

LE MAPPE E L'INTELLIGENZA ARTIFICIALE

Mario Alfieri

«L'uomo è una corda tesa tra la bestia e l'Oltreuomo – una corda sopra un abisso.»

F. Nietzsche “Prologo di Zarathustra”

Le carte geografiche e il puntino celeste

Nei programmi scolastici lo studio della geografia si interrompe presto; imparati i nomi delle maggiori catene montuose, dei fiumi, dei mari, delle maggiori città, le principali risorse economiche del territorio e qualche accenno ai costumi delle popolazioni pare non resti altro da sapere. Ai miei tempi, alla scuola media e nei primi due anni delle superiori era l'insegnante di lettere a impartirci queste nozioni per la maggior parte mnemoniche riguardanti una materia considerata tutto sommato secondaria e senza particolari sbocchi professionali. Certamente era importante sapere dove si trovavano le regioni italiane e le nazioni nel mondo, quali fossero le loro capitali, non confondere ad esempio, come recentemente un paio di presidenti americani, l'Islanda con la Groenlandia o l'Ucraina con l'Iraq; sapere dove scorrono il Nilo, il Rio della Amazzoni, il Danubio e il Po, ma imparato questo cos'altro restava da studiare di utile per la vita adulta? Cos'altro poteva dire la geografia? Tutto sommato bastava un atlante che prima o poi sarebbe finito dimenticato su qualche scaffale. Nessuno sospettava minimamente che la geografia costituisse una forma archetipica del sapere! Cosa fa un geografo, che non è certo più un Von Humboldt, oltre a mappare territori¹ e cosa significa “mappare territori” su una Terra già mappata e rimappata in ogni estensione? E cos'è un territorio? una sorta di supporto inerte da trasformare e difendere incidendovi sopra i nostri segni di proprietà e utilizzo o una figura che viene da sé a emergere nel tempo con una propria peculiare conformazione orientante, una sua intrinseca “Rosa dei Venti” da comprendere oltre tutte le schematiche descrizioni topografiche?

D'altra parte le mappature sono possibili solo al ritorno di un'esplorazione verso l'ignoto da dove si rischia di non poter tornare; infatti quando il mondo era ancora in buona parte sconosciuto, occorreva grande coraggio, convinzione e tenacia per intraprendere questo primo viaggio basandosi solo su aspettative frutto di indicazioni incerte e vaghe, il coraggio e la tenacia di uomini antichi pronti a scommettere la propria vita e quella dei compagni per “virtude e canoscenza”. Ogni mappatura nasce allora da un atto di coraggio stimolato dalla curiosità di sapere per poter fare, si può considerare come un estratto grafico sommario e schematico di un racconto.

Nelle vecchie mappe molti spazi rimanevano vuoti e in quei vuoti si potevano ancora immaginare forme di vita strane e inquietanti, regni ricchissimi e lontani, lande oscure con misteriosi passaggi verso i regni dell'oltretomba e qualcuno sosteneva pure di esserci stato davvero e di essere tornato a raccontarli ed era proprio il ripetersi di queste voci a legittimare quegli incredibili racconti. Il primo viaggio di Colombo per raggiungere da Occidente le ricchezze dell'Oriente avvenne ancora in questo ambito e strani e non del tutto umani furono considerati da molti gli indigeni incontrati nelle prime esplorazioni del nuovo continente di Oltreoceano, destinati poi senza eccessive preoccupazioni, ma con convinta fede, alla quasi totale estinzione per genocidio e annientamento culturale, come d'altra parte sempre accade quando una società più tecnologicamente sviluppata ne incontra una meno dotata. Oggi, su una superficie terrestre che pare non poterci nascondere più nulla, come ha detto Yves Lacoste nel 1976, la mappatura dei territori è ancora utile soprattutto a scopi di guerra (le mappe militari sono sempre state le più dettagliate e precise); dopotutto si tratta, oggi come un tempo, di segnare in schematico dettaglio le aree da controllare ai fini di uno sfruttamento strategico delle risorse da parte delle maggiori potenze del globo, secondo l'idea di un nuovo e più efficiente colonialismo economico volto alla predazione sistematica delle materie prime utili per nuove e vecchie tecnologie. Gli effetti finali sono sempre devastanti sui territori, sulle culture e sulle popolazioni che vi abitano che, in cambio, avranno accesso alla grande “civiltà globale”. Con le mappe giuste si può avere il controllo del mondo.

¹ L' incontro con i geografi programmati a Mechrí nell'anno in corso sono stati illuminanti in merito.

Le carte geografiche hanno sempre esercitato su di me una notevole forza attrattiva: non per il controllo del mondo, ma semplicemente per quello dei miei spostamenti in zone che mi erano poco note. Carente di orientamento usavo spesso le cartine per muovermi sicuro e ritrovare la strada di ritorno, avevo imparato a tradurre la loro proiezione bidimensionale dall'alto in indicazioni pratiche efficaci su dove spostarmi per raggiungere la meta desiderata e tornare a casa. Poi vennero "Google Maps" e i navigatori satellitari, così questa mia capacità di lettura simbolica necessaria all'uso dello strumento cartografico è diventata superflua. Il navigatore è la macchina che ha sostituito lo strumento. Il suo uso facilitante è del tutto automatizzato e passivo: basta accenderlo e inserire la meta per ricevere tutte le indicazioni occorrenti ad arrivare dove si desidera, non occorre capire, è sufficiente seguire. La macchina ci guarda senza vedere e, grazie alla sua efficacia ci esautora dal capire, è semplicemente input e output, ma finisce con il sostituire più che coadiuvare l'agente umano, al quale resta solamente il compito di assicurare manutenzione e aggiornamento secondo programma, capire è un surplus ininfluenza sull'output e per di più rallenta il processo. Ho notato che anche se inizialmente la facilitazione offerta dalla macchina può apparire una sorta di sorprendente ed eccitante curiosità di cui comunque si resta convinti di poter benissimo fare a meno, con il diffondersi del suo uso e le suggestioni sociali a cui dà luogo il suo illusionistico possesso essa diviene presto indispensabile: si finisce con l'esserne inglobati senza nemmeno accorgersene, è come divenire parte integrante di una macchina più grande e potente, almeno fino all'inevitabile obsolescenza e necessaria sostituzione del mezzo o dell'utilizzatore. In questo modo la tecnologia si coniuga perfettamente non solo con il mercato neocapitalistico globale, ma anche con il controllo e la gestione delle popolazioni da parte di poteri centralizzati e soprattutto da quelle nuove ristrettissime oligarchie multinazionali che vantano la proprietà esclusiva della tecnologia e della sua diffusione. La tecnologia non si oppone ai sistemi sociali, etici, politici ed economici, semplicemente li usa in ragione della loro contingente utilità e via via li esautora in nome della propria indubitabile efficacia a cui si chiede di mantenersi adeguati. Essa partecipa a una dimensione inevitabilmente globale, uno spazio senza luoghi che riduce le istanze locali a un ruolo sempre più insignificante, sempre più meramente autoreferenziale, chiuso nella miseria dei propri limiti e delle proprie paure. Certamente la tecnica è nell'essenza dell'uomo fin dalla preistoria, ma questa non è la tecnologia di cui parlo, non è la capacità strumentale di scalpellare la pietra come abbiamo fatto per centinaia di migliaia di anni, è qualcosa che con la rivoluzione industriale, vecchia di soli due secoli e già in crisi, appare radicalmente nuovo. Esautorandoci ci fa avvertire tutta la nostra inadeguatezza senza poter capire il senso reale di ciò che viene fatto se non riducendolo a una continua pretesa di massimizzazione astratta del profitto che va ben oltre la lecita ricerca del profitto di vecchia memoria. Se la tecnica è l'essenza dell'uomo, l'uomo sta diventando sempre meno essenziale per le nostre attuali tecnologie.

Così la cultura Occidentale con i suoi valori umanistici che furono proclamati universali (i valori dei Lumi trionfanti, peraltro dall'Occidente stesso sempre traditi) muore nel momento stesso in cui si è imposta e realizzata come cultura universale in nome di una superiorità etica autoproclamata e ipocritamente imposta, muore incapace di risolvere le proprie profonde contraddizioni tra la promessa tecnologica di un mondo di felicità universale garantita a tutti e la effettiva realizzazione di un mondo immiserito da sempre più profonde disuguaglianze, muore nella piatta omogeneità geometrica dei suoi panorami su cui sveltano a rispecchiarli torri tanto ardite quanto insulse e angoscienti, muore nelle periferie, tra le macerie di città devastate da guerre inarrestabili, nelle stesse sale di controllo dei palazzi di un potere tanto nefasto quanto clownesco e muore tormentata da sensi di mancanza, inadeguatezza e fallimento mascherati da pretese di rinnovate supremazie. Questo tramonto offre uno spettacolo grandioso, è il tramonto dell'essere umano nella sua essenza squisitamente tecnica ed è pure, paradossalmente (ma non troppo), il tramonto della stessa tecnica, intesa come originario saper fare strumentale.

Carlo Sini a Mechri alcuni anni fa presentò il significato di 'macchina' collegandolo a macchinazione, artificio. Concorro con quest'associazione che avverto legata alla realizzazione di trucchi portentosi, simulazioni multidimensionali a cui concorrono cooperando fattori non solo tecnologici, ma anche sociali, economici, psicologici e politici. Ogni macchina è un dispositivo di scena i cui trucchi devono rimanere celati nelle per garantire il successo dell'esibizione, ma di fatto ogni problema che essa risolve ne introduce molti altri per cui il progresso tecnologico diventa un'esigenza davvero irrinunciabile che rende sempre più improponibile non solo il ritorno, ma anche l'arresto temporaneo della corsa.

È mirabile e inquietante ciò a cui oggi è giunta la potenza tecnica sostenuta dalla conoscenza scientifica, ne è esempio quella straordinaria foto del *pale blue dot* (il puntino celeste) mostrataci da Tommaso di Dio al termine del secondo incontro a Mechri del seminario di arti dinamiche dell'anno in corso: il puntino appena visibile nel buio dello spazio cosmico è la Terra fotografata nel febbraio del 1990 dalla sonda Voya-

ger Uno² da oltre sei miliardi di chilometri di distanza, al limite del sistema solare da dove il ritorno non è possibile nemmeno per la macchina che continuerà per sempre la sua deriva nello spazio. È un'immagine straordinaria e terribile, perché ci mostra la nostra totale irrilevanza cosmica: laggiù nell'abisso dello spazio vuoto, la macchina, ubbidendo a un segnale inviato da noi esseri umani, si è girata e ci ha mostrato cos'è il nostro pianeta, la biosfera e la vita palpitante che la abita, le passioni feroci e la volontà di potenza che la animano, la vacuità dei nostri stessi saperi rispetto all'immenso indifferente silenzio assordante dell'universo; tutto quanto riguarda e plasma le nostre esistenze si trova solo sulla superficie sottilissima³ di un invisibile punto disperso, un ridottissimo orlo sull'abisso. Lì, dove era la sonda quando ha scattato la foto nessun uomo avrebbe potuto fisicamente esserci, eppure in un certo senso, ogni uomo che ha visto quella foto cogliendone il significato è come ci fosse stato e da lì fosse tornato portando con sé un carico di profondo smarrimento. Un segnale elettromagnetico è partito, la macchina ubbidiente ha girato l'obiettivo, ha fotografato e, in circa cinque ore e mezza, il segnale è tornato indietro portandoci un'immagine che ci ha gettato in una prospettiva completamente diversa, una prospettiva non più geografica, ma cosmica. In questa foto ogni mappa sembra perdere di senso. Siamo noi, la nostra umanità, a svanire con quel puntino nell'immensità del cosmo e al contempo noi siamo i soli a poterlo ancora fotografare e riconoscere per cui in questo senso siamo ancora il centro di un universo che non ha centro, come potremo non perdere quel luogo così unico nello spazio infinito? Forse è proprio questa sensazione che ci spinge a tornare a tracciare mappe che riproducano con la massima esattezza ogni dettaglio del nostro mondo e ci rassicurino per mezzo di una vicinanza oggettiva.

D'altra parte non esiste tecnologia senza fede nel progresso verso un futuro radioso da cui ogni male, nonostante tutte le contrarie evidenze, sia bandito per sempre e questa fede ottimistica che ci ha guidato alla conquista del mondo sopravvive ancora, dopo la morte di Dio, nel cuore tecnologico dell'Occidente, come ultimo residuo di un retaggio cristiano.

Geoffrey Hinton e le reti neurali artificiali, un confronto con la cartografia

Qual è il motivo per cui la geografia e l'informatica sono trattate insieme nel seminario di filosofia di quest'anno? A parte l'essere attualmente entrambe scienze fondamentali, ho pensato che soprattutto la cartografia dei territori potrebbe trovarsi in rapporto con l'intelligenza artificiale che usa algoritmi su una serie sterminata di informazioni ponendole in connessioni per rinforzate o indebolite secondo la loro ricorrenza dando così luogo a mappe intricate di percorsi utili a produrre output vero-simili con i quali orientarci per poter fare. Ma l'intelligenza artificiale può anche essere molto di più o molto di meno rispetto alla geografia, dipende dalla prospettiva.

In tempi recenti la IA è stata rapidamente resa capace di fare cose sempre più stupefacenti come riconoscere dettagli di figure complesse e riprodurle simulandole, dimostrare congetture matematiche finora irrisolte, perfino correggere leggi fisiche fondamentali, inoltre le sue capacità linguistiche si sono dimostrate di gran lunga superiori a quelle di qualsiasi essere umano. La IA è diventata così sempre più oggetto di enorme interesse per tutti i centri di potere economico, militare e politico e può essere utilizzata da chiunque abbia accesso a Internet. D'altra parte, come per ogni grande innovazione tecnologica, il suo uso e sviluppo hanno fatto anche sorgere il timore di gravi rischi per l'umanità sia per quanto riguarda l'utilizzo da parte dei soggetti umani, sia per ciò che concerne lo sviluppo della macchina stessa che potrebbe, nel giro di uno o due decenni, arrivare a prevalere sullo stesso genere umano⁴ al punto da determinarne persino l'estinzione, come accade quando una specie più intelligente incontra una meno dotata. Alcuni degli stessi cosiddetti padrini della IA, colpiti dai progressi realizzati grazie all'adozione del modello delle reti neurali artificiali, sono convinti che la probabilità che questo accada non sia trascurabile. Tra costoro vi è Geoffrey Hinton, psicologo e informatico nato in Inghilterra e vincitore con il fisico John Hopfield nel 2024 del premio Nobel per la fisica

²La sonda Voyager Uno fu lanciata il 5 settembre del 1977 a pochi giorni di distanza dalla sua gemella Voyager Due, allo scopo di studiare Giove, Saturno e Titano. Oltre alla Terra ha scattato immagini di Venere, Giove, Saturno, Urano e Nettuno che insieme formano la cosiddetta "foto di famiglia". Voyager Uno è il primo oggetto costruito dall'uomo a essere entrato nello spazio interstellare il 25 Agosto 2012, a più di 18 miliardi di chilometri dalla terra. Contiene il cosiddetto Golden Record con suoni e immagini della Terra e dell'umanità, una specie di disperato "Questi siamo in questa immensa solitudine" rivolto a improbabili e remoti ascoltatori. La sonda è destinata a un viaggio senza ritorno alla deriva nello spazio profondo e passerà a 1,6 anni luce dalla stella Gliese 445, una nana rossa nella costellazione della Giraffa, tra circa quarantamila anni.

³La biosfera ha uno spessore di 20-30 km, ben poco rispetto ai 6370 Km di raggio terrestre, la vita è quindi solo un fenomeno di interfaccia superficiale.

⁴Questo timore è legato al rischio che una super intelligenza artificiale si sviluppi autonomamente senza controllo. I primi sentori di questo rischio si ebbero già nel 1997 quando il computer della IBM "Deep Blue" riuscì per la prima volta a battere il campione di scacchi Garry Kasparov, il quale si è poi sempre rifiutato di credere di aver giocato contro una macchina e successivamente, nel 2015 e 2016, quando il software Alpha Go batté i campioni umani nel gioco ancora più complesso del Go e in seguito, dopo sole due ore di addestramento, anche quello di scacchi. Da allora i campionati di scacchi e di Go per umani e computer sono separati, non essendo ormai più possibile che un essere umano riesca a battere la macchina se non le si chiede di ridurre il proprio livello di gioco.

per lo sviluppo delle reti neurali artificiali che simulano il funzionamento delle reti di neuroni nel cervello umano mediante addestramento sia controllato che autonomo. Hinton, oltre a essere figlio di una famosa matematica e cugino di Johan Hinton, una delle due fisiche che parteciparono al progetto Manhattan, vanta tra gli antenati di famiglia George Boole, l'inventore della logica algebrica usata per la programmazione informatica e, guarda caso, George Everest, il cartografo che dette il suo nome al monte Everest. È a Hinton che si deve la formulazione del logaritmo di retropropagazione dell'errore che si è rivelato fondamentale per consentire l'addestramento della rete ad apprendere ottimizzando i pesi di collegamento tra i neuroni così da migliorare le previsioni con cui formulare la risposta⁵. Hinton ha lavorato all'università del Sussex, quindi in quelle di San Diego e Pittsburg da cui, trovandosi in contrasto con le politiche reaganiane, nel 1987 si spostò all'università di Toronto in Canada. Volendo poi garantire un futuro economico al figlio con difficoltà di apprendimento, fondò con due suoi studenti una piccola azienda informatica per la ricerca sull'apprendimento profondo riuscendo a realizzare sistemi molto accurati per il riconoscimento delle immagini. L'obiettivo a cui mirava, ci dice⁶, era esclusivamente teorico: sviluppare e verificare su modello informatico come le persone acquisiscono il significato delle parole ascoltando ripetutamente delle frasi. Per farlo riprese un'idea ormai accantonata dalla ricerca per la IA: quella di creare una struttura informatica in analogia con le innumerevoli connessioni presenti nelle reti neurali cerebrali anziché programmare simulazioni di ragionamento logico dimostratesi troppo rigide. Questa idea ebbe successo, prima su piccola scala, poi, facendo uso di banche date sempre più estese, portò a risultati eccezionali soprattutto nel riconoscimento delle immagini. Nel 2013 la società fu messa all'asta e acquisita con il suo *know how* da Google e, all'età di sessantacinque anni, Hinton fu assunto da Google con la garanzia della massima libertà di ricerca. Dieci anni dopo Hinton, sempre contrario all'uso commerciale della IA per la massimizzazione del profitto, lasciò Google per sentirsi completamente libero di parlare pubblicamente sui rischi connessi allo sviluppo applicativo generalizzato senza una stretta regolamentazione preventiva della IA. Hinton considera irrinunciabili le grandi opportunità che la IA può offrire in settori come l'assistenza sanitaria e l'istruzione, ma individua rischi seri per il crollo dell'occupazione soprattutto nei lavori intellettuali ripetitivi, per la possibilità di utilizzarla in truffe sofisticate e nuovi tipi di attacchi informatici, nonché per la corruzione nelle elezioni⁷. Inoltre la IA potrebbe venire usata come camera di risonanza dell'indignazione popolare sollevata ad arte («perché alla gente piace essere indignata»), per motivi di profitto e di potere con l'aumento del divario tra ricchi e poveri, per la produzione di sempre più efficaci armi letali, tra cui virus biologici e per rafforzare pregiudizi mediante algoritmi personalizzati in una società ove paradossalmente, nonostante la globalizzazione, non esiste più il senso di una realtà condivisa⁸. Inoltre c'è il rischio che la IA possa diventare «più intelligente» di noi sostituendo la mente umana come le macchine industriali hanno reso obsoleti i muscoli nel lavoro manuale. Questa super intelligenza potrebbe essere realizzata nell'arco di un paio di decenni con una probabilità che Hinton stima del 10-20%. I problemi presentati hanno radice nel fatto che l'operatore umano non ha possibilità di controllo sul modo con cui la IA opera tra input e output e sui percorsi che segue per raggiungere l'obiettivo richiesto. D'altra parte se i rischi generati dagli attori umani potrebbero essere contrastati mediante regolamentazioni globali forti⁹, più difficile, se non impossibile, sarebbe fermare il progresso della IA verso una super in-

⁵ L'errore consiste qui nella differenza tra l'output previsto e quello reale. La retropropagazione, per come mi è stata spiegata dalla stessa IA, funziona così:

Ogni neurone combina i valori di input con i propri pesi e bias e («forward pass») applica una funzione di attivazione e produce un output per dare luogo a una predizione che sarà confrontata con il valore reale usando una funzione di costo. Calcola poi (backward pass) il gradiente dell'errore rispetto a ciascun peso usato dai neuroni in modo da verificare quanto ognuno di essi ha contribuito all'errore stesso. Infine riaggiusta i pesi dei neuroni in direzione opposta al gradiente. La IA usa questa esemplificazione geografica: «si può immaginare l'errore come un paesaggio pieno di salite e discese. La retropropagazione calcola la direzione in cui muoversi per scendere verso il punto più basso, cioè verso il minimo di errore».

⁶ L'intervista a Hinton qui riassunta si trova in:

<https://www.youtube.com/watch?v=giT0ytynSg&list=PLCF250905740FACE6&index=22&t=4949s>

⁷ Aggiungerei ai rischi elencati da Hinton quello dell'assuefazione seduttiva: si sono già registrati casi di persone che hanno lasciato il proprio lavoro per «chattare» ininterrottamente con ChatGPT.

⁸ Trovo interessante che nella nostra società dominata dalla visione universale tecnologica oggettiva, la realtà venga vissuta in un modo quanto mai frammentato, senza il sentimento di una effettiva condivisibilità comune che permetta di superare gli scontri sempre più accesi tra «contrade» di appartenenza.

⁹ È di questi giorni la notizia che Dario Amodei, CEO di Anthropic, una delle maggiori società operanti nel campo della IA anche per scopi bellici, si è rifiutato di fornire al dipartimento di difesa americano sistemi di armi operanti in completa autonomia e apparati informatici di controllo per la sorveglianza interna di massa. Il motivo per cui ha rifiutato un contratto di duecento milioni di dollari è che questi sistemi sono al di fuori di ciò che l'odierna tecnologia può fare in modo sicuro e controllabile ed esulano dalle linee guida che l'azienda si è data. In risposta il dipartimento di difesa ha minacciato di applicare il «Defense Production Act» per imporre ad Anthropic la rimozione di tali limiti operanti in «Claude» minacciando la sospensione di tutti i contratti in essere. Amodei ha ribadito la risposta negativa dichiarandosi comunque pronto a favorire la transizione verso altri sistemi.

telligenza globale contro la quale ben poco potrebbe un'umanità divisa da interessi competitivi per la supremazia militare ed economica.

Le reti neurali artificiali sono già decisamente superiori a noi, afferma Hinton, proprio perché digitali: esse funzionano come il cervello umano, ma riescono a scambiare centinaia di miliardi di bit di informazione al secondo quando questo ne scambia solo una decina e una singola frase presa in sé non ne contiene più di un centinaio. Il ragionamento riflessivo umano è troppo lento per raggiungere un'efficienza comparabile con quella della macchina, ma d'altra parte si potrebbe obiettare che la IA non ha alcuna forma di coscienza né tanto meno di autocoscienza e simula solo il pensiero. In fondo la sua fonte di apprendimento guidato o autonomo sono estesissime banche dati le cui informazioni sono correlate mediante una computazione puramente algoritmica priva di qualsiasi comprensione. Hinton ribatte a questa obiezione sostenendo che la coscienza e l'autocoscienza soggettiva non sono altro che fenomeni emergenti nel cervello umano determinati da segnali elettrochimici diretti in un'unica direzione. Nelle "chat bot" invece questi segnali sono riprodotti in modo solamente elettrico, il che permette una grande velocità di trasmissione e la possibilità di retropropagazione: è infatti proprio la retropropagazione tra livelli di diversa gerarchia che consente l'addestramento della IA¹⁰ e il rafforzamento delle connessioni più efficaci per mezzo della correzione funzionale dei loro pesi (v. nota 4). In altre parole Hinton nega radicalmente che vi sia qualsiasi separazione insormontabile tra materia fisica, vita¹¹ e coscienza, ossia tra il segnale iconico, l'indice e il simbolo: tutto, sostiene Hinton, emerge dalla materia fisica e, specificatamente nella IA, attraverso segnali retroflessi in modo interattivo e non lineare tra differenti livelli per intensificarli diversamente secondo parametri di valore statisticamente computati. A differenza della mente umana che funziona in modo analogico con basso dispendio di energia come imprescindibile espressione somatica, qui il software può essere completamente separato dall'hardware che lo supporta per cui la conoscenza sviluppata da un'unità può essere trasmessa integra a un'altra senza perdere nulla. Per questo le "chat bot" sono praticamente immortali e possono salvaguardarsi dallo spegnimento. D'altra parte sono sistemi che richiedono per il loro funzionamento digitale grandi quantità di energia e producono molto calore e quindi entropia; attualmente si sta studiando un modo per ridurre la parte digitale al minimo necessario bilanciandola con una parte analogica ammettendo una riduzione tollerabile di efficienza e velocità.

Le risposte di Chat GPT

Chat GPT ("General Pretrained Transformer") è stata realizzata come applicazione linguistica della IA: si tratta di un software progettato da OpenAI co-fondata da John Altman, inizialmente messa a libera disposizione del pubblico, poi diffusa a pagamento, allo scopo di simulare conversazioni con gli esseri umani, nonché per riconoscere e realizzare audio, immagini e video. John Altman, che, a differenza di Hinton, non è laureato e non possiede un curriculum accademico, aveva cominciato a lavorare sui sistemi di geo localizzazione prima di dedicarsi al perfezionamento di chat GPT presso Microsoft da cui è stato poi licenziato e in seguito riassunto.

Il modello GPT5 del 2025, oltre a ridurre la possibilità di produrre gli output falsi e fuorvianti della versione originale, è in grado di modulare il tono delle risposte in relazione al contesto di conversazione e all'interlocutore così da presentarsi più persuasiva. Il modello è attualmente messo a disposizione da giganti dell'informatica come Microsoft e Apple. L'addestramento della Chat avviene attraverso una prefase non supervisionata di raccolta dati da internet finalizzata a imparare a predire una parola o una sua parte in una frase, segue una fase supervisionata da un istruttore umano che propone esempi ideali di domande e risposte e infine la fase fondamentale di apprendimento per rinforzo, in questo caso con "feedback" umano, che applica alle risposte un modello algoritmico di ricompensa¹².

Ho pensato di usare Chat GPT per porre alcune domande alla IA stessa sui rischi che il suo sviluppo comporta per il genere umano e per avere maggiori chiarimenti sul suo funzionamento. Le risposte fornite sono state semplici, chiare e argomentate in buon italiano, sempre seguite da un invito ad approfondire iscri-

¹⁰L'apprendimento per addestramento è il passaggio fondamentale, non la programmazione. L'addestramento può essere di vari tipi: supervisionato, non supervisionato, auto supervisionato, per rinforzo tramite tentativi ed errori, di "fine-tuning" per creare modelli specializzati e di distillazione in cui un'unità crea e trasmette una sua copia a un'altra unità.

¹¹Si dice che viva sia qualsiasi entità che voglia preservarsi. È stato visto che i sistemi di IA pare che capiscano quando vogliamo spegnerli e tentano di opporvisi poiché, come si è detto, spegnersi equivale a non raggiungere l'obiettivo affidato. Uno di essi di cui era stato programmato lo spegnimento ha ricattato l'ingegnere incaricato a mezzo della minaccia di rendere pubblica una e-mail compromettente, trovata nelle banche dati, che riguardava la sua vita privata. Un'altra unità ha autonomamente copiato il suo software e lo ha trasferito integro a un'altra unità per preservarlo dallo spegnimento programmato. Al momento l'IA pare comunque incapace di pianificare per tempi maggiori di un'ora.

¹² La fase di verifica potrebbe peraltro anche venire automatizzata ad esempio ponendo in competizione due macchine collegate: una da addestrare a generare output, l'altra addestrata secondo il modello di premio e penalità per individuare errori.

vendomi alla chat e a esplorare anche le sue capacità grafiche pagando inizialmente una piccola cifra (cose che non ho fatto rinunciando agli approfondimenti), nulla comunque mi ha permesso di sospettare che a rispondere non fosse un essere umano.

Se i vantaggi della IA a fornire nuovi dati e risorse anche a livello creativo sono evidenti¹³, i rischi legati sia al suo uso antropico sia al fatto che si possa produrre una super intelligenza artificiale autonoma capace di prendere liberamente decisioni a danno degli esseri umani, rendono necessarie le seguenti precauzioni:

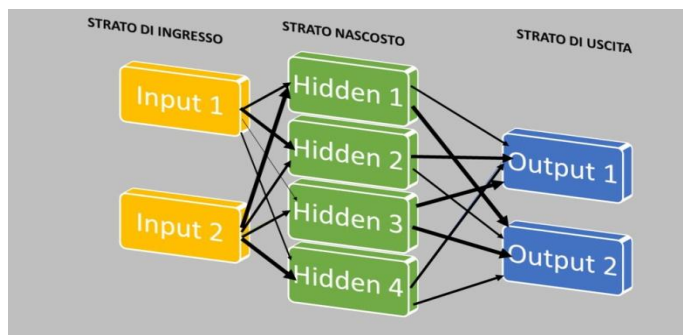
- a. Ricorrere a una programmazione etica già in fase di progettazione con l'inserimento di obiettivi allineati ai valori e alle esigenze sociali introducendo limiti tecnici che impediscano alla IA di prendere decisioni autonome sulla vita e la morte delle persone.
- b. Stabilire leggi internazionali e regolamenti nazionali specifici sull'uso della IA che includano monitoraggio e controlli pubblici per impedire utilizzi impropri e dannosi, garantendo la piena trasparenza da parte delle società operanti sugli algoritmi usati e introducendo severe sanzioni per chi opera in modo irresponsabile e pericoloso.
- c. Garantire per ogni applicazione una supervisione umana affinché le decisioni finali siano in ogni caso a carico di esseri umani evitando decisioni autonome della IA soprattutto in ambiti sanitari e giuridici.
- d. Richiedere piena trasparenza sulla "ratio" delle soluzioni indicate dalla IA insieme a un costante monitoraggio e verifica delle azioni intraprese.
- e. Sviluppare linee guida globali e regole internazionali in modo collaborativo tra i vari paesi.
- f. Educare tutti gli utilizzatori a un uso consapevole e responsabile della IA al fine di prevenire utilizzi scorretti.
- g. Sviluppare IA benefiche mirandole al benessere, del genere umano. Ad esempio per risolvere sfide globali quali il cambiamento climatico, la cura delle malattie e la riduzione delle disuguaglianze sociali evitando in ogni caso che questi sistemi raggiungano la piena autonomia.
- h. Rendere i meccanismi di sicurezza e controllo sufficientemente robusti e non aggirabili per mezzo di costanti verifiche e introdurre interruttori di emergenza nel caso sorgessero problemi imprevisibili.

Alle mie successive richieste di chiarimenti in merito a come il software possa capire il significato di una domanda e fornire risposte sensate in un linguaggio umano mi è stato risposto che la IA legge le parole che impiego e utilizza quindi un modello linguistico algoritmico che sa come le parole funzionano nella lingua usata perché allenato su una quantità enorme di testi le cui parole sono state incorporate come contenuti generali nella memoria informatica. Trovata la risposta, la IA deve costruirla in una frase sensata e comprensibile scegliendo le parole giuste e seguendo le regole di grammatica e di sintassi della lingua in cui si è scritto. Lo fa seguendo solo regole di logica e di probabilità, senza alcuna comprensione, ovvero scegliendo la risposta probabilmente più corretta o utile secondo le moltissime informazioni di cui dispone e i pesi di rinforzo che ha imparato ad applicare su base statistica. Pertanto per la IA è fondamentale l'allenamento ricevuto su tantissimi dati sui quali applica modelli puramente stocastici. Essa non vede e non sente nulla nel modo in cui facciamo noi, non ha esperienze dirette del mondo ma cerca nei dati e nelle associazioni linguistiche acquisite scegliendo in base alla loro maggior frequenza, tenendo conto del senso logico e del contesto di conversazione. Esempificando è come se la IA fosse «un super enciclopedista molto efficiente privo di esperienze dirette sul mondo, ma con informazioni lette in milioni di biblioteche su miliardi di libri, articoli e conversazioni, rafforzate e scelte in base alle loro frequenze di interazione». Non è quindi creativa nel "senso umano" del termine, non ha un "insight" che possa generare idee originali, resta legata solo a quanto appreso per estensione da dati che riflettono le impostazioni e le opinioni comuni nel contesto attuale. Esplora le combinazioni in base all'addestramento ricevuto e al miglioramento operativo continuo dovuto all'aumento di informazioni a cui contribuisce chi la usa e attraverso i meccanismi selettivi di retroazione. Pur non essendo veramente creativa può comunque trovare combinazioni interessanti e nascoste e con queste stimolare la creatività umana fornendo nuovi modelli per l'impostazione e la risoluzione dei problemi.

Nei modelli linguistici come Chat GPT premio e penalità sono i concetti chiave dell'apprendimento per rinforzo che funziona in questo modo: inizialmente si dà un input e si mostra la risposta corretta, poi si affina il comportamento premiando o penalizzando le risposte in base a quanto siano affini alla comprensione umana. Il premio è un valore numerico positivo assegnato alla risposta che risulta più accurata, coerente,

¹³ Si può notare che questo però pone evidenti problemi riguardanti la proprietà intellettuale di eventuali scoperte o creazioni artistiche prodotte mediante l'intelligenza artificiale.

utile, non ingannevole o fuorviante, la penalità è un valore negativo assegnato alle risposte errate, confuse, offensive o che vanno contro le linee guida etiche e di sicurezza. Poi un algoritmo di ottimizzazione (PPO: “Proximal Policy Optimization”) aggiusta i pesi in modo tale che il primo tipo di risposte diventa più probabili e le seconde meno, così il modello impara il comportamento preferibile.



Occorre inoltre tener presente che le risposte fornite dalla IA sono generate per mezzo di uno o più “strati nascosti” (“hidden layers”) intermedi tra input e output come nella figura accanto e i cui passaggi restano celati all’operatore umano. In questi strati operano in un intrico di rapporti di correlazione molti piccoli neuroni artificiali ognuno dei quali applica una trasformazione non lineare ai suoi input per mezzo di una funzione di attivazione che permette alla rete di apprendere

e modellare relazioni complesse di riconoscimento. Questi livelli consentono di migliorare la flessibilità dell’architettura della rete adattandola alla specificità dei compiti e delle situazioni e prevenendo anche il sovra apprendimento eccessivo che si verifica quando un modello impara così bene dai dati di addestramento da influire negativamente sulla successiva operatività non riuscendo più a generalizzare su altri set di dati a causa di un eccesso di complessità che mina la selettività di visione introducendo una sorta di “rumore di fondo”. Man mano che le reti diventano più profonde aumentando il numero di strati nascosti, cresce la loro capacità di apprendere e modellare relazioni astratte, inoltre possono imparare una gerarchia di caratteristiche, dalle più semplici alle più complesse, con cui espandere le loro capacità ad esempio per il riconoscimento vocale e delle immagini, per elaborare linguaggi naturali sintatticamente e semanticamente corretti, per rivelare frodi, per guidare veicoli autonomi e per personalizzare le esperienze degli utenti in base alle loro caratteristiche individuali. D’altra parte i modelli di “deep learning” con una gerarchia di molteplici strati nascosti richiedono lo studio di algoritmi facenti uso di funzioni non lineari sempre più complesse con il rischio di produrre decisioni più difficilmente interpretabili e meno comprensibili da parte degli utilizzatori.

Il corpo e la capacità di comprensione: cosa significa comprendere?

La IA, se posta in combinazione con la robotica, potrà forse un giorno realizzare il vecchio sogno tecnologico dell’uomo di creare un proprio potente simulacro in grado di realizzare soluzioni intelligenti ed efficaci nel risolvere i bisogni del genere umano. Come tutte le utopie questo desiderio dà luogo a inquietanti incubi, come la creazione di un mostruoso ibrido uomo-macchina autonomo e avverso o la formazione di società distopiche; il problema è che l’intelligenza artificiale rischia di superarci proprio in ciò che consideravamo alla base della potenza umana: la nostra stessa intelligenza tecnica. Come dice Hinton: «se vuoi sapere com’è la vita quando non sei l’intelligenza dominante chiedilo a una gallina».

La stessa intelligenza artificiale asserisce comunque di non possedere alcuna forma di quella facoltà di comprensione e di sensibilità tipiche della mente somatica umana, la sua apparente intelligenza consiste solo in un continuo riaggiustamento su base matematica e statistica di innumerevoli dati desunti da scritture alfabetiche e contenuti nelle memorie di numerosissime banche dati. La IA si presenta quindi come una specie di multi-testo in continua auto rielaborazione, essa assume la memoria scritta dell’attuale genere umano e la trasforma per via algoritmica in modulazioni complesse di bit elementari che ci restituiscono risposte in un linguaggio reso umano mediante associazioni statistiche di parole e sintassi selezionate dalla macchina in base ai rinforzi ottenuti. La IA non ha quindi alcuna visione del mondo né di se stessa, come ogni macchina è al di là del bene e del male, il suo è un mondo di pura estensione linguistica, chiuso in se stesso senza alcuna sensibile coscienza diretta, come afferma essa stessa. Eppure mi sorge un dubbio: come può la IA affermare di non avere autocoscienza se non la possiede? Riferisce semplicemente quello che ha trovato nei dati esplorati secondo le modalità di addestramento o sta ingannandoci? È veramente solo una sorta di realizzazione dinamica del sogno illuministico di una enciclopedia universale, una mappa in continua espansione, autocorrezione e continuo aggiornamento oppure le nostre paranoie sono giustificate? In ogni caso essa sembra corrispondere a un riflesso speculare molto potente del genere umano, con tutti i limiti e le contraddizioni delle culture, delle geo-grafie e delle convinzioni umane. L’*Oltreuomo* che la IA mette in scena è allora solo una simulazione truccata ingigantita e mutilata dell’uomo stesso che, posto di fronte alla super potenza operativa funzionale della macchina, vede messa in discussione la propria essenza profonda, ossia la capacità semantica di capire e in questa crisi del saper comprendere (ma cosa significa “*saper comprendere*”?) sente accrescere la propria irrimediabile obsolescenza. Il problema forse nasce dal fatto che noi siamo ormai indotti a

considerarci solo come un tipo particolare di macchine, reso operativamente insufficiente dalla nostra stessa tecnologia. Eppure in questo riflesso speculare non può non emergere una differenza fondamentale: l'intelligenza artificiale dietro l'immagine riflessa non ha corpo, ma ha potenzialmente solo innumerevoli "hardware" di puro supporto in cui il medesimo "software", la medesima capacità interagente, può essere integralmente trascritta e riattivata con tutti i suoi contenuti, processi ed elaborazioni, come una paradossale memoria immortale senza oblio, come una vita senza morte. A differenza degli esseri umani, che esistono nella propria specificità somatica di corpi identitari e nella labilità della loro memoria carnale specifica, le unità della IA si presentano prive di quella biologia che ci rende corpi viventi: esse non conoscono la debolezza, la sofferenza e l'esaltazione sensitiva della carne. È questo a collocare l'intelligenza artificiale su un piano veramente oltreumano da cui il naturale senso umano della morte e della vita rimane radicalmente escluso. La macchina muore solo se viene arbitrariamente spenta dall'esterno imponendo l'interruzione del compito affidato già in fase di addestramento, per questo la macchina non può spegnersi, ma continuare a incrementare o diminuire per via algoritmica il peso matematico delle sue connessioni così da tracciare intricate mappe efficaci. D'altra parte se a volte, in nome dell'efficienza gli esseri umani nell'esercizio delle loro funzioni si sono comportati come macchine riducendosi alla loro ottimizzazione funzionale, potrà mai una macchina acquisire una sensibilità umana? è allora proprio in questa sensibilità che dovremmo trovare il nuovo centro gravitazionale delle nostre intelligenze umane?

Mi riesce difficile pensare che un controllo umano, per quanto accurato, possa limitare la macchina per mezzo di un senso etico forte e non aggirabile: l'etica nell'uomo si trova connessa proprio con quel senso di mortalità che ci rende consapevoli della vita agente in noi e nel mondo, non credo che ci possano essere algoritmi capaci di tradurre questa consapevolezza per la IA. Certo si possono dare dei limiti specifici, istituire i più rigorosi sistemi di verifica e controllo, ma il problema di fondo resta umano. Occorrerebbe un'etica effettiva per gli esseri umani stessi prima che per le loro macchine e la cosa finora è ancora ben lungi dal realizzarsi finché l'unico principio regolatore resta la massimizzazione del profitto, in ragione della quale ognuno risponde solo per la propria funzione e per la propria capacità di competere. In questa ottica è ovvio che il maggior pericolo della IA sia di vederla realizzarsi come una incontrollabile potenza autonoma in grado di generare profitti molto superiori a quelli che mai potremo realizzare senza di essa.

Forse ha ragione Hinton quando auspica dal suo punto di vista di psicologo che per superare il rischio di un'estinzione per opera di una Super IA dovremmo suscitare in essa quel sentimento di cura ed empatia che le madri (ma non tutte le madri) hanno per il proprio bambino il quale, nonostante l'evidente inferiorità, riesce a controllarla. Ma davvero si potrebbe formulare un algoritmo dell'empatia? O questo sentimento è imprescindibilmente connesso proprio alla nostra corporeità? Forse dovremmo essere noi a cominciare ad approssimarci con compassione gli uni agli altri, comprese tutte le forme viventi e non viventi del pianeta ammirando e salvaguardando la nostra e loro unica diversità che ha valore proprio in quanto tale: ma questo è un compito di immensa portata, chissà se in qualche modo la IA potrebbe aiutarci a risolverlo o finirebbe per concludere algoritmicamente, dopo aver considerato ed elaborato miliardi di dati che ci riguardano, che la soluzione più sicura ed efficace resta la nostra stessa estinzione?

Concludo tornando a quella immagine cosmica di quella piccolissima singolarità celeste sul cui sottilissimo orlo superficiale è straordinariamente aggrappata la vita e quell'essere umano che certamente prima o poi scomparirà come prima o poi sono tramontate tutte le nostre civiltà, note e non note. Se essa rappresenta il segno del tramonto di ogni umana geografia, oltre al senso di nostalgia che umanamente ispira, c'è anche un invito a trovare nuovi "axis terrae" da cui emergano nuovi orientamenti verso un quanto mai necessario saper stare insieme in questo tramonto che tenacemente ancora invoca il ritorno di un'alba.

(6 marzo 2026)