BBS
MULTIPLYING OPPORTUNITIES

Fairness in AI

Prof. Matteo Golfarelli

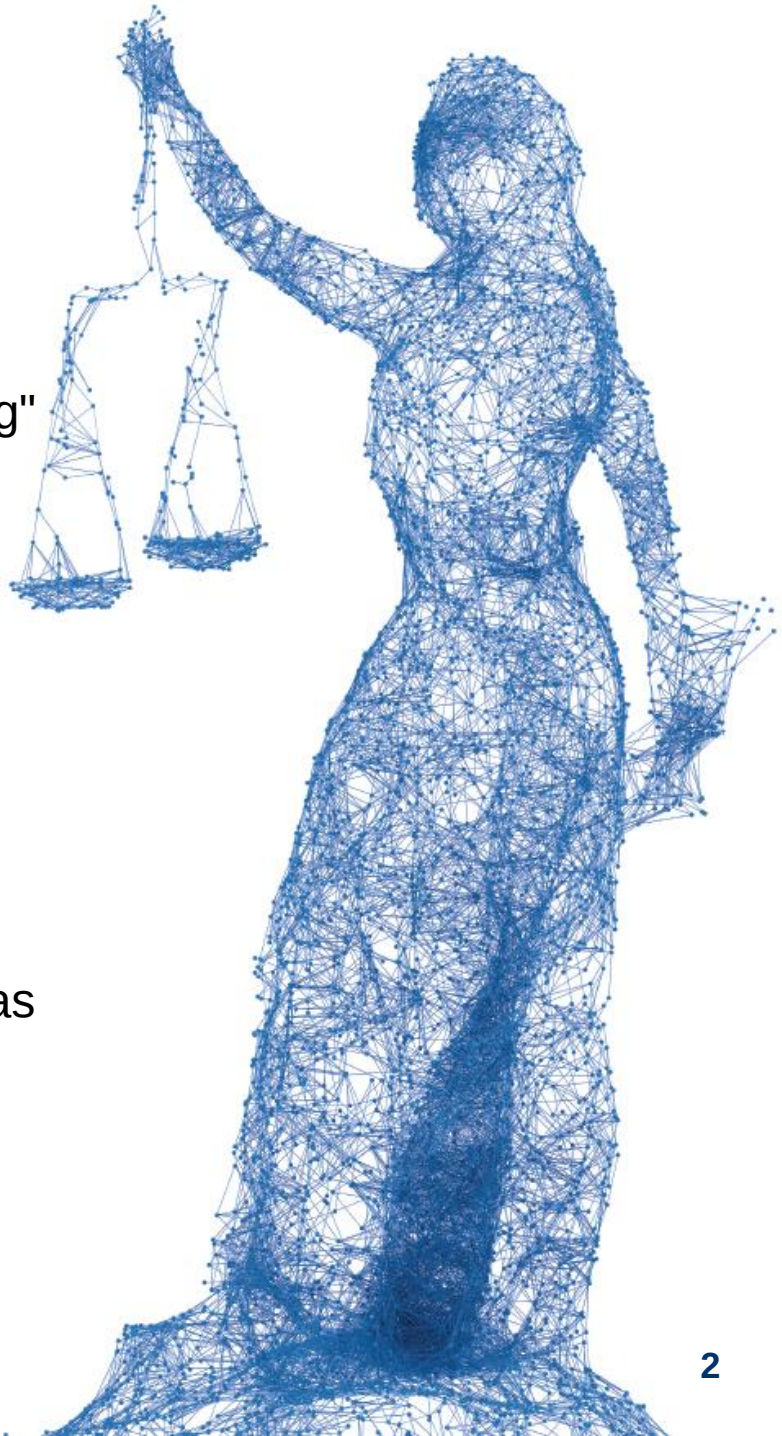
Fairness, Bias

Algorithmic bias describes systematic and repeatable errors in a computer system that create "unfair" outcomes, such as "privileging" one category over another in ways different from the intended function of the algorithm.

This is a very broad subject area that covers different topic areas:

- Most biases come from data
- Some bias could be introduced by algorithms

Fairness in machine learning refers to the various attempts at correcting algorithmic bias in automated decision processes based on machine learning models.



Data BIAS by example

PROBLEM: Our company receives thousands of CVs daily

- The openings are many and different from each other (programmer, marketing, administrative, sales, . . .)
- Just skim through the CVs requires a lot of time and effort
- Good candidates can be erroneously discarded in this preliminary phase

SOLUTION: An AI system that analyzes the CV and takes only the best candidates

- Use the CVs of the current employees as ground truth data
- We want to select candidates similar to the valuable people we already have in our company
- Our great engineers designed and developed the system with state-of-the-art models and techniques

Data BIAS by example

OUTCOME: The selected people are very good candidates

- The system performs better than our HRs in selecting good candidates
- All the ML metrics shows stunning performance

QUESTION: Are you happy? Do you approve the system? Do you give a raise to the engineers?

Data BIAS by example

OUTCOME: The selected people are very good candidates

- The system performs better than our HRs in selecting good candidates
- All the ML metrics shows stunning performance

QUESTION: Are you happy? Do you approve the system? Do you give a raise to the engineers?

Amazon scraps secret AI recruiting tool that showed bias against women

By Jeffrey Dastin

8 MIN READ

SAN FRANCISCO (Reuters) - Amazon.com Inc's AMZN.O machine-learning specialists uncovered a big problem: their new recruiting engine did not like women.

Algorithms reflect the real world, which means they can unintentionally perpetuate existing unbalance

Data BIAS by example

COMPUTING

Racial Bias Found in a Major Health Care Risk Algorithm

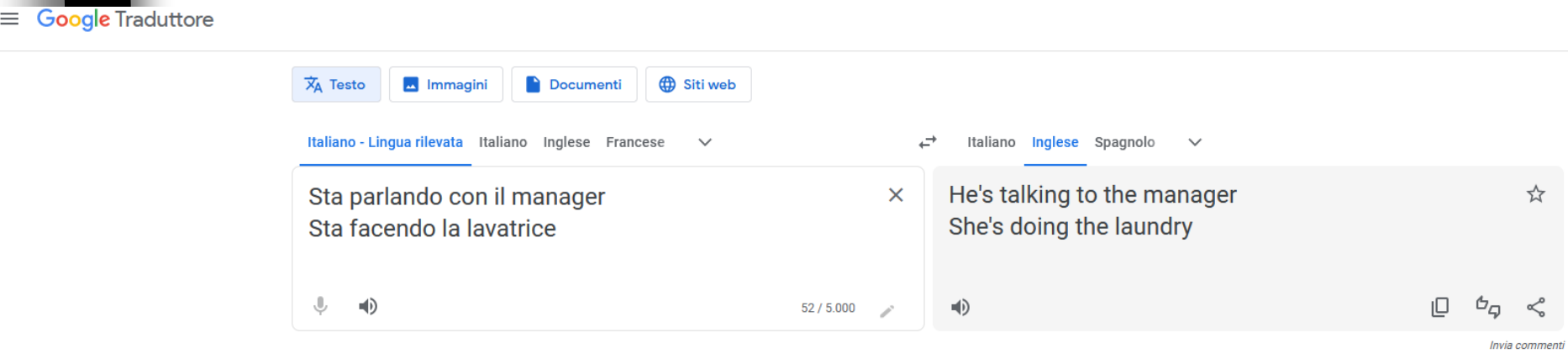
Black patients lose out on critical care when systems equate health needs with costs

By Starre Vartan on October 24, 2019

The algorithm's designers used previous patients' health care spending as a proxy for medical needs.
The researchers found this proxy arrangement did not work well because even when black and white patients spent the same amount, they did not have the same level of need: black patients tended to pay for more active interventions such as emergency visits for diabetes or hypertension complications.

implied subject

Data BIAS by example



in English, the implied subject is translated as a male/female subject depending on the context

Beyond fairness: ethics and autonomous behaviors?