volerelaluna LA POLITICA PUNTOACAPO

Intelligenza artificiale: tra etica e profitto

Autore: Norberto Patrignani

Il 9 dicembre 2023 l'annuncio che l'Europa (Parlamento, Commissione e Consiglio) ha raggiunto un accordo (*AI Act*) per la regolamentazione della cosiddetta "intelligenza artificiale" arriva sulle prime pagine di tutto il mondo. Seppur tra mille discussioni tra chi vuole maggiore protezione dei diritti (il Parlamento EU) e chi invece non vuole mettere troppi "limiti all'innovazione" (come Germania, Francia e Italia), viene **per la prima volta approvata una normativa che pone dei limiti alle applicazioni "ad alto rischio", come il riconoscimento biometrico, richiedendo la supervisione umana**. Ci sarà tempo per approfondimenti, ma sicuramente questo primo passo rappresenta un importante tentativo di bilanciare l'enorme potere accumulato dalle Big Tech a partire dagli anni '90, quando il mercato e le tecnologie hanno iniziato a dominare incontrastati e l'educazione e le norme sono stati completamente abbandonati per favorire la "nuova frontiera" dell'*e-business*. Forse è tempo di trovare la giusta miscela tra questi quattro domini riscoprendo l'importanza dell'educazione al corretto uso delle tecnologie (sobrietà e saggezza digitale) e l'utilità delle norme condivise ed approvate democraticamente. Come diceva il grande giurista Stefano Rodotà: in assenza di norme la rete è il "Far West", dove vince il più forte. Questo primo passo dell'Europa in termini di norme è importante, pur con tutti i suoi limiti fatti emergere ad esempio da Amnesty International, dalla coalizione European Digital Rights Network etc. Una delle decisioni più controverse ad esempio è quella di non aver proibito totalmente il riconoscimento facciale, con gravi rischi per i diritti umani (Amnesty, 2023).

Ma **di cosa parliamo quando parliamo di "intelligenza artificiale"?** Il termine viene introdotto dal 1955 in campo informatico e da allora sono passate molte stagioni di alti e bassi. In effetti raggruppa molte aree di ricerca come: il cosiddetto **"apprendimento" delle macchine, l'elaborazione del linguaggio e della voce umana, la pianificazione e l'ottimizzazione di attività, la robotica, l'elaborazione delle immagini** etc. Anche la definizione formale data dal National Institute of Standard USA è ancora un po' generica: «un sistema che può, per un dato insieme di obiettivi, generare output come previsioni, raccomandazioni o decisioni, influenzare ambienti reali o virtuali. I sistemi di intelligenza artificiale sono progettati per funzionare con diversi livelli di autonomia» (*nist. gov.*). Eppure, negli ultimi mesi, praticamente tutto l'universo mediatico su scala mondiale è stato inondato dal dibattito sulla cosiddetta "intelligenza artificiale". Andando a vedere bene, si parla in realtà dei primi due campi elencati sopra: l'apprendimento delle macchine e l'elaborazione del linguaggio. L'amichevole interfaccia proposta dai sistemi conversazionali (come ChatGPT) ha scatenato una delle "tempeste perfette" sui media. In poche settimane dal lancio (novembre 2022) il numero di utenti collegati è balzato a centinaia di milioni, dopo un anno è arrivato a 1,7 miliardi.

Ovviamente anche il valore di mercato delle imprese che rilasciano questi sistemi è aumentato esponenzialmente. Infatti le imprese in grado di memorizzare ed elaborare le

quantità di dati necessari per la “calibrazione” di questi algoritmi si contano sulle dita di una mano. Questi sistemi conversazionali, per costruire i “modelli del linguaggio” (i *Large Language Model*, LLM) sui quali sono basati, hanno bisogno di elaborare quantità gigantesche di dati. Questo serve per costruire un testo comprensibile e coerente, statisticamente simile a quelli contenuti nei dati immessi dai progettisti. Molto azzeccata è infatti la definizione “pappagallo stocastico” data da alcune ricercatrici (Bender e al., 2021). A seguito dell’acceso dibattito scatenato dall’articolo, una di loro, Timnit Gebru, dell’*Ethical Artificial Intelligence Team* di Google è stata licenziata.

Ecco quindi emergere la realtà: **le tecnologie non sono mai neutre, sono *sistemi socio-tecnici*. Tecnologie e società si plasmano a vicenda**. Infatti i progettisti scelgono gli insiemi di dati da usare (*dataset*) nella fase di “addestramento” (ma è preferibile parlare di “calibrazione”): da quali fonti attingere? Con quali criteri selezionano i dati e arricchirli con metadati? Come viene affinato il modello? Verso quali obiettivi? Tutte queste scelte sono nelle mani dei progettisti, ovvero delle imprese che investono in questo settore (ormai si parla di centinaia di miliardi di dollari). Come se non bastasse, questo flusso va filtrato a mano da esseri umani (tipicamente reclutati nel Sud del pianeta per pochi dollari al giorno), i cosiddetti “schiavi del clic” (Casilli, 2020). Il risultato di tutto questo è forse **la più grande concentrazione di potere e ricchezza in poche mani della storia dell’umanità**.

Indubbiamente queste interfacce conversazionali sono molto attraenti e il loro tono amichevole per molte persone rappresenta addirittura una risposta alla collezione di solitudini del Web. Ma anche questa è una scelta, perché non presentare agli utenti connessi una semplice interfaccia come per i motori di ricerca? Forse perché i miliardi di utenti connessi alle Big Tech, che forniscono servizi online gratuiti, forniscono a loro volta un continuo flusso di dati innescando un gigantesco circuito chiuso. Infatti questo sistema a “feedback dopaminico” (si parla esplicitamente di *persuasive technologies*) serve a incassare dagli inserzionisti pubblicitari: più persone sono connesse per più tempo, più aumentano i profitti. Come dice Shoshana Zuboff «il capitalismo manageriale andava a caccia del nostro corpo per automatizzarlo. Il capitalismo della sorveglianza va a caccia della nostra mente per automatizzarla» (Zuboff, 2019).

Gli enormi pericoli sociali di queste tecnologie sono evidenti. Il più grave forse è il caos epistemico: **la concentrazione in poche mani di un potere di persuasione** (tramite mala-informazione, mis-informazione, dis-informazione) **mai visto nella storia rischia di mettere seriamente in pericolo i sistemi democratici**. Si pensi alle campagne elettorali. Chi ha in mano strumenti interattivi sofisticati e personalizzati avrà in mano un enorme potere di influenza. Perché il miliardario Elon Musk ha acquistato Twitter (ora si chiama X)? Forse perché nelle sue basi di dati “raffinati” ci sono anni e anni di conversazioni politiche soggette alla supervisione di umani: una preziosa base di dati che risulterà molto utile durante la campagna elettorale USA del 2024. **Un algoritmo calibrato con dati di questo tipo sarà in grado di scrivere milioni di commenti sui social, interventi**

impeccabili sui media etc.

Non a caso l'*AI Act* europeo individua questi rischi e i sistemi che verranno classificati ad "alto rischio" (es. la cosiddetta "AI generativa") dovranno **rispettare regole di trasparenza, fornire gli elenchi dei materiali usati in input nella fase di "calibrazione", indicare chiaramente che il contenuto è stato generato da una macchina**, etc. ma su tutto questo bisognerà ancora approfondire nei prossimi mesi. Il dibattito ha fatto emergere molti aspetti sociali, etici e ambientali legati a questa evoluzione dell'informatica. Aspetti che devono essere approfonditi al più presto. Tra questi:

– l'**impatto sul mondo del lavoro** (come verrà gestita la inevitabile disoccupazione di *white collar*?);

– l'**impatto ambientale** (una fase di "addestramento" di ChatGPT3 produce una quantità di CO_2 equivalente a 600 voli New York – Parigi);

– il progressivo slittamento verso una **delega completa alla macchina** (passando dalla semplice previsione, alla predizione, alla prescrizione, l'umano, moderno "apprendista stregone" finisce per delegare alla macchina anche scelte di vita o di morte, es. i droni militari);

– la **mancanza di trasparenza delle fonti e degli algoritmi utilizzati** (diffondendo nella società meccanismi automatici basati su tecnologie chiuse, inaccessibili ad un controllo democratico).

Sono almeno immaginabili scenari alternativi? Certamente. Iniziano a diffondersi modelli di linguaggio aperti e verificabili (Bloom, OpenLM etc.). Questo fa emergere anche eventuali usi positivi di queste nuove tecnologie (non chiamiamole "intelligenze artificiali", chiamiamole semplicemente per quello che sono: algoritmi dinamici che si calibrano con tanti dati). Si pensi a strumenti in mano alle persone come strumenti di lavoro, non strumenti che sostituiscono gli umani. Si pensi alle possibili nuove interfacce che facilitano l'accesso alla conoscenza, alla consultazione di basi di dati finora inaccessibili a chi non conosce linguaggi di interrogazione sofisticati. Si pensi a dei "cobot" (collaborative robot) che aiutano le persone nelle loro attività, anche creative, inclusive. Perché non pensare a strumenti per stimolare la riflessione etica e il pensiero critico, migliorare le intuizioni, aiutare ad incorporare valori negli artefatti, usare euristiche affidabili, considerare l'impatto a lungo termine delle nostre azioni? Perché non pensare a strumento aperti, riscoprendo il valore delle "tecnologie conviviali" (Illich, 1973; Milani, 2022). Tecnologie conviviali sviluppate e sostenute da comunità responsabili e con una profonda preparazione etica oltre che tecnica. Si pensi alla Rete Italiana Open Source (RIOS), ad associazioni come Informatici Senza Frontiere, o *Sloweb*, a riferimenti accademici come *Nexa Center for Internet & Society* del Politecnico di Torino.

Persone esperte di informatica che conoscono anche l'esistenza di un *Code of Ethics and Professional Conduct*, definito nel 2018 dalla più grande associazione di informatica l'*Association for Computing Machinery*, che introduce il principio guida di «usare le competenze per il bene della società e dell'ambiente» (ACM, 2018). Le persone esperte di informatica sanno come è fatto un sistema e come funziona: è tempo di chiedersi come e perché progettarlo? Chi lo userà, per quali scopi? Fino alla domanda cruciale: se progettarlo. **Iniziare a lavorare a scenari alternativi e meno centralizzati, oltre a riscoprire la bellezza di immaginare l'utopia, potrebbe aiutare a ridimensionare le campagne marketing martellanti** (se non le vere e proprie speculazioni finanziarie, tipo lo "*snake oil*" del vecchio Far West) che parlano di superamento dell'umano, che continuano ad inseguire l'antico mito del *Golem*, della creazione di una *Artificial General Intelligence* (AGI).

Però è necessaria una riflessione etica, una domanda profonda sugli scopi, sul senso di queste ricerche. Come giustamente scrive Luca De Biase su *IlSole24Ore* a proposito della stucchevole saga attorno a *OpenAI*, l'impresa che ha distribuito *ChatGPT,* «la mentalità tra tecnologo etico e quella dell'imprenditore che vende la sua merce erano destinate scontrarsi» (De Biase, 2023). Ricorda molto il "Discorso ai lavoratori" di Adriano Olivetti (1901-1960): «Può l'industria darsi dei fini? Si trovano questi semplicemente nell'indice dei profitti?» (Olivetti, 1955).

Riferimenti

– ACM (2018), ACM Code of Ethics and Professional Conduct, https://www.acm.org/code-of-ethics

– Amnesty (2023, 9 Dicembre), *EU: Bloc's decision to not ban public mass surveillance in AI Act sets a devastating global precedent*. amnesty.org

– Bender E.M., Gebru T., McMillan-Major A., and Mitchell M. (2021), *On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?* Fairness, Accountability, and Transparency (FAccT '21), March 3–10, 2021

– Casilli, A.A. (2020), *Schiavi del clic. Perché lavoriamo tutti per il nuovo capitalismo?* Feltrinelli

– De Biase, L. (2023, 19 novembre), *Il nodo è lo scontro tra etica e profitti*, *IlSole24Ore*

– Illich, I. (1973), *Tools for Conviviality*, Harper & Row

– Milani, C. (2022), *Tecnologie Conviviali*, Eleuthera

– Olivetti, A. (1955). Discorso "Ai lavoratori di Pozzuoli" per l'inaugurazione dello

stabilimento di Pozzuoli, 23 aprile 1955, in A.Olivetti (1959), *Città dell'uomo*, Edizioni di Comunità

– Zuboff S. (2019), *The Age of Survellance Capitalism*, PublicAffairs