

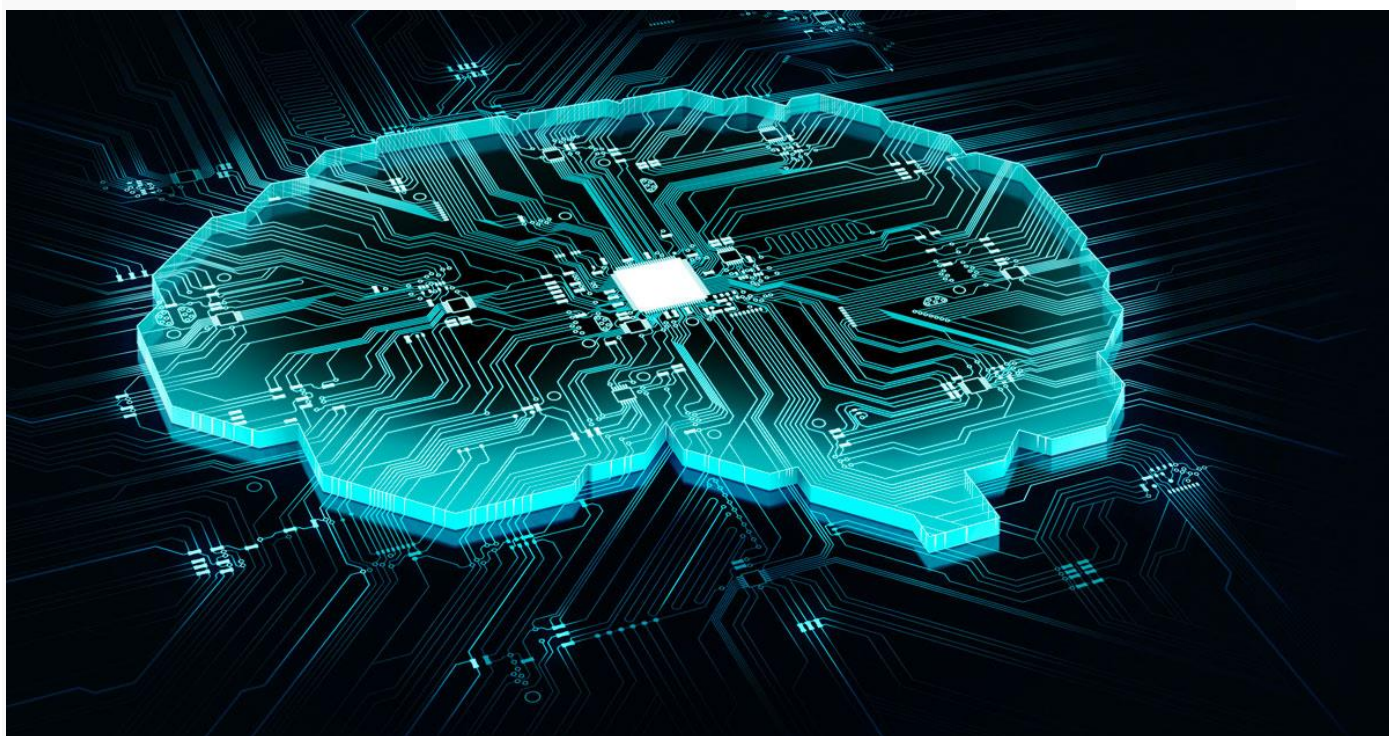
SOCIETÀ

14 APRILE 2023

AI, etica e società: c'è ancora tanto da fare (e da riflettere)

di **Antonella De Robbio**

<https://ilbolive.unipd.it/it/news/etica-societa-ce-ancora-tanto-fare-riflettere>



Non solo ChatGPT: negli ultimi anni l'intelligenza artificiale (IA) è diventata sempre più pervasiva, con un aumento delle applicazioni in settori come salute, agricoltura, energia e ambiente, mentre il numero di pubblicazioni scientifiche sull'argomento è più che raddoppiato dal 2010. Temi come i sistemi generativi di linguaggio, apprendimento automatico e visione artificiale sono sempre più discussi anche dal grande pubblico, tuttavia rimangono da affrontare ancora molte sfide per quanto riguarda la trasparenza, l'efficienza e l'accessibilità delle tecnologie di IA.

È quanto emerge dalla sesta edizione dell'[Artificial Intelligence Index Report](#), iniziativa indipendente dello Stanford Institute for Human-Centered Artificial Intelligence, guidata dall'AI Index Steering Committee, un gruppo interdisciplinare di esperti provenienti da tutto il mondo accademico e industriale. Il documento fornisce un'ampia panoramica e copre una vasta gamma di argomenti tra cui la ricerca sul campo, le applicazioni commerciali e l'impatto sociale delle nuove tecnologie, stimolando un confronto che oltre ai ricercatori e agli operatori economici coinvolge ormai sempre più anche le istituzioni e l'opinione pubblica.

Non è un caso che agli aspetti di politica e governance sia dedicato l'intero capitolo 6. La regolamentazione dell'uso dei sistemi IA e la relativa integrazione nelle normative dei singoli Paesi è infatti sempre più una priorità per i governi in tutto il mondo: se da una parte si vede nell'IA la chiave per migliorare l'efficienza delle politiche e la qualità dei servizi e per aumentare la produttività, dall'altra ci sono anche preoccupazioni riguardo la possibile perdita di posti di lavoro, la sicurezza dei dati personali e la responsabilità etica. Per questo un po' dappertutto è in atto il tentativo di bilanciare opportunità e criticità attraverso un ritrovato protagonismo della politica e della sfera pubblica rispetto agli attori economici.

Facendo un passo indietro i primi due capitoli del rapporto, riguardanti ricerca e sviluppo e prestazioni tecniche, evidenziano che non sono più solo gli Stati Uniti ad investire in IA ma che oggi sono diversi i Paesi in rapida crescita, Cina in testa. A questo riguardo vale ricordare che anche l'Italia ha lanciato il suo primo piano strategico nazionale per l'IA nel 2019, stanziando fondi per supportarne lo sviluppo in settori come robotica, automazione industriale, mobilità intelligente e sicurezza cibernetica.

È interessante notare come sia l'industria a prendere sempre più il sopravvento nella creazione di modelli di *machine learning*. Costruire sistemi all'avanguardia richiede sempre più dati, potenza di calcolo e investimenti economici, risorse alle quali gli attori economici hanno un accesso privilegiato rispetto alle organizzazioni non profit e alle università. Nel 2022 c'erano 32 significativi modelli di *machine learning* prodotti dai privati rispetto a soli tre prodotti dal mondo accademico: una situazione che rischia di avere conseguenze sulle applicazioni pratiche dell'IA, dato che l'industria tende a concentrarsi su obiettivi commerciali e sul profitto, piuttosto che sull'innovazione e sulla creazione di soluzioni etiche e socialmente responsabili (anche se ci sono comunque ottimi esempi di collaborazioni tra pubblico e privato).

Uno spazio importante è dedicato agli aspetti etici ma anche all'utilizzo dell'IA in campo legale e in particolare nella giustizia penale, dove solleva preoccupazioni riguardo a temi come la discriminazione razziale e giustizia equa. L'impatto sul sistema giudiziario e sui diritti degli individui è uno dei temi enucleati da qualche anno, in quanto l'IA ha già avuto un impatto significativo sul modo in cui vengono effettuate le ricerche giuridiche, analizzati i documenti legali e condotti i processi. Le tendenze e le sfide nell'utilizzo dell'IA nel sistema legale, come ad esempio l'etica e la trasparenza nella selezione dei dati di addestramento per gli algoritmi di giudizio, toccano aspetti relativi alla privacy dei dati personali e/o sensibili e ciò solleva non poche preoccupazioni. C'è poi il tema enorme della responsabilità per le decisioni prese dall'IA, specialmente quando esse hanno un impatto significativo sulla vita delle persone come ad esempio nei casi relativi all'assistenza sanitaria o al mercato del lavoro. Da tempo, l'intelligenza artificiale sta cambiando il campo legale, tramite l'introduzione e l'impiego crescente di strumenti di analisi predittiva per valutare la probabilità di vittoria di una causa o di recidiva di un condannato. Si parla ad esempio di rischio di discriminazione algoritmica per quanto riguarda le decisioni automatiche o semiautomatiche che possono avere un impatto sui diritti e le libertà dell'individuo. L'importanza di trovare un equilibrio tra efficienza e accuratezza che l'IA può offrire e il rispetto esigenze delle persone sta generando un dibattito sul modo in cui la tecnologia può essere utilizzata in modo responsabile e giusto in ambito giudiziario.

Il vero focus del rapporto è il capitolo 3, centrato sull'etica. Poiché giganti tecnologici come Google, Microsoft e Facebook continuano a investire enormi quantità di risorse nella

ricerca sull'IA, in particolare nella creazione di modelli di apprendimento automatico sempre più avanzati, è necessario garantire che questi siano utilizzati in modo etico e responsabile. Sviluppare algoritmi sempre più robusti e affidabili è una preoccupazione crescente per evitare minacce come i cyber-attacchi e la manipolazione dell'informazione, oltre ad arginare il fenomeno delle *deepfake*, contenuti falsi o manipolati come video, audio e immagini. Le implicazioni del fenomeno *deepfake* è di estrema gravità in quanto ad essere manipolata è l'opinione pubblica, diffondendo disinformazione. Per affrontare il problema serve un'efficace regolamentazione ma soprattutto educare le persone. Da una parte sarà importante sviluppare strumenti e tecniche in grado di rilevare e contrastare i falsi, dall'altra si dovrà garantire la trasparenza dei percorsi di creazione delle news e soprattutto responsabilizzare all'uso dell'IA a tutti i livelli.

Un altro rischio è la discriminazione, sia intenzionale sia involontaria. Le questioni etiche relative all'IA sono diventate più evidenti al grande pubblico, poiché la barriera tecnica all'ingresso per la creazione e l'implementazione di sistemi di intelligenza artificiale generativa si è drasticamente abbassata: ormai è evidente come l'IA abbia un impatto significativo sulla società con possibili effetti su disuguaglianza, la privacy, il lavoro, tutte questioni che riguardano aspetti etici.

Una delle sfide principali è il campo dell'educazione e della formazione, che il rapporto tratta nel capitolo 5. Le istituzioni dovranno necessariamente rivedere i loro programmi di formazione, sviluppando metodologie educative al passo con la tecnologia. L'istruzione relativa all'IA si sta diffondendo a livello K-12, ovvero nelle scuole primarie e secondarie, tramite corsi di informatica gratuiti per insegnanti, che possono poi insegnare queste competenze ai loro studenti. Il numero di programmi di laurea e di dottorato in informatica con una componente di AI è inoltre in costante aumento. Nel 2021 negli Stati Uniti il 65,4% dei dottori di ricerca formati sull'intelligenza artificiale è stato impiegato nell'industria, più del doppio rispetto a quelli (il 28,2%) che hanno accettato un lavoro in ambito accademico: un dato che suggerisce che l'industria sta diventando sempre più attraente per i laureati in questo campo e che vi è una forte domanda da parte delle aziende.

L'accesso alla conoscenza passa attraverso l'accesso all'informazione e alla capacità di usare strumenti tecnologici: strettamente correlato all'istruzione e alla formazione è quindi il capitolo dedicato alla diversità, che inizia sottolineando l'importanza di quest'ultima nella comunità dell'IA e la sua possibile influenza su ricerca e sviluppo. Si evidenziano a questo riguardo la scarsa inclusività che sembra ancora caratterizzare il settore e se ne discutono le ragioni: nei corsi di informatica ad esempio i tassi di abbandono degli studenti di colore e delle donne sono ancora molto alti rispetto a quelli dei loro colleghi maschi bianchi; nel settore tecnologico ci sono inoltre ancora disparità salariali e di accesso alle opportunità di carriera per le minoranze. Il rapporto rileva anche che la rappresentanza delle persone di genere non binario nella disciplina è estremamente bassa, solo lo 0,1%.

Vengono infine le considerazioni riguardo l'impatto sull'opinione pubblica, in generale ancora curiosa e positiva ma preoccupata riguardo a temi come occupazione, privacy, sicurezza e modo in cui vengono prese decisioni in settori sensibili come salute, giustizia penale e mercato del lavoro. L'AI può ad esempio essere utilizzata per analizzare grandi quantità di dati medici e identificare possibili diagnosi o terapie: sarà però necessario coinvolgere il pubblico ed educare i pazienti non solo sulle opportunità, ma anche sui possibili rischi derivanti essenzialmente dalla cessione dei propri dati sensibili. C'è insomma nell'opinione pubblica una crescente consapevolezza, ma anche il desiderio diffuso di essere informati e di capire meglio come questa straordinaria tecnologia possa

essere utilizzata per migliorare effettivamente la vita delle persone e non per controllarle o limitarne le possibilità.