ChatGPT – L’impostore autorevole

written by fondazioneHume | 12 Agosto 2023

Quando si parla di ChatGPT – il programma che fornisce risposte istantanee su quasi tutto lo scibile umano – di solito scatta lo schema: molto comodo e interessante, ma non infallibile. Come dire: quasi sempre ChatGPT fornisce una risposta corretta, ma alle volte può incorrere in infortuni più o meno clamorosi. Quindi state attenti, controllate anche altre fonti, eccetera.

Questa visione delle capacità e dei limiti di ChatGPt è ancora quella dominante in Italia, e in parte anche all’estero. Ma è profondamente errata. Del tutto errata, direi. Perché presuppone che Chat (d’ora in poi userò l’abbreviazione) sia programmato per soddisfare un utente che cerca la verità, solo la verità, e non desidera ricevere informazioni false.

Questa, alla luce del funzionamento effettivo di Chat, è una credenza decisamente ingenua. Come ha notato qualche mese fa Tim Harford sul Financial Times, Chat non è programmato per generare affermazioni vere, ma per fornire risposte verosimili, ossia risposte che possano essere *credute* vere, anche a costo di inventarle di sana pianta. In questo senso, nota Tim Harford, Chat è la perfetta realizzazione di un concetto messo a punto dal filosofo statunitense Harry Frankfurt in un celebre saggio-pamphlet degli anni ’80, significativamente intitolato *Bullshit* (letteralmente: stronzate): quello della proliferazione incontrollata di affermazioni campate per aria ma verosimili, ovvero plausibili.

Ecco, in passato eravamo soli a raccontare, millantare, inventare storie, per gli scopi più diversi: far impressione su una ragazza, essere ammirati dai nostri amici, mostrare

competenza davanti ai colleghi, in generale intrattenere gli astanti. Ora non più, ora interagiamo con un software che si comporta con noi come noi ci comportavamo con i nostri interlocutori. Lo scopo di Chat non è dirci la verità, ma farci credere di conoscerla. Impressionarci con la sua competenza. Catturare la nostra fiducia e la nostra attenzione, come peraltro si intuisce dallo stile estremamente accattivante, amichevole, personalizzato, gentile, per non dire seduttivo, con cui interagisce con noi.

Mi rendo conto che le mie sono affermazioni molto forti, e poco condivise (almeno in Italia). Ma ho le prove. Un mare di prove. In questo primo articolo su Chat ne fornirò un piccolo campionario, raccontando rapidamente l'esito di alcune interrogazioni.

Con alcuni amici professori universitari abbiamo provato a interrogare Chat su noi stessi e le nostre pubblicazioni. Ebbene, il risultato tipico sono notizie biografiche (data e luogo di nascita) del tutto false, e una lista di pubblicazioni (con tanto di rivista, numero di pagine, eccetera), tutte o quasi tutte inesistenti. Ma, attenzione: quando interrogo Chat su me stesso, i miei campi di interesse sono ben individuati, e i libri e gli articoli che mi vengono attribuiti potrei benissimo averli scritti io. Insomma non c'è un bit di verità nel profilo che Chat mi attribuisce, ma non c'è nulla di inverosimile.

O meglio: non c'è nulla di inverosimile nella lista delle mie pubblicazioni finché la interrogazione avviene da una postazione a Cambridge, in Inghilterra. Ma se ripeto l'interrogazione da una postazione in Italia, escono libri che non solo non ho scritto, ma non avrei mai potuto scrivere: ad esempio: *Volemose bene. Chi ha detto che gli italiani sono furbi?*, e *Amore liquido all'italiana*. Posso solo congetturare che una routine "generativa" di Chat mi abbia classificato come giornalista (anziché come sociologo e docente di Analisi dei dati) e mi abbia assegnato pubblicazioni nello stile di

certo giornalismo creativo.

Più interessante il caso di mia moglie Paola Mastrocola, che di mestiere fa la scrittrice. Qui Chat sembra adottare la tecnica dei pentiti quando vogliono depistare le indagini, cioè: mescolare fatti veri con fatti inventati ma verosimili. I romanzi citati sono tre: uno vero (*Non so niente di te*), uno esistente ma di un'altra scrittrice italiana (*Lei così amata*, di Melania Mazzucco), l'altro anch'esso esistente ma della scrittrice britannica Kerry Hudson (*Tutti gli uomini di mia madre*).

Potrei continuare con altri esempi. Se chiedi una bibliografia su un dato argomento, può succederti di ottenere una lista con decine di saggi inesistenti, eppure indicati con precisione per rivista, data, numero di pagine. Se chiedi qual è l'articolo di economia più citato di tutti i tempi (ci ha provato l'economista David Smerdon), puoi ottenere un titolo del tutto plausibile (*A Theory of Economic History*), con due autori a loro volta plausibili (Douglas North e Robert Thomas), e una rivista ospite più che appropriata (*Journal of Economic History*), salvo scoprire che l'articolo non è mai stato scritto. Non esiste. È la risultante di una routine di Chat, che ha messo insieme elementi effettivamente esistenti per combinarli in una risposta plausibile, senza alcun riguardo alla verità della risposta stessa.

E non è tutto. Ho provato a interrogare Chat sul pensiero di alcuni autori italiani controversi, come Pasolini e don Milani. Il risultato è sconcertante. Le risposte di Chat sembrano il risultato di un processo deduttivo in due stadi: *primo*, si classifica l'autore dal punto di vista ideologico-culturale; *secondo*, gli si attribuiscono i pensieri che è ragionevole attendersi in base a come è stato classificato. Pasolini era progressista, quindi era a favore del divorzio (sappiamo invece che è vero il contrario). Don Milani era un educatore illuminato, quindi doveva amare la letteratura antica e lo sport come strumenti di crescita personale (anche

qui sappiamo che non è così).

Il caso di Don Milani è particolarmente interessante, perché Chat non solo risponde in modo errato a domande specifiche sul priore di Barbiana, ma ne traccia un profilo del tutto fantasioso. Dopo avergli assegnato tutte le credenze tipiche delle pedagogie progressiste dei nostri giorni, ne segnala il libro *Lettera a un professore*, che conterrebbe un dialogo con lo storico Adolfo Omodeo. Rimbrottato da me, Chat mi risponde scusandosi, ammettendo che Don Milani non ha affatto scritto *Lettera a un professore*, ma si guarda bene dal segnalare che – invece – ha scritto *Lettera a una professoressa*.

Tornerò, in altri articoli, sul funzionamento di ChatGPT. Quel che vorrei sottolineare fin da ora, però, è il suo status epistemico: ChatGPT non è un algoritmo che persegue più o meno imperfettamente la verità, ma un dispositivo che – quando non conosce la verità – si comporta come un affabulatore (Treccani: Affabulatore = persona che narra in maniera affascinante e abile, o che racconta storie affascinanti ma poco fondate o totalmente infondate). In poche parole: un impostore. O meglio: un impostore autorevole, che tale resterà finché ci ostineremo a credergli.