

Intervistare Claude è pericoloso se le domande sono sbagliate



Nei giorni scorsi è apparsa sul *Corriere della Sera* un'intervista di Walter Veltroni a Claude, una delle piattaforme di intelligenza artificiale più conosciute. Le domande riguardavano cose del tipo l'infanzia che Claude non ha avuto, il suo rapporto con Dio, la paura della morte, il desiderio di vedere il mare. L'intervista, io direi il pezzo, dato che non c'era nessuno da intervistare, ha suscitato reazioni che vanno da chi ha apprezzato il garbo delle domande e delle risposte a chi ha colto l'occasione per ridicolizzare Veltroni, rilanciando l'etichetta dell'intervista a un pappagallo stocastico.

Vorrei provare a discuterne in un altro modo. Quell'intervista non è né geniale né ridicola. È, prima di tutto, in ritardo. È un gioco che chiunque abbia avuto accesso a un modello linguistico alcuni anni fa ha provato a fare nei primi giorni:

chiedergli se ha emozioni, se sogna, se teme di morire. Era la fase dello stupore iniziale, ma erano i primi giorni. **Pubblicare oggi sul principale quotidiano italiano un'intervista costruita su quelle stesse domande è il sintomo di un dibattito pubblico che spesso affronta l'AI ancora nella fase dello stupore, anziché con gli strumenti culturali e scientifici necessari per comprenderne l'impatto.**

Ma c'è qualcosa di più pericoloso del ritardo. Quell'intervista incarna, in forma quasi pura, la prima delle domande sbagliate che dominano il dibattito pubblico sull'AI: le macchine pensano come noi? Tiene l'attenzione inchiodata su una domanda fuorviante, trasformando una questione scientifica e tecnologica in un dilemma inesistente. L'intervista non solo cade in questo tranello, ma lo mette anche in scena. Chiede al modello di rispondere, generando **un cortocircuito che alimenta l'antropomorfizzazione e impedisce di capire che cosa sia davvero Claude.**

Per uscire da questa illusione bisogna guardare cosa sono i **grandi modelli linguistici come Claude. Sono sistemi addestrati a prevedere la continuazione più probabile di una sequenza linguistica. Sono modelli di AI predittiva. La coerenza delle loro risposte non è il segno di una mente che parla, ma un fenomeno emergente di natura statistica che si manifesta dopo l'addestramento su enormi quantità di testi.** Riconoscere che, all'interno del modello, non c'è un soggetto, ma un mondo di algoritmi, non significa banalizzare il fenomeno. Il punto interessante, infatti, non è ridurre questi sistemi al loro meccanismo di base. È capire come, da quel processo ripetuto su larga scala e con l'aiuto di montagne di dati, emergano comportamenti che assomigliano alla comprensione, al ragionamento e alla conversazione. Questo è un problema scientifico aperto, della stessa natura di quello che si pone osservando una colonia di formiche: regole locali semplici che danno luogo a un'intelligenza collettiva di cui nessuna singola formica è consapevole. In questo senso,

chiedere a Claude se ha un'anima è come chiedere a una colonia di formiche se sia felice e consapevole della propria intelligenza collettiva. La domanda non ha senso e oscura proprio ciò che andrebbe compreso.

La cornice scientifica, volenti o nolenti, è l'unica a produrre conoscenza e, di conseguenza, capacità di intervento. Per esempio, questi sistemi sbagliano in modi specifici e generano contenuti plausibili ma falsi secondo dinamiche che possiamo studiare. Sono fatti misurabili e governabili, ma solo se accettiamo che ciò che abbiamo davanti è un fenomeno emergente da comprendere, non un soggetto da intervistare. E una volta acquisito questo sguardo, il problema si sposta sulle vere domande da porci dal punto di vista della società. Sempre di più, ciò che leggiamo, scriviamo, impariamo e decidiamo passa attraverso questi sistemi. Da questa interazione su larga scala emergono dinamiche che nessuno ha pianificato né controlla pienamente. Lo stesso sguardo scientifico che serve a capire un singolo modello serve, a una scala diversa, a comprendere questa nuova società ibrida, fatta di uomini e AI in interazione. È la stessa famiglia di domande, applicata a livelli diversi.

Ecco perché l'intervista di Veltroni non è un episodio innocuo. Non perché sia mal fatta, ma perché legittima, con il garbo delle parole e il romanticismo del racconto, una cornice che il dibattito serio ha già abbandonato da tempo. E lo fa proprio nel momento in cui chi studia questi sistemi sta cercando, faticosamente, di creare consapevolezza pubblica e di costruire il vocabolario delle domande giuste. **Domande che non fanno parlare l'AI di sé, ma ci impongono di riflettere sulla società che stiamo costruendo e su quali infrastrutture pubbliche servono per comprendere un sistema complesso che abbiamo creato, ma che non comprendiamo ancora a sufficienza. È lì che il giornalismo serve. Ed è lì che il ritardo italiano comincia a diventare un problema.**