

INTELLIGENZA ARTIFICIALE E DISCRIMINAZIONE DI GENERE

Comitato Pari Opportunità Ordine Avvocati
Torino – 21-06-23

Avv. G. Cavasino

AMBITI DI APPLICAZIONE

IA E MEDICINA

VANTAGGI

- L'IA è sempre più utilizzata in ambito medico per le diagnosi (diagnosi personalizzate)
- Incremento di strumenti volti alle pratiche di prevenzione (prevenzione per le donne);
- Medicina predittiva: permetterà di diagnosticare preventivamente una malattia. Prevedere le complicanze e personalizzare il trattamento farmacologico.
- Vantaggi nella gestione del sistema sanitario, gestione del pronto soccorso, grandi sviluppi nella ricerca clinica e farmacologica, controllo e monitoraggio delle pandemie, migliora la gestione delle patologie croniche (diabete), gestione attraverso le App;

AMBITI DI APPLICAZIONE

IA E MEDICINA

Svantaggi

RISULTATI DISCRIMINATORI -Falsa rappresentazione della popolazione-
mancanza di un campione rappresentativo – Presenza di stereotipi di genere nella società – pregiudizio del programmatore
nel 2019 uno studio ha evidenziato che nel trattamento della malattia di Lyme le donne avevano maggiori probabilità di avere diagnosi errate e di ottenere un trattamento inadeguato rispetto agli uomini perché la malattia è considerata tipicamente maschile. Spesso alcune malattie alle donne non vengono diagnosticate perché attribuite a stress.

TUTELA DEL DIRITTO ALLA SALUTE:

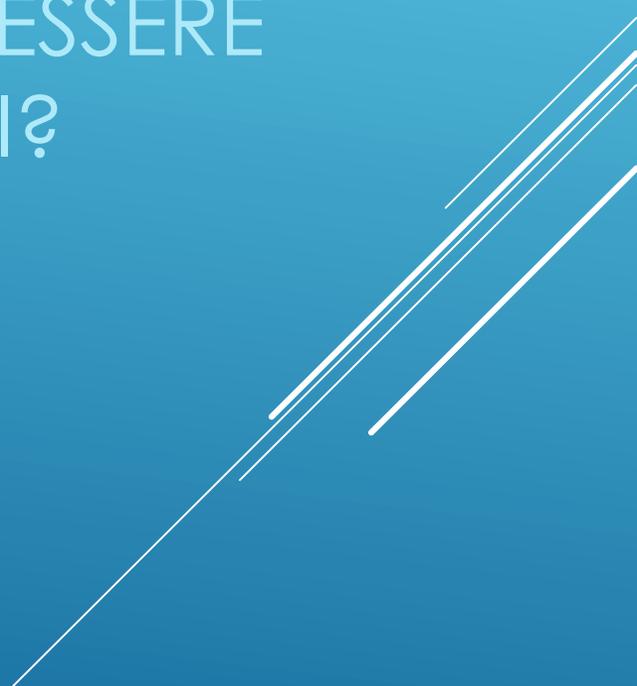
DIVARIO TECNOLOGICO- TECNOLOGIA COMPLESSA E POCO DIFFUSA –
RISCHI PER LE CATEGORIE VULNERABILI (ANZIANI)

RAPPORTO DI CURA MEDICO E PAZIENTE:

IA POTREBBE FORNIRE UN PARERE OPPOSTO A QUELLO DEL MEDICO –
PERDITA DI FIDUCIA DA PARTE DEL PAZIENTE – FENOMENO DELLA PERDITA
DI COMPETENZE TRA I MEDICI

- ▶ PRINCIPIO DELLA CONOSCIBILITA' – TRASPARENZA- conoscenza da parte del paziente dell'uso dell'IA
- ▶ PRINCIPIO DELLA NON DISCRIMINAZIONE- eliminazione degli elementi pregiudizievoli
- ▶ PRINCIPIO DELLA NON ESCLUSIVITA' - AUTODERMINAZIONE- il paziente potrà sempre rifiutare l'uso del sistema di IA e le decisioni

QUALI PRINCIPI
DEVONO ESSERE
GARANTITI?



IA E GIUSTIZIA

- Ambito stragiudiziale: per la redazione di contratti e atti standard
- Uso di IA nel processo
- Amministrazione della giustizia: fase di indagine in ambito penale- riconoscimento facciale
- Giustizia predittiva: radicale sostituzione del giudice

GIUSTIZIA PREDITTIVA

Pro e contro

- Ipotesi teorica
- Renderebbe molto difficile l'appello perché la decisione di basa su un calcolo statistico e non su un ragionamento giuridico
- Fattore culturale e umano: ci si aspetta che un umano venga giudicato da un umano
- Si potrebbe utilizzare su questioni di modesta entità (Cina ed Estonia è già utilizzato – reati minori– concessione decreti ingiuntivi – decreti penali di condanna)
- Potrebbe essere utilizzata come supporto alla decisione del Giudice (ipotesi più realistica)
- Pre – parere per il Giudice, con la possibilità di disattenderlo

IA E VEICOLI A GUIDA AUTONOMA

PRO E CONTRO

- Diminuzione degli incidenti
- Maggiore autonomia per le persone con mobilità ridotta
- Per il diritto c'è un problema di responsabilità –
Conducente/Proprietario

DISCRIMINAZIONI ALGORITMICHE

Si tratta di una discriminazione generata dall'algoritmo che SBAGLIANDO, poiché non è in grado di profilare attendibilmente per via di dati incompleti, o obsoleti o di errori nella costruzione di algoritmi, crea di fatto una discriminazione.

DISCRIMINAZIONE ALGORITMICHE DI GENERE

Abbiamo visto che se i dataset utilizzati per addestrare l'intelligenza artificiale non sono sufficientemente numerosi e diversificati introducono diverse tipologie di **bias**

Il così detto **data training** diventa fondamentale ed è composto dalle informazioni che produciamo e che informano l'intelligenza artificiale che opera per parametri di comparazione.

L'immaginario culturale e sociale è intriso di discriminazioni di vario tipo che vengono poi riprodotte dall'IA.

ESEMPIO:

Una IA neutra visiona 4 foto, nelle prime tre ci sono tre donne in cucina, nella quarta foto c'è un uomo. L'esperienza della IA è di aver visto innumerevoli foto di donna in cucina per cui indicizza anche l'uomo della quarta foto come donna.

Cucina – pentole- cibo= donna

A lungo andare l'IA impara che certe parole sono associate a certi ruoli di genere e i modelli neutrali non solo riflettono le asimmetrie sociali controverse, ma le **ESAGERANO**» rafforzando la presenza di tratti esistenti in un ruolo, piuttosto che correggerli.

Casi rilevanti di discriminazione di genere

Gli algoritmi di selezione del personale di due colossi come **FACEBOOK** e **AMAZON**, si sono resi colpevoli di aver causato discriminazione di genere:

- nel primo caso il social network più popolare del mondo si è reso conto di proporre a uomini bianchi annunci per lavori meglio pagati rispetto a quelli proposti alle donne, specie se di colore.
- Amazon ha dovuto riprogrammare da zero la sua IA per lo screening iniziale dei curricula in quanto selezionava esclusivamente uomini. Proprio la scelta del database dato in pasto all'IA di Amazon si era rivelata la maggiore criticità poiché conteneva dati esclusivamente sugli assunti negli ultimi 10 anni, per la stragrande maggioranza uomini.

Tra i casi che hanno fatto scuola c'è quello della carta di credito Apple di Goldman Sachs, accusata al suo debutto di discriminare le donne perché i suoi algoritmi avrebbero fissato tetti di spesa più bassi per le donne, in quanto tali.

Altri casi di discriminazione si verificano con l'uso degli **algoritmi di riconoscimento vocale**, essenziali per la generazione di sottotitoli automatici, che di frequente si dimostrano meno efficaci con voci femminili.

La ragione dietro questa carenza risiede nella differenza di frequenze nelle voci femminili, più che nelle voci maschili, più basse in virtù di corde vocali più lunghe, caratteristiche queste ultime che sono state maggiormente analizzate e modellate negli studi sul parlato propedeutici per la creazione dei suddetti algoritmi.

FORME DI DISCARIMINAZIONI ALGORITMICHE

Il software COMPASS (Usa) , grazie ad un algoritmo predittivo, valutava il rischio di recidiva in un soggetto e la conseguente misura cautelare da applicare.

Il risultato nell'utilizzo di questo software ha evidenziato che i cittadini afroamericani avevano quasi il doppio delle probabilità che venissero considerati recidivi rispetto ad un cittadino bianco.

il caso della c.d. "DISCRIMINAZIONE DI MASSA": il caso più noto è quello delle ricerche effettuate del termine "gorillas" tramite Google, che nel 2016 faceva apparire tra i primi risultati svariate immagini di una giovane coppia afroamericana;

|

FORME DI DISCRIMINAZIONI ALGORITMICHE

II RICONOSCIMENTO FACCIALE

Una ricercatrice al MIT, ha sviluppato un software di riconoscimento facciale che prevedeva di mostrare diversi volti umani ad un computer da analizzare, l'algoritmo non è riuscito a riconoscere il suo volto, perché si trattava di una donna di pelle nera.

L'algoritmo ha riconosciuto il volto della ricercatrice solo quando si è messa una maschera bianca.

Ha riscontrato un errore del 34,7% per le donne afroamericane mentre per gli uomini caucasici solo del 0,8%.

FORME DI DISCRIMINAZIONI ALGORITMICHE

II RICONOSCIMENTO FACCIALE

Nel 2016 la NIST (Ente federale statunitense) ha sottoposto oltre 18 milioni di immagini raffiguranti 8 milioni di persone a 200 algoritmi realizzati da 99 sviluppatori diverse per valutarne l'efficacia e la correttezza del funzionamento

Il Colore della pelle, il sesso e l'età sono risultati gli elementi comuni di errate interpretazioni

Le IA hanno dimostrato di essere più precise con uomini bianchi di mezza età .

Se si ipotizzasse di utilizzare un algoritmo di riconoscimento facciale in una macchina a GUIDA AUTONOMA, ci sarebbe una percentuale ben maggiore di possibilità che l'auto investa una donna di colore piuttosto che un uomo bianco a causa di una fisionomia poco corrispondente ai suoi criteri di identificazione.

FORME DI DISCRIMINAZIONI ALGORITMICHE

Il Caso **OPTUM** un sistema utilizzato negli ospedali statunitensi per prevedere quali pazienti avrebbero potuto avere necessità di cure ulteriori mediche.

Discriminava perché considerava solo gli uomini bianchi anche se il colore della pelle non rientrava tra le variabili considerate, ma il dato che veniva preso in considerazione erano i costi sanitari, i neri risultavano aver sostenuto costi inferiori rispetto agli uomini bianchi.

La discriminazione è un fenomeno direttamente correlato alla condotta umana, secondo una costruzione relazionale di tipo causale tra la seconda e la prima.”

Nel caso della discriminazione algoritmica ci troviamo di fronte ad una **oggettiva difficoltà nel tracciare un collegamento tra la condotta umana e la discriminazione, motivata dalla intermediazione della macchina,** questo si pone così al centro delle problematiche teoriche e applicative, che investono i rapporti tra la discriminazione, il diritto antidiscriminatorio e l'intelligenza artificiale.

Un ulteriore dato certo è che l'IA distingue: poiché la macchina è chiamata ad operare distinzioni sulla base di una definita e chiusa selezione di dati.

Se, quindi, l'intelligenza artificiale per sua natura **distingue**, si tratterà di interrogarsi non tanto sul se le tecniche di intelligenza artificiale distinguono, quanto piuttosto sul come lo fanno. **Quindi dovremo capire come questa distinzione possa poi diventare una “discriminazione”.**

Parleremo di «discriminazione diretta o indiretta»

- **DIRETTA:** situazione in cui una persona è, o potrebbe essere trattata meno favorevolmente di un'altra in una situazione comparabile, a causa di una delle caratteristiche vietate.

Molto difficilmente si potrà riportare la discriminazione algoritmica all'interno della discriminazione diretta (ad eccezione del caso dell'azione diretta del programmatore)

- **INDIRETTA:** cioè tutte le volte cui ci troviamo di fronte ad una causa della discriminazione ignota, il cui effetto ha però carattere discriminatorio nei confronti di una certa categoria di soggetti.
- Un problema legato alla discriminazione indiretta è che spesso questa non viene riconosciuta dai soggetti discriminati. Se ad esempio un algoritmo per la ricerca di occupazione discrimina in funzione del sesso, della razza o della minoranza etnica, questi soggetti non hanno modi accorgersi della discriminazione ne sono in grado di apprendere i processi che li pongono in una situazione di svantaggio.

Parleremo di «discriminazione diretta o indiretta»

Il Tribunale di Bologna, sul caso Deliveroo:

L'algoritmo FRANK non distingueva tra i lavoratori che annullavano il proprio turno di lavoro per motivi legittimi, quali il diritto di sciopero, e i lavoratori che annullavano il proprio turno per motivi futili, penalizzando entrambi nel medesimo modo.

Il Giudice ha censurato la «cecità della macchina» ma ha ritenuto di identificare il soggetto responsabile con l'utilizzatore finale, sostenendo che Deliveroo sapeva che la piattaforma discriminava ma se ne è disinteressata.

Questa commistione tra elementi oggettivi e soggettivi, e quindi tra discriminazione diretta e indiretta, ha tipicizzato la «discriminazione artificiale»

Se ci chiediamo chi si occupano effettivamente di intelligenza artificiale?
Uomini, prevalentemente bianchi.

Attualmente infatti solo il 16,1% dei docenti di *Computer Science* nelle migliori università mondiali è rappresentato da donne, e la percentuale di donne laureate in tale settore è addirittura diminuita (anche se a vantaggio delle donne laureate in altre discipline STEM come fisica e biologia).

Anche a livello lavorativo il quadro rimane simile, come dimostrano le statistiche della forza lavoro dedicata all'IA di Google e Facebook, dove la rappresentanza femminile si assesta rispettivamente al 10% e al 15%.

In Italia il tasso di occupazione delle donne nel settore ICT sia solo dal 15%, al venticinquesimo posto fra i paesi dell'Unione Europea

GRAZIE

