Insights

protiviti®
Global Business Consulting

Artificial Intelligence

*Dicembre 2023*

# *Etica e Rischi dell'Intelligenza Artificiale: che cosa c'è da sapere e da fare*

a cura di:

Emma Marcandalli
Managing Director

Le soluzioni di Intelligenza Artificiale sono impiegate in misura crescente, ma nascondono insidie di natura etica che le organizzazioni farebbero bene a considerare.

In questa pubblicazione Protiviti spiega a che punto siamo, quali casi hanno fatto scuola e che cosa possono fare le imprese per mitigare i rischi.

Né buona né cattiva in sé. L'Artificial Intelligence (AI) offre soluzioni straordinarie ma, in termini etici, non fa eccezione rispetto a molte altre cose: il giudizio dipende dal modo in cui le applicazioni sono programmate e utilizzate, da come i rischi sono identificati e mitigati e da come gli output sono monitorati per assicurare la conformità alle linee guida etiche stabilite.

Se da un lato aumentano la potenza computazionale, i dati a disposizione e l'accuratezza degli output, dall'altro crescono anche le criticità legate all'uso dell'AI da parte delle imprese. Considerata l'estensione degli usi nel business, il dibattito è destinato a durare a lungo.

Non vogliamo qui parlare dei possibili errori che l'AI può commettere senza un'adeguata supervisione (come peraltro accade agli esseri umani). Vogliamo piuttosto parlare delle questioni etiche che si pongono quando l'algoritmo, che rappresenta l'anima razionale dell'AI:

- replica o amplifica i pregiudizi (bias) di chi l'ha programmata, cosa che può accadere quando gli algoritmi di apprendimento e decisionali considerano parametri potenzialmente discriminatori, come età, sesso, nazionalità, etnia, etc.;
- utilizza dati o serie storiche distorti che riflettono i bias umani (per esempio, il dataset di addestramento non rappresenta correttamente le minoranze e riflette quindi situazioni che in partenza contengono discriminazioni).

In tutti questi casi, l'output può generare disparità di trattamenti, arrivando fino a configurare una violazione dei princìpi di non-discriminazione e tutela della privacy. È il fenomeno conosciuto come discriminazione algoritmica o algoretica, di cui ci sono già alcuni casi eclatanti che fanno scuola.

Eccone tre.

## 1. L'algoritmo discriminatorio di Deliveroo

Nel dicembre 2020, il Tribunale di Bologna, in funzione di giudice del lavoro, ha giudicato discriminatorio l'algoritmo utilizzato da Deliveroo - la piattaforma per la consegna del cibo a domicilio - per stabilire le priorità d'accesso al sistema di prenotazioni delle sessioni di lavoro settimanali da parte dei rider. L'azienda è stata condannata a pagare alle organizzazioni sindacali un risarcimento di 50mila euro, oltre al danno di immagine che ne è conseguito.

L'algoritmo classificava i rider in base ad affidabilità e partecipazione (c.d. tasso reputazionale). Ai fini della valutazione, il calcolo considerava in primis il tasso di presenze lavorative ponendo più in alto nella classifica i rider sempre disponibili, trascurando i motivi delle assenze. Per il Tribunale, quindi, il sistema era implicitamente programmato per negare il "diritto alla malattia" e il "diritto allo sciopero" del lavoratore.

## 2. Le insidie nei processi di selezione del personale

Valutazioni automatizzate, colloqui digitali e analisi dei dati per selezionare curricula e candidati sono usate in misura crescente dalle imprese. I benefici sono facilmente intuibili: processi più efficienti, migliore corrispondenza tra candidati e posizioni ricercate, riduzione dei bias cognitivi che normalmente influenzano le assunzioni (il Cognitive Bias Codex ne stima oltre 180).

I rischi, tuttavia, vanno presi in adeguata considerazione. Per esempio, il rischio di violare il diritto alla privacy se si ricorre al riconoscimento facciale, che potrebbe essere considerato un eccesso rispetto alla finalità di selezione del candidato, e come tale non etico. L'analisi di post, story e video pubblicati sui social media per conoscere orientamenti, gusti, comportamenti potrebbe, analogamente, essere considerato discriminatorio se volto ad escludere dal processo di selezione determinate categorie di candidati.

Ancora: un algoritmo costruito con input impropri potrebbe minare le politiche e gli obiettivi di Diversity & Inclusion di un'organizzazione. Incrociando dati come età, area di residenza, nazionalità / origine etnica, i sistemi potrebbero escludere candidati appartenenti a minoranze aggravando la disuguaglianza sociale.

Qualcuno ricorderà un caso di qualche anno fa (2017) accaduto in Amazon. L'azienda utilizzava un algoritmo, poi dismesso, che assegnava ai potenziali candidati un punteggio da 1 a 5, calcolato però sulla base dei profili di operatori (in questo caso informatici) che nei 10 anni precedenti erano stati considerati per posizioni e ruoli analoghi a quelli oggetto di ricerca. E la maggior parte di questi proveniva da uomini. L'algoritmo, dunque, "aveva insegnato a sé stesso" che i candidati uomini erano preferibili e aveva penalizzato i curriculum che includevano riferimenti al genere femminile. Così facendo, l'algoritmo escludeva dalla ricerca di profili informatici i CV delle donne.

## 3. La mancanza di trasparenza nell'assegnazione di punteggi sociali

Nel 2018, una coalizione di associazioni ha portato il governo olandese in tribunale per l'uso di SyRI, un algoritmo di AI che incrociava dati personali sensibili da 17 grandi basi di dati (fisco, servizi sociali, storia medica, utenze elettriche e telefoniche) all'insaputa dei cittadini. SyRI (acronimo olandese per «sistema indicatore di rischio») serviva a capire se i percettori di sussidi o altre forme di welfare fossero inclini a frodi e abusi attribuendo a ciascun cittadino un punteggio di rischio. In pratica, la versione digitale di un ispettore per controllare, con finalità preventiva, il corretto accesso allo stato sociale.

Questo sistema, tuttavia, fece fin da subito sollevare più di un sopracciglio: il governo olandese, infatti, aveva definito la lista dei quartieri dove sarebbe entrato in funzione, ed erano tutti quartieri poveri di Rotterdam, Eindhoven e Haarlem.

L'algoritmo profilava anche cittadini che non percepivano sussidi; bastava che vivessero nelle zone attenzionate. Oltre a ciò, nessun cittadino poteva sapere il proprio punteggio di rischio né come questo era calcolato.

Non c'è stata alcuna trasparenza: di quali dati per ciascun cittadino disponesse il governo, come venissero incrociati, quali modelli e fattori l'algoritmo impiegasse per valutare il rischio. E non è neppure chiaro se venissero impiegati anche sistemi di riconoscimento facciale.

Il Ministero degli Affari Sociali olandese ha usato SyRI dal 2014 all'inizio del 2020, quando è arrivata la sentenza del Tribunale dell'Aja che ne ha disposto la sospensione. Mentre, infatti, la necessità di un governo di controllare che non vi siano abusi nell'utilizzo dello stato sociale è comprensibile, è invece «contrario al diritto considerare i cittadini, particolarmente i più poveri, come sospettati da spiare anziché come detentori di diritti». Non a caso, il giudice ha basato la sentenza non sul GDPR, ma sulla Convenzione europea dei diritti umani, che all'articolo 8 sancisce il diritto individuale a una vita privata.

## Che cosa possono fare le aziende per mitigare i rischi di un uso non etico dell'AI?

I casi sopra citati dimostrano come uno sviluppo poco attento o un utilizzo non etico dell'AI possano causa danni non solo alle persone, ma anche agli enti e alle imprese.

Chi sviluppa e utilizza sistemi di AI deve dunque definire e attuare, durante tutto il ciclo di vita dell'applicazione, differenti strategie di mitigazione dei rischi, che devono ispirarsi ai framework di AI etica e responsabile allo scopo di garantire la massima sicurezza dell'applicazione per l'individuo e per i suoi diritti fondamentali.


Ci sono vari toolkit e template che possono ispirare le aziende: segnaliamo, ad esempio, il *Responsible AI Tools and Practices di Microsoft AI* o, ancora, il *Responsible AI – Tools and resources to build responsibly di AWS (amazon.com)*.

In linea con tali framework, **strategie preventive di "ETHIC BY DESIGN"** impongono un'attenta verifica durante la fase iniziale di costruzione dell'algoritmo di apprendimento, in particolare sugli attributi considerati dall'algoritmo, così da escludere il più possibile regole di analisi discriminatorie.

Allo stesso modo, si può lavorare sul dataset di apprendimento allo scopo di eliminare quanto più possibile o pulire le informazioni che possono creare distorsioni o discriminazioni. In questa fase, la “spiegabilità” e trasparenza dell’algoritmo è fondamentale per lo sviluppo di un’AI responsabile.

**EX-POST** è essenziale monitorare e valutare nel continuo l’output generato dall’AI e come l’algoritmo auto-apprende dall’analisi dei dati che riceve, garantendo un’adeguata sorveglianza umana e prevedendo degli audit indipendenti, che consentano di intercettare eventuali distorsioni e correggerle tempestivamente.

Tutto questo rende trasparente e comprensibile l’algoritmo e consente di poterlo ottimizzare e correggere nel tempo.

Ma le strategie sopra menzionate da sole potrebbero non bastare. In un mondo “perfetto”, queste ultime dovrebbero innestarsi all’interno di un **sistema più articolato**, fatto di regole, strutture organizzative, responsabilità, processi e procedure chiaramente definite e correttamente applicate, che assicurino un governo e un controllo adeguato su sviluppo e utilizzo dell’AI in azienda.

Spetterà ai Business Leader delle singole realtà aziendali decidere il livello di articolazione e complessità di tale sistema, che dovrà essere calibrato in relazione al profilo di rischio specifico, alle peculiarità organizzative nonché alla propensione al rischio-rendimento atteso.

Parola d’ordine: accelerare l’innovazione sprigionabile con l’AI minimizzando i rischi ad essa connessi.

## Cosa può fare Protiviti

Protiviti dispone delle competenze, delle professionalità, delle metodologie e degli strumenti necessari a supportare i Clienti nell’analisi e mitigazione dei rischi connessi allo sviluppo e all’utilizzo dell’AI in azienda, nonché nel disegno e nell’implementazione di sistemi di governo e controllo dell’AI allineati alle best practice.

## CONTATTI


**Emma Marcandalli**

Managing Director

emma.marcandalli@protiviti.it

LinkedIn

**Francesco Monini**

Managing Director

francesco.monini@protiviti.it

LinkedIn

**Luca Risi**

Managing Director

luca.risi@protiviti.it

LinkedIn