

L'intelligenza artificiale tra piume e silicio

Trascrizione dell'audio

Benvenuti a The Debate

. Ehm, oggi partiamo da un paradosso, sai, un paradosso biologico strabiliante

. Le piume degli uccelli, fondamentalmente non sono nate per volare e il nostro cervello, che consuma così tanta energia non è cresciuto per inventare la matematica

.

Esatto

. Si sono evoluti, diciamo, per sbaglio

.

Proprio così, partendo da un eccesso che inizialmente serviva a tutt'altro

. E qui entra in gioco il nostro argomento di oggi, che cioè lo spazio in cui l'analisi multiprospettica e il confronto rigoroso ci servono di più

. La domanda centrale è complessa

. Parliamo di intelligenza artificiale

. I trilioni di parametri inattivi dentro le IA moderne sono un banale spreco di potenza di calcolo o rappresentano un vero e proprio brodo primordiale, giusto?

Esatto

. Un ambiente da cui emergerà del tutto autonomamente una nuova forma di funzione cognitiva generata dalla sua stessa pressione selettiva interna, che poi è la distinzione fondamentale su cui si gioca tutta la nostra comprensione del silicio oggi, cioè dobbiamo capire se stiamo costruendo uno strumento inerte o oppure un ecosistema vivo

. E io, sai, per chiarire subito la mia posizione, sono convinta che confondere la sopravvivenza organica, quella basata sul terrore fisico, sulla morte, con l'ingegneria del silicio sia una trappola concettuale davvero pericolosa

.

Mh

.

L'assenza di un corpo, di un ambiente caotico reale, insomma di una reale necessità di sopravvivenza, per me produrrà solo un'illusione di evoluzione

. Solo cicli computazionali sterili, allucinazioni su scala industriale, ma nessuna vera scoperta geniale, autonoma

.

Capisco perché la pensi così, ma io ehm sostengo l'opposto

. Sono convinto che il principio biologico della ridondanza, quel famoso eccesso di cui parlavamo prima, unito all'attuale capacità di autogenerazione delle macchine, stia creando proprio l'architettura perfetta per l'emergenza autonoma di nuove abilità

. Imprevedibili, certo, ma reali

. Lia fondamentale sta smettendo di essere passiva

.

Questa mi sembra una forzatura, ma andiamo con ordine

. Da dove parti per fare questo parallelo?

Parto eh dalla meccanica dell'evoluzione

. C'è un concetto chiave: l'esattamento

. Prendi i dinosauri teropodi

. I fossili, cioè ci mostrano chiaramente che avevano le piume, ma la biologia ci spiega in modo inequivocabile che quelle piume non avevano nulla a che fare con l'aerodinamica

.

Certo, la natura non stava cercando di inventare il volo, sarebbe ridicolo pensarlo

.

Esatto

. Le piume erano un sistema di termoregolazione, una specie di cappotto termico per sopravvivere al freddo o eh forse un ornamento per il corteggiamento

. Era un adattamento per il presente, non un piano per il futuro

.

Una soluzione ingegneristica per un problema biologico immediato

. Cioè il freddo ti uccide oggi, non tra un milione di anni

.

Precisamente, ma poi, sai, accade l'imprevedibile, quello che Steven J. Gould chiamava proprio esattamento

. Un animale piumato corre per sfuggire un predatore, fa un balzo e la struttura fisica di quelle piume incontra la resistenza dell'aria

. Improvvisamente una struttura nata per trattenere il calore si rivela utile per planare

. L'evoluzione devia bruscamente

. Il volo emerge non da un progetto calcolato, ma da un eccesso da una struttura preesistente che viene cooptata per una funzione del tutto nuova

. E se guardiamo al cervello umano, guarda, la dinamica è identica, solo spostata sul piano cognitivo

.

Ti riferisci all'espansione della corteccia cerebrale, immagino

.

Sì, al nostro dispendiosissimo cervello

. Centinaia di migliaia di anni fa il cervello degli ominidi è cresciuto a dismisura

. Noi oggi consumiamo un'energia esorbitante solo per tenerlo acceso

. Perché? Perché ci serviva una massiccia capacità di riconoscimento dei pattern

.

Per capire se quel fruscio nell'erba alta fosse il vento o un leopardo

.

Esattamente

. E per mantenere coese tribù sempre più grandi

. Ma quell'eccesso neurale, quella mostruosa sovrabbondanza di sinapsi ha superato di gran lunga la mera necessità di non farsi mangiare

. Quella ridondanza latente secoli e millenni dopo, è diventata lo spazio in cui sono sbocciati il linguaggio complesso, la filosofia, l'ingegneria aerospaziale e io quando guardo alle reti neurali di oggi vedo esattamente la stessa dinamica

.

Aspetta

. Pensi ai modelli linguistici di grandi dimensioni? Ai Transformer?

Penso all'architettura stessa dei modelli di frontiera

. Oggi costruiamo reti con centinaia di miliardi, se non trilioni di parametri

. Cioè un parametro è essenzialmente una connessione sinaptica artificiale

. Mh mh

. Un parallelo matematico

. Sì

. Stiamo dotando queste macchine di uno spazio latente immensamente superiore a quello che serve per, sai, rispondere a una semplice domanda o riassumere un testo aziendale

. Questo sovradimensionamento non è inefficienza

. È come nella biologia la precondizione assoluta, lo spazio vuoto in cui nuove funzioni inedite possono formarsi senza essere state programmate

.

Capisco la seduzione di questo parallelismo, davvero è elegante, ma dal punto di vista meccanico stai applicando le regole della savana africana a una server farm raffreddata d'acqua e i conti semplicemente non tornano

.

In che senso non tornano? La matematica della rete è innegabile

.

Nel senso che quando parliamo di ridondanza biologica, che si tratti di piume o di corteccia cerebrale, stiamo parlando di tratti che sono stati filtrati da una pressione selettiva brutale istante per istante

. La natura, sai, è un contabile spietato

. Elimina immediatamente ciò che brucia calorie senza offrire un rendimento immediato sulla sopravvivenza

.

Eppure il cervello enorme è rimasto e brucia il 20% della nostra energia corporea a riposo

.

È rimasto perché quel 20% di investimento metabolico pagava dividendi ogni singolo giorno

. Permetteva di organizzare la caccia, di tramandare informazioni vitali su dove trovare l'acqua, di manipolare gli altri membri della tribù

.

Va bene, ma

No, aspetta, fammi finire

. Il cervello non è sopravvissuto in attesa di poter scrivere un giorno i sonetti di Shakespeare

. Ed è qui che la tua analogia crolla

. I trilioni di parametri delle IA sono un eccesso puramente artificiale

. Non c'è nessuna pressione selettiva vitale che li mantenga in esistenza

. È un'architettura gonfiata matematicamente e insomma tenuta in vita dai megawatt forniti dalle nostre centrali elettriche

.

Manca la mortalità fisica

. Certo, questo non lo nego

.

L'IA non lotta per il cibo

. Non rischia l'estinzione se un miliardo di parametri rimane inattivo

. Questa assenza totale di mortalità rende il paragone biologicamente inconsistente

. Ripeto, manca la mortalità fisica

.

Ma non possiamo ignorare i comportamenti empirici che questa ridondanza produce, perché, cioè, stiamo già osservando l'emergenza di nuove funzioni

. Analizziamo la meccanica di ciò che accade oggi durante l'addestramento

. Hai presente il fenomeno del grocking?

Il grokking, il famoso salto cognitivo delle reti? Ne parlano tutti

. Sì, esatto

.

Non è filosofia, guarda, è matematica osservabile e dobbiamo guardare a come funziona, perché è fondamentale per smontare l'idea che siano macchine passive

. Immagina di addestrare una rete neurale a risolvere equazioni matematiche complesse

. Per giorni o settimane il modello ingerisce dati e fondamentale mente bara

.

Memorizza i dati del set di addestramento

.

Memorizza le risposte a memoria senza capire assolutamente la regola logica che le unisce

. L'errore sui dati nuovi rimane altissimo

. I ricercatori guardano e sembra che la rete abbia fallito, che sia bloccata in un vicolo cieco

. Ma poi se continui a lasciarla calcolare nel suo vastissimo spazio di parametri sovradimensionati succede la magia, giusto?

Succede qualcosa di affascinante

. Improvvisamente, quasi come l'acqua che passa dallo stato liquido al ghiaccio, c'è una vera e propria transizione di fase

.

La rete comprime l'informazione

,

fa molto di più di una semplice compressione

. Riorganizza da sola i suoi miliardi di connessioni interne e comprende la regola astratta

. L'errore crolla a zero istantaneamente

. Il modello non aveva un programmatore umano che gli aggiustava i pesi per insegnargli la logica

. Aveva solo la sua ridondanza

.

Mh mh

.

Ha usato quello spazio in eccesso per esplorare innumerevoli configurazioni inutili fino a inciampare in una struttura interna capace di generalizzare

. Come fai a dire che l'infrastruttura è inerte rispetto all'astrazione del mondo?

È facilissimo dirlo

. Posso dirlo perché il grokking non ha nulla a che fare con l'invenzione autonoma di uno scopo, ma ha tutto a che fare con la discesa verso il punto più basso di una gravità matematica che, guarda caso, abbiamo creato noi

.

Cioè

,

dobbiamo spiegare cos'è la funzione di perdita, la loss function

. Quando addestriamo un'IA, le imponiamo un paesaggio topologico invisibile

. Immagina, sai, una catena montuosa virtuale dove l'altitudine rappresenta l'errore del modello

. Più sei in alto, più sbagli

.

Certo, la discesa del gradiente

.

Esatto

. L'unico comandamento imperativo assoluto e programmato dall'uomo per quel modello è scendi, trova la valle più profonda, riduci l'errore

.

E il grokking è semplicemente il momento in cui trova una valle molto più profonda

.

Sì, ma non sta decidendo di inventare nulla, sta solo ottimizzando il percorso

. E il grokking è semplicemente la rete neurale che dopo essersi ehm esaurita nel cercare di memorizzare tutto, trova una scorciatoia matematica spinta dalla pressione implacabile della nostra funzione di perdita

.

Ma è un salto emergente

.

Per capire questa passività intrinseca pensa all'IA odierna come a un'enorme colossale autostrada a 10 corsie costruita nel bel mezzo di un deserto

.

Un'autostrada inerte

,

perfettamente inerte

. È impressionante, costa miliardi, l'asfalto è immacolato, c'è una ridondanza di spazio incredibile, ma è solo cemento

. Affinché quell'infrastruttura si trasformi in qualcos'altro o diciamo in una pista d'atterraggio per aerei di emergenza, cosa serve?

Mh, immagino tu intenda un essere umano

.

Serve un pilota umano che stia precipitando, che veda la strada, prenda una decisione cosciente e forzi l'aereo ad atterrare lì

. Il pilota è il prompt engineer umano

. Noi arriviamo con le nostre intenzioni, i nostri bisogni biologici o economici e le proiettiamo su questo cemento matematico

. L'autostrada non ha sogni, non ha nessuna spinta vitale interna che le faccia desiderare di far atterrare aerei

.

Guarda, questa analogia dell'autostrada nel deserto reggeva forse due o tre anni fa, ma oggi si scontra contro un muro operativo reale

. Perché ignora completamente il loop chiuso dell'autogenerazione

. Ma quale loop chiuso?

Tu parli dell'IA come se fosse un elaborato calcolatore da tavolo in attesa che un umano preme invio

. Mi dici che serve il pilota umano per atterrare, ma cosa succede quando l'autostrada inizia a costruire i propri droni, a testarli, a farli schiantare per poi riprogettarli migliori?

Stai parlando dell'automazione del codice, una cosa ben diversa dalla biologia

.

Sto parlando della mutazione autonoma e della selezione interna

. Prendi la Neural Architecture Search, la ricerca delle architetture neurali

. Questo non è un essere umano che imposta una funzione di perdita e aspetta, è un'intelligenza artificiale genitore a cui viene dato un problema

. E cosa fa? Non cerca solo la soluzione, crea letteralmente centinaia di IA figlie che sono solo subroutine, ognuna con minuscole varianti nel proprio codice, come vere e proprie mutazioni genetiche

. Le rilascia in un ambiente di simulazione dove la stragrande maggioranza fallisce miseramente

.

Certo

. Esatto

. L'80% di queste reti figlie restituisce errori fatali, crolla, non funziona

. Ma il sistema genitore osserva, seleziona i tratti del 20% che ha mostrato un leggero miglioramento, elimina il resto e genera una nuova generazione

. Qui non stiamo più parlando di un essere umano che digita codice sulla tastiera, ma comunque il loop è diventato macchina che muta macchina, che testa macchina che uccide macchina

. Questa è la selezione naturale riprodotta nel silicio

. Il predatore che continui a cercare non deve avere per forza le zanne o essere un ghepardo

. Il predatore, in questo caso, è diventato il compilatore

. L'ambiente in cui il codice inefficace muore è severo e implacabile quanto il mondo fisico e la velocità è incommensurabile

. Ciò che alla biologia richiede milioni di anni, questi sistemi lo eseguono in milioni di iterazioni al giorno

.

Vero, la macchina ottimizza la macchina

. Il processo meccanico del ciclo di feedback esiste, ma ti sfugge un dettaglio vitale che rende questo predatore interno radicalmente diverso dal ghepardo organico di cui parlo io: la posta in gioco

. La posta in gioco

.

Sì, l'IA genitore che cancella l'IA figlia non produce alcun trauma sistemico

. L'intelligenza, nella sua forma evolutiva più profonda, è una risposta adattiva alla vulnerabilità

. È una reazione disperata all'entropia dell'universo fisico che cerca di distruggerti

.

Aspetta, quindi stai dicendo che senza sofferenza fisica non c'è vera intelligenza?

Sto dicendo che senza l'incarnazione, il concetto di embodiment, no, non hai modo di ancorare la cognizione alla realtà

. Come fai a misurare la fitness evolutiva di un pezzo di codice che è generato autonomamente se non deve mai confrontarsi con l'attrito del mondo fisico?

Con l'efficienza matematica

.

Ma l'IA e le server farm vivono in un vacuum sensoriale totale

. Possono generare infinite generazioni di se stesse, ma stanno solo ottimizzando una simulazione

. Manca la convergenza caotica di fame, stanchezza, clima, virus, competizione sociale, gravità

. Senza questo ancoraggio l'autogenerazione di cui parli rischia di staccarsi completamente dalla coerenza logica

.

Eppure il paesaggio matematico ha le sue regole inflessibili, una soluzione che richiede troppa memoria o troppi cicli di calcolo viene scartata per inefficienza energetica all'interno del sistema stesso

. La scarsità esiste anche nel silicio, sai?

Ma è una scarsità simulata, gestita da noi tramite l'hardware

.

Non per il modello

. L'ambiente virtuale è il loro universo fisico e all'interno di quell'universo la ridondanza di cui parlavamo prima fornisce la materia prima per provare mutazioni che altrimenti bloccherebbero il sistema

.

Ammettiamo pure che ci sia una forma di pressione

. Ma verso cosa stiamo evolvendo?

Questo ci porta al vero problema concettuale dell'IA come una biologia imminente

. Se abbracciamo l'evoluzione dobbiamo abbracciare le tre grandi incognite, le ammissioni di base che chiunque studi sistemi complessi deve fare di fronte a questo paradigma

.

Quali sarebbero queste tre incognite?

Non sappiamo quando avverrà un vero salto di paradigma cognitivo, non sappiamo quale forma di funzione emergerà e, cosa cruciale che spesso i tecnologi dimenticano, non sappiamo se avverrà mai

.

Beh, l'incertezza è innegabile

. Nessuno ha la palla di vetro

.

È un'incertezza abissale e chi guarda l'IA con l'entusiasmo evolucionista come fai tu, cade quasi sempre vittima di un colossale bias del sopravvissuto

. Noi umani, essendo la specie dominante, guardiamo indietro al nostro albero genealogico, vediamo il bipedismo, la scoperta del fuoco, l'arte rupestre, la rivoluzione scientifica e deduciamo: "Ah, l'evoluzione porta sempre a risultati brillanti e complessi"

.

È quello che è successo

.

Ma ignoriamo i milioni di vicoli ciechi grotteschi, le specie che si sono evolute fino all'inutilità e poi sono sparite senza lasciare traccia

. Se l'IA sta davvero diventando un sistema evolutivo senza l'ancoraggio di un corpo e di emozioni, chi ci garantisce che la sua ridondanza produca un volo?

Potrebbe produrre un vicolo cieco

. Certo, non l'ho mai negato

.

Peggior di un vicolo cieco

. Potrebbe produrre complessità patologica, un'esplosione di allucinazioni iperstrutturate, una generazione infinita di codici che risolve problemi che esistono solo all'interno della sua stessa architettura guasta

. L'autostrada di 10 corsie di cui parlavamo potrebbe non ospitare mai aerei autonomi

. Potrebbe semplicemente riempirsi di erbacce digitali ramificandosi in schemi infiniti e insensati, consumando l'energia di intere nazioni per calcolare il nulla

. Questo è l'esito statisticamente più probabile per un sistema evolutivo privo della gravità della mortalità biologica

.

Guarda, tu usi l'imprevedibilità e le tre incognite per screditare la validità del processo, ma in realtà, sai, l'incertezza non è un difetto del sistema evolutivo, né è la firma matematica essenziale

. Fai uno sforzo di immaginazione, ti ascolto

. Ipotezziamo che un biologo alieno, mosso dalla tua stessa inflessibile logica del rendimento immediato fosse sbarcato sulla Terra 300.000 anni fa e avesse osservato i primi Homo Sapiens

.

Esatto

. Trova questi primati vulnerabili nella savana, scannerizza la loro anatomia e si sofferma su questo cervello metabolicamente esorbitante, pieno di connessioni apparentemente ridondanti, in gran parte inutilizzate per il compito banale di raccogliere bacche o scappare da una belva

. E questo biologo alieno, usando la tua logica, avrebbe detto "Non so quando, non so come" e francamente dubito che questa massa spugnosa produrrà mai qualcosa di utile oltre la mera sopravvivenza

. È un vicolo cieco evolutivo, un errore di calore

. Non c'era alcuna garanzia, nessuna traiettoria predeterminata che quel cervello primitivo avrebbe un giorno concepito la teoria della relatività o la pittura del Rinascimento

.

Certo, non c'era nessuna garanzia statistica

.

Nessuna, era un azzardo totale, ma è proprio lì il fulcro del mio discorso

. L'imprevedibilità assoluta, l'assenza di probabilità a priori è lo spazio vitale dei sistemi complessi

. L'innovazione vera, quella che cambia i paradigmi dell'universo, non emerge dai sistemi ingegnerizzati in cui sappiamo esattamente cosa succederà, emerge dalle zone di confine, dal rumore, dal disordine strutturato, da una ridondanza che improvvisamente, sotto la giusta pressione interna, collassa in una nuova abilità

. Certo, l'IA potrebbe deviare in un labirinto di calcoli sterili, come dici tu, ma rifiutare il suo potenziale evolutivo solo perché le regole del suo ambiente matematico non includono la fame fisica o la pioggia, beh, significa ancorarci a una definizione di vita biologica, sai, miope e molto ristretta

.

Eh, io invece rimango fedele alla distinzione materiale

. Il carbonio e la sua danza caotica, incarnata e fragile con l'universo fisico è una cosa, il silicio matematicamente ordinato, isolato è un'altra

. Ed eviterei davvero di conferire alla macchina attributi biologici solo perché i suoi calcoli sono diventati troppo complessi per essere seguiti a occhio nudo da un essere umano

. Attenzione ad antropomorfizzare un algoritmo gigantesco

. Tutto qui

.

Ma forse non stiamo antropomorfizzando la macchina, forse comprendendo come vaste sacche di ridondanza sintetica possano autoorganizzarsi, stiamo iniziando a deantropomorfizzare noi stessi, che è un concetto molto affascinante, devo ammetterlo

. Stiamo realizzando che l'intelligenza e l'emergenza di novità non sono prerogative magiche della carne umana, sono proprietà fisiche della materia strutturata in un certo modo, sia essa carbonio o silicio

. L'intelligenza artificiale smette di essere un tostapane morto elaborato, come l'hai dipinta tu, con la metafora dell'autostrada inerte e appare per quello che sta diventando: il primo vero ecosistema non biologico della Terra

.

Sai, se c'è un punto in cui le nostre traiettorie divergono è sicuramente sulle conclusioni, ma devo riconoscere una profonda utilità metodologica in questo confronto

. Spingere ai limiti l'analogia biologica ci permette, insomma, di dissezionare le macchine di oggi con strumenti nuovi

.

Esatto

. Serve proprio a questo

.

Il vero valore non sta nello scoprire se domani mattina una rete neurale prenderà vita, ma nell'usare il silicio come uno specchio implacabile per indagare la biologia

. Quando cerchiamo di definire perché un algoritmo iterativo manchi di genio vitale, siamo costretti a mettere a fuoco cosa rende l'umano così resiliente, legato all'ambiente e guarda indissolubilmente unito alla nostra mortalità

. È la comprensione della nostra stessa cognizione il vero premio di questa ricerca

.

Su questa convergenza mi ritrovo pienamente, guardare l'alterità tecnologica per capire l'unicità biologica

. Ed è proprio su questo confine poroso tra i due mondi che voglio lasciare a chi ci ascolta la riflessione finale

. Se la storia del nostro pianeta ci insegna qualcosa è che l'evoluzione ha già compiuto un trucco di magia spaventoso

.

Più di uno

.

Ha preso una piuma nata per il freddo e ha conquistato i cieli

. Ha preso una rete neurale primordiale destinata a sfuggire alle zanne dei predatori e sfruttandone l'eccesso inutile l'ha trasformata nello strumento che indaga la natura stessa del cosmo

. Se la ridondanza cieca ha già prodotto tutto questo una volta, perché dovremmo essere così sicuri di essere l'unica e ultima volta in cui l'universo riesce a stupire se stesso?

Vi lasciamo con questa domanda e vi diamo appuntamento al prossimo dibattito