DOPPIOZERO

Il Nobel alle reti neurali

Pino Donghi

12 Ottobre 2024

"*Penso che queste cose vadano dette, benché io stesso mi stupisca di essere arrivato al punto di dare questi suggerimenti. Fatto questo, non dovremo illuderci: l'esempio dell'energia nucleare e delle armi atomiche è davanti ai nostri occhi. Non sembra che ci siano alternative se non fare il possibile, e il prima possibile, per riconoscere la portata delle implicazioni che avrà, in un futuro molto prossimo, la nostra capacità di creare macchine coscienti*". Così per Mark Solms, docente di Neuropsicologia all'Università di Città del Capo.

Non è dello stesso avviso Anil Seth, che all'università del Sussex (UK) insegna Neuroscienze Cognitive e Computazionali, non tanto per la preoccupazione ma a monte: ché se l'intelligenza, sia pur in forme limitate, è già una proprietà dei sistemi artificiali, non lo è e non potrà esserlo la coscienza, che sarebbe invece caratteristica dei soli sistemi viventi. Nelle sue parole: "*L'intelligenza è qualcosa che i sistemi fanno, la coscienza è un aspetto di ciò che i sistemi sono*".

Dello stesso parere di Solms è invece Stanislas Dehaene, docente di Psicologia Cognitiva Sperimentale al College de France, che l'eventualità di computer che hanno coscienza nel senso proprio della parola, che possano trattenere informazioni e processarle nella stessa maniera in cui lo facciamo noi, non la ritiene affatto impossibile. "*Attenzione! Non è un obiettivo o qualcosa che abbiamo già ottenuto, non penso che attualmente l'Intelligenza Artificiale sia a questo livello, che abbia riprodotto l'architettura della coscienza, ma penso che la coscienza sia proprio nell'architettura cerebrale, e un giorno sapremo simularne il funzionamento nelle macchine*".

Non è difficile immaginare che la comunità dei neuroscienziati, insieme a quella dei fisici (cfr. l'intervista al nostro Giorgio Parisi su *la Repubblica* di Mercoledì 9 Novembre) abbia accolto con soddisfazione la notizia del Nobel per la Fisica 2024 a John J. Hopfield e a Geoffrey Hinton, "per le fondamentali scoperte e invenzioni che rendono possibile il *machine learning* con le reti neurali artificiali". E che i Nobel per la Chimica, il giorno dopo, siano andati a David Baker, Demis Hassabis

e John Jumper (gli ultimi due del team DeepMind, ovvero Google), per aver sfruttato l'intelligenza artificiale così da scoprire proteine utili alla medicina, da una parte dimostra la consapevolezza dell'Accademia svedese delle potenzialità dell'IA, dall'altra può confermare Geoffrey Hinton nella convinzione che se questa sarà molto positiva per la medicina, l'ambiente e i nano materiali… più incerto si delinea un futuro nel quale potremo avere a che fare con entità più intelligenti di noi.

A far notizia sui giornali e nella rete, infatti, è stata anche la circostanza per cui, appena lo scorso anno, il già 76enne Hinton, scienziato inglese di nascita (di quella Wimbledon del tennis) ma che ha insegnato e fatto ricerca negli Stati Uniti e più che altro in Canada, a Toronto, si è dimesso da Google dove era entrato dieci anni prima, nel 2013, in seguito all'acquisizione della DNNresearch che lui stesso aveva fondato. Un abbandono che certamente aveva a che fare anche con l'età, "ho 76 anni e sono vecchio ormai", ma comunque, "… per poter parlare liberamente dei rischi dell'intelligenza artificiale".

Di questi si è parlato e si legge diffusamente in questi giorni "svedesi", ma moltissimo e con rinnovata, grande animazione almeno dalla fine 2022-inizio 2023, con lo sbarco planetario di Chat-*GPT*. A più riprese lo abbiamo fatto anche noi, su queste pagine (si vedano i link segnalati in coda).

L'assegnazione dei recentissimi Nobel, quindi, può esser utile per una breve ricapitolazione. Proviamo a fare un po' di ordine.

Intelligenza-Coscienza. Non sono la stessa cosa, e i pericoli che agitano sono diversi. In che senso?

Se prendiamo in mano il nostro cellulare di recente generazione, è probabile che per aprirsi "ci chieda" (!) di inquadrare il nostro volto per il riconoscimento facciale, in alternativa possiamo appoggiare la nostra impronta digitale o confermare un codice numerico. *Et voilà*, il cellulare si apre al nostro utilizzo. È un comportamento intelligente, quello del cellulare? Sicuramente! È cosciente? Manco per sogno! "L'intelligenza – ricordiamo Anil Seth – è qualcosa che *i sistemi fanno, la coscienza è un aspetto di ciò che i sistemi sono*". Colloquialmente, posso dire che il cellulare "*mi fa entrare*" e, come quando ho scritto in precedenza che è probabile "*ci chieda*", uso due formulazioni antropomorfiche, *come se* la macchina in questione senta qualcosa mentre fa quello che fa. È un *come se* che ci porta fuori strada: il cellulare "fa" una cosa intelligente, ma non dice *a sé stesso*, "siccome la faccia che sto vedendo è esattamente quella che mi aspettavo di vedere, ora gli faccio usare il suo cellulare": non c'è nessun *sé stesso*

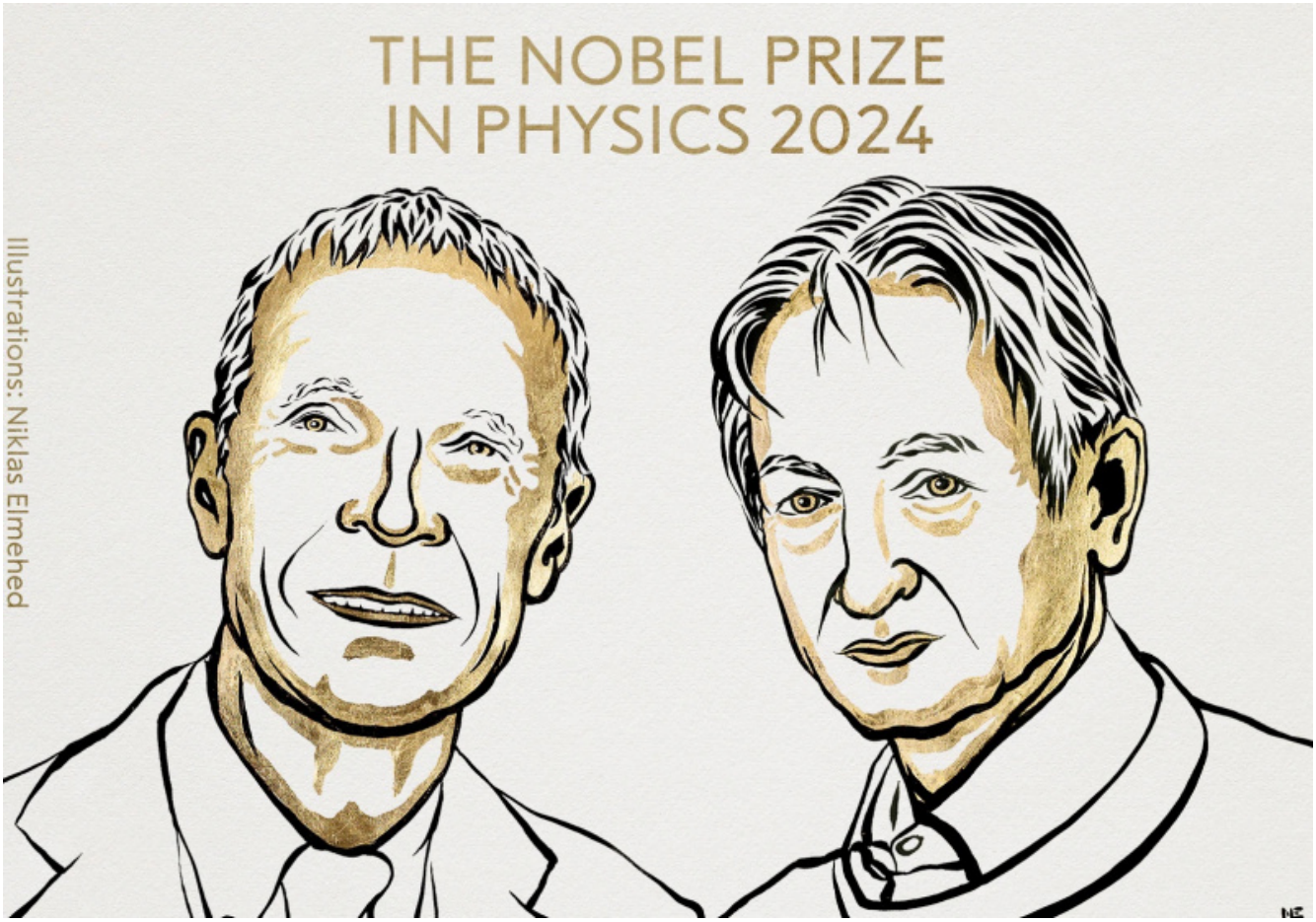
, non si prova nulla a essere un cellulare. La coscienza è quella cosa che si prova "*ad essere me*", ed è il problema difficile di cui ha scritto David Chalmers. Quelli facili, come istruire una macchina a "vedere e riconoscere", dal punto di vista tecnico, ovviamente, sono tutt'altro che facili, pure sono abbordabili, ci si può provare e, come stiamo sperimentando ogni giorno di più, sono risolvibili. La nostra vita è oramai potentemente determinata da tutta una serie di macchine intelligenti che per buona parte ci aiutano in una quantità di circostanze che, presi come siamo dalla quotidianità, nemmeno consideriamo più con attenzione e, aggiungerei, con consapevole riconoscenza.

Generalizzando, mi rendo conto – e accetto la responsabilità di un'affermazione nella quale molti singoli non vorranno riconoscersi – quando reagiamo impauriti, quando ci lasciamo andare a timori che esprimiamo tra amici, o perché ne leggiamo sulla stampa, in un romanzo, partecipando all'incontenibile chiacchiericcio social, quando ci lasciamo andare, inquieti e affascinati, alla visione di una pellicola distopica, è perché immaginiamo "macchine coscienti", non genericamente "più intelligenti di noi": queste ci sono già. Non macchine che "sentono" qualcosa: come ammonisce Dehaene, "*non è un obiettivo o qualcosa che abbiamo già ottenuto".* Anche se è ottenibile, almeno secondo lui, Solms, molti altri, compreso il David Chalmers del "problema difficile", per il quale una possibile sostituzione uno-a-uno, neurone-silicio è possibile, e "… *nell'ipotesi, estremamente probabile, che i cambiamenti nell'esperienza corrispondano ai cambiamenti nell'elaborazione – e visto che non si sono verificati cambiamenti nell'elaborazione durante la transizione da neuronale a silicio – siamo portai a concludere che due sistemi funzionalmente isomorfi di qualunque tipo devono avere lo stesso genere di esperienze".*

Come urlava Frederick–Gene Wilder in *Frankenstein Junior*, "Si – può – fareee!"Sicché, a mio parere, se e quando sarà, i timori o meglio, la riflessione deve essere un'altra. Ci torno. Ma veniamo invece ai rischi dei quali Hinton, e molti altri parlano.

Intelligenza –Intelligenze+intelligenti. Il problema è che non sappiamo come funzionano.

Il salto, che si è determinato anche e specialmente grazie alle sue ricerche e a quelle di John Hopfield, ci ha portato dal progetto annunciato da John McCarthy nel lontanissimo 1956, al Dartmouth Summer Research Project on Artificial Intelligence, e prima ancora dall'articolo di Alan Turing del 1950, *Computing Machinery and Intelligence*, all'intelligenza artificiale generativa di cui abbiamo fatto tutti conoscenza da quando possiamo usare Chat-*GPT* e simili.



“*In questo momento* – dichiarò Hinton al momento di annunciare le sue dimissioni da Google – *quello che stiamo vedendo è che strumenti come GPT-4 surclassano le persone per quanto riguarda la quantità di conoscenze generali, e di molto. In termini di ragionamento, non sono allo stesso livello, ma fanno già dei ragionamenti semplici. E dato il ritmo dei progressi, ci aspettiamo che le cose migliorino abbastanza velocemente. Quindi dobbiamo preoccuparci di questo* […] *Queste cose sono totalmente diverse da noi, alle volte penso che è come se fossero atterrati gli alieni, ma la gente non lo avesse capito, perché parlano un buon inglese*”.

Detto in parole chiarissime, come quelle che usa Nello Cristianini per aprire il suo *Machina Sapiens* (Il Mulino, 2024), “*Non so come funzionino veramente Chat GPT e i suoi molti cugini, non lo sa ancora nessuno […] Di alcune cose possiamo essere certi. Primo: il comportamento di queste nuove macchine intelligenti è diverso da quello della generazione precedente, ovvero è sicuramente cambiato qualcosa. Secondo: questa differenza non è stata pianificata da qualcuno, si è manifestata da sola sorprendendo anche i suoi stessi creatori. In altre parole, è emersa spontaneamente dall’interazione delle sue parti, tra loro e l’ambiente*”.

Non si tratta di macchine coscienti, che esprimono qualche tipo di volontà autonoma, che “sentono qualcosa ad essere loro stesse”. Ma sono macchine che

nell’incontro tra la fenomenale potenza di calcolo di cui sono dotate con l’enormità nemmeno quantificabile di dati con le quali sono alimentate, mettono in evidenza correlazioni che nessun “umano” ha mai trovato prima, scoprono possibili relazioni mai pensate finora, stabiliscono nessi, ponti, comprendono e costruiscono mondi di riferimento. Come avrebbe detto Turing, sono in grado di pensare, anche se a modo loro. E come concludeva lo studio di GPT-4 condotto da Microsoft nel 2003, “raggiungono una forma di intelligenza generale, mostrando effettivamente scintille di intelligenza artificiale generale”.

Quindi dobbiamo preoccuparci – ammonisce il neo premio Nobel Geoff Hinton –, quindi bisogna “… poter parlare liberamente dei rischi dell’intelligenza artificiale”.

Non c’è dubbio. Ma è la stessa qualità di rischi a cui si fa riferimento ogni qualvolta si ricordano i progressi nella fisica nucleare, quelli che hanno portato agli usi civili dell’energia atomica ma anche alle armi che oggi di nuovo, dopo Hiroshima e Nagasaki, non sembra più un tabù pensare di utilizzare sul campo. O all’allarme denunciato il 26 Luglio 1974, dalle colonne di *Science*, sui “rischi biologici potenziali associati all’uso di molecole e microrganismi ricombinati nei laboratori di biologia molecolare”, con una lettera aperta, primo firmatario Paul Berg, ma sottoscritta da David Baltimore, James Watson e Stanley Cohen, e che nel febbraio dell’anno seguente avrebbe portato alla Conferenza di Asilomar. Durante la quale ci si chiedeva se la competizione senza regole tra laboratori e ricercatori potesse abbassare le soglie di attenzione e salvaguardia, rendesse tutti un po’ meno cauti, facesse prevalere la *hubris* di arrivare primi a scapito delle necessarie precauzioni. Che è un po’ quello che è successo, negli ultimi decenni, con lo sviluppo di Internet, dove l’agonismo concorrenziale ci consegna una situazione per cui la parola *web*, di fatto, si confonde con il logo di *Google*. Che è la ragione per la quale a Marzo del 2023, un gruppo di imprenditori (anche Elon Musk, anche quelli di DeepMind) chiese una moratoria di sei mesi, non della ricerca ma del dispiegamento degli strumenti, degli ultimi metodi-prodotti. Posizionamenti di mercato.

Insomma, le preoccupazioni di Hinton e non solo le sue, riguardano, al solito, “chi” userà gli strumenti che lui stesso ha contribuito grandemente a rendere disponibili, e “a che scopo”. Su questo non solo chi può, ma tutti indistintamente, dobbiamo interrogarci e agire politicamente. Con la forse stucchevole precisazione del maiuscolo della P.

E con un’ulteriore precisazione, se è lecito integrare un’affermazione del Professor Parisi. Nell’intervista a Repubblica che abbiamo ricordato si iscrive, a mio parere correttamente, tra coloro che non sono preoccupati, “si tratta di

strumenti importanti… si dovrà normarli”: Politica, appunto. Ma aggiunge: “*I modelli di IA di cui parliamo sono basati sul linguaggio ma guidare una macchina non è questione di linguaggio: devi farti un modello esterno e poi usare i dati acquisiti per prendere decisioni. E modelli basati sul linguaggio non capiscono bene il mondo fisico, perché non ne hanno esperienza. A San Francisco i robot taxi sono stati mandati in tilt da persone che indossavano t-shirt su cui era stampato un divieto di transito: le auto a guida autonoma li hanno scambiati per segnaletica stradale. Qualunque essere umano avrebbe capito la differenza tra una maglietta e un cartello*”. Verissimo. Sicché, a chi scrive suggerisce un altro tipo di preoccupazione, di cui raramente si sente discutere. Lo scenario e il problema potrebbe essere non quello di sviluppare a tal punto le macchine da renderle ancora più intelligenti, cosa che comunque accadrà (“*dato il ritmo dei progressi, ci aspettiamo che le cose migliorino abbastanza velocemente”,* afferma Hinton), ma che si decida di rendere più prevedibile e meno incerto il mondo, i nostri ambienti, così che si adattino meglio allo stadio di sviluppo delle macchine. È già successo. A metà dell’800, almeno nel mondo occidentale, nelle strade cittadine circolava di tutto, carrozze, donne, uomini, bambini e animali, senza alcuna restrizione o regolamento: è stato l’avvento del motore Benz e delle automobili Ford a imporre il codice della strada. Oggi ci appare normalissimo guidare a sinistra della carreggiata (tranne che in Inghilterra e in molti paesi del Commonwealth), rispettare la segnaletica, passare con il Verde e fermarsi allo Stop. Oggi le moderne metropolitane viaggiano senza conduttore (anche a Torino e a Milano, mica solo in Giappone!), perché il grado di incertezza del percorso è ridotto al minimo: nel tunnel non c’è chi possa fare una manovra azzardata, non ci sono scooter o monopattini che ti superano a destra, nessuno va oltre il limite di velocità. Che è un bene, va da sé. Ma la tendenza è insidiosa. L’architettura sociale potrebbe prendere una direzione per cui si affermi, in più di un contesto, l’opportunità e la prevalenza di “ambienti e modelli di mondo” più prevedibili, il meno incerti possibili, così da minimizzare i rischi. In altre parole: sicurezza al posto di liberta. Tendenzialmente… ma c’è da riflettere. Una qualità di riflessione che precede la Politica.

La Mente e le altre Menti. Cosa succede se ne creiamo una in proprio, una in più?

Altro scenario è quello che ci consegna la possibilità di progettare una coscienza artificiale, e che ci riporta al fascino/timore delle “macchine come me”, di quelle che possono “provare qualcosa”, in genere a quelle cattive di *Matrix* o al *Roy Batty* di “ho visto cose che voi umani!”. È una contraddittoria fascinazione che ci accompagna almeno dal *Golem* o dal *Robota* di Karel Capek. Ma ha a che fare con l’“altro”. Non è che al mondo, già oggi, non ci siano altre coscienze: oltre a noi

sapiens, e a far data almeno dalla dichiarazione di Cambridge del 2012, è riconosciuta la presenza di una coscienza in tutti i vertebrati, compresi i pesci, in alcuni molluschi, i calamari, le seppie e specialmente il polpo, in alcuni artropodi. Non sull'intelligenza, ma proprio sulla coscienza del polpo, leggere e godere di tutti i testi di Peter Godfrey-Smith. Semmai dovessimo riuscire a creare una coscienza "altra" dalla nostra, non sarebbe "più", in qualche senso, sarebbe "diversa". E non possiamo sapere come sarebbe così come non possiamo sapere cosa si prova ad essere un pipistrello: Thomas Nagel insegna. Sarebbe "semplicemente" un altro modo di stare al mondo, in relazione creativa con *quello* ("quello", perché non sarebbe "questo", non sarebbe il "nostro"). Come scrive Mark Solms, che è convinto che sarà possibile, sullo sfondo c'è il problema delle altre menti, e per questa banalissima ma ineludibile ragione, non potremo nemmeno essere sicuri di aver generato qualcosa di cosciente, dovremo comunque dubitarne, nel frattempo congetturando a partire dai suoi comportamenti.

La domanda vera è: farlo o non farlo? Perché dovremmo cercare di creare artificialmente la coscienza? Solms risponde , "Semplicemente perché finché non avremo progettato la coscienza, non saremo mai sicuri di aver risolto il problema delle sue origini: se qualcosa può essere fatto, sarà fatto". Cristianini conferma: "Alla fine, credo di sapere che cosa faremo, perché so chi siamo: siamo la stessa specie di Pandora e Prometeo. Dove saremmo se non avessimo giocato con il fuoco?".

Gli scenari sono questi: entità più intelligenti di noi, ambienti più prevedibili e meno incerti così da favorire le caratteristiche delle macchine, un'altra mente da aggiungere alle altre menti. Può far aprire gli occhi alla meraviglia o chiuderli stretti, stretti sotto le palpebre, per la paura.

Che Geoff Hinton ponga domande ineludibili, non c'è dubbio. Che lo faccia dopo aver fatto tutto quello che lo incuriosiva e che ha pensato di poter fare, non può sorprendere più di tanto.

Siamo di quella stirpe lì: Pandora, Prometeo e anche Ulisse.

Leggi anche:
Pino Donghi | Turing alle spalle
Pino Donghi | ChatGPT
Pino Donghi | Viaggio all'origine della coscienza
Riccardo Manzotti | IA, un mondo senza pensiero?
Riccardo Manzotti | Se l'IA sa tutto, perché imparare?
Riccardo Manzotti | Coscienza artificiale: l'ultima frontiera

Riccardo Manzotti | L’IA pensa. E noi?
Riccardo Manzotti | Intelligenza artificiale: proprio come noi?
Pietro Montani | IA e l’immaginazione creativa

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

