

# EQUALITY, ARTIFICIAL INTELLIGENCE, ALGORITHMS AND AUTOMATED DECISIONS: HOW TO DETECT AND ADDRESS DISCRIMINATION IN THE CONTEXT OF AI?

Marzia Barbera

Department of Labour and Anti-Discrimination Law

Department of Law

University of Brescia



Funded under the 'Citizens, Equality, Rights and Values programme 2021-2027' of the European Commission

1

## SUMMARY

1. What is an algorithm
2. What we know about management by algorithms  
The risks of AI
3. What we know about anti-discrimination law  
How to use existing law
4. How to ascertain algorithmic discrimination
  - 4.1. A concrete example: the case of the Frank algorithm
5. What form of discrimination: indirect or direct?

2

## WHAT IS AN ALGORITHM

- The algorithm is one of the various applications of artificial intelligence and gives digital machines the ability to handle huge masses of data with which to process information to enable 'intelligent' operation by machines and to identify a specific solution to a given problem
- A complex class of algorithms performs what is called 'machine learning', i.e. the ability for machines to acquire their own knowledge
- The machine learning algorithm is exposed to training data to 'teach' an algorithm what to look for

3

## CONT WHAT IS AN ALGORITHM

Two separate algorithmic processes at work

1. The evaluator algorithm (or screener) takes into account the characteristics of a person and returns a prediction about the outcome for this person
2. The training algorithm (or trainer) is what produces the screening algorithm



The screening algorithm is the mechanical result of applying the training algorithm.

4



## WHAT WE KNOW ABOUT MANAGEMENT BY ALGORITHMS THE RISKS OF AI

*European Commission - White Paper on Artificial Intelligence - A European Approach, 2020*

"Artificial intelligence is developing rapidly and in doing so will change our lives for the better, but it will also entail a number of potential risks, such as opaque *decision-making* mechanisms, *discrimination* based on gender or other grounds, intrusion into our private lives or use for criminal purposes."

5



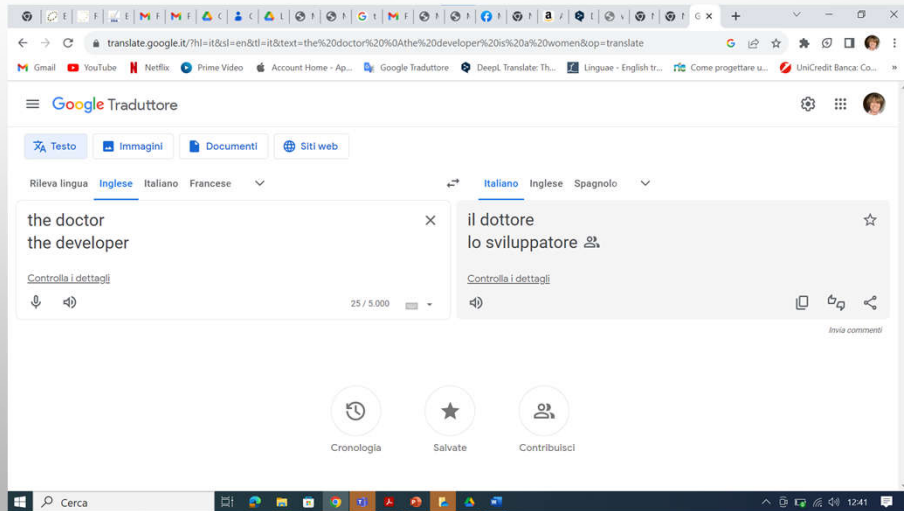
*"An algorithm is only as good as the data it works with"*

S. Barocas , A.D. Selbst, *Big Data's Disparate Impact*, California Law Review, 2016, 671.

6

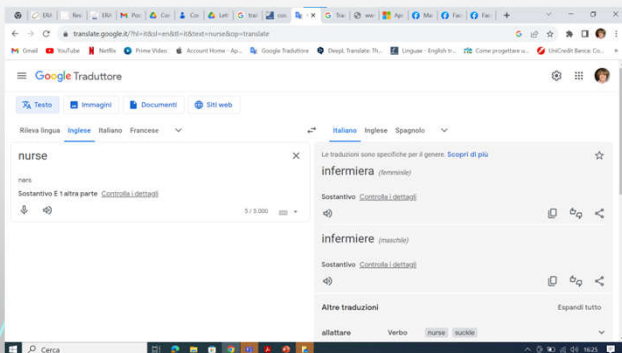
# TWO EXAMPLES OF DISCRIMINATORY MECHANISMS

## a) The gendered language of the automatic translator

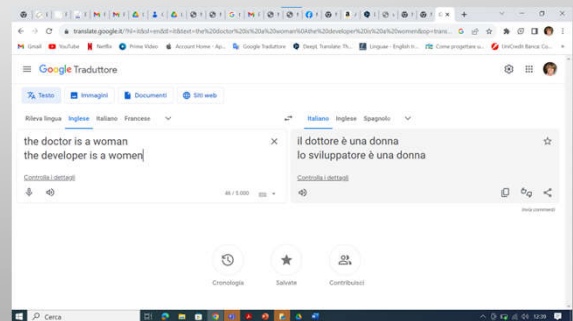


7

## attempts to correct it...



## ...with dubious successes



8

## CONT TWO EXAMPLES OF DISCRIMINATORY MECHANISMS

### b) "Three black teenagers"



In 2016, a Google Image search for 'three black teenagers' led to mugshots, while a search for 'three white kids' led mainly to images of happy people. In response to the reactions, Google stated: Our image search results reflect content from across the web. But how much do these images in turn influence the perception of the characteristics of the two groups?

*"Technology and its consequences are human phenomena and are closely dependent on the social and regulatory fabric on which they reverberate".*

A. Aloisi -V. De Stefano, *Il tuo capo è un algoritmo. Contro il lavoro disumano*, La Terza, 2020, XVII

## WHAT ANTI-DISCRIMINATION LAW KNOWS HOW TO USE EXISTING LAW

*From the point of view of discrimination theory, the sources of algorithmic discrimination can be categorised into two overarching types of inequality (re)producing mechanisms.*

*On the one hand, stereotyping and prejudice affect the equal representation of groups in society.*

*On the other, past discrimination that has been institutionalised and reified over the course of history is reflected in structural forms of inequality.*

*Algorithmic discrimination therefore both arises from, and further entrenches, hierarchising status beliefs and stereotypes as well as structural institutionalised patterns of inequality. From the perspective of discrimination law, the mechanisms through which discrimination might invade algorithms can thus be classified along these two main axes.*

R. Xenidis, L. Senden, *EU non-discrimination law in the era of artificial intelligence: Mapping the challenges of algorithmic discrimination*, in Ulf Bernitz et al (eds), *General Principles of EU law and the EU Digital Order*, Kluwer Law International, 2020

11

## A TAXONOMY OF ALGORITHMIC DISCRIMINATION

### ○ *Intentional algorithmic discrimination*

Concerns the most direct hypothesis, in which the operators in the construction of the algorithm have **intentionally** used **protected** discriminatory factors.

Es. *National Non-Discrimination and Equality Tribunal, Finland, 21 March 2018*, <https://perma.cc/ZKS8-SFNJ>: a consumer credit company used gender, age, language and their combined effect as indices of the applicant's financial reliability.

12

## CONT. A TAXONOMY OF ALGORITHMIC DISCRIMINATION

### ○ *Cont. Intentional algorithmic discrimination*

In other cases, **proxies** are **intentionally** used, i.e. data that are linked to particular social categories and can therefore lead to discriminatory behaviour on the part of the machine. The discriminatory result is not determined by the data entered into the machine, but by the characteristics attached to it.

From a legal point of view, these are clear cases of direct discrimination.

13

## CONT. A TAXONOMY OF ALGORITHMIC DISCRIMINATION

### ○ *Systemic 'historical' discrimination*

No deliberately discriminatory element was introduced, but in previous procedures someone had systematically discriminated against particular social categories, so the data used to train the algorithm was already biased from the outset.

Here again, the discriminatory effect is determined both through data directly referring to the prohibited factor and through *proxies*.

14

## CONT. A TAXONOMY OF ALGORITHMIC DISCRIMINATION

A well-known  
example of  
'historical'  
discrimination



In 2014, Amazon adopted a system that did not select applications in a gender-neutral manner, for roles as software developers and other high-tech positions. It basically excluded **women's** CVs because the AI system was programmed to choose candidates by looking at the patterns of CVs submitted to the company over a 10-year period, and most of these were from **men**.

Once corrected, the algorithm continued to produce more unfavourable treatment of women by using a **proxy**, i.e. downgrading CVs that contained references to women's colleges or typically female sports or skills. In 2017, the system was decommissioned.

15

## CONT. A TAXONOMY OF ALGORITHMIC DISCRIMINATION

### ○ *Systemic discrimination by under-representative sampling*

Dependent on inexact sampling of the data to be used in training the algorithm (training data), in particular training data sets that are not representative of the conditions of categories protected by anti-discrimination law.

16



## CONT. A TAXONOMY OF ALGORITHMIC DISCRIMINATION

An example of discrimination systemic discrimination by sampling under-representation



### *App Street Bump* from Boston

The app used the GPS information of the users to report to the public authorities which streets were to receive maintenance. In this case, the sampling was distorted by the fact that the reports were made for streets where the number of smartphones was higher, so streets in better-off neighbourhoods received more assistance than those in less well-off neighbourhoods, where the percentage of new generation phones was lower

17

## HOW TO ASCERTAIN ALGORITHMIC DISCRIMINATION THE CASE OF THE **FRANK ALGORITHM**

The company Deliveroo had adopted the Frank algorithm to establish priority conditions of riders' access to the booking of work sessions through the digital platform. The access conditions were based on static relating to "participation and reliability". of workers



18

## CONT. HOW TO ASCERTAIN ALGORITHMIC DISCRIMINATION THE CASE OF THE **FRANK ALGORITHM**

- The algorithm 'profiled' workers on an ongoing basis, based on a mechanism that established a correlation between non-participation in a work session, late cancellation and the two reliability indices. On this correlation, the 'reputational' profile of the workers was constructed, which in turn was functional to the riders' lesser or greater possibility of accessing slots and, therefore, job opportunities.
- As a result of this mechanism, workers who went on strike and were therefore absent from work were penalised in the allocation of slots and, in some cases, were excluded from the system
- In 2019 Filt, Filcams and Nidil Cgil brought a lawsuit before the Court of Bologna for discriminatory conduct based on personal beliefs (Legislative Decree 216/03)

19

## CONT. HOW TO ASCERTAIN ALGORITHMIC DISCRIMINATION THE DECISION OF THE COURT OF BOLOGNA

**Order of 31 December 2020:** The company has engaged in "indirect discrimination by enforcing an apparently neutral provision (the contractual provisions on the early cancellation of booked sessions), which, however, puts a certain category of workers (those participating in union walkouts) at a particular potential disadvantage".

20

## CONT. HOW TO ASCERTAIN ALGORITHMIC DISCRIMINATION ALGORITHM TRACEABILITY

### The evidential cost of lack of transparency

- *Danfoss*, Case 109/89: lack of transparency of decision-making mechanisms is itself of evidentiary relevance (even if not decisive).
- *Meister*, Case 415/10: in discharging the burden of proof on a party alleging discrimination, it must be ensured that the defendant's refusal to provide information does not risk undermining the achievement of the objectives pursued by Union law

21

## CONT. HOW TO ASCERTAIN ALGORITHMIC DISCRIMINATION THE FOCUS OF ANTI-DISCRIMINATION LAW: CONSEQUENCES

- The lack of evidence concerning the functioning of the Frank algorithm did not prevent the case from being decided in favour of the applicants.
- Why?
- Anti-discrimination law acts on the **consequences of actions** (different treatment; particular disadvantage) (Court of Genoa, ord. 24/12/2012), when the existence of a causal link between the different treatment and disadvantage produced and membership of a protected category leads to the presumption of discrimination.

22

## CONT. HOW TO ASCERTAIN ALGORITHMIC DISCRIMINATION THE STAGES OF JUDGEMENT

○ *Establishment of Disadvantageous Effect Treatment*

○ *Presence of one of the protected factors*

possessed by **most**, not necessarily all, of the persons concerned

*Chez*, Case 83/14: 'that measure was put in place and/or maintained for reasons related to the ethnic origin common to most of the residents of the neighbourhood in question'

23

## HOW TO ASCERTAIN ALGORITHMIC DISCRIMINATION CONT. THE STAGES OF JUDGEMENT

○ *Existence of a casual link* → **probabilistic calculation**  
**sufficient**

*O'Flynn*, case 237/94: "It is sufficient to note that the said provision is capable of producing such an effect".

**but not exclusive**

*Chez*, case 83/14: "racial motives combined with motives of another nature"

In Italian jurisprudence: *Cass. 5/4/2016, n. 657* (in the case of dismissal, it can concur with a just cause or a justified reason)

In English jurisprudence: *O'Neill v Governors of St Thomas More Roman Catholic Voluntary Aided Upper School* [1996] EAT/1180/94 (it is sufficient that it was one of the causes)

24

## HOW TO ASCERTAIN ALGORITHMIC DISCRIMINATION CONT. THE STAGES OF JUDGEMENT

### ○ *Justification* (in indirect algorithmic discrimination)

#### legitimacy of the aims pursued

The Court of Bologna took into consideration the need for the company to adopt a system that incentivised participation and rewarded the reliability of workers, concluding that 'there is no question of the legitimacy of the system in itself, nor of the fact that it incentivises the prior cancellation of booked sessions that are no longer intended to be used'.

#### proportionality of the means adopted

The Court held that the means were not proportionate because the company, while correcting other distortions in the use of the algorithm (e.g. absence due to injury or log in defect), had taken into account the fact that the negative consequences of any late cancellation or non-participation in the non-cancelled session could preclude the exercise of the constitutionally guaranteed right to strike.

25

## DIRECT OR INDIRECT DISCRIMINATION?

Cfr. M. Barbera, *Discriminazioni algoritmiche e forme di discriminazione*, *Labor and Law Issue*, 1/2021, <https://labourlaw.unibo.it/article/view/13127>

J.Adams-Prassl, R. Binns e A. Kelly-Lyth, *Directly Discriminatory Algorithms*, in *Modern Law Review*, 2023, 144–175

Is algorithmic discrimination necessarily indirect discrimination?

In many cases of algorithmic discrimination, direct discrimination is actually possible. The advantage: direct discrimination admits of no justification.

### ○ Intentional covert discrimination

There are some passages in the Bologna ordinance that suggest that a form of intentional covert discrimination could have been envisaged (this is direct discrimination).

Discrimination determined by a protected factor

*Chez*, Case 43/14, in relation to discrimination based on ethnic origin, "for there to be discrimination it is sufficient that this ethnic origin has determined the decision in question".

26

## CONT. DIRECT OR INDIRECT DISCRIMINATION

- What does «determined» mean? Being the direct cause of something
- Some jurisprudential examples:
- Implicit stereotypes or prejudices(no intent required)

*Chez*, case 83/14: there is direct discrimination if «the practice at issue **is based** on ethnic stereotypes or prejudices».

« **Inextricably linked** » to the prohibited factor: high or significant correlation with a prohibited factor (applying that criterion means applying the prohibited factor)

*Dekker*, case C-177/88: "when the reason (for refusing to employ a woman) is that she is pregnant, the decision is directly determined by the sex of the applicant."

*Maruko*, case 27/06: marriage status was considered to be inextricably linked to sexual orientation if marriage is allowed only for heterosexual couples

**Thank you for your attention!**

marzia.barbera@unibs.it