

PARITÀ DI GENERE

IMPRESA, INNOVAZIONE ED INFRASTRUTTURE

RIDURRE LE DISUGUAGLIANZE

Intelligenza artificiale e responsabilità sociale

Fairness e Intelligenza Artificiale: come perseguire l'imparzialità nei Sistemi di decisione automatizzati

Viviamo in un mondo governato da algoritmi digitali, il cui effetto segregatorio, prodotto nell'elaborazione dei sistemi di Intelligenza Artificiale (IA), è sotto l'analisi incrociata delle sfere scientifiche, giuridiche ed etiche.

L'indagine che conduco durante il mio percorso di dottorato assume una prospettiva olistica. Il fulcro della mia ricerca è lo studio approfondito dei fallimenti algoritmici, al fine di comprendere quali siano le persone colpite da tali fallimenti e le conseguenze sociali che ne derivano.



Ceci n'est pas une couple

Cosa significa concretamente per un sistema di IA essere 'equo' o 'non discriminatorio'? L'equità nell'IA si traduce nella formulazione di decisioni basate su criteri matematici oggettivi, evitando di produrre risultati diseguali per individui specifici.

Ne deriva che un elemento cruciale di questa analisi riguarda le metriche utilizzate per valutare l'equità (Fairness) dei modelli algoritmici, poiché queste misure matematiche determinano come la stessa viene misurata e compresa. Gli sforzi mirano, quindi, a valutare non solo la propensione al pregiudizio nei modelli algoritmici, ma anche l'adeguatezza delle metriche scelte e le strategie per mitigare tali pregiudizi quando necessario.

Parallelamente, le decisioni progettuali influenzano profondamente le interpretazioni giuridiche ed etiche dei risultati ottenuti nella società. Risulta fondamentale comprendere quali valori sociali e concetti democratici siano connessi all'attuale formalizzazione matematica dell'equità, al fine di trarre inferenze che possano soddisfare le diverse definizioni tecniche di Fairness. Di conseguenza, l'obiettivo di questa attività è valutare il significato della scelta di un tipo di equità rispetto a un altro in un determinato contesto sociale, e valutare il grado di accettabilità in base al contesto e al criterio di equità scelto. Per questi motivi esistono

numerose definizioni di Fairness, legate a diverse nozioni intuitive di 'decisioni ingiuste', definite considerando l'impatto differenziato sulle persone a causa di varie caratteristiche personali come il genere, l'etnia, l'età, l'orientamento sessuale, politico o religioso, considerate protette o sensibili. Pertanto, per poter arrivare ad attribuire la responsabilità degli effetti discriminatori prodotti dai sistemi di IA, è necessario comprendere quali siano i criteri per stabilire cosa si debba richiedere a una metrica affinché sia funzionale nell'ambiente sociale predefinito.

Nello specifico, allo stato dell'arte, esistono tre principali tipologie di categorizzazione della Fairness: (i) group Fairness: garantisce che gli individui appartenenti a gruppi protetti ricevano, in media, lo stesso trattamento o risultati della popolazione generale; (ii) individual Fairness: si concentra sul garantire che due individui simili, fatta eccezione per gli attributi protetti, ricevano un trattamento o risultati uguali o simili, e la (iii) causality-based Fairness, la quale richiede che gli attributi protetti, come il genere o l'etnia, non abbiano alcun effetto causale sui risultati.

In conclusione, un adeguato livello di sostenibilità sociale può essere raggiunto nel mondo della tecnocrazia algoritmica solo se si sarà in grado di circoscrivere la tipologia di outcome discriminatorio che il sistema di IA ha prodotto in relazione all'ecosistema sociale e agli obblighi giuridici nazionali o sovranazionali di riferimento. Questo sebbene le norme a tutela dei soggetti colpiti da un'attività discriminatoria non si traducano sempre

direttamente in codice, le prassi attuali concedono ai progettisti una certa flessibilità nel determinare quali strategie siano accettabili per tenere conto delle caratteristiche protette durante l'implementazione numerica degli algoritmi. Pertanto, l'analogia tra la formulazione giuridica e l'interpretazione computazionale, determinata dalla formulazione della metrica adeguata, deve diventare centrale nell'analisi dei processi considerati iniqui o ingiusti.



Veronica Paternolli

Giurista, Dottoranda PNRR presso il Dip. di Informatica dell'Università di Verona, si dedica allo studio del principio di Fairness nei sistemi di IA, per contribuire a soluzioni più etiche e responsabili. Con la startup BuildTrust, analizza l'impatto rivoluzionario degli Smart Contract nelle opere civili, al fine di implementare l'efficienza e la trasparenza dei progetti ingegneristici.

