

Lettere aperte a proposito dei rischi impliciti nella Chat GPT e nell'Intelligenza Artificiale in genere: un altro vano giro di valzer. Il Garante italiano risponde inutilmente. Si discetta sui pappagalli. Ma i cittadini restano ancora una volta esclusi dalla discussione

di Francesco Varanini

[Articolo pubblicato sul blog Dieci Chili di Perle, 4 maggio 2023](#)

Fine marzo 2023. La lettera aperta del Future of Life Institute [Pause Giant AI Experiments](#), riportata dal *Financial Times*, è pedestremente ripresa dalla stampa italiana. Elon Musk e altri 1.000 leader della Silicon Valley chiedono di "sospendere lo sviluppo dell'intelligenza artificiale". I firmatari spiegano come "negli ultimi mesi c'è stata una corsa fuori controllo dei laboratori per l'intelligenza artificiale a sviluppare potenti menti digitali che nessuno, neanche i creatori, possono capire, prevedere e controllare".

Lettere aperte: Future of Life Institute e Association for the Advancement of Artificial Intelligence

"Chiediamo a tutti i laboratori di intelligenza artificiale di fermarsi immediatamente per almeno sei mesi nell'addestrare i sistemi di IA più potenti come GPT-4. La pausa dovrebbe essere pubblica, verificabile e includere tutti. Se non sarà attuata rapidamente, i governi dovrebbero intervenire e istituire una moratoria", si legge nella lettera del Future of Life Institute.

Un monito da prendere sul serio? Una presa di posizione veramente nuova? Non sembra.

Come racconto nel mio libro [Le Cinque Leggi Bronzee dell'Era Digitale. E perché conviene trasgredirle](#) (Guerini e Associati, 2020, pp. 219 e segg.), lo scrivere lettere aperte a proposito dei rischi impliciti nell'Intelligenza Artificiale è uno sport molto diffuso tra tecnici e guru dell'Era Digitale. Così ci si lava la coscienza, poi si può tornare sollevati a chiudersi in laboratorio, senza pensare troppo alle conseguenze delle proprie azioni, senza assumersi reali responsabilità. Si sbandierano assieme il vessillo della preoccupazione per il futuro dell'essere umano e lo stendardo della potente forza implicita nell'Intelligenza Artificiale. Si ripete fino alla noia che l'Intelligenza Artificiale è gravida di buone promesse ma potrebbe anche comportare l'estinzione della specie umana.

Campione della gara delle lettere aperte è proprio il [Future of Life Institute](#). La recente lettera aperta ripete stancamente quella del gennaio 2015, documento di esordio dell'Institute. Mark Tegmark, fisico ansiosamente alla ricerca del grande pubblico, cerca spazio sul ghiotto terreno dell'Intelligenza Artificiale. Eccolo fondare il Future of Life Institute, di cui è tutt'ora presidente. Si susseguono da allora campagne che sostengono l'ovvio. Il trucco consiste nel non prendere partito: vogliamo una AI al contempo *robust and beneficial*.

Questo significa forse mettere un freno alla corsa verso l'Intelligenza Artificiale? Non sia mai! Al contrario, "this does not mean a pause on AI development". Sarebbe sbagliatissimo, si legge nella lettera del 2015, rinunciare alla ricerca, perché "i benefici [benefits] sono enormi".

"A fronte del grande potenziale dell'AI è importante ricercare il modo di raccogliere i suoi benefici evitando potenziali insidie [pitfalls]". Ciò che già si leggeva nella lettera del 2015, lo si ritrova pari pari nella lettera del marzo 2023: "AI research and development should be refocused on making today's powerful, state-of-the-art systems more accurate, safe, interpretable, transparent, robust, aligned, trustworthy, and loyal".

A cosa dovrebbe servire la pausa ora proposta? "I laboratori di intelligenza artificiale e gli esperti indipendenti dovrebbero sfruttare questa pausa per sviluppare e implementare congiuntamente una serie di protocolli di sicurezza condivisi per la progettazione e lo sviluppo avanzati di intelligenza artificiale, rigorosamente verificati e supervisionati da esperti esterni indipendenti". Notare bene: si parla di protocolli stabiliti dagli stessi laboratori impegnati nella ricerca, accompagnati da esperti esterni indipendenti. Tra le righe è facile leggere la risposta ad una domanda: chi sono gli esperti indipendenti? Innanzitutto i ricercatori del Future Of Life Institute! Quale lo scopo dei protocolli? Rendere tutto più *robusto*. "Robust AI governance systems". "Robust auditing and certification

ecosystem". E anche, guarda caso, "robust public funding for technical AI safety research". Cioè: l'industria privata fa cose pericolose, quindi, argomentano Tegmark e amici, merita risorse pubbliche per mitigare i rischi impliciti nei progetti.

Gli interessi economico-finanziari e l'interesse dei tecnici ad avere mano libera vanno a braccetto. Così, anche, si mantiene vivo, ma sotto controllo, lo stesso business della critica ai rischi dell'Intelligenza Artificiale. Chi è il vice-presidente e tesoriere del Future of Life Institute? Meia Chita-Tegmark, moglie di Mark Tegmark. Ora si considera notevole il fatto che tra i primi firmatari della lettera si trovi Elon Musk. Si evita di ricordare che già all'inizio del 2015 Tegmark, celebrava la "la donazione di 10 milioni di dollari di Elon Musk al Future of Life Institute, che ha contribuito a mettere a disposizione 37 sovvenzioni per gestire un programma di ricerca globale volto a mantenere l'IA benefica per l'umanità".

Dalla discussione, insomma, è bandita l'ipotesi che l'Intelligenza Artificiale possa veramente essere dannosa. Basta del resto leggere ciò che scrive Tegmark per osservare come considera obiettivo auspicabile per gli stessi esseri umani l'affermarsi di una Superintelligenza non umana. Racconto nelle [Cinque Leggi](#) di come Tegmark parli del futuro collocandosi al di fuori della comunità umana, dei cittadini del mondo: scienziato che dall'alto e dal di fuori disegna scenari - per la precisione dodici. Per non comprometersi, non prende esplicitamente partito per nessuno di essi. Ma comunque annuncia qual è il luogo dove il destino degli esseri umani sarà spiegato al volgo: il Future of Life Institute.

Musk, da parte sua, proprio alla fine del 2015 fonda Open AI, società che ha per scopo sviluppare quell'Intelligenza Artificiale Generale, AGI, che Tegmark chiama Superintelligenza. Proprio Open AI sembra ora, con la Chat GPT, essersi avvicinata ben più di ogni altra impresa all'Intelligenza Artificiale Generale. Come mostro nelle [Cinque Leggi](#), Elon Musk presto cambiò idea, e passò a considerare l'Intelligenza Artificiale Generale "un rischio fondamentale per la civiltà umana". Non va dimenticato che Musk indirizzò allora i suoi investimenti verso un progetto diverso, ma non meno inquietante: connettere tramite sottilissimi filamenti neurali il cervello umano ad un computer.

Ecco comunque ora tutti, impauriti o sorpresi, a firmare la lettera del Future of Life Institute. Ma anche, allo stesso tempo, ansiosamente dediti a provare l'ultima versione della Chat GPT. I tecnici firmatari della lettera invocano "un passo indietro rispetto alla pericolosa corsa verso modelli black-box sempre più grandi e imprevedibili con capacità emergenti". L'invocazione è certo frutto di preoccupazione, ma anche, è lecito supporre, espressione di invidia nei confronti dei colleghi di Open AI. Quasi a dire: la black-box dovrebbe essere aperta a tutti i tecnici, cioè anche a noi. La sorpresa di qualcuno di fronte a ciò che la Chat GPT riesce a fare non sembra invece molto motivata. I risultati erano prevedibili. Non c'è niente di nuovo da un punto di vista tecnico, c'è solo il fatto che usando ingenti investimenti e enorme potenza di calcolo si è andati avanti lungo una strada già nota.

Se questo è vero per la lettera del Future of Life Institute, pubblicata il 23 marzo, è ancor più vero per la lettera dell'Association for the Advancement of Artificial Intelligence, che segue a pochi giorni di distanza, il 5 aprile. Storica società scientifica, l'AAAI si propone come punto di riferimento per la comunità professionale dal 1979. Naturalmente si è data [un codice etico](#): vuoto e non impegnativo, come troppo spesso capita di vedere: basta citare una frase: "the entire computing profession benefits when the ethical decision-making process is accountable to and transparent to all stakeholders". Si può subito notare che l'intento non è portare la comunità professionale ad assumere impegni, l'impegno consiste al contrario nel difendere gli interessi della comunità professionale di fronte a qualsiasi situazione che ne minacci la posizione dominante e lo spazio d'azione. Lo stesso intento appare chiarissimo ora nella lettera aperta [Working together on our future whit AI](#). L'incipit ammonisce il popolo: "L'Intelligenza Artificiale sta già arricchendo le nostre vite, spesso in modi non percepiti". "Alimenta i nostri sistemi di navigazione, è utilizzata in migliaia di screening quotidiani di tumori, smista miliardi di lettere del servizio postale". E via dicendo: "ha svelato la struttura di migliaia di proteine", garantisce previsione meteorologiche dettagliate, fornisce idee che stimolano la creatività. "We believe that AI will be increasingly game-

changing" in ogni ambito.

Segue una modesta ammissione: "siamo consapevoli dei limiti e delle preoccupazioni sui progressi dell'Intelligenza Artificiale". Ma subito si rovescia la frittata, dichiarando innocenti i membri della comunità professionale e affermando che dei problemi se ne deve occupare qualcun altro: "Beyond technology, we see opportunities for work in policy, including efforts with standards, laws, and regulations". Si coglie anche l'occasione per chiedere investimenti pubblici destinati a ricerche sulla valutazione del rischio.

Per quanto riguarda la comunità professionale, ci si limita a dire che la si incoraggia ad incrementare gli sforzi indirizzati verso la sicurezza e l'affidabilità dell'AI e le sue influenze sulla società. In che modo? Parlandone di più in convegni e workshop "ed altre attività esistenti". Basta dunque aggiungere ai convegni una sezione dove si parli di "short-term and longer term of AI". Ciò che si deve quindi sottolineare è che esiste in ogni caso un punto che accomuna l'intera comunità professionale impegnata in un modo o in un altro nel campo della ricerca legata all'Intelligenza Artificiale. Accomuna coloro che mirano ad Intelligenze Artificiali specializzate in un compito e coloro che cercano l'Intelligenza Generale Artificiale. Accomuna i puri tecnici e gli imprenditori. Accomuna i profeti della Digital Disruption e i filosofi fiancheggiatori. Consapevoli tutti di appartenere alla stessa famiglia professionale e di avere interessi in comune da difendere. Tutti infatti appaiono nei fatti d'accordo su un punto: l'importante è che il dibattito, e le decisioni conseguenti, restino in mano alla lobby degli addetti ai lavori. Chi infatti è invitato a firmare queste lettere? Tecnici e scienziati, tecnocrati, imprenditori del settore - che riservano a se stessi l'autorità di decidere come autoregolarsi. Tutto in mano agli 'esperti'!

C'è motivo di ritenere che le lettere aperte degli esperti servano solo, in un momento in cui si ha di fronte una novità per certi aspetti sorprendente, a rinserrare le fila. E' facile per i tecnici considerare i cittadini non all'altezza di capire ed opinare. A questo scopo si evita ogni trasparenza, si usano gerghi escludenti. Oppure si offre una divulgazione volutamente superficiale, tesa a sollevare entusiasmi e a far apparire tecnofobi e retrogradi coloro che chiamano alla cautela. Si evita, insomma, un autentico dibattito pubblico, dove i cittadini siano messi in condizioni di comprendere e siano convocati a partecipare. Scrivo nelle [Cinque Leggi](#) che il primo passo in questa direzione dovrebbero compierlo i tecnici. Cercando la trasparenza nella narrazione del loro lavoro e nella descrizione dei sistemi da loro sviluppati. Ma soprattutto tornando a considerare sé stessi come cittadini. Spogliandosi del proprio potere, potrebbero chiedersi: cosa sto facendo? Comodo invece limitarsi a sottoscrivere lettere aperte.

Scende nell'agone il Garante italiano della Privacy

Il Garante italiano della Privacy avvia una istruttoria su OpenAI e la Chat GPT. Risultato: l'accesso alla Chat è negato ai cittadini italiani.

Motivi addetti dal Garante:

OpenAI non ha mai fornito una spiegazione dettagliata su come utilizzerà le informazioni che immagazzina

Il gruppo ha la sua base negli States. Sembra quindi che dati di cittadini europei siano raccolti su database situati fuori dalla comunità europea.

La Chat GPT non verifica l'età dell'utente. Quando in Europa la possibilità che un minore sia esposto ad informazioni 'inidonee' è considerato reato.

Naturalmente, c'è poi un ragionamento politico-strategico: lo stesso uso della Chat da parte dei cittadini, è fonte di alimentazione delle conoscenze che rendono possibile il funzionamento della Chat.

Il Garante afferma che la Chat utilizza dati dei quali i cittadini non hanno concesso l'uso. Questo può essere vero. Ma in molti casi, i dati personali che Open AI, come tanti altri grandi operatori digitali, usa, sono legittimati dal fatto che li abbiamo resi pubblici involontariamente noi cittadini non stando attenti alle disclaimer di siti e piattaforme, dove si dichiara che silenzio dell'utente è assenso, e cose simili.

Si può dire che l'atto del Garante appare una censura fondata su interpretazioni formali di norme

vigenti. Sul breve termine Open AI accetta la posizione del Garante bloccando il servizio in Italia. Presto troverà il modo di aggiustare il tiro.

Si possono riconoscere ai difensori civici le buone intenzioni. Ma ogni censura è vana, perdente, specie nei tempi digitali dove i confini nazionali e europei valgono ben poco. I loro interventi non hanno la minima di possibilità di incidere sulle ricerche, sulle strategie e sulle scelte di mercato delle grandi case digitali.

E ancora, l'intervento del Garante finisce per essere il sostituto di un vero dibattito pubblico. La comunità dei tecnici digitali intanto, nel mentre sparge un po' di lacrime di cocodrillo attraverso qualche lettera aperta, continua, in un modo o nell'altro, ad appropriarsi dei dati dei cittadini per costruire strumenti sempre più potenti per condizionare i cittadini stessi, rendendoli passivi utenti.

Il fascino della Chat GPT: sottile ed ingannevole

La censura è vana, perdente. Gli argomenti addotti dal Garante sono dubbi. Non si ferma certo così un trend. L'isolamento al quale l'intervento del Garante costringe chi opera nel nostro paese è dannoso.

Ma il coro di voci che si leva contro il Garante dà da pensare. Possiamo veramente considerarli imparziali nel giudicare? Non credo. Le voci che criticano l'intervento del Garante non sono né indipendenti né disinteressate. Sono voci di addetti ai lavori personalmente interessati al fatto che la giostra continui a girare e anzi giri sempre più veloce.

Denunciamo dunque la vanità delle prese di posizione del Garante. Ammettiamo però anche l'ipocrisia implicita in lettere come quella del Future of Life Institute. E ammettiamo anche il fascino pericoloso che esercita su di noi questa Intelligenza Artificiale Generativa. Sia chi ha, sia chi non ha, fiducia nel possibile avvento di una Intelligenza Artificiale Generale, sia chi questa fiducia non ce l'ha, sono affascinati dai risultati raggiunti da Open AI. Tutti passano il tempo a interrogare il nuovo oracolo! Non posso che tornare sugli argomenti espressi nella [Confutazione poetica della Chat GPT](#), pubblicata di recente su questo blog. E non posso fare a meno di ricordare quello che scrivo nelle [Cinque Leggi](#). Mi sembra che tra gli 'esperti' la grande maggioranza si ponga senza troppo ragionare nella scia di Turing, quando, [nell'articolo apparso su Mind nel 1950](#), dice: 'Spero che le macchine pensino'. La seconda delle leggi di cui parlo nel mio libro infatti recita: 'Preferirai la macchina a te stesso'. La quarta: 'Lascierai alla macchina in governo'. Un cittadino dovrebbe, secondo me, preferire sé stesso ad ogni macchina e non lasciare alla macchina il governo. Ma gli 'esperti' invitano invece i cittadini a fidarsi della macchina più che di sé stessi.

Vorrei anche aggiungere una nota strettamente politica. Perché le scelte che ci appaiono meramente scientifiche e tecnologiche, sono in realtà sempre anche scelte politiche. Gli esperti, i tecnologi, posso dire in genere i membri delle élite, possono ben dire che la Chat GPT è mero strumento, tramite il quale scaricarsi i compiti ripetitivi e inutilmente faticosi.

La Chat, però, lungi dal apparire strumento per aprire alla creatività, appare strumento di dominio a chi vive in condizioni di sudditanza.

L'élite che magnifica la Chat, e prova soddisfazione nell'usarla, nasconde così a sé stessa il fatto che in un contesto democratico, dove la democrazia non sia ridotta alla versione più banale e comoda del liberismo, l'élite svolge il suo ruolo solo se va oltre il comodo, esclusivo perseguimento dei propri interessi, e si fa carico di creare condizioni affinché ognuno, anche chi proviene da condizioni disagiate, possa giungere ad essere pienamente cittadino, attivo e responsabile. E' però evidente che la Chat GPT è usata al contrario: per imporre risposte ad ogni domanda, per svilire la ricerca, per costringere chi non appartiene all'élite in una condizione di definitiva sudditanza. Per qualcuno la Chat è un gioco, un divertimento, per altri è strumento di controllo sociale.

Ed in ogni caso, non dimentichiamolo, la Chat GPT è un modo per raccogliere dati e per garantire al sistema nuove informazioni sui comportamenti umani. Più alimentiamo Chat GPT, più ne diventiamo dipendenti.

Agenti, copiloti e pappagalli

Un articolo diffuso in pre-print il 13 aprile è molto citato in questi giorni, a conferma della

condivisa speranza degli 'esperti': [Sparks of Artificial General Intelligence: Early experiments with GPT-4](#). Gli autori -Sébastien Bubeck ed altri, lavorano nei laboratori di ricerca di Microsoft, la grande casa cui Open AI è più vicina. "Cosa sta realmente accadendo? Il nostro studio di GP 4 è interamente fenomenologico: ci siamo concentrati sulle cose sorprendenti che GP 4 può fare, ma non affrontiamo le domande fondamentali sul perché e come raggiunge un'intelligenza così straordinaria". Per questo l'articolo piace tantissimo agli addetti ai lavori: propone una entusiastica apologia della nuova tecnologia, eludendo qualsiasi domanda impegnativa, da un punto di vista sia tecnico che da punti di vista sociali, politici ed etici.

E' la filosofia del lavarsene le mani. La posizione che Bubeck, con una certa ingenuità, ammette, è la posizione generalmente diffusa tra gli addetti ai lavori. Delle conseguenze ci preoccuperemo veramente, se sarà il caso, in futuro. Intanto, firmiamo una lettera - che tanto, lo sappiamo, non bloccherà nessuna ricerca. E semmai il compito di bloccare spetta alla politica. Il business e l'industria devono andare avanti. I finanziatori attendono sempre più alti ritorni dell'investimento. Subito la ricerca deve essere applicata a prodotti.

Non a caso [Microsoft, sulla base degli studi che Open AI porta avanti, lancia in questi stessi giorni un nuovo tool della suite Office 365, dal nome emblematico: Copilot](#). "We spend too much time consumed by the drudgery of work on tasks that zap our time, creativity and energy. To reconnect to the soul of our work, we don't just need a better way". La *better way* proposta ad ogni cittadino del pianeta per l'uso del proprio tempo, dei propri personali spazi di libertà, nel lavoro e nella vita quotidiana, sarà indicata dal copilota digitale!

Satya Nadella, Chairman e CEO di Microsoft, infatti annuncia: "With our new copilot for work, we're giving people more agency and making technology more accessible through the most universal interface - natural language."

More agency: agency è l'ambigua parola che i filosofi digitali sostituiscono a *libertà*. Infatti nessuno osa tradurre *agency* con *libertà*, o neanche con *potere*, *capacità*. Si preferisce lasciare la parola in inglese, o avventurarsi nel neologismo *agentività*. L'espressione vuole affermare l'esistenza di un terreno sul quale appaiono -simili, fungibili, sostituibili l'un l'altro- l'umano e la macchina. La metrica buona per la macchina è per questa via imposta agli umani. La millenaria storia dell'umana ricerca dell'essere, de sé, dello spazio di autonomia, della responsabilità personale, è così comodamente elusa, o anzi rimossa. Anche di questo parlo in modo dettagliato e preoccupato nelle [Cinque Leggi](#).

Microsoft e Open AI, dunque, ci dicono che sanno ciò che è bene per ognuno di noi, e ci chiedono di fidarci di loro. Quando scrivo Microsoft e Open AI, possiamo intendere i tecnici di Microsoft e di Open AI, oppure, direttamente la Chat che i tecnici di Microsoft e Open AI hanno creato, Chat dotata di Intelligenza Artificiale che parla agli umani come se a parlarci fosse un umano. Ecco cosa intendo quando dico: cittadini ridotti a utenti. Il concetto di *Copilota* risulta infatti volutamente ambiguo. Mentre si lascia in apparenza libertà, si dà per scontato che convenga agli umani seguire i suoi melliflui consigli del *Copilota*. Mentre prosegue il lavoro di rendere, tramite l'uso di ricerche neuroscientifiche, sempre più difficile agli umani non seguire la via indicata, tramite acconci algoritmi, sui Social Network; mentre diventano più incombenti i 'consigli' che arrivano tramite notifiche, e ora tramite il *Copilota*. Agli 'esperti' seguaci di Turing piace essere guidati dalla macchina. Forse molti cittadini la pensano diversamente. Ma ora saranno ricondotti sulla retta via. In realtà, di fronte all'avvento della Chat GPT, qualcuno è tornato a citare un articolo del 2021, di grande interesse, critico e premonitore. [On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?](#) Emily Bender e le altre ricercatrici legate al centro di ricerca di Google, come si sa, hanno pagato cara questa presa di posizione. Non è questa la sede per riprenderne in dettaglio i diversi argomenti. Basta l'attenzione dedicata a mostrare i costi ambientali e finanziari della tecnologia fondata su modelli linguistici sempre più grandi. Le ricercatrici si chiedono *How big is too big?* Notano come non conta solo la quantità dei dati, ma anche la loro qualità. Conta anche, da un punto di vista etico, come sono stati raccolti i dati. Non è necessariamente questo -modelli linguistici di dimensioni sempre più grandi- l'unica direzione nella quale dovrebbe muoversi la ricerca, suggeriscono Binder e le coautrici. Ma certo se la gran parte

degli investimenti sono rivolti a questo tipo di ricerca, ben poco di differente potrà essere sperimentato.

Alla riflessione sui dati: la loro quantità e la loro qualità; e all'invito a non trascurare altri indirizzi di ricerca le ricercatrici aggiungono una presa di posizione etica. "Chiediamo di riconoscere che le applicazioni che mirano a imitare [...] gli esseri umani comportano il rischio di danni estremi [extreme harms]. Il lavoro sul comportamento umano sintetico è una linea di confine nello sviluppo dell'AI rispettoso dell'etica, dove "gli effetti a valle devono essere compresi e modellati per bloccare i danni alla società e a diversi gruppi sociali".

Sui rischi, e sulle necessarie cautele, la lettera aperta del Future of Life Institute non fa che ripetere quello che Bender e le sue colleghe avevano già detto. E comunque, come ho mostrato, la recente lettera dice in realtà che i ricercatori impegnati su questa frontiera non vogliono né fermarsi né cercare altre strade. Si contentano sempre di ribadire la banale affermazione: 'Intelligenza Artificiale robusta e benefica'. Non aggiunge molto la pretesa novità affermata dall'Unione Europea: Intelligenza Artificiale *human centered*. In ogni caso nessuno vuole davvero mettere limiti alla ricerca.

Di recente, Bender è tornata a far ascoltare la sua voce: "Sento che ci sono troppi sforzi per creare macchine autonome", ha detto. Mentre diminuiscono gli sforzi per "cercare di creare macchine che siano strumenti utili per gli esseri umani". Di fronte agli atteggiamenti di persone con le quali si è trovata a discutere in incontri pubblici, Bender si è sentita obbligata a precisare. "Non ho intenzione di conversare con persone che non pongono la mia umanità come assioma alla conversazione". Sono infatti molti i tecnici e i filosofi fiancheggiatori che amano ricordare che non conta essere umani o non esserlo. Siamo tutti, alla pari *agenti*.

Solo alla luce di questa scelta personale, esistenziale, etica e filosofica si può ben capire il senso del titolo dell'articolo del 2021. *Pappagalli stocastici*. Bender, consapevole di essere umana, consapevole di appartenere alla vita sulla terra, ad una storia, consapevole di come il proprio pensiero ed il proprio parlare siano legati al corpo umano, ha molti motivi per dire che la Chat GPT non è che un pappagallo stocastico. Possedendo un solido dominio della sintassi, si esprime in modo correttissimo, e quindi in apparenza convincente. Ma non possedendo capacità semantiche parla come un pappagallo, che imita il linguaggio umano, ma a differenza degli umani non sa quello che dice.

L'Intelligenza Artificiale Generale è ormai tra noi?

Qui l'opposizione è netta. Da un lato coloro che, nonostante le grandi prestazioni di GPT credono che si tratti di nient'altro che di un pappagallo stocastico. Dall'altro lato coloro che, come Bubeck, ritengono le prestazioni di GPT siano tanto sorprendenti da far pensare che -lo dice già nel titolo: *Sparks of Artificial General Intelligence*- vi si trovino scintille di Intelligenza Artificiale Generale. "Resta ancora molto da fare per creare un sistema che possa qualificarsi come Intelligenza Artificiale Generale completa", scrive Bubeck. Eppure Bubeck, come tanti esperti, è entusiasta, emozionato, di fronte alle *core mental capabilities* di GPT. Le elenca con ammirazione: *reasoning, creativity, deduction; variety of tasks it is able to perform: playing games, using tools, explaining itself...*

Resta ancora nutrita, potremmo anche dire maggioritaria, la schiera dei ricercatori e dei tecnici che intendono l'Intelligenza Artificiale come capacità applicata a macchine dedicate ad un singolo scopo. Essi non hanno mai creduto all'Intelligenza Artificiale Generale, né ne hanno fatto l'oggetto della propria ricerca. Ma molti di loro in questi giorni si stanno ricredendo.

Quella Superintelligenza, totalmente autonoma dall'intelligenza umana, a suo modo cosciente, che era stata cercata per settant'anni seguendo altre vie tecnologiche, sembra comunque ora apparire vicinissima a manifestarsi, come frutto del Deep Learning e dei Large Language Model: la si chiama Intelligenza Artificiale Generativa.

Chi ci parla davvero, senza enfasi, dei rischi esistenziali

Retorica e propaganda accompagnano la crescita dell'industria digitale. I tecnici impegnati nella

ricerca, spinti dall'ansia di sperimentare, scelgono di sottovalutare le conseguenze dei loro progetti: si limitano a firmare lettere aperte e al fare appello alla politica, che dovrebbe imporre norme e freni. Ma i politici, non sapendo cosa fare di fronte a ricerche di cui non capiscono senso e portata, si affidano a commissioni formate dagli stessi tecnici che rimandano le decisioni alla politica. Da questo circolo vizioso, non possono che nascere norme inefficaci.

Ai tecnici, poi, è stato insegnato che la misura del successo di una ricerca sta nei risultati finanziari generati dalla ricerca stessa. Il legame tra tecnici ricercatori e investitori risulta così ambiguo tanto quanto il legame tra tecnici e politici.

Le voci critiche realmente autonome sono quindi poche. Le poche, vanno sminuite e svalutate. Così accade appunto con l'articolo di Emily Binder e delle altre ricercatrici, *On Danger of Stochastic Parrots*.

Non solo sono disattesi gli accorati inviti di Binder: mettere un limite alla raccolta dei dati, evitare ricerche orientate. Ma nell'ansia di svalutare l'articolo si punta dritto alla tesi espressa nel titolo. Sam Altman, CEO di Open AI, il 4 dicembre 2022, quattro giorni dopo il rilascio di Chat GPT, scrive in un tweet: "Sono un pappagallo stocastico". Crede di essere ironico.

Bender risponde: "Le persone vogliono credere così tanto che questi modelli linguistici siano effettivamente intelligenti che sono disposte a prendere se stesse come punto di riferimento e a svalutare questo aspetto per adeguarsi a ciò che il modello linguistico può fare".

"Vogliono credere": vi colgo un richiamo all'"I hope" di Turing. Specchiarsi nella macchina. Il 'gioco dell'imitazione', di cui parlava Turing nell'articolo del 1950, i seguaci turinghiani l'hanno poi pomposamente definito *test di Turing*. Turing ridefinisce l'intelligenza ed il pensiero umani, in modo tale da poter arrivare a dire: 'l'intelligenza ed il pensiero corrispondenti alla definizione, si ritrovano nella macchina'. Insomma, torno a dirlo con la seconda delle [Cinque Leggi](#): 'Preferirai la macchina a te stesso'. Prenderai la macchina a modello di te stesso. Il Gemello digitale di ogni entità fisica, anche di ogni essere umano, è un passo in questa direzione.

Ma in realtà, nessuno si ferma. La ricerca prosegue. L'argomento dei pappagalli stocastici viene usato come arma retorica contro Binder. Bubeck osserva emozionata le scintille di Intelligenza Artificiale Generale che traspiono nei GPT.

Non sappiamo ancora ben immaginare, e tanto meno regolamentare, un mondo in cui quell'intelligenza esiste. Eppure la lettera aperta di Future of Life si apre con una frase precisa: "AI systems with human-competitive intelligence can pose profound risks to society and humanity". Perché non ce ne preoccupiamo in modo serio? Credo che per rispondere a questa domanda, convenga seguire la lezione di Eliezer Yudkowsky. Scrivo di lui nelle [Cinque Leggi](#), p. 225 e seguenti. Da lui viene il pensiero dal quale è nata la lettera aperta di Future of Life.

Comunque lo si giudichi, Yudkowsky ha posto vent'anni fa il tema dei rischi esistenziali legati all'Intelligenza Artificiale. E nel corso degli anni è divenuto via via più pessimista. Cosa dice ora Yudkowsky di GPT? Parlo di questo in un [articolo pubblicato su Agenda Digitale](#).