DOPPIOZERO

Turing alle spalle

Pino Donghi

9 Aprile 2024

Presi come siamo, nel dibattito pubblico, dalle lezioni della storia, infinitamente rivisitata così da provare a capire il presente, capita ci si distragga dalla immediata lettura dell'oggi, e ancor più dalla necessaria prefigurazione del futuro dove vogliamo (o vorremmo) vivere. Lo scriveva già Eco in "A passo di gambero": sguardo in avanti, in realtà retrocedendo.

Sicché ogni tanto suona la sveglia. Come scrive Nello Cristianini a pag. 59 del suo fresco di stampa *Machina sapiens* (il Mulino, 2024): "Il 9 dicembre 2023 il CEO di OpenAI Sam Altman scrisse su Twitter: "*Ci siamo lasciati alle spalle il test di Turing, e nessuno ci ha fatto caso*".

Marginale (ma insomma…), una notazione sintattica: per buona parte del libro, Cristianini declina i verbi al passato remoto, "Sam Altman *scrisse*…", "*lo chiamarono GPT* ", riferendosi ad avvenimenti appena accaduti, di fatto riferibili a un presente appena dilatato, raccontando come "storia" passaggi che al più indietreggiano o seguono Covid-19 di un paio d'anni. Una misura, non banale, della velocità a cui marcia il progresso tecnologico.

Testo necessario, *Machina sapiens*, poiché orientarsi in quello che sta accadendo, alla velocità in cui succede, è imprescindibile, un obbligo sociale, più che una curiosità dell'intelletto, a meno di contentarsi dell'inesauribile nostalgia di tutte le dispute del '900 o – meglio/peggio – della rilettura di un'antichità greco-latina dove tutto sarebbe già stato compreso, detto e risolto: note a margine del pensiero di Platone.

Di tutte le infinite profezie (retorica ma, anche in questo caso, insomma…) di Alan Turing, quella secondo cui "… *alla fine del secolo* [il XX] *l'uso delle parole e l'opinione generale colta saranno cambiati a tal punto che si potrà parlare di macchine pensanti senza aspettarsi di essere contraddetti*", è forse l'ultima e la più definitiva. E certamente ci riguarda, se è vera la frase con cui Cristianini apre il libro: "Non so come funzionino veramente ChatGPT e i suoi molti cugini, non lo

sa ancora nessuno”.

La prima tessera del domino a cadere, secondo la ricostruzione dell’autore, data l’annuale conferenza dell’allora NIPS, ora NeurIPS, *Neural Information Processing* System, fu il Dicembre del 2017 a Long Beach, quando si seppe che AlphaZero, il nuovo algoritmo di DeepMind aveva imparato a giocare a Go interamente da solo e meglio di qualsiasi altro algoritmo (un anno e mezzo prima, nel Marzo 2016 AlphaGo aveva battuto Lee Se-dol, il professionista sudcoreano, campione mondo). In pochi, ricorda Cristianini, fecero caso a una presentazione/articolo di alcuni ricercatori di Google che descriveva un nuovo modo per risolvere i compiti di traduzione. Un metodo che metteva insieme un nuovo algoritmo, *Transformer*, dotato di un meccanismo di “attenzione” che gli permetteva di scoprire quali altre parole devono essere considerate quando se ne traduce una data – ovvero, *le dipendenze* – con un nuovo processore GPU, di quelli utilizzati per le applicazioni grafiche (*Graphic Precessing Unit*): “Sia la descrizione teorica di Transformer sia il codice che lo implementava furono distribuiti in open source da Google, come è oramai d’uso nella ricerca scientifica […] Siccome ciò che faceva era generare testi, il metodo era un esempio di quella che si chiama “*IA generativa*”. Ci siamo.

Per capire il secondo passo, la caduta della seconda tessera del domino, bisogna sapere o ricordare che ogni agente intelligente ha bisogno di un modello del suo ambiente, che sia implicito o esplicito, e così da poter scegliere nel repertorio delle azioni quella più opportuna. Quando si parla di IA, bisogna sempre distinguere l’Agente (*bot* o macchina intelligente, per esempio ChatGPT) dal modello del mondo che lo anima (per esempio, GPT-3) e dall’algoritmo che produce quel modello partendo dall’esperienza (Transformer, o simili). “Gli algoritmi di *machine learning* hanno lo scopo di produrre modelli del mondo a partire dall’osservazione di esso”. Seconda tessera: in un articolo pubblicato nel 2018 da OpenAI (la società di Sam Altman, la mamma di ChatGPT) dal titolo, *Improving Language Understanding by Generative Pre-Training*, i ricercatori di quella società raccontavano come avessero “insegnato una serie di compiti diversi a un agente intelligente senza dover ripartire da zero ogni volta”. Ora, una delle regole che governano il *machine learning* è che più il modello è complesso, più dati sono richiesti per addestrarlo. Usando Transformer, la capacità di generare le parole (come nel test *cloze* usato in psicologia della Gestalt, quando si cancella una parola, “La Terra orbita intorno al …”, o “Il … Mese dell’anno è Marzo”, test che correla bene con quelli a *multiple choise*), diventa generativo: “le conoscenze linguistiche apprese semplicemente addestrandosi su un testo e un compito generici si rivelavano trasferibili ad altri compiti, che normalmente richiedono dati costosi […] Siccome questo *Language Model* era stato creato pre-

addestrando un Transformer in modo generativo, lo chiamarono *Generatively Pretrained Transformer*, o GPT per gli amici". La biografia di cui discutiamo da un anno e poco più, nasce con questo *fiat lux*.

L'eccitazione o l'inquietudine, a ognuno la sua risposta emotiva, nasce dalla consapevolezza che un *agente* nato per predire parole mancanti, per risolvere un test *cloze*, sfrutta (nell'uso del verbo si insinua la tentazione antropomorfica) l'abilità acquisita per imparare compiti diversi. Che è la differenza tra istruire una macchina, ogni volta da capo, per ogni singolo compito, e l'istruire un bambino o un animale che a "primo colpo" o in pochi esempi, è poi capace di trasferire idee da un ambito a un altro. Si chiama *transfer learning* o, come lo spiegava Gregory Bateson, il deutero apprendimento: "imparare qualcosa da pochissimi esempi non è un cambiamento quantitativo, ma un nuovo tipo di comportamento". Come è successo tutto questo? Quello che si sa è che l'abilità in sé non dipende dall'algoritmo, ma dal modo in cui questo interagisce con l'enorme quantità di dati usati per addestrarlo. In altre parole: "non si sa ancora che cosa GPT sappia del mondo". Che ci riporta all'inizio di questa recensione e a quello di *Machina sapiens*, il cui sottotitolo, non a caso, recita, "L'algoritmo che ci ha rubato il segreto della conoscenza".



Quello che abbiamo riassunto finora, lo trovate nel prologo e nella prima parte, dedicata ai protagonisti *sine qua non*, "*Gli scienziati: costruire macchine pensanti* "; seguono gli "utenti", nella seconda parte dal titolo "*Le persone: quando l'uomo incontrò la macchina*", e infine ciò che generalmente inquieta, "*Le macchine: cosa sanno di noi, cosa sappiamo di loro*". Non è solo che vale la pena leggerlo, mi azzarderei a dire che è necessario.

In *Machina sapiens* troviamo la storia di ciò che sta accadendo, più o meno a far data dalle prime intuizioni di Turing, giusto a metà del secolo scorso; c'è l'attualità di tutto ciò che stiamo vivendo: è dell'inizio di Marzo 2024 il varo dell'*AI Act* da parte del Parlamento Europeo, il primo atto politico che assume come inderogabile una responsabilità regolativa del settore; e ci sono le domande per il futuro. Oggi ChatGPT-4 è in grado di scrivere programmi in Python di ottima qualità e, come sottolinea l'autore, "la rapidità dell'evoluzione tecnologica è importante tanto quanto la sua direzione". Ricordando *Erewhon*, il romanzo satirico e distopico scritto da Samuel Butler, pubblicato nel 1872, e che già Turing menzionava in un'intervista alla BBC del 1951, segnatamente nel passo dove gli "antimacchinisti" affermano allarmati: "*… gli esseri coscienti esistono da circa venti milioni di anni: guardate quali passi hanno fatto le macchine negli ultimi mille anni!"*, Cristiani si chiede e ci chiede: "le macchine intelligenti di oggi sono un prodotto degli ultimi cinque anni. Siamo pronti per quello che verrà dopo?".

Il vaso di Pandora, avverte l'autore, è già aperto: è possibile non guardarci dentro? Non è mai stato così, l'abbiamo sempre fatto, piaccia o non piaccia, che sia di Esopo o meno, come si racconta nella fiaba, "è nella nostra natura". E se solo una ventina d'anni fa potevamo pensare di risolvere separatamente due compiti, come modellare il linguaggio e come modellare il mondo, per poi combinare i modelli, le cose invece sono andate diversamente, e oggi "… mi chiedo se la distinzione tra comprendere il mondo e comprendere il linguaggio non sia arbitraria, e se un altro tipo di mente non possa tracciare confini ben diversi". Da vecchio, di anagrafe, appassionato di semiotica, direi che è sicuramente, radicalmente arbitrario, e un altro modo di "essere", non può che consegnarci un altro modo di vedere il mondo. Su questo, la più avvertita ricerca neuroscientifica incontra – dati misurabili alla mano – le riflessioni che furono già di De Saussure, Hjelmslev e Greimas.

Ma c'è di più, forse di meglio. Recensendo su queste pagine il fenomenale *La fonte nascosta* di Mark Solms, chiudevamo così come lui si interrogava, ovvero se sarà possibile, in linea di principio, progettare un sistema capace di autoverifica dotato di una coscienza non umana: in altre parole, se sarà possibile costruire

una coscienza artificiale. E secondo Solms, la risposta più ovvia alla domanda più conseguente, “perché farlo?”, consiste paradossalmente nel comprendere come, fintanto che non progetteremo la coscienza, non saremo mai sicuri di aver risolto il problema della sua origine.

Alla stessa maniera, Alan Turing, rispondendo alla BBC nel 1951, dichiarava: “L’intero processo di pensare è ancora piuttosto misterioso per me, ma credo che i tentativi di costruire una macchina pensante ci aiuteranno molto a scoprire il modo in cui pensiamo”.

Siamo la stessa specie di Pandora e Prometeo, ci ricorda Cristianini, e non saremmo dove siamo arrivati se non avessimo giocato con il fuoco. Citando, come fa l’autore in chiusura, i versi dell’*Ulisse* di Alfred Tennyson:

Anche se molto ci è stato preso, molto ci rimane;
e anche se noi non siamo più quella forza che in giorni antichi
muoveva la terra e il cielo, siamo ciò che siamo;
una tempra di cuori eroici, indeboliti dal tempo e dal fato, ma forti nella volontà di lottare, di cercare, di trovare, e di non cedere mai.

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


NELLO
CRISTIANINI
macrolibrarsi
ALTERNATIVA NATURALE
MACHINA
SAPIENS
L'ALGORITMO CHE
CI HA RUBATO IL SEGRETO
DELLA CONOSCENZA
il Mulino