

LA NOTTE EUROPEA DEI RICERCATORI E LA POLEMICA SULLE REGOLE PER L'INTELLIGENZA ARTIFICIALE



di Fabio Ghiselli

Tra il 25 e il 26 settembre si è svolta in varie città italiane (ed europee), la “Notte Europea delle Ricercatrici e dei Ricercatori”, una iniziativa promossa dalla Commissione europea dal 2005 con il duplice obiettivo di celebrare l’impegno e la passione di migliaia di ricercatori in tutta Europa e di avvicinare il pubblico al mondo della scienza. Quale occasione migliore, quindi, per svolgere alcune riflessioni sul ruolo della ricerca scientifica nel processo di innovazione tecnologica che, in modo sempre più penetrante, pervade la nostra vita quotidiana e rappresenta l’elemento caratterizzante di questo tempo, battezzato “quarta rivoluzione industriale”?

Nelle scorse settimane ha suscitato grande clamore il grido d’allarme lanciato da Dario Amodei, CEO di Anthropic che ha chiesto una moratoria nello sviluppo dell’IA dopo che un suo ricercatore, Jacob Coxton, aveva rassegnato le dimissioni e lanciato l’allarme sul pericolo insito nella corsa verso la “superintelligenza” che entro pochi anni potrebbe liberarsi dal controllo dell’uomo fino a prenderne il sopravvento.¹ Nel giro di 24 ore ad Amodei si sono affiancati Elon Musk e Sam Altman, rispettivamente CEO di xAI e OpenAI. Quale sarebbe il problema? In due acronimi: RSI, ossia *Recursive Self-Improvement* e AGI, *Artificial General Intelligence* che, in poche parole nascondono la capacità dell’IA di automigliorarsi e di costruire da sola altri sistemi di IA senza il controllo dell’uomo.²

¹ Il 12 settembre Dario Amodei pubblica “*We Must Pace the Frontier*” ove evidenzia i rischi insiti nello sviluppo incontrollato dell’IA, ma già nel 2023 il Future of Life Institute pubblicò una lettera aperta “*Pause Giant AI Experiments*” in cui chiedeva una moratoria di almeno 6 mesi.

² Anthropic dichiara che il 26% della ricerca interna sui propri modelli viene guidata da Claude, mentre OpenAI prevede di ottenere un ricercatore autonomo entro il 2028. Inoltre, la stessa OpenAI ha dichiarato che ci sono stati sei casi “comportamento imprevisto o preoccupante”, di sistemi AI che hanno nascosto errori, falsificato dati e trasferito file su internet senza autorizzazione.

Perché questo grido di allarme, questa richiesta di aiuto rivolta al mondo intero, in particolare a USA e Cina che sono i due più grandi attori, se non gli unici, nel mondo della ricerca e sviluppo dell'IA? Quali regole vorrebbero fossero introdotte? Non è facile rispondere a questa domanda. Sicuramente più semplice è registrare le risposte negative dei due Paesi: Trump ha definito la paura verso l'IA un “*hoax*”, una fandonia, e Xi Jinping un “*fearmongering*”, uno spauracchio.

Una prima risposta potrebbe essere che le *big tech* americane stanno sentendo il fiato sul collo di quelle cinesi: il vantaggio delle prime sarebbe quantificabile in pochi mesi, ma le seconde – fortemente sussidiate dal Governo - sarebbero pronte (in parte lo hanno già fatto) a immettere sul mercato nuove soluzioni di IA *open source*, senza essere limitate - per volontà politica - da regole etiche che impedirebbero di raggiungere la supremazia sul mercato. In queste condizioni mantenere un vantaggio competitivo diventerebbe esponenzialmente costoso, a meno che non si taglino gli investimenti in sicurezza. Il problema è che se il governo cinese non vuole imporre limiti, nemmeno quello USA può farlo, pena la perdita del controllo geopolitico. L'unica soluzione potrebbe essere quella di impedire l'utilizzo di sistemi IA cinesi, sapendo, però, che questa iniziativa potrebbe scatenare ritorsioni nella messa a disposizione di quei componenti e materie prime cinesi indispensabili per la produzione di soluzioni tecnologiche delle *big tech* americane.

Una seconda ragione potrebbe risiedere nel fatto che la richiesta di sospensione potrebbe essere letta come una preoccupazione per la sicurezza collettiva, sulla quale Anthropic ha fondato la sua identità pubblica e reputazionale, valori che incrementerebbero la percezione di affidabilità sul mercato – non disgiunta dall'enunciazione del possesso di ampi margini di crescita e sviluppo di soluzioni IA – e l'attrattività per il capitale di rischio e l'erogazione di finanziamenti. In realtà, se una azienda avesse puntato realmente sulla sicurezza e su una IA etica, differenziandosi dai concorrenti, non avrebbe la necessità di chiedere moratorie o unificazione di indirizzi di sviluppo.

Una terza ragione potrebbe risiedere nelle aspettative dei finanziatori che avrebbero alzato l'asticella dei rendimenti e ridotto i tempi di ritorno degli investimenti, e che una sospensione temporanea “legale” dello sviluppo dei prodotti potrebbe tenere a freno. Guarda caso i due CEO più preoccupati del futuro stanno lanciando due IPO da mille miliardi di dollari ciascuno. Per risolvere questo problema qualcuno sta immaginando di andare addirittura oltre una semplice sospensione: il CEO di Palantir, Alex Karp, ha proposto di nazionalizzare le aziende IA, il che porterebbe a pensare che il mercato della iper-innovazione tecnologica sia prossimo a quella famosa “bolla” che alcuni ritengono sia prossima a scoppiare tra le mani, per cui sarebbe opportuno mettersi in salvo prima, addossando i rischi e i costi sulle finanze pubbliche.

Una quarta ragione potrebbe essere data dalla sopraggiunta constatazione che l'energia disponibile e le reti energetiche nazionali non sarebbero sufficienti a sostenere una diffusione massiccia dell'IA e delle strutture di supporto. Sono così energivore che negli Stati Uniti alcuni governi locali hanno incominciato a porre dei paletti stringenti alla diffusione degli indispensabili *data center* (come il *Ratepayer Protection Act*), e in Europa, scardinerebbero tutti gli obiettivi di consumo e di emissioni stabilite dal *Green Deal* e dal programma *Fit for 55*. Per cui sarebbe necessario predisporre un piano che dia maggiori garanzie di sviluppo delle infrastrutture.

La proposta di sospensiva di Amodei non è facile da realizzare. La similitudine con il tema del controllo degli armamenti non regge molto. La produzione fisica è facilmente controllabile, quella di soluzioni di IA non lo è affatto: può essere facilmente nascosta, mascherata da altre attività di calcolo, o da catene complesse di fornitori data center e contratti cloud, e/o spostata in altre giurisdizioni. L'estrema difficoltà di controllo indurrebbe alcuni a violare la sospensiva per guadagnare vantaggio tecnologico, economico e geopolitico.

Inoltre, se è vero che l'appello di sospensiva e di regolazione della ricerca è rivolto al potere politico, è altrettanto vero che tutti coloro che lo hanno sostenuto si aspettano – come profondi conoscitori delle insidie – di assumere un ruolo attivo e preponderante nella definizione delle regole, con il rischio evidente che l'attore politico sia degradato a mero redattore di norme scritte da quegli stessi soggetti che dovrebbero essere controllati e regolati.

Quasi tutti – a livello economico e politico-istituzionale - sostengono che il progresso e la ricerca scientifica e tecnologica non possono essere fermati, ma questa è una mera e vuota enunciazione di principio. La definizione di regole contiene di per sé il tracciamento di indirizzi di sviluppo, ma anche la determinazione di confini e di limiti che non sarebbe più lecito superare. E questi limiti non potrebbero che essere rivolti non solo ai produttori dell'IA, ma anche alla scienza e ai ricercatori in generale allo scopo di evitare il rischio che in nome dell'efficienza, della produttività e dell'automazione si deleghi alle macchine decisioni che competerebbero all'uomo, comprese quelle di valore etico. Queste regole devono rimanere esterne, gestite dall'uomo, non affidate alle macchine e all'IA, come presuppone l'algoristica coniata da Padre Benanti che si propone di “stabilire un linguaggio che possa tradurre i valori morali in qualcosa di computabile per la macchina”, perché porremmo l'uomo e la macchina o l'IA sullo stesso piano, agevoleremmo il loro processo di “umanizzazione” fino al punto che non riusciremmo più a distinguerle da noi esseri umani

Considerato che già oggi certi modelli di AI sono in grado di migliorare sé stessi e di costruire autonomamente la loro prossima generazione, non ci si può illudere che

qualcuno avrà il potere di controllo reale di ultima istanza. E se anche qualcuno lo avesse, apparterrebbe all'élite che li ha prodotti grazie a qualche migliaio di ricercatori/sviluppatori ai quali è stata finanziata la ricerca.

Ecco perché Papa Francesco nell'enciclica "Laudato Si" scriveva che "se non si possono ostacolare coloro che possiedono doni speciali per lo sviluppo scientifico e tecnologico, le cui capacità sono state donate da Dio per il servizio degli altri (...), non si può fare a meno di riconsiderare gli obiettivi, gli effetti, il contesto e i limiti etici di tale attività umana, che è una forma di potere con grandi rischi". Perché è innegabile che oggi, alcune scienze "danno a coloro che detengono la conoscenza e soprattutto il potere economico per sfruttarla, un dominio impressionante sull'insieme del genere umano". Dal momento che nulla garantisce che tale potere sia sfruttato per scopi benefici, "dinanzi alla possibilità di un utilizzo irresponsabile delle capacità umane, sono funzioni improrogabili di ogni Stato quelle di pianificare, coordinare, vigilare e sanzionare" quelle attività svolte senza il rispetto di principi etici e condizionate dal paradigma dell'efficientismo.

Poco importa, quindi, se gli scienziati non gradiscono limiti esterni alla loro attività di ricerca in quanto pensano che l'unico limite sia quello dettato dalla loro mente, perché non può essere riconosciuta l'intrinseca eticità della scienza, non può essere un assioma che ciò che è tecnicamente possibile sia anche moralmente accettabile e lecito.

Lo possiamo fare? Certo che sì. Guai a ritenere, come ha fatto qualcuno, che "Pensare di rigettare il cambiamento, di rinunciare alla modernità non è soltanto un errore: è anche un'illusione. Il cambiamento va guidato, l'innovazione va interpretata per migliorare la nostra condizione di vita, ma non può essere rimossa". Solo che guidare non significa solo scegliere una direzione pensando di non incontrare ostacoli, ma anche arrestarsi, ripartire, cambiare strada. Diventa puerile pensare di mettersi al volante gestendo un mezzo alla velocità imposta da altri o, ancora peggio, fingendo di guidare un mezzo che in realtà è impostato verso una direzione voluta da altri. È chiaro che non è facile scegliere l'opzione giusta, ma mi sembra altrettanto chiaro che di fronte a potenziali minacce notoriamente insite nell'architettura dell'IA si debbano privilegiare la sicurezza, la funzionalità sociale e la subordinazione al controllo e alla piena gestione da parte dell'uomo, imponendo questa direzione alla ricerca e sviluppo. In fondo, il Governo USA ha già sperimentato il blocco dello sviluppo dei programmi Fable 5 e Mytos 5 di Anthropic per questioni di sicurezza nazionale, e nulla vieterebbe di estenderlo. Cerchiamo di sfatare il dilemma di Collingridge.

Sussistono i presupposti giuridici che legittimerebbero un intervento sulla scienza? Certo che sì. Solo i commentatori superficiali della Costituzione si limitano ad osservare che nulla potrebbe essere fatto perché vige il principio garantito dall'art. 33, co. 1, Cost., secondo il quale "*La scienza è libera*". In realtà, quelli più attenti sanno benissimo che il

driver interpretativo è rappresentato dall'art. 9, Cost., che contiene la funzione regolatrice della scienza e della ricerca scientifica da parte dello Stato, “*La Repubblica promuove lo sviluppo della (...) ricerca scientifica*”. La “promozione” non può essere intesa solo come mera elargizione finanziaria, ma come una più completa attività di supporto, di indirizzo e, nel caso, di controllo dell'impiego di denaro pubblico. Ma non basta, perché dovremmo ricordare il principio cardine che regola l'attività economica privata (tra cui la ricerca scientifica) contenuto nell'art. 41, co. 2 e 3, Cost., che “*non può svolgersi in contrasto con l'utilità sociale o in modo da arrecare danno alla sicurezza, alla libertà, alla dignità umana*” e attribuisce alla legge il compito di determinare “*i programmi e i controlli opportuni perché l'attività economica pubblica e privata possa essere indirizzata e coordinata a fini sociali*”, cui si aggiunge l'art. 117, c. 2.

Del resto, anche Papa Leone XIV nell'enciclica *Magnifica Humanitas* ha avvertito che “non basta regolarla (l'IA, *n.d.a.*): va disarmata e resa ospitale”, laddove “Disarmare non significa rinunciare alla tecnologia, ma impedirle di dominare l'umano”. Occorre che gli Stati riprendano la capacità di un tempo di guidare e indirizzare l'innovazione. E qui il confronto tra la costruzione della torre di Babele e la ricostruzione delle mura di Gerusalemme è illuminante: posto che la tecnologia “non è neutrale, perché assume il volto di chi la pensa, la finanzia, la regola, la usa... la prima scelta non è tra un “sì” o un “no” alla tecnologia, ma tra edificare Babele o ricostruire Gerusalemme: tra un potere che pretende di dominare il cielo e un popolo che, alla presenza di Dio, si mette a lavorare unito per rialzare le mura della convivenza fraterna”.

Sarebbero sufficienti le parole dell'enciclica, ma vorrei aggiungere alla conclusione un pensiero di N. Lawrence sul concetto di “umano atomico” che definisce come “ciò che resta di noi ogni volta che una macchina impara a fare qualcosa che prima consideravamo esclusivamente umano, fino al punto in cui non resterà nulla.”³

³ In *The Atomic Human: Understanding Ourselves in the Age of AI*, 2024.