Il paradosso del potere dell'IA - Gli Stati possono imparare a governare l'intelligenza artificiale, prima che sia troppo tardi?

foreignaffairs.com/world/artificial-intelligence-power-paradox

16 agosto 2023

Daniele Lievano

È il 2035 e l'intelligenza artificiale è ovunque. I sistemi di intelligenza artificiale gestiscono ospedali, gestiscono compagnie aeree e si combattono in aula. La produttività è aumentata a livelli senza precedenti e innumerevoli aziende prima inimmaginabili sono cresciute a una velocità vertiginosa, generando enormi progressi nel benessere. Nuovi prodotti, cure e innovazioni arrivano quotidianamente sul mercato, mentre la scienza e la tecnologia entrano in azione. Eppure il mondo sta diventando più imprevedibile e più fragile, mentre i terroristi trovano nuovi modi per minacciare le società con armi informatiche intelligenti e in evoluzione e i colletti bianchi perdono il lavoro en
massa.

—

Solo un anno fa, quello scenario sarebbe sembrato puramente fittizio; oggi sembra quasi inevitabile. I sistemi di intelligenza artificiale generativa possono già scrivere in modo più chiaro e persuasivo della maggior parte degli esseri umani e possono produrre immagini originali, opere d'arte e persino codici di computer basati su

semplici suggerimenti linguistici. E l'IA generativa è solo la punta dell'iceberg. Il suo arrivo segna un momento del Big Bang, l'inizio di una rivoluzione tecnologica che cambierà il mondo e che rimodellerà la politica, le economie e le società.

Come le ondate tecnologiche del passato, l'intelligenza artificiale abbinerà una crescita e opportunità straordinarie a sconvolgimenti e rischi immensi. Ma a differenza delle ondate precedenti, avvierà anche un cambiamento sismico nella struttura e nell'equilibrio del potere globale, poiché minaccia lo status degli stati-nazione come principali attori geopolitici del mondo. Che lo ammettano o no, i creatori dell'IA sono essi stessi attori geopolitici e la loro sovranità sull'IA rafforza ulteriormente l'emergente ordine "tecnopolare", un ordine in cui le società tecnologiche esercitano il tipo di potere nei loro domini un tempo riservato agli stati-nazione. Nell'ultimo decennio, le grandi aziende tecnologiche sono effettivamente diventate attori indipendenti e sovrani nei regni digitali che hanno creato. L'intelligenza artificiale accelera questa tendenza e la estende ben oltre il mondo digitale. La complessità della tecnologia e la velocità del suo avanzamento renderanno quasi impossibile per i governi emanare norme pertinenti a un ritmo ragionevole. Se i governi non si mettono al passo presto, è possibile che non lo faranno mai.

Per fortuna, i responsabili politici di tutto il mondo hanno iniziato a prendere coscienza delle sfide poste dall'intelligenza artificiale e lottare per governarla. Nel maggio 2023, il G-7 ha lanciato il "processo AI di Hiroshima", un forum dedicato all'armonizzazione della governance dell'IA. A giugno, il Parlamento europeo ha approvato una bozza della legge sull'IA dell'UE, il primo tentativo completo dell'Unione europea di erigere salvaguardie intorno al settore dell'IA. E a luglio, il segretario generale delle Nazioni Unite Antonio Guterres ha chiesto l'istituzione di un watchdog globale per la regolamentazione dell'IA. Nel frattempo, negli Stati Uniti, i politici su entrambi i lati della navata chiedono un'azione normativa. Ma molti sono d'accordo con Ted Cruz, il senatore repubblicano del Texas, che a giugno ha concluso che il Congresso "non sa cosa diavolo sta facendo".

Sfortunatamente, gran parte del dibattito sulla governance dell'IA rimane intrappolato in un pericoloso falso dilemma: sfruttare l'intelligenza artificiale per espandere il potere nazionale o soffocarla per evitarne i rischi. Anche coloro che diagnosticano accuratamente il problema stanno cercando di risolverlo inserendo l'IA in strutture di governance esistenti o storiche. Tuttavia, l'IA non può essere governata come qualsiasi tecnologia precedente e sta già modificando le nozioni tradizionali di potere geopolitico.

La sfida è chiara: progettare un nuovo framework di governance adatto a questa tecnologia unica. Se si vuole che la governance globale dell'IA diventi possibile, il sistema internazionale deve superare le concezioni tradizionali di sovranità e accogliere al tavolo le aziende tecnologiche. Questi attori potrebbero non trarre legittimità da un contratto sociale, dalla democrazia o dalla fornitura di beni pubblici, ma senza di loro un'efficace governance dell'IA non avrà alcuna possibilità. Questo è un esempio di come la comunità internazionale dovrà ripensare i presupposti di base sull'ordine geopolitico. Ma non è l'unico.

Una sfida insolita e urgente come l'IA richiede una soluzione originale. Prima che i responsabili politici possano iniziare a elaborare una struttura normativa appropriata, dovranno concordare i principi di base su come governare l'IA. Per cominciare, qualsiasi quadro di governance dovrà essere precauzionale, agile, inclusivo, impermeabile e mirato. Costruire su

questi principi, i responsabili politici dovrebbero creare almeno tre regimi di governance sovrapposti: uno per stabilire i fatti e consigliare i governi sui rischi posti dall'intelligenza artificiale, uno per prevenire una corsa agli armamenti a tutto campo tra di loro e uno per gestire le forze dirompenti di una tecnologia diversa tutto ciò che il mondo ha visto.

Piaccia o no, il 2035 sta arrivando. Il fatto che sia definito dai progressi positivi consentiti dall'IA o dalle interruzioni negative causate da essa dipende da ciò che fanno ora i responsabili politici.

PIÙ VELOCE PIÙ ALTO PIÙ FORTE

L'intelligenza artificiale è diversa, diversa dalle altre tecnologie e diversa nei suoi effetti sul potere. Non pone solo sfide politiche; la sua natura iperevolutiva rende anche la risoluzione di queste sfide progressivamente più difficile. Questo è il paradosso del potere dell'IA.

Il ritmo del progresso è sbalorditivo. Prendiamo la legge di Moore, che ha predetto con successo il raddoppio della potenza di calcolo ogni due anni. La nuova ondata di intelligenza artificiale fa sembrare strano quel ritmo di progresso. Quando OpenAI ha lanciato il suo primo modello di linguaggio di grandi dimensioni, noto come GPT-1, nel 2018, aveva 117 milioni di parametri, una misura della scala e della complessità del sistema. Cinque anni dopo, si pensa che il modello di quarta generazione dell'azienda, GPT-4, ne abbia oltre un trilione. La quantità di calcolo utilizzata per addestrare i modelli di intelligenza artificiale più potenti è aumentata di un fattore dieci ogni anno negli ultimi dieci anni. In altre parole, i modelli di intelligenza artificiale più avanzati di oggi, noti anche come modelli "di frontiera", utilizzano cinque miliardi di volte il

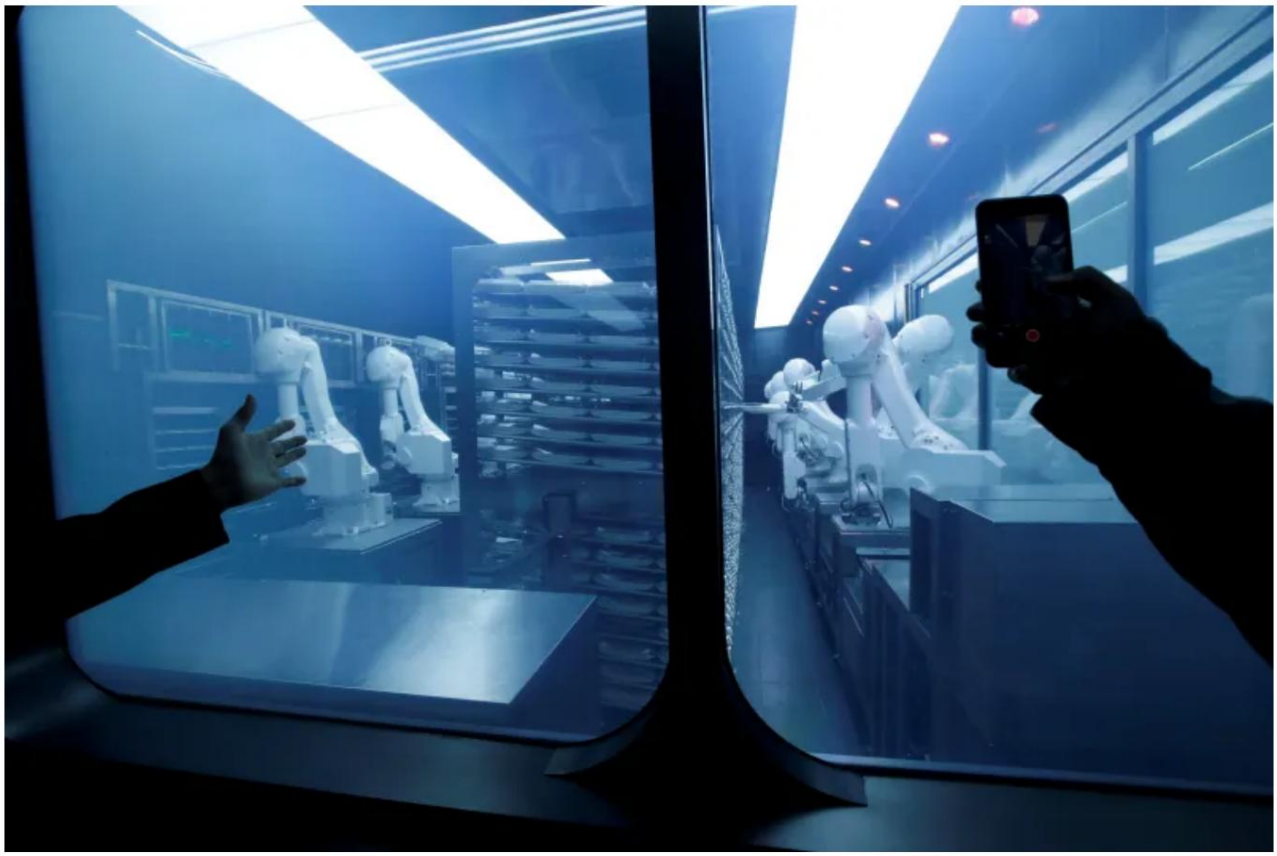
potenza di calcolo di modelli all'avanguardia di un decennio fa. L'elaborazione che una volta richiedeva settimane ora avviene in pochi secondi. Nei prossimi due anni arriveranno modelli in grado di gestire decine di trilioni di parametri. I modelli di "scala del cervello" con oltre 100 trilioni di parametri, all'incirca il numero di sinapsi nel cervello umano, saranno fattibili entro cinque anni.

Con ogni nuovo ordine di grandezza emergono capacità inaspettate. Pochi prevedevano che la formazione sul testo grezzo avrebbe consentito a modelli linguistici di grandi dimensioni di produrre frasi coerenti, nuove e persino creative. Ancora meno si aspettavano che i modelli linguistici fossero in grado di comporre musica o risolvere problemi scientifici, come alcuni ora possono fare. Presto, gli sviluppatori di intelligenza artificiale riusciranno probabilmente a creare sistemi con capacità di auto-miglioramento, un punto critico nella traiettoria di questa tecnologia che dovrebbe far riflettere tutti.

Anche i modelli di intelligenza artificiale stanno facendo di più con meno. Le funzionalità all'avanguardia di ieri vengono eseguite oggi su sistemi più piccoli, più economici e più accessibili. Appena tre anni dopo che OpenAI ha rilasciato GPT-3, i team open source hanno creato modelli capaci dello stesso livello di prestazioni che sono meno di un sessantesimo delle sue dimensioni, ovvero 60 volte più economici da eseguire in produzione, completamente gratuiti e disponibile a tutti su Internet. I futuri modelli di linguaggi di grandi dimensioni seguiranno probabilmente questa traiettoria di efficienza, diventando disponibili in forma open source solo due o tre anni dopo che i principali laboratori di intelligenza artificiale hanno speso centinaia di milioni di dollari per svilupparli.

Come con qualsiasi software o codice, gli algoritmi AI sono molto più facili ed economici da copiare e condividere (o rubare) rispetto alle risorse fisiche. I rischi di proliferazione sono evidenti. Il potente modello di linguaggio di grandi dimensioni Llama-1 di Meta, ad esempio, è trapelato su Internet pochi giorni dopo il suo debutto

a marzo. Sebbene i modelli più potenti richiedano ancora hardware sofisticato per funzionare, le versioni di fascia media possono essere eseguite su computer che possono essere noleggiati per pochi dollari l'ora.
Presto, tali modelli funzioneranno su smartphone. Nessuna tecnologia così potente è diventata così accessibile, così ampiamente, così rapidamente.



Robot che preparano il cibo in un ristorante hotpot a Pechino, novembre 2018
Jason Lee/Reuters

L'intelligenza artificiale differisce anche dalle tecnologie precedenti in quanto quasi tutte possono essere caratterizzate come "doppio uso", con applicazioni sia militari che civili. Molti sistemi sono intrinsecamente generali e, in effetti, la generalità è l'obiettivo principale di molte aziende di intelligenza artificiale. Vogliono che le loro applicazioni aiutino quante più persone possibile nel maggior numero di modi. Ma gli stessi sistemi che guidano le auto possono guidare i carri armati. Un'applicazione AI creata per diagnosticare le malattie potrebbe essere in grado di crearne e armarne una nuova. I confini tra il civile sicuro e il distruttivo militarmente sono intrinsecamente sfocati, il che spiega in parte perché gli Stati Uniti hanno limitato l'esportazione dei semiconduttori più avanzati verso la Cina.

Tutto questo si gioca su un campo globale: una volta rilasciati, i modelli di intelligenza artificiale possono e saranno ovunque. E ci vorrà solo un modello maligno o "breakout" per provocare il caos. Per questo motivo, la regolamentazione dell'IA non può essere fatta in modo patchwork. È di scarsa utilità regolamentare l'IA in alcuni paesi se rimane non regolamentata in altri. Poiché l'IA può proliferare così facilmente, la sua governance non può presentare lacune.

Inoltre, il danno che l'IA potrebbe causare non ha un limite evidente, anche se gli incentivi per costruirlo (ei vantaggi di farlo) continuano a crescere. L'intelligenza artificiale potrebbe essere utilizzata per generare e diffondere disinformazione tossica, erodendo la fiducia sociale e la democrazia; sorvegliare, manipolare e sottomettere i cittadini, minando la libertà individuale e collettiva; o per creare potenti

armi digitali o fisiche che minacciano vite umane. L'intelligenza artificiale potrebbe anche distruggere milioni di posti di lavoro, aggravando le disuguaglianze esistenti e creandone di nuove; radicare modelli discriminatori e distorcere il processo decisionale amplificando i circuiti di feedback delle informazioni errate; o innescare escalation militari involontarie e incontrollabili che portano alla guerra.

Né è chiaro il lasso di tempo per i rischi maggiori. La disinformazione online è un'ovvia minaccia a breve termine, proprio come la guerra autonoma sembra plausibile a medio termine.
Più lontano all'orizzonte si nasconde la promessa dell'intelligenza artificiale generale, il punto ancora incerto in cui l'IA supera le prestazioni umane in un determinato compito e il pericolo (certamente speculativo) che l'AGI possa diventare autodiretto, auto-replicante e auto-migliorante. al di fuori del controllo umano. Tutti questi pericoli devono essere presi in considerazione nell'architettura di governance sin dall'inizio.

L'intelligenza artificiale non è la prima tecnologia con alcune di queste potenti caratteristiche, ma è la prima a combinarle tutte. I sistemi di intelligenza artificiale non sono come automobili o aeroplani, che sono costruiti su hardware suscettibile di miglioramenti incrementali e i cui guasti più costosi si presentano sotto forma di incidenti individuali. Non sono come le armi chimiche o nucleari, che sono difficili e costose da sviluppare e immagazzinare, figuriamoci condividere o schierare segretamente. Man mano che i loro enormi vantaggi diventano evidenti, i sistemi di intelligenza artificiale diventeranno solo più grandi, migliori, più economici e più onnipresenti. Diventeranno persino capaci di quasi autonomia - in grado di raggiungere obiettivi concreti con una supervisione umana minima - e, potenzialmente, di auto-miglioramento.
Ognuna di queste caratteristiche sfiderebbe i modelli di governance tradizionali; tutti insieme rendono questi modelli irrimediabilmente inadeguati.

TROPPO POTENTE PER UNA PAUSA

Come se non bastasse, spostando la struttura e l'equilibrio del potere globale, l'IA complica lo stesso contesto politico in cui è governata. L'intelligenza artificiale non è solo lo sviluppo di software come al solito; è un mezzo completamente nuovo per proiettare il potere. In alcuni casi, sovvertirà le autorità esistenti; in altri, li trincererà. Inoltre, il suo progresso è spinto da incentivi irresistibili: ogni nazione, azienda e individuo ne vorrà una versione.

All'interno dei paesi, l'intelligenza artificiale consentirà a coloro che la esercitano di sorvegliare, ingannare e persino controllare le popolazioni, potenziando la raccolta e l'uso commerciale dei dati personali nelle democrazie e affinando gli strumenti di repressione che i governi autoritari usano per sottomettere le loro società. In tutti i paesi, l'intelligenza artificiale sarà al centro di un'intensa competizione geopolitica. Che sia per le sue capacità repressive, il potenziale economico o il vantaggio militare, la supremazia dell'IA sarà un obiettivo strategico di ogni governo con le risorse per competere. Le strategie meno fantasiose pompano denaro in campioni di intelligenza artificiale nostrani o tentano di costruire e controllare supercomputer e algoritmi. Strategie più sfumate favoriranno vantaggi competitivi specifici, come la Francia cerca di fare sostenendo direttamente le startup di intelligenza artificiale; il Regno Unito, sfruttando le sue università di livello mondiale e l'ecosistema del capitale di rischio; e l'UE, plasmando il dibattito globale su regolamentazione e norme.

La stragrande maggioranza dei paesi non ha né i soldi né il know-how tecnologico per competere per la leadership dell'IA. Il loro accesso all'IA di frontiera sarà invece determinato dalle loro relazioni con una manciata di aziende e stati già ricchi e potenti. Questa dipendenza minaccia di aggravare gli attuali squilibri di potere geopolitico. I governi più potenti gareggeranno per il controllo della risorsa più preziosa del mondo mentre, ancora una volta, i paesi del Sud del mondo saranno lasciati indietro. Questo non vuol dire che solo i più ricchi beneficeranno della rivoluzione dell'IA. Come Internet e gli smartphone, l'IA proliferà senza rispetto per i confini, così come i guadagni di produttività che scatenerà. E come l'energia e la tecnologia verde, l'intelligenza artificiale andrà a vantaggio di molti paesi che non la controllano, compresi quelli che contribuiscono alla produzione di input di intelligenza artificiale come i semiconduttori.

All'altra estremità dello spettro geopolitico, tuttavia, la competizione per la supremazia dell'IA sarà feroce. Alla fine della Guerra Fredda, paesi potenti avrebbero potuto cooperare per placare le reciproche paure e arrestare una corsa agli armamenti tecnologici potenzialmente destabilizzante.
Ma il teso ambiente geopolitico di oggi rende tale cooperazione molto più difficile. L'intelligenza artificiale non è solo un altro strumento o un'arma che può portare prestigio, potere o ricchezza. Ha il potenziale per consentire un significativo vantaggio militare ed economico sugli avversari.
A torto oa ragione, i due attori che contano di più, Cina e Stati Uniti, vedono entrambi lo sviluppo dell'IA come un gioco a somma zero che darà al vincitore un vantaggio strategico decisivo nei decenni a venire.

> Sia la Cina che gli Stati Uniti vedono lo sviluppo dell'IA come un gioco a somma zero.

Dal punto di vista di Washington e Pechino, il rischio che l'altra parte ottenga un vantaggio nell'intelligenza artificiale è maggiore di qualsiasi rischio teorico che la tecnologia potrebbe porre alla società o alla propria autorità politica interna. Per questo motivo, sia il governo degli Stati Uniti che quello cinese stanno investendo immense risorse nello sviluppo delle capacità di intelligenza artificiale mentre lavorano per privarsi a vicenda degli input necessari per le scoperte di prossima generazione.
(Finora, gli Stati Uniti hanno avuto molto più successo della Cina nel fare quest'ultimo, specialmente con i suoi controlli sulle esportazioni di semiconduttori avanzati.) Questa dinamica a somma zero - e la mancanza di fiducia da entrambe le parti - significa che Pechino e Washington stanno focalizzato sull'accelerazione dello sviluppo dell'IA, piuttosto che sul rallentamento. A loro avviso, una "pausa" nello sviluppo per valutare i rischi, come richiesto da alcuni leader del settore dell'IA, equivarrebbe a uno sciocco disarmo unilaterale.

Ma questa prospettiva presuppone che gli stati possano affermare e mantenere almeno un certo controllo sull'IA. Questo potrebbe essere il caso della Cina, che ha integrato le sue società tecnologiche nel tessuto dello stato. Tuttavia, in Occidente e altrove, è più probabile che l'intelligenza artificiale indebolisca il potere statale piuttosto che rafforzarlo. Al di fuori della Cina, una manciata di grandi aziende specializzate in intelligenza artificiale attualmente controlla ogni aspetto di questa nuova ondata tecnologica: cosa possono fare i modelli di intelligenza artificiale, chi può accedervi, come possono essere utilizzati e dove possono essere distribuiti. E poiché queste aziende custodiscono gelosamente la loro potenza di calcolo e i loro algoritmi, solo loro comprendono (la maggior parte) cosa stanno creando e (la maggior parte) cosa possono fare quelle creazioni. Queste poche aziende potrebbero mantenere il loro vantaggio per il prossimo futuro, oppure potrebbero

essere eclissato da una serie di attori più piccoli poiché le basse barriere all'ingresso, lo sviluppo open source e i costi marginali quasi nulli portano a una proliferazione incontrollata dell'IA. In ogni caso, la rivoluzione dell'IA avverrà al di fuori del governo.

In misura limitata, alcune di queste sfide assomigliano a quelle delle precedenti tecnologie digitali. Le piattaforme Internet, i social media e persino i dispositivi come gli smartphone operano tutti, in una certa misura, all'interno di sandbox controllate dai loro creatori. Quando i governi hanno fatto appello alla volontà politica, sono stati in grado di implementare regimi normativi per queste tecnologie, come il regolamento generale sulla protezione dei dati dell'UE, la legge sui mercati digitali e la legge sui servizi digitali. Ma tale regolamentazione ha impiegato un decennio o più per concretizzarsi nell'UE, e non si è ancora completamente concretizzata negli Stati Uniti. L'intelligenza artificiale si muove troppo rapidamente perché i responsabili politici possano rispondere al ritmo abituale. Inoltre, i social media e altre vecchie tecnologie digitali non aiutano a crearsi da soli, e gli interessi commerciali e strategici che li guidano non si sono mai combinati allo stesso modo: Twitter e TikTok sono potenti, ma pochi pensano che possano trasformare l'economia globale.

Tutto ciò significa che, almeno per i prossimi anni, la traiettoria dell'IA sarà in gran parte determinata dalle decisioni di una manciata di imprese private, indipendentemente da ciò che faranno i politici a Bruxelles o Washington. In altre parole, i tecnologi, non i politici oi burocrati, eserciteranno l'autorità su una forza che potrebbe alterare profondamente sia il potere degli stati-nazione sia il modo in cui si relazionano tra loro. Ciò rende la sfida di governare l'IA diversa da qualsiasi cosa i governi abbiano affrontato prima, un atto di bilanciamento normativo più delicato - e con una posta in gioco più alta - di quanto qualsiasi politico abbia tentato.

BERSAGLIO IN MOVIMENTO, ARMA IN EVOLUZIONE

I governi sono già dietro la curva. La maggior parte delle proposte per governare l'IA la trattano come un problema convenzionale riconducibile alle soluzioni stato-centriche del ventesimo secolo: compromessi sulle regole elaborate da leader politici seduti attorno a un tavolo. Ma questo non funzionerà per l'IA.

Gli sforzi normativi fino ad oggi sono agli inizi e ancora inadeguati. La legge sull'IA dell'UE è il tentativo più ambizioso di governare l'IA in qualsiasi giurisdizione, ma si applicherà integralmente solo a partire dal 2026, quando i modelli di intelligenza artificiale saranno avanzati oltre il riconoscimento. Il Regno Unito ha proposto un approccio volontario ancora più flessibile alla regolamentazione dell'IA, ma non ha i denti per essere efficace. Nessuna delle due iniziative tenta di governare lo sviluppo e l'implementazione dell'IA a livello globale, cosa che sarà necessaria per il successo della governance dell'IA. E mentre gli impegni volontari a rispettare le linee guida sulla sicurezza dell'IA, come quelle fatte a luglio da sette importanti sviluppatori di intelligenza artificiale, tra cui Inflection AI, guidato da uno di noi (Suleyman), sono i benvenuti, non sostituiscono la regolamentazione nazionale e internazionale legalmente vincolante.

I sostenitori degli accordi a livello internazionale per domare l'IA tendono a raggiungere il modello del controllo delle armi nucleari. Ma i sistemi di intelligenza artificiale non solo sono infinitamente più facili da sviluppare, rubare e copiare rispetto alle armi nucleari; sono controllati da compagnie private, non da governi. Mentre la nuova generazione di modelli di intelligenza artificiale si diffonde più velocemente che mai, il confronto nucleare sembra sempre più obsoleto. Anche se i governi possono controllare con successo l'accesso ai materiali necessari per costruire i modelli più avanzati - come sta tentando di fare l'amministrazione Biden impedendo alla Cina di acquisire chip avanzati - possono fare ben poco per fermare la proliferazione di quei modelli una volta che sono stati addestrati e richiedono quindi molti meno chip per funzionare.

Affinché la governance globale dell'IA funzioni, deve essere adattata alla natura specifica della tecnologia, alle sfide che pone e alla struttura e all'equilibrio di potere in cui opera. Ma poiché l'evoluzione, gli usi, i rischi e i benefici dell'IA sono imprevedibili, la governance dell'IA non può essere completamente specificata all'inizio o in qualsiasi momento, se è per questo.
Deve essere tanto innovativa ed evolutiva quanto la tecnologia che cerca di governare, condividendo alcune delle caratteristiche che rendono l'IA una forza così potente in primo luogo. Ciò significa partire da zero, ripensare e ricostruire un nuovo quadro normativo dalle fondamenta.

L'obiettivo generale di qualsiasi architettura normativa globale dell'IA dovrebbe essere quello di identificare e mitigare i rischi per la stabilità globale senza soffocare l'innovazione dell'IA e le opportunità che ne derivano. Chiamate questo approccio "tecnoprudenziale", un mandato piuttosto simile al ruolo macroprudenziale svolto da istituzioni finanziarie globali come il Consiglio per la stabilità finanziaria, la Banca dei regolamenti internazionali e il Fondo monetario internazionale. Il loro obiettivo è identificare e mitigare i rischi per la stabilità finanziaria globale senza compromettere la crescita economica.

Guardie a una conferenza Huawei a Shanghai, settembre 2019
Aly Song / Reuters

Un mandato tecnoprudenziale funzionerebbe in modo simile, richiedendo la creazione di meccanismi istituzionali per affrontare i vari aspetti dell'IA che potrebbero minacciare la stabilità geopolitica. Questi meccanismi, a loro volta, sarebbero guidati da principi comuni che sono entrambi adattati alle caratteristiche uniche dell'IA e riflettono il nuovo equilibrio di potere tecnologico che ha messo le aziende tecnologiche al posto di guida. Questi principi aiuterebbero i responsabili politici a elaborare quadri normativi più granulari per governare l'IA man mano che si evolve e diventa una forza più pervasiva.

Il primo e forse più vitale principio per la governance dell'IA è la precauzione. Come suggerisce il termine, il tecnoprudentialismo è fondamentalmente guidato dal credo precauzionale: primo, non nuocere. Limitare al massimo l'IA significherebbe rinunciare ai suoi lati positivi che alterano la vita, ma liberarla al massimo significherebbe rischiare tutti i suoi svantaggi potenzialmente catastrofici. In altre parole, il profilo rischio-rendimento per l'AI è asimmetrico. Data la radicale incertezza sulla portata e sull'irreversibilità di alcuni dei potenziali danni dell'IA, la governance dell'IA deve mirare a prevenire questi rischi prima che si materializzino piuttosto che mitigarli dopo il fatto.

Ciò è particolarmente importante perché l'intelligenza artificiale potrebbe indebolire la democrazia in alcuni paesi e rendere loro più difficile l'adozione di regolamenti. Inoltre, l'onere di dimostrare che un sistema di intelligenza artificiale è sicuro al di sopra di una soglia ragionevole dovrebbe ricadere sullo sviluppatore e sul proprietario; non dovrebbe spettare esclusivamente ai governi affrontare i problemi una volta che si presentano.

La governance dell'IA deve anche essere agile in modo da potersi adattare e correggere la rotta man mano che l'IA si evolve e migliora se stessa. Le istituzioni pubbliche spesso si calcificano al punto da non essere in grado di adattarsi al cambiamento. E nel caso dell'intelligenza artificiale, la velocità assoluta del progresso tecnologico aumenterà rapidamente

sopraffare la capacità delle strutture di governance esistenti di recuperare e tenere il passo. Ciò non significa che la governance dell'IA dovrebbe adottare l'etica del "muoviti velocemente e rompi le cose" della Silicon Valley, ma dovrebbe rispecchiare più da vicino la natura della tecnologia che cerca di contenere.

Oltre ad essere precauzionale e agile, la governance dell'IA deve essere inclusiva, invitando la partecipazione di tutti gli attori necessari per regolamentare l'IA nella pratica. Ciò significa che la governance dell'IA non può essere esclusivamente centrata sullo stato, poiché i governi non comprendono né controllano l'IA. Le aziende tecnologiche private possono mancare di sovranità nel senso tradizionale del termine, ma esercitano un potere e un'agenzia reali, persino sovrani, negli spazi digitali che hanno creato e governano efficacemente. A questi attori non statali non dovrebbero essere concessi gli stessi diritti e privilegi degli Stati, che sono internazionalmente riconosciuti come agenti per conto dei propri cittadini. Ma dovrebbero essere parti di vertici internazionali e firmatari di qualsiasi accordo sull'IA.

Un tale ampliamento della governance è necessario perché qualsiasi struttura normativa che escluda i veri agenti del potere dell'IA è destinata a fallire. Nelle precedenti ondate di regolamentazione tecnologica, alle aziende è stato spesso concesso così tanto margine di manovra che hanno oltrepassato il limite, portando i responsabili politici e le autorità di regolamentazione a reagire duramente ai loro eccessi. Ma questa dinamica non ha giovato né alle aziende tecnologiche né al pubblico. Invitare gli sviluppatori di intelligenza artificiale a partecipare fin dall'inizio al processo decisionale aiuterebbe a stabilire una cultura più collaborativa della governance dell'IA, riducendo la necessità di frenare queste società dopo il fatto con una regolamentazione costosa e contraddittoria.

> L'intelligenza artificiale è un problema dei beni comuni globali, non solo appannaggio di due superpotenze.

Le aziende tecnologiche non dovrebbero sempre avere voce in capitolo; è meglio lasciare alcuni aspetti della governance dell'IA ai governi, e va da sé che gli stati dovrebbero sempre mantenere il potere di veto finale sulle decisioni politiche. I governi devono anche proteggersi dalla cattura normativa per garantire che le aziende tecnologiche non utilizzino la loro influenza all'interno dei sistemi politici per promuovere i propri interessi a scapito del bene pubblico. Ma un modello di governance inclusivo e multistakeholder assicurerebbe che gli attori che determineranno il destino dell'IA siano coinvolti e vincolati dai processi normativi. Oltre ai governi (soprattutto, ma non solo, Cina e Stati Uniti) e le aziende tecnologiche (in particolare, ma non solo, i giocatori di Big Tech), scienziati, esperti di etica, sindacati, organizzazioni della società civile e altre voci con conoscenza, potere o interesse nei risultati dell'IA dovrebbero avere un posto al tavolo. La Partnership on AI, un gruppo senza scopo di lucro che riunisce una serie di grandi aziende tecnologiche, istituti di ricerca, enti di beneficenza e organizzazioni della società civile per promuovere un uso responsabile dell'IA, **è** un buon esempio del tipo di forum misto e inclusivo necessario.

Anche la governance dell'IA deve essere il più impermeabile possibile. A differenza della mitigazione del cambiamento climatico, in cui il successo sarà determinato dalla somma di tutti gli sforzi individuali, la sicurezza dell'IA è determinata dal minimo comune denominatore: un singolo algoritmo di breakout potrebbe causare danni incalcolabili. Poiché la governance globale dell'IA è valida solo quanto il paese, l'azienda o la tecnologia peggio governati, deve essere impermeabile ovunque, con l'accesso

abbastanza facile da costringere alla partecipazione e all'uscita abbastanza costosa da scoraggiare la non conformità. Una singola scappatoia, un anello debole o un disertore canaglia aprirà la porta a fughe di notizie diffuse, malintenzionati o una corsa al ribasso della regolamentazione.

Oltre a coprire l'intero globo, la governance dell'IA deve coprire l'intera catena di fornitura, dalla produzione all'hardware, dal software ai servizi e dai fornitori agli utenti. Ciò significa regolamentazione e supervisione tecnicoprudenziale lungo ogni nodo della catena del valore dell'IA, dalla produzione di chip AI alla raccolta dei dati, dalla formazione del modello all'uso finale e attraverso l'intero stack di tecnologie utilizzate in una determinata applicazione. Tale impermeabilità assicurerà che non vi siano zone grigie normative da sfruttare.

Infine, la governance dell'IA dovrà essere mirata, piuttosto che adatta a tutti. Poiché l'intelligenza artificiale è una tecnologia generica, pone minacce multidimensionali. Un unico strumento di governance non è sufficiente per affrontare le varie fonti di rischio di IA. In pratica, determinare quali strumenti sono appropriati per affrontare quali rischi richiederà lo sviluppo di una tassonomia vivente e respirante di tutti i possibili effetti che l'IA potrebbe avere e come ognuno può essere governato al meglio.
Ad esempio, l'intelligenza artificiale sarà evolutiva in alcune applicazioni, esacerbando problemi attuali come le violazioni della privacy, e rivoluzionaria in altre, creando danni completamente nuovi.
A volte, il posto migliore per intervenire sarà dove vengono raccolti i dati. Altre volte, sarà il punto in cui vengono venduti chip avanzati, assicurandosi che non cadano nelle mani sbagliate. Affrontare la disinformazione e la disinformazione richiederà strumenti diversi rispetto all'affrontare i rischi dell'AGI e di altre tecnologie incerte con potenziali ramificazioni esistenziali. Un leggero tocco normativo e una guida volontaria funzioneranno in alcuni casi; in altri, i governi dovranno imporre rigorosamente la conformità.

Tutto ciò richiede una profonda comprensione e una conoscenza aggiornata delle tecnologie in questione. Le autorità di regolamentazione e altre autorità avranno bisogno di supervisione e accesso ai principali modelli di intelligenza artificiale. In effetti, avranno bisogno di un sistema di controllo in grado non solo di monitorare le capacità a distanza, ma anche di accedere direttamente alle tecnologie di base, che a loro volta richiederanno il talento giusto. Solo tali misure possono garantire che le nuove applicazioni di intelligenza artificiale siano valutate in modo proattivo, sia per i rischi evidenti che per le conseguenze potenzialmente distruttive di secondo e terzo ordine. Una governance mirata, in altre parole, deve essere una governance ben informata.

L'IMPERATIVO TECNOPRUDENZIALE

Sulla base di questi principi dovrebbero esserci almeno tre regimi di governance dell'IA, ciascuno con mandati, leve e partecipanti diversi. Tutti dovranno essere nuovi nel design, ma ognuno potrebbe trarre ispirazione dagli accordi esistenti per affrontare altre sfide globali, vale a dire il cambiamento climatico, la proliferazione degli armamenti e la stabilità finanziaria.

Il primo regime si concentrerebbe sull'accertamento dei fatti e assumerebbe la forma di un organismo scientifico globale per consigliare obiettivamente i governi e gli organismi internazionali su questioni fondamentali come cos'è l'IA e quali tipi di sfide politiche pone. Se nessuno può concordare sulla definizione di IA o sulla possibile portata dei suoi danni, sarà impossibile elaborare politiche efficaci. Qui, il cambiamento climatico è istruttivo. Per creare una base di conoscenza condivisa per i negoziati sul clima, le Nazioni Unite hanno istituito il Gruppo intergovernativo su

Cambiamento climatico e gli ha dato un semplice mandato: fornire ai responsabili politici "valutazioni regolari delle basi scientifiche del cambiamento climatico, dei suoi impatti e dei rischi futuri e delle opzioni per l'adattamento e la mitigazione". L'intelligenza artificiale ha bisogno di un organismo simile per valutare regolarmente lo stato dell'IA, valutare in modo imparziale i suoi rischi e i potenziali impatti, prevedere scenari e prendere in considerazione soluzioni politiche tecniche per proteggere l'interesse pubblico globale. Come l'IPCC, questo organismo avrebbe un imprimatur globale e un'indipendenza scientifica (e geopolitica). E i suoi rapporti potrebbero informare i negoziati multilaterali e multilaterali sull'IA, proprio come i rapporti dell'IPCC informano i negoziati sul clima delle Nazioni Unite.

Il mondo ha anche bisogno di un modo per gestire le tensioni tra le principali potenze di intelligenza artificiale e prevenire la proliferazione di pericolosi sistemi di intelligenza artificiale avanzati. La relazione internazionale più importante nell'IA è quella tra Stati Uniti e Cina.
La cooperazione tra i due rivali è difficile da ottenere nelle migliori circostanze.
Ma nel contesto di una maggiore concorrenza geopolitica, una corsa incontrollata all'IA potrebbe condannare ogni speranza di creare un consenso internazionale sulla governance dell'IA. Un'area in cui Washington e Pechino potrebbero trovare vantaggioso lavorare insieme è il rallentamento della proliferazione di potenti sistemi che potrebbero mettere in pericolo l'autorità degli stati-nazione. All'estremo, la minaccia di AGI incontrollate e autoreplicanti, se dovessero essere inventate negli anni a venire, fornirebbe forti incentivi per coordinarsi sulla sicurezza e il contenimento.

Su tutti questi fronti, Washington e Pechino dovrebbero mirare a creare aree di comunanza e persino guardrail proposti e sorvegliati da una terza parte. Qui, gli approcci di monitoraggio e verifica spesso presenti nei regimi di controllo degli armamenti potrebbero essere applicati ai più importanti input di intelligenza artificiale, in particolare quelli relativi all'hardware informatico, inclusi i semiconduttori avanzati e i data center. La regolamentazione dei punti di strozzatura chiave ha contribuito a contenere una pericolosa corsa agli armamenti durante la Guerra Fredda e potrebbe aiutare a contenere una corsa all'IA potenzialmente ancora più pericolosa ora.

> Pochi potenti elettori favoriscono il contenimento dell'IA.

Ma poiché gran parte dell'IA è già decentralizzata, è un problema dei beni comuni globali piuttosto che appannaggio di due superpotenze. La natura devoluta dello sviluppo dell'IA e le caratteristiche fondamentali della tecnologia, come la proliferazione open source, aumentano la probabilità che venga utilizzata come arma da criminali informatici, attori sponsorizzati dallo stato e lupi solitari. Ecco perché il mondo ha bisogno di un terzo regime di governance dell'IA in grado di reagire quando si verificano interruzioni pericolose.
Per i modelli, i responsabili politici potrebbero guardare all'approccio utilizzato dalle autorità finanziarie per mantenere la stabilità finanziaria globale. Il Consiglio per la stabilità finanziaria, composto da banchieri centrali, ministeri delle finanze e autorità di vigilanza e regolamentazione di tutto il mondo, lavora per prevenire l'instabilità finanziaria globale valutando le vulnerabilità sistemiche e coordinando le azioni necessarie per affrontarle tra le autorità nazionali e internazionali. Un organismo altrettanto tecnocratico per il rischio di intelligenza artificiale, chiamiamolo Geotechnology Stability Board, potrebbe lavorare per mantenere la stabilità geopolitica in mezzo a rapidi cambiamenti guidati dall'intelligenza artificiale. Supportato dalle autorità nazionali di regolamentazione e dagli organismi internazionali di definizione degli standard, riunirebbe competenze e risorse per prevenire o rispondere alle crisi legate all'IA, riducendo il rischio di contagio. Ma coinvolgerebbe anche

direttamente con il settore privato, riconoscendo che i principali attori tecnologici multinazionali svolgono un ruolo fondamentale nel mantenimento della stabilità geopolitica, proprio come fanno le banche di rilevanza sistemica nel mantenere la stabilità finanziaria.

Un tale organismo, con autorità radicata nel sostegno del governo, sarebbe ben posizionato per impedire agli attori tecnologici globali di impegnarsi in arbitraggio normativo o nascondersi dietro i domicili aziendali. Riconoscere che alcune aziende tecnologiche sono di importanza sistemica non significa soffocare le start-up o gli innovatori emergenti. Al contrario, la creazione di un'unica linea diretta da un organismo di governance globale a questi colossi tecnologici migliorerebbe l'efficacia dell'applicazione delle normative e della gestione delle crisi, a vantaggio dell'intero ecosistema.

Un regime progettato per mantenere la stabilità geotecnologica colmerebbe anche un vuoto pericoloso nell'attuale panorama normativo: la responsabilità di governare l'IA open source. Sarà necessario un certo livello di censura online. Se qualcuno carica un modello estremamente pericoloso, questo organismo deve avere la chiara autorità - e capacità - per rimuoverlo o indirizzare le autorità nazionali a farlo. Questa è un'altra area di potenziale cooperazione bilaterale. La Cina e gli Stati Uniti dovrebbero voler lavorare insieme per incorporare vincoli di sicurezza nel software open source, ad esempio limitando la misura in cui i modelli possono istruire gli utenti su come sviluppare armi chimiche o biologiche o creare agenti patogeni pandemici. Inoltre, potrebbe esserci spazio per Pechino e Washington per cooperare sugli sforzi globali di antiproliferazione, anche attraverso l'uso di strumenti informatici interventisti.

Ciascuno di questi regimi dovrebbe operare universalmente, godendo del consenso di tutti i principali attori dell'IA. I regimi dovrebbero essere abbastanza specializzati da far fronte ai sistemi di IA reali e abbastanza dinamici da continuare ad aggiornare la loro conoscenza dell'IA man mano che si evolve. Lavorando insieme, queste istituzioni potrebbero compiere un passo decisivo verso la gestione tecnoprudenziale del mondo emergente dell'IA. Ma non sono affatto le uniche istituzioni di cui avremo bisogno. Altri meccanismi normativi, come gli standard di trasparenza "conosci il tuo cliente", i requisiti di licenza, i protocolli di test di sicurezza e i processi di registrazione e approvazione dei prodotti, dovranno essere applicati all'IA nei prossimi anni.
La chiave di tutte queste idee sarà creare istituzioni di governance flessibili e sfaccettate che non siano vincolate dalla tradizione o dalla mancanza di immaginazione: dopotutto, i tecnologi non saranno vincolati da queste cose.

PROMUOVERE IL MEGLIO, PREVENIRE IL PEGGIO

Nessuna di queste soluzioni sarà facile da implementare. Nonostante tutto il brusio e le chiacchiere provenienti dai leader mondiali sulla necessità di regolamentare l'IA, manca ancora la volontà politica di farlo. In questo momento, pochi potenti elettori favoriscono il contenimento dell'IA e tutti gli incentivi puntano verso una continua inazione. Ma progettato bene, un regime di governance dell'IA del tipo descritto qui potrebbe soddisfare tutte le parti interessate, sancire principi e strutture che promuovono il meglio dell'IA prevenendo il peggio. L'alternativa - IA non contenuta - non solo porrebbe rischi inaccettabili per la stabilità globale; sarebbe anche dannoso per gli affari e sarebbe contrario all'interesse nazionale di ogni paese.

Un forte regime di governance dell'IA mitigherebbe sia i rischi sociali posti dall'IA sia allenterebbe le tensioni tra Cina e Stati Uniti riducendo la misura in cui l'IA è un'arena - e uno strumento - della competizione geopolitica. E un tale regime otterrebbe qualcosa di ancora più profondo e duraturo: stabilirebbe un modello su come affrontare altre tecnologie emergenti dirompenti. L'intelligenza artificiale può essere un catalizzatore unico per il cambiamento, ma non è affatto l'ultima tecnologia dirompente che l'umanità dovrà affrontare. Anche l'informatica quantistica, la biotecnologia, la nanotecnologia e la robotica hanno il potenziale per rimodellare radicalmente il mondo. Governare con successo l'IA aiuterà il mondo a governare con successo anche quelle tecnologie.

Il ventunesimo secolo presenterà poche sfide così scoraggianti o opportunità così promettenti come quelle presentate dall'IA. Nel secolo scorso, i politici hanno iniziato a costruire un'architettura di governance globale che, speravano, sarebbe stata all'altezza dei compiti dell'epoca.
Ora devono costruire una nuova architettura di governance per contenere e imbrigliare la forza più formidabile e potenzialmente determinante di questa era. L'anno 2035 è proprio dietro l'angolo. Non c'è tempo da perdere.

IAN BREMMER è presidente e fondatore di Eurasia Group e GZERO Media. È l'autore di ***The Power of Crisis: How Three Threats — and Our Response — Will Change the World.***

MUSTAFA SULEYMAN è CEO e co-fondatore di Inflection AI. Un co-fondatore di DeepMind, è l'autore di ***The Coming Wave: Technology, Power, and the Twenty-first Century's Greatest Dilemma.***