

«La discriminazione degli algoritmi non è solo una questione tecnica, ma è anche sociale e politica»

L'intelligenza artificiale rischia di minare i traguardi raggiunti sul piano dell'uguaglianza e dell'inclusione a livello globale. Responsabili sono gli algoritmi che amplificano le disparità contenute nei dati con cui sono allenati. Quali le contromisure della comunità internazionale e degli Stati per arginare questo problema? Ne parliamo con l'esperta **Ivana Bartoletti**, docente e ricercatrice, co-curatrice dello studio del Consiglio d'Europa sull'impatto dei sistemi di intelligenza artificiale in materia di parità e discriminazione.

Intervista: Isabella Visetti

Il Consiglio d'Europa ha adottato a metà maggio 2024 una Convenzione quadro sull'intelligenza artificiale e i diritti umani, la democrazia e lo Stato di diritto.¹ Come valuta questa convenzione?

In modo positivo. Sono identificati i problemi di rispetto dei diritti umani dovuti all'intelligenza artificiale (IA) e la necessità di avere più trasparenza e più controlli per garantire un suo uso responsabile in ogni fase: dalla progettazione, allo sviluppo, all'applicazione, al controllo dei sistemi che la utilizzano. È un documento che chiede agli Stati un impegno per individuare, valutare e attenuare i rischi di quei sistemi basati sull'IA che sono fonte di discriminazione e ledono il principio della parità di trattamento, la protezione della privacy e il diritto delle persone a essere informate quando interagiscono con servizi messi a punto con l'IA. Gli Stati dovranno inoltre garantire possibilità di ricorso alle vittime di discriminazione algoritmica e dovranno anche attivare strumenti di controllo indipendente per verificare la conformità alla convenzione. In oltre, dovranno sensibilizzare e informare il pubblico sulle tecnologie legate all'IA.

Per quello che riguarda gli aspetti che toccano la democrazia, gli Stati dovranno vigilare affinché l'IA non sia usata per danneggiare le istituzioni e i principi democratici di separazione dei poteri e di accesso alla giustizia. L'aspetto più critico, che è stato anche fonte di accese discussioni, ed è forse il limite di questa convenzione, riguarda la sua validità che tocca solo il settore pubblico e le sue aziende fornitrici di beni o servizi. Per quello privato, ogni Stato potrà decidere in modo autonomo, sce-

gliendo due opzioni: adottare la regolamentazione della convenzione in modo automatico anche per l'ambito privato o varare delle misure ad hoc per questo ambito per adeguarsi ai principi enunciati dalla convenzione.

Lei fa parte di un gruppo di persone esperte che ha l'obiettivo di elaborare delle raccomandazioni per gli Stati che sottoscriveranno la Convenzione, aperta alla firma il 5 settembre 2024 a Vilnius, in Lituania. Qual è lo scopo di queste raccomandazioni?

La Commissione per l'uguaglianza di genere (CEG) e il Comitato direttivo sull'antidiscriminazione, la diversità e l'inclusione (CDADI) del Consiglio d'Europa hanno incaricato un gruppo di persone esperte per elaborare le raccomandazioni da fornire al Comitato dei ministri, che è l'organo decisionale del Consiglio d'Europa ed è composto dai Ministri degli Affari esteri di tutti gli Stati membri. Queste raccomandazioni, ispirate allo studio sull'impatto dei sistemi di intelligenza artificiale in materia di parità e discriminazione² che ho redatto con Raphaële Xenidis per conto della CEG e del CDADI nel 2023, toccano differenti ambiti dove gli Stati possono agire per rendere operativa la Convenzione. C'è l'aspetto della prevenzione, della trasparenza e della responsabilità; quello importante dell'accesso alla giustizia; la questione della diversità delle comunità professionali, dell'inclusione e della promozione della parità e infine quello della partecipazione democratica e della sensibilizzazione dell'opinione pubblica. Nel processo di definizione di queste raccomandazioni si ritrova il modus



Ivana Bartoletti

«Si dovrebbe considerare ogni sistema basato sull'IA discriminatorio per definizione.»

«Siamo dunque di fronte a un problema sia tecnico, sia sociopolitico, la cui soluzione richiede una risposta che non sia solo tecnica.»

operandi del Consiglio d'Europa che prevede un dialogo con gli Stati membri, che a loro volta coinvolgono la società civile, i diversi gruppi d'interesse, oltre ai ricercatori e agli esperti sulla materia. Ovviamente il ruolo principale e decisivo spetta ai singoli Stati, ma si può dire che c'è un allineamento su quanto emerso nel nostro studio.

Quali sono i passi necessari per evitare che l'IA contrasti l'uguaglianza attraverso gli algoritmi?

È fondamentale una maggiore conoscenza sui bias algoritmici, cioè quegli errori sistematici di giudizio o di interpretazione, che portano a valutazioni sbagliate o poco oggettive. Questi bias emergono proprio dalla natura intrinseca dell'IA e dal suo funzionamento: l'IA usa dati che non sono neutri ma influenzati dal contesto e rispecchiano le disparità presenti nella società. Quando questi dati vengono immessi in una macchina per allenare un algoritmo a prendere delle decisioni o a formulare delle previsioni, si replica questa disparità e la si conferma con un software: è una softwarizzazione che codifica e perpetua le disuguaglianze esistenti nella realtà su una scala molto più grande. Un artefatto di intelligenza artificiale non è però solo frutto della tecnologia, perché i parametri con cui addestrare l'algoritmo sono scelti da persone. La domanda da porsi è semplice: chi decide di inserire determinati parametri in un sistema di IA? Chi compie questa scelta che ne esclude per forza altre? Questo processo decisionale, questa interazione fra la macchina e l'essere umano, può essere fonte di discriminazione ed è spesso difficile da individuare. Questa scelta è spesso operata da team di profes-

sionisti, che non hanno un background misto, non sono rappresentativi di tutti i gruppi sociali, tantomeno di quelli minoritari o svantaggiati. La mancanza di diversità e inclusione nei team di programmatori di sistemi IA è un fattore che gioca un ruolo importante. Siamo dunque di fronte a un problema sia tecnico, sia sociopolitico, la cui soluzione richiede una risposta che non sia solo tecnica, come spesso è accaduto in passato. Serve un approccio multidisciplinare, che inglobi le scienze sociali, il diritto, l'etica, le norme sulla non discriminazione, ma occorre anche la volontà politica, la volontà degli Stati, di volere da questi sistemi IA un risultato equo e non una fotocopia di una situazione diseguale.

Quali sono le difficoltà su questo cammino?

Quando si affronta questo tema, fa sempre capolino l'obiezione: ma se gli esseri umani sono portatori di un bias, che problema c'è se lo sono anche gli algoritmi? Il punto è che quando si inserisce la discriminazione in un software e la si automatizza, diventa poi complicato contrastarla. È anche difficile capire l'origine del bias, cioè identificare in quale tappa della pipeline – quella catena di elementi elaborati a cascata in un software – si produca. Inoltre, questi algoritmi agiscono all'interno di macchine, che vengono chiamate scatole nere, che imparano da sole e diventano incomprensibili anche ai professionisti del settore. Scatole nere verso le quali è complicato sollevare contestazioni da parte di chi si sente danneggiato, perché sono poco accessibili e protette dalla proprietà intellettuale.

Finora sono mancati gli strumenti legali e tecnici che possano aiutare a contrastare l'opacità di funzionamento e le discriminazioni dell'IA. A livello europeo, c'è ora la Legge sull'intelligenza artificiale (*Artificial Intelligence Act*) e il Regolamento generale per la protezione dei dati personali (in Svizzera la nuova Legge sulla protezione dei dati è entrata in vigore il 1° settembre del 2023), che pongono l'accento sulla *fairness*, cioè sull'equità e la correttezza delle decisioni prese da algoritmi. Questa *fairness* va costruita. Per esempio, se in una selezione del personale si vuole avere un esito equo dal punto di vista del genere, ma i dati a disposizione considerano poche donne, si possono introdurre dati sintetici, inventati, al fine di ottenere il risultato auspicato. Questo presuppone una chiara volontà di rendere il sistema di IA giusto e questa volontà non può che venire da una decisione da parte di chi produce questi sistemi. In questi anni, il Regolamento generale per la protezione dei dati personali è stato importante per appellarsi alla non trasparenza dell'algoritmo e dunque rifiutarsi di fornire il consenso per il trattamento dei dati personali. Al contrario, la normativa sulla discriminazione non è così efficace a contrastare il bias di correlazione prodotto dall'algoritmo, la cosiddetta *proxy discrimination*, che per esempio avviene quando si discrimina non solo perché un individuo è donna, ma anche perché naviga in determinati siti e perché si abita in un certo quartiere. Questa correlazione matematica che compie l'algoritmo, associando dati diversi, crea nuove categorie di discriminazione, che non sono quelle tradizionali (l'essere donna, l'essere disabile, l'appartenere a un gruppo religioso...), verso le quali la legge attuale non è così incisiva.

Dal punto di vista del quadro legale cosa servirebbe?

Uno degli elementi più forti che abbiamo proposto nel nostro studio è l'inversione dell'onere della prova. Si dovrebbe considerare ogni sistema basato sull'IA discriminatorio per definizione, a meno che i suoi sviluppatori non provino il contrario. Questo cambio di paradigma faciliterebbe il ricorso alla giustizia in caso di discriminazione e andrebbe a riequilibrare un'asimmetria di potere creata dai sistemi IA anche in termini di accesso alle informazioni e alla conoscenza sul loro funzionamento. Per

una persona normale è infatti un'impresa quasi impossibile raccogliere le prove di una discriminazione algoritmica. Questa presunzione di un bias discriminatorio si giustifica con l'alta frequenza di questo fenomeno e si dovrebbe applicare solo quando non ci sono delle misure preventive, come per esempio un'analisi di impatto, un audit che verifichi le procedure o una certificazione dell'algoritmo utilizzato. In questa architettura preventiva, dovrebbe essere presente un controllo pubblico da parte delle istituzioni di difesa dei diritti umani e degli enti contro la discriminazione, ma anche delle azioni positive per incentivare lo sviluppo di sistemi IA che soddisfino standard di qualità ed equità. Il quadro è ora facilitato dall'adozione dall'*AI Act* dell'Unione europea, che prevede tutta una serie di requisiti e di obblighi e fra questi anche una valutazione di conformità sull'equità degli algoritmi.

Lei ha fondato e dirige Women Leading in AI³: di cosa si occupa questa associazione?

È nata nel 2018 ed è una rete di donne attive in tutti gli ambiti dell'intelligenza artificiale e quindi non solo in quello della programmazione. Raggruppamento professioniste provenienti dal settore tecnico, programmatrici, informatiche, scienziate, ma anche giornaliste, politiche, avvocate, esperte di etica e di diritti umani, donne attive nelle istituzioni pubbliche e nelle imprese private, ricercatrici e docenti universitarie. L'idea è quella di allargare il discorso sulla governance dell'IA e sul suo sviluppo responsabile, sulla definizione di linee guida, sugli strumenti per renderla un motore di innovazione e non di discriminazione. Abbiamo messo a punto un documento con dieci principi e altrettante raccomandazioni sulla regolamentazione dell'IA, che abbiamo indirizzato ai governi; vogliamo mobilitare la politica, ma anche sensibilizzare l'opinione pubblica, affinché sia meglio informata su questo importante tema. Crediamo nella costruzione di un'intelligenza artificiale che sostenga i nostri obiettivi e sia ancorata ai nostri valori umani.

Isabella Visetti è una giornalista RSI, conduce il magazine radiofonico *Millevoci* su Rete Uno e da tempo si occupa di temi legati alla parità di genere.

«Occorre anche la volontà politica, la volontà degli Stati, di volere da questi sistemi IA un risultato equo e non una fotocopia di una situazione diseguale.»

«Gli Stati dovranno inoltre garantire possibilità di ricorso alle vittime di discriminazione algoritmica.»

Note

- 1 Consiglio d'Europa: Council of Europe Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law (Convezione quadro del Consiglio d'Europa sull'intelligenza artificiale e i diritti umani, la democrazia e lo Stato di diritto), 2024. <https://rm.coe.int/1680afae3c> (consultato il: 25.06.2024).
- 2 Bartoletti, Ivana / Xenidis, Raphaële: Étude sur l'impact des systèmes d'intelligence artificielle, leur potentiel de promotion de l'égalité, y compris l'égalité de genre, et les risques qu'ils peuvent entraîner en matière de non-discrimination. Conseil de l'Europe, 2023. <https://rm.coe.int/study-on-the-impact-of-artificial-intelligence-systems-their-potential/1680ac99e3> (consultato il: 25.06.2024).
- 3 Women Leading in AI: <https://womenleadinginai.org/> (consultato il: 25.06.2024).

Abstracts

«Diskriminierungen durch Algorithmen sind nicht nur eine technische, sondern auch eine soziale und politische Frage»

Künstliche Intelligenz kann Gleichberechtigung und Inklusion untergraben. Eine Massnahme dagegen ist das *Rahmenabkommen des Europarats über die Künstliche Intelligenz, Menschenrechte, Demokratie und Rechtsstaatlichkeit*. **Ivana Bartoletti** ist Mit-herausgeberin der Studie des Europarats zu den Auswirkungen von KI-Systemen auf Gleichstellung und Diskriminierung, die Empfehlungen für fairere Algorithmen und KI-Systeme enthält. Die Probleme sind sowohl technisch als auch soziopolitisch und um sie zu lösen, braucht es einen multidisziplinären Ansatz. Auf gesetzlicher Ebene müssten alle KI-Systeme standardmässig als diskriminierend eingestuft werden, solange sie nicht das Gegenteil beweisen – dies wäre eine Umkehrung der Beweislast, die Opfern den Zugang zur Justiz erleichtern würde.

« La discrimination algorithmique n'est pas qu'une question technique, mais aussi sociale et politique »

L'intelligence artificielle risque de mettre fin aux espoirs d'égalité et d'inclusion. Pour y faire face, le Conseil de l'Europe a adopté la *Convention cadre sur l'intelligence artificielle, les droits de l'homme, la démocratie et l'État de droit*. **Ivana Bartoletti** est co-responsable de l'étude du Conseil de l'Europe sur l'impact des systèmes d'intelligence artificielle en matière d'égalité et de discrimination, qui formule des recommandations pour rendre les algorithmes et ces systèmes plus équitables. Il s'agit certes d'un problème technique, mais aussi sociopolitique, et il faut adopter une approche multidisciplinaire pour le résoudre. Sur le plan juridique, tout système d'IA devrait être considéré par défaut, sauf preuve du contraire, comme discriminatoire : un renversement de la charge de la preuve qui faciliterait l'accès à la justice pour les victimes.