

EGUAGLIANZA, INTELLIGENZA ARTIFICIALE, ALGORITMI E DECISIONI AUTOMATIZZATE: COME RILEVARE E AFFRONTARE LA DISCRIMINAZIONE NEL CONTESTO DELL'IA?

Marzia Barbera

Ordinaria di Diritto del lavoro e Diritto antidiscriminatorio

Dipartimento di Giurisprudenza

Università di Brescia



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

1

SOMMARIO

1. Cos'è un algoritmo
2. Quel che sappiamo del *management* tramite algoritmi
I rischi dell'IA
3. Quel che ne sa il diritto antidiscriminatorio
Come usare il diritto vigente
4. Come si accerta una discriminazione algoritmica.
 - 4.1. Un esempio concreto: il caso dell'algoritmo Frank
5. Che forma di discriminazione: indiretta o diretta?

2

COS'È UN ALGORITMO

- L'algoritmo è una delle diverse applicazioni dell'intelligenza artificiale e dà la possibilità alle macchine digitali di gestire enormi masse di dati mediante i quali elaborare informazioni capaci di consentire un funzionamento "intelligente" da parte delle macchine e di individuare una specifica soluzione ad un determinato problema
- Una classe complessa di algoritmi esegue quello che viene chiamato "apprendimento automatico", ovvero la capacità per le macchine di acquisire la proprie conoscenze (il cd. "machine learning")
- L'algoritmo di apprendimento automatico è esposto a dati di addestramento che servono a "insegnare" a un algoritmo cosa cercare

3

SEGUE COS'È UN ALGORITMO

Vi sono due processi algoritmici separati al lavoro

1. L'algoritmo valutatore (o *screener*) prende in considerazione le caratteristiche di una persona e restituisce una previsione circa il risultato relativo a questa persona
2. L'algoritmo di addestramento (o *trainer*) è ciò che produce l'algoritmo di *screening*



L'algoritmo di *screening* è il risultato meccanico dell'applicazione dell'algoritmo di addestramento

4



QUEL CHE SAPPIAMO DEL *MANAGEMENT* TRAMITE ALGORITMI I RISCHI DELL'IA

Commissione europea - Libro bianco sull'intelligenza artificiale - Un approccio europeo, 2020

*«L'intelligenza artificiale si sta sviluppando rapidamente e nel far ciò cambierà le nostre vite migliorandola, ma comporterà anche una serie di rischi potenziali, quali meccanismi **decisionali opachi**, **discriminazioni** basate sul genere o di altro tipo, intrusioni nelle nostre vite private o utilizzi per scopi criminali»*

5



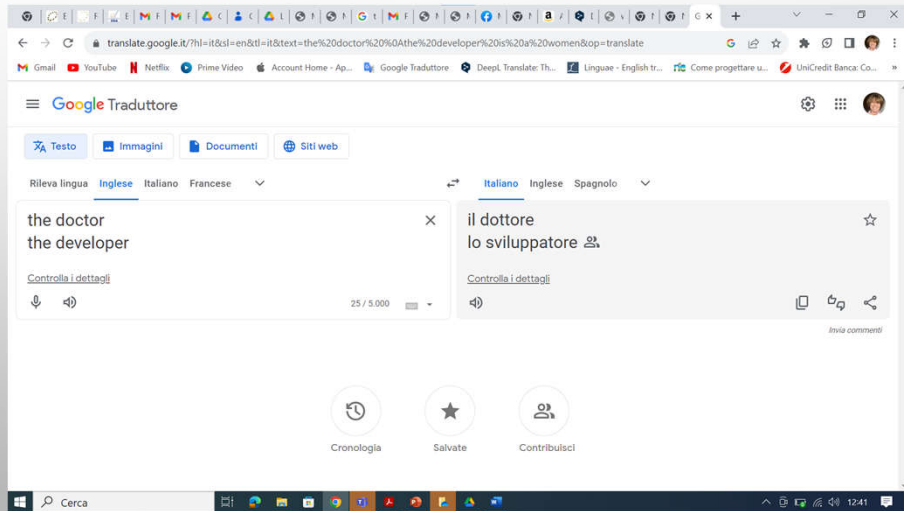
“An algorithm is only as good as the data it works with”

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

6

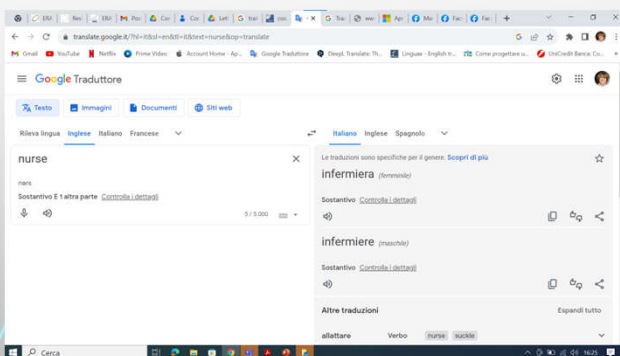
DUE ESEMPI DI MECCANISMI DISCRIMINATORI

a) Il linguaggio sessuato del traduttore automatico

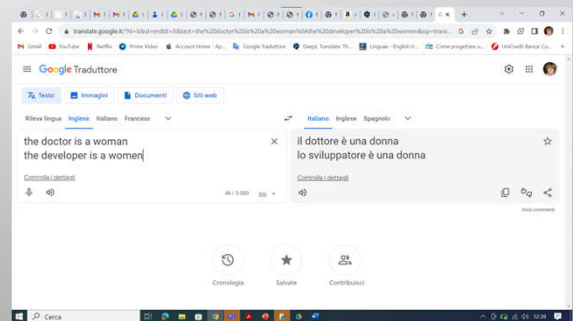


7

segue i tentativi di correggerlo...



...con successi dubbi



8

SEGUE DUE ESEMPI DI MECCANISMI DISCRIMINATORI

b) "Three black teenagers"



Nel 2016, una ricerca su Google Immagini di "three black teenagers" portava a foto segnaletiche, mentre la ricerca di "three white kids" portava principalmente a immagini di persone felici.

In risposta alle reazioni, Google dichiarò: *I nostri risultati di ricerca di immagini riflettono i contenuti provenienti da tutto il web.*

Ma queste immagini a loro volta quanto influenzano la percezione delle caratteristiche dei due gruppi?

«La tecnologia e le sue conseguenze sono fenomeni umani e dipendono strettamente dal tessuto sociale e regolamentare su cui si riverberano»

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

COSA NE SA IL DIRITTO ANTIDISCRIMINATORIO COME USARE IL DIRITTO VIGENTE

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

UNA TASSONOMIA DELLE DISCRIMINAZIONI ALGORITMICHE

○ *Discriminazione algoritmica volontaria*

Riguarda l'ipotesi più diretta, nella quale gli operatori nella costruzione dell'algoritmo abbiano **intenzionalmente** usato **fattori protetti** discriminatori

Es. *National Non-Discrimination and Equality Tribunal, Finlandia, 21 Marzo 2018*, <https://perma.cc/ZKS8-SFNJ>: una società di credito al consumatore usava quali indici di affidabilità finanziaria del/la richiedente il sesso, l'età, la lingua e il loro effetto combinato.

12

SEGUE UNA TASSONOMIA DELLE DISCRIMINAZIONI ALGORITMICHE

○ segue *Discriminazione algoritmica volontaria*

In altri casi sono usati **intenzionalmente** dei *proxies*, ovvero dati che sono collegati a particolari categorie sociali e possono quindi comportare un comportamento discriminatorio da parte della macchina. Il risultato discriminatorio non si determina per effetto dei dati immessi nella macchina, ma per le caratteristiche ad essi ricollegati

Da un punto di vista giuridico, questi sono casi evidenti di discriminazione diretta

13

SEGUE UNA TASSONOMIA DELLE DISCRIMINAZIONI ALGORITMICHE

○ *Discriminazione sistemica "storica"*

Non è stato introdotto alcun elemento volontariamente discriminatorio, ma nelle precedenti procedure qualcuno aveva sistematicamente discriminato particolari categorie sociali, di conseguenza i dati utilizzati per l'addestramento dell'algoritmo erano già di partenza alterati

Anche in questo caso l'effetto discriminatorio si determina sia attraverso dati che fanno diretto riferimento al fattore vietato sia attraverso i *proxies*

14

SEGUE UNA TASSONOMIA DELLE DISCRIMINAZIONI ALGORITMICHE

Un esempio molto conosciuto di discriminazione «storica»



Nel 2014 Amazon ha adottato un sistema che non selezionava in modo neutrale rispetto al genere le candidature, per ruoli di sviluppatori di software e altri posti ad alto contenuto tecnologico. In pratica escludeva i cv delle **donne** perché il sistema di AI era stato programmato per scegliere i candidati, osservando i modelli di cv presentati alla società nell'arco di 10 anni e la maggior parte di questi proveniva da **uomini**.

Una volta corretto, l'algoritmo ha continuato a produrre un trattamento più sfavorevole per le donne, usando un **proxy** e cioè declassando i curricula che contenevano riferimenti a college femminili o a sport o abilità tipicamente femminili. Nel 2017 il sistema è stato dismesso.

15

SEGUE UNA TASSONOMIA DELLE DISCRIMINAZIONI ALGORITMICHE

- *Discriminazione sistemica da campionamento sotto-rappresentativo*

Dipende dall'inesatto campionamento dei dati da utilizzare nell'addestramento dell'algoritmo (training data), in particolare da set di training dati non rappresentativi delle condizioni di categorie protette dal diritto antidiscriminatorio

16

SEGUE UNA TASSONOMIA DELLE DISCRIMINAZIONI ALGORITMICHE

Un esempio di discriminazione sistemica da campionamento sotto-rappresentativo



L'app *Street Bump* di Boston

L'app utilizzava le informazioni GPS degli utenti per segnalare alle amministrazioni pubbliche quali erano le strade che dovevano ricevere manutenzione.

In questo caso il campionamento risultava distorto dal fatto che le segnalazioni avvenivano per le strade dove il numero di smartphone era maggiore, perciò le strade dei quartieri più abbienti ricevevano maggiore assistenza rispetto a quelle di quelli meno abbienti, dove la percentuale di telefoni di nuova generazione era inferiore

17

COME SI ACCERTA UNA DISCRIMINAZIONE ALGORITMICA IL CASO DELL' ALGORITMO FRANK

L'azienda Deliveroo aveva adottato l'algoritmo Frank per stabilire le condizioni di priorità di accesso dei riders alla prenotazione delle sessioni di lavoro tramite la piattaforma digitale. Le condizioni di accesso erano basate su statiche relative alla "partecipazione e affidabilità" dei lavoratori



18

SEGUE COME SI ACCERTA UNA DISCRIMINAZIONE ALGORITMICA COME FUNZIONAVA L' ALGORITMO FRANK

- L'algoritmo "profilava" i lavoratori in maniera continuativa, sulla base di un meccanismo che stabiliva una correlazione fra mancata partecipazione a una sessione di lavoro, cancellazione tardiva e i due indici di partecipazione affidabilità. Su questa correlazione veniva costruito il profilo "reputazionale" dei lavoratori, funzionale, a sua volta, alla minore o maggiore possibilità dei riders di accedere agli slot e, dunque, alle occasioni di lavoro
- Per effetto di questo meccanismo i lavoratori che scioperavano e risultavano dunque assenti dal lavoro venivano penalizzati nell'assegnazione degli slot e, in alcuni casi, venivano esclusi dal sistema
- Nel 2019 Filt, Filcams e Nidil Cgil nazionali hanno promosso innanzi al Tribunale di Bologna un'azione legale per condotta discriminatoria basata sulle convinzioni personali (D.lgs 216/03)

19

SEGUE COME SI ACCERTA UNA DISCRIMINAZIONE ALGORITMICA LA DECISIONE DEL TRIBUNALE DI BOLOGNA

Ordinanza 31 dicembre 2020: La società ha messo in atto «una discriminazione indiretta, dando applicazione ad una disposizione apparentemente neutra (le disposizioni contrattuali sulla cancellazione anticipata delle sessioni prenotate) che però mette una determinata categoria di lavoratori (quelli partecipanti a iniziative sindacali di astensione dal lavoro) in una posizione di potenziale particolare svantaggio»

20

SEGUE COME SI ACCERTA UNA DISCRIMINAZIONE ALGORITMICA TRACCIABILITÀ DELL'ALGORITMO

Il costo probatorio della mancata trasparenza

- *Danfoss*, causa 109/89: la mancanza di trasparenza dei meccanismi decisionali assume essa stessa rilevanza probatoria (anche se non decisiva).
- *Meister*, causa 415/10: nell'assolvimento dell'onere della prova gravante su chi allega una discriminazione deve essere garantito che il diniego di fornire informazioni da parte del convenuto non rischi di compromettere la realizzazione degli obiettivi perseguiti dal diritto dell'Unione

21

SEGUE COME SI ACCERTA UNA DISCRIMINAZIONE ALGORITMICA IL FOCUS DEL DIRITTO ANTIDISCRIMINATORIO: LE CONSEGUENZE

- La carenza di elementi relativi al funzionamento dell'algoritmo Frank non ha impedito che la causa fosse decisa a favore dei ricorrenti.
- Perché?
- Il diritto antidiscriminatorio agisce sulle *conseguenze delle azioni* (trattamento diverso; particolare svantaggio) (Tribunale Genova, ord. 24/12/2012), quando l'esistenza di un nesso causale fra il diverso trattamento e lo svantaggio prodotto e l'appartenenza a una categoria protetta fa presumere l'esistenza di una discriminazione.

22

SEGUE COME SI ACCERTA UNA DISCRIMINAZIONE ALGORITMICA LE FASI DEL GIUDIZIO

- Accertamento del trattamento dell'effetto svantaggioso
- Presenza di uno dei fattori protetti

posseduto dalla **maggior parte dei soggetti interessati**, non necessariamente da tutti

Chez, causa 83/14: "detta misura è stata posta in essere e/o mantenuta per ragioni connesse all'origine etnica comune alla maggior parte dei residenti del quartiere di cui trattasi"

23

COME SI ACCERTA UNA DISCRIMINAZIONE ALGORITMICA SEGUE LE FASI DEL GIUDIZIO

- Esistenza di un nesso causale → è un calcolo probabilistico
sufficiente

O'Flynn, causa 237/94: "Basta rilevare che detta disposizione è in grado di produrre un effetto del genere"

ma non esclusivo

Chez, causa 83/14: "i motivi razziali che si combinano con motivi di altra natura"

Nella giurisprudenza italiana: *Cass. 5/4/2016, n. 657* (in caso di licenziamento può concorrere con una giusta causa o un giustificato motivo)

Nella giurisprudenza inglese: *O'Neill v Governors of St Thomas More Roman Catholic Voluntary Aided Upper School* [1996] EAT/1180/94 (è sufficiente che sia stata una delle cause)

24

COME SI ACCERTA UNA DISCRIMINAZIONE ALGORITMICA SEGUE LE FASI DEL GIUDIZIO

○ Giustificazioni (nelle discriminazioni algoritmiche indirette)

legittimità delle finalità perseguite

il Tribunale di Bologna ha preso in considerazione la necessità per la società di adottare un sistema che incentivasse la partecipazione e premiasse l'affidabilità dei lavoratori, per concludere che «non si discute della legittimità del sistema in sé, né del fatto che sia incentivata la cancellazione preventiva delle sessioni prenotate che non si intendono più utilizzare».

proporzionalità dei mezzi adottati

il Tribunale ha ritenuto che i mezzi non siano proporzionati perché la società, pur correggendo altre distorsioni nell'uso dell'algoritmo (es. le assenze per infortunio o difetto del *log in*), aveva tenuto conto del fatto che le conseguenze negative dell'eventuale cancellazione tardiva o mancata partecipazione alla sessione non cancellata potesse precludere l'esercizio del diritto di sciopero costituzionalmente garantito.

25

DISCRIMINAZIONE INDIRETTA O DIRETTA?

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

Le discriminazioni algoritmiche sono necessariamente discriminazioni indirette?

In molti casi di discriminazioni algoritmiche in realtà è possibile prospettare un' ipotesi di discriminazione diretta. Il vantaggio: la discriminazione diretta non ammette giustificazioni

○ Discriminazione **occulta intenzionale**

Vi sono alcuni passaggi dell'ordinanza di Bologna che fanno pensare che si sarebbe potuta prospettare una forma di discriminazione occulta intenzionale (è una discriminazione diretta)

○ Discriminazione **determinata** da un fattore protetto

Chez, causa 43/14, in relazione a una discriminazione basata sull'origine etnica, «perché si abbia discriminazione è sufficiente che tale origine etnica abbia **determinato** la decisione in questione»

26

SEGUE DISCRIMINAZIONE INDIRETTA O DIRETTA ?

○ Cosa vuol dire «determinata»? essere la causa diretta di qualcosa

○ Alcuni esempi giurisprudenziali:

○ Stereotipi o pregiudizi impliciti (non occorre intento)

Chez, causa 83/14: vi è discriminazione diretta se «*la prassi controversa si fonda su stereotipi o pregiudizi di ordine etnico*»

○ «**Inestricabilmente connessa**» al fattore vietato: alta o significativa correlazione con un fattore vietato (applicare quel criterio vuol dire applicare il fattore vietato)

Dekker, causa C-177/88: «*allorché il motivo (del rifiuto di assumere una donna) è costituito dal fatto che l'interessata è incinta, la decisione è direttamente determinata dal sesso del candidato*»

Maruko, causa 27/06: lo stato matrimoniale è stato considerato inestricabilmente connesso all'orientamento sessuale se il matrimonio è permesso solo alle coppie eterosessuali

Grazie dell'attenzione!

marzia.barbera@unibs.it