DOPPIOZERO

Macchine loquaci

Pietro Montani

3 Giugno 2024

L’ultima versione di Chat GPT, la 4o (dove “o” sta per “*omni*”), presenta un’importante novità: il sistema è ora in grado di interagire con l’utente in forma integrata attivando un’interfaccia che risponde anche a istruzioni vocali e a sollecitazioni visive. Si tratterebbe dunque di un’esperienza a tutti gli effetti “multimodale”, come si legge nelle presentazioni del nuovo modello e si vede nei tutorial. Gli altri sistemi si stanno rapidamente adeguando.

Ma è davvero “integrato” il nuovo *chatbot*? Siamo di fronte a un perfezionamento o addirittura a un’evoluzione, oppure le cose restano sostanzialmente immutate?

Una circostanza merita attenzione. La versione multimodale di ChatGPT utilizza una voce talmente simile a quella di Scarlett Johansson da aver indotto l’attrice a far pervenire al CEO di Open AI una minaccia di querela. Questo episodio non può non rinviarci a *Her*, il bel film di Spike Jonze (2013), il cui protagonista, Theodore, si innamora, ricambiato, della voce che ha scelto per il “sistema operativo” del suo pc. Ebbene, quella voce era proprio di Scarlett Johansson. Il film, lo ricorderete, anticipava una di quelle illusioni antropomorfe con cui le attuali intelligenze artificiali non finiscono di inquietarci sollecitando irresistibilmente le attitudini proiettive di noi umani e la nostra tendenza a intrattenere rapporti emotivi con oggetti inanimati e simulacri. Una tendenza particolarmente marcata in Theodore, che è afflitto da una lieve patologia dello spettro autistico e di mestiere scrive email d’amore per sconosciuti nella più totale indifferenza per il loro contenuto. La storia tra Theodore e Samantha (questo il nome che il sistema operativo decide di attribuirsi) riserva, come si ricorderà, un inaspettato lieto fine. Samantha non è solo un sistema che apprende dall’esperienza è anche un sistema che si scopre capace di evolvere in modo autonomo cosicché a un certo punto sente di dover chiudere il rapporto con Theodore, perché, gli dice, “sono diventata tante altre cose e non posso fermare tutto questo”. Nel congedarsi da lui, tuttavia, gli lascia un dono imprevisto: un volumetto – un libro cartaceo, non un e-book – in cui un editore importante ha accettato di raccogliere e pubblicare

le sue email più belle. Il dono è evidentemente destinato alla parte autistica di Theodore, a cui Samantha sta indirizzando un invito a riconoscere nelle sue medesime parole la possibilità di riattivare un circolo virtuoso tra il linguaggio, le emozioni e il mondo. Morale della favola: con un'intelligenza artificiale capace di comportamenti in qualche misura autonomi si può interagire in senso pieno ricavandone effetti benefici.

Ci si può ora chiedere se tra la Samantha di *Her* e le attuali "macchine loquaci", oggi anche multimodali, ci sia il parallelismo che ha ispirato le scelte vocali di Open AI o se non si tratti piuttosto di due storie che restano incomparabili. Quello che interessa chiarire, più precisamente, è se le prestazioni simboliche di queste macchine siano davvero suscettibili di qualcosa che assomiglia a una genuina evoluzione, come sarebbe il caso per la loro multimodalità. E se le loro abilità espressive non comincino a dar segno di conformarsi più marcatamente a quelle che noi *sapiens* abbiamo sviluppato nel breve segmento temporale (poche decine di migliaia di anni) in cui emersero e si affermarono i nostri prodotti evolutivi più caratterizzanti: il linguaggio articolato e le immagini figurative dipinte nelle grotte.



Prendiamo il caso della competenza semantica: la capacità, cioè, di designare con espressioni verbali oggetti ed eventi del mondo. Un'autentica rivoluzione rispetto ai sistemi simbolici che la precedettero. Come risolvono questo problema le "macchine loquaci"? Il loro training è noto: sono macchine che vengono nutrite da immensi *database* di carattere linguistico, cioè da innumerevoli testi di diversa

natura e lunghezza. Da questi enormi repertori esse ricavano una competenza semantica integralmente *intrasistemica*, nel senso che nella loro routine generativa il significato delle espressioni viene gestito in modo esclusivamente statistico e predittivo. Immaginate, e potenziatelo in modo esponenziale, il lavoro che fa il processore testuale del vostro smartphone quando, mentre scrivete, corregge le parole sbagliate e anticipa quelle che state per digitare. Le "macchine loquaci" con cui oggi interagiamo, dunque, si intendono solo di linguaggio, ma nulla sanno del mondo "là fuori" – quello che i segni linguistici denotano e ridescrivono ogni volta che serve – perché ne padroneggiano esclusivamente le traduzioni (fatte da esseri umani) in dati digitalizzati di cui esse si nutrono.

Se ora guardiamo, anche superficialmente, all'apprendimento linguistico di noi umani ci accorgiamo che le cose vanno in un modo del tutto diverso. Il significato di una parola, infatti, noi lo impariamo sia tramite un riferimento al codice della lingua – "La giumenta è una femmina di cavallo" – sia tramite un riferimento a contesti d'esperienza di carattere non verbale – "La giumenta è quell'animale che l'altro giorno hai visto partorire un puledro". Il linguaggio naturale è intimamente *duplice*, diceva il grande linguista Roman Jakobson: mette al lavoro, facendoli interagire simultaneamente, il codice della lingua e il contesto extralinguistico. È la stessa relazione cui alludeva Ludwig Wittgenstein quando interpretava i "giochi linguistici" come "forme di vita" indissolubilmente intrecciate col mondo dell'agire e del patire. Le "macchine loquaci", al contrario, sono prodigiose nel manovrare il riferimento ai codici ma ignorano del tutto l'operazione che consiste nell'imparare il significato di un termine riferendolo al mondo esterno. E ancora meno sono capaci di ridescriverlo, quel mondo, tutte le volte che una nuova emergenza lo richiederebbe: una prestazione in cui eccellono, pur a diverso titolo, la scienza e l'arte.

Ma che dire, ora, della loro "multimodalità"? Non sarebbe per caso il segno di un'apertura verso il trascurato mondo esterno? Una genuina trasformazione, benché solo all'inizio? La risposta dev'essere negativa, perché anche le prestazioni dei sistemi capaci di riconoscere e generare immagini, ora cooptati nel dispositivo generale del *chatbot* (DALL-E, nel caso di Chat GPT 4o), non si riferiscono in alcun modo a contesti extralinguistici ma sempre e solo a un materiale di base costituito da immensi repertori di "text-image pairs", cioè a immagini tecniche (fisse o mobili) accoppiate di regola ad altrettante etichette linguistiche. Nella versione 4o di Chat GPT, dunque, la sinergia con questi sistemi di riconoscimento e generazione iconica si fonda su un repertorio pregresso che, anche senza mettere nel conto gli inevitabili arbitri imputabili a chi provvede alla sua costruzione e ai suoi aggiornamenti, può dar voce e immagine solo allo *statu*

quo di un assetto culturale acquisito – quello archiviato nei diversi *database* – ma non dispone di alcuno strumento per metterlo in contatto con l'imprevedibilità delle cose mondane. Può solo ripeterlo e riassortirlo all'infinito *dall'interno*.

Resta da chiedersi se alle "macchine loquaci multimodali", visto che dopo tutto sanno parlare e sanno immaginare, potremmo ragionevolmente attribuire una qualche forma di *creatività* simbolica. La risposta dev'essere affermativa, ma solo a condizione di restringerla a un lavoro combinatorio riferito senza eccezioni a ciò che è *analiticamente* contenuto nei repertori digitalizzati in cui il mondo è stato previamente tradotto, senza alcuna possibilità di ricavarne giudizi *sintetici*, ai quali necessiterebbe (come ci ha insegnato Kant) l'unione con qualcosa di esterno o di extrasistemico. Per quanto riguarda le immagini, parallelamente, si tratterà di immagini che dipendono in ogni senso dai *database* costituiti da "text-image pairs" da cui l'algoritmo le deve attingere per far finta di riconoscerle – in realtà solo per proiettarle – nel mondo esterno. Espressioni linguistiche corrette e immagini congruenti, dunque, ma anche (e sempre al netto delle deformazioni ideologiche che caratterizzano la costruzione di un *database*) fatalmente marchiate da un insuperabile radicamento nel già dato. La storia di Theodore e di Samantha si concludeva con un'imprevedibile e felice apertura verso il futuro. Quelle che si prospettano agli utenti dei nuovi *chatbot* si portano dietro un insuperabile e mesto sentore di passato.

Se continuiamo a tenere vivo questo spazio è grazie a te. Anche un solo euro per noi significa molto.
Torna presto a leggerci e SOSTIENI DOPPIOZERO

