Intelligenza Artificiale:chi controlla i “significati”? Non fidiamoci troppo della IA

Data 17 maggio 2027
Categoria Medicinadigitale

Le Intelligenze Artificiali strutturate su reti neurali potentissime che conferiscono le ben note proprietà dei Large Language Models sono state oggetto, oltre che di meritate lodi, anche di critiche basate in particolare sugli errori e sulle “allucinazioni”.
Vi è tuttavia un grave limite intrinseco evidenziato e correttamente criticato da una grande filosofa britannica, Miranda Fricker(1) nel “lontano” 2007 e sviluppato e diffuso in Francia ed ora in Italia da Gloria Origgi filosofa italiana poco apprezzata in patria ma molto in Francia.
Sintetizziamo un interessante articolo della Origgi, che riteniamo approfondisca le corrette riflessioni della Fricker, aggiornandole alla nuova era dei LLM.

Per comprendere meglio le riflessioni delle due filosofe, ed ovviamente individuarne limiti ed eventuali errori è necessario definire un concetto base della Fricker, ovvero “La ingiustizia epistemica”, teorizzata dalla filosofa nel 2007, che è una forma di ingiustizia che colpisce un soggetto nella sua capacità di conoscere e trasmettere sapere, a causa di pregiudizi sociali. Fricker ne identifica due tipi principali: l'ingiustizia testimoniale (deficit di credibilità per pregiudizio) e quella ermeneutica (mancanza di risorse interpretative collettive per comprendere un'esperienza). *La Ingiustizia Testimoniale* si verifica quando un ascoltatore assegna un livello di credibilità inferiore alle parole di un parlante a causa di un pregiudizio, spesso legato a identità sociali (genere, razza, classe). (Esempio: non credere a una testimonianza a causa del genere o della etnia del testimone). *La Ingiustizia Ermeneutica* avviene quando una lacuna nel bagaglio concettuale collettivo impedisce a un soggetto di dare senso alla propria esperienza. Esempio storico: l'assenza del concetto di "molestie sessuali" prima degli anni '70, impediva alle donne di etichettare e denunciare tali atti, così come la assenza di riconoscimento dei pregiudizi razziali in molti paesi porta a non riconoscere discriminazioni ed ingiustizie di vario genere e tipo…

Il cuore dell’articolo di Gloria Origgi può essere riassunto in una tesi forte: il potere dell’intelligenza artificiale non consiste soltanto nel raccogliere, possedere o controllare dati, ma nel controllare i significati che quei dati assumono nella vita sociale. Questo è il punto decisivo. Un dato, di per sé, non “parla”: diventa rilevante quando viene interpretato, classificato, collegato ad altri dati, trasformato in profilo, previsione, raccomandazione o decisione. L’IA non si limita dunque a osservare la realtà: contribuisce a definirla, perché stabilisce categorie, produce etichette, attribuisce probabilità, costruisce identità operative. **In questo senso controllare i significati vuol dire influenzare il modo in cui una società comprende le persone, distribuisce opportunità, riconosce diritti, assegna responsabilità e produce esclusione.**
L’articolo parte da un confronto implicito tra due rivoluzioni digitali. La prima è stata quella del web e di Google: l’accesso all’informazione sembrava ampliarsi enormemente, liberando tempo, energie e possibilità di ricerca.
La seconda è quella degli LLM e dell’IA generativa, che non si limita più a cercare informazioni, ma conversa, interpreta, sintetizza, produce testi, immagini, classificazioni e decisioni. La tentazione è vedere in questa nuova fase una frattura assoluta, quasi l’arrivo di “alieni intelligentissimi” capaci di sostituire il lavoro umano e governare le nostre società. Origgi però invita a spostare l’attenzione: il problema non nasce solo oggi con ChatGPT o con i modelli generativi. Esisteva già prima, nella progressiva integrazione degli algoritmi in ogni dimensione della vita sociale.
La questione centrale è: che cosa controllano davvero questi sistemi? Non solo dati, non solo decisioni, ma i criteri stessi con cui diamo senso alla realtà.
Per chiarire questo punto, l’autrice ricorre al concetto di ingiustizia ermeneutica, elaborato dalla filosofa Miranda Fricker. L’ingiustizia ermeneutica si verifica quando una persona vive un’esperienza di disagio, torto o esclusione, ma non dispone delle parole, delle categorie o dei concetti socialmente riconosciuti per comprenderla e farla comprendere agli altri. È una forma di ingiustizia epistemica: riguarda cioè la nostra capacità di conoscere, interpretare e comunicare la realtà. **Se una società non possiede le risorse concettuali per nominare una certa esperienza, chi la subisce resta in una posizione di impotenza: sente che qualcosa non va, ma non riesce a tradurlo in una comprensione ed in una reazione adeguata e riconosciuta.**

Origgi applica questo concetto al mondo dell’IA. Gli algoritmi, soprattutto quando sono opachi ed integrati nelle istituzioni, producono una nuova forma di ingiustizia ermeneutica: i cittadini vengono classificati, profilati e valutati secondo criteri che non conoscono, non comprendono e non possono discutere. Possono subire una decisione automatizzata — per esempio in ambito assicurativo, bancario, sanitario, lavorativo, amministrativo — senza sapere esattamente quale immagine di loro sia stata costruita dal sistema. Il problema non è soltanto che l’algoritmo possa sbagliare; è che spesso il soggetto non possiede gli strumenti per capire quale significato, ruolo e valore gli sia stato attribuito.

Qui entra in gioco il concetto di risorse ermeneutiche collettive. Esse sono l’insieme dei significati, delle categorie, dei racconti, delle parole e dei concetti attraverso cui una società comprende sé stessa, gli individui e il mondo. In passato queste risorse venivano prodotte, almeno idealmente, nella sfera pubblica: istituzioni, media, diritto, università, dibattito politico, cultura condivisa. Erano comunque oggetto di conflitti e rapporti di

forza, ma almeno potevano essere discusse. L'autrice sottolinea infatti che i significati non sono mai neutrali: nominare, classificare e definire la realtà è sempre anche un "esercizio di potere", non necessariamente negativo, ma pur sempre "potere".

La novità dell'IA consiste nel fatto che una quota crescente di queste risorse ermeneutiche viene oggi prodotta da sistemi tecnici privati, opachi e accessibili solo a pochi: grandi corporation, ingegneri, tecnocrati, proprietari delle infrastrutture digitali. Gli algoritmi raccolgono comportamenti online e offline, li aggregano, li etichettano, li trasformano in profili e previsioni. Così facendo, producono nuove categorie sociali: utenti affidabili o inaffidabili, consumatori ad alto valore, pazienti a rischio, lavoratori promettenti, soggetti sospetti, profili fragili, persone assicurativamente costose. Queste categorie non sempre sono visibili, ma incidono concretamente sulla vita delle persone.

L'articolo insiste su un punto molto importante: l'IA non interpreta solo ciò che diciamo consapevolmente, ma anche ciò che facciamo. I nostri spostamenti, acquisti, ricerche, testi, immagini, interazioni, perfino il ritmo e la frequenza delle nostre attività digitali diventano materiale interpretativo. I dati vengono classificati e collegati tra loro fino a produrre un'immagine operativa della persona. Questa immagine può essere più influente della narrazione che la persona fa di sé stessa. **In altre parole: non siamo più soltanto ciò che diciamo di essere, ma soprattutto ciò che i sistemi predittivi inferiscono che siamo.**
Da qui nasce un senso di alienazione. Gli algoritmi costruiscono significati su di noi usando risorse interpretative che non ci appartengono e che non possiamo controllare. Non sappiamo quali dati siano stati raccolti, come siano stati assemblati, quali correlazioni siano state considerate, quali etichette siano state applicate. Il risultato è una società in cui le persone sono continuamente interpretate da sistemi che non possono interrogare davvero. L'autrice parla perciò di risorse ermeneutiche "artificialmente cristallizzate": significati che, invece di restare aperti alla negoziazione sociale, vengono congelati in classificazioni automatiche.
Se concordiamo con questa analisi non basta chiedere algoritmi più efficienti o meno discriminatori. Occorre sviluppare nuove competenze epistemologiche, filosofiche e semiotiche capaci di analizzare i significati prodotti dall'IA.
Serve un "meta-livello" di controllo, svolto da strutture pubbliche rappresentative che tutelino anche le minoranze: concetti con cui valutare altri concetti, categorie con cui discutere le categorie automatizzate, strumenti culturali per rendere negoziabili i significati generati dalle macchine. La posta in gioco non è soltanto tecnica, ma democratica e antropologica: mantenere il controllo sulla produzione dei significati condivisi significa difendere una delle attività più proprie dell'essere umano come specie storica, sociale e politica.

Concludendo alla luce di queste riflessioni la domanda decisiva non è più soltanto: "Chi possiede i dati?", ma: "Chi controlla i significati prodotti dai dati?". E da questa domanda dipende una parte rilevante del nostro futuro di libertà, democrazia e giustizia.

Riccardo De Gobbi e Giampaolo Collecchia

Bibliografia
1) Fricker M.: Epistemic Injuistice Oxford University Press 2007
2) Origgi G.: "Chi controlla i significati? Le nuove ingiustizie della IA" in:Umana , Troppo Umana La IA e noi. MICROMEGA n.6 2024