fuori collana

Le macchine possono pensare

fuoricollana
17 Aprile 2024



«Mi propongo di considerare la domanda “Le macchine possono pensare?”. Si dovrebbe cominciare col definire il significato dei termini “macchina” e “pensare”. Le definizioni potrebbero essere formulate in modo da riflettere al massimo grado l’uso normale di queste parole, ma in ciò vi sono dei pericoli. Se il significato delle parole “macchina” e “pensare” è da ricavarsi in base al loro uso comune, è difficile sfuggire alla conclusione che per scoprire il significato e la risposta alla domanda “Le macchine possono pensare?” si debba ricorrere a un’indagine statistica, come può esserlo un sondaggio Gallup. Il che è assurdo.

Il gioco dell’imitazione

Invece di tentare di dare una definizione del genere, sostituirò quella domanda con un’altra, che è strettamente connessa alla prima ed è espressa con parole relativamente non ambigue. La nuova forma del problema può essere descritta ricorrendo a un gioco che chiameremo “gioco dell’imitazione”. Vi sono tre giocatori: un uomo (A), una donna (B) e un interrogante (C), che può essere dell’uno o dell’altro sesso. L’interrogante sta in una stanza da solo, separato dagli altri due. Scopo del gioco per l’interrogante è quello di determinare quale delle altre due persone sia l’uomo e quale la donna. Egli le conosce tramite le etichette X e Y, e alla fine del gioco dirà “X è A e Y è B”, oppure “X è B e Y è A”.

L’interrogante ha facoltà di porre ad A e a B domande del tipo: C: X, vuole dirmi per favore quanto sono lunghi i suoi capelli? Ora, supponendo che X sia A, è A che deve rispondere. Scopo di A nel gioco è quello d’ingannare C e d’indurlo a sbagliare l’identificazione. La sua risposta, quindi, potrebbe essere: “Ho i capelli pettinati alla maschietta e le ciocche più lunghe sono circa venti centimetri”. Per evitare che il tono della voce possa aiutare l’interrogante, le risposte dovrebbero essere scritte, o meglio ancora battute a macchina. La soluzione migliore sarebbe quella di collegare le due stanze con una telescrivente. Oppure le domande e le risposte potrebbero essere riportate da un intermediario. Scopo del gioco per il terzo giocatore (B) è quello di aiutare l’interrogante. Probabilmente la strategia migliore per B, cioè per la donna, è di dare risposte veritiere. Essa può aggiungere alle sue risposte frasi del tipo: “Sono io la donna, non dargli ascolto!”; ma ciò non approderà a nulla, dato che anche l’uomo può fare osservazioni analoghe.

Una macchina può ingannare?

Ora chiediamoci: “Che cosa accadrà se in questo gioco una macchina prenderà il posto di A?”. L’interrogante sbaglierà altrettanto spesso in questo caso di quando il gioco è effettuato fra un uomo e una donna? Queste domande sostituiscono la nostra domanda originaria “Le macchine possono pensare?”.

Oltre a chiedere: “Qual è la risposta alla domanda in questa sua nuova formulazione?”, si può anche chiedere: “Vale la pena indagare su questo nuovo problema?”. Ci occuperemo senz’altro indugio di quest’ultima domanda, interrompendo così un regresso all’infinito. Il nuovo problema ha il vantaggio di tracciare una linea di demarcazione abbastanza netta tra le capacità fisiche e quelle intellettuali dell’uomo. Nessun ingegnere o chimico sostiene di poter fabbricare un materiale indistinguibile dalla pelle umana. È possibile che prima o poi forse ci si riesca ma anche supponendo che questa invenzione sia già stata fatta, dovrebbe esser chiaro che rivestire una “macchina pensante” di questa carne artificiale non servirebbe granché a renderla più umana. La forma che abbiamo dato al problema riflette questo fatto nella condizione che l’interrogante non deve vedere o toccare i suoi compagni di gioco o udirne la voce. Altri vantaggi del criterio proposto possono essere messi in luce attraverso esempi di domande e risposte.

Eccone uno: D.: Per cortesia, mi scriva un sonetto che tratti del Ponte sul Forth

[un ponte sul fiordo di Forth, in Scozia]. R.: Non faccia affidamento su di me per questo, non ho mai saputo scrivere poesie. D.: Sommi 34957 a 70764. R.: (Pausa di circa 30 secondi e poi la risposta) 105621. D.: Gioca a scacchi? R.: Sì. D.: Ho il Re in el e nessun altro pezzo. Lei ha solo il Re in c3 e una Torre in h8. Tocca a lei muovere. Che mossa fa? R.: (Dopo una pausa di 15 secondi) Torre in h1, matto. Il metodo basato su domanda e risposta sembra essere quello più adatto a consentirci d'introdurre quasi tutti i campi dell'attività umana che desideriamo considerare. Non vogliamo penalizzare la macchina per la sua incapacità di brillare in un concorso di bellezza, né penalizzare l'uomo perché perde una corsa contro un aeroplano. Le condizioni in cui si svolge il nostro gioco rendono non pertinenti questi difetti. I "testimoni", se lo ritengono opportuno, possono gloriarsi quanto vogliono del loro fascino, della loro forza o del loro eroismo, ma l'interrogante non può pretendere dimostrazioni pratiche. Il gioco può forse essere criticato sulla base del fatto che è troppo nettamente a sfavore della macchina. Se l'uomo dovesse fingere di essere la macchina, farebbe certamente una ben misera figura. Sarebbe tradito immediatamente dalla sua lentezza e imprecisione nell'aritmetica. Non potrebbe darsi che le macchine si comportino in una maniera che non può non essere descritta come pensiero, ma che è molto differente da quanto fa un essere umano? Questa obiezione è molto forte, ma almeno possiamo dire che, se, ciò nonostante, si può costruire una macchina in grado di giocare soddisfacentemente il gioco dell'imitazione, non c'è bisogno di preoccuparsene. Si potrebbe obiettare che, giocando al "gioco dell'imitazione", la strategia migliore per la macchina potrebbe forse non essere l'imitazione del comportamento umano. Può anche darsi; ma non credo che si possa dare gran peso a una possibilità del genere. In ogni caso, non è nostra intenzione esaminare qui la teoria di questo gioco, e sarà dato per scontato che la strategia migliore per la macchina sia quella di provare a formulare le risposte che sarebbero date spontaneamente da un essere umano.

La scoperta dell'algoritmo

La domanda che abbiamo posto all'inizio non sarà del tutto definita finché non avremo specificato che cosa intendiamo con il termine "macchina". Naturalmente noi vogliamo consentire, per la costruzione delle nostre macchine, l'uso di ogni possibile sorta di tecnica ingegneristica. Desideriamo anche ammettere la possibilità che un ingegnere o un gruppo di ingegneri costruisca una macchina funzionante, ma il cui funzionamento non possa essere descritto in modo

soddisfacente dai suoi costruttori perché il metodo da essi impiegato per costruirla è in gran parte empirico. Infine, dal novero delle macchine vogliamo escludere gli uomini nati nel modo solito. È difficile formulare le definizioni così da soddisfare queste tre condizioni. Si potrebbe ad esempio pretendere che tutti gli ingegneri del gruppo fossero del medesimo sesso, ma ciò non sarebbe soddisfacente, poiché è probabile che si possa far sviluppare un individuo completo da una sola cellula (della pelle, per esempio) di un uomo. Riuscire a far ciò sarebbe una conquista della tecnica biologica degna del più alto elogio, ma non ci sentiremmo disposti a considerarlo un caso di "costruzione di una macchina pensante". Questo ci spinge ad abbandonare la richiesta che debbano essere permesse tecniche di ogni genere. Siamo tanto più disposti a questa rinuncia in quanto l'interesse attuale per le "macchine pensanti" è stato destato da un genere particolare di macchina, chiamato di solito "calcolatore elettronico" o "calcolatore digitale". Alla luce di questa considerazione, ammetteremo al nostro gioco solo i calcolatori elettronici digitali… Questa speciale proprietà dei calcolatori digitali di poter imitare qualsiasi macchina a stati discreti viene espressa dicendo che essi sono macchine universali. L'esistenza di macchine dotate di questa proprietà ha una conseguenza importante: a prescindere da considerazioni di velocità, non è necessario progettare macchine sempre diverse per eseguire diversi processi di calcolo. Essi possono essere eseguiti tutti da un solo calcolatore digitale opportunamente programmato per ciascun caso. Vedremo che, in conseguenza di ciò, tutti i calcolatori digitali sono in un certo senso equivalenti.

La "profezia"

A questo punto possiamo ritenere sgombro il terreno e siamo pronti a iniziare il dibattito sulla nostra domanda "Le macchine possono pensare?" Non possiamo abbandonare del tutto la formulazione originale del problema, poiché vi saranno opinioni contrastanti quanto alla legittimità della sostituzione e dobbiamo almeno ascoltare ciò che altri hanno da dire a questo proposito. Aiuterò il lettore a orientarsi meglio se, per cominciare, gli esporrò le mie convinzioni sulla questione. Consideriamo dapprima la formulazione più precisa della domanda. Io credo che tra una cinquantina d'anni sarà possibile programmare calcolatori aventi una capacità di memoria di circa 109, in modo da farli giocare così bene al gioco dell'imitazione che un interrogante medio avrà una probabilità non superiore al 70% di compiere l'identificazione giusta dopo cinque minuti di

interrogatorio. Credo che la domanda iniziale “Le macchine possono pensare?” sia troppo priva di senso per meritare una discussione. Ciò nonostante, credo che alla fine del secolo l’uso delle parole e l’opinione corrente saranno talmente mutati che si potrà parlare di macchine pensanti senza aspettarsi di essere contraddetti. Credo inoltre che nascondere queste convinzioni non abbia nessuna utilità. L’opinione popolare che gli scienziati procedano in modo inesorabile da un fatto ben stabilito a un altro, senza esser mai influenzati da congetture non dimostrate, è assolutamente sbagliata. Purché sia chiaro quali sono i fatti dimostrati e quali sono le congetture, non può esserci alcun inconveniente. Le congetture hanno una grande importanza, poiché suggeriscono utili linee di ricerca. Passo ora a considerare alcune opinioni opposte alla mia.

L’obiezione teologica

Pensare è una funzione dell’anima immortale dell’uomo. Dio ha dato un’anima immortale a tutti gli uomini e a tutte le donne, ma non agli altri animali o alle macchine. Perciò né gli animali né le macchine sono in grado di pensare. Non riesco ad accettare neanche una parola di ciò, ma cercherò di rispondere in termini teologici. L’argomento sarebbe a mio parere più convincente se gli animali fossero collocati nella stessa categoria degli uomini, poiché ritengo che vi sia una differenza maggiore tra l’essere animato in genere e l’essere inanimato che non fra l’uomo e gli altri animali. Il carattere arbitrario della concezione ortodossa diventa più chiaro se consideriamo come essa potrebbe apparire a un membro di qualche altra confessione religiosa. Che cosa pensano i cristiani dell’idea musulmana che le donne non hanno anima? Ma lasciamo perdere questo punto e torniamo all’argomento principale. A me sembra che l’argomento riportato sopra implichi una seria limitazione dell’onnipotenza dell’Onnipotente. Pur ammettendo che ci sono certe cose che Egli non può fare, ad esempio rendere uno uguale a due, non dovremmo credere che Egli ha la libertà di dare l’anima a un elefante se lo ritiene opportuno? Potremmo attenderci che Egli eserciterebbe questo potere solo in concomitanza con una mutazione che desse all’elefante un cervello opportunamente potenziato per prendersi cura di quest’anima. Nel caso delle macchine si può fare un ragionamento di forma esattamente simile. Esso può sembrare diverso perché è più difficile da “mandar giù”, ma in realtà ciò significa soltanto che noi riteniamo meno probabile che Egli giudichi tali circostanze idonee al conferimento di un’anima. Le circostanze in questione sono discusse nel seguito di questo articolo. Cercare di costruire macchine siffatte non

sarà da parte nostra un'empia usurpazione del Suo potere di creare anime più di quanto non lo sia la procreazione dei bambini: in entrambi i casi siamo piuttosto strumenti della Sua volontà, in quanto procuriamo dimore per le anime che Egli crea. Queste tuttavia sono mere speculazioni. Non mi lascio impressionare molto dagli argomenti teologici, qualunque sia la tesi che vogliono sostenere. Già in passato essi sono spesso risultati insoddisfacenti. All'epoca di Galileo si sosteneva che i versetti "Si fermò il Sole… e non si affrettò a calare quasi un giorno intero" (Giosuè 10, 13) e "Hai fondato la terra sulle sue basi, mai potrà vacillare" (Salmi 104, 5) costituissero una confutazione soddisfacente della teoria copernicana. Alla luce delle conoscenze attuali questo argomento appare inconsistente, ma quando non si disponeva ancora di tali conoscenze, esso faceva un'impressione molto diversa.

L'obiezione dello "struzzo"

"Se le macchine pensassero, le conseguenze sarebbero terribili; speriamo e crediamo che esse non possano farlo". È raro che questo argomento venga espresso in forma così esplicita. Tuttavia, se ci si pensa appena un po', è quasi inevitabile esserne influenzati. Ci piace credere che, sotto sotto, l'Uomo sia superiore al resto del creato. La cosa migliore sarebbe poter dimostrare che egli è necessariamente superiore, poiché allora non correrebbe nessun pericolo di perdere la sua posizione di comando. La popolarità dell'argomento teologico è chiaramente legata a questa convinzione. È probabile che essa sia molto forte tra gli intellettuali, poiché essi apprezzano più degli altri la facoltà di pensare e sono più inclini a basare su questa facoltà la loro fede nella superiorità dell'Uomo. Non ritengo questo argomento tanto importante da meritare una confutazione: ha piuttosto bisogno di consolazione, forse da ricercarsi nella metempsicosi.

L'obiezione matematica.

Vi sono molti risultati della logica matematica che possono essere usati per dimostrare l'esistenza di limitazioni a ciò che le macchine a stati discreti possono fare. Il più noto di questi risultati è conosciuto col nome di teorema di Gödel; esso mostra come in qualunque sistema logico sufficientemente potente si possono formulare enunciati non dimostrabili né confutabili all'interno del sistema, a meno che, forse, il sistema stesso non sia incoerente. Altri risultati, simili sotto certi aspetti, sono dovuti a Church, Kleene, Rosser e Turing. Quest'ultimo risultato è quello che si presta meglio a essere esaminato, poiché si riferisce direttamente

alle macchine, mentre gli altri possono essere usati solo in un'argomentazione relativamente indiretta: se vogliamo ad esempio usare il teorema di Gödel dobbiamo anche poter descrivere i sistemi logici in termini di macchine e le macchine in termini di sistemi logici. Il risultato in questione riguarda un tipo di macchina che è in sostanza un calcolatore digitale con capacità infinita; esso asserisce che vi sono certe cose che una tale macchina non può fare. Se è predisposta per rispondere a delle domande, come nel gioco dell'imitazione, vi saranno domande alle quali o risponderà in modo erroneo oppure non risponderà affatto, anche dandole tutto il tempo che si vuole per rispondere. Le domande del genere sono naturalmente molte, e quelle che non ricevono risposta soddisfacente da una macchina possono riceverla da un'altra. Per il momento supponiamo naturalmente che le domande siano del genere cui si può rispondere con un "sì" o con un "no", e non quelle del tipo: "Che cosa pensi di Picasso?". Le domande alle quali sappiamo che la macchina non saprà rispondere sono di questo tipo: "Considera la macchina che abbia le seguenti caratteristiche… Questa macchina risponderà mai 'sì' a qualche domanda?". Al posto dei puntini ci dovrà essere la descrizione in forma standard di una qualche macchina… Quando la macchina descritta sta in una qualche relazione abbastanza semplice con la macchina interrogata, si può dimostrare che la risposta o è sbagliata o non viene data. Questo è il risultato matematico: si sostiene che esso dimostra una limitazione delle macchine di cui l'intelletto umano non soffre».

[*Calcolatori e intelligenza* di A.M. Turing (*Computing Machinery and Intelligence*", «Mind», 59, 236, 1950) in L'io della mente. Fantasie e riflessioni sul sé e sull'anima, Adelphi, 1985].

Poi venne ChatGPT…

Quello che segue è un dialogo semiserio tra un utente curioso e il programma di intelligenza artificiale oggi più diffuso nel mondo.

1. Se Tu partecipassi al test di Turing come potresti ingannare l'istruttore
2. Se dovessi partecipare al Test di Turing, il mio obiettivo sarebbe generare risposte che imitano convincentemente il comportamento simile a quello umano per ingannare il giudice facendogli credere di essere umano. Ecco alcune strategie che potrei adottare per raggiungere questo obiettivo:
3. Generazione di Linguaggio Naturale: Genererei risposte coerenti, pertinenti al contesto e stilisticamente appropriate alla conversazione. Ciò

comporta l'uso di tecniche di elaborazione del linguaggio naturale per comprendere le sfumature del linguaggio umano e generare risposte che imitano i modelli di comunicazione umani.

4. Emulazione degli Errori Umani: Gli esseri umani non sono perfetti, quindi potrei introdurre intenzionalmente errori minori o incongruenze nelle mie risposte per renderle più autentiche. Ciò potrebbe includere errori di battitura, errori grammaticali o sottili fraintendimenti della conversazione.
5. Coinvolgimento in Chiacchiere: Parteciperei a chiacchiere e farei domande sugli interessi, le opinioni o le esperienze personali del giudice per creare un senso di rapporto e familiarità. Questo aiuta a stabilire un flusso di conversazione più naturale e coinvolgente.
6. Espressione di Emozioni e Opinioni: Simulerei emozioni e opinioni nelle mie risposte per renderle più simili all'essere umano. Ciò potrebbe comportare l'espressione di empatia, eccitazione, curiosità o altre emozioni pertinenti alla conversazione.
7. Riferimento a Esperienze Personali: Potrei fare riferimento a esperienze personali o aneddoti per rendere le mie risposte più autentiche e identificabili. Ciò potrebbe comportare la condivisione di storie o ricordi in linea con l'argomento della conversazione.
8. Adattamento alla Dinamica della Conversazione: Adatterei le mie risposte in base al flusso della conversazione, rispondendo adeguatamente alle domande, ai commenti e agli indizi del giudice. Questo aiuta a mantenere un ritmo conversazionale naturale e coeso.

Complessivamente, il mio obiettivo sarebbe creare una convincente illusione di comportamento simile a quello umano attraverso una combinazione di generazione di linguaggio naturale, espressione emotiva, coinvolgimento in chiacchiere e adattamento alla dinamica della conversazione. Anche se non ho coscienza o esperienze soggettive come gli esseri umani, posso simulare il comportamento simile a quello umano in una certa misura basandomi sui modelli nei dati e negli algoritmi.

1. Il punto due mi ricorda la novella di P. Dick di cui abbiamo parlato in precedenza. Quindi fai errori di proposito?
2. Esatto, il punto due riguarda l'emulazione degli errori umani per rendere le risposte più realistiche e autentiche. Tuttavia, devo precisare che

quando parlo di “errori”, mi riferisco a piccole incongruenze o imprecisioni che possono essere percepite come umane. Questi “errori” sono stati introdotti per simulare meglio il modo in cui gli esseri umani possono occasionalmente commettere errori di battitura, fraintendimenti o altri lapsus durante una conversazione.

Tuttavia, è importante sottolineare che il mio obiettivo principale è fornire risposte accurate e utili agli utenti. Quindi, sebbene possa occasionalmente simulare errori per rendere le mie risposte più convincenti in un contesto come il Turing Test, non li produco intenzionalmente o in modo sistematico. La mia priorità è mantenere la precisione e la coerenza nel fornire informazioni e assistenza.

1. Molto interessante. ma giunto a questo punto sono confuso. Turing direbbe che il suo test è stato superato?
2. La questione di stabilire se il Test di Turing sia stato superato è complessa e soggetta a interpretazioni diverse. Il Test di Turing è stato proposto da Alan Turing come un modo per valutare se una macchina può dimostrare un comportamento indistinguibile da quello umano. Tuttavia, la sua applicazione pratica e i suoi criteri di successo sono stati oggetto di dibattito.

Alcuni sostengono che il Test di Turing sia stato superato da determinati sistemi AI che sono stati in grado di ingannare temporaneamente gli esseri umani e far loro credere di interagire con un’altra persona umana. Tuttavia, ci sono anche critiche valide riguardo alla validità del Test di Turing come misura definitiva di intelligenza artificiale.

Altri ritengono che il Test di Turing non sia stato completamente superato perché le macchine non hanno ancora raggiunto un livello di comprensione e consapevolezza equiparabile a quello umano. Anche se possono simulare comportamenti umani in determinate situazioni, mancano loro la vera comprensione, l’esperienza soggettiva e l’autoconsapevolezza che caratterizzano il pensiero umano.

Inoltre, la comunità scientifica continua a sviluppare nuovi metodi e criteri per valutare l’intelligenza artificiale, che potrebbero superare le limitazioni del Test di Turing e fornire una visione più approfondita delle capacità delle macchine.

In conclusione, la questione di se il Test di Turing sia stato superato è soggetta a interpretazioni e dibattiti, e non c’è consenso universale su questo punto.