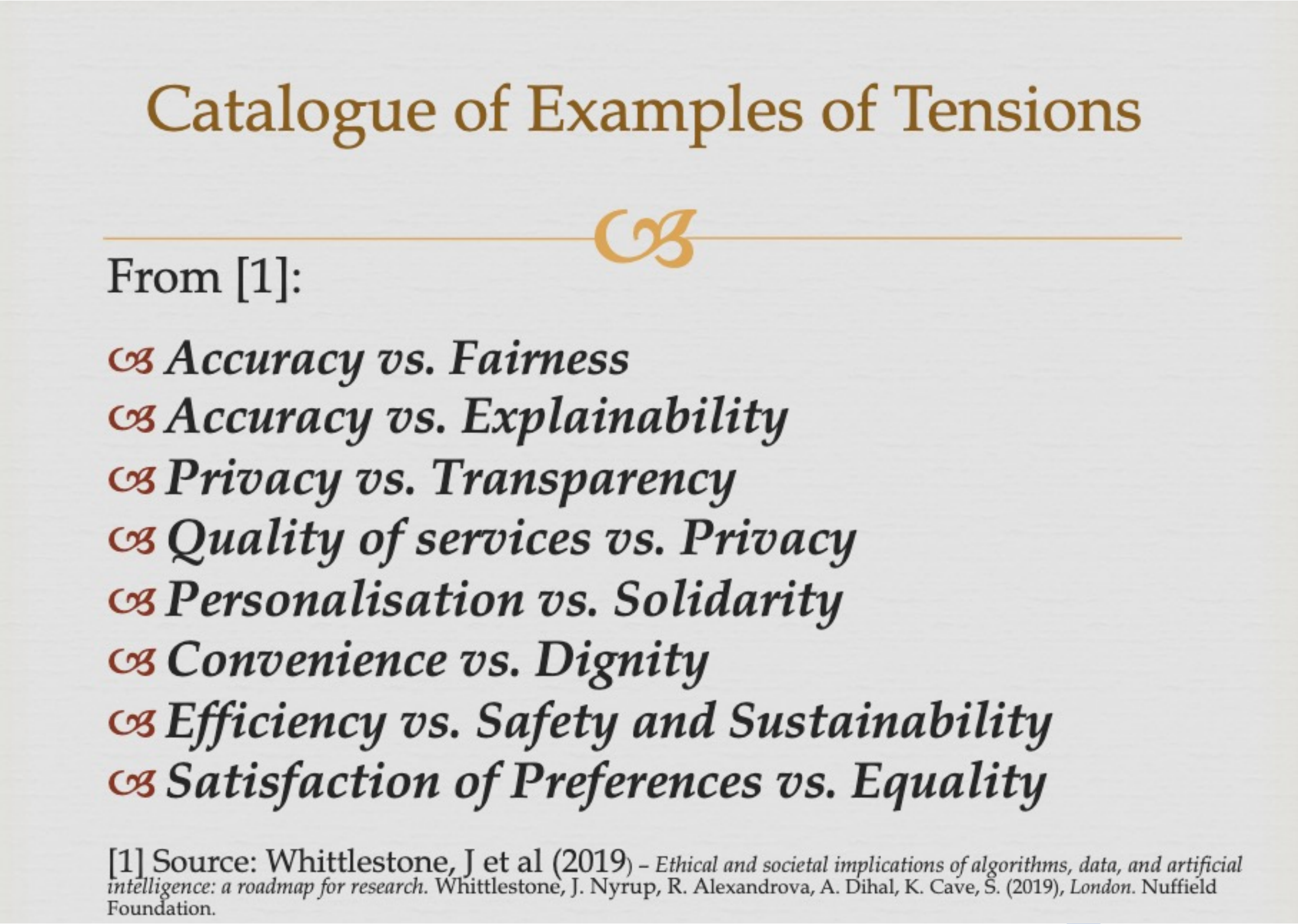
University of Copenhagen



Ethical Tension: Definition

- “We use the umbrella term ‘tension’ to refer to different ways in which values can be in conflict, some more fundamentally than others.”

Source:[1] *Ethical and societal implications of algorithms, data, and artificial intelligence: a roadmap for research.* Whittlestone, J. Nyrup, R. Alexandrova, A. Dihal, K. Cave, S. (2019), *London.* NuffieldFoundation.

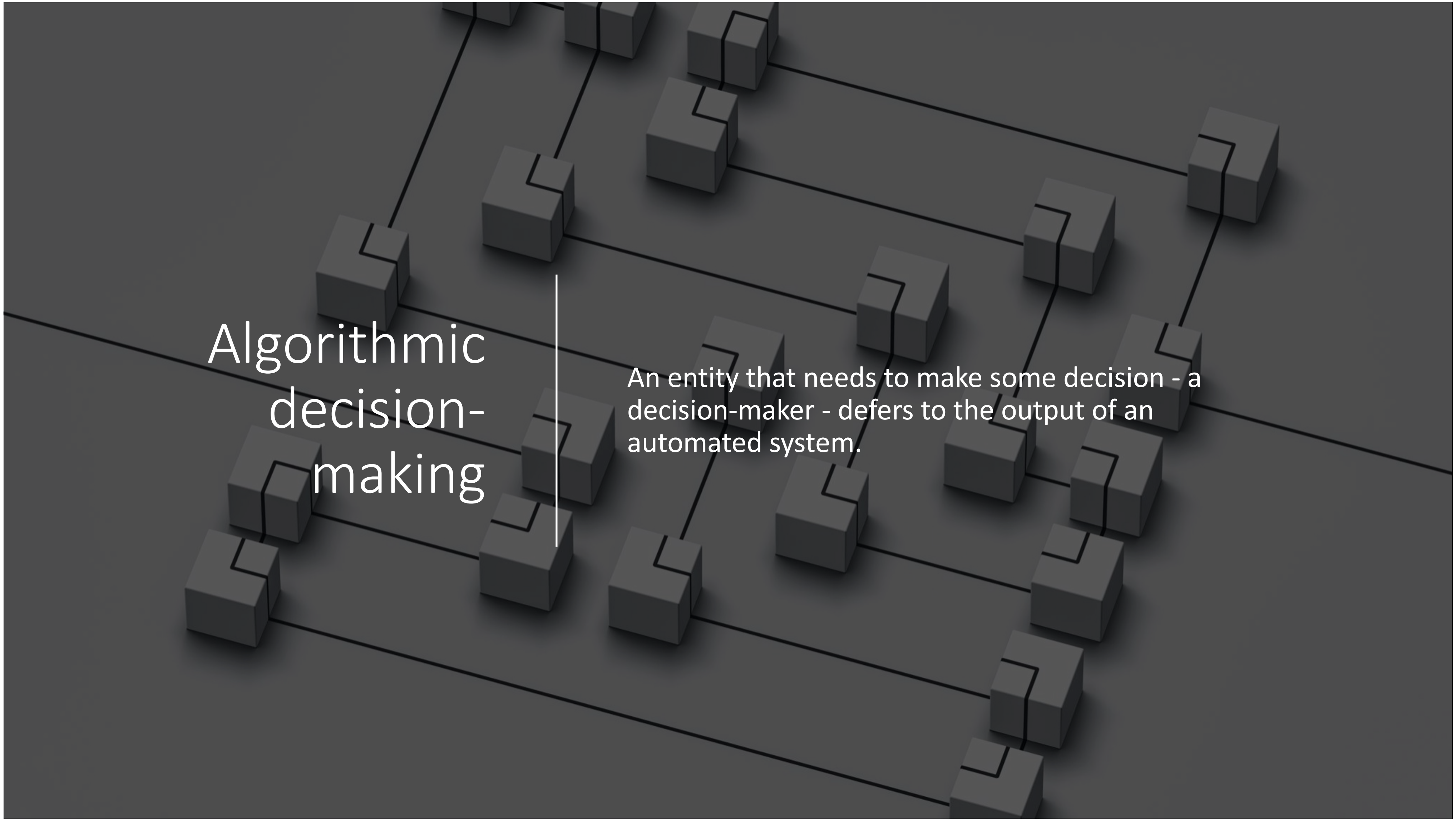


Catalogue of Examples of Tensions

From [1]:

- *Accuracy vs. Fairness*
- *Accuracy vs. Explainability*
- *Privacy vs. Transparency*
- *Quality of services vs. Privacy*
- *Personalisation vs. Solidarity*
- *Convenience vs. Dignity*
- *Efficiency vs. Safety and Sustainability*
- *Satisfaction of Preferences vs. Equality*

[1] Source: Whittlestone, J et al (2019) – *Ethical and societal implications of algorithms, data, and artificial intelligence: a roadmap for research.* Whittlestone, J. Nyrup, R. Alexandrova, A. Dihal, K. Cave, S. (2019), *London.* Nuffield Foundation.

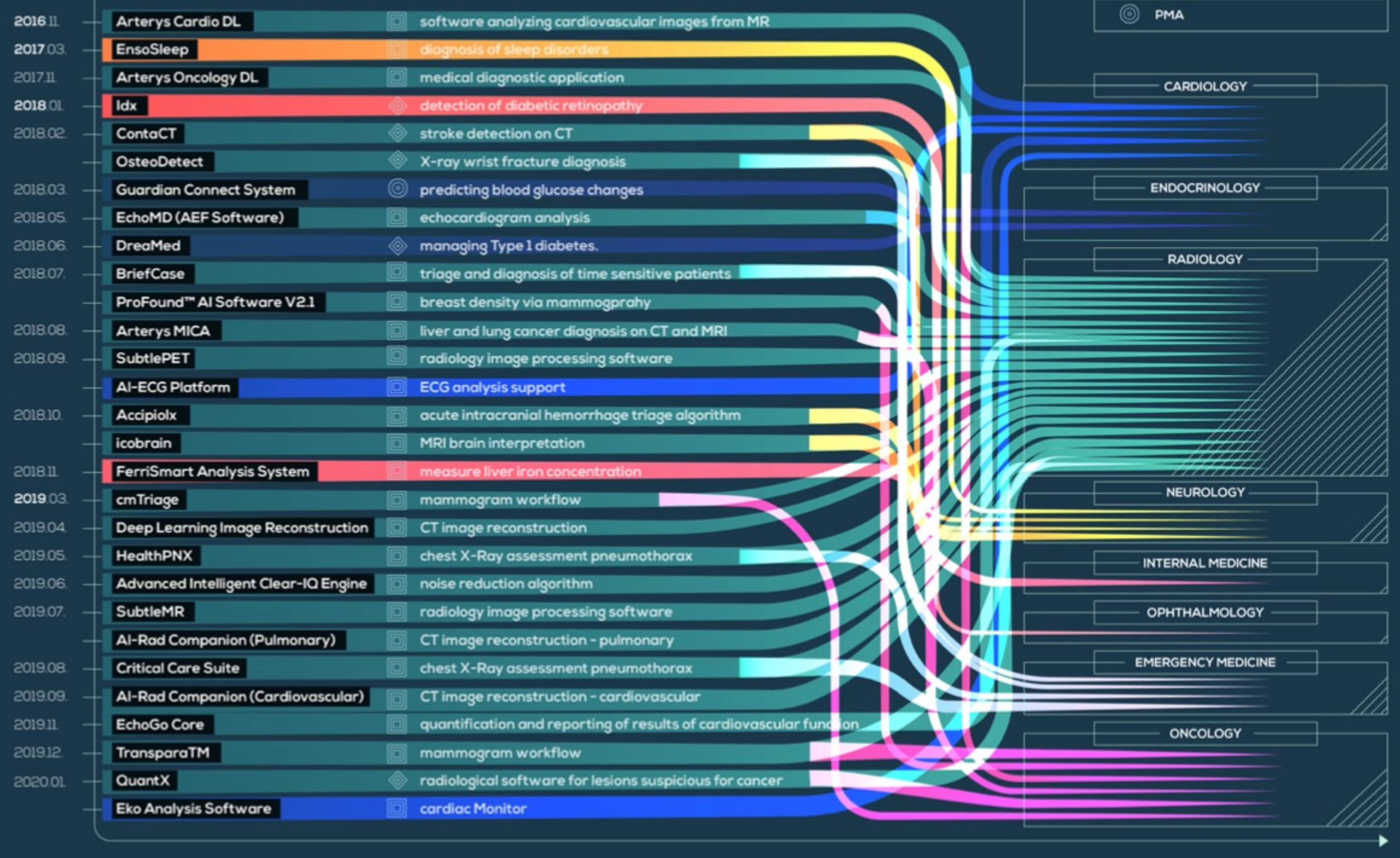


Algorithmic decision-making

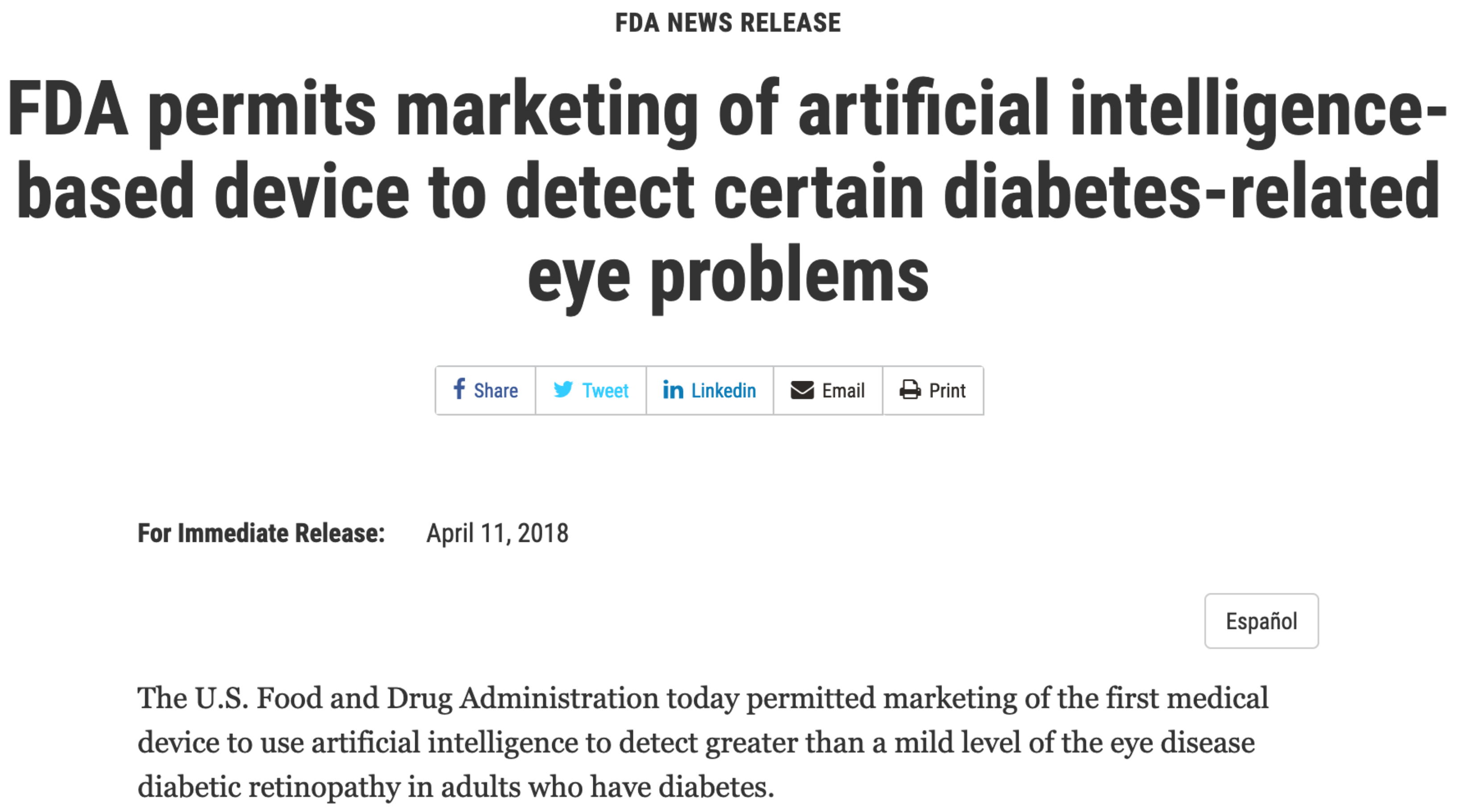
An entity that needs to make some decision - a decision-maker - defers to the output of an automated system.

FDA APPROVALS FOR ARTIFICIAL INTELLIGENCE-BASED DEVICES IN MEDICINE

TYPE OF FDA APPROVAL
□ 510(K) PREMARKET NOTIFICATION
◇ DE NOVO PATHWAY
◎ PMA



2016.11.	Arterys Cardio DL	□	software analyzing cardiovascular images from MR
2017.03.	EnsoSleep	□	diagnosis of sleep disorders
2017.11.	Arterys Oncology DL	□	medical diagnostic application
2018.01.	Idx	◇	detection of diabetic retinopathy
2018.02.	ContaCT	◇	stroke detection on CT
	OsteoDetect	◇	X-ray wrist fracture diagnosis
2018.03.	Guardian Connect System	◎	predicting blood glucose changes
2018.05.	EchoMD (AEF Software)	□	echocardiogram analysis
2018.06.	DreaMed	◇	managing Type 1 diabetes.
2018.07.	BriefCase	□	triage and diagnosis of time sensitive patients
	ProFound™ AI Software V2.1	□	breast density via mammogprahy
2018.08.	Arterys MICA	□	liver and lung cancer diagnosis on CT and MRI
2018.09.	SubtlePET	□	radiology image processing software
	AI-ECG Platform	□	ECG analysis support
2018.10.	Accipiolx	□	acute intracranial hemorrhage triage algorithm
	icobrain	□	MRI brain interpretation
2018.11.	FerriSmart Analysis System	□	measure liver iron concentration
2019.03.	cmTriage	□	mammogram workflow
2019.04.	Deep Learning Image Reconstruction	□	CT image reconstruction
2019.05.	HealthPNX	□	chest X-Ray assessment pneumothorax
2019.06.	Advanced Intelligent Clear-IQ Engine	□	noise reduction algorithm
2019.07.	SubtleMR	□	radiology image processing software
	AI-Rad Companion (Pulmonary)	□	CT image reconstruction - pulmonary
2019.08.	Critical Care Suite	□	chest X-Ray assessment pneumothorax
2019.09.	AI-Rad Companion (Cardiovascular)	□	CT image reconstruction - cardiovascular
2019.11.	EchoGo Core	□	quantification and reporting of results of cardiovascular function
2019.12.	TransparaTM	□	mammogram workflow
2020.01.	QuantX	◇	radiological software for lesions suspicious for cancer
	Eko Analysis Software	□	cardiac Monitor

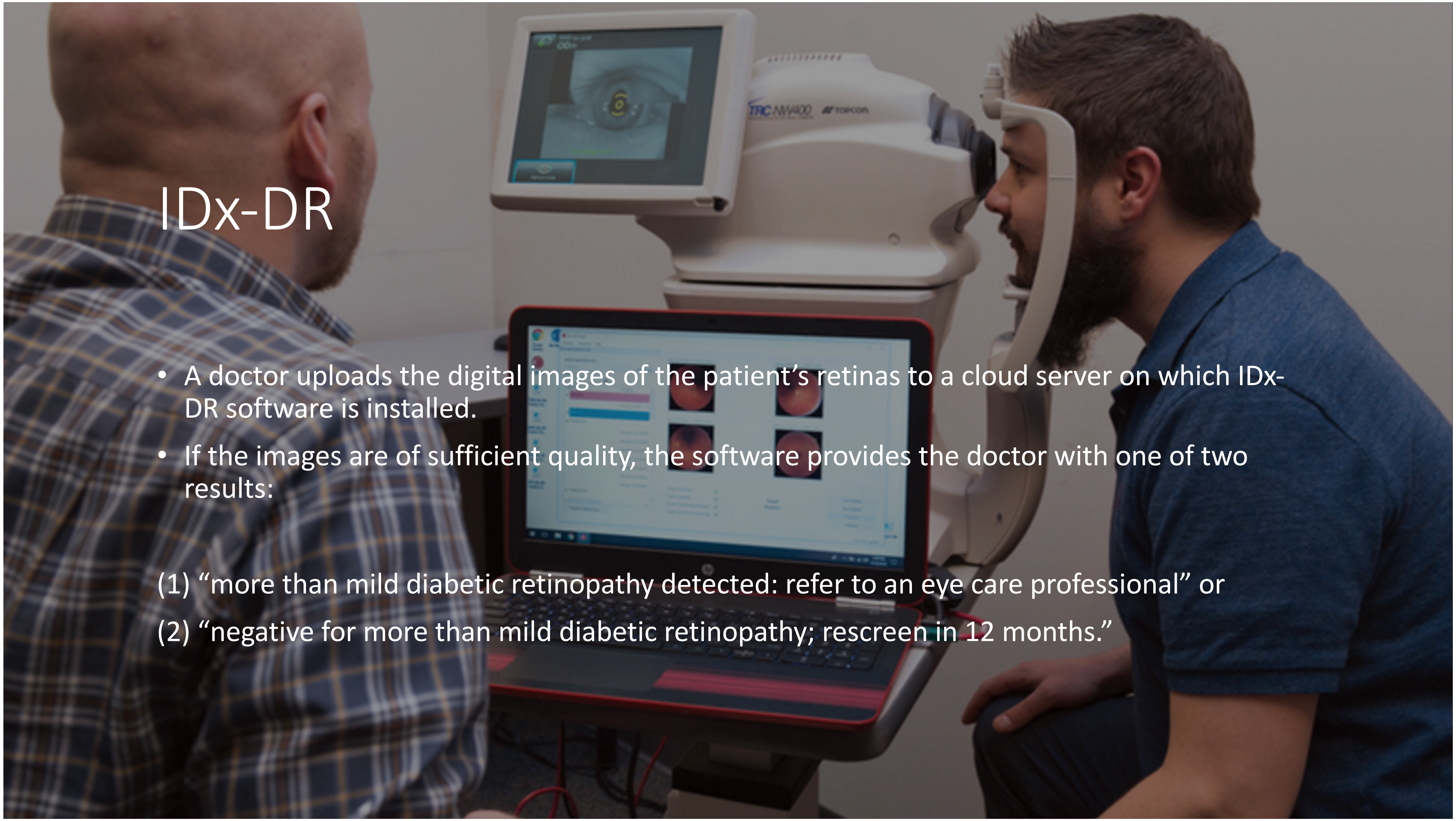


FDA NEWS RELEASE

FDA permits marketing of artificial intelligence-based device to detect certain diabetes-related eye problems

For Immediate Release: April 11, 2018

The U.S. Food and Drug Administration today permitted marketing of the first medical device to use artificial intelligence to detect greater than a mild level of the eye disease diabetic retinopathy in adults who have diabetes.

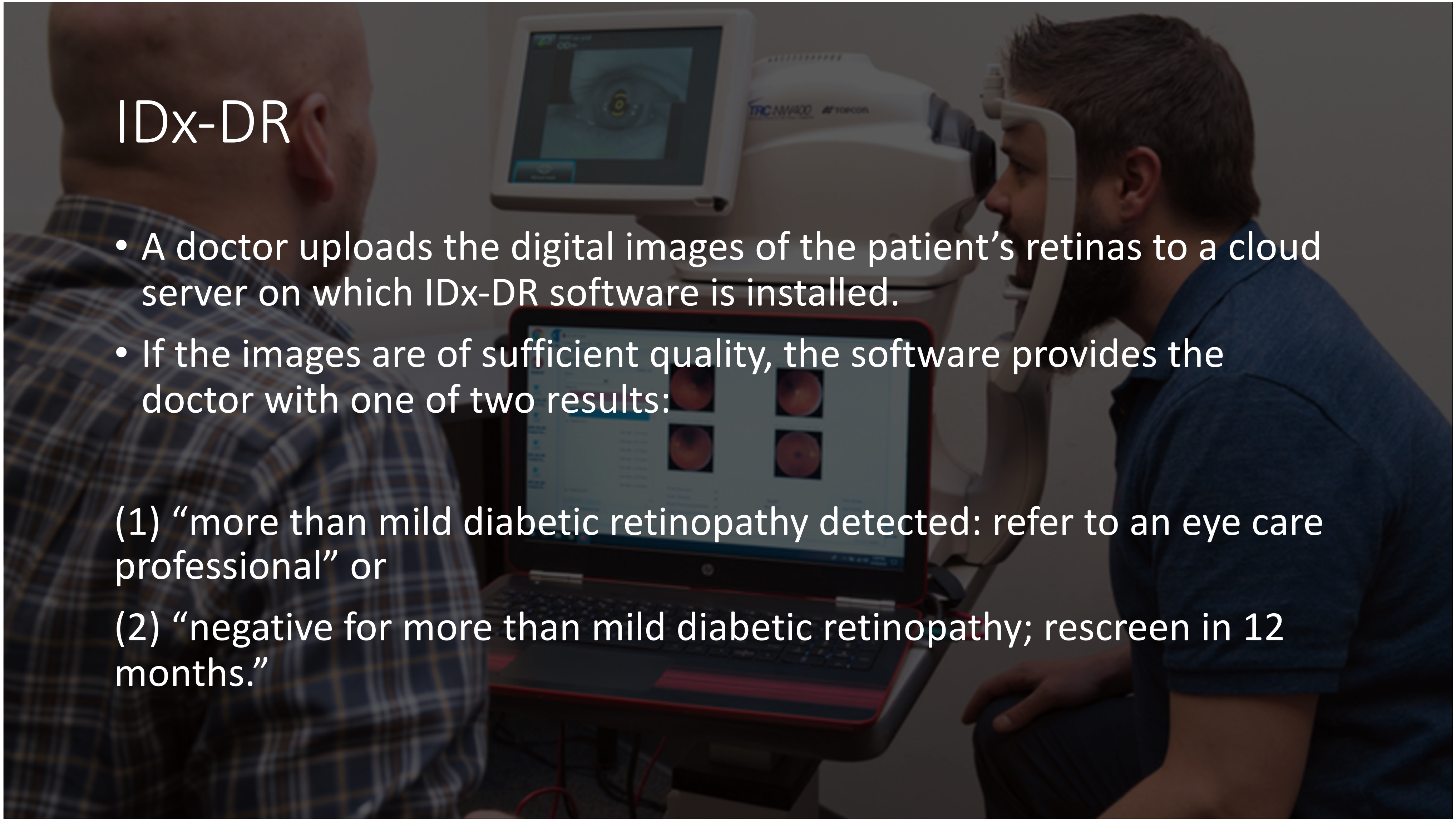


IDx-DR

- A doctor uploads the digital images of the patient's retinas to a cloud server on which IDx-DR software is installed.
- If the images are of sufficient quality, the software provides the doctor with one of two results:

(1) "more than mild diabetic retinopathy detected: refer to an eye care professional" or

(2) "negative for more than mild diabetic retinopathy; rescreen in 12 months."

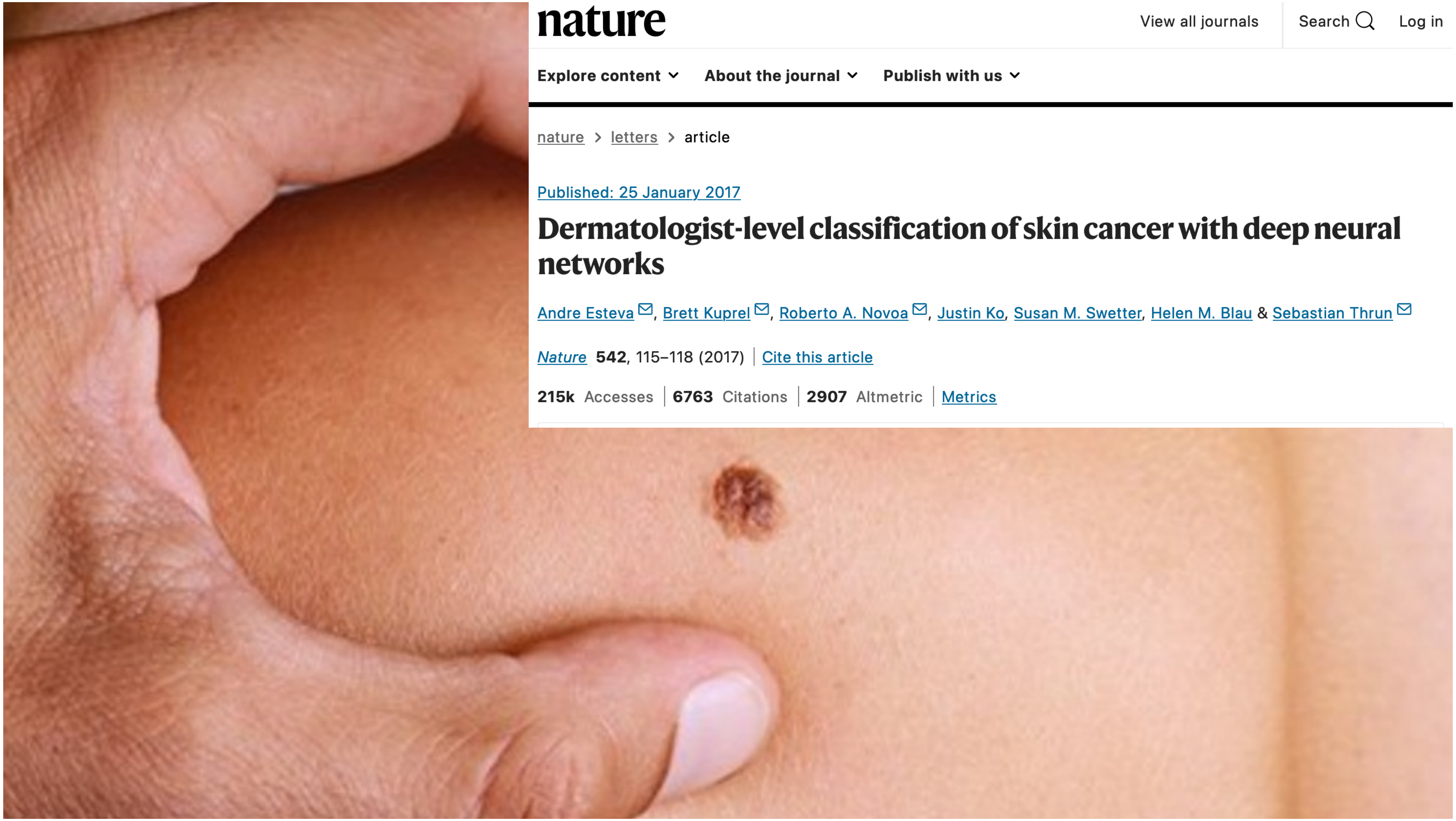


IDx-DR

- A doctor uploads the digital images of the patient’s retinas to a cloud server on which IDx-DR software is installed.
- If the images are of sufficient quality, the software provides the doctor with one of two results:

(1) “more than mild diabetic retinopathy detected: refer to an eye care professional” or

(2) “negative for more than mild diabetic retinopathy; rescreen in 12 months.”



Published: 25 January 2017

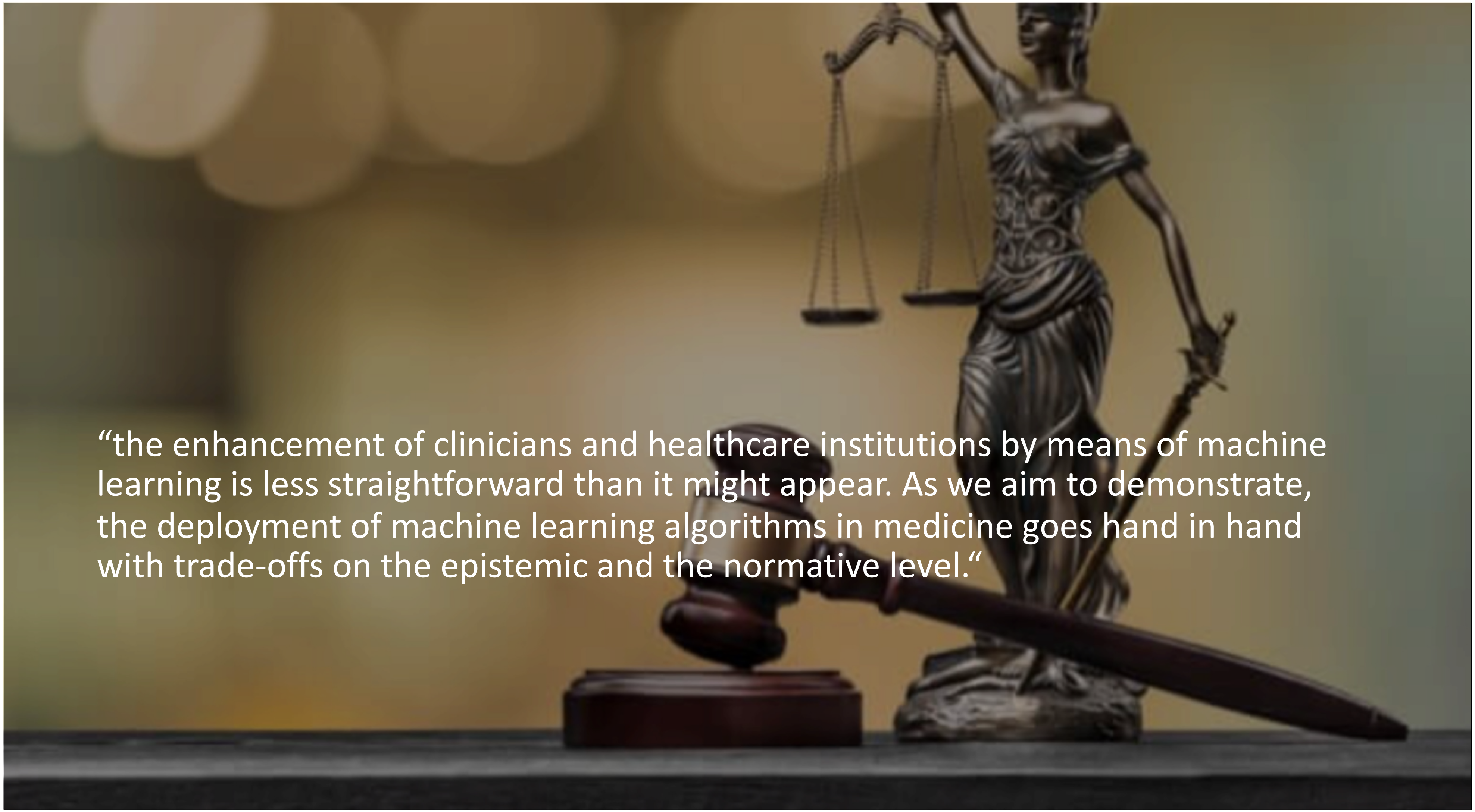
Dermatologist-level classification of skin cancer with deep neural networks

Andre Esteva, Brett Kuprel, Roberto A. Novoa, Justin Ko, Susan M. Swetter, Helen M. Blau & Sebastian Thrun

Nature **542**, 115–118 (2017) | Cite this article

215k Accesses | **6763** Citations | **2907** Altmetric | Metrics

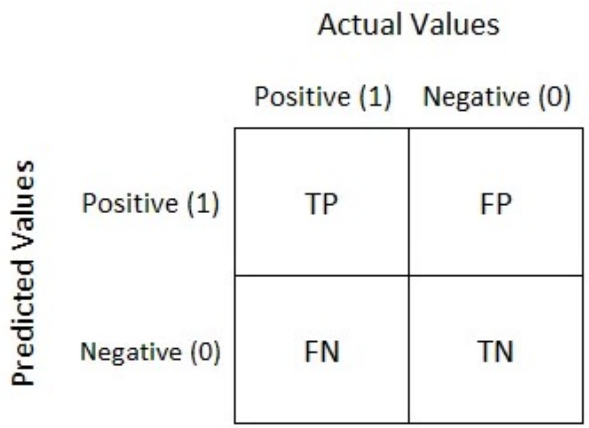
20
1
18
4
13
6
15
2
17
3
19
7
16
11
14
9
12
5



“the enhancement of clinicians and healthcare institutions by means of machine learning is less straightforward than it might appear. As we aim to demonstrate, the deployment of machine learning algorithms in medicine goes hand in hand with trade-offs on the epistemic and the normative level.“

- Fairness
- Data
- Explainability
- Responsibility

Different performance measures



		Actual Values	
		Positive (1)	Negative (0)
Predicted Values	Positive (1)	TP	FP
	Negative (0)	FN	TN

Performance measures:

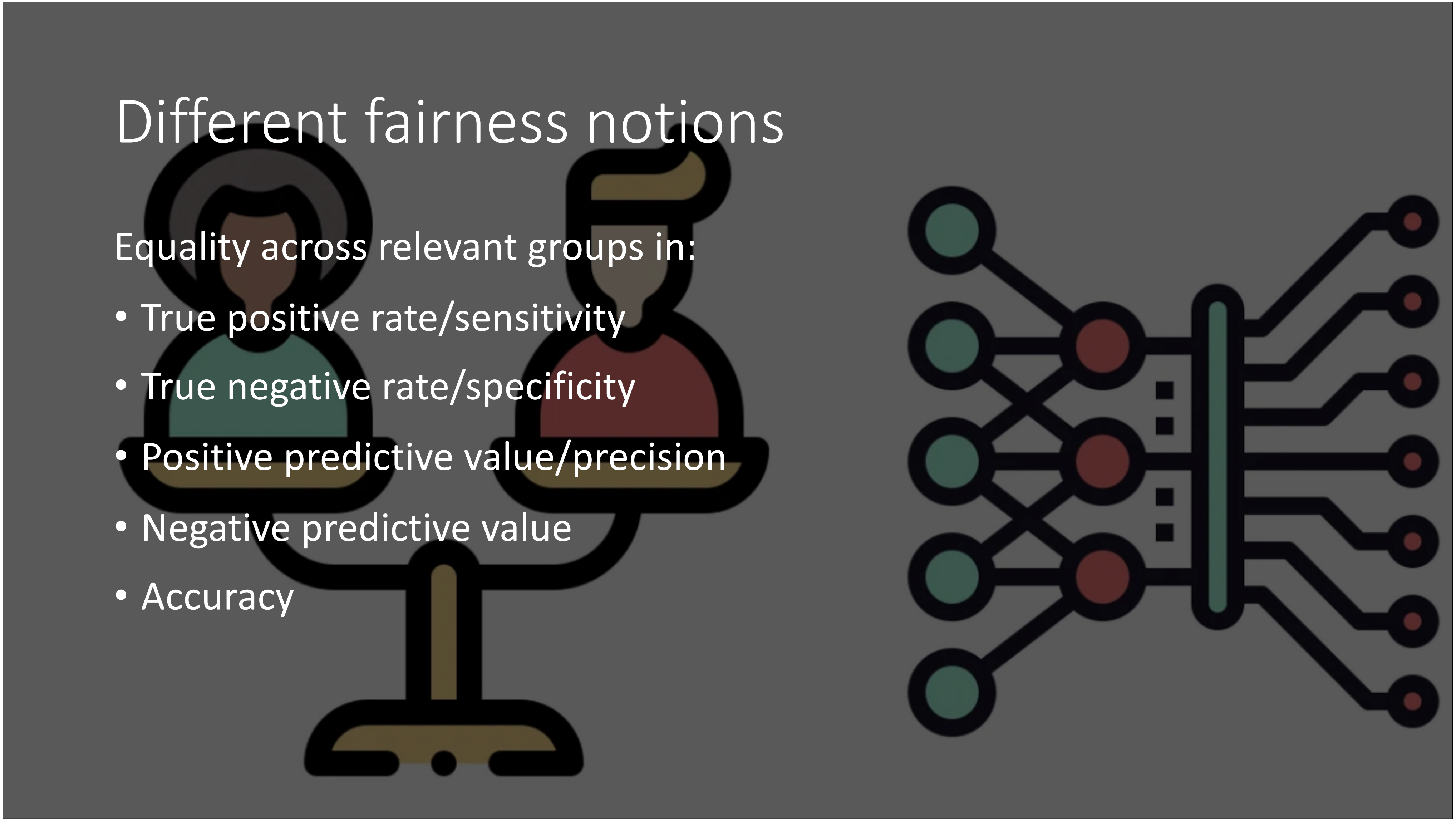
True positive rate/sensitivity

True negative rate/specificity

Positive predictive value/precision

Negative predictive value

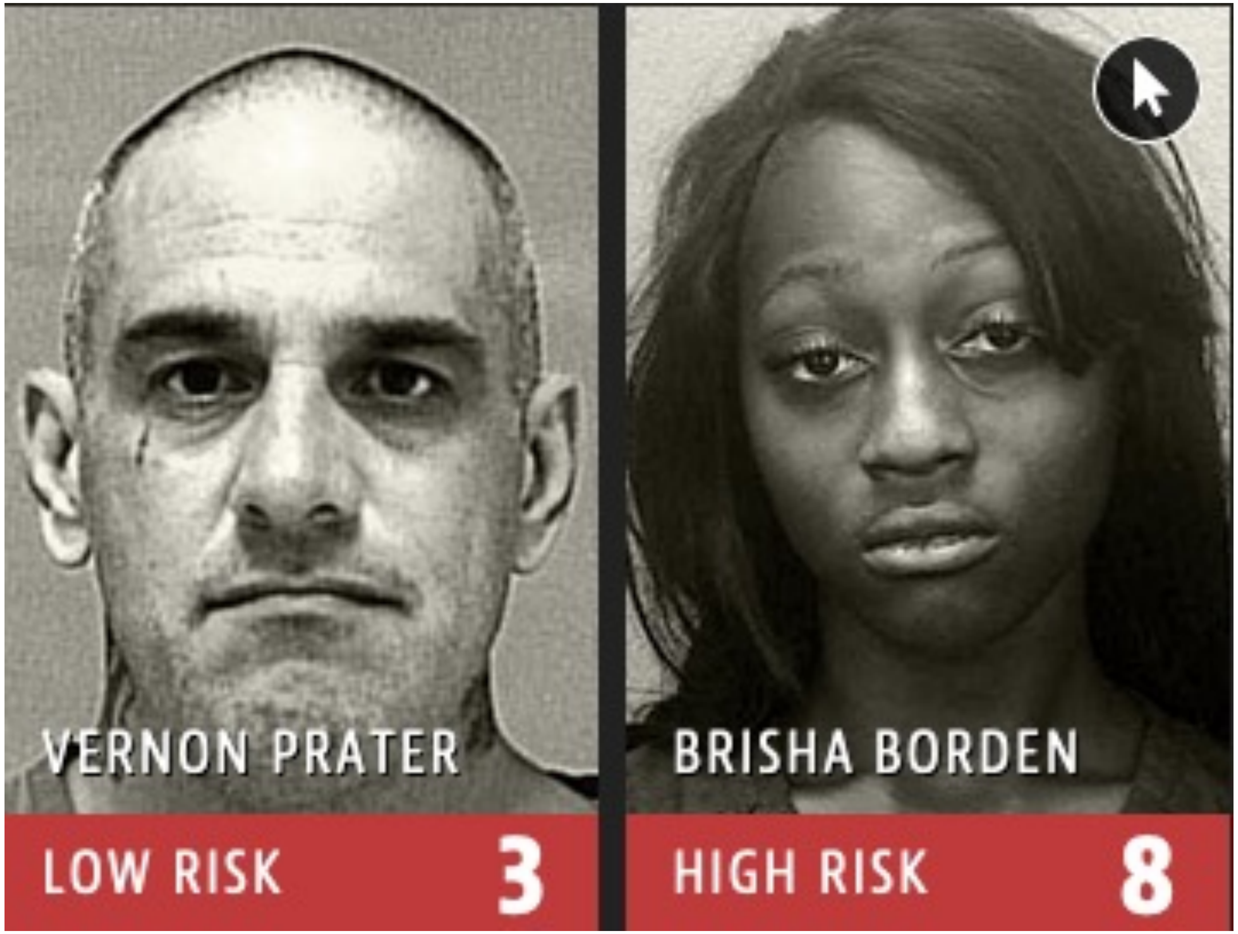
Accuracy



Different fairness notions

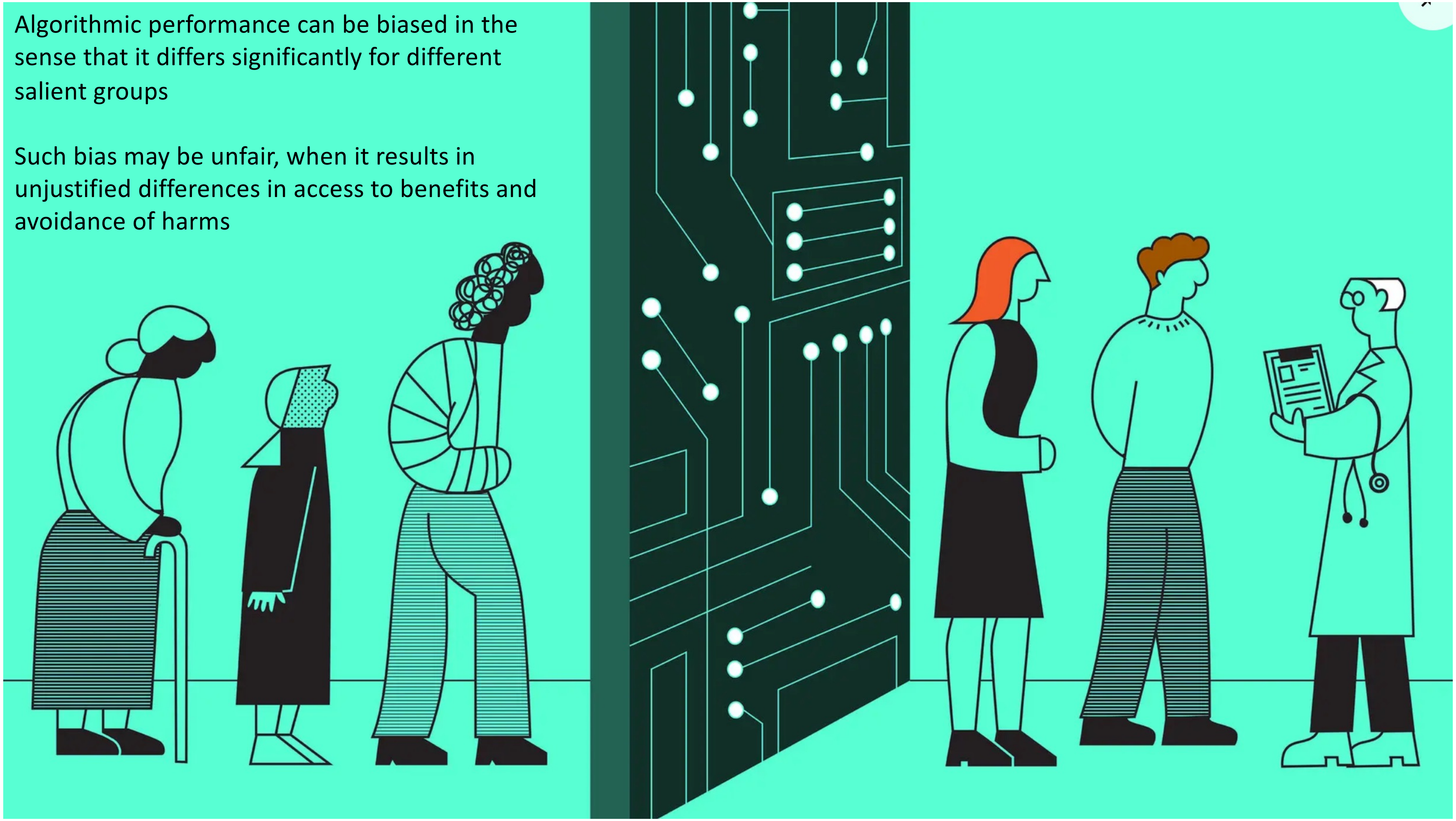
Equality across relevant groups in:

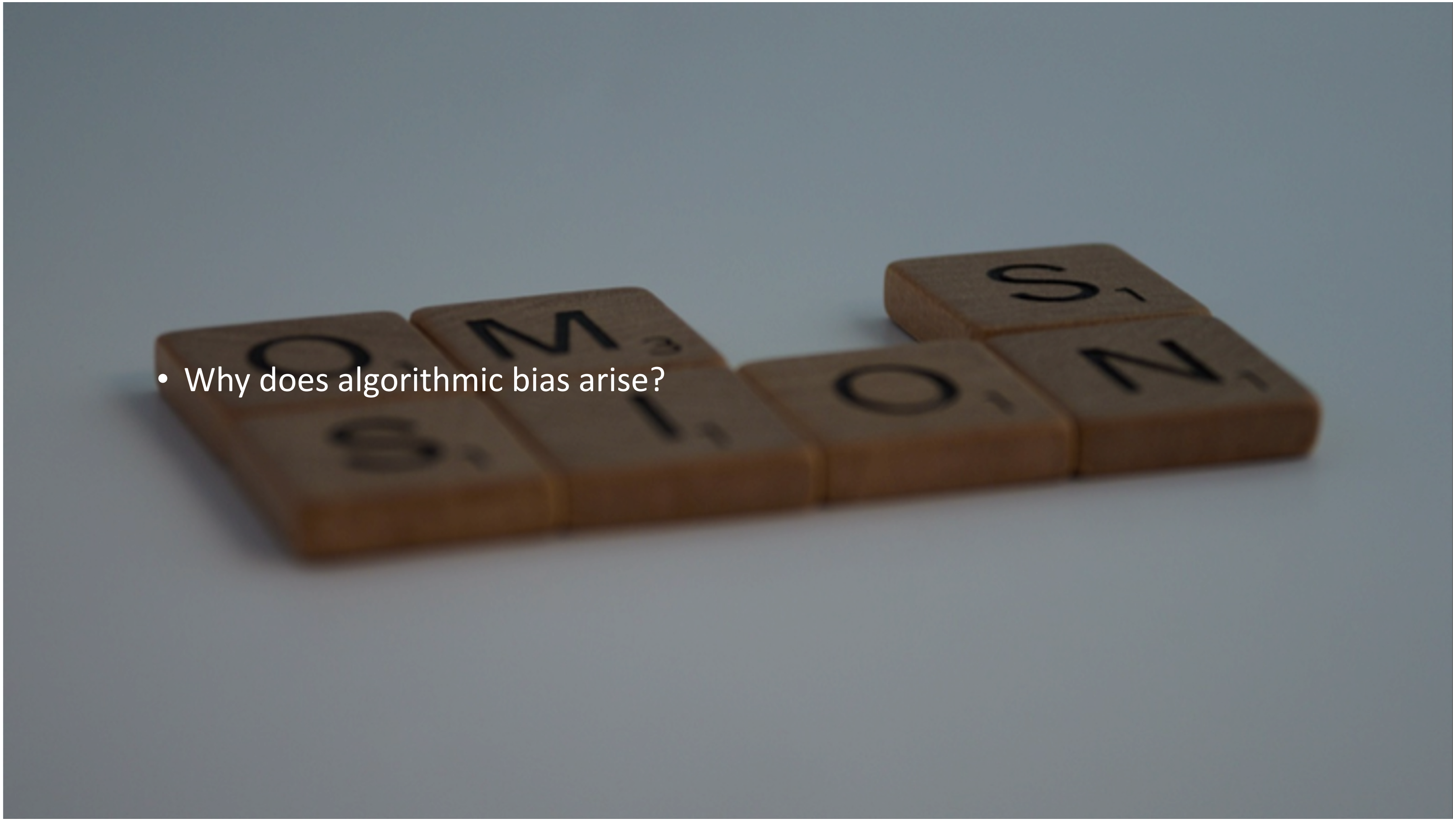
- True positive rate/sensitivity
- True negative rate/specificity
- Positive predictive value/precision
- Negative predictive value
- Accuracy


VERNON PRATER
LOW RISK
3
BRISHA BORDEN
HIGH RISK
8

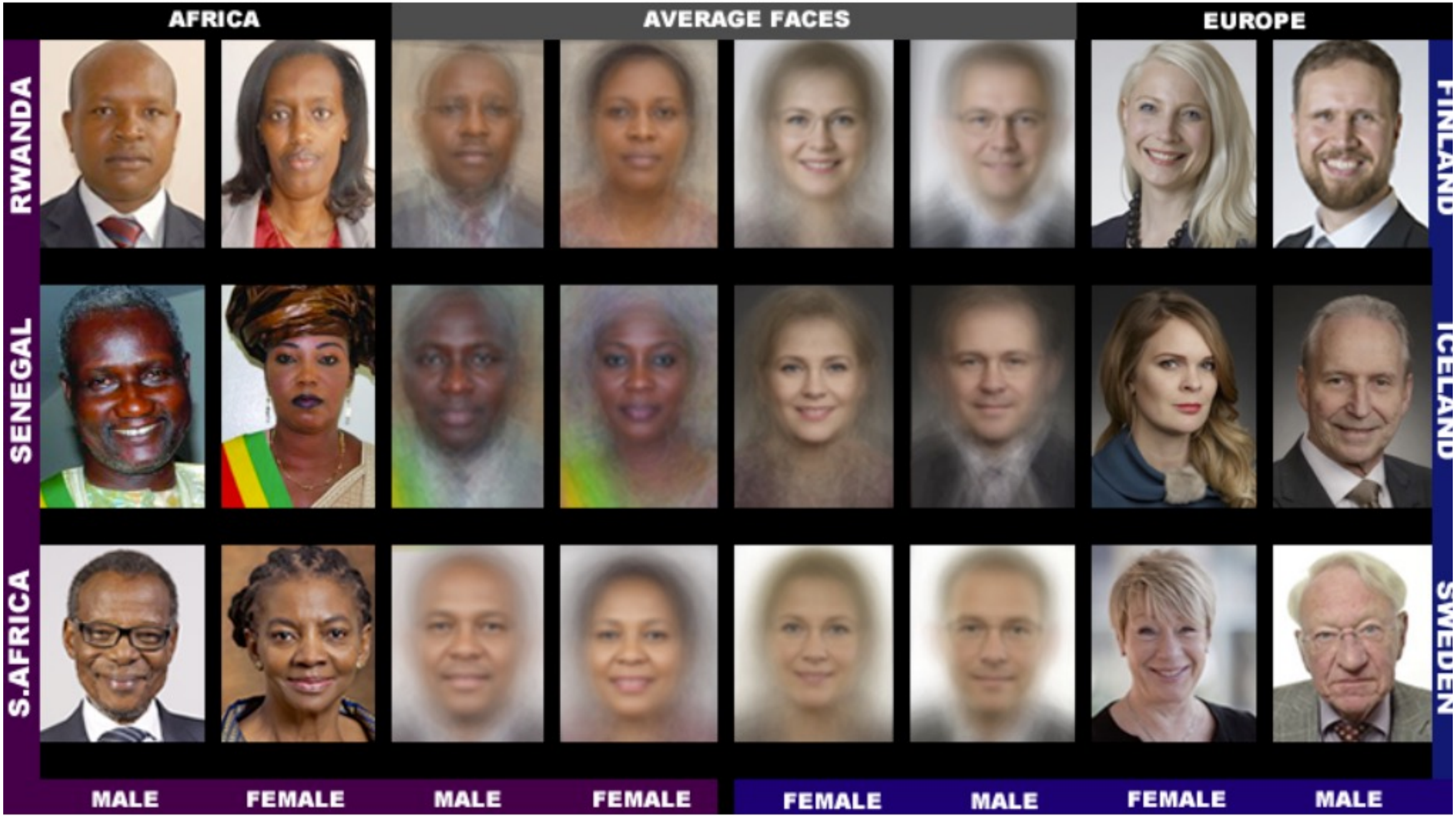
Algorithmic performance can be biased in the sense that it differs significantly for different salient groups

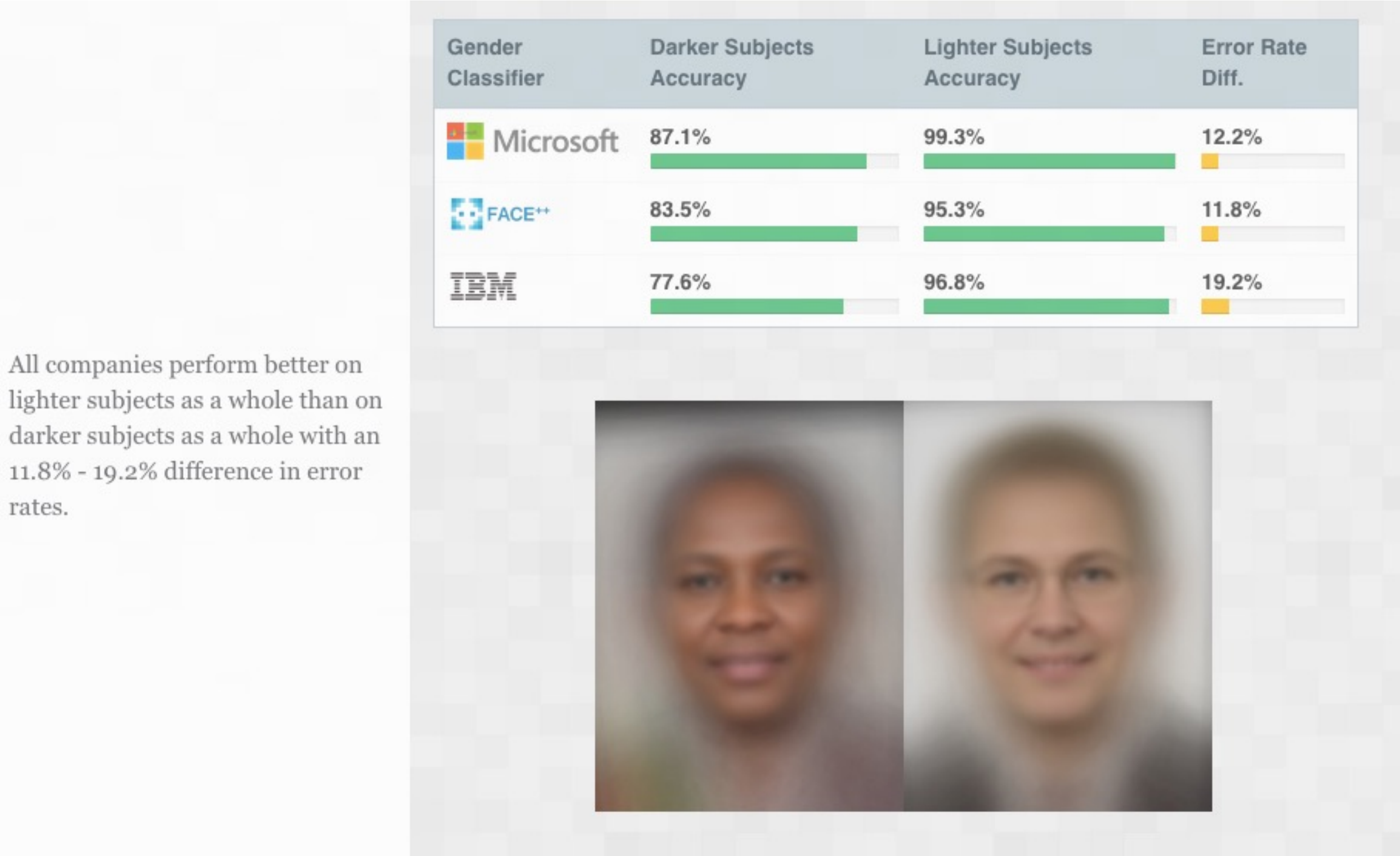
Such bias may be unfair, when it results in unjustified differences in access to benefits and avoidance of harms





- Why does algorithmic bias arise?

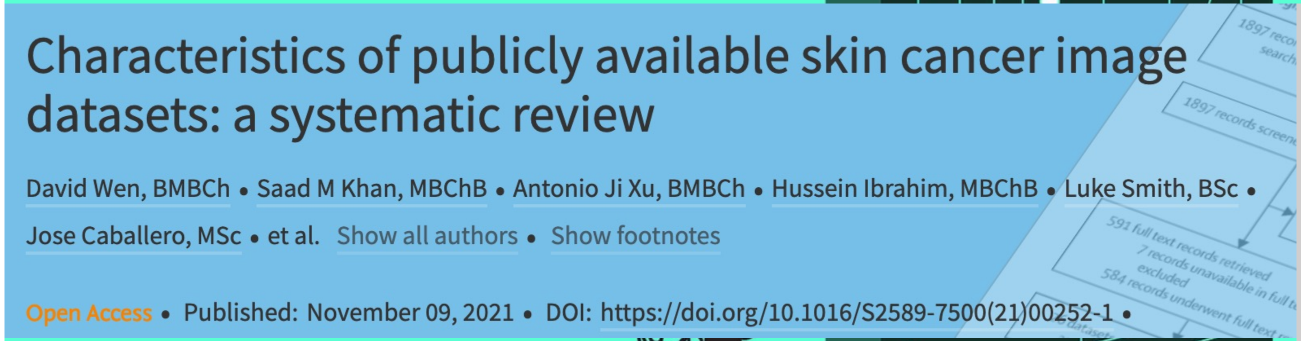

AFRICA
AVERAGE FACES
EUROPE
RWANDA
SENEGAL
S.AFRICA
FINLAND
ICELAND
SWEDEN
MALE
FEMALE
MALE
FEMALE
FEMALE
MALE
FEMALE
MALE



Gender Classifier	Darker Subjects Accuracy	Lighter Subjects Accuracy	Error Rate Diff.
Microsoft	87.1%	99.3%	12.2%
FACE++	83.5%	95.3%	11.8%
IBM	77.6%	96.8%	19.2%

All companies perform better on lighter subjects as a whole than on darker subjects as a whole with an 11.8% - 19.2% difference in error rates.

http://gendershades.org/

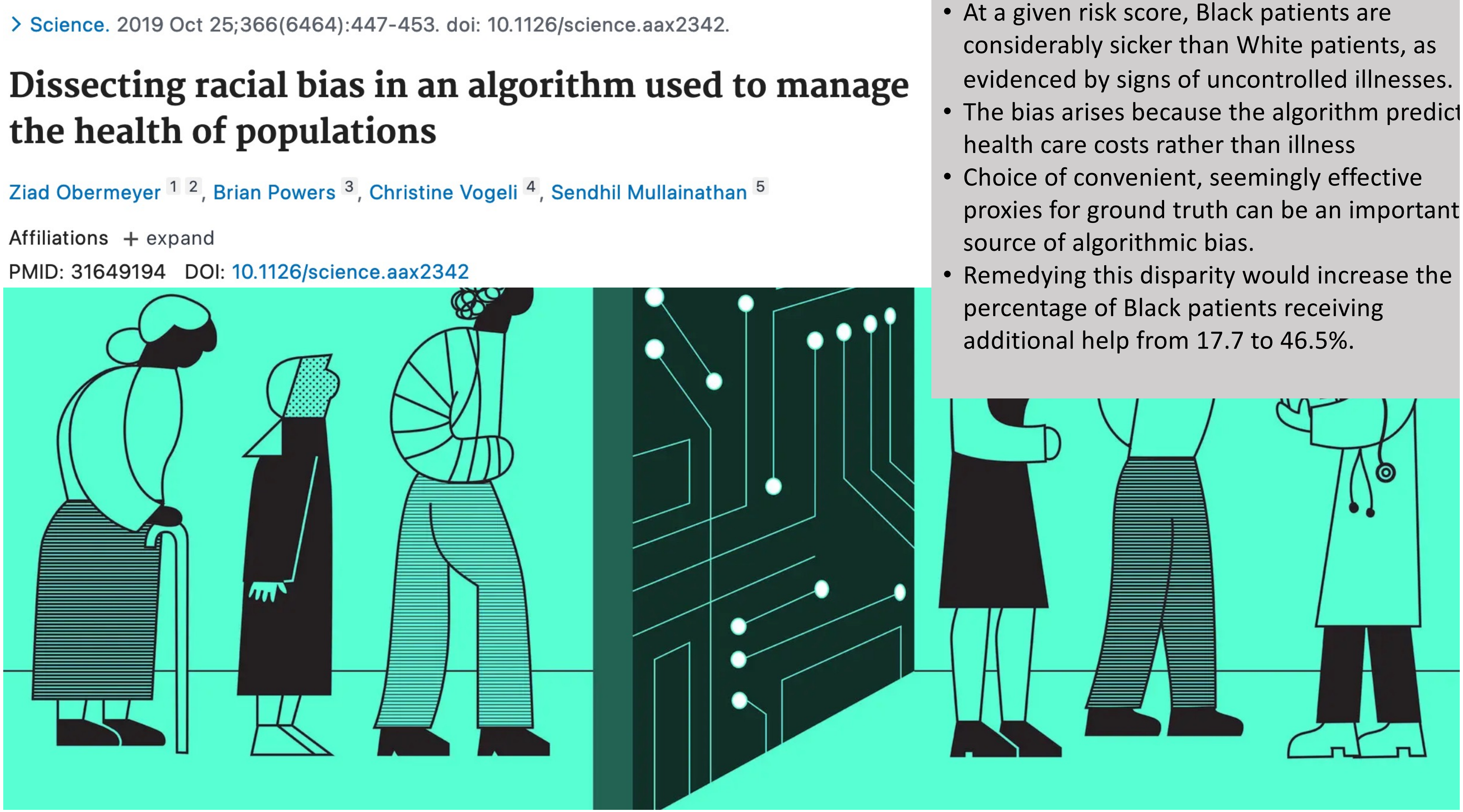


Characteristics of publicly available skin cancer image datasets: a systematic review

David Wen, BMBCh • Saad M Khan, MBChB • Antonio Ji Xu, BMBCh • Hussein Ibrahim, MBChB • Luke Smith, BSc • Jose Caballero, MSc • et al. Show all authors • Show footnotes

Open Access • Published: November 09, 2021 • DOI: https://doi.org/10.1016/S2589-7500(21)00252-1 •

- 2,436 out of 106,950 images within 21 databases had skin type recorded.
- Of these, only 10 images were from people recorded as having brown skin and one was from an individual recorded as having dark brown or black skin.
- No images were from individuals with an African, African-Caribbean or South Asian background.
- Coupled with the geographical origins of datasets, there was massive under-representation of skin lesion images from darker-skinned populations.”



> Science. 2019 Oct 25;366(6464):447-453. doi: 10.1126/science.aax2342.

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

Ziad Obermeyer [1] [2], Brian Powers [3], Christine Vogeli [4], Sendhil Mullainathan [5]

Affiliations + expand

PMID: 31649194 DOI: 10.1126/science.aax2342

- At a given risk score, Black patients are considerably sicker than White patients, as evidenced by signs of uncontrolled illnesses.
- The bias arises because the algorithm predict health care costs rather than illness
- Choice of convenient, seemingly effective proxies for ground truth can be an important source of algorithmic bias.
- Remedying this disparity would increase the percentage of Black patients receiving additional help from 17.7 to 46.5%.

Black-Box Concerns

It is impossible for humans to comprehend the mechanism by which some types of algorithms produce their output from an input.

The algorithmic decision-maker can’t explain decisions about individuals.

This threatens the possibility of holding algorithmic decision-makers accountable.

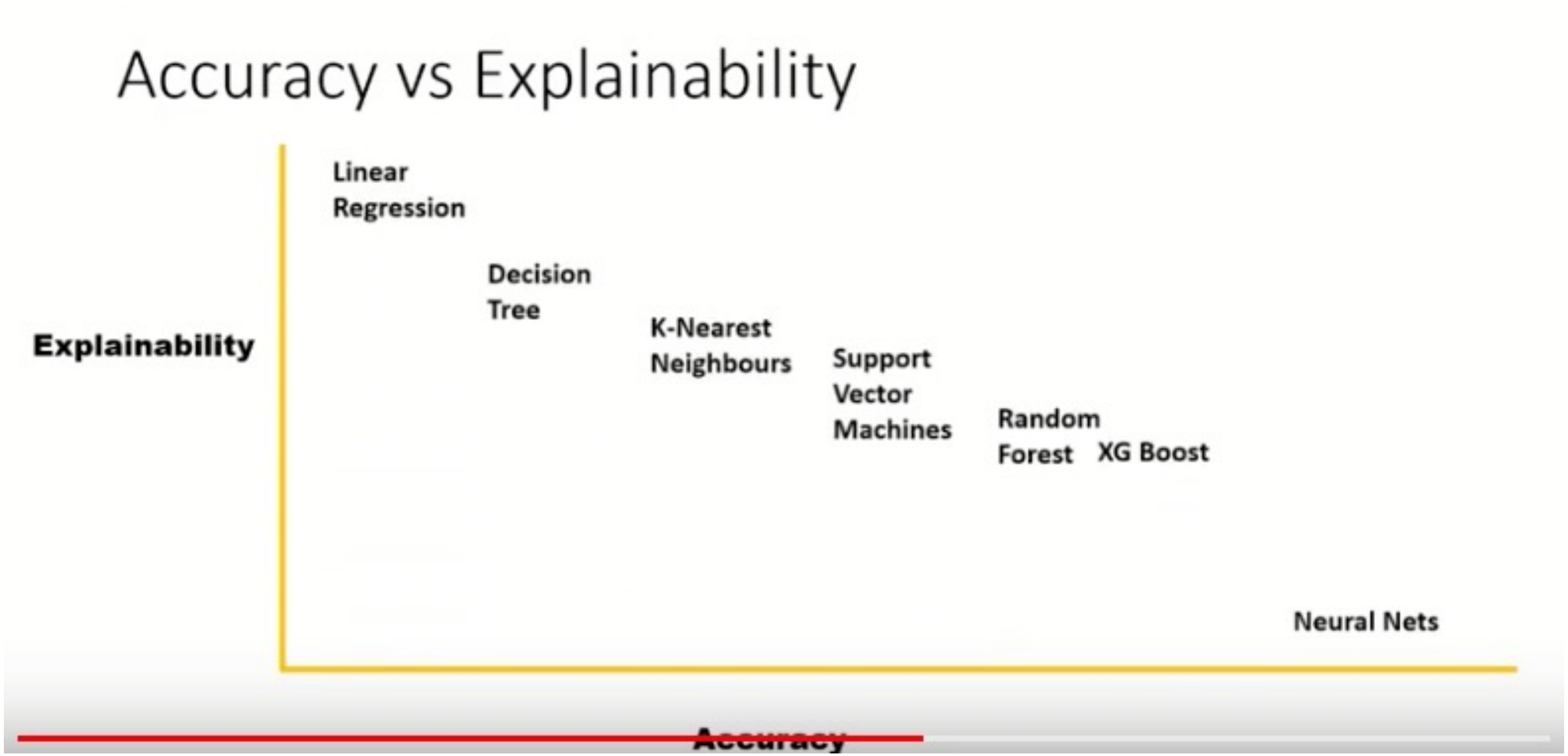
Black-Box Concerns

Lipton (2017, 1) on the call for making interpretable models:

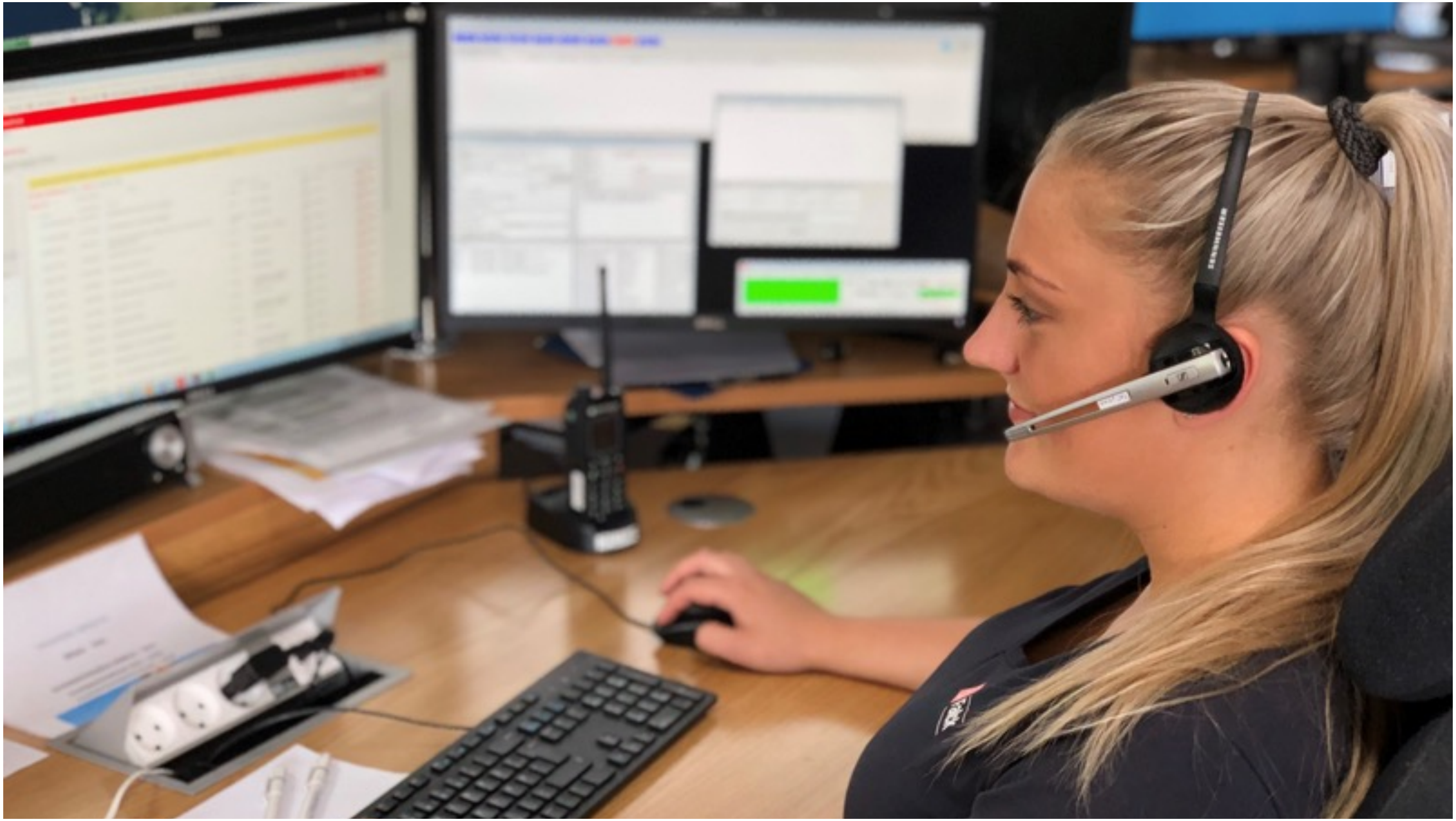
“We can train a model, and it can even give us the right answer. But we can’t just tell the doctor “my neural network says this patient has cancer!”

The doctor just won’t accept that!

They want to know why the neural network says what it says. They want an explanation. They need interpretable models.”



Accuracy vs Explainability



Original Investigation | Emergency Medicine

Effect of Machine Learning on Dispatcher Recognition of Out-of-Hospital Cardiac Arrest During Calls to Emergency Medical Services

A Randomized Clinical Trial

Stig Nikolaj Blomberg, MsC; Helle Collatz Christensen, MD, PhD; Freddy Lippert, MD; Annette Kjær Ersbøll, MsC, PhD; Christian Torp-Petersen, MD, PhD; Michael R. Sayre, MD; Peter J. Kudenchuk, MD; Fredrik Folke, MD, PhD

BWK1428

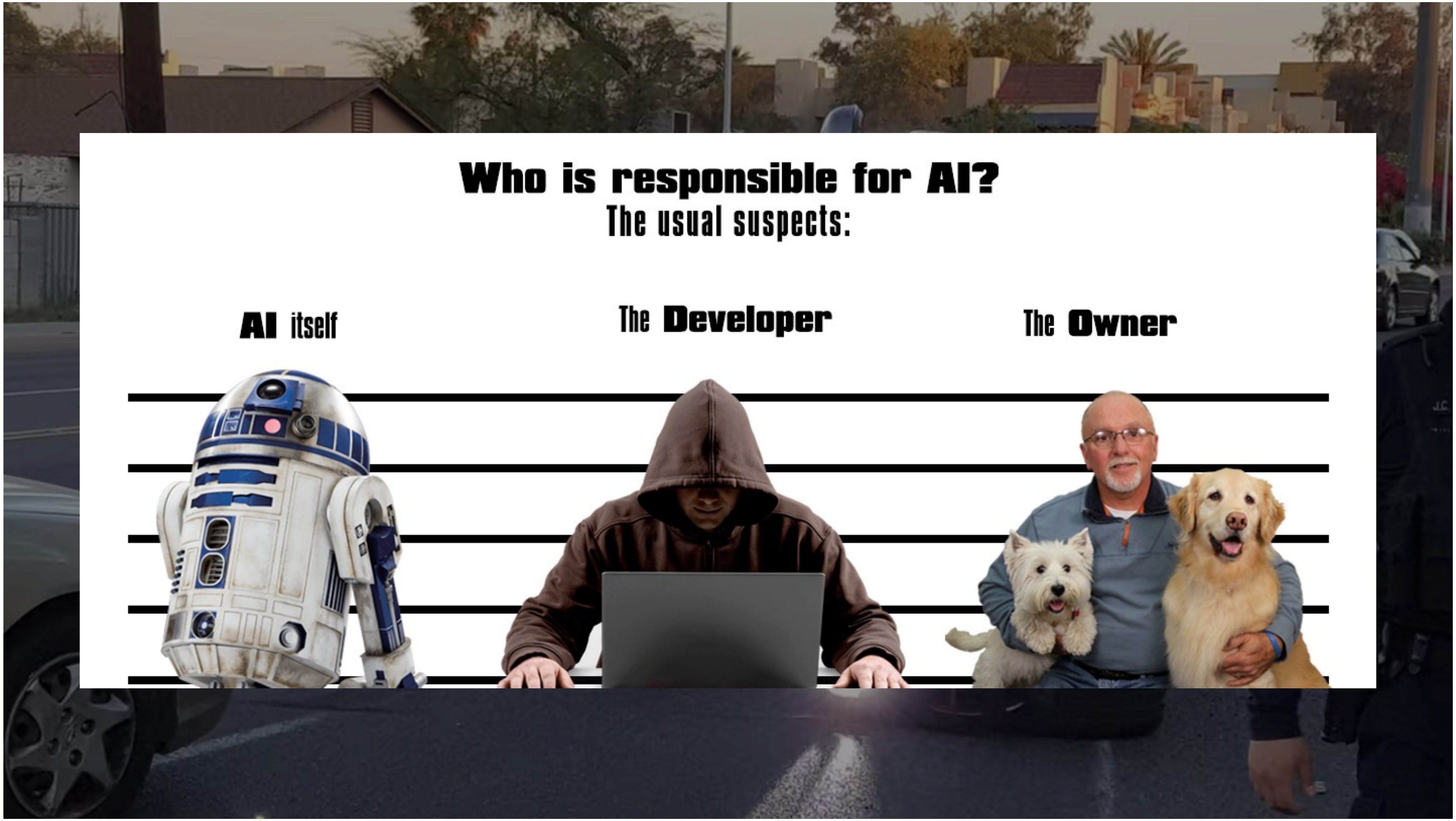
Who is responsible for AI?

The usual suspects:

AI itself

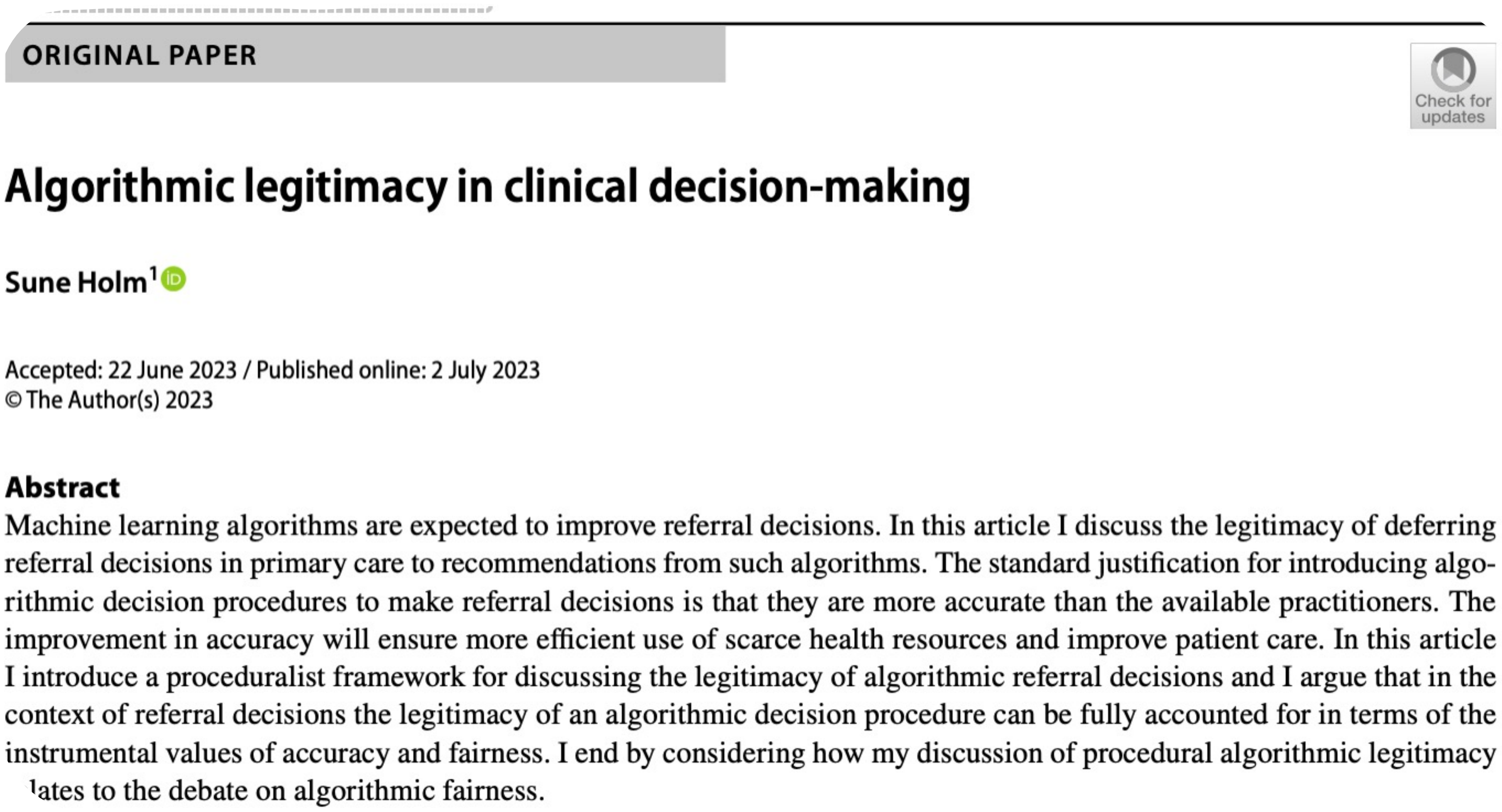
The **Developer**

The **Owner**



Some Key Questions

- What is the model supposed to predict and why?
- What type of model has been used?
- What training and test data were used?
- What performance measures have been used to assess the model?
- How is fairness understood?
- Has the system been clinically validated?



ORIGINAL PAPER

Algorithmic legitimacy in clinical decision-making

Sune Holm[1]

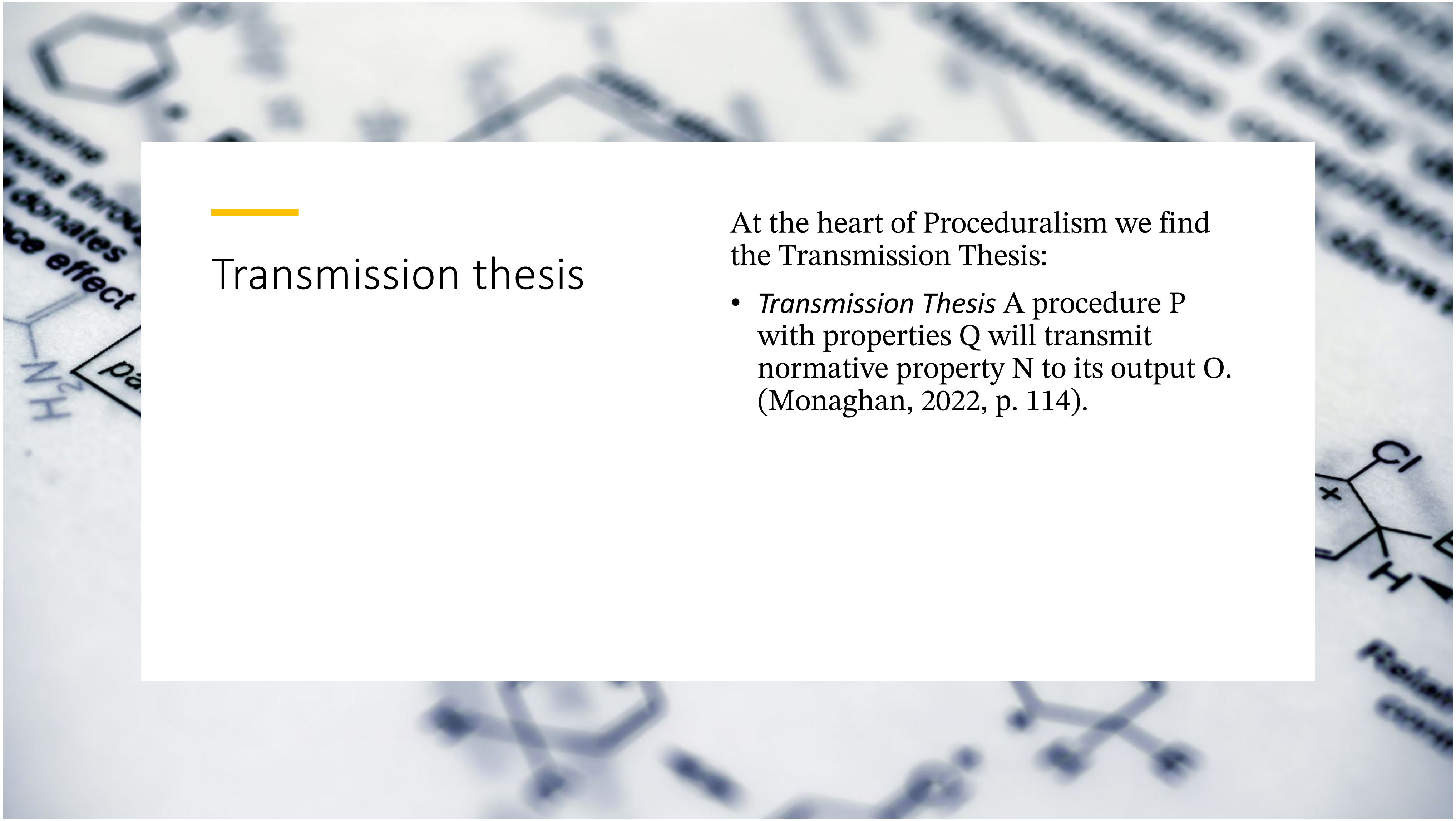
Accepted: 22 June 2023 / Published online: 2 July 2023

Abstract

Machine learning algorithms are expected to improve referral decisions. In this article I discuss the legitimacy of deferring referral decisions in primary care to recommendations from such algorithms. The standard justification for introducing algorithmic decision procedures to make referral decisions is that they are more accurate than the available practitioners. The improvement in accuracy will ensure more efficient use of scarce health resources and improve patient care. In this article I introduce a proceduralist framework for discussing the legitimacy of algorithmic referral decisions and I argue that in the context of referral decisions the legitimacy of an algorithmic decision procedure can be fully accounted for in terms of the instrumental values of accuracy and fairness. I end by considering how my discussion of procedural algorithmic legitimacy lates to the debate on algorithmic fairness.

Proceduralism and legitimacy

- According to Proceduralism a decision is legitimate if it is produced by an appropriate procedure (Monaghan, 2022, p. 110).
- This allows Proceduralism to recognize that incorrect decisions can be legitimate.
- For example, proceduralists may argue that even an incorrect guilty verdict is legitimate and therefore should be accepted because of features of the criminal procedure that produced it.



Transmission thesis

At the heart of Proceduralism we find the Transmission Thesis:

- *Transmission Thesis* A procedure P with properties Q will transmit normative property N to its output O. (Monaghan, 2022, p. 114).

The question I ask is this:

Under what conditions, if any, are decisions based on the output of an algorithm legitimate?

To decide on this question, I distinguish between instrumental and non-instrumental Q properties.

Instrumental: Accuracy & fairness

Non-instrumental: Explainability