

Intelligenza artificiale: perché pone interrogativi etici e sui nostri diritti di cittadini?



di Rodolfo Marchisio
e Stefano Penge

Dopo aver [introdotta](#) il tema, averlo [approfondito](#) vediamo di spiegare un po' di più gli **aspetti critici dal punto di vista della cittadinanza**.

Un intervento educativo (formazione di competenze e cultura digitale) è possibile solo se si *supera la visione della digitalizzazione come un processo di democratizzazione spontanea dell'accesso delle informazioni oggi non più possibile*. Agenda digitale. **E come un processo "magico" e spontaneo di riforma della scuola in senso democratico grazie alle tecnologie**

Quale IA

L'Europa ha stabilito i limiti da porre allo sviluppo della IA per tutelare i cittadini: [AI Act](#).

Riguardano i *modelli fondativi alla base di grandi sistemi di AI* e il *ricorso alla sorveglianza biometrica e alla (ipotizzata) polizia predittiva*. L'AI Act, **inquadra i diversi sistemi di intelligenza artificiale** pone paletti, **proibisce alcune applicazioni** e introduce procedure di salvaguardia per mettere al riparo i cittadini dell'Unione da **abusi e violazioni** dei diritti fondamentali. L'attenzione è:
1- sui **modelli fondativi**, quelle forme di intelligenza artificiale generali in grado di svolgere compiti diversi (come creare un testo o un'immagine) e allenati attraverso un'enorme mole di dati non [categorizzati](#).

Si è lavorato ad una applicazione preventiva delle regole su **sicurezza** informatica, **trasparenza dei processi di addestramento** e **condivisione della documentazione tecnica** prima di arrivare sul mercato. Wired.

2- La UE è arrivata ad un **compromesso** sull'uso dell'AI per **compiti di polizia** e di **sorveglianza**.

Sul riconoscimento biometrico in tempo reale, ma si era discusso anche di **polizia predittiva** poi vietata. Ossia usare gli algoritmi per prevedere le probabilità con cui **può essere commesso un reato, da chi e dove?** Quali diritti verrebbero violati con questa delega? **AI Act Europa** (Wired)

Oltre a questo ci sono nodi e modi di vedere la IA (che è mondo complesso) che emergono anche dal fatto che il termine sia sempre più spesso accompagnato da un aggettivo (generativa, etica, spiegabile ...).

Si riflette su questi **temi e interrogativi**:

- L'intelligenza artificiale può **prendere decisioni** basate sui **dati** e sul **passato**. Talora **errate** a causa di una mancanza di consapevolezza del contesto. Esiste una possibilità di sapere come? Dovrebbe essere la **IA explainable**. (spiegabile). Ma funziona? Per tutta la IA?
- Una decisione può influenzare la vita delle persone, essere utilizzata per scopi maligni, essere **influenzata dai pregiudizi e dai valori dei suoi creatori**. Gli algoritmi utilizzati nell'IA possono essere influenzati da **pregiudizi, anche involontari**. O "premiati" in modo da imparare in una certa direzione prefissata da chi li progetta. Ciò può portare a decisioni discriminatorie, sbagliate, diverse da quelle che avrebbero preso degli umani (processi di selezione del personale, prestito bancario...). Anche se le stesse decisioni umane sono **diverse se prese da persone competenti, da potenti, da una maggioranza più o meno informata**. Come constatiamo da anni. Ma il meccanismo si potenzia ed opacizza con IA.

- L' IA comporta la raccolta e l'elaborazione **di grandi quantità di dati personali**. Questo solleva preoccupazioni sulla **privacy** delle persone e sulla **sicurezza dei dati**.
- L'IA sta automatizzando e **sostituendo sempre più lavori umani**. Ciò solleva preoccupazioni sulla perdita di posti di lavoro e sulla necessità di una riqualificazione delle persone.
Vedi [studi del possibile impatto sulla occupazione](#).
- Nonostante i progressi nell'IA, esistono ancora criticità sui sistemi **di controllo decisionale affidabili e trasparenti**. Questi problemi richiedono una **rigorosa regolamentazione, un'etica e una responsabilità riguardo all'uso dell'IA, nonché un'attenzione costante alla sua implementazione e sviluppo**.

Stefano Penge, filosofo e informatico, avverte:

“L'interesse recente – da un anno a questa parte – per l'intelligenza artificiale, con tutti i discorsi a favore e contro, si è concentrato su un sottoinsieme particolare di abilità che attraggono e spaventano apparentemente più delle altre: quelle linguistiche, che permettono ad un software di analizzare un testo, di tradurlo, di prepararne un riassunto, ma anche di continuarlo o riscriverlo secondo un altro stile; il tutto, attraverso un'interazione continua con le persone (il famoso 'prompt').

*Il riconoscimento facciale e la guida autonoma (le altre abilità di cui si occupa l'AI ACT) hanno in comune con questo sottoinsieme linguistico l'abbandono della **logica tradizionale**, quella basata su regole e deduzione, a favore di una **logica induttiva**. Niente di nuovo, insomma: tutto sommato anche la nostra specie funziona così la maggior parte del tempo e solo in casi speciali usiamo la logica classica e il ragionamento formale. Vedi anche <https://www.stefanopenge.it/wp/intellig-enti/>.*

*Quelli citati sopra sono tutti esempi di **machine learning**, che indica una maniera di costruire modelli artificiali che*

simulano una parte del mondo (per esempio, un testo) a partire dalla raccolta di dati relativi a miliardi di situazioni, per induzione appunto.

*Questo sistema è **inerentemente soggetto ad errore**, perché i dati di partenza non sono tutti i dati ma solo una selezione; perché la maniera in cui sono stati selezionati quei dati (e la maniera in cui vengono “premiati” i modelli migliori) potrebbe essere non oggettiva ma guidata da preferenze culturali o sociali. Ma soprattutto perché il futuro non è per forza uguale al passato, mentre il machine learning per sua natura si basa proprio su questa fede.*

*Ma anche se decidessimo di fidarsi delle competenze dei software basati sul machine learning, il problema nasce quando queste competenze vengono usate non a fianco, ma **al posto** di quelle umane come se fossero equivalenti in qualità, **perché costano di meno e rendono di più a chi le governa.**”*

Anche **Sasha Luccioni**, ricercatrice nell'AI etica e sostenibile -HuggingFace – sottolinea:

1. **Consumi energetici troppo elevati, i bias cognitivi che vengono riproposti dagli algoritmi, e la tutela del copyright sono le tre sfide che [l'intelligenza artificiale dovrà affrontare](#).**

*La diversità deve essere presente all'interno dei modelli dell'intelligenza artificiale: “La tecnologia non è neutra, i bias dell'[intelligenza artificiale](#) altro non sono che quello che noi vediamo nella società. **Qualsiasi stereotipo si applichi viene peggiorato dal sistema, che tende ad amplificare la distorsione. Solo il 12% di chi lavora nell'AI è donna** ...è fondamentale che questa tecnologia “non sia nelle mani di pochi soggetti, **deve essere open**, per intervenire su questi aspetti in modo condiviso”. Wired*

2. **Per tutelare la proprietà intellettuale e il copyright (Luccioni) bisogna “implementare meccanismi per capire quali siano le fonti che sono state utilizzate,**

analizzare i [dataset](#) usati da un'intelligenza artificiale. Un lavoro immane che va fatto fare dalle macchine, ma **come e da chi controllate?** E i cittadini/utenti in tutto questo?

3. **L'impatto ambientale non è trascurabile:**

Ad esempio, sono state emesse 500 tonnellate di CO₂ per la creazione di Gpt-3". Wired

A proposito della possibilità di controllare l'IA Penge paragonava la IA ad una piccola bomba atomica: ha un potere enorme, ma può essere costruita (e smontata) solo da quelli che hanno sufficiente potere per raccogliere, selezionare e gestire i materiali (radioattivi!) necessari. Il semplice cittadino non avrà mai questa possibilità; al massimo può giocare con le interfacce che permettono di interagire con i software di IA.

"Per dire meglio ci sono tre aspetti:

*– uno è quello della **chiusura del software**, in questo caso della **proprietà** dei modelli linguistici utilizzati, che hanno grande valore perché sono stati prodotti con molte ore di calcolo e molto lavoro di revisione umana;*

*– il secondo è quello delle **dimensioni di questi modelli**, che per essere creati e trattati richiedono risorse informatiche enormi, che consumano elettricità, scaldano, etc.;*

*– il terzo è quello dell'**opacità**: un Large Language Model è costruito con miliardi di parametri. Significa che **non è analizzabile da umani, non si può capire perché risponde in un modo o in un altro**. Questi tre aspetti insieme **impediscono di mettere le mani in questi strumenti** come si potrebbe fare con un software libero, ad esempio con LibreOffice (ma non con MS Word)."*

Tornando alla scuola un esempio di preoccupante ottimismo in questo [podcast](#) che ipotizza una scuola alla cinese. Paese che notoriamente è da tempo l'incarnazione

del “grande fratello” che viola costantemente i diritti dei cittadini, sudditi [ipercontrollati](#) e prigionieri. Concluderemo sulla scuola.