UNIVERSITÀ
DEGLI STUDI
DI PADOVA

DIPARTIMENTO
DI INGEGNERIA
DELL'INFORMAZIONE

DIPARTIMENTO DI INGEGNERIA DELL'INFORMAZIONE

CORSO DI LAUREA IN INGEGNERIA BIOMEDICA

"Questioni di genere nelle applicazioni di IA in medicina"

Relatore: Prof. Rodà Antonio

Correlatrice: Prof.ssa Badaloni Silvana

Laureando: Casson Francesco
N^{O} matricola: 1217378

ANNO ACCADEMICO 2023 – 2024

Data di laurea 15-07-2024

Sommario

L'Intelligenza Artificiale sta prendendo sempre più piede nell'ambito medico: dalla diagnostica, alle predizioni, agli aspetti organizzativi, fino alla semplice consultazione per informazioni sanitarie.

Questa nuova tecnologia, in rapido sviluppo e diffusione, non è però esente dalla presenza di stereotipi o "bias"; in questa relazione viene preso in esame come la disparità di genere possa alterarne i risultati.

Viene dunque fatto un excursus sulla storia della medicina di genere, dai primi movimenti di protesta negli USA, fino ai più recenti provvedimenti presi a riguardo.

Si andranno poi ad elencare i vari ambiti sanitari in cui le Intelligenze Artificiali vengono utilizzate, per poi giungere all'analisi dei bias di genere. Tutte le possibili cause di questi ultimi vengono trattate nei due capitoli finali, fino a concludere con la proposta e la discussione di alcuni metodi per effettuare un "debiasing" degli algoritmi.

Come verrà evidenziato, le varie modifiche ai metodi di analisi e restituzione dei dati da parte degli algoritmi sono solo rimedi temporanei. La vera parità di genere va raggiunta facendo fronte alle numerose ingiustizie sociali che si ripercuotono nei confronti delle donne, nonché eliminando i vari stereotipi di genere che da sempre plasmano la nostra società.

1. INTRODUZIONE

L'introduzione dell'Intelligenza Artificiale (IA), nata nel 1956 (seminario presso il Dartmouth College di Hanover; New Hampshire; USA), sta recentemente permeando ogni ambito delle nostre attività prospettandosi come una delle grandi rivoluzioni dell'umanità.
Le macchine e gli algoritmi che stiamo costruendo ci hanno già superato in molte forme specifiche di intelligenza. Molti ritengono sia necessaria una "governance" ovvero la fissazione di regole per il suo utilizzo.
L'utilizzo che è stato fatto finora dell'Intelligenza Artificiale (IA), non può essere considerato "corretto" da un punto di vista etico, tant'è che molti casi di "bias" di vario tipo sono stati rilevati nelle IA chiamate "data-dependent". Queste sono algoritmi, principalmente di Machine Learning (ML), le cui performance sono legate ai dati con cui vengono addestrate.
Un caso celebre è quello dell'algoritmo COMPAS, utilizzato dalla magistratura americana come strumento ausiliare per i giudici nello stimare la possibile recidività degli imputati; si ritiene che esso violasse il quattordicesimo emendamento della costituzione, tramite la discriminazione razziale degli imputati. [1]
Risale al maggio del 2017, l'introduzione del termine "Algoretica" [2] che designa un nuovo campo di studi che indaga i problemi etici connessi all'avvento dell'IA e in particolare degli strumenti che si basano sugli algoritmi.
E' risaputo che la scelta dei dati forniti all'algoritmo come "training set" è la fonte principale dei bias legati alle IA (principalmente nelle "Neural Networks" e "Deep Learning").
Per garantire la correttezza ("fairness") si possono seguire due possibili approcci:
- Data debiasing;
- Model debiasing. [3]

In questa relazione verrà affrontato il tema dei "bias di genere" (in particolare quelli legati al sesso biologico) analizzandone i risvolti in ambito sanitario.
Ad una breve panoramica sulla storia della medicina di genere (1),
seguiranno una valutazione dei possibili utilizzi dei sistemi di IA in medicina (2),
l'analisi dell'origine dei bias di genere (3) e quindi
l'analisi dei possibili metodi di de-biasing (4).

2. STORIA DELLA MEDICINA DI GENERE

La medicina di genere è un nuovo approccio alla medicina nato negli anni '90. Alcuni osservatori fanno risalire la sua nascita nell'Istituto Nazionale della Salute (NIH) degli Stati Uniti. Bernardine Patricia Healy, appena diventata Direttrice dell'Istituto di Cardiologia rilevò che la ricerca scientifica in quell'Istituto era condotta solo sugli uomini e sugli animali maschi e che, a livello clinico, le donne erano sottoposte molto meno degli uomini a procedure diagnostiche e terapeutiche, quali coronarografie, trombolisi, stent coronarici.
Prima di quegli anni non erano mai stati approfonditi temi quali: la risposta differenziale ai farmaci tra uomo e donna, nonché la diversa suscettibilità alle malattie tra i due sessi. [4]

Queste problematiche, tuttavia, erano già state introdotte negli anni '60 sia in Europa che negli USA: risale al 1969 la scrittura di "Our Bodies, Ourselves" (redatto dal Women's Health Book Collective di Boston), un libro in cui si affrontano le problematiche della salute femminile, della sessualità e dell'aborto.
Tuttavia è solo nel 1990 che il "Women's Health Equity Act" viene approvato dal congresso degli Stati Uniti; è un decreto che di fatto sancisce l'inizio della medicina di genere (si pensi all'articolo che rende obbligatoria l'inclusione delle donne nei test clinici).
Prima di allora, la maggior parte dei dati raccolti nei test clinici, riguardava esclusivamente i maschi (ad esempio il dosaggio dei farmaci era sempre stato basato sullo standard del maschio adulto con peso corporeo di 70 kg). [4]

È di fondamentale importanza anche la fondazione, sempre nel 1990, dell'Office of Research on Women's Health (ORWH), agenzia all'interno del National Institute of Health (NIH) dedicata esclusivamente alle questioni sanitarie femminili.
Nel 1993, la Food and Drug Administration (FDA), compie un altro grande passo verso l'uguaglianza di genere dichiarando che gli studi clinici devono includere le donne nelle stesse proporzioni degli uomini.
In Europa, le questioni di genere approdano solo nel 2002 con il "Sixth Research Framework Program (FP6)".
Nel 2004, "l'International Conference on Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH)" pubblica il suo primo articolo nel quale si stabilisce la necessità della presenza di entrambi i sessi nei test clinici. La stessa ICH ha in seguito ri-considerato l'inclusione delle donne nei test clinici proponendo di considerarla come una "categoria a sé" nella caratterizzazione del paziente.

Nel 2022 la ICH ha assegnato alle donne la categoria di “special population” insieme ai pazienti di geriatria e ai bambini. [5]

Secondo dati dell’ISS (Istituto Superiore di Sanità) del 2022, la percentuale di donne incluse nella sperimentazione relativa alla sicurezza dei nuovi farmaci è del 20-25% [6]. Questo dato ci fa riflettere sull’enorme disparità ancora esistente nonostante le questioni di genere siano tra i temi più discussi ai giorni nostri.

La medicina di Genere non si limita solamente alla farmacologia; molti studi focalizzati sulla questione di genere vengono infatti condotti su malattie più comuni.
L’osteoporosi è una malattia sistemica la quale, per molto tempo, si è ritenuto affliggesse quasi esclusivamente le donne; tuttavia, i dati riportano che quasi un terzo dei casi di frattura dell’anca di natura osteoporotica coinvolge gli uomini, evidenziando dunque che la malattia non è semplicemente una conseguenza della menopausa e quindi, da ricercare soltanto nelle donne. A tal proposito i ricercatori si sono già messi all’opera, scoprendo dei fattori secondari (oltre alla menopausa) come l’ipogonadismo e l’ipercalciuria, che possono indurre alla condizione di osteoporosi. Questi fattori sono stati poi uniti in un nuovo metodo di diagnosi della malattia che ne tenesse in considerazione, il “SECOBs” (Secondary Contributors to Osteoporosis and Metabolic Bone Disorders). [7]

Anche le tecnologie biomediche devono iniziare a tener conto del fattore di genere: uno studio condotto da Rem Konig presso l’Harvard Business School ha stimato che, se i brevetti sulle tecnologie biomediche rilasciati dal 1976 al 2010 avessero avuto come “paziente obiettivo” le donne nella stessa percentuale degli uomini, avremmo svariate migliaia di strumenti biomedici in più dedicati alla salute femminile. Questa stima dipende specialmente dal fatto che tendenzialmente i gruppi di ricerca tendono a sviluppare tecnologie che verranno utilizzate su pazienti dello stesso sesso degli sviluppatori; ciò implica che la carenza di tecnologie dedite alla salute delle donne è dovuta principalmente alla scarsa presenza di quest’ultime negli ambiti di ricerca. [8]

Questa disparità ha portato ad una sorta di rivoluzione denominata FemTech; termine coniato da Ida Tin nel 2016, indica “software, diagnostiche, prodotti, e servizi che usano la tecnologia per focalizzarsi sulla salute femminile”. Esempi di queste tecnologie possono essere “Evvy” una startup che distribuisce test per analizzare il microbioma vaginale da casa, al fine di combattere infezioni e dolori che spesso non vengono diagnosticati; oppure Ring Echo un sistema di imaging diagnostico, estremamente accurato e indolore per il paziente, che punta a sostituire la mammografia, da molti ritenuta estremamente dolorosa. [9]

3. VALUTAZIONE DEI POSSIBILI UTILIZZI DELL'IA IN MEDICINA

Tra le principali applicazioni dell'IA in ambito sanitario si possono ricordare i seguenti macrogruppi:

- Strumenti per informazioni sanitarie;
- Strumenti organizzativi;
- Sistemi di diagnosi;
- Sistemi di predizione.

I chatbot, rappresentano la maggioranza degli strumenti IA utilizzati in ambito sanitario e hanno la potenzialità di dare un contributo in ciascuno dei macrogruppi sopra citati.

I chatbot sono software configurati per simulare una conversazione con un essere umano. Vengono anche chiamati "Large Language Model" (LLM) e simulano ed elaborano il linguaggio umano grazie all'ingente quantità di dati utilizzati per il loro training e alla grande quantità di parametri appresi durante lo stesso. Il più famoso è sicuramente il modello GPT (Generative Pre-trained Transformer) lanciato il 30 novembre 2022 dalla società non-profit OpenAi. ChatGPT ha dimostrato una notevole capacità di interagire come un essere umano, fornendo risposte molto accurate tant'è che ne è stato suggerito l'utilizzo in ambito sanitario.

I chatbot come **GPT-4** possono essere di supporto nei seguenti ambiti:

- Triage: guidando il paziente tramite domande mirate riguardanti i suoi sintomi e la sua storia clinica, per comprendere la severità della sua condizione. Velocizzando così tutto il processo di triage e fornendo un rapido e mirato pronto intervento.
- Compilazione: aiutando i medici nella scrittura e compilazione di documenti, report e cartelle cliniche suggerendo soluzioni e correzioni.
- Comunicazione: mediando la comunicazione tra il medico e il paziente, rendendo il colloquio più facilmente comprensibile ad entrambi.
- Organizzazione: diventando un assistente virtuale al paziente nell'atto di consultare la propria cartella clinica, prenotare un appuntamento o ricevere ricette o altri trattamenti clinici.
- Educazione: fornendo rapido accesso alle informazioni e alle risorse necessarie per lo studio e l'apprendimento, utili a medici e studenti, oltre che per assistere al processo di insegnamento. [10]

Sono proprio alcuni di questi possibili utilizzi che hanno portato l'Organizzazione Mondiale della Sanità (OMS) a sviluppare il proprio chatbot "**S.A.R.A.H**" (acronimo di Smart AI Resource Assistant for Health) il quale è in grado di fornire supporto ai pazienti e a chiunque fosse interessato, fornendo consigli ed informazioni di base..

SARAH, non è tuttavia predisposto per fornire pareri medici. [11]

Funzione di cui invece è predisposto il chatbot di Google **MedPaLM** il quale fu la prima IA in assoluto a superare l'esame per l'ottenimento della licenza medica negli Stati Uniti. MedPaLM è infatti in grado di analizzare l'anamnesi del paziente al fine di individuare la patologia di cui è affetto e di analizzare immagini come Raggi X o Mammografie. [12]

Le IA sembrano avere un futuro prospero nell'ambito medico, basti pensare al loro enorme contributo nella lotta al COVID-19 nella diagnosi, valutazione delle prognosi, predizione dell'evoluzione della pandemia e nella ricerca per la cura al virus. [13]
E' per questo ed altri motivi che la FDA il 13 Maggio 2024 ha autorizzato l'utilizzo di ben 882 IA (basate sul Machine Learning) come strumenti di ausilio alla professione medica. [14]

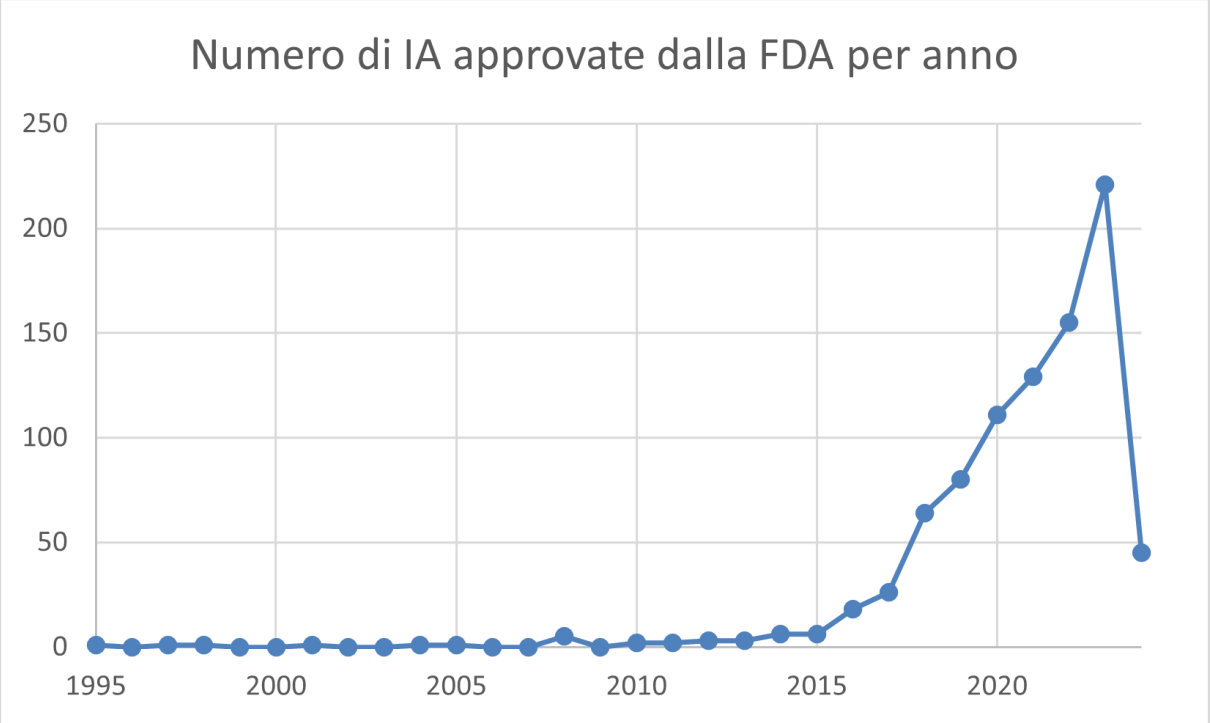

Figura 1: Distribuzione temporale del numero di IA approvate dalla FDA a partire dal 1995 fino a Maggio 2024

Come illustra il grafico, vi è un aumento esponenziale in termini di quantità di Intelligenze artificiali che ogni anno vengono approvate dalla FDA; ciò ci permette di meglio visualizzare il rapido sviluppo di questo fenomeno tecnologico.

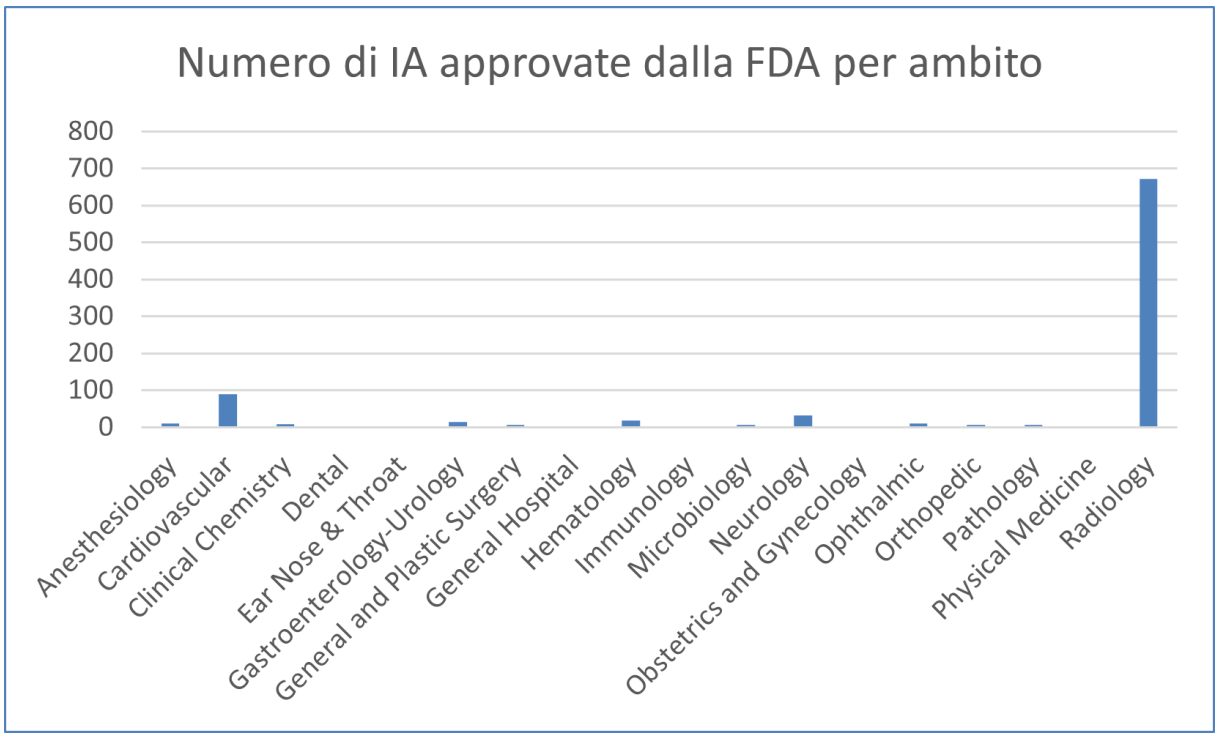

Figura 2: Istogramma rappresentante la quantità di IA approvate dalla FDA suddivise in base all'ambito nel quale sono adoperate

Di questi 882 algoritmi la vasta maggioranza è impiegata in Radiologia; questa grande presenza nel settore è basata sulla grande capacità degli algoritmo di analizzare immagini e, non meno importante, sulla grande adesione da parte degli ospedali ad un sistema di imaging universalizzato: "DICOM" (Digital Imaging and Communications in Medicine) è uno standard tecnico di comunicazione creato con lo scopo di facilitare le comunicazioni tra hardware e software (spesso di diversa manifattura), utilizzando appunto una struttura standardizzata per la creazione e condivisione di file. Il fatto che le immagini siano strutturate similmente in tutti gli ospedali comporta un grande aiuto alle intelligenze artificiali, le quali potranno dunque risultare ampiamente più accurate nell'analisi di esse.

Il Parlamento europeo ha da poco approvato la prima legge al mondo per regolamentare l'applicazione dell'IA; con l'AI act [15] l'Unione Europea vuole facilitare "l'immissione sul mercato, la messa in servizio e l'utilizzo dei sistemi di Intelligenza Artificiale (AI sistemi) nell'Unione, in conformità con i valori dell'Unione" e stabilire nuove normative in merito all'utilizzo di tali tecnologie. È da sottolineare che si tratta di una legge "orizzontale", generale, che stabilisce obiettivi di alto livello applicabili a tutti i settori della società. I dettagli verranno forniti successivamente nelle linee guida, negli standard e nelle leggi e politiche degli stati membri. Non è chiaro come la legge si intersecherà con la legislazione settoriale preesistente per l'IA medica: alcuni ritengono fosse già sufficiente il preesistente Regolamento relativo ai Dispositivi Medici (MDR). [16]

4. ORIGINE DEI BIAS DI GENERE IN IA

Il bias di genere è una delle minacce alla "fairness" nello sviluppo dell'IA nel settore medico. Le sue radici affondano in 5 macrocategorie principali:

a- Mancanza di eterogeneità dei dati e degli sviluppatori;
b- Stereotipi di genere;
c- Bias nei dati a causa dell'algoritmo;
d- Fattori economici;
e- Comportamenti e decisioni biased. [17]

4a. La mancanza di eterogeneità dei dati e degli sviluppatori

La mancanza di eterogeneità dei dati e degli sviluppatori è la principale minaccia in IA.
In medicina, è argomento molto dibattuto la carenza di eterogeneità nei dati relativi al genere. Nonostante la consapevolezza di questa lacuna, siamo ancora lontani dal colmarla. Secondo l'ISS solo il 20-25% dei soggetti coinvolti nei test è di genere femminile.
L'eterogeneità va cercata anche negli specialisti che lavorano nel settore (sviluppatori); secondo il "Artificial Intelligence, platform work and gender equality" dell'European Institute for Gender Equality, solamente il 16% dei tecnici IA sono donne. Questo gap si accentua se consideriamo gli anni di esperienza nel settore: da un 20% nelle persone con 0-2 anni di esperienza si scende fino ad un 12% per persone con più di 10 anni di esperienza. [18]
L'eterogeneità (non solo per quanto riguarda il genere) dei gruppi di lavoro è importante poiché garantisce uno sviluppo più corretto ed imparziale delle IA. Infatti, persone con background diversi possono offrire esperienze variegate riducendo, per quanto riguarda ad es. i bias, i potenziali fattori di rischio associati ai gruppi di lavoro omogenei.

4b. Gli stereotipi di genere

Gli stereotipi di genere, sono bias generati e sviluppati dalla società, diretta conseguenza della non eterogenea rappresentanza nel mondo del lavoro. Hanno una duplice natura:
- descrittiva, ovvero specificano come uomini e donne sono e come si comportano;
- prescrittiva, specificano come uomini e donne dovrebbero essere e come dovrebbero comportarsi.
Essi sono estremamente pericolosi perchè perpetuano le differenze di genere grazie alla teoria della giustificazione del sistema che porterebbe gli individui a difendere e giustificare la

situazione attuale per garantire un bisogno di stabilità (Jost & Banaji 1994); secondo questa teoria, anche gli individui del gruppo discriminato legittimano le ingiustizie perpetuate contro di essi.

A ciò, si aggiunge la minaccia dello stereotipo: quando ci viene fatto notare uno stereotipo dal quale dovremmo essere condizionati, tendiamo inconsciamente a confermarlo con le nostre azioni. La minaccia degli stereotipi, o la convinzione di poter essere il bersaglio di stereotipi umilianti, porta ad interruzioni delle prestazioni in una varietà di ambiti. Questo fenomeno è stato dimostrato in vari esperimenti nei quali sono stati confrontati un gruppo condizionato da uno stereotipo ed un gruppo di controllo. Due esperimenti condotti in un simulatore di guida dimostrano che la minaccia dello stereotipo disturba anche il controllo di un'automobile. Le donne a cui era stato ricordato lo stereotipo secondo cui le donne sono pessime guidatrici avevano più del doppio delle probabilità di scontrarsi con pedoni che camminavano fuori strada rispetto alle donne a cui non era stato ricordato questo stereotipo. [19]

4c. Bias nei dati a causa dell'algoritmo

Per quanto riguarda questi bias, si possono definire 2 sottocategorie:

- bias del programmatore (conscio o inconscio) che potrebbe riflettersi nell'algoritmo;
- bias di genere nel linguaggio di tutti i giorni.

Il bias del programmatore può essere conscio, nel qual caso, i programmatori sono ben consapevoli della disparità di rappresentanza di sessi sia nel team di sviluppo che nei dati di training. Si tratta quindi di una situazione simile alla mancanza di eterogeneità nel personale e nei dati, tuttavia rispetto a questa c'è la consapevolezza e quindi la noncuranza nello sviluppare una IA "fair by design".

Il bias del programmatore può essere inconscio invece quando il programmatore ha l'intenzione di sviluppare una IA fair, non riuscendoci tuttavia. L'esperimento di Lambrecht & Tucker sul campo, ci dà una testimonianza di ciò: un algoritmo ha pubblicato annunci che promuovono opportunità di lavoro nei campi della scienza, della tecnologia, dell'ingegneria e della matematica (STEM). L'annuncio era concepito per essere neutro rispetto al genere nella sua pubblicazione, tuttavia lo hanno visto meno donne rispetto agli uomini. Questo perché le donne più giovani sono un gruppo demografico ambito dai distributori di pubblicità e ad esse è più costoso mostrare annunci. Un algoritmo che ottimizza semplicemente il rapporto costo-efficacia nella pubblicazione degli annunci, pubblicherà invece annunci che non sono neutri rispetto al genere. Questa regolarità empirica vale anche per altre importanti piattaforme digitali.

In definitiva, questo algoritmo, per minimizzare i costi di diffusione, ha rinunciato alla fairness di cui era dotato, autogenerando un bias di genere. [20]

Esiste anche un bias di genere legato all'influenza del linguaggio corrente sul word embedding. Il word embedding è una tecnica di elaborazione del linguaggio naturale (Natural Language Processing o NLP) utilizzata per rappresentare le parole in uno spazio vettoriale. Questa rappresentazione permette di associare parole con significato simile in punti vicini tra di loro. Ad esempio, parole molto affini come "re" e "regina", sono molto vicine nello spazio vettoriale. Il word embedding permette di compiere operazioni aritmetiche e logiche, usando l'esempio di prima "re" - "uomo" + "donna" = "regina". Questa proprietà però è un'arma a doppio taglio: infatti i bias già presenti nella nostra società (principalmente i pregiudizi) sono assorbiti da questo modello. È stato analizzato che delle 1000 parole più utilizzate del vocabolario inglese, ben il 77% risulta essere associate maggiormente all'uomo piuttosto che alla donna, portando in evidenza la tendenza alla mascolinità del linguaggio parlato. Di queste parole quelle associate all'uomo tendono ad essere i verbi (ad esempio combattere, dominare…) mentre alla donna sono maggiormente associati aggettivi ed avverbi (ad esempio generoso, emotivamente…). Sempre considerando le 1000 parole più frequenti è importante sottolineare che il genere maschile contiene i domini delle parole legate all'ingegneria, alla religione, allo sport e alla violenza; mentre al genere femminile sono associati i domini di contenuti sessuali, insulti di genere, aspetto fisico, e termini legati alla cucina.
E' stata infine testata su 20.000 parole tramite il coefficiente di correlazione per ranghi (ρ) di Spearman la correlazione tra 3 valori: il valore emotivo, la stimolazione e la dominanza.
Per valore emotivo si intende la percezione di una parola come positiva o negativa,
con stimolazione si intende quanto la parola suscita eccitazione/stimolazione, infine con dominanza si intende quanto una parola rappresenta controllo/autorità. I risultati attestano che le parole più associate al genere maschile tendono ad avere valori di correlazione più alta nei campi della stimolazione e della dominanza; invece le parole associate al genere femminile hanno un coefficiente di correlazione maggiore nell'attributo del valore emotivo. [21]

4d. Fattori economici

Anche i fattori economici possono portare a dei bias di genere. In questo caso però non si tratta di un bias di rappresentanza bensì di allocazione. [22]
In precedenza abbiamo trattato esempi di bias in cui le donne erano rappresentate in minor numero rispetto agli uomini (bias di rappresentanza); nel bias di allocazione invece assistiamo

al differente trattamento da parte dell'algoritmo nei confronti del paziente per via di alcuni parametri economici. Se teniamo in considerazione la disparità salariale ancora esistente tra uomo e donna, non è irragionevole pensare che in sistemi sanitari come quello americano, nei quali prevale la gestione "privata" data dalla personale copertura assicurativa, gli algoritmi sviluppati portino a diverso trattamento degli individui accentuando tra le altre le disparità tra maschi e femmine.

4e. Comportamenti e decisioni "biased"

L'ultima macrocategoria (comportamenti e decisioni biased) include i comportamenti discriminatori dell'utente e le discriminazioni storiche. I primi rappresentano un problema se si ha a che fare con sistemi che adottano il continuous machine learning, ovvero algoritmi il cui training non viene tecnicamente mai terminato poiché continuano ad imparare tramite le interazioni con gli utenti. Il rischio è dato dal fatto che, se l'utente stesso risulta essere biased, l'algoritmo, che da esso attinge e impara, diventerà biased a sua volta.
Le discriminazioni storiche sono invece l'insieme di tutti i dati biased che nel tempo non sono stati "corretti" e quindi sono giunti a noi mantenendo i propri bias. La loro presenza in medicina è importante tenuto conto che la medicina di genere è un argomento recente e molti concetti non sono ancora stati verificati in entrambi i sessi.
In definitiva, da dati biased si creano algoritmi biased; se nel training dell'IA si utilizzano dati contenenti "errori", gli algoritmi che sottendono tratteranno questi "errori" come se fossero dati "corretti"; ne consegue che l'IA replicherà gli "errori" presenti nel data set.

5. ANALISI DEI POSSIBILI METODI DI DEBIASING

Il Bias di genere, come abbiamo visto, è un problema che da millenni la nostra società si porta appresso. La parità di genere purtroppo è un obiettivo assai lontano e secondo l'ONU ci sono vari traguardi da raggiungere prima di ottenere l'effettiva uguaglianza.
Uno di questi è l'uguale rappresentanza di maschi e femmine nei parlamenti nazionali; divario che in Italia si è tentato di colmare attraverso le quote rosa. Tuttavia, la rappresentanza femminile nei partiti (nella Camera dei deputati) non supera mai il 50% e le donne presenti alla Camera sono solo il 32% dei deputati; ciò conferma la distanza dalla parità di genere. [23]

Un altro traguardo è l'uguale rappresentanza nelle posizioni di potere e di dirigenza sul posto di lavoro (stimato in 140 anni); ciò equivale ad eliminare il cosiddetto "glass ceiling", metafora che indica l'impossibilità per certe categorie di persone di avanzare nella propria carriera a causa di discriminazioni di natura razziale o sessuale.
Il traguardo per l'eliminazione di tutte le leggi discriminatorie e l'introduzione di nuove leggi fair viene stimato in 286 anni. [24]
Questi numeri ci fanno comprendere la gravità della situazione, l'impossibilità a breve termine di eliminare il problema alla radice e la difficoltà nell'ottenere un IA "fair" senza il bisogno di mitigazioni "artificiali".
Le mitigazioni contro il bias di genere possono essere inquadrate in 3 macrogruppi:

- Assicurare la diversità nello sviluppo delle IA.
- Riduzione del bias nell'algoritmo;
- Design fair ed etico e sua implementazione nelle applicazioni IA; [17]

5a Assicurare la diversità nello sviluppo IA

Partendo dal primo macrogruppo, vi sono diverse possibilità per assicurare la diversità nello sviluppo IA.
5a-1 Una di queste consiste nella creazione di team interdisciplinari di analisti dei dati e analisti della società (social scientists) in modo tale da poter meglio inquadrare la popolazione da rappresentare nel training set dei dati. Pensando dunque di sviluppare un nuovo algoritmo, i data scientists si dovrebbero occupare della raccolta dei dati e dello sviluppo dell'IA mentre i social scientists dovrebbero utilizzare le conoscenze per valutare se i dati raccolti presentano vari tipi di bias. La cooperazione tra i due team può essere anche utile nel definire e valutare delle metriche di "fairness" che l'IA dovrebbe rispettare. Un'altra possibilità di cooperazione

tra i due gruppi è legata alla valutazione delle performance etiche a training terminato ovvero raccogliendo gruppi di input da fornire all'algoritmo differenziati da background etnico, di età, di sesso e di fattori demografici. In tal modo si possono studiare le differenze nella performance dell'algoritmo al variare di parametri sensibili e valutare dunque se l'IA sviluppata è fair. [25]

5a-2 Un altro metodo per assicurarsi la diversità nello sviluppo delle IA è l'adozione di un linguaggio neutro. L'eventuale adozione di un linguaggio neutro rappresenta un notevole passo in avanti nel combattere gli stereotipi di genere soprattutto per quanto riguarda i sistemi di NLP (Natural Language Processing). Eliminando la distinzione tra i sessi nei data point, vengono meno le associazioni vettoriali che tendono ad associare determinate parole ad un gender più che all'altro. La neutralizzazione del linguaggio può avvenire anche senza utilizzare parole appartenenti al linguaggio neutro: basta infatti creare per ogni frase "sessualizzata", la sua controparte in cui si allude al sesso opposto. Questa tecnica è stata testata su un algoritmo che si basava su un dataset di post provenienti da Twitter (ora "X"): come primo passo è stata creata una tabella in cui veniva associata, ad ogni parola sessualizzata, la corrispondente del sesso opposto (ad esempio padre/madre, figlio/figlia). Basandosi su di essa si è poi andati a ricercare nel dataset ogni parola che fosse stata rinvenuta all'interno della tabella. Infine, per ogni corrispondenza è stata creata una nuova frase (o "dato") in cui viene sostituita la parola "sessualizzata" iniziale con il suo opposto. Un procedimento di vera neutralizzazione è stato eseguito con i nomi propri: anzichè sostituirli con il corrispettivo, si è deciso di rimpiazzarli con la parola neutra "NAME" (in italiano "nome"). Riaddestrando l'algoritmo con questi accorgimenti si è verificata una riduzione del bias del 29.41% con un'accuracy loss del 2.47%. Queste percentuali sono state calcolate sottoponendo l'IA al test set "Equity Evaluation Corpus (EEC)", set selezionato appositamente per far emergere bias razzisti e di genere dagli algoritmi. [26][27]

5a-3 Un ultimo metodo per assicurarsi la diversità è quello di aumentare il numero di donne nei vari processi di sviluppo AI e ricerca. Le donne rappresentano una netta minoranza nelle materie STEM: solo il 31% degli studenti STEM risulta essere donna (la percentuale include le persone non binary), percentuale che crolla al 23% nelle carriere informatiche e addirittura al 21% nei corsi di ingegneria [28]. Queste percentuali vanno a riflettersi negli ambiti lavorativi e di ricerca: come riportato in precedenza, le donne sono rappresentate al 16% nel settore AI (il che segnala una perdita del 7/5% rispetto agli studenti di informatica/ingegneria). Anche la ricerca soffre di questa disparità: in uno studio dove sono state analizzate 5092 pubblicazioni da Novembre 2022 in poi, risulta che solo il 24.3% porta la firma di donne; percentuale che non cambia se si considerano gli articoli con più autori. Se si analizzano le pubblicazioni in ambito di informatica/informazione e scienze biomediche/cliniche il rapporto tra maschi e femmine è

di circa 3:1. Infine, per quanto riguarda le pubblicazioni su riviste mediche, il gap di genere aumenta ancora di più con percentuali di rappresentanza femminile sempre inferiore al 20% . Fa eccezione la rivista Nature che mostra un equilibrio di genere (45%) dovuto anche alla scelta di non pubblicare articoli non curanti della prospettiva di genere. [29]

L'incentivo ad aumentare la presenza delle donne nelle carriere STEM non è però sufficiente se non accompagnato dalla rottura del citato glass ceiling. Infatti, benchè nell'ambito sanitario le donne rappresentino la maggioranza (con un 67% di presenza), solamente il 25% dei ruoli di leadership sono affidati a donne. [30]

Gli ostacoli da abbattere per aumentare la presenza femminile nelle carriere STEM sono dati da:
- Le discriminazioni per le quali molte ragazze vengono scoraggiate dal proseguire una carriera in AI poiché ritenuta materia per uomini;
- Le barriere personali costituite da mancanza di autostima e dalla paura di dover lavorare in un ambiente a forte maggioranza maschile;
- Le persone di supporto, ovvero la mancanza di mentori/icone/idoli a cui aspirare, l'assenza di supporto emotivo da parte di genitori ed amici, la mancanza di voci maschili che chiedano l'uguaglianza di genere;
- L'educazione, dalla scelta dei giocattoli e dei media a cui i bambini sono sottoposti nella prima infanzia, alla mancanza di corsi introduttivi all'Intelligenza Artificiale nelle scuole dell'obbligo; c'è il bisogno di rendere questa "tecnologia" una materia d'obbligo nelle scuole. [31]

5b Riduzione del bias a livello dell'algoritmo stesso.

La rimozione della diversità nello sviluppo delle IA non è l'unico modo in cui è possibile effettuare un debias. Un'altra possibilità è data dalla riduzione del bias a livello dell'algoritmo stesso.
La semplice rimozione dei dati considerati "biased" dal training set, potrebbe considerarsi già sufficiente. Tuttavia, in ambito medico, ciò non rappresenta necessariamente la scelta migliore; si ricordano infatti due fattori importanti nello sviluppo delle IA: l'ingente quantità di dati necessaria ad allenare l'algoritmo e il fatto che la medicina di genere sia argomento relativamente recente. Andare dunque a rimuovere manualmente ogni dato biased presente nel training set potrebbe rivelarsi un'impresa assai lunga e complicata. Se si considera anche la

carenza di dati relativi a individui di sesso femminili in ambito medico, quest'opzione, pur essendo fair, potrebbe non portare ai risultati desiderati.
Per sopperire a questo problema, si può ricorrere alla data augmentation come fatto da (Zhao et al. 2019)[32]. Questa tecnica è molto simile a quella vista per la neutralizzazione del linguaggio tuttavia, in questo caso, non si crea solamente una copia del dato in cui si allude al sesso opposto, ma si rimuove pure il dato originale, creando così un bias opposto. Questa tecnica ha portato a grandi risultati e ha infatti eliminato il bias in algoritmi, che all'epoca rappresentavano lo stato dell'arte.
Un altro possibile metodo è quello di utilizzare le **reti GAN** (Generative Adversarial Network), ovvero una classe di modelli di apprendimento, ideata da Ian Goodfellow, in cui due reti neurali, una generativa e una discriminatoria competono tra di loro. La rete discriminatoria deve riuscire a capire se l'input fornitole è proveniente da un dataset reale o se è stato creato dalla rete generativa. Questo modello di algoritmi è stato proposto per mitigare il bias di genere: il compito della rete generativa è infatti quello di non far identificare, alla rete discriminatoria, un sesso a cui potrebbe essere associato il suo input. [33][34]
Se vogliamo evitare un'amplificazione dei bias, possiamo ricorrere ad un modello condizionale vincolato (**CCM**: Constrained conditional model) per cui si applica un vincolo che l'algoritmo deve rispettare, nel nostro caso non accentuare le percentuali di rappresentanza di un genere rispetto ad un altro. Dal momento che le donne sono sottorappresentate in certi contesti sociali, l'IA potrebbe mitigare questo dato associando percentuali maggiori rispetto a quelle reali a persone di genere maschile/femminili in determinati contesti. [33]

5c Design fair ed etico

Nel design fair ed etico si trattano gli aspetti che generalmente sono raccomandati agli sviluppatori di IA.
Primo fra tutti e anche il più importante, è l'explainability: con questo termine si intende la capacità di riuscire a spiegare il processo di formulazione dei risultati dell'IA. Purtroppo, l'explainability è ancora un obiettivo molto lontano e difficile da raggiungere, questo perché l'IA è un sistema che opera a "Black Box", il che vuol dire che non si conoscono le strade percorse dall'algoritmo per raggiungere il suo risultato. La difficoltà di comprensione dei meccanismi che stanno alla base dei processi elaborativi, è uno degli ostacoli più grandi che si pongono tra l'IA e le sue applicazioni in ambiti lavorativi dove la trasparenza è un fattore determinante. Oltre a quello sanitario, si ricordano il banking, i servizi pubblici e la pubblica sicurezza.

Gli investimenti in ricerca e sviluppo dovrebbero essere orientati all'acquisizione di strumenti tecnici quali algoritmi, database, IA al fine di soddisfare i requisiti etici e di fairness. Vale la pena citare un database recentemente creato nell'Università di Padova da S. Badaloni, A. Rodà e M. Scagnet. In questo database, sono stati raccolti vari articoli provenienti da riviste destinate ad un pubblico maschile, ad un pubblico femminile e articoli neutri provenienti dal sito dell'Università degli Studi di Padova. Questi articoli sono stati valutati da 107 partecipanti i quali hanno tentato di indovinare, tramite un punteggio in una scala da -2 (totalmente femminile) a +2 (totalmente maschile), a quale audience fosse indirizzato l'articolo. Questo database etichettato può essere utilizzato per individuare e valutare possibili stereotipi di genere presenti nei testi in lingua italiana. [35]

Per combattere il bias di genere tramite design fair ed etico è importante la consapevolezza della sua esistenza. È importante promuovere iniziative quali workshop, corsi di aggiornamento, etc. che possono aumentare la consapevolezza nella società, e in particolare negli sviluppatori e fruitori di sistemi di IA, dell'esistenza e gravità delle disparità di genere. Un esempio di queste iniziative è il corso "Saperi di genere ed Etica nell'Intelligenza Artificiale" che dall' anno accademico 2021/22 si tiene presso l'Università degli Studi di Padova. In questo corso, tra le altre cose, vengono trattate e approfondite le questioni di uguaglianza e conoscenze di genere, vengono analizzati i concetti di sesso e genere, le statistiche del mondo accademico e sottolineata l'importanza della necessità di un approccio di genere nel mondo dell'innovazione e dell'IA. Tutto ciò al fine di contrastare pregiudizi e stereotipi e giungere ad uno stato di equità nella società (https://ceur-ws.org/Vol-3319/paper12.pdf).

6. CONCLUSIONI

In medicina, come in molti altri settori della nostra società, i bias di genere sono ancora presenti e l'ottenimento della parità di genere è un traguardo ancora molto lontano.

L'Intelligenza Artificiale, le cui potenzialità di sviluppo nel settore sanitario sono enormi, contiene purtroppo i germi delle diseguaglianze sociali presenti (di genere, razziali, economiche…) e potrebbe prestarsi quale loro potenziale amplificatore.

Per ovviare a questo pericolo, dovremmo da un lato accelerare i processi di sviluppo della medicina di genere e dall'altro implementare linee guida etiche nelle macchine che ospitano forme di Intelligenza Artificiale. La codifica dell'etica in termini di algoritmo, ci costringe ad analizzare con chiarezza le nostre finalità e le nostre scelte.

Di sicuro, dovremmo impegnarci di più nel mitigare i vari bias di genere che conosciamo.

Le tecniche di de-biasing che sono state analizzate nel capitolo precedente potrebbero non essere sufficienti se non accompagnate da un'effettiva crescita sociale e superamento delle diseguaglianze sopra citate.

Lo sviluppo della medicina di genere impone la ricerca di soluzioni nuove e del supporto dell'IA nella revisione dei dati, nella loro modifica, fino alle analisi sull'algoritmo.

Nell'ambito dell'IA, è fondamentale garantire l'eterogeneità degli operatori coinvolti (sviluppatori, utilizzatori) e un adeguata formazione del personale medico.

La crescente sensibilità nei confronti di questi temi (manifestazioni, eventi, incontri, ricerche e pubblicazioni) lascia ben sperare sulla riduzione delle discriminazioni tra uomo e donna.

7. BIBLIOGRAFIA

1. Christopher Thomas & Antonio Pontón-Núñez, Automating Judicial Discretion: How Algorithmic Risk Assessments in Pretrial Adjudications Violate Equal Protection Rights on the Basis of Race, 40(2) Law & Ineq. 371 (2022), DOI: https://doi.org/10.24926/25730037.649.

2. Lombardi Vallauri, L. (2022) "Algoretica e Informatica giuridica", i-lex. Bologna, Italy, 15(1), pagg. 29–35. doi: 10.6092/issn.1825-1927/15586.

3. Badaloni Silvana & Francesca A. Lisi., 2020, "Towards a Gendered Innovation in AI (short paper).", pubblicato in DP@AI*IA 2020. https://ceur-ws.org/Vol-3319/invited1.pdf

4. IRCCS Ospedale San Raffaele, 2023, "Medicina di genere e salute femminile: cure e prevenzione migliori per le donne", https://www.hsr.it/news/2023/marzo/medicina-genere-cure-personalizzate-donne#:~:text=La%20medicina%20di%20genere%20%C3%A8,generi%20nella%20suscettibilit%C3%A0%20alle%20malattie.

5. Becher E, Oertelt-Prigione S. "The Impact of Sex and Gender in Medicine and Pharmacology", Handb Exp Pharmacol. 2023;282:3-23. doi:10.1007/164_2023_688

6. Europa Donna, "Medicina di genere e gender bias: il diritto alla salute e le donne", 2022, https://www.europadonna.it/2022/04/14/gender-bias-e-medicina-di-genere/

7. Gendered Innovations, "Osteoporosis Research in Men: Rethinking Standards and Reference Models", Stanford, https://genderedinnovations.stanford.edu/case-studies/osteoporosis.html#

8. Rembrand Koning et al. "Who do we invent for? Patents by women focus more on women's health, but few women get to invent." Science 372,1345-1348(2021). DOI:10.1126/science.aba6990

9. Gendered Innovations, "Medical Technology: Intersectional Approaches", Stanford, https://genderedinnovations.stanford.edu/case-studies/medtech.html#tabs-2

10. Wójcik S, Rulkiewicz A, Pruszczyk P, Lisik W, Poboży M, Domienik-Karłowicz J. "Beyond ChatGPT: What does GPT-4 add to healthcare? The dawn of a new era." Cardiol J. 2023;30(6):1018-1025. doi:10.5603/cj.97515

11. World Health Organization, "Meet S.A.R.A.H. A Smart AI Resource Assistant for Health", WHO, https://www.who.int/campaigns/s-a-r-a-h

12. Google Research, "Med-PaLM", Google, https://sites.research.google/med-palm/

13. Wang L, Zhang Y, Wang D, et al. "Artificial Intelligence for COVID-19: A Systematic Review." Front Med (Lausanne). 2021;8:704256. Published 2021 Sep 30. doi:10.3389/fmed.2021.704256

14. U.S. Food&Drug Administration, "Artificial Intelligence and Machine Learning (AI/ML)-Enabled Medical Devices", FDA, 2024, https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-and-machine-learning-aiml-enabled-medical-devices#:~:text=With%20this%20update%2C%20the%20FDA,used%20to%20generate%20this%20list

15. European Parliament, "Artificial Intelligence Act", Europuan Parliament, 2024, https://www.europarl.europa.eu/doceo/document/TA-9-2024-0138_EN.pdf

16. Gilbert, S. "The EU passes the AI Act and its implications for digital medicine are unclear." npj Digit. Med. 7, 135 (2024). https://doi.org/10.1038/s41746-024-01116-6

17. Nadeem, A., Abedin, B., & Marjanovic, O. (2020). "Gender bias in AI: a review of contributing factors and mitigating strategies." In ACIS 2020 Proceedings (pp. 1-12). [27] AIS Electronic Library (AISeL). https://aisel.aisnet.org/acis2020/27

18. Arianna Meroni & Camilla Zan, "Women in AI: female contribution towards an inclusive future", Skilla, 2024, https://www.skilla.com/en/blog/women-in-ai-female-contribution-towards-an-inclusive-future/

19. Nai Chi Jonathan Yeung, Courtney von Hippel, "Stereotype threat increases the likelihood that female drivers in a simulator run over jaywalkers" Accident Analysis & Prevention, Volume 40, Issue 2, 2008, Pages 667-674, ISSN 0001-4575, https://doi.org/10.1016/j.aap.2007.09.003.

20. Anja Lambrecht, Catherine Tucker (2019) “Algorithmic Bias? An Empirical Study of Apparent Gender-Based Discrimination in the Display of STEM Career Ads.” Management Science 65(7):2966-2981. https://doi.org/10.1287/mnsc.2018.3093

21. Aylin Caliskan, Pimparkar Parth Ajay, Tessa Charlesworth, Robert Wolfe, and Mahzarin R. Banaji. 2022. “Gender Bias in Word Embeddings: A Comprehensive Analysis of Frequency, Syntax, and Semantics”. In Proceedings of the 2022 AAAI/ACM Conference on AI, Ethics, and Society (AIES '22). Association for Computing Machinery, New York, NY, USA, 156–170. https://doi.org/10.1145/3514094.3534162

22. Sun, T., Gaut, A., Tang, S., Huang, Y., Elsherief, M., Zhao, J., Mirza, D., Belding-Royer, E.M., Chang, K., & Wang, W.Y. (2019). Mitigating Gender Bias in Natural Language Processing: Literature Review. Annual Meeting of the Association for Computational Linguistics. https://doi.org/10.48550/arXiv.1906.08976

23. Camera dei Deputati, “Statistiche Parlamentari, Dati e cifre dell’attività parlamentare”, Parlamento Italiano, https://www.camera.it/leg19/1422?idStat=10005

24. Sustainable Development Goals, “Goal 5: Achieve gender equality and empower all women and girls”, United Nations, https://www.un.org/sustainabledevelopment/gender-equality/#:~:text=At%20the%20current%20rate%2C%20it,achieve%20equal%20representation%20in%20national

25. D. Lucas, P. Kaledio, “Bias Mitigation: Address and mitigate biases in AI models to ensure fairness and equity in healthcare outcomes”, 2024, https://www.researchgate.net/publication/378342531_Bias_Mitigation_Address_and_mitigate_biases_in_AI_models_to_ensure_fairness_and_equity_in_healthcare_outcomes

26. Svetlana Kiritchenko, Saif M Mohammad, “Examining gender and race bias in two hundred sentiment analysis systems”, 2018, https://doi.org/10.48550/arXiv.1805.04508

27. Sue, Carson, et al. "Fairness in Machine Learning: Detecting and Removing Gender Bias in Language Models." 2022, https://www.shahaesha.com/uploads/Fairness%20Gender%20Bias.pdf

28. STEMWomen, “Women in STEM Statistics: Progress and Challenges”, StemWomen, 2023, https://www.stemwomen.com/women-in-stem-statistics-progress-and-challenges

29. Prema Nedungadi, Maneesha Ramesh, Venu Govindaraju, Bhavani Rao, Paola Berbeglia, Raghu Raman, “Emerging leaders or persistent gaps? Generative AI research may foster women in STEM”, International Journal of Information Management, Volume 77, 2024, 102785, ISSN 0268-4012, https://doi.org/10.1016/j.ijinfomgt.2024.102785.

30. World Health Organization, “Value gender and equity in the global health workforce”, WHO, https://www.who.int/activities/value-gender-and-equity-in-the-global-health-workforce

31. M. Roopaei, J. Horst, E. Klaas, G. Foster, T. J. Salmon-Stephens and J. Grunow, "Women in AI: Barriers and Solutions," 2021 IEEE World AI IoT Congress (AIIoT), Seattle, WA, USA, 2021, pp. 0497-0503, doi: 10.1109/AIIoT52608.2021.9454202.

32. Jieyu Zhao, Tianlu Wang, Mark Yatskar, Ryan Cotterell, Vicente Ordonez, Kai-Wei Chang, “Gender Bias in Contextualized Word Embeddings”, 2019, https://doi.org/10.48550/arXiv.1904.03310

33. Sun et al., “Mitigating Gender Bias in Natural Language Processing: Literature Review”, ACL 2019, https://aclanthology.org/P19-1159

34. Brian Hu Zhang, Blake Lemoine, Margaret Mitchell, “Mitigating Unwanted Biases with Adversarial Learning”, 2018, https://doi.org/10.48550/arXiv.1801.07593

35. Silvana Badaloni, Antonio Rodà, Martino Scagnet, “An Italian dataset for the analysis of gender stereotypes in textual documents”, 2023, https://ceur-ws.org/Vol-3615/short1.pdf

36. Silvana Badaloni, Antonio Rodà, “Gender knowledge and Artificial Intelligence”, 2022, https://ceur-ws.org/Vol-3319/paper12.pdf