

Il panorama globale dell'IA nel 2026 è segnato da una frattura filosofica e industriale ormai insanabile tra le due superpotenze tecnologiche. Da un lato, gli Stati Uniti hanno consolidato un modello di business definibile "a castello": un ecosistema chiuso, centralizzato e protetto da altissime barriere d'ingresso economiche, legali e computazionali. Giganti come OpenAI, Microsoft, Amazon, Anthropic e Google vendono l'accesso alla "conoscenza" come un servizio cloud (SaaS), mantenendo un controllo totale sui "pesi" dei modelli (i pesi o weights sono i valori numerici – miliardi di parametri – che un modello ha appreso durante la fase di addestramento; sono, in pratica, la sua memoria e la sua capacità di ragionamento), sugli algoritmi e sui preziosi dati di addestramento. Questo approccio closed-source crea una dipendenza tecnologica globale dai server della Silicon Valley, accentrando il valore economico e decisionale nel cloud, in una sorta di neofeudalesimo digitale dove l'utente è un semplice abbonato e produce dati senza essere retribuito, alla stregua di un servo della gleba (le persone vengono smaterializzate in dati e attraverso la banca dati sono trasformate in segmenti di mercato, in campioni statistici, in archivi diversi interoperabili). I dati di cui i sistemi di IA hanno bisogno spesso non vengono acquisiti con mezzi del tutto leciti, per non parlare di equi. Le aziende che si occupano di IA si appropriano della conoscenza umana, automatizzano i processi lavorativi, li brevettano e poi tentano di rivenderci tutto questo. Il governo statunitense asseconda gli oligarchi della tecnologia per timore di perdere la corsa al vantaggio tecnologico, nonostante la diffusa ostilità dell'opinione pubblica verso i sistemi di IA¹.

Dall'altro lato, la Cina ha adottato una strategia "a sciame", agile e distribuita. Consapevole delle limitazioni imposte dalle sanzioni occidentali sull'hardware avanzato, Pechino ha spinto i propri campioni nazionali — Alibaba, DeepSeek, Huawei — a rilasciare modelli open-weight estremamente efficienti. I modelli open-weight rappresentano una via di mezzo tra i modelli totalmente "chiusi" (proprietary) e quelli puramente open source. Questa scelta non è filantropica, ma tattica: distribuire il "motore" dell'IA permette una diffusione capillare su dispositivi locali (Edge AI), garantendo sovranità dei dati e indipendenza dalle infrastrutture straniere.

Questa divergenza comporta conseguenze geopolitiche profonde. Gli Stati Uniti guidano nettamente la corsa globale alla potenza di calcolo per l'intelligenza artificiale e il calcolo ad alte prestazioni (HPC), mentre la Cina si posiziona al secondo posto nel mondo. Insieme, le due superpotenze controllano circa il 90% di tutta la potenza computazionale globale. Il modello USA punta al controllo e all'alta redditività, limitando l'accesso ai modelli più avanzati per motivi strategici. Il modello cinese mira a una rapida adozione, creando un ecosistema democratico che aggira la necessità di GPU top di gamma (Graphics Processing Unit, unità di elaborazione grafica). Tale scenario crea una balkanizzazione tecnologica, in cui il mondo si divide tra un'IA chiusa e iper-centralizzata e un'IA aperta e distribuita. Il controllo dell'infrastruttura definisce nuove alleanze, rendendo l'IA non solo un settore economico, ma il perno della sicurezza nazionale e della sovranità digitale per le due, ormai contrapposte, superpotenze.

La strada cinese verso l'IA

Innanzitutto, la strategia cinese comporta conseguenze che ridisegnano il concetto stesso di sicurezza informatica. L'uso di modelli in "locale" permette ad aziende, istituzioni e persone di elaborare dati sensibili senza doverli inviare a server esterni o stranieri. Le informazioni non lasciano mai il dispositivo dell'utente, eliminando il rischio di fughe di dati durante il transito verso il cloud. I sistemi funzionano anche in assenza di una connessione internet stabile (offline) o in caso di sanzioni che blocchino l'accesso ai servizi cloud esteri. Inoltre, la democratizzazione dell'IA tramite i modelli open-weight cinesi, come quelli della famiglia Qwen (Alibaba) o DeepSeek (con le serie V3, R1 e il recente V4-Pro), stanno trasformando l'IA da servizio di lusso a utility a basso costo. Gli sviluppatori indipendenti e le startup (piccole software house) possono scaricare le app del modello gratuitamente, evitando i costi ricorrenti delle API (Application Programming Interface) tipici dei fornitori USA. Essendo progettati per girare su dispositivi locali, questi modelli richiedono meno potenza di calcolo rispetto ai giganti cloud, rendendo l'IA accessibile anche su hardware meno costoso. La riduzione di prezzo, stimata tra l'80% e il 95%, consente agli sviluppatori indipendenti di implementare logiche complesse come cicli di validazione multipli o "chain of thought" profonde (ragionamento sequenziale), che prima avrebbero prosciugato il budget di una startup in poche ore.

L'ecosistema open-weight cinese ha fornito il carburante computazionale per la nascita di framework applicativi rivoluzionari nati dal basso. Un esempio lampante è OpenClaw, un framework per agenti autonomi locali integrabili su app di messaggistica (Telegram, Discord, WhatsApp). Nato come prototipo indipendente, OpenClaw ha sfruttato l'efficienza di calcolo e l'ampia finestra di contesto (fino a 1 milione di token nei modelli cinesi più recenti) per scalare a tempo di record, superando librerie storiche e registrando oltre 280.000 stelle su GitHub. Questo dimostra che l'innovazione software nel 2026 non è più solo una prerogativa dei laboratori della Silicon Valley, ma appartiene anche alle community open-source indipendenti.

Rilasciando modelli aperti, la Cina sta costruendo un'infrastruttura di influenza globale che minaccia il dominio statunitense. Nel 2025, i modelli Qwen hanno superato i modelli Llama di Meta per numero di download su piattaforme come Hugging Face². Un tipico flusso di lavoro di uno sviluppatore indipendente permette di gestire la sintesi dei dati e la scrittura preliminare del codice in locale su istanze Qwen a costo zero, inviando ai modelli chiusi statunitensi solo la rifinitura finale. Ciò riduce il costo operativo medio di sviluppo di un'applicazione da \$10 a meno di \$ 2 per utente. Quando migliaia di programmatori creano applicazioni basate su modelli cinesi, questi diventano lo standard "de facto", rendendo difficile per i competitor chiusi statunitensi riconquistare quote di mercato.

Infine, questa strategia funge da volano per l'industria dei semiconduttori interna cinese, costretta a innovare a causa delle restrizioni USA. Aziende come Huawei sviluppano chip specificamente ottimizzati per far girare i modelli open-weight cinesi, creando un ecosistema integrato e autosufficiente.

In sintesi, se gli USA vendono il “servizio” di IA, la Cina sta distribuendo il “motore” per costruirla, puntando a diventare l’ossatura invisibile, ma onnipresente, della prossima rivoluzione tecnologica locale e industriale. Se gli USA sono un “petro-Stato” che vende gas liquefatto per produrre l’elettricità in grandi centrali, la Cina vende i progetti per produrre energia elettrica rinnovabile distribuita in casa ed è diventato un “elettro-Stato”. Inoltre, la Cina si concentra su come l’IA interagisce con l’economia nel suo complesso, non come tecnologia a sé stante, ma come moltiplicatore di forza integrato (IA applicata) in fabbriche, reti logistiche, reti elettriche, reti sanitarie e in tutte le altre principali arterie del commercio e della produzione, piuttosto che operare esclusivamente nei chatbot di consumo.

La rete e le funzioni dei data center cinesi

Nonostante l’enfasi sui modelli “locali”, la Cina non ha rinunciato alla potenza di calcolo centralizzata. Al contrario, sta costruendo una delle reti di data center più grandi al mondo (la rete dei data center rappresenta il sistema nervoso centrale dei prodotti di IA). L’enfasi sui modelli open-weight non sostituisce la necessità di infrastrutture centralizzate, ma ne cambia l’uso strategico. I grandi data center cinesi restano indispensabili per due motivi principali:

1. addestramento (training): anche se un modello cinese come Qwen può “girare” su un laptop, per essere creato (addestrato) ha comunque richiesto migliaia di potenti chip (GPU) attivi per mesi in un immenso data center;
2. infrastruttura di Stato: la Cina considera la potenza di calcolo un’“utility” pubblica, come l’elettricità o l’acqua.

A giugno 2026, fonti internazionali come Bloomberg e Reuters hanno riportato l’intenzione di Pechino di investire \$ 295 miliardi nei prossimi cinque anni per creare una rete nazionale di data center IA. L’obiettivo è ridurre la dipendenza dai chip USA e puntare ad un hardware nazionale (come i chip Huawei Ascend) per almeno l’80% delle infrastrutture. Inoltre, la Cina ha un vantaggio energetico: aggiunge capacità elettrica rinnovabile (fotovoltaico ed eolico) ogni anno a ritmi che gli USA non riescono a pareggiare, risolvendo il problema dei consumi enormi dell’IA. Mentre negli USA i data center sono “fabbriche di servizi” per i giganti del tech, in Cina sono concepiti come “motori” energetici e computazionali distribuiti per alimentare l’intera economia nazionale, inclusi i dispositivi locali. Nonostante sia la Cina sia gli USA stiano costruendo massicciamente reti di data center, l’approccio è radicalmente diverso:

Caratteristica	Modello statunitense	Modello cinese
----------------	----------------------	----------------

Proprietà	Prevalentemente privata (Microsoft, Google, Meta).	Prevalentemente statale o a guida pubblica (China Mobile, China Telecom).
Obiettivo	Vendere servizi cloud "chiusi" (SaaS).	Creare un'infrastruttura nazionale interconnessa per l'industria.
Localizzazione	Vicino ai nodi di rete e ai mercati.	Spostata a Ovest (zone rurali, desertiche e montane) per sfruttare energia solare, eolica e idroelettrica a basso costo.
Piano Strategico	Basato sulla competizione di mercato.	Progetto "East Data, West Computing": calcolare a Ovest i dati prodotti a Est.

Il progetto cinese "East Data, West Computing"

Il progetto "East Data, West Computing" (EDWC) è il cuore pulsante della strategia cinese per sviluppare l'IA, trasformando le aree desertiche e montuose occidentali (come Ningxia, Gansu, Guizhou e Mongolia Interna) in una gigantesca "batteria computazionale" per l'industria delle aree orientali (come Pechino, Shanghai e Shenzhen). Questa iniziativa sta plasmando il futuro delle rinnovabili e i dati non urgenti, l'archiviazione a lungo termine e l'addestramento dei modelli di IA vengono così delocalizzati a ovest, ottimizzando le risorse del Paese.

A differenza dei modelli occidentali, la Cina sta eliminando gli intermediari della rete elettrica per alimentare i supercomputer. A maggio 2026, è entrato in funzione a Ningxia il primo progetto su larga scala da 500 MW (solare ed eolico³) collegato direttamente a un data center tramite linee di trasmissione dedicate. Si tratta del primo progetto su larga scala in Cina basato sul modello di "coordinamento computazione-elettricità". Entro la fine dell'anno, la quota di energia verde nei nuovi hub di calcolo nazionali dovrà superare l'80%. Il progetto sposta il carico di lavoro dove l'energia costa meno e abbonda. I data center fungono da consumatori stabili per l'energia solare ed eolica prodotta in eccesso nella Cina occidentale che andrebbe altrimenti sprecata. In regioni come la Mongolia Interna e lo Xizang (Tibet) si usano le basse temperature naturali per il raffreddamento, abbattendo drasticamente il consumo elettrico.

Per le città costiere dove lo spazio manca, la Cina ha lanciato soluzioni di frontiera. Operativo da maggio 2026, è il primo data center sottomarino al mondo - Shanghai Lingong - alimentato esclusivamente da un parco eolico offshore. Sfrutta l'acqua marina per il raffreddamento, riducendo le emissioni di carbonio di 12.000 tonnellate all'anno. Questa sinergia permette alla Cina di offrire una potenza di calcolo più economica e sostenibile rispetto ai competitor statunitensi, rendendo l'IA non solo una sfida di algoritmi,

ma una sfida di infrastruttura energetica.

Numeri della transizione cinese (Dati 2025-2026)

Obiettivo	Stato nel 2026
Capacità solare totale	Superati i 1.000 GW (record storico).
Quota rinnovabili	Oltre il 60% della capacità installata nazionale.
Efficienza (PUE - Power Usage Effectiveness)	Target di 1.25 o inferiore per i grandi data center.
Nuova potenza (2026)	Previsti altri 300 GW tra eolico e solare.

Elementi di criticità della strategia cinese per lo sviluppo dell'IA

La strategia cinese, sebbene competitiva sul fronte dei costi e della diffusione, presenta criticità speculari a quelle statunitensi, legate principalmente alla dipendenza tecnologica residua, ai vincoli politici e a una crisi demografica che minaccia il capitale umano. Nonostante gli sforzi verso l'autosufficienza, la Cina soffre ancora per le restrizioni sulle tecnologie di frontiera. Le sanzioni statunitensi continuano a limitare l'importazione di GPU di fascia altissima (come le serie H100 e Blackwell di Nvidia). Sebbene esistano alternative domestiche come i chip Huawei Ascend, il divario prestazionale rispetto ai leader USA rimane un ostacolo per l'addestramento dei modelli più grandi (Frontier Models). Inoltre, la Cina dipende ancora da tecnologie estere (come ASML) per produrre chip sotto i 7 nanometri (nm), rendendo la sua catena di approvvigionamento hardware vulnerabile a ulteriori inasprimenti normativi statunitensi.

La diffusione dell'IA in Cina non è libera, ma strettamente regolamentata per allinearsi alla sicurezza nazionale e alla stabilità sociale. I modelli cinesi devono superare rigidi controlli governativi per garantire che le risposte siano conformi ai valori socialisti. Questo "filtro" ideologico può limitare la creatività e l'imprevedibilità necessaria per scoperte scientifiche radicali⁴. La protezione ossessiva dei dati come patrimonio strategico nazionale può rallentare la collaborazione internazionale, isolando l'ecosistema cinese in una "bolla" tecnologica.

L'IA richiede un flusso costante di talenti di alto livello, un'area dove la Cina affronta sfide strutturali. La crisi demografica cinese è il nodo più critico per il 15° Piano Quinquennale. Una popolazione in calo riduce il numero di giovani ingegneri e ricercatori disponibili. Nonostante la produzione di milioni di laureati STEM, i migliori talenti cinesi continuano spesso a preferire carriere o dottorati negli Stati Uniti, attratti da un ambiente di ricerca

meno vincolato e da stipendi più competitivi.

La strategia basata su modelli open-weight a basso costo potrebbe trasformarsi in una trappola economica. Abbattendo i prezzi dei modelli (come hanno fatto DeepSeek e Qwen), le aziende cinesi rischiano di non generare margini sufficienti per finanziare la prossima generazione di ricerca fondamentale. Secondo gli analisti, nel settore affollato dell'IA cinese si sta scatenando una guerra dei prezzi, con le aziende che riducono le tariffe o lanciano promozioni in un momento cruciale in cui la diminuzione dei costi e la convergenza delle capacità dei modelli stanno intensificando la pressione competitiva. ByteDance, la società madre di TikTok, e Tencent Holdings, il colosso dei videogiochi con sede a Shenzhen, sono le ultime aziende ad aver lanciato offensive di prezzo nel settore dell'IA.

Mentre la Cina domina nell'adozione di massa, i modelli chiusi statunitensi (come quelli di OpenAI o Anthropic) conservano un vantaggio nelle capacità di ragionamento puro e nei flussi di lavoro più complessi.

Sintesi delle criticità cinesi

Area	Criticità principale	Impatto strategico
Tecnologia	Dipendenza da macchinari esteri per chip <7nm.	Rischio di rimanere indietro sulla frontiera della potenza di calcolo.
Politica	Necessità di allineamento ideologico degli output.	Possibile limitazione della creatività del modello.
Sociale	Calo demografico e invecchiamento della popolazione.	Scarsità di capitale umano specializzato a lungo termine.
Economica	Guerra dei prezzi sui modelli open source.	Difficoltà nel sostenere costi di ricerca e sviluppo miliardari.

La strategia statunitense per l'IA

La strategia statunitense per l'IA si è evoluta in una dottrina che mescola libero mercato aggressivo, sicurezza nazionale e protezione del vantaggio tecnologico. A differenza del modello a guida pubblica cinese, quella americana è una strategia a guida privata e supporto pubblico. La visione dell'IA proposta dalle grandi aziende tecnologiche statunitensi e sostenuta dall'amministrazione Trump è che l'IA è un fine in sé. L'obiettivo è implementarla il più rapidamente possibile, a prescindere dalle conseguenze per la società e il pianeta. Si contrapponga a ciò la recente enciclica di Papa Leone XIV sull'IA, la quale afferma che laddove "lo sviluppo tecnologico avanza senza un corrispondente progresso etico e sociale, il risultato può essere un aumento dei mezzi senza una crescita

dell'umanità". Ha denunciato la "cultura del potere" che alimenta l'ascesa dell'IA e ha chiesto il "disarmo" dell'IA (si veda il nostro articolo qui). I cambiamenti apportati dall'IA interessano l'intera società e sono estremamente dirompenti – possono mettere a repentaglio il lavoro equo, perpetuare stereotipi razziali e di genere ed esacerbare le crisi ambientali. Inoltre, stanno avvenendo anche molto rapidamente. Le persone non hanno il tempo di sviluppare resilienza o di immunizzarsi da questi cambiamenti. I punti di vista tradizionali dell'industria dell'IA (da Sam Altman a Alex Karp, Peter Thiel ed Elon Musk) – ovvero che questa tecnologia rivoluzionerà il mondo, se non addirittura lo distruggerà (con scenari sia da incubo in cui l'IA autoreplicante annienterà l'umanità sia utopici in cui le macchine senzienti ci condurranno a un paradiso post-umano) – si prestano anche a una narrazione radicale distopica e apocalittica secondo cui l'IA rappresenta una minaccia esistenziale e deve essere fermata a tutti i costi⁵.

Gli Stati Uniti si rifiutano di considerare in modo critico i potenziali benefici, rischi e limiti tecnici dell'IA, dando invece per scontato il suo impatto positivo senza fornire prove concrete⁶. Si sostiene che l'IA sia uno sviluppo completamente nuovo, destinato a sostituire del tutto il lavoro umano, in quanto in grado di superare l'intelligenza umana. Le prove a sostegno di questa tesi sono tuttavia dubbie. Gran parte dell'IA consiste semplicemente nell'elaborazione rapida di conoscenze umane preesistenti e non può sostituire la natura immaginativa dell'intelligenza umana. Inoltre, ci vorrà del tempo, persino decenni, prima che l'IA diffonda i suoi effetti positivi sulla produttività nelle economie.

Gli USA puntano a mantenere la leadership attraverso il controllo dei modelli di frontiera delle piattaforme di distribuzione. Aziende come OpenAI, Google e Anthropic sviluppano modelli "chiusi". L'obiettivo è creare un ecosistema in cui l'IA è un servizio venduto in abbonamento, rendendo i clienti globali dipendenti dalle infrastrutture cloud americane (Azure, AWS, Google Cloud). L'obiettivo è quello di monetizzare il vantaggio competitivo, sfruttando il primato temporale per imporre standard di mercato prima che i competitor possano reagire.

La strategia hardware è diventata una questione di sicurezza nazionale prioritaria. Gli USA utilizzano il controllo sulle esportazioni (tramite il Dipartimento del Commercio) per negare alla Cina e ai competitor l'accesso ai chip più avanzati (Nvidia H100/Blackwell) e ai macchinari per produrli (ASML). L'obiettivo è mantenere un distacco tecnologico di almeno due generazioni, cercando di strangolare e invertire il progresso tecnologico ed economico della Cina, considerata la principale minaccia all'egemonia statunitense. È in esecuzione un piano di sussidi massicci (finanziato dal CHIPS Act, approvato durante la presidenza Biden) per riportare la produzione di semiconduttori sul suolo statunitense (reshoring), riducendo la vulnerabilità legata a Taiwan e alla catena di approvvigionamento asiatica.

Al tempo stesso, il Pentagono ha accelerato l'integrazione dell'IA nelle forze armate per contrastare la superiorità numerica cinese. Punta sullo sviluppo di sciame di droni economici e altre armi autonome capaci di operare in ambienti contestati. Nonché sull'uso dell'IA per l'analisi dei dati satellitari e la gestione del comando e controllo in tempo reale. Sono stati raggiunti accordi con otto aziende leader nel settore dell'IA: SpaceX, OpenAI, Google, Nvidia, Reflection AI, Microsoft, Oracle e Amazon Web Services – AWS. "Questi accordi

accelerano la trasformazione verso la creazione di una forza armata statunitense incentrata sull'IA e rafforzeranno la capacità dei nostri soldati di mantenere la superiorità decisionale in tutti gli ambiti della guerra", ha dichiarato il Pentagono in un comunicato. Secondo il Pentagono, le aziende hanno acconsentito all'impiego della loro tecnologia da parte dell'esercito statunitense per "qualsiasi uso lecito". Solo Anthropic si è rifiutata di includere la clausola dell'uso lecito nel suo contratto con il Dipartimento della Difesa, aprendo una controversia.

Il punto più critico della strategia statunitense attuale è la corsa all'approvvigionamento energetico. Per alimentare i nuovi maxi-data center da 5-10 GW, i giganti tech (Microsoft, Amazon) stanno investendo direttamente in nuove centrali elettriche a gas e reattori nucleari di nuova generazione (SMR - Small Modular Reactors) e nella riapertura di centrali dismesse (come Three Mile Island in Pennsylvania). Forti sono le pressioni politiche per accelerare le licenze per la costruzione di queste infrastrutture critiche, considerate essenziali per non perdere la gara contro la Cina.

A differenza dell'AI Act europeo (rigido e vincolante), la strategia USA si basa su accordi non vincolanti con le Big Tech per la sicurezza e il watermarking dei contenuti generati dall'IA (letteralmente "filigrana"; una tecnica utilizzata per inserire un segnale distintivo invisibile all'interno di un contenuto generato da un algoritmo - testo, immagine, audio o video). Si punta, come sempre, anche su una esportazione dei "valori", per cui si sostiene che quella statunitense sarebbe un'IA "democratica" in contrapposizione a quella "autoritaria" cinese, cercando di attrarre i Paesi alleati nell'orbita tecnologica statunitense.

Se la Cina sta vincendo sulla diffusione con modelli locali e hardware efficiente, gli USA scommettono sulla potenza bruta e sul monopolio dell'accesso. Tuttavia, questa strategia è vulnerabile alla saturazione della rete elettrica e alla potenziale "bolla" finanziaria se i profitti dei data center non dovessero materializzarsi rapidamente.

Le criticità della strategia statunitense

Mentre i modelli statunitensi restano i più potenti in termini di ragionamento puro, la loro strategia mostra profonde criticità. Innanzitutto, c'è l'imponente drenaggio di capitali finanziari: l'IPO di SpaceX (che produce modelli di IA e razzi spaziali) avvenuta il 12 giugno, ha segnato un record storico per i mercati finanziari, drenando \$ 75 miliardi di capitali freschi, con Anthropic e OpenAI che seguiranno presto. Alphabet, la società madre di Google, prevede inoltre di raccogliere \$ 80 miliardi in finanziamenti azionari, la sua prima offerta pubblica iniziale in oltre vent'anni. Goldman Sachs, la mega banca d'investimento, ritiene che l'IA sia solo "una grande operazione speculativa sull'economia statunitense". E la bolla degli investimenti nell'IA si sta gonfiando ulteriormente. Il mercato azionario è sempre più concentrato in una manciata di titoli legati all'IA, che ora rappresentano oltre il 40% della capitalizzazione di mercato dell'S&P 500, secondo i dati di Bank of America⁷. La redditività complessiva viene gonfiata da una piccola fetta dell'economia che ottiene rendimenti straordinari dalla corsa allo sviluppo di capacità di IA. Il rischio, quindi, è che

l'economia, il ciclo degli utili e il mercato azionario siano tutti appoggiati allo stesso stretto pilastro. Se i rendimenti attesi dalle infrastrutture e dalle piattaforme di IA venissero messi in discussione, le conseguenze potrebbero non limitarsi a pochi titoli tecnologici dalle valutazioni elevate.

Inoltre, la strategia statunitense presenta diversi punti critici infrastrutturali, geopolitici e sociali che potrebbero comprometterne la sostenibilità nel lungo periodo. Il collo di bottiglia principale per gli USA nel 2026 non è più il software o il capitale, ma l'elettricità. I data center richiedono quantità di elettricità che un tempo erano associate a intere città, sollevando interrogativi su chi paga per le nuove infrastrutture, chi ha accesso alla scarsa energia e con quale rapidità i progetti possono connettersi alla rete. La rete elettrica statunitense attuale è obsoleta. Gli Stati Uniti ospitano più data center di qualsiasi altro Paese, circa 4.200, secondo Data Center Map, pari a circa il 40% del totale mondiale. Altri migliaia sono in fase di progettazione o costruzione, mentre solo quattro colossi della Silicon Valley spendono circa \$ 700 miliardi in infrastrutture per l'IA quest'anno. Il governo federale si è espresso a favore dell'assenza di limiti alla crescita, con Trump che a dicembre ha emanato un ordine esecutivo volto a impedire agli Stati di emanare leggi che limitino l'IA in quasi ogni sua forma.

I tempi per connettere nuovi data center alla rete (interconnessione) sono balzati a 4-7 anni in molti hub chiave. Un'analisi di JP Morgan ha rilevato che oltre il 60% della capacità dei data center prevista per il completamento nel 2027 non è ancora in costruzione e un ulteriore 7% è in ritardo. Sempre più numerose sono le comunità locali e le città che si organizzano contro la costruzione di data center a seguito della forte opposizione dei residenti, con candidati politici e legislatori che promettono una maggiore supervisione e limitazione o addirittura un blocco delle costruzioni. A livello nazionale, sette americani su dieci si oppongono alla costruzione di data center per l'intelligenza artificiale nelle proprie zone, secondo un sondaggio Gallup pubblicato a maggio. Gli americani di tutto il paese sono preoccupati per l'aumento dei prezzi delle utenze e per i danni ambientali causati da questi impianti ad alto consumo energetico. Entro il 2028, i data center potrebbero consumare fino al 12% dell'elettricità totale degli Stati Uniti. L'aumento della domanda rischia di far impennare i prezzi delle bollette per i cittadini, creando una potenziale crisi politica. Il "Ratepayer Protection Pledge" è un'iniziativa lanciata da Trump il 4 marzo 2026 per impedire che l'espansione dei data center per l'IA faccia aumentare le bollette elettriche delle famiglie statunitensi. Sette giganti tecnologici — Amazon, Google, Meta, Microsoft, OpenAI, Oracle e XAI — hanno firmato questo impegno formale accettando di farsi carico interamente dei costi energetici necessari alle loro infrastrutture.

Il modello centralizzato statunitense crea "punti unici di fallimento" critici. Il dominio di pochi giganti (Microsoft, Google, AWS) concentra enormi quantità di dati sensibili in poche infrastrutture, rendendole bersagli primari per attacchi informatici su larga scala. Molti alleati (in particolare in Europa) iniziano a temere la dipendenza dal cloud USA, paventando un possibile "interruttore di spegnimento" (kill switch) geopolitico, e quindi favorendo lo sviluppo di sovranità tecnologiche locali.

L'Europa sta iniziando a fare i conti: l'approccio politico conflittuale dell'amministrazione

Trump nei confronti dell'UE ha messo a nudo la pericolosa dipendenza del blocco dalla tecnologia statunitense. Il pericolo crescente è che questa potenza tecnologica possa essere usata contro l'Europa a livello politico. Elon Musk ha già sfruttato la sua proprietà di X e Starlink (satelliti) per interferire nel dibattito pubblico europeo e influenzare la guerra in Ucraina. E il 12 giugno (lo stesso giorno dell'IPO di SpaceX) il Dipartimento del Commercio statunitense ha ordinato all'azienda di IA Anthropic di limitare l'accesso dei cittadini stranieri ai suoi prodotti di frontiera come Mythos/Fable 5 per motivi di sicurezza (senza fornire dettagli tecnici specifici, ha sostenuto l'esistenza di un jailbreak, cioè di un metodo per aggirare i filtri dei modelli e accedere a capacità altrimenti bloccate). Una mossa straordinaria "capricciosa e caotica" (così l'ha definita la rivista The Economist) che sottolinea la ritrovata volontà dell'amministrazione Trump di esercitare il controllo su un settore considerato cruciale⁸. Non essendo in grado di distinguere tra statunitensi e stranieri, Anthropic ha bloccato l'accesso a tutti. Cosa accadrebbe se Washington, durante una disputa commerciale, interrompesse l'accesso dell'Europa ai chip avanzati statunitensi, o sfruttasse il suo controllo sui social media e sul cloud computing per spiare i governi europei e influenzare le elezioni? Dato che l'UE dipende da Paesi extra-UE per oltre l'80% della sua tecnologia e per il 70% del suo cloud computing, e considerando l'impegno dell'amministrazione Trump a "coltivare la resistenza" in Europa (sostenendo le forze politiche nazionaliste di estrema destra), nulla di tutto ciò sembra poi così inverosimile. Secondo la rivista Wired, in Europa sempre più aziende e istituzioni (governi, ong e istituzioni scolastiche e universitarie) stanno abbandonando le tecnologie statunitensi, scegliendo alternative open source o locali. La Finlandia ha scelto di non trasferire i propri dati elettorali sui servizi cloud di AWS, mentre l'organizzazione che gestisce il dominio di primo livello belga (.be) ha dichiarato che lascerà AWS. In Francia, il servizio di intelligence interno abbandonerà gli strumenti di analisi dei dati basati sull'IA di Palantir a favore di un fornitore nazionale. Il passaggio all'azienda francese ChapsVision (che sarebbe stata selezionata anche dal BfV, il servizio di sicurezza interno tedesco) è volto a evitare "dipendenze strategiche", ha affermato il Primo Ministro Sébastien Lecornu, esprimendo preoccupazione per la dipendenza da strumenti controllati dagli Stati Uniti. "Dobbiamo usare i nostri modelli di IA; non possiamo accettare nuove dipendenze strategiche nella sfera digitale, ha scritto Lecornu sui social media. "Non possiamo affidarci a strumenti sviluppati da potenze straniere. La Francia deve avere i propri strumenti". La Francia deve "costruire una vera autonomia" e "non dipendere dalla buona volontà di certi partner, capaci di chiudere il rubinetto dell'accesso" all'IA, ha affermato il primo ministro. Lecornu ha anche dichiarato che la Francia prevede di investire 655 milioni di euro nell'IA e di creare un chatbot condiviso per tutti i servizi statali. Verrà inoltre creato un chatbot per la salute pubblica destinato all'agenzia statale di assicurazione sanitaria Ameli. Le restrizioni sui chip (come quelle verso la Cina) stanno avendo conseguenze impreviste. Obbligando la Cina a non dipendere più da Nvidia, gli USA hanno involontariamente accelerato la creazione di un ecosistema tecnologico cinese autosufficiente. Le aziende tecnologiche statunitensi stanno perdendo ricavi miliardari in uno dei mercati più grandi al mondo, facilitando di fatto lo sviluppo di futuri competitor globali.

Nonostante gli investimenti, mancano le persone per costruire fisicamente il futuro dell'IA. Migliaia di progetti per data center e impianti di semiconduttori competono per lo stesso limitato bacino di manodopera specializzata, causando ritardi cronici e costi fuori controllo. Se la Cina sta costruendo un "motore distribuito" che corre sulle proprie strade (energia e hardware locali), gli USA stanno cercando di far volare un aereo potentissimo ma con piste di decollo (la rete elettrica) troppo corte e una manutenzione (forza lavoro) insufficiente.

Sintesi dei punti critici della strategia IA statunitense

Area	Criticità principale	Impatto previsto (2026-2027)
Energia	Rete elettrica obsoleta e tempi di allaccio biblici.	Ritardo o cancellazione del 50% dei nuovi data center pianificati.
Geopolitica	Export controls sui chip.	Nascita di un ecosistema IA cinese totalmente indipendente dagli standard USA.
Sicurezza	Monopolio del cloud.	Vulnerabilità a guasti a cascata e resistenza da parte degli alleati (EU).
Sociale	Costo dell'energia.	Possibile opposizione popolare alla costruzione di nuove centrali e data center.

Il modello di business statunitense è sostenibile?

Il modello di business dei giganti dell'IA statunitense è attualmente in una fase di forte tensione finanziaria e strutturale che ne mette in discussione la sostenibilità a lungo termine. I principali attori coinvolti, spesso definiti "hyperscalers" o "AI labs", includono Microsoft, Google (Alphabet), Amazon (AWS), Meta, Anthropic e OpenAI. Il cuore del problema è il divario tra gli investimenti colossali e i ritorni economici reali. Aziende come OpenAI e Anthropic stanno bruciando miliardi di dollari. Secondo proiezioni recenti, OpenAI potrebbe perdere fino a \$ 16 miliardi solo nel 2026, con perdite cumulative previste di \$ 44 miliardi prima di raggiungere la redditività nel 2029. Dati finanziari recentemente trapelati confermano una traiettoria persino più complessa: nel 2025 OpenAI ha registrato costi totali per \$ 34 miliardi a fronte di \$ 13,07 miliardi di ricavi, generando una perdita operativa di \$ 20,92 miliardi. Il picco di perdita netta (salita a \$ 38,5 miliardi) è dovuto alla transizione societaria in entità for-profit, che ha comportato enormi adeguamenti contabili non monetari sulle quote azionarie.

A differenza del software tradizionale, l'IA ha costi marginali elevatissimi. Ogni query ha un costo computazionale reale: i chatbot consumano un'immensa potenza di calcolo, che a sua

volta richiede elettricità, acqua per il raffreddamento dei server, terreni per i data center e miliardi di dollari di investimenti in hardware. Al momento, OpenAI, ad esempio, perde circa \$ 1,22 per ogni dollaro di ricavo generato e lo scorso anno ha speso \$ 34 miliardi in vista della prevista IPO. Molti sono consapevoli degli elevati costi fissi, ma non dei costi variabili di inferenza che si verificano ogni volta che il modello genera un'immagine. OpenAI prevede di spendere oltre \$ 150 miliardi solo per i costi di inferenza entro il 2030. Il consumo di cassa di Anthropic è molto inferiore, ma si attestava comunque a \$ 3 miliardi nel 2025, anno in cui ha triplicato la sua valutazione da \$ 61,5 miliardi a \$ 183 miliardi (a fine maggio 2026 ha raggiunto una valutazione di \$ 965 miliardi).

Mentre la stragrande maggioranza degli utenti continua ad accedere alle piattaforme gratuitamente, la domanda è come si colmerà il divario tra risorse e ricavi e chi ne sopporterà i costi. La concorrenza tra i giganti sta abbattendo i prezzi delle API. Se l'IA diventa una "commodity" a basso costo, recuperare le centinaia di miliardi spesi in infrastrutture diventa un'impresa titanica. Secondo la società di consulenza McKinsey, mentre il 90% delle aziende usa l'IA (la percentuale era il 33% nel 2023), per ora solo il 5,5% vede un reale impatto economico positivo. Se le aziende non vedono un ROI (Ritorno sull'Investimento), potrebbero tagliare i budget per gli abbonamenti cloud, lasciando i data center "vuoti"⁹. La promessa intrinseca dell'utilizzo dell'IA è che il denaro che un'azienda spende per questi strumenti venga ampiamente ripagato dal miglioramento della produttività, un indicatore di efficienza economica, dove una maggiore produttività significa ottenere più risultati da ogni lavoratore. Se questo obiettivo non si verifica, allora i presupposti alla base delle valutazioni e delle politiche sull'IA vengono compromessi. Anthropic ha iniziato a guadagnare terreno su OpenAI verso la fine dello scorso anno, quando il suo strumento Claude Code è diventato virale soprattutto tra gli sviluppatori di software dell'area di San Francisco, prima di diffondersi più ampiamente¹⁰. Claude Code ha rappresentato un cambiamento nel modo in cui vengono utilizzati i grandi modelli linguistici (LLM) - la tecnologia alla base dei chatbot - inaugurando una transizione verso agenti di IA autonomi che svolgono compiti senza intervento umano, consentendo anche ai meno esperti di tecnologia di creare software e svolgere un'ampia gamma di attività.

Le aziende sono ancora incerte su come utilizzare l'IA in modo efficace e sono sempre più preoccupate dei potenziali svantaggi derivanti dal suo utilizzo. Tra questi svantaggi rientrano le "allucinazioni" (ovvero le finzioni create dal modello di IA), intrinseche ai modelli di apprendimento LLM. Uno studio ha rilevato che, per un set di addestramento di 32.000 parole, il tasso medio di allucinazioni negli LLM era del 6,8%. Quando il set è stato ampliato a 128.000 parole, il tasso medio di allucinazioni è salito al 10%. Ciò comporta un notevole dispendio di tempo per la correzione e il monitoraggio da parte degli operatori umani.

Un altro problema è che, poiché i LLM sono progettati per essere bravi in tutto, non eccellono in nessun ambito specifico rispetto alle app specializzate. Un rapporto sull'utilizzo dell'IA nello sviluppo di software ha rilevato un impatto esplosivo all'inizio, con i programmatori che creavano o modificavano quasi il 300% in più di file, ma tale incremento si è dimezzato al 150% quando le aziende hanno ricevuto il numero di elaborati da

sottoporre a revisione, e a sua volta si è ridotto di cinque volte, arrivando a un aumento di circa il 30% al momento del rilascio completo del software. Inoltre, quando i ricercatori hanno esaminato se l'aumento della produzione di software facilitato dall'IA abbia portato a un maggiore utilizzo da parte dei clienti, hanno trovato scarse prove. Il notevole aumento delle app per dispositivi mobili rilasciate nell'ultimo anno non è stato accompagnato da alcun aumento dei download: la maggior parte delle nuove app non riesce a conquistare nemmeno un pubblico modesto.

La crescita del PIL statunitense è ormai trainata quasi esclusivamente dall'aumento della spesa tecnologica¹¹. Se questa dovesse iniziare a diminuire, l'economia statunitense entrerebbe in recessione molto rapidamente, anche se gli investimenti tecnologici calassero solo leggermente, diciamo del 4-6%, come è accaduto dopo i boom tecnologici, di entità molto minore, degli anni '60 e durante la recessione del 2008¹².

Per sopravvivere, i giganti statunitensi stanno cambiando pelle. Microsoft e Google stanno sviluppando propri chip (come i chip Azure Maia e i Google TPU) per smettere di pagare i margini altissimi a Nvidia. Microsoft sta riducendo la dipendenza da OpenAI sviluppando i propri modelli interni per ottimizzare i costi. Wall Street è sempre più nervosa. Nel 2026, la pazienza per la crescita "a ogni costo" sta finendo, spingendo le aziende oltre l'indebitamento tramite l'emissione di obbligazioni in tutto il mondo o lo schema del "finanziamento circolare"¹³, verso IPO (quotazioni in borsa) anticipate per raccogliere nuova liquidità.

Il modello attuale basato su perdite miliardarie non è sostenibile indefinitamente. Assisteremo a una "selezione naturale" dove solo chi riuscirà a trasformare l'IA in un agente autonomo capace di sostituire processi lavorativi interi – un gergo aziendale per indicare l'esecuzione di un'attività completa dall'inizio alla fine – (aumentando la produttività e generando valore reale) riuscirà a rendere i conti profittevoli. Per il capitale e le grandi aziende mediatiche, l'obiettivo è rendere l'IA una tecnologia redditizia, ma ciò può essere fatto solo riducendo drasticamente la forza lavoro e bloccando qualsiasi tentativo di regolamentarne le applicazioni e l'utilizzo. Se l'IA deve avere successo per il capitale, ciò avverrà solo a spese della maggior parte dei lavoratori e delle loro famiglie¹⁴.

I punti critici: vulnerabilità incrociate

Nel 2026, la corsa alla supremazia tecnologica rivela che nessuna delle due superpotenze è immune da rischi sistemici, delineando un quadro di vulnerabilità incrociate. La Cina, nonostante i passi da gigante nell'autosufficienza, soffre ancora per la dipendenza critica da tecnologie estere insostituibili, come i macchinari per la litografia di ASML, necessari per produrre chip avanzati sotto la soglia dei 7nm. Oltre al limite fisico, Pechino deve affrontare un vincolo politico: l'obbligo di sottoporre ogni algoritmo a rigidi controlli ideologici (il cosiddetto "filtro socialista"). Questa censura preventiva non limita solo il discorso politico, ma rischia di soffocare la "serendipità" e la creatività laterale indispensabili per le scoperte scientifiche più radicali e dirompenti.

Dall'altro lato, gli Stati Uniti mostrano crepe strutturali opposte. Il loro modello iper-centralizzato, basato su pochi grandi nodi cloud, crea "punti unici di fallimento" (single points of failure) estremamente pericolosi: un singolo attacco informatico massivo o un guasto infrastrutturale potrebbe paralizzare l'intera economia digitale dipendente da quelle API. Inoltre, la strategia aggressiva delle restrizioni sulle esportazioni sta producendo un effetto boomerang: ha accelerato la nascita di un ecosistema cinese totalmente indipendente e resiliente, privando contemporaneamente le aziende statunitensi di ricavi miliardari necessari per finanziare la propria ricerca ed espansione. In questo stallo, la vulnerabilità di un Paese diventa spesso lo specchio delle debolezze dell'altro.

La condizione dell'UE, schiacciata tra Cina e USA: colonia o arbitro?

L'Unione Europea si trova in mezzo a questa morsa, in una posizione paradossale: è il "campo di battaglia" normativo e commerciale dei due giganti, ma nel 2026 sta cercando di tracciare una "terza via" per evitare di essere stritolata.

Sebbene l'Europa sia in netto ritardo sugli investimenti (l'UE rappresenta solo il 9% dei capitali globali in IA contro il 43% degli USA), la sua strategia non è di pura sottomissione, ma di "resistenza sovrana". L'Europa ha capito che non può battere gli USA sui modelli chiusi o la Cina sulla scala industriale. La sua risposta si articola su due fronti:

1. modelli open-source ad alte prestazioni: aziende come Mistral AI (Francia), Bending Spoons (Italia), Helsing (Germania), Wayve (Regno Unito) e Aleph Alpha (Germania) sono diventate l'alternativa europea a OpenAI, raccogliendo finanziamenti miliardari per costruire modelli aperti e personalizzabili che piacciono alle aziende che temono il controllo USA;
2. infrastrutture sovrane: nel 2026 sono nati i primi "AI Sovereign Cloud", come l'Industrial AI Cloud di Monaco. L'obiettivo è permettere alle industrie europee di usare l'IA senza che i dati escano dai confini UE o siano soggetti a leggi straniere.

L'Europa usa il suo mercato da 450 milioni di persone come leva - il cosiddetto "Bruxelles effect" del regolare per sopravvivere. L'AI Act è entrato pienamente in vigore tra il 2025 e il 2026. Impone standard di sicurezza mondiali. Le aziende statunitensi e cinesi devono conformarsi se vogliono vendere in Europa (ad esempio, la nuova IA Siri per la serie 27 dell'iPhone di Apple non funziona a causa delle normative europee), esportando di fatto le regole europee nel resto del mondo. Inoltre, consapevole del rischio di eccessiva burocrazia, a metà 2026 l'UE ha approvato il "Digital Omnibus Package" per tagliare i costi di compliance per le startup IA europee, cercando di non soffocare l'innovazione interna. L'Europa sta iniziando a prendere le distanze dalle grandi aziende tecnologiche statunitensi, ma continua a rispettare le regole della Silicon Valley, senza sviluppare una propria visione su come l'IA dovrebbe essere al servizio della società. Lo scorso 3 giugno la Commissione europea ha presentato alcune nuove proposte di legge che vanno a costituire un

“pacchetto per la sovranità digitale” (l'European Technological Sovereignty Package) per promuovere le tecnologie europee autoctone e proteggere l'UE dalle interferenze straniere. Nel suo complesso, il pacchetto rappresenta un gradito, seppur tardivo, riconoscimento del fatto che la dipendenza dalle aziende tecnologiche statunitensi non è solo un problema economico, ma una minaccia diretta all'indipendenza, alla resilienza e alla sicurezza del blocco. I provvedimenti andrebbero a regolamentare ed indirizzare vari aspetti dell'industria tecnologica europea, a partire dalla produzione di semiconduttori, arrivando allo sviluppo di sistemi cloud e IA, fino all'ammodernamento della rete elettrica europea. Il pacchetto si compone di quattro diverse iniziative: il Chips Act 2.0, il Cloud and AI Development Act (CADA), la Open Source Strategy e la Strategic Roadmap for Digitalisation and AI in the Energy Sector¹⁵. Se dal lato del software libero e della ricerca scientifica open access gli attori europei possono attingere ad una pila di software e conoscenze molto estesa e anche tecnologicamente avanzata, non si può fare lo stesso con le fabbriche di semiconduttori¹⁶ e la costruzione di data center. Queste infrastrutture, infatti, richiedono risorse ingenti e anni per essere costruite e per entrare a regime.

Il fulcro del pacchetto è il CADA che creerebbe un sistema di classificazione per i fornitori di servizi cloud che gestiscono dati del settore pubblico, come Amazon Web Services, Microsoft Azure o la francese OVHCloud. In teoria, le operazioni e i dati più sensibili, in particolare quelli relativi alla sicurezza nazionale e all'applicazione della legge, sarebbero riservati ai fornitori che soddisfano i più elevati standard di sovranità, stabilendo una chiara preferenza per i fornitori europei.

La Commissione si impegna a triplicare la capacità dei data center europei in un periodo compreso tra 5 e 7 anni. Un impegno incentrato su misure contenute nel CADA che obbligano ogni Paese dell'UE a istituire “zone di accelerazione per i data center”. All'interno di queste zone, le autorità locali dovrebbero approvare le richieste di autorizzazione per i data center entro 12 mesi, anche semplificando le valutazioni ambientali e urbanistiche per agevolare il rilascio dei permessi. Le zone di accelerazione sollevano serie preoccupazioni in merito a trasparenza, responsabilità democratica e sostenibilità, in un momento in cui l'opposizione pubblica ai data center sta esplodendo a causa del loro impatto sull'ambiente e sulle bollette elettriche delle famiglie. Rischiano inoltre di compromettere gli obiettivi di sovranità della Commissione; non includendo criteri relativi alle dimensioni o alla nazionalità delle aziende, per cui le zone potrebbero finire per consolidare ulteriormente il dominio degli hyperscalers statunitensi che dominano il mercato europeo del cloud.

Nonostante gli sforzi, l'UE resta esposta a rischi critici. L'Europa non produce i chip avanzati necessari per l'IA. Anche se sta costruendo data center, i “motori” (Nvidia o Huawei) restano stranieri. Gli USA continuano ad attrarre i migliori ricercatori europei con stipendi tripli, rendendo l'UE una sorta di “vivaio” per la Silicon Valley. A differenza di Cina e USA, l'Europa è divisa in 27 mercati con lingue e regole diverse, il che rende difficilissimo per una startup crescere velocemente quanto un competitor cinese o statunitense.

In sintesi, l'UE non è ancora “finita”, ma è in una corsa disperata contro il tempo. Se riuscirà a imporre i suoi standard etici/normativi e a far crescere campioni nazionali come Mistral AI¹⁷ o Aleph Alpha, diventerà l'arbitro della partita. Se fallirà, diventerà semplicemente una

colonia digitale: un soggetto che si adegua alle decisioni altrui, consumando tecnologia altrui e pagando royalties altissime, piuttosto che un soggetto che prende delle decisioni in base allo sviluppo di una propria base tecnologica.

Alessandro Scassellati

1. La reazione contro l'IA sta assumendo anche una piega estremista, seguendo le orme dei militanti tecnopessimisti dei passati periodi di cambiamento tecnologico (a partire dalla ribellione Luddista dei lavoratori tessili britannici all'inizio del XIX secolo, quando distrussero le macchine tessili automatizzate per rivendicare maggiori diritti lavorativi). Quando un ventenne texano è stato arrestato all'inizio di quest'anno con l'accusa di aver tentato di incendiare la sede centrale di OpenAI e la casa di Sam Altman, le autorità hanno trovato un manifesto anti-IA accanto al suo accendino e a una tanica di cherosene. Si è trattato di uno dei tanti attacchi che hanno destato allarme tra i ricercatori, l'industria tecnologica e le forze dell'ordine per la crescente diffusione dell'estremismo anti-tecnologico. Che si tratti di gruppi antigovernativi violenti che si oppongono alla sorveglianza di massa, di ecofascisti con rivendicazioni ambientali, di accelerazionisti neonazisti decisi a far collassare le infrastrutture tecnologiche critiche o dell'uomo che avrebbe preso di mira la casa di Altman, preoccupato che un'IA superpotente possa distruggere l'umanità, l'IA è diventata un'ossessione per tutto lo spettro estremista.[↔]
2. La famiglia di modelli Qwen (Alibaba) ha catturato oltre il 50% di tutti i download globali di modelli open-source su piattaforme di riferimento come Hugging Face. Sono presenti oltre 180.000 modelli derivati basati sulle fondamenta di Qwen (più di Meta e Google messi insieme). Il 63% dei nuovi modelli personalizzati dagli sviluppatori nasce ormai su architetture di origine cinese.[↔]
3. I 500 MW solari costituiscono la prima fase di un maxi-complesso da 2 GW complessivi, che include un'estensione di ulteriori 1,5 GW di energia eolica.[↔]
4. Il framework normativo della Repubblica Popolare Cinese impone linee rosse insuperabili. Questo condiziona l'output dei modelli non solo sulle piattaforme cloud ufficiali, ma persino quando vengono eseguiti in locale dagli sviluppatori indipendenti. Le aziende tecnologiche cinesi operano sotto la stretta vigilanza della Cyberspace Administration of China (CAC). Normative cruciali come le Generative AI Interim Measures (2023), gli emendamenti del 2026 alla Cybersecurity Law e le nuovissime regole sui servizi interattivi (Anthropomorphic Interaction Services) impongono un principio inderogabile: l'output dell'IA deve allinearsi ai valori socialisti fondamentali e non può minare la sicurezza nazionale o l'unità dello Stato. Di conseguenza, i modelli cinesi integrano un sistema di censura e moderazione multilivello. Gli sviluppatori indipendenti devono distinguere tra l'utilizzo dei modelli tramite API ufficiali e l'esecuzione on-premise (locale) tramite i pesi scaricati. Se si interrogano i server di DeepSeek o Alibaba Cloud, la moderazione è rigida, centralizzata e applicata tramite filtri di sistema (guardrail esterni). Uno studio condotto dalla piattaforma di testing Promptfoo nel 2026 ha dimostrato che le API di DeepSeek V4 rifiutano l'85% delle domande riguardanti argomenti politicamente sensibili. Richieste sullo status politico di Taiwan, i fatti di Piazza Tienanmen, la gestione della regione dello Xinjiang, la Rivoluzione Culturale o qualsiasi critica diretta alla leadership del PCC attivano risposte di rifiuto standardizzate. Quando uno sviluppatore scarica i pesi di Qwen o DeepSeek per farli girare sul proprio server offline, il filtro API esterno scompare, ma la moderazione rimane parzialmente impressa nei paesi stessi.[↔]
5. Nell'ultimo anno si sono manifestati segnali di un cambiamento di retorica da parte del settore dell'IA, alle prese con la diffusa diffidenza del pubblico. Il mese scorso Altman ha affermato che l'IA probabilmente non porterà all'"apocalisse del lavoro" di cui aveva parlato in passato, nonostante aziende come Meta stiano licenziando decine di migliaia di dipendenti. Nel frattempo, OpenAI e Anthropic hanno entrambe annunciato quest'anno fondi e think tank volti ad aiutare le istituzioni civili ad adattarsi all'IA, con l'organizzazione no-profit di OpenAI che ha stanziato \$ 250 milioni in sovvenzioni per programmi che aiutino i lavoratori ad affrontare i cambiamenti introdotti dall'IA.[↔]
6. La chiusura dei canali legittimi per affrontare l'opposizione pubblica all'IA, così come la sensazione che la tecnologia venga imposta alla società, sta creando quello che i ricercatori definiscono un vuoto di responsabilità

che può incentivare il terrorismo e la violenza politica. Donald Trump, in sintonia con i leader del settore tecnologico, ha emesso lo scorso anno un ordine esecutivo nel tentativo di bloccare qualsiasi legislazione statale volta a limitare lo sviluppo dell'IA e ha affermato che nulla rallenterà gli Stati Uniti nella corsa globale all'IA. I miliardari del settore tecnologico stanno inoltre investendo milioni di dollari in attività di lobbying e spese politiche nel tentativo di impedire la regolamentazione dell'IA.[↔]

7. L'indice S&P 500, che monitora le 500 maggiori aziende statunitensi, ha registrato una crescita vertiginosa negli ultimi 5 anni, con un aumento di quasi l'80%. Questo balzo è stato trainato dai titoli delle grandi aziende tecnologiche che hanno investito nel boom dell'IA, ovvero le "magnifiche sette" - Alphabet, Amazon, Apple, Meta, Microsoft, Nvidia e Tesla -, insieme a quelle di altre 34 aziende che operano nell'IA.[↔]
8. La situazione, per certi versi, rispecchia lo scontro di quest'anno tra Anthropic e il Dipartimento della Difesa, culminato con il Pentagono che definì l'azienda un rischio per la sicurezza nazionale nella catena di approvvigionamento. All'epoca, i rappresentanti del settore avvertirono che la disputa avrebbe ostacolato gli Stati Uniti nella corsa globale allo sviluppo di un'IA avanzata. Tuttavia, i rapporti tra Anthropic e il governo si sono presto in parte distesi, spinti anche dalla volontà del governo di accedere a Mythos, un nuovo e potente modello di IA di Anthropic la cui versione pubblica è stata rilasciata la scorsa settimana con il nome di Fable 5.[↔]
9. Ogni volta che un chatbot o un agente basato sull'IA fornisce una risposta, questa viene misurata in "token", ovvero elementi costitutivi del linguaggio che possono essere parole, segni di punteggiatura o sillabe. Ad esempio, OpenAI afferma che la frase "Perdi il 100% dei tiri che non fai" vale 11 token. I token vengono utilizzati anche per misurare gli input, come ad esempio il testo che si digita in ChatGPT. I costi variano a seconda del modello; OpenAI li propone a \$ 5 per milione di token di input per GPT-5.5 e a \$ 30 per milione di token di output (ovvero la risposta fornita al prompt). Il problema per gli abbonati è che i costi dei token stanno aumentando vertiginosamente, anche se le aziende ovunque incoraggiano i dipendenti a "massimizzare l'utilizzo dei token", ovvero a sfruttare al massimo l'IA. Il problema per le aziende che si occupano di IA è che non applicano ancora tariffe adeguate.[↔]
10. La competizione spietata tra Anthropic (con Claude Code) e OpenAI sta innescando una guerra al ribasso sui prezzi delle API. Vendere i token a prezzi stracciati per conquistare quote di mercato sta erodendo i margini operativi proprio nel momento in cui le aziende tech si preparano a sbarcare in borsa (IPO) e a mostrare i propri bilanci al pubblico dei mercati azionari. D'altra parte, l'aggressiva politica commerciale asiatica ha scardinato il modello di business basato sulla vendita dei token a caro prezzo. Con il consolidamento di famiglie di modelli come DeepSeek V4 (con le varianti Pro e Flash), i prezzi delle API sono stati tagliati in modo permanente fino al 75%. DeepSeek V4 Pro si attesta a circa \$ 0.43 per milione di token in input e \$ 0.87 in output. Strumenti di frontiera commerciali statunitensi equivalenti (come GPT-5.5 o Claude Opus) risultano da 7 a 12 volte più costosi a parità di performance di calcolo. Il "fossato economico" dei laboratori della Silicon Valley si è svuotato. Le startup occidentali non competono più sul costo di accesso ai modelli, ma sull'esperienza d'uso (Vibe Coding) e sull'orchestrazione degli agenti.[↔]
11. L'economia statunitense odierna è in realtà composta da due economie distinte. C'è l'economia tecnologica e poi c'è tutto il resto. Negli ultimi quattro trimestri, fino alla fine del primo trimestre del 2026, il 93% della crescita del PIL statunitense è stato dovuto esclusivamente agli investimenti nel settore tecnologico (sebbene gran parte degli acquisti siano importazioni e non prodotti internamente). Si tratta di una bolla pronta a scoppiare.[↔]
12. Gita Gopinath, ex capo economista del FMI, ha calcolato che un crollo del mercato azionario dell'IA, equivalente a quello che ha posto fine alla bolla delle dot-com alla fine degli anni '90, cancellerebbe circa \$ 20 trilioni di ricchezza delle famiglie statunitensi e altri \$ 15 trilioni all'estero, una cifra sufficiente a strangolare la spesa dei consumatori e a indurre una recessione globale. Questa è anche l'opinione del FMI che teme che le aziende di IA non riescano a generare utili commisurati alle loro elevate valutazioni. Il crollo dei precedenti boom degli investimenti ha ridotto in media di circa 1 punto percentuale la crescita del PIL reale statunitense. Anche una correzione moderata delle valutazioni azionarie dell'IA ridurrebbe la crescita globale dello 0,4%. [↔]
13. Lo schema del "finanziamento circolare" si basa sugli investimenti incrociati tra Microsoft, OpenAI e altri. In sostanza, un colosso come Microsoft, ricco di liquidità, acquista hardware da Nvidia, AMD e altri fornitori. Nvidia utilizza poi questi ricavi per acquisire una partecipazione multimiliardaria in OpenAI. OpenAI, a sua volta, utilizza questi fondi per assicurarsi capacità di calcolo nei data center di Microsoft. Anche Microsoft investe in OpenAI e

stipula un accordo di condivisione dei ricavi, in base al quale parte dei ricavi di OpenAI confluisce in Microsoft e viceversa, man mano che le due aziende utilizzano i prodotti l'una dell'altra. Ipotizzando che Microsoft spenda \$ 100 miliardi per ordinare hardware per i data center, Nvidia, AMD e altri fornitori possono contabilizzare questi \$ 100 miliardi come ricavi. Utilizzano quindi questi fondi per investire in OpenAI (ad esempio), che a sua volta li utilizza per prenotare capacità di calcolo nei data center di Microsoft. Microsoft contabilizza questo investimento in OpenAI come ricavo, trasformando di fatto la sua spesa di \$ 100 miliardi in miliardi di ricavi.[↔]

14. Le misurazioni di METR, un'organizzazione di ricerca che misura le capacità dell'IA, si basano sulla capacità dei modelli di IA di svolgere un'attività di programmazione, quantificata dal tempo che impiegherebbe un essere umano a farlo. Secondo questo parametro, le capacità dei modelli di IA raddoppiano ogni 4 mesi. Ad esempio, si stima che il modello Claude Mythos di Anthropic raggiunga un tasso di successo del 50% in attività che richiederebbero a un esperto umano dalle 8 ore ai 2 giorni. Tuttavia, finora non si è registrato un impatto proporzionale sul mercato del lavoro. Un rapporto di Anthropic pubblicato a marzo conteneva una ricerca che dimostrava come, in teoria, l'IA potrebbe svolgere una vasta gamma di lavori, dall'informatica al settore legale, ma finora non lo ha fatto in modo significativo. Esistono degli ostacoli all'adozione dell'IA nel mondo del lavoro. Ad esempio, quanta parte del lavoro di un amministratore delegato o di un dirigente senior può essere tranquillamente esternalizzata a un bot? Le attività legalmente sensibili possono essere svolte da qualcosa di diverso da un essere umano? Ciononostante, il cambiamento è in arrivo. Molte persone svolgono mansioni che potrebbero essere eseguite da un'IA. L'entità dei cambiamenti a cui assisteremo sarà enorme. Per ora, secondo i dati elaborati dall'Organizzazione Internazionale del Lavoro (ILO), il numero di lavoratrici e lavoratori licenziati negli ultimi 3 anni come conseguenza dell'IA sono circa 425.000 a livello globale. Di questi 142mila sono in Europa. Secondo l'ILO, il 25% dell'occupazione globale rientra in professioni potenzialmente esposte all'IA, con percentuali più elevate nei Paesi ad alto reddito (34%), dove un lavoratore su 4 potrebbe essere sostituito da una macchina nei prossimi anni. I lavori più a rischio sono quelli che prevedono attività ripetitive e componenti digitali e testuali facilmente automatizzabili: call center, assistenza amministrativa, assistenza clienti, impiegati di banche e poste, cassieri e traduttori.[↔]
15. È importante puntualizzare che questa comunicazione della Commissione Europea agisce attualmente come soft law (quadro programmatico). Per diventare leggi vincolanti, i due pilastri normativi principali — il Chips Act 2.0 e il CADA — dovranno affrontare l'iter di negoziazione e approvazione formale da parte del Parlamento Europeo e del Consiglio dell'UE.[↔]
16. Il Chips Act 2.0 punta esplicitamente a mobilitare 120 miliardi di euro entro il 2035 per colmare il vuoto della manifattura hardware sotto i 7nm. Introduce inoltre i cosiddetti "Demand Accelerators" per obbligare l'industria acquirente europea a connettersi direttamente con i produttori di chip interni, nel tentativo di forzare la nascita di un mercato autoctono.[↔]
17. La Francia ha iniziato a implementare uno strumento governativo basato sull'IA, che offre un chatbot a un milione dei suoi 2,6 milioni di dipendenti pubblici. Basato su modelli della startup francese Mistral AI, il sistema è pensato per fornire supporto in situazioni come l'accelerazione dei procedimenti legali o l'ottenimento di finanziamenti per i ricercatori, con i ministri desiderosi di contrastare i rischi per la sicurezza posti dagli strumenti di IA commerciali.[↔]