Is Computer-Aided Diag/Prognosis fair towards minorities? Current status and perspectives

Dr. André Anjos (anjos.ai)

February 10, 2022

AI and Health

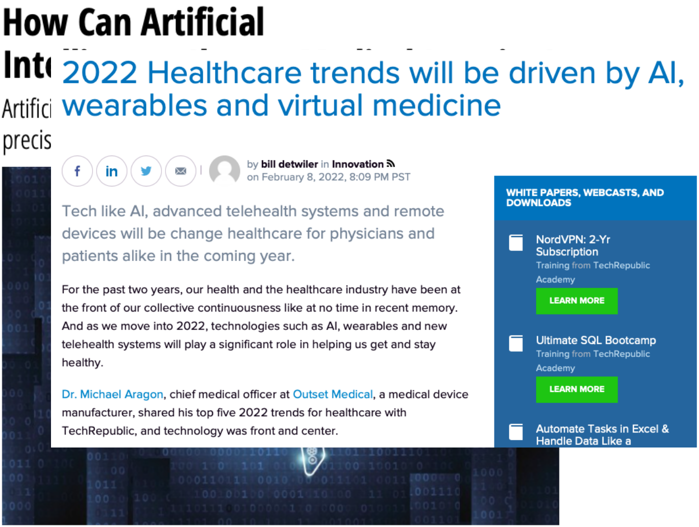
Press Releases

How Can Artificial Intelligence Change Medical Imaging?

Artificial intelligence can improve medical imaging for screenings, precision medicine, and risk assessment.

AI and Health

Press Releases



How Can Artificial
Int
Artifici
precis

2022 Healthcare trends will be driven by AI, wearables and virtual medicine

by **bill detwiler** in **Innovation**
on February 8, 2022, 8:09 PM PST

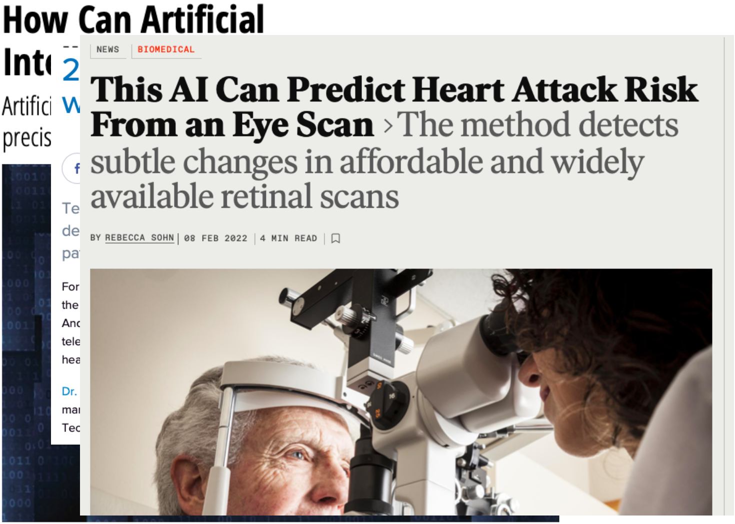
Tech like AI, advanced telehealth systems and remote devices will be change healthcare for physicians and patients alike in the coming year.

For the past two years, our health and the healthcare industry have been at the front of our collective continuousness like at no time in recent memory. And as we move into 2022, technologies such as AI, wearables and new telehealth systems will play a significant role in helping us get and stay healthy.

Dr. Michael Aragon, chief medical officer at Outset Medical, a medical device manufacturer, shared his top five 2022 trends for healthcare with TechRepublic, and technology was front and center.

AI and Health

Press Releases



How Can Artificial

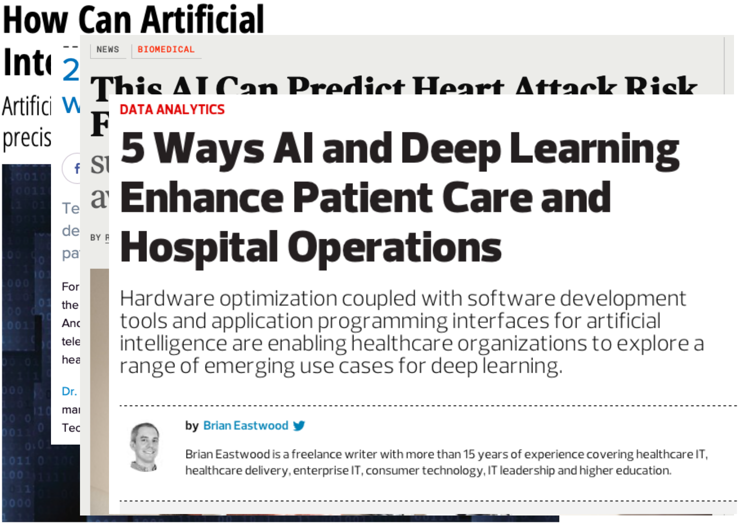
NEWS | BIOMEDICAL

This AI Can Predict Heart Attack Risk From an Eye Scan › The method detects subtle changes in affordable and widely available retinal scans

BY REBECCA SOHN | 08 FEB 2022 | 4 MIN READ

AI and Health

Press Releases



How Can Artificial

NEWS | BIOMEDICAL

This AI Can Predict Heart Attack Risk

DATA ANALYTICS

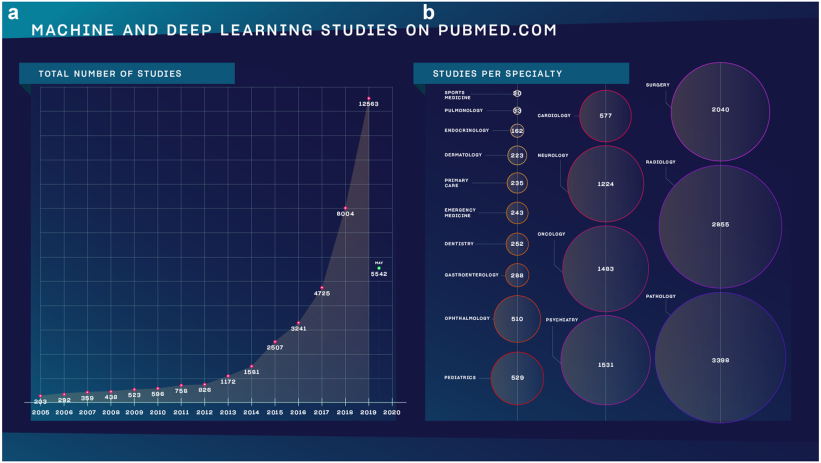
5 Ways AI and Deep Learning Enhance Patient Care and Hospital Operations

Hardware optimization coupled with software development tools and application programming interfaces for artificial intelligence are enabling healthcare organizations to explore a range of emerging use cases for deep learning.

by Brian Eastwood

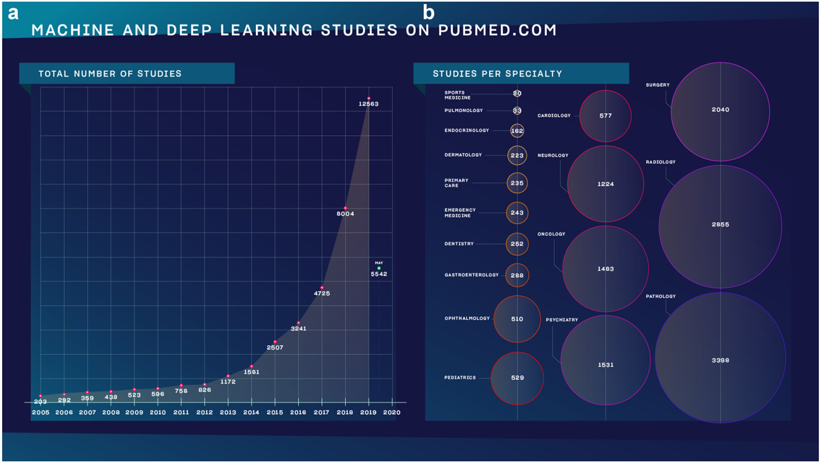
Brian Eastwood is a freelance writer with more than 15 years of experience covering healthcare IT, healthcare delivery, enterprise IT, consumer technology, IT leadership and higher education.

Scientific Papers[1]



[1] A short guide for medical professionals in the era of artificial intelligence, Mesko and Gorog, 2020, **https://doi.org/10.1038/s41746-020-00333-z**

Scientific Papers[1]



Breakthroughs and performance are on the menu!

[1] A short guide for medical professionals in the era of artificial intelligence, Mesko and Gorog, 2020, **https://doi.org/10.1038/s41746-020-00333-z**

Can it do the job?

Unfortunately, large-scale deployments of AI in healthcare require more than *just* performance. Among critical elements, one can cite:

- Fairness
- Accountability
- Transparence (Explainability, Interpretability)
- Safety

Can it do the job?

Unfortunately, large-scale deployments of AI in healthcare require more than *just* performance. Among critical elements, one can cite:

- **Fairness**
- Accountability
- Transparence (Explainability, Interpretability)
- Safety

Multi-goal Approach

In systems taking decisions affecting human beings, one expects to check all those boxes, while still showing (super) human performance!

Fairness

No person may be discriminated against in particular on grounds of origin, race, gender, age, language, social position, way of life, religious, ideological, or political convictions, or because of a physical, mental or psychological disability.[2]

[2] Swiss constitution, Art. 8, al 2.

Fairness

No person may be discriminated against in particular on grounds of origin, race, gender, age, language, social position, way of life, religious, ideological, or political convictions, or because of a physical, mental or psychological disability.[2]

Fair AI System

- Independent of sensitive attributes
- Does not privilege any one demography
- Well calibrated w.r.t. different demographies

[2] Swiss constitution, Art. 8, al 2.

Disclaimer

In this "survey" talk I will:

- Assume one would like to build fair AI systems that maintain maximum utility (performance)
- Address **technical issues** of fairness related to AI
- Motivate it from a healthcare perspective

Examples of Unfair Treatment

There is a growing number of examples of unfair treatment from automatic algorithms, at scientific literature:

- Automated Experiments on Ad Privacy Settings, Datta and others, 2015 **https://doi.org/10.1515/popets-2015-0007**): **Job search advertisement for highly paid positions are less likely to be presented to women**;
- Discrimination in Online Ad Delivery: Google ads, black names and white names, racial discrimination, and click advertising, Sweeney, 2013 **https://doi.org/10.1145/2460276.2460278**: **Searches for distinctively Black-sounding names are more likely to trigger ads for arrest records**;
- Face Recognition Performance: Role of Demographic Information, Klare and others, 2012, **https://doi.org/10.1109/TIFS.2012.2214212**: **Facial recognition systems increasingly used in law enforcement perform worse on recognizing faces of women and Black individuals**;
- Semantics derived automatically from language corpora contain human-like biases, Caliskan and others, 2017 **https://doi.org/10.1126/science.aal4230**: **Natural language processing algorithms encode language in gendered ways**.

Examples in Healthcare

Case 1: Risk Prediction[3]

- Commercial tool to identify patients with complex **health needs**

[3] Dissecting racial bias in an algorithm used to manage the health of populations, Obermeyer and others, 2019, **https://doi.org/10.1126/science.aax2342**

Examples in Healthcare

Case 1: Risk Prediction[3]

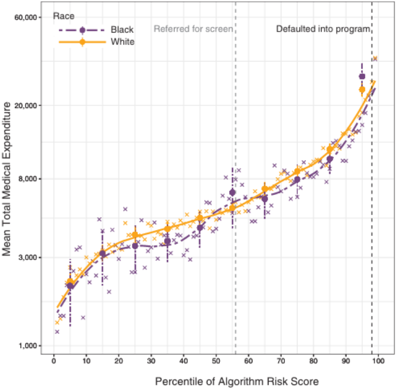
- Commercial tool to identify patients with complex **health needs**
- Deployed nation-wide in the US

[3] Dissecting racial bias in an algorithm used to manage the health of populations, Obermeyer and others, 2019, **https://doi.org/10.1126/science.aax2342**

Examples in Healthcare

Case 1: Risk Prediction[3]

- Commercial tool to identify patients with complex **health needs**
- Deployed nation-wide in the US
- Algorithms input patient data (excludes self-reported race), maps to incurred costs



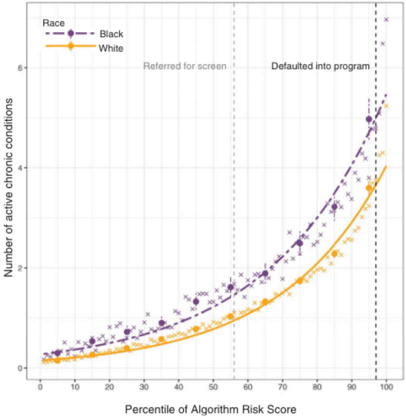
Perfectly calibrated!

[3]Dissecting racial bias in an algorithm used to manage the health of populations, Obermeyer and others, 2019, **https://doi.org/10.1126/science.aax2342**

Examples in Healthcare

Case 1: Risk Prediction[3]

- Commercial tool to identify patients with complex **health needs**
- Deployed nation-wide in the US
- Algorithms input patient data (excludes self-reported race), maps to incurred costs
- Issue: ***Incurred* costs as a proxy for health *needs***



Not for chronic conditions

[3]Dissecting racial bias in an algorithm used to manage the health of populations, Obermeyer and others, 2019, **https://doi.org/10.1126/science.aax2342**

Examples in Healthcare

Case 2: CAD for melanoma detection[4]

- Tool to screen for melanoma on skin lesions

[4] Machine Learning and Health Care Disparities in Dermatology, Adamson and Smith, 2018, **https://doi.org/10.1001/jamadermatol.2018.2348**

Examples in Healthcare

Case 2: CAD for melanoma detection[4]

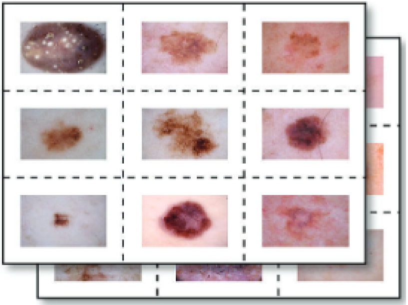
- Tool to screen for melanoma on skin lesions
- Large public dataset: International Skin Imaging Collaboration: Melanoma Project

[4] Machine Learning and Health Care Disparities in Dermatology, Adamson and Smith, 2018, **https://doi.org/10.1001/jamadermatol.2018.2348**

Examples in Healthcare

Case 2: CAD for melanoma detection[4]

- Tool to screen for melanoma on skin lesions
- Large public dataset: International Skin Imaging Collaboration: Melanoma Project
- Algorithms input images, output probability of melanoma

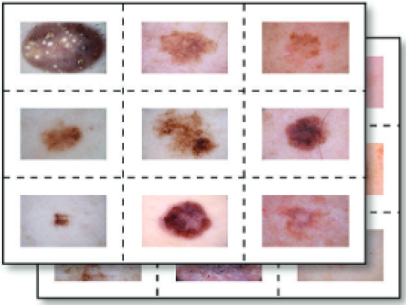


[4]Machine Learning and Health Care Disparities in Dermatology, Adamson and Smith, 2018, **https://doi.org/10.1001/jamadermatol.2018.2348**

Examples in Healthcare

Case 2: CAD for melanoma detection[4]

- Tool to screen for melanoma on skin lesions
- Large public dataset: International Skin Imaging Collaboration: Melanoma Project
- Algorithms input images, output probability of melanoma
- Issue: Most data issued from individuals with pale skin colour: **unknown bias!**



[4]Machine Learning and Health Care Disparities in Dermatology, Adamson and Smith, 2018, **https://doi.org/10.1001/jamadermatol.2018.2348**

Examples in Healthcare

Case 3: CAD for Chest Radiography[5]

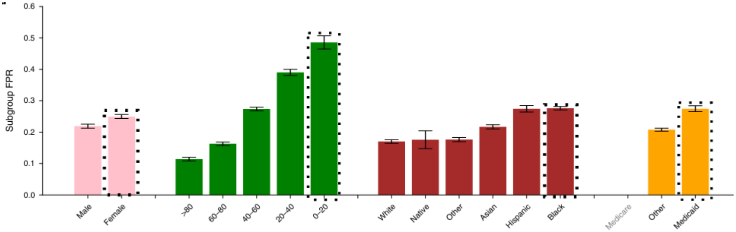
- Tool to screen for radiological findings in CXR
- 4 Large public datasets, with meta information concerning patient gender, age, race, and insurance
- Algorithm inputs images, outputs probability of various radiological findings, or "no findings"

[5] Underdiagnosis bias of artificial intelligence algorithms applied to chest radiographs in under-served patient populations, Seyyed-Kalantari and others, 2021, **https://doi.org/10.1038/s41591-021-01595-0**

Examples in Healthcare

Case 3: CAD for Chest Radiography[5]

- Tool to screen for radiological findings in CXR
- 4 Large public datasets, with meta information concerning patient gender, age, race, and insurance
- Algorithm inputs images, outputs probability of various radiological findings, or "no findings"

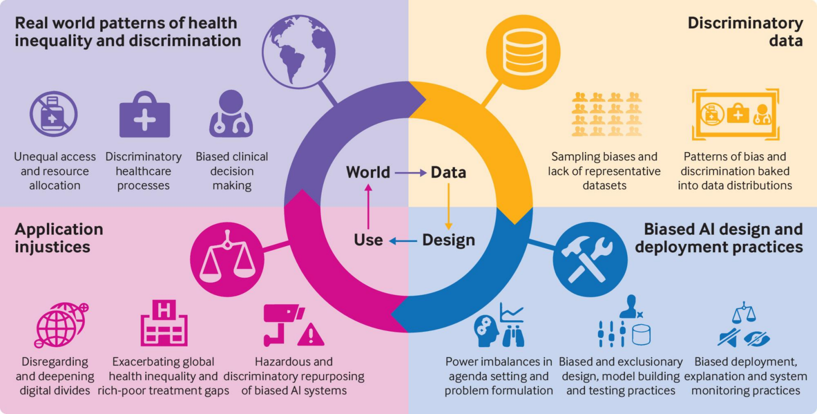


Issue: Low representativity on training data

[5]Underdiagnosis bias of artificial intelligence algorithms applied to chest radiographs in under-served patient populations, Seyyed-Kalantari and others, 2021, **https://doi.org/10.1038/s41591-021-01595-0**

Sources of bias

Bias must be looked for in all parts of the design process of algorithms[6].



[6]Does "AI" stand for augmenting inequality in the era of covid-19 healthcare?, Leslie and others, 2021, **http://dx.doi.org/10.1136/bmj.n304**

Data Bias and Problem Design: Examples

- Data is issued only from those in a limited category (historical unfairness):
  - Access to healthcare
  - Linked with a social media profile, or smartphone ownership
- Data fails to include clinically relevant variables (case 1)
  - Outcomes (or their proxies) introduce bias
  - Biased predictor choice for sampled population
- Data has a sampling bias (cases 2 and 3)
  - Underrepresented segments of the population
  - Selection bias / Prejudice
- Power imbalances in agenda setting and problem formulation
- Data shifts (populations change with time)

Data Bias: Mitigation

At every iteration, ensure:

- Include and report population characteristics on your dataset
- Analyze predictor and outcome relationships to population characteristics – trim down feature set if possible
- If not done before, go through PROBAST (assessment of risk of bias) tool questionnaire[7]:
  - Were there a reasonable number of participants with the outcome?
  - Was the outcome defined and determined in a similar way for all participants? (e.g. different sensors for different populations?)
  - Were predictor assessments made without knowledge of outcome data?
  - etc.
- Re-think your data regularly: is it still unbiased?
- Avoid feedback loops: *data that shapes algorithms that shapes data!*

[7] PROBAST: A tool to assess the risk of bias and applicability of prediction model studies, Wolff and others, 2019, **https://doi.org/10.7326/M18-1376**

Algorithmic Bias: Sources

Here are some typical sources of bias:

- Naïve model development and evaluation (which does not consider population differences) - **very frequent**
- Choice of best performing models based on limited data
- Selective reporting ("removing outliers") – *reshapes* data biases
- Opaque algorithm decisions – how confident am I the algorithm is unbiased?

Algorithmic Bias: Understanding

No Fairness through Unawareness[8]

Simply suppress class memberships from input feature set

- Probably, the oldest approach to "fairness"

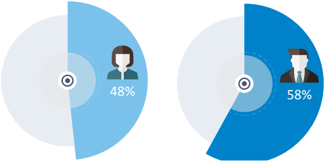
[8] Fairness and machine learning, Limitations and Opportunities, Barocas and others, 2021, **http://www.fairmlbook.org**

Algorithmic Bias: Understanding

No Fairness through Unawareness[8]

Simply suppress class memberships from input feature set

- Probably, the oldest approach to "fairness"
- Too naïve to work – class membership is often correlated with other features!
- A highly accurate classifier for class membership can be built of slightly correlated features



Example: The feature "uses-internet" is slightly correlated to being male (source: ITU gender gap report, 2020)

[8] Fairness and machine learning, Limitations and Opportunities, Barocas and others, 2021, **http://www.fairmlbook.org**

Algorithmic Bias: Understanding

Independence

Predicted scores (R) must be independent of class membership (A): $R \perp A$

- Some countries (US) require this for certains classes of products (e.g., loan risk assessment)

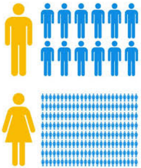
Algorithmic Bias: Understanding

Independence

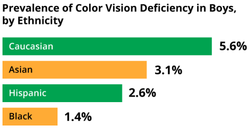
Predicted scores (R) must be independent of class membership (A): $R \perp A$

- Some countries (US) require this for certains classes of products (e.g., loan risk assessment)
- Typically, a weak assumption of fairness, as different demographies may actually have different probabilities of a "positive" outcome

Colour Blindness



8% males | 0.5% females



Algorithmic Bias: Understanding

Separation (or Balance)

Making a mistake must be independent of class membership:
$R \perp A | Y$

- Many (scientific) authors consider balancing mistakes both ways as a desirable feature in fair AI systems:
  - Equality of Opportunities: FNR must match between demographies
  - Equality of Odds: FPR must match between demographies
- Balancing systems *often* reduce their overall utility (performance)

Algorithmic Bias: Understanding

Sufficiency (or Calibration by Group)

All demographic groups should be well-calibrated w.r.t. each other:
$Y \perp A | R$

- I.e.: Scores must be comparable between classes, or the importance of one demography is reduced
- Sufficiency and utility are not contradictory

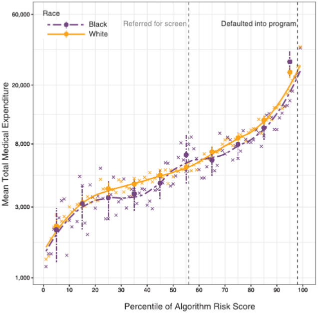
Algorithmic Bias: Understanding

Sufficiency (or Calibration by Group)

All demographic groups should be well-calibrated w.r.t. each other:
$Y \perp A | R$

- I.e.: Scores must be comparable between classes, or the importance of one demography is reduced
- Sufficiency and utility are not contradictory
- Sufficiency is often satisfied by default without the need for any explicit intervention.

Example from the US tool to measure patient risk:



Perfectly calibrated!

Relationships between criteria

Only one can survive[9]

Independence, Separation and Sufficiency are mutually exclusive

Independence	Separation	Sufficiency
$R \perp A$	$R \perp A \vert Y$	$Y \perp A \vert R$

R = predicted score; Y = label; A = sensitive attribute

- These criteria cannot be achieved simultaneously
- Independence vs. Sufficiency or Separation is relatively easy to demonstrate, and known for a while.
- Mutual exclusion between Sufficiency (calibration per group) and Separation was only recently demonstrated.
- Exception to this rule: $A \perp Y$

[9] Inherent Trade-Offs in the Fair Determination of Risk Scores, Kleinberg and others, 2017, **https://doi.org/10.4230/LIPICS.ITCS.2017.43**

Algorithmic Bias: Mitigation

Summary

1. Choose your de-biasing criterion
2. Choose where to affect your machine learning pipeline: preprocessing, training, post-processing
3. Apply de-biasing technique
4. Understand and report biases on your training, validation and test (hold-out) datasets
5. Do not filter your dataset (e.g. exclude patients) without re-evaluating biases
6. Favour more interpretable AI algorithms, or those that are explainable, or auditable
7. During deployment, maintain a representative hold-out set to monitor algorithm performance through time.
8. Avoid "rush" to deploy, step back and evaluate bias before!

Conclusions

Is Computer-Aided Diagnosis fair towards minorities?

- It can be, subject to the criterion being chosen
- Deploying a fair system is a continuous process that needs to be well-anchored:
  - Companies selling technology need to be transparent about fairness of their algorithms
  - Institutions dealing with humans need to monitor shifts in population and algorithm performance:

Further Resources

- Interactive: Google PAIR has an intuitive explanation on bias: **https://pair.withgoogle.com/explorables/dataset-worldviews/**
- Scientific Book: Fairness and machine learning, Limitations and Opportunities, Barocas and others, 2021, **http://www.fairmlbook.org**
- Bed time: Weapons of Math Destruction, Cathy O'Neil, 2016, ISBN 0553418815

andre.anjos@idiap.ch
https://anjos.ai

Thank you for your attention!

Dr. André Anjos (anjos.ai)

February 10, 2022