

26. dicembre

Elogio e necessità dell'ignoranza

Attenzione: leggere può danneggiare la tua ignoranza



Voltaire nelle sue opere racconta dei tentativi della ragione umana per liberarsi dai pregiudizi. Nel 1777 quando incontrò Filippo D'Orleans reggente del re di Francia ironicamente commentò *"Deve essere molto ignorante perché risponde ad ogni domanda che gli viene fatta"*

Mentre un sapiente è consapevole della piccolezza del suo sapere, un ignorante non conosce l'enorme estensione della sua ignoranza. L'ignoranza è oggettivamente pericolosa, ma l'ignoranza dell'ignoranza di più. Un tempo l'ignoranza era assenza di sapere; ora è, prevalentemente, presenza di un sapere "presunto", per di più erroneo. I palinsesti televisivi sono programmi di approfondimento dell'ignoranza.

In un'epoca iperconnessa in cui la ricchezza della conoscenza umana è a portata di mano, l'ignoranza difficilmente può essere considerata una virtù. Nonostante il clamore e il terrore diffusi sui pericoli dell'IA, i modelli attuali sono per molti versi sviluppati ***in modo simile a come viene addestrato un animale.***

Un sistema di IA come ChatGPT potrebbe avere accesso a quantità inimmaginabili di informazioni, ma richiede ***l'addestramento da parte degli esseri umani*** su quali informazioni siano preziose o meno, se abbia compreso correttamente la richiesta ricevuta e se la sua risposta sia corretta.

Il paradigma prevalente è che il feedback umano aiuti l'IA ad affinare il suo modello attraverso feedback positivi per le risposte corrette e feedback negativi per le risposte errate, in modo che mantenga qualsiasi metodo abbia portato a feedback positivi e modifichi qualsiasi metodo abbia portato a feedback negativi. In realtà non è così lontano da come vengono addestrati gli animali.

Tuttavia, sta emergendo un problema.

I sistemi di intelligenza artificiale sono diventati abili nel dare risposte coerenti e convincenti che tuttavia sono **del tutto errate.**
Come è successo?

Approfondendo il metodo di addestramento per l'IA, i ricercatori hanno scoperto che gli umani che addestravano l'IA segnalavano le risposte "Non lo so" ***come insoddisfacenti.*** Da un certo punto di vista, questo ha senso. L'intero scopo di questi sistemi è fornire risposte, dopotutto. Ma anziché far sì che l'IA tornasse e ripensasse ai suoi dati, sviluppava invece risposte sempre più convincenti che non erano affatto vere, al punto che i supervisori umani non segnalavano risposte

sufficientemente convincenti come sbagliate perché loro stessi non si rendevano conto che erano sbagliate. Il risultato è che **"più è difficile la domanda e più avanzato il modello che usi, più è probabile che tu ottenga una risposta assurda, plausibile e ben confezionata"**.

Scoprire qualcosa di ciò che sta accadendo nei sistemi di IA dissipa sia il *fervente clamore* che l'intelligenza artificiale potrebbe essere la *nostra salvatrice*, sia la *profonda paura* che potrebbe essere la *nostra rovina sociale*. Questo è uno strumento; è bravo in alcuni compiti e meno bravo in altri. E, come tutti gli strumenti, non ha una moralità intrinseca. Se venga usato per il bene o per il male dipende dall'approccio degli umani che lo usano.

Ma gli studi svelano anche il nostro rapporto teso con l'ignoranza. I problemi sorgono nelle risposte fornite da sistemi come *ChatGPT* perché una risposta convincente è più apprezzata dell'ammettere l'ignoranza, anche se la risposta convincente non è affatto corretta. Poiché l'IA è stata addestrata a *evitare di ammettere di non sapere qualcosa*, tutte le sue risposte sono meno affidabili, anche quelle che sono effettivamente corrette.

Questo non è un problema limitato all'intelligenza artificiale.

Avevo un amico (Salvatore) che sembrava incapace di ammettere di non sapere qualcosa, e ogni volta che veniva corretto da qualcun altro, faceva sembrare che la sua prima risposta fosse effettivamente quella corretta, piuttosto che qualsiasi cosa avesse detto. Non so quanto fosse consapevole di averlo fatto, ma il risultato era che non mi fidavo particolarmente di qualsiasi cosa dicesse come se fosse corretta. Paradossalmente, se avesse ammesso la sua ignoranza più prontamente, avrei creduto che fosse meno ignorante.

È strano che *l'ammissione di ignoranza* venga così evitata. Dopotutto, è per molti versi il nostro stato predefinito. Nessuno biasima un neonato o un bambino per non sapere le cose. Se non altro, ci aspettiamo che l'ignoranza sia un potenziale carburante per la curiosità.

La nostra ricerca di risposte inizia nel riconoscimento di non averle. E in un'epoca in cui circa 500 ore di video vengono caricate su YouTube ogni minuto, la somma di ciò che non sappiamo deve per forza essere di gran lunga maggiore di tutto ciò che sappiamo. Ciò che ognuno di noi può sapere è solo una piccola frazione di tutto ciò che c'è da sapere.

Uno dei doni della teologia cristiana è la capacità di riconoscere ciò che ci rende umani. Una di queste cose è il fatto che ogni cosa creata è, per definizione, limitata.



Solo **Dio** è l'unico che può essere descritto dall'"omnis". È onnipotente, onnipresente e onnisciente. Non c'è limite al suo potere, alla sua presenza e alla sua conoscenza. La distinzione tra creatore e creazione significa che le cose create hanno limiti al loro potere, alla loro presenza e alla loro conoscenza. Non possiamo fare tutto ciò che vogliamo. Non possiamo essere ovunque allo stesso tempo. E non possiamo sapere tutto ciò che c'è da sapere. Proiettare o ipotizzare una conoscenza infinita è essenzialmente affermare di essere Dio.

Ammettere la nostra ignoranza è quindi semplicemente riconoscere la nostra natura di esseri creati, riconoscendoci a vicenda che non siamo Dio e quindi non possiamo sapere tutto. Ma, cosa fondamentale, **ammettere di non sapere tutto non è la stessa cosa che dire di non sapere nulla** .

La nostra natura data da **Dio** è una natura di scoperta e apprendimento. A volte mi piace immaginare la gioia di Dio nella nostra scoperta di qualche aspetto precedentemente sconosciuto della sua creazione, mentre riesce a condividere con noi tutto ciò che ha creato.

Forse ciò che conta davvero è cosa facciamo con la nostra ignoranza. Resteremo semplicemente soddisfatti di non sapere, o ci volgerà verso l'esterno per deliziarsi nelle cose nuove che si nascondono dietro ogni angolo?

Per gli sviluppatori di **ChatGPT** e simili, c'è anche un promemoria qui che non dovremmo aspettarci che l'IA assuma gli attributi di **Dio**. L'IA usata bene nelle mani degli umani potrebbe ancora fare cose straordinarie per noi, ma non sarà veramente in grado di fare nulla, essere ovunque o sapere tutto. Forse se fosse addestrata a dire "non lo so" un po' di più, potremmo tutti imparare un po' di più sulla natura del mondo che **Dio** ha creato.

Ammettere la propria ignoranza è un segno di intelligenza.

