

Perché investire nell'infrastruttura per l'IA



WisdomTree.eu
+44 (0) 207 448 4330

Perché investire nell'infrastruttura per l'IA: la base fisica per un futuro digitale

L'infrastruttura per l'intelligenza artificiale (IA) sta diventando un pilastro fondamentale dell'economia globale. Con il miglioramento dell'intelligenza dei modelli e l'affermarsi degli agenti di IA nei carichi di lavoro digitali, la domanda di potenza di calcolo sta crescendo a un ritmo sostenuto.

Il valore non risiede più esclusivamente nelle GPU¹ o nei modelli, ma anche nell'ecosistema di calcolo che li circonda e li rende possibili. L'IA si sta evolvendo in un sistema fisico che dipende dall'energia, dall'hardware e da sistemi e catene di approvvigionamento sempre più complessi. Questo sta determinando una crescita costante nei seguenti ambiti:



Semiconduttori e produzione



Collegamenti di rete



Centri dati ed energia

1 GPU: unità di elaborazione grafica. Progettate originariamente per il rendering grafico, le GPU sono oggi i principali processori utilizzati per la formazione e l'inferenza dell'IA grazie alle loro capacità di elaborazione parallela.

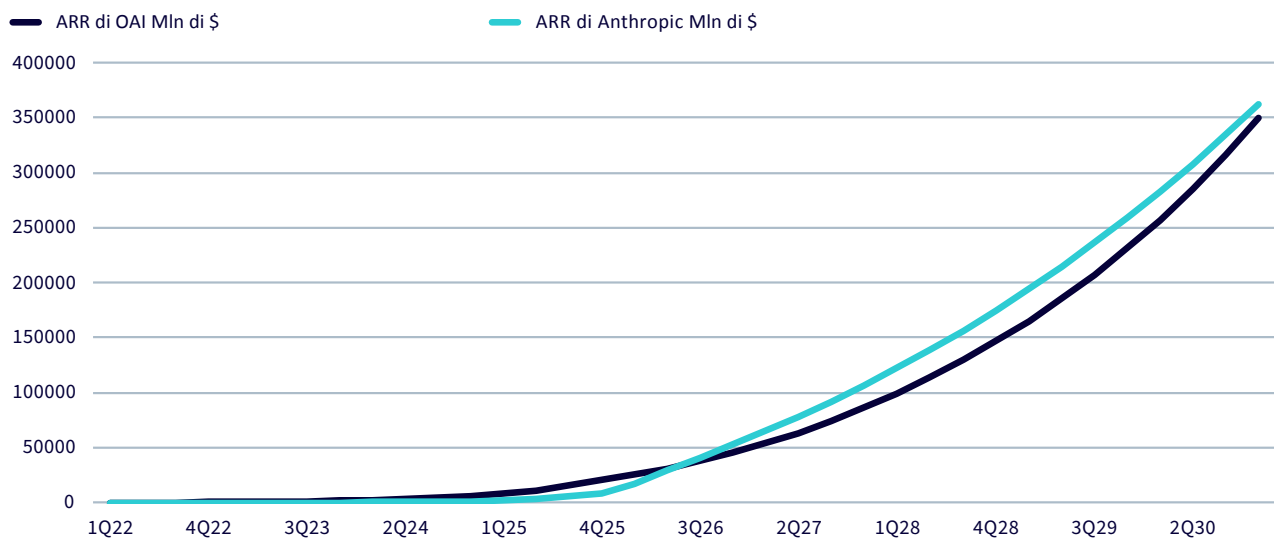
Domanda di inferenza e scalabilità della potenza di calcolo

L'elemento distintivo di questo ciclo è rappresentato dalla natura della domanda. Quella di potenza di calcolo per l'IA è ormai evidente e costante in tutti i casi d'uso, sia aziendali che consumer. La formazione di un modello richiede una potenza di calcolo enorme per un periodo di tempo determinato, mentre l'inferenza è un processo continuo.

Al livello più elementare, questa domanda si misura in token, ovvero le unità di testo, codice o dati elaborati dai modelli di IA. Ogni query, flusso di lavoro e attività guidata da un agente genera token, ciascuno dei quali richiede potenza di calcolo per essere elaborato. Man mano che l'utilizzo cresce, la generazione di token aumenta di pari passo, creando una domanda strutturale di potenza di calcolo.

Questo si riflette nella crescita dei principali laboratori di modelli di IA, ovvero i “produttori di token” della tecnologia.

Figura 1: Ricavi annualizzati dei principali fornitori di LLM²



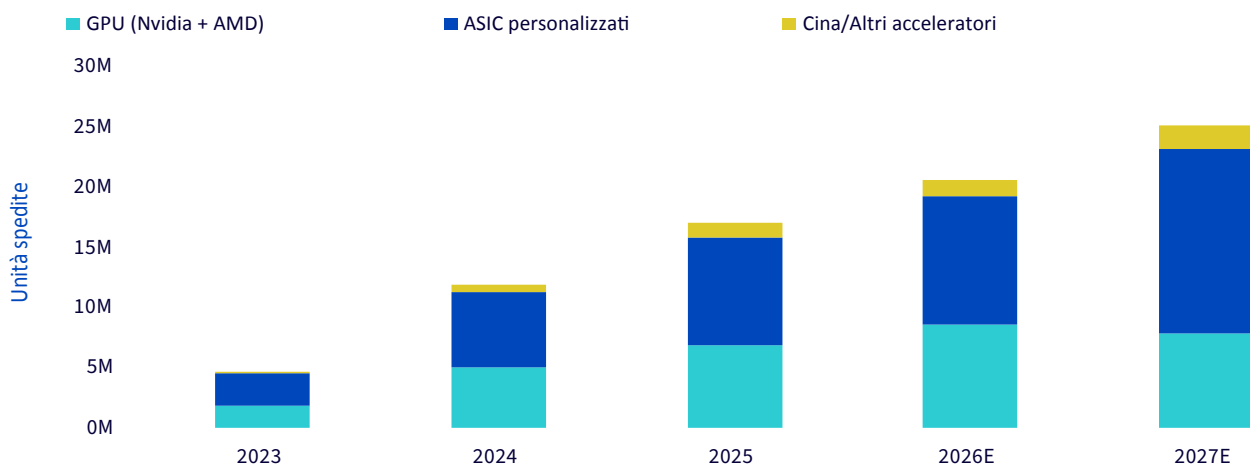
Fonte: SemiAnalysis, aprile 2026. Le previsioni non sono indicative della performance futura e qualsiasi investimento è soggetto a rischi e incertezze. La performance storica non è indicativa di quella futura e qualsiasi investimento può diminuire di valore.

2 LLM: modello linguistico di grandi dimensioni. Un tipo di modello di IA formato su grandi set di dati testuali per generare, comprendere e ragionare sul linguaggio. Tra gli esempi figurano la serie GPT di OpenAI e Claude di Anthropic.

I ricavi ricorrenti annualizzati (ARR)³ dei fornitori di modelli di IA, come OpenAI e Anthropic, sono passati da zero a oltre 30 miliardi di dollari nel giro di pochi anni. Il ritmo dell'adozione supera quello dei precedenti cicli tecnologici⁴ e segue una curva di crescita esponenziale, con SemiAnalysis che prevede ARR superiori ai 300 miliardi di dollari entro il 2030.

Gli hyperscaler⁵ e i fornitori di neocloud⁶ offrono gran parte della potenza di calcolo su cui fanno affidamento questi laboratori di IA, mentre sviluppano anche i propri modelli concorrenti. Entrambe le parti stanno investendo centinaia di miliardi in capitali con lo scopo di espandere la capacità. Si tratta di decisioni lungimiranti con orizzonti pluriennali, incentrate sulla distribuzione di chip sempre più specializzati per ottenere la massima efficienza. Questo sta dando vita alla progettazione di nuovi chip e a un aumento degli ordini per i produttori partner.

Figura 2: Spedizioni di unità XPU⁷



Fonte: SemiAnalysis Accelerator Model, aprile 2026. Le previsioni non sono indicative della performance futura e qualsiasi investimento è soggetto a rischi e incertezze. La performance storica non è indicativa di quella futura e qualsiasi investimento può diminuire di valore.

Si prevede che le spedizioni di acceleratori di IA passeranno dai circa 5 milioni di unità del 2023 a oltre 25 milioni entro il 2027, con un aumento di cinque volte nel corso del periodo. Questo riflette sia la crescente necessità di supportare i carichi di lavoro dei sistemi di calcolo dell'IA, sia un ecosistema in espansione che va oltre un singolo fornitore o una singola architettura.

3 ARR: ricavi ricorrenti annualizzati. Una misura normalizzata dei ricavi prevedibili e ricorrenti su base annuale, comunemente utilizzata per monitorare la crescita delle aziende tecnologiche che operano su abbonamento.

4 Al momento del suo lancio, ChatGPT è stata ampiamente riconosciuta come l'applicazione consumer in più rapida crescita nella storia, con 100 milioni di utenti mensili attivi raggiunti in appena due mesi.

5 Gli hyperscaler sono grandi fornitori di servizi cloud che gestiscono enormi infrastrutture per centri dati a livello globale.

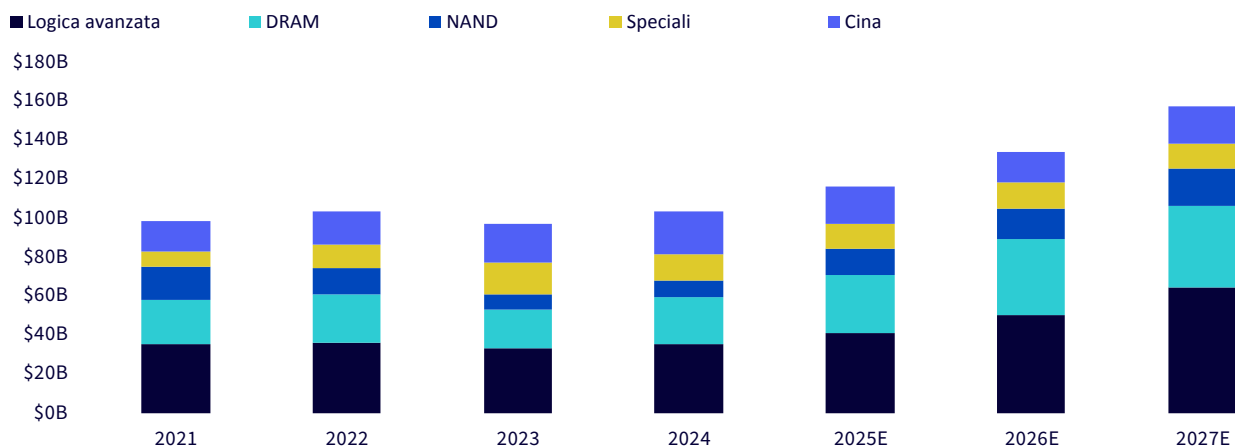
6 I neocloud sono fornitori di servizi cloud specializzati e incentrati sull'IA che offrono GPU-as-a-Service on-demand ad alte prestazioni per la formazione e l'inferenza dell'IA.

7 XPU: termine generico che indica qualsiasi unità di elaborazione acceleratrice utilizzata nei carichi di lavoro dell'IA, tra cui GPU, ASIC personalizzati e altre soluzioni in silicio specializzate.

Scalabilità dei semiconduttori e strozzature nella catena di approvvigionamento

Soddisfare questa domanda di chip richiede una forte espansione della produzione di semiconduttori. Attualmente, le fonderie non dispongono di una capacità sufficiente a produrre il volume richiesto dai clienti. Svilupparla significa costruire nuovi impianti e investire in attrezzature specializzate, note come apparecchiature per la fabbricazione di wafer (WFE), utilizzate in ogni fase della trasformazione di un wafer di silicio grezzo in un chip finito.

Figura 3: Spesa globale per WFE⁸, per applicazione



Fonte: SemiAnalysis WFE Model, febbraio 2026. Le previsioni non sono indicative della performance futura e qualsiasi investimento è soggetto a rischi e incertezze. La performance storica non è indicativa di quella futura e qualsiasi investimento può diminuire di valore.

Si prevede che la spesa globale per le apparecchiature per la produzione di wafer (WFE) raggiungerà i 160 miliardi di dollari entro il 2027, a testimonianza di un cambiamento strutturale orientato all'espansione della capacità produttiva guidata dall'IA. Nonostante questo aumento, permangono alcuni ostacoli fondamentali, in particolare per quanto riguarda la litografia⁹ avanzata e la capacità delle camere bianche¹⁰, che continuano a limitare la produttività.

8 WFE: apparecchiature per la produzione di wafer. I macchinari specializzati utilizzati per la produzione di chip semiconduttori, compresi i sistemi di litografia, deposizione, incisione e ispezione.

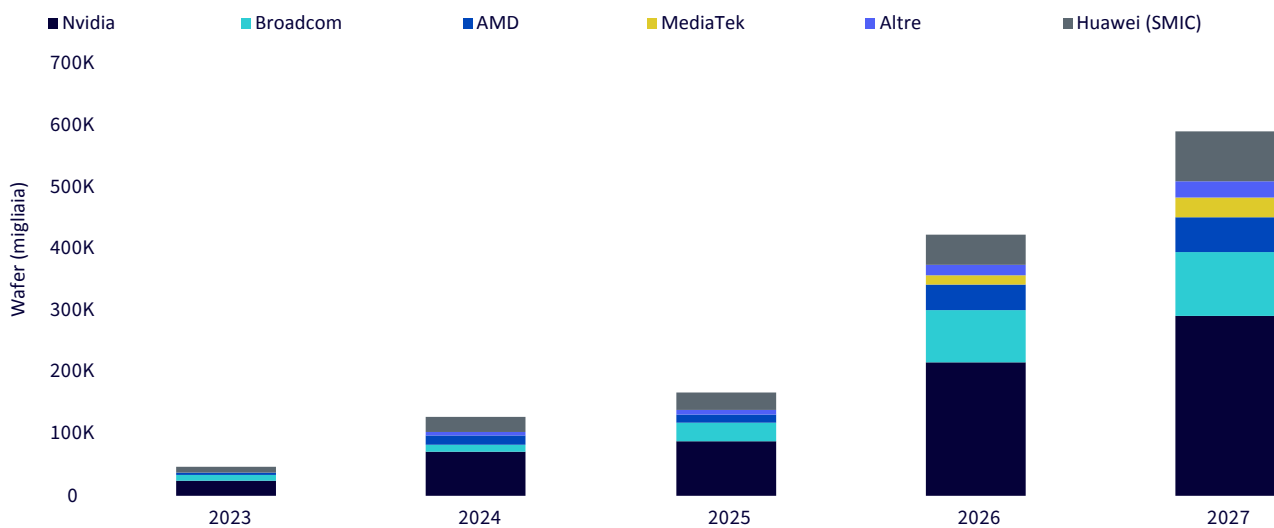
9 Litografia: il processo di trasferimento dei pattern dei circuiti su un wafer di silicio tramite l'uso della luce. In particolare, la litografia a ultravioletti estremi (EUV) è essenziale per la produzione dei chip di IA più avanzati e rimane una delle fasi più dispendiose in termini di capitale nella fabbricazione dei semiconduttori.

10 Una camera bianca è un ambiente utilizzato nella produzione di semiconduttori con livelli estremamente bassi di particelle sospese nell'aria, fluttuazioni di temperatura, variazioni di umidità e altre forme di contaminazione o instabilità ambientale. Piccole particelle contaminanti o deviazioni dalle condizioni ideali possono rovinare un wafer semiconduttore o persino un intero lotto di chip.

Allo stesso tempo, i chip stessi stanno diventando più complessi. Per stare al passo con le crescenti esigenze di calcolo, ogni nuova generazione di acceleratori di IA richiede più memoria e una maggiore larghezza di banda, il tutto entro vincoli di spazio e potenza simili.

Questa crescente complessità significa che i singoli componenti non possono più essere semplicemente affiancati su un tradizionale package per chip. Devono invece essere strettamente integrati utilizzando tecniche di packaging avanzate come il CoWoS (Chip-on-Wafer-on-Substrate), che impila e collega più chiplet in un'unica unità. Di conseguenza, il packaging avanzato è diventato uno degli ostacoli più urgenti nella catena di approvvigionamento.

Figura 4: Spedizioni di wafer CoWoS¹¹ per azienda



Fonte: SemiAnalysis WFE Model, febbraio 2026. Le previsioni non sono indicative della performance futura e qualsiasi investimento è soggetto a rischi e incertezze. La performance storica non è indicativa di quella futura e qualsiasi investimento può diminuire di valore.

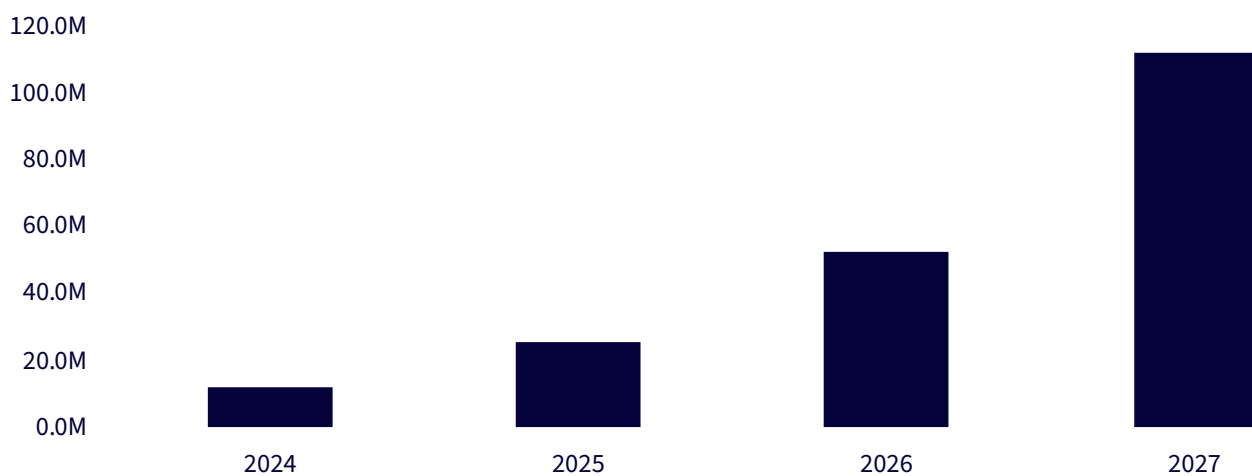
Si prevede che le spedizioni di wafer CoWoS aumenteranno di oltre dieci volte tra il 2023 e il 2027, ma la domanda continua a superare l'offerta. I progettisti di chip si stanno ora accaparrando la capacità di packaging con anni di anticipo. Le aziende dotate di queste competenze specializzate si trovano in una posizione strategica, poiché poter accedere a un packaging avanzato significa anche riuscire a produrre chip più avanzati.

11 CoWoS: Chip-on-Wafer-on-Substrate. Una tecnologia avanzata di packaging dei semiconduttori sviluppata da TSMC che consente di integrare più chip su un unico substrato, fondamentale per gli acceleratori di IA ad alte prestazioni.

Limiti a livello di rete e di sistema

Gli acceleratori non sono gli unici chip che contano. Con l'espansione dei sistemi di IA, è necessario che migliaia di processori operino in modo coordinato e l'infrastruttura di rete che li collega diventa altrettanto fondamentale. I dati devono spostarsi tra i chip a velocità e volumi enormi, rendendo i chip di collegamento, i ricetrasmittitori ottici e le apparecchiature di commutazione componenti essenziali del sistema. Se la rete non riesce a tenere il passo, nemmeno gli acceleratori di IA più potenti possono funzionare.

Figura 5: Domanda di ricetrasmittitori per l'IA



Fonte: SemiAnalysis, aprile 2026. Le previsioni non sono indicative della performance futura e qualsiasi investimento è soggetto a rischi e incertezze. La performance storica non è indicativa di quella futura e qualsiasi investimento può diminuire di valore.

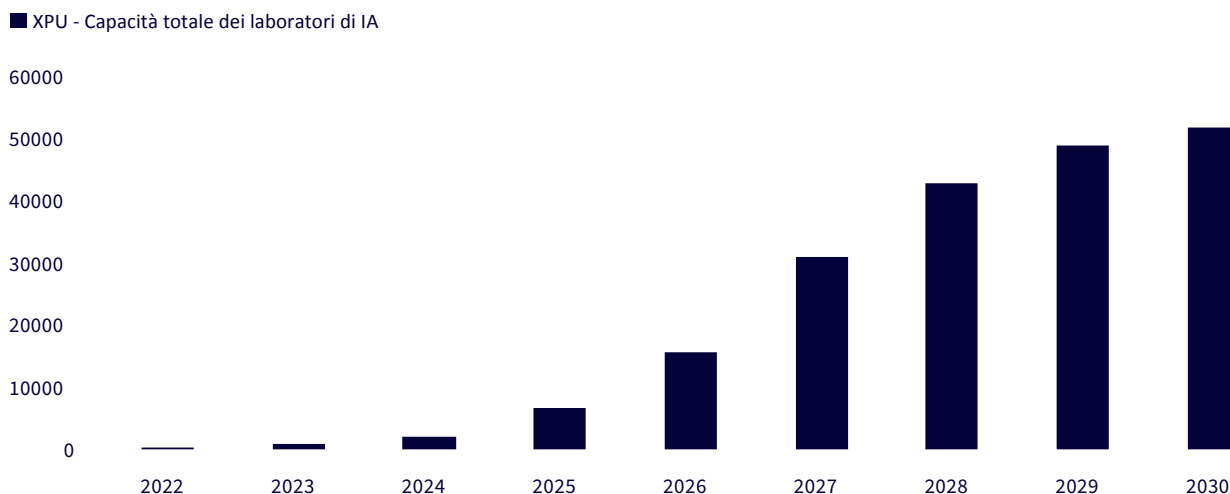
Spesso, a collegare questi processori ad alta velocità all'interno dei centri dati sono ricetrasmittitori ottici da 800 G¹². Si prevede che la domanda di tali componenti passerà dai 25 milioni di unità del 2025 a 112 milioni entro il 2027, con una rapida espansione anche dei collegamenti di nuova generazione da 1,6 T. Questa crescita riflette la sempre maggiore importanza delle reti come fattore determinante per le prestazioni e i costi complessivi dei sistemi.

Fabbisogno energetico ed espansione dei centri dati

Anche quando i chip e le infrastrutture di rete sono disponibili, occorre uno spazio dove farli funzionare. Per fornire i token agli utenti finali, ogni acceleratore richiede alimentazione, raffreddamento e spazio fisico nel centro dati. La portata di questa domanda diventa tangibile se vista attraverso la lente del consumo energetico, poiché i carichi di lavoro legati all'inferenza dell'IA determinano un aumento sostenuto della domanda di energia.

12 800 G: si riferisce ai ricetrasmittitori ottici da 800 Gigabit al secondo, l'attuale standard ad alta velocità utilizzato nelle reti dei centri dati per l'IA per trasferire i dati tra server e switch. I ricetrasmittitori di nuova generazione da 1,6 T (1.600 Gbps) stanno iniziando a essere distribuiti per cluster di IA che richiedono una larghezza di banda maggiore.

Figura 6: Capacità XPU dei fornitori di LLM (MW¹³)



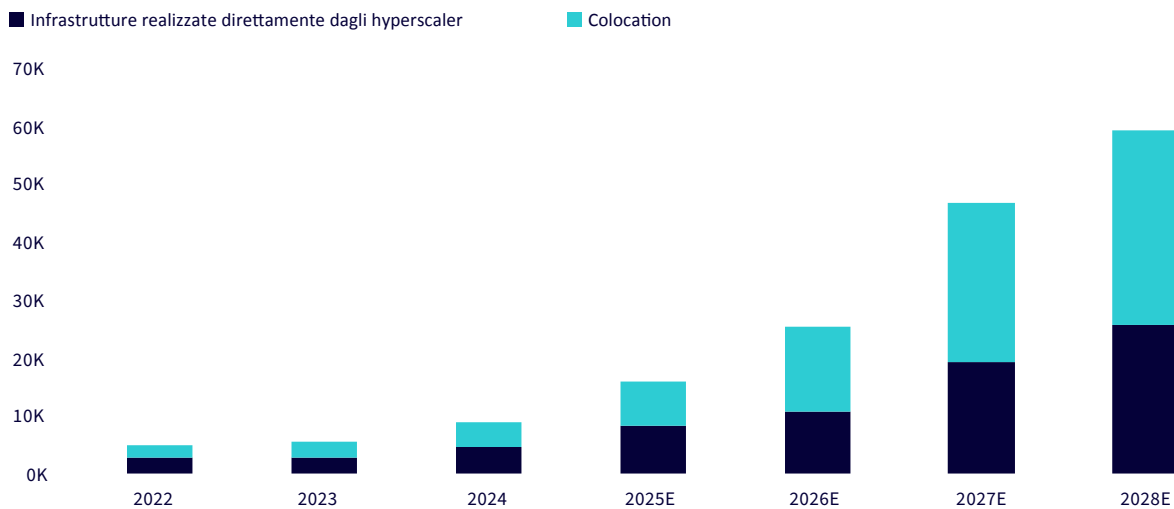
Fonte: SemiAnalysis AI Datacentre Industry Model, aprile 2026. Le previsioni non sono indicative della performance futura e qualsiasi investimento è soggetto a rischi e incertezze. La performance storica non è indicativa di quella futura e qualsiasi investimento può diminuire di valore.

Secondo le previsioni, entro il 2030 la domanda di energia per l'inferenza supererà abbondantemente i 50.000 MW. A tali livelli, l'IA non rappresenta più un carico marginale per le infrastrutture energetiche tradizionali. Diventa invece uno dei principali motori della domanda di elettricità a livello di sistema e, di conseguenza, i sistemi stessi si comportano sempre più come processi industriali. Lo dimostra il termine “fabbriche di IA”, utilizzato dall'AD di NVIDIA, Jensen Huang, per descrivere i moderni centri dati dedicati all'IA.

I crescenti requisiti di potenza di calcolo ed energia stanno imponendo una rapida espansione della capacità dei centri dati. L'unico modo per soddisfare questa domanda è mettere in funzione un numero maggiore di fabbriche di IA.

13 MW = megawatt.

Figura 7: Incremento globale di MW dei centri dati (esclusa la Cina)



Fonte: SemiAnalysis AI Datacentre Industry Model, aprile 2026. Le previsioni non sono indicative della performance futura e qualsiasi investimento è soggetto a rischi e incertezze. La performance storica non è indicativa di quella futura e qualsiasi investimento può diminuire di valore.

Si prevede che la capacità aggiuntiva dei centri dati a livello globale passerà dai circa 16.000 MW del 2025 a quasi 60.000 MW entro il 2028. Gli hyperscaler stanno costruendo le proprie strutture e collaborando con i neocloud e altri operatori di centri dati per attivare la capacità il più rapidamente possibile.

Tuttavia, le infrastrutture energetiche e di rete hanno tempi di realizzazione pluriennali e si stanno rivelando il limite determinante per la velocità con cui è possibile implementare nuova capacità. Per ovviare a questo problema, gli operatori stanno ricorrendo sempre più alla generazione in loco, con turbine a gas, sistemi basati su motori e celle a combustibile, privilegiando la velocità rispetto alla tradizionale dipendenza dalla rete. In definitiva, la capacità di fornire IA su larga scala è direttamente legata alla disponibilità di infrastrutture fisiche, e i centri dati e l'energia elettrica saranno probabilmente uno dei principali colli di bottiglia del prossimo decennio.

Il ciclo di investimenti nelle infrastrutture per l'IA è in corso

L'IA si sta evolvendo in un sistema fisico e l'infrastruttura necessaria per supportarla abbraccia le catene di approvvigionamento di semiconduttori e produzione, reti, centri dati e sistemi di alimentazione. Con l'introduzione dell'IA nei processi produttivi in tutto il mondo, la domanda sta accelerando, mentre le limitazioni a ogni livello dello stack stanno diventando più evidenti.

L'IA sta guidando uno dei più grandi cicli di investimenti in capitali nella storia del settore tecnologico. La domanda sta aumentando, ma il ritmo di distribuzione non è determinato solo da quest'ultima, ma anche dalla velocità con cui è possibile costruire le infrastrutture fisiche. Questo prolunga la durata del ciclo e rende le infrastrutture per l'IA un tema di investimento determinante per il prossimo decennio.

Informazioni importanti

Comunicazioni di marketing emesse all'interno dello Spazio economico europeo ("SEE"):

Il presente documento è stato emesso e approvato da WisdomTree Ireland Limited, società autorizzata e regolamentata dalla Central Bank of Ireland.

Comunicazioni di marketing emesse in giurisdizioni non appartenenti al SEE: Il presente documento è stato emesso e approvato da WisdomTree UK Limited, società autorizzata e regolamentata dalla Financial Conduct Authority del Regno Unito.

Per fare riferimento a WisdomTree Ireland Limited e a WisdomTree UK Limited si utilizza per entrambe la denominazione "WisdomTree" (come applicabile). La nostra politica sui conflitti d'interesse e il nostro inventario sono disponibili su richiesta.

Le informazioni contenute nel presente documento sono fornite a titolo meramente informativo e non costituiscono né un'offerta di vendita né una sollecitazione di un'offerta di acquisto di titoli o azioni. Il presente documento non deve essere utilizzato come base per una qualsiasi decisione d'investimento. Gli investimenti possono aumentare o diminuire di valore e si può perdere una parte o la totalità dell'importo investito. Le performance passate non sono necessariamente indicative di performance future. Qualsiasi decisione d'investimento deve essere basata sulle informazioni contenute nel Prospetto informativo di riferimento e deve essere presa dopo aver richiesto il parere di un consulente d'investimento, fiscale e legale indipendente.

L'applicazione di regolamenti e leggi fiscali può spesso portare a una serie di interpretazioni diverse. Eventuali punti di vista o opinioni espresse in questa comunicazione rappresentano le opinioni di WisdomTree e non devono essere interpretate come consulenza normativa, fiscale o legale. WisdomTree non fornisce alcuna garanzia o dichiarazione circa l'accuratezza di qualsiasi punto di vista o opinione espressa in questa comunicazione. Qualsiasi decisione di investimento dovrebbe essere basata sulle informazioni contenute nel prospetto appropriato e dopo aver richiesto una consulenza finanziaria, fiscale e legale indipendente.

Il presente documento non è, e in nessun caso deve essere interpretato come, una pubblicità o qualsiasi altro strumento di promozione di un'offerta pubblica di azioni o titoli negli Stati Uniti o in qualsiasi provincia o territorio degli Stati Uniti. Né il presente documento né alcuna copia dello stesso devono essere acquisiti, trasmessi o distribuiti (direttamente o indirettamente) negli Stati Uniti.

Benché WisdomTree si adoperi per garantire l'esattezza del contenuto del presente documento, WisdomTree non garantisce né assicura la sua esattezza o correttezza. Qualsiasi terzo fornitore di dati di cui ci si avvalga per reperire le informazioni contenute nel presente documento non rilascia alcuna garanzia o dichiarazione di sorta in relazione ai suddetti dati. Laddove WisdomTree abbia espresso dei pareri relativamente al prodotto o all'attività di mercato, si ricorda che tali pareri possono cambiare. Né WisdomTree, né alcuna consociata, né alcuno dei rispettivi funzionari, amministratori, partner o dipendenti, accetta alcuna responsabilità per qualsiasi perdita, diretta o indiretta, derivante dall'utilizzo del presente documento o del suo contenuto.



WisdomTree.eu
+44 (0) 207 448 4330