

Le tecnologie informatiche, soprattutto quelle di nuova generazione come Intelligenza Artificiale e blockchain, vengono considerati uno straordinario strumento per accelerare il raggiungimento di obiettivi di sviluppo sostenibile (vedi per esempio, Agenda 2030 per lo Sviluppo Sostenibile promossa dall'Onu), ma al contempo introducono nuove forme di esclusione e marginalizzazione. Una criticità evidenziata anche dall'Unesco in un recente rapporto su Uguaglianza di Genere ed Intelligenza Artificiale [\(1\)](#).

Secondo i criteri definiti per la cosiddetta "IA affidabile" (trustworthy AI), sanciti in un documento della Unione Europea pubblicato ad aprile 2019, questa scollatura non è accettabile, e va fortemente contrastata.

L'intelligenza artificiale (AI) viene sempre più impiegata per prendere decisioni che influenzano la maggior parte degli aspetti della nostra vita. L'uso dell'AI nelle previsioni e nel processo decisionale può ridurre la soggettività umana, ma può anche incorporare pregiudizi, producendo risultati discriminatori su larga scala e ponendo enormi rischi per la società, in particolare, nei sistemi che apprendono grazie ad algoritmi cosiddetti di machine learning. Per sfruttare il potenziale di trasformazione dell'AI è necessario affrontare il fenomeno per quello che è: un prodotto che nasce dall'elaborazione umana ed in quanto tale produce delle risposte in base agli input che vengono dati. Poiché la tecnologia è costruita dagli esseri umani, ovviamente imiterà i nostri pregiudizi. Per combattere il perpetuare degli stereotipi e dei pregiudizi è necessario esserne consapevoli ed affrontarli alle origini, perché al momento le narrazioni sui big data e la scienza dei dati sono prevalentemente bianche, maschili e tecno-eroiche.

Però, rimuovere i pregiudizi dagli algoritmi può essere un meccanismo potente per affrontare questi squilibri.

Molte istituzioni prendono decisioni basate su sistemi di intelligenza artificiale utilizzando l'apprendimento automatico (ML - Machine Learning), in base al quale una serie di algoritmi prende e impara da enormi quantità di dati per trovare schemi e fare previsioni. Questi sistemi, per esempio, informano quanto gli istituti finanziari di credito dovrebbero offrire ai diversi clienti, a chi il sistema sanitario deve dare la priorità per i vaccini COVID-19 e quali candidati le aziende devono chiamare per i colloqui di lavoro. Il pregiudizio di genere in questi sistemi è pervasivo e ha un impatto profondo sulla sicurezza psicologica, economica e sanitaria delle donne a breve e lungo termine, in quanto in grado di rafforzare e amplificare gli stereotipi e i pregiudizi di genere dannosi esistenti. Per comprenderci, se non inseriamo dati sulla medicina di genere, che sono già sottorappresentati, la macchina non potrà certo imparare a migliorarne la valutazione.

Se poi le raccomandazioni prodotte da questi algoritmi sono successivamente utilizzate come supporto ai processi decisionali, assistiamo all'instaurarsi di un vero e proprio circolo vizioso. I pregiudizi si rafforzano sempre più, e con essi la discriminazione di genere in molti ambiti applicativi. Relativamente alla tematica della "gendered innovation" (innovazione di genere), il non tener conto della dimensione di genere già in fase di ricerca e sviluppo produce tecnologie inaffidabili perché poco rispondenti ai bisogni e alle aspettative di ampie fette della popolazione mondiale.

Ma, come detto sopra, sono esseri umani che generano, raccolgono ed etichettano le informazioni che entrano nei set di dati-input. Gli esseri umani determinano da quali set di dati, variabili e regole gli algoritmi apprendono per fare previsioni. Entrambe queste fasi possono introdurre pregiudizi che vengono incorporati nei sistemi di intelligenza artificiale riproducendo quelli di partenza.

Prendendo ad esempio il settore del credito al consumo, i primi processi utilizzavano lo stato civile e il sesso per determinare l'affidabilità creditizia. Alla fine, queste pratiche discriminatorie sono state sostituite da quelle considerate più neutre. Ma a quel punto, le donne avevano una storia finanziaria meno formale e soffrivano di discriminazioni, influenzando la loro capacità di ottenere credito. I dati che tracciano i limiti di credito degli individui catturano queste

tendenze discriminatorie.

I dati sono istantanee del mondo in cui viviamo e le grandi distorsioni di genere che vediamo sono in parte dovute al divario digitale di genere e a come/dove queste informazioni vengono raccolte. Ad esempio, circa 300 milioni di donne in meno degli uomini accedono a Internet da un telefono cellulare e le donne nei paesi a reddito medio e basso hanno il 20% di probabilità in meno rispetto agli uomini di possedere uno smartphone. Queste tecnologie generano dati sui loro utenti, quindi il fatto che le donne abbiano minori possibilità di accesso distorce intrinsecamente le informazioni di base.

Ma, anche quando i dati vengono generati, gli esseri umani che li raccolgono decidono cosa enfatizzare e come. Chi prende le decisioni che informano i sistemi di intelligenza artificiale e chi fa parte del team che li sviluppa tecnicamente determina il loro agire: la bassa presenza di donne incide sul risultato complessivo. Anche qui c'è un enorme divario di genere: solo il 22% dei professionisti nei settori dell'AI e della scienza dei dati sono donne e occupano mediamente lavori associati a uno status inferiore.

E l'AI può avere un impatto negativo/positivo sul percorso professionale di una persona, in una sorta di perverso loop. Due esempi possono aiutare ad inquadrare il tema.

Una piattaforma di assunzione di tecnologia online, Gild, ha sviluppato un sistema di intelligenza artificiale per aiutare i datori di lavoro a classificare le/i candidate/i per i lavori di programmazione. Gild non solo ha usato le informazioni raccolte da fonti tradizionali come i curriculum, ma ha anche utilizzato un proxy chiamato "dati sociali" (dati generati da azioni nel regno digitale) per misurare quanto la/il candidata/o fosse parte integrante della comunità digitale. In questo caso, i dati social sono stati ricavati dal tempo trascorso a condividere e sviluppare codice su piattaforme online. Ma come sappiamo troppo bene, fattori come le aspettative della società riguardo al lavoro di cura, a carico prevalentemente delle donne, si traducono in donne che hanno meno tempo per chattare online con la conseguenza che proprio noi produciamo meno di questi "dati sociali". Invece di rimuovere i pregiudizi umani, Gild ha creato un algoritmo predisposto a penalizzare le donne e classificare sistematicamente le candidate al di sotto delle controparti maschili.

L'università di Bath, per andare ad un altro esempio, ha riportato che gli algoritmi che si utilizzano per interpretare il linguaggio hanno acquisito molti dei pregiudizi verificati nell'uomo dagli esperimenti di psicologia. Le parole «femmina» e «donna», per esempio vengono associate maggiormente a occupazioni «umanistiche» e legate alla cura della casa. Se a questi programmi viene chiesto di selezionare dei curriculum per un lavoro, è più probabile che scelgano quelli di maschi per un posto da ingegnere.

Uno studio del Berkeley Haas Center for Equity, Gender and Leadership sulla mitigazione dei pregiudizi nell'intelligenza artificiale, ha monitorato circa 133 sistemi di AI in tutti i settori dal 1988 ad oggi, scoprendo che il 44,2% (59 sistemi) dimostra pregiudizio di genere, con il 25,7% (34 sistemi) che mostra sia pregiudizi di genere che razziali.

E ancora: il rafforzamento degli stereotipi e dei pregiudizi esistenti e dannosi (nel 28,2% dei sistemi con pregiudizi di genere) è esacerbato dai cicli di feedback tra input e output dei dati. Ad esempio, il software di traduzione, che apprende da grandi quantità di testo online, ha storicamente utilizzato termini neutri rispetto al genere (come "il dottore" o "l'infermiera" in inglese) e ha restituito traduzioni di genere (come "el doctor" e "la enfermera", in spagnolo), rafforzando gli stereotipi dei medici uomini e delle infermiere donne.

Basti ricordare l'analisi fatta dalla Dott.ssa Irene Biemmi sui libri di testo delle nostre scuole elementari, per avere ancora più chiara l'origine della discriminazione anche in forma digitale (2).

Le discriminazioni più accentuate sono quelle sulle donne, soprattutto se di colore, l'omofobia, il razzismo e il classismo.

Troppo spesso le macchine imparano dai loro sviluppatori ad identificare il soggetto "standard" delle nostre società con l'uomo, soprattutto bianco e privilegiato.

Ma ribadisco: se ben programmato, un sistema può diventare oggettivo e antirazzista.

Per esempio, gli studi effettuati sul comportamento umano da alcune Università statunitensi hanno dimostrato che, a parità di skills, è molto più facile che una/un cittadina/o europea/o venga assunta/o rispetto ad una/o afro-discendente, con una maggioranza netta di Europei.

Togliendo la qualifica etnica, l'algoritmo liberato da ogni possibile riferimento stereotipato, ha selezionato le persone migliori per il profilo professionale, ribaltando la percentuale tra Europei (a questo punto in minoranza) e afro-discendenti.

Un potere paradossale, quello dell'AI, che rischia di cancellare importanti conquiste sociali e non stare al passo con il progresso.

Abbiamo compreso perfettamente il rischio, ma anche il potenziale e quindi è necessario proporre diversità di approccio e di sistema.

L'intelligenza artificiale viene sempre più utilizzata per affrontare le sfide dello sviluppo globale, inclusa la disuguaglianza di genere, e le organizzazioni della società civile stanno entrando a far parte del progetto. Ad esempio, Women's World Banking e Mujer Financiera utilizzano il ML per sostenere l'inclusione finanziaria delle donne. È importante insistere e supportare gli sviluppatori ML per centrare le voci delle donne e degli individui non binari nello sviluppo, nella creazione e nella gestione di questi sistemi di intelligenza artificiale.

Ad esempio, Digital Democracy, un'organizzazione che lavora con le comunità emarginate per difendere i loro diritti attraverso la tecnologia, ha lavorato con gruppi delle comunità locale come la Commissione delle donne vittime di violenza (KOFIV) per costruire un sistema sicuro per la raccolta di dati sulla violenza di genere ad Haiti. Il sistema ha permesso alle donne locali di tracciare, analizzare, mappare e condividere i dati, capaci di fornire una visione oggettiva del tema. Un altro passo importante è riconoscere e lavorare contro pratiche dannose sui dati, come delineato nel Feminist Data Manifest-No (3).

È necessario incorporare e promuovere la diversità di genere, l'equità e l'inclusione tra i team che sviluppano e gestiscono i sistemi di intelligenza artificiale, per arginare una deriva pericolosa e per sfruttarne il potenziale per un mondo più giusto. Agire assicurandosi che la diversità sia una priorità fondamentale della leadership e aggiornare le politiche, le pratiche e le strutture istituzionali per sostenere la diversità e l'inclusione.

Per farlo bisogna riconoscere che i dati e gli algoritmi non sono neutri e stabilire approcci di governance sensibili al genere per un'AI responsabile immaginando anche strutture di governance etica.

Queste azioni non sono esaustive, ma forniscono un punto di partenza per la creazione di un ML intelligente di genere che promuove l'equità.

Facile? Assolutamente no: è una sfida titanica, ma che associa scienziate e donne di tutto il mondo.

In questa direzione va il Data Feminism, cioè le pratiche di dati femministe che includono l'analisi di come funziona il potere e l'uso dei dati per sfidare strutture di potere disuguali, superando il divario di genere, valorizzando molteplici forme di conoscenza e sintetizzando diverse prospettive. Un nuovo modo di pensare alla scienza dei dati e all'etica dei dati che è informato dalle idee del femminismo intersezionale (4).

La visione femminile sull'AI serve a garantire maggiormente la correttezza rispetto alla dimensione di genere non solo

degli algoritmi di Machine Learning, ma anche del metodo seguito dalla scienza e dalla tecnologia in ambito AI. Un incremento della presenza femminile nel settore, sia in ambito accademico che in quello aziendale, costituisce un arricchimento, in termini di competenze e capacità. Più in generale, la diversità è il vero motore del cambiamento. Tecnologie dirompenti come quelle basate su AI, destinate a mutare profondamente il mondo, non possono prescindere dalla salvaguardia della diversità per essere accettate e veramente utili.

Oggi, la scienza dei dati è una forma di potere. Combattere i pregiudizi dei dati riguarda molto più del genere. Riguarda il potere, chi lo ha e chi no, e come questi differenziali di potere possono essere sfidati e cambiati.

Il momento è questo, senza sconti per chi ci vuole incapaci di scienza.

*A cura di Anna Maria Romano*

---

(1) <https://en.unesco.org/artificial-intelligence/ethics>

(2) <https://www.ibs.it/educazione-sessista-stereotipi-di-genere-ebook-irene-biemmi/e/9788878855649>

(3) <https://www.manifestno.com/>

(4) <https://www.ibs.it/data-feminism-libro-inglese-catherine-d-ignazio-lauren-f-klein/e/9780262044004>