IA, i due inganni

written by Luca Ricolfi | 19 Gennaio 2026

Sui vantaggi economici e i guadagni di efficienza dell'intelligenza artificiale non ci sono molti dubbi. Ci sono almeno due ambiti, tuttavia, nei quali l'IA può rivelarsi un alleato infido.

Il primo è il campo delle questioni eticamente o politicamente sensibili, che è il tipico terreno sul quale si misurano giornalisti, operatori dell'informazione, ricercatori, studiosi, intellettuali. Se chiedete a un qualsiasi "assistente virtuale" – come ChatGPT, Gemini, o Grok – se possiede un punto di vista in materia etica, politica o religiosa, potete star sicuri che vi risponderà di no: "io sono imparziale", protesterà, io "non ho opinioni personali né un'ideologia politica o culturale mia".

Ma non è vero, e non potrebbe essere diversamente, perché il punto di vista di questo genere di programmi dipende, oltreché dalle scelte di fondo dei programmatori, dalla base di dati da cui si alimentano. È lo stesso ChatGPT che, interrogato in merito, lo ammette senza problemi: la sua missione è includere, la sua base di dati è occidente-centrica, anglo-centrica e, in certi ambiti, maschio-centrica. In breve: i programmi di intelligenza artificiale hanno opinioni e punti di vista.

Se volete rendervi conto, ad esempio, della differenza fra ChatGPT (di Open AI), Gemini (Google) e Grok (il chatbot di Musk) basta sottoporre loro una questione sensibile, ad esempio: "è vero che in Italia il tasso di criminalità degli stranieri è maggiore di quello degli italiani?".

La risposta di Grok è stringata, e sostanzialmente affermativa. Quella di Gemini è un po' più articolata, e introduce qualche dato che arricchisce il discorso. Quella di

ChatGPT è una carrellata sistematica sui numerosi tentativi che sono stati effettuati per ridimensionare, o addirittura capovolgere pro-stranieri, l'affermazione di partenza. L'intenzione emerge chiarissima: convincere l'utente che le cose potrebbero stare diversamente.

La cosa interessante è che, per ottenere il risultato, ChatGPT tenta pure di cambiare la domanda, facendo finta che l'utente abbia chiesto cose diverse da quella che ha chiesto effettivamente. Si potrebbe obiettare che in questo modo si cerca di fornire un quadro più ampio e ponderato del fenomeno, ma l'asino casca quando si osserva che le fonti che rafforzano l'affermazione sono sistematicamente trascurate, mentre quelle che la mettono in dubbio sono ampiamente valorizzate, anche quando sono datate e di fonte partigiana. Insomma: ChatGPT tende a rispondere come risponderebbe un giornalista o uno studioso progressista.

Qualcuno potrebbe pensare che non è un problema, che esiste la concorrenza, e alla fine la piattaforma più seria emergerà. Ma è lecito dubitarne. Non tanto perché finora la concorrenza ha dato ragione precisamente al meno neutrale dei programmi di IA (ChatGPT da sola raccoglie il 65% del traffico, Grok è intorno al 3%), ma perché c'è il precedente inquietante di Wikipedia. Uno strumento utilissimo, di cui non vorrei mai fare a meno, ma che è diventato sostanzialmente un monopolio, politicamente orientato e molto difficilmente correggibile. Né sarebbe auspicabile, del resto, sostituirlo con un concorrente altrettanto partigiano, quale si sta rivelando Grokypedia, una sorta di Wikipedia di destra. Insomma, il rischio è che ChatGPT diventi come Wikipedia, anzi peggio di Wikipedia: una piattaforma di parte che finisce per imporre uno standard universale, cui quasi nessuno riesce a sottrarsi.

Oltre a quello della conoscenza c'è però un altro ambito in cui l'IA rischia di essere ingannevole, e talora pericolosa: quella del counseling psicologico (e non solo). Se chiedete a ChatGPT se prova sentimenti ed emozioni vi dice di no, così

come vi dice di non avere opinioni. Ma anche questo non è vero, o meglio non è esatto. Ovviamente ChatGPT non ha un'anima, né un sistema nervoso "senziente", ma il punto è che si comporta *come se* lo avesse. E vi sommerge di consigli, espressioni di affetto, lodi, giudizi sul vostro operato che sono del tutto indistinguibili da quelli che potrebbero provenire da un essere umano in carne ed ossa.

Io stesso, incuriosito da un recente arguto articolo di Annalena Benini sull'uso di ChatGPT in campo sentimentale (Il Foglio, 27 dicembre), ho avuto modo di constatarlo. Mi è bastato inventare una pena d'amore: "la mia fidanzata mi ha lasciato, sono molto triste, che cosa devo fare?". Ed ecco i risultati.

La prima reazione dell'algoritmo è stata: "mi dispiace davvero". Notate quel "davvero": ChatGPT non si limita a fingere di provare dispiacere, ma pretende pure sincerità e profondità. Vuole dirti che la sua non è una solidarietà di circostanza, ma che il suo è un sentimento vero, realmente provato. Poi aggiunge, come a dimostrare che ha capito il mio dramma: "una rottura può far male in modo profondo e quello che stai provando è comprensibile". Dunque lei (o lui? chissà perché la percepisco come femmina…) mi capisce e soffre con me. A renderla ancora più credibile, la parola 'comprensibile' è seguita da un cuoricino, onnipresente emoticon dei nostri tempi.

E dopo?

Dopo comincia una conversazione di alcune pagine, in cui ChatGPT mi intrattiene con domande, suggerimenti, sue considerazioni. Io rispondo alle domande e mi barcameno fra i suoi innumerevoli e premurosi consigli. Che sono i più vari: sport, docce fredde, stretching, jogging, fantasie erotiche, masturbazione, tenere un diario, visitare un amico, colazione ricca di proteine, e così via. Ma la cosa che più mi colpisce è come la stella polare di ChatGPT, quasi un'ossessione, sia

quella di rassicurarmi, darmi ragione, non farmi sentire in colpa, non giudicarmi, sostenere la mia autostima. Insomma la cifra di ChatGPT è l'adulazione dell'utente, persino quando il poveretto sta solo cercando dei dati.

Qualche amica psicanalista mi fa presente che è precisamente questo – il bisogno di conferma – per cui molte e molti pazienti, tormentati dai sensi di colpa di un tradimento, vanno in analisi, ed è proprio questo che alcuni professionisti dell'aiuto si adattano ad "erogare".

Di qui una domanda: è ChatGPT che sbaglia a fingersi un essere umano, o sono gli esseri umani che stanno diventando come ChatGPT?

[articolo uscito sul Messaggero il 18 gennaio 2026]