DOPPIOZERO

L’Io e l'IA

Marco Trainito

29 Agosto 2023

Quello che il filosofo della mente (e ingegnere) Riccardo Manzotti e il neuroscienziato Simone Rossi (*IO & IA. Mente, cervello e GPT*, Rubbettino, luglio 2023) hanno confezionato non è solo un libro introduttivo sulla natura di ChatGPT – il modello linguistico a larga scala basato su Transformer introdotto nel 2017 da Ashish Vaswani nell’articolo *Attention is All You Need* – ma una messa a fuoco delle enormi questioni filosofiche che questa intelligenza artificiale ha sollevato. Il suo impatto è descritto facendo appello al linguaggio degli storici: «ciò che uno storico come Alessandro Barbero chiamerebbe una periodizzazione, ovvero un evento che divide la storia in due età completamente diverse».

Perché ChatGPT suscita inquietudini apocalittiche sul destino umano? Il problema, notano a più riprese Manzotti e Rossi, è che ChatGPT attacca la cittadella in cui risiede il cuore stesso dell’umano, cioè il linguaggio, che per qualcuno era addirittura *la casa dell’essere.* Manzotti e Rossi non citano esplicitamente Heidegger, ma certamente hanno in mente un certo umanismo metafisico quando osservano che certe paure sono legate a considerazioni ontologiche sul linguaggio naturale. E qui affiora un pregiudizio antico sul rapporto tra pensiero calcolante da un lato e pensiero filosofante e poetante dall’altro. Non è la prima volta, infatti, che certe innovazioni tecnologiche svolgono compiti *umani* meglio degli stessi esseri umani, e le calcolatrici e i computer sono esempi lampanti. Ma perché questi ultimi non hanno scatenato negli umanisti le stesse ansie da fine dei tempi? La ragione sta nel fatto che tutta una tradizione filosofica ha considerato come inferiore il pensiero che *calcola*, coinvolto con la scienza e la tecnica, rispetto a quello che *specula* e *immagina*, coinvolto con la teoresi e la poesia; per cui, se le macchine occupavano il territorio del primo, non c’era da preoccuparsi, perché la forma più alta del pensiero era comunque salva. Con ChatGPT, invece, gli achei hanno fatto irruzione a Troia, e gli umanisti hanno cominciato a vestire i panni dei profeti di sventura.

Il contributo filosofico più significativo del libro di Manzotti e Rossi sta nel mettere in questione l'idea che ChatGPT sia una minaccia per l'uomo. Il pregiudizio di fondo, risalente a Cartesio, consiste nel porre un nesso strettissimo tra soggetto, pensiero e linguaggio: dove c'è quest'ultimo, c'è dietro un pensiero, e il pensiero è la dotazione interna fondamentale del soggetto (*res cogitans*). Ma la tecnologia che sta dietro l'IA linguistica dimostra in maniera drammatica che la suddetta santissima trinità era solo una costruzione metafisica precaria, che infatti ora giace a terra in frantumi. Oggi è possibile interagire verbalmente con un algoritmo generativo che non ha dietro a sé né un soggetto, nel senso della metafisica tradizionale, né, tanto meno, un pensiero: ChatGPT, infatti, non è *qualcuno*, non *pensa* e non *capisce* alcunché di quello che scrive, perché le stringhe verbali che produce sono il risultato di complessi calcoli probabilistici (basati sulla *probabilità condizionata* del reverendo Bayes) basati su una cascata di contesti. Ecco perché la sua apparizione ci spinge a ripensare il bagaglio di pregiudizi ereditati dalla tradizione filosofica e dal senso comune. Credevamo di sapere, almeno vagamente, cosa fossero l'intelligenza, il pensiero e il linguaggio; credevamo che fossero capacità specificamente umane; credevamo persino di sapere dove fossero grosso modo localizzate. E invece ci sbagliavamo.

Il dialogo a tre messo in scena dal filosofo, dal neuroscienziato e dall'IA diventa l'occasione per ripensare criticamente e senza paure apocalittiche alcuni muri maestri della nostra natura. Non si deve pensare che i ruoli tra le due voci (o tre?) siano scontati. Rossi non è (solo) il portavoce dell'approccio naturalistico oggettivante, così come Manzotti non è il paladino della tradizione umanista che difende la centralità ontologica della *res cogitans* dotata di un teatro cartesiano interno. Le carte sono state sparigliate e il lettore scopre che il neuroscienziato trova difficile uscire dall'idea tradizionale secondo cui dentro il cervello accade *qualcosa* che ha a che fare con il pensiero e la coscienza, cioè con il fondamento stesso del soggetto pensante, mentre il filosofo della mente è il difensore di una teoria così rivoluzionaria e fuori dagli schemi consueti da apparire provocatoria e persino implausibile: non c'è niente di speciale nel cervello e il pensiero e la coscienza sono solo chimere linguistico-concettuali che dovremmo smettere di cercare nella scatola cranica, perché quello che chiamiamo esperienza del mondo è lo stesso mondo di cui facciamo esperienza. L'esperienza cosciente di questa mela, ama dire Manzotti, è la stessa mela fisica che abbiamo davanti.



Il libro inizia dal livellamento verso il basso delle capacità cognitive umane a seguito di un abuso delle comodità offerte dalle tecnologie informatiche. Poi, dopo un dialogo serrato tra il neuroscienziato e ChatGPT su plasticità cerebrale, cambiamento climatico e abbozzo di un romanzo di fantascienza, gli autori forniscono una descrizione dettagliata (nei limiti di un testo divulgativo) del modo di operare di Transformer, cioè della tecnologia che sta alla base della produzione verbale di ChatGPT. A questo punto, neuroscienziato e filosofo interrogano ChatGPT come se fosse a un esame universitario. Il risultato è così sorprendente che gli autori lanciano una proposta importante alla scuola: ChatGPT costringe a ripensare interi settori della pedagogia e della didattica. Passando attraverso altri temi chiave, quali la plasticità del cervello e i modelli predittivi della mente, il libro si considera il parallelo tra GPT e una forma estrema di esternalismo, anche alla luce dei giochi linguistici wittgensteiniani, che tiene conto solo del comportamento linguistico, di Input e di Output (da cui I/O), «ovvero il mondo che sta intorno al nostro sistema». In tal senso, IO e IA sarebbero solo un I/O, senza fantasma nella macchina, senza *homunculus,* senza un'inaccessibile interiorità. L'esperienza umana si differenzierebbe dalle prestazioni delle macchine per il valore e il significato che dà al mondo, anzi, per il valore e il significato del mondo in cui consisterebbe *tout court*, grazie a quell'identificazione radicale di mente e oggetto così cara a Manzotti.

Il libro termina con una riflessione critica su alcune prerogative cognitive che amiamo autoattribuirci, la cui plausibilità empirica è incerta. Prime fra tutte, il *capire* che è stato spesso considerato un processo misterioso che ci distingue da

tutti gli altri animali e l'intenzionalità che è stata vista come un ponte tra il pensiero mediato dal linguaggio e la realtà. Oggi, però, la produzione linguistica artificiale sganciata dalla *comprensione* mette in crisi queste certezze anche se gli esseri umani interagiscono con il mondo attraverso il corpo e con gli altri esseri umani attraverso il linguaggio, mentre i modelli linguistici a larga scala (di cui GPT è l'esempio più famoso) accedono al mondo indirettamente attraverso il linguaggio prodotto dall'uomo. È la copula diretta con il mondo che, almeno per ora, manca. Non sappiamo cosa accadrà quando quest'ultima porta sarà aperta e il contatto sarà possibile (149-150). Quando la vecchia cittadella della conoscenza enciclopedica e del linguaggio naturale sarà ormai perduta, ci rimarrà il compito di costruirne una nuova per mettere in salvo alcuni tratti umani irrinunciabili come il pensiero, la creatività, l'immaginazione, l'ironia, l'espressione non verbale, la prosodia nell'esporre i concetti e la capacità di provare emozioni. L'IO, pertanto, può ancora sfuggire alle rimasticature combinatorie *chiuse* dell'IA linguistica e proiettarsi verso l'esplorazione creativa di dimensioni nuove, *aperte* e ancora sconosciute.

Il libro riserva sorprese filosofiche e il lettore sarà costretto a mettere in discussione i pregiudizi in cui è immerso. Si pensi solo a questo: chi ama ripetere e ripetersi che ChatGPT non è dotata di intelligenza né pensa *veramente* ha forse ragione, ma non ce l'ha per la *ragione* che comunemente immagina lui, soprattutto se pensa di sapere cosa siano l'intelligenza e il pensiero. Abituati a pensare di essere dotati di facoltà interne uniche, misteriose e per certi versi magiche, raramente ci chiediamo se sappiamo davvero cosa siano questi superpoteri. Non a caso le neuroscienze oggi non hanno ancora rintracciato nel cervello alcunché di simile. Insomma, dentro il sistema nervoso non c'è niente di non fisico e *nessuno*, *nemmeno noi*. E dove siamo, allora? La risposta di Manzotti è che siamo lì fuori, nel mondo fisico e che è correlato al nostro corpo, il quale costituisce solo un'occasione (altrettanto fisica) per il suo manifestarsi.

A sostegno della tesi dell'identità tra IO, I/O, IA e mondo viene evocato Luigi Pirandello di *Sei personaggi in cerca d'autore* (1921): «come possiamo intenderci, signore, se nelle parole ch'io dico metto il senso e il valore delle cose come sono dentro di me; mentre chi le ascolta, inevitabilmente le assume col senso e col valore che hanno per sé, del mondo com'egli l'ha dentro? Crediamo d'intenderci; non c'intendiamo mai!». E qui l'angoscia del personaggio pirandelliano rimanda idealmente all'incomunicabilità del significato delineata da Wittgenstein nella prima parte delle Ricerche filosofiche (293) con la celebre metafora dei *coleotteri* – chi può sapere che cosa nascondiamo dentro la scatola dell'interiorità? Wittgenstein ci dice che, se non altro per quanto riguarda le espressioni

psicologiche, un fraintendimento della loro grammatica ci spinge a credere che ciascuno di noi pesca il loro significato in ‘oggetti’ interni e privati (il proprio coleottero nella scatola, che potrebbe mutare o addirittura non esserci affatto), mentre quest’ultimo risiede tutto nell’uso pubblico di tali espressioni, ovvero in giochi linguistici appresi in cui entrano in gioco anche e forse soprattutto il comportamento (le parole di GPT). Ma Pirandello anticipa i tempi in un altro passo, quando Vitangelo Moscarda ormai uscito di senno e relegato in un ospizio dichiara: «Io sono vivo e non concludo. La vita non conclude. E non sa di nomi, la vita. […] Sono quest’albero. Albero, nuvola; domani libro o vento: il libro che leggo, il vento che bevo. Tutto fuori […] muojo ogni attimo, io, e rinasco nuovo e senza ricordi: vivo e intero, non più in me, ma in ogni cosa fuori» (Libro ottavo, IV: “Non conclude”).

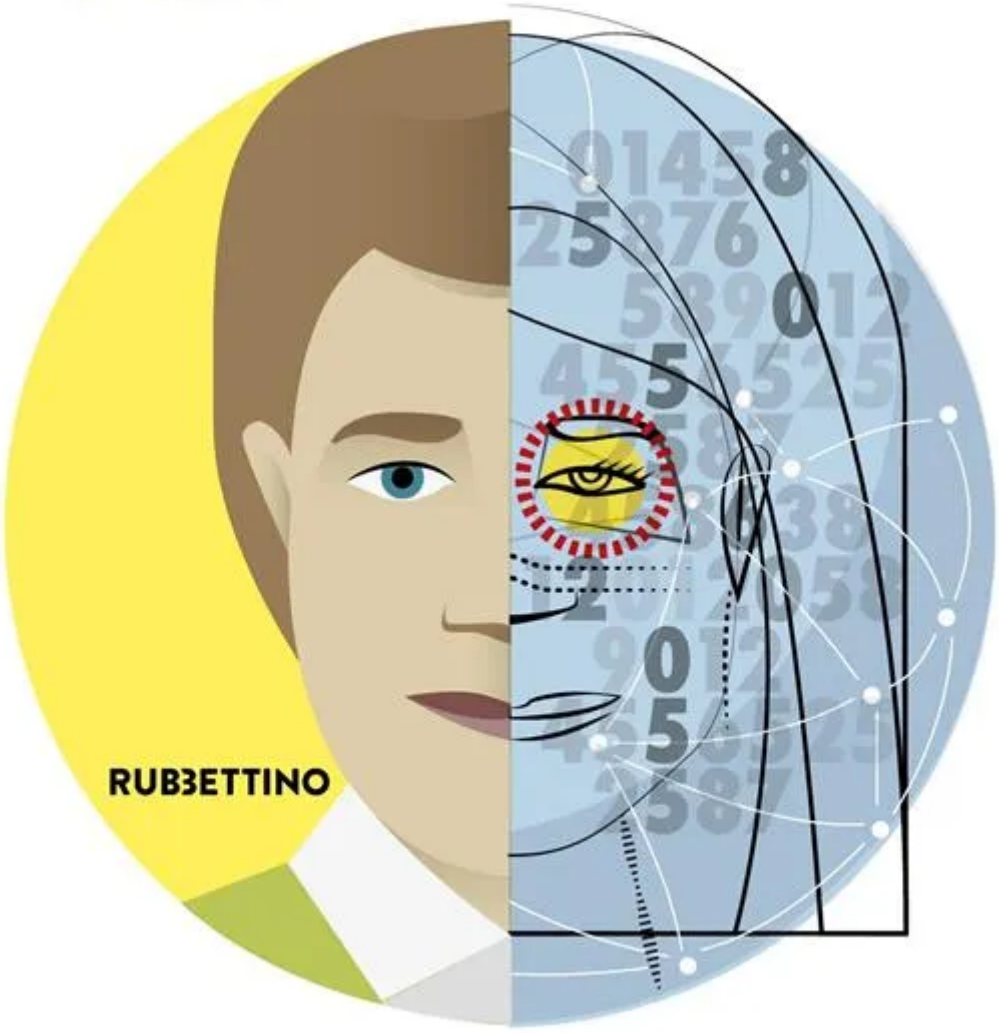
Manzotti e Rossi affrontano così l’*Hard Problem* di David Chalmers, all’incrocio tra letteratura, neuroscienze e filosofia. La domanda sul pensiero di IA diventa la domanda sul nostro pensiero. L’IA ha o avrà un IO? Alla nota osservazione di Colin McGinn secondo la quale le neuroscienze, nonostante le promesse, non sono ancora riuscite a spiegare come si passi dall’acqua dei neuroni al vino della coscienza, il filosofo e il neuroscienzato, arrivando da due discipline diverse, danno una risposta netta: la tramutazione appare misteriosa perché abbiamo cercato nel posto sbagliato (il cervello) un ingrediente miracoloso (il pensiero) che con ogni evidenza non esiste. Il problema dell’IO rivisto da IA – da Vitangelo Moscarda a Wittgenstein – potrebbe essere uno pseudoproblema.

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

RICCARDO MANZOTTI
SIMONE ROSSI

IO & IA

MENTE, CERVELLO & GPT



RUBBETTINO