Il vento di DeepSeek riuscirà a gonfiare le vele delle aziende di software cloud?

Pubblicato il 12 febbraio 2025

Baoqi Zhu
Associate Director, Quantitative Research & Multi Asset Solutions

Elvira Kuramshina
Associate Director, Quantitative Research

Principali insegnamenti

- Il mix di esperti e l'apprendimento per rinforzo (RL) puro di DeepSeek migliorano l'efficienza, riducendo i costi di calcolo senza compromettere la portata.
- L'"effetto pesce gatto" costringe i fornitori di intelligenza artificiale (IA) a ridurre i prezzi, rendendo l'IA più accessibile per le aziende di Software-as-a-Service (SaaS).
- L'integrazione di DeepSeek R1 nelle principali piattaforme cloud riduce i timori legati alla privacy e alle normative, accelerando l'adozione dell'IA da parte delle aziende di software.
- Le aziende di software cloud sono pronte a sfruttare i progressi dell'IA attraverso la differenziazione dei prodotti, l'efficienza dei costi e nuovi flussi di entrate.
- Il passaggio all'IA agenziale e all'IA verticale sta ridisegnando il software cloud, consentendo sistemi autonomi e applicazioni di IA specifiche per ogni settore, che offrono un vantaggio competitivo.
- Prodotti correlati WisdomTree Cloud Computing UCITS ETF - USD Acc, WisdomTree Megatrends UCITS ETF - USD Acc Scopri di più

DeepSeek fa ormai notizia dalla vigilia dell'Anno del Serpente (24/01/2025) e i mercati finanziari hanno reagito in modo forte e contrastante. Nel frattempo, sono sorte anche discussioni sulla riservatezza dei dati, la proprietà intellettuale e le tensioni geopolitiche. Tuttavia, vorremmo fare un passo indietro rispetto a questi dibattiti più ampi e concentrarci sui progressi tecnici dei modelli di DeepSeek e sul loro potenziale impatto sul settore del software.

Quali innovazioni ha sviluppato DeepSeek?

Il modello di intelligenza artificiale (IA) di DeepSeek si basa sull'architettura Transformer, ma integra diverse ottimizzazioni nel design e nella progettazione a vantaggio dell'efficienza. Piuttosto che introdurre metodi completamente nuovi, DeepSeek combina le innovazioni esistenti e affronta sfide pratiche, come la limitata potenza di calcolo, per trovare un equilibrio tra progressi teorici e un'esecuzione ingegneristica solida.

In particolare, riteniamo che i modelli DeepSeek V3 e DeepSeek R1 Zero presentino due innovazioni fondamentali:

mix di esperti (MoE) e strategia senza perdite ausiliarie - DeepSeek-V3 divide il modello in più moduli "esperti" per distribuire il carico di lavoro in modo efficiente. Durante la formazione, i diversi esperti vengono eseguiti su GPU diverse, migliorando l'efficienza. Nell'inferenza, vengono utilizzati solo 37 miliardi di parametri attivi per token invece dei 671 miliardi totali, riducendo in modo significativo i costi di calcolo. Una sfida comune con il MoE è lo sbilanciamento del carico di lavoro tra i moduli esperti; DeepSeek affronta questo problema utilizzando un instradamento basato sulla rilevanza invece di funzioni di perdita ausiliaria, garantendo la conservazione di tutti i token durante la formazione e l'inferenza.

Apprendimento per rinforzo senza feedback umano - DeepSeek-R1 Zero è un modello di ragionamento come GPT-o1 e dimostra che l'apprendimento per rinforzo (RL) puro può sviluppare capacità di ragionamento senza affidarsi all'apprendimento per rinforzo da feedback umano (RLHF). La maggior parte dei modelli linguistici di grandi dimensioni (LLM) prevede l'etichettatura e il feedback umani tramite fine-tuning supervisionato (SFT) e RLHF. Tuttavia, DeepSeek R1 Zero sostituisce SFT[1] e RLHF con un RL puro senza feedback umano.

L'apprendimento per rinforzo tradizionale ha un costo computazionale elevato e una scalabilità limitata rispetto a diversi compiti. DeepSeek R1 Zero affronta queste sfide introducendo un metodo innovativo chiamato Group Relative Policy Optimisation (GRPO); inoltre, i progressi di DeepSeek indicano che un metodo RL puro potrebbe essere un'alternativa per comprendere come migliorare continuamente i modelli di intelligenza artificiale, dato che i dati di alta qualità generati dall'uomo sono limitati e potrebbero esaurirsi. Questo aumenta la sostenibilità dei miglioramenti delle capacità dei modelli.

L'impatto sulle aziende di software

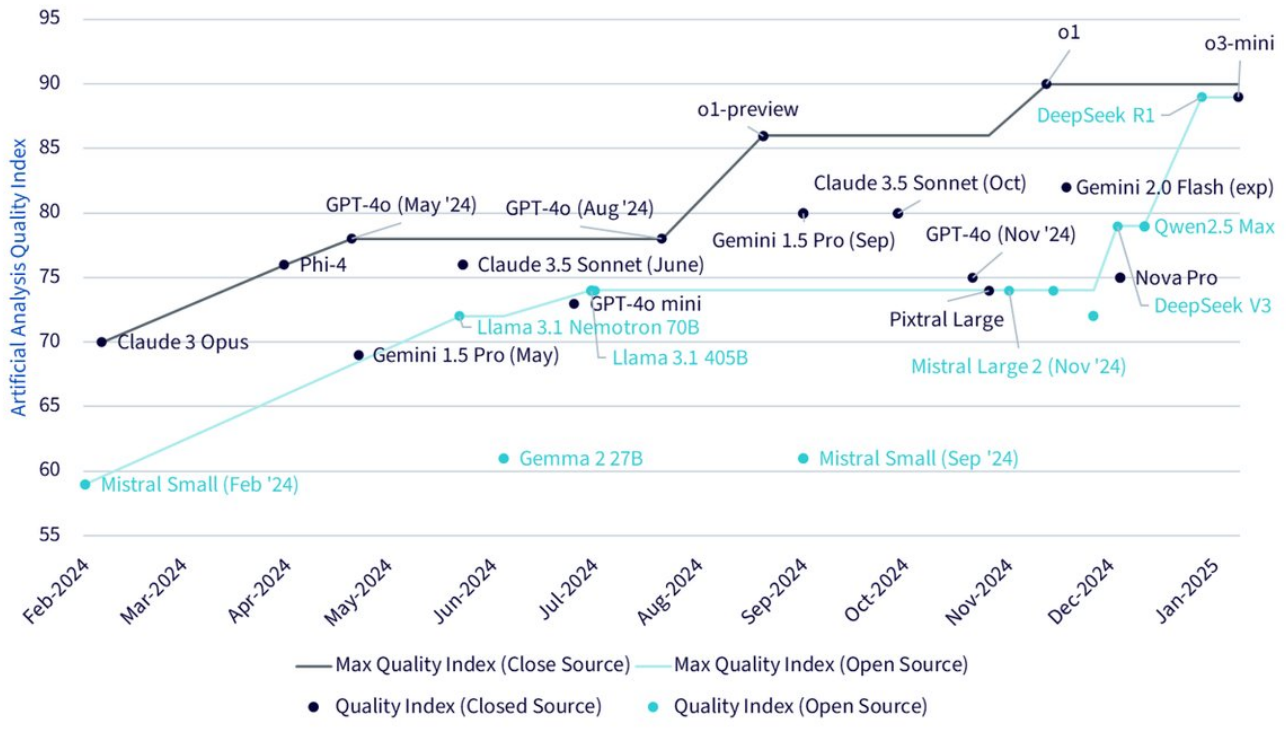
Dal lancio di ChatGPT, gli investitori nel settore del software si trovano ad affrontare una domanda impegnativa: quanto valore aggiunge l'IA alle aziende SaaS? Gli sviluppi dell'IA hanno sollevato preoccupazioni sulla sostenibilità della crescita del software e, in particolare, sulla possibilità che l'intelligenza artificiale possa sostituire il software tradizionale. Inoltre, l'integrazione dei modelli di IA nelle offerte SaaS ha introdotto nuovi costi di utilizzo, suscitando dubbi in merito alla potenziale pressione sui margini. Tuttavia, i recenti sviluppi suggeriscono una prospettiva più favorevole, attenuando alcune di queste preoccupazioni.

Un elemento chiave di tale cambiamento è rappresentato da DeepSeek, i cui modelli open-source sono in grado di competere con le prestazioni dei principali modelli linguistici di grandi dimensioni, risultando però più economici da utilizzare. Questa convenienza riduce gli ostacoli che le aziende SaaS incontrano nell'aggiungere funzionalità di IA, consentendo loro di integrarla nelle aziende con spese relativamente inferiori. Inoltre, le principali piattaforme cloud, tra cui Microsoft Azure, AWS e Google Cloud, hanno già aggiunto DeepSeek R1 ai loro servizi, risolvendo così le preoccupazioni relative alla privacy e alle

normative, grazie alla distribuzione al di fuori della Cina. Diverse aziende statunitensi hanno già adottato DeepSeek. Perplexity, nota per i suoi servizi di risposta alle domande basati sull'intelligenza artificiale, ha integrato il modello DeepSeek R1 nel suo motore di ricerca. Anche Cerebras e Groq si sono unite a questa tendenza. Nel frattempo, ZoomInfo si è detta interessata ad acquistare l'accesso a DeepSeek R1 attraverso le piattaforme cloud statunitensi[2].

Il prezzo competitivo di DeepSeek potrebbe innescare un “effetto pesce gatto”, spingendo altri fornitori a presentare modelli più economici e convenienti. Ad esempio, OpenAI ha già lanciato GPT-o3 mini a un prezzo molto basso, con un costo di produzione di soli 4,4 dollari per 1 milione di token, ovvero circa il 63% in meno rispetto al prezzo di GPT-o1 mini. Tale concorrenza favorisce le aziende di software, consentendo loro di adottare l'IA senza erodere i margini di profitto. Inoltre, DeepSeek ha pubblicato i suoi metodi di ottimizzazione, favorendo le comunità open-source e riducendo ulteriormente il divario tra i modelli open-source e quelli closed-source. Nel tempo, quest’apertura dovrebbe promuovere la concorrenza e mantenere i costi dell'IA a livelli gestibili.

Figura 1: Il divario di prestazioni tra i modelli closed-source e quelli open-source si sta riducendo



Fonte: Artificial Analysis. L'Artificial Analysis Quality Index fa riferimento al risultato medio delle nostre valutazioni, che coprono diverse dimensioni dell'intelligenza dei modelli. Attualmente include MMLU, GPQA, Math e HumanEval. I dati del modello OpenAI o1 sono preliminari e si basano sui dati dichiarati

da OpenAI. Dati aggiornati al 31 gennaio 2025. La performance storica non è indicativa di quella futura e qualsiasi investimento può diminuire di valore.

Figura 2: Le API di DeepSeek sono relativamente convenienti rispetto ai principali modelli di IA

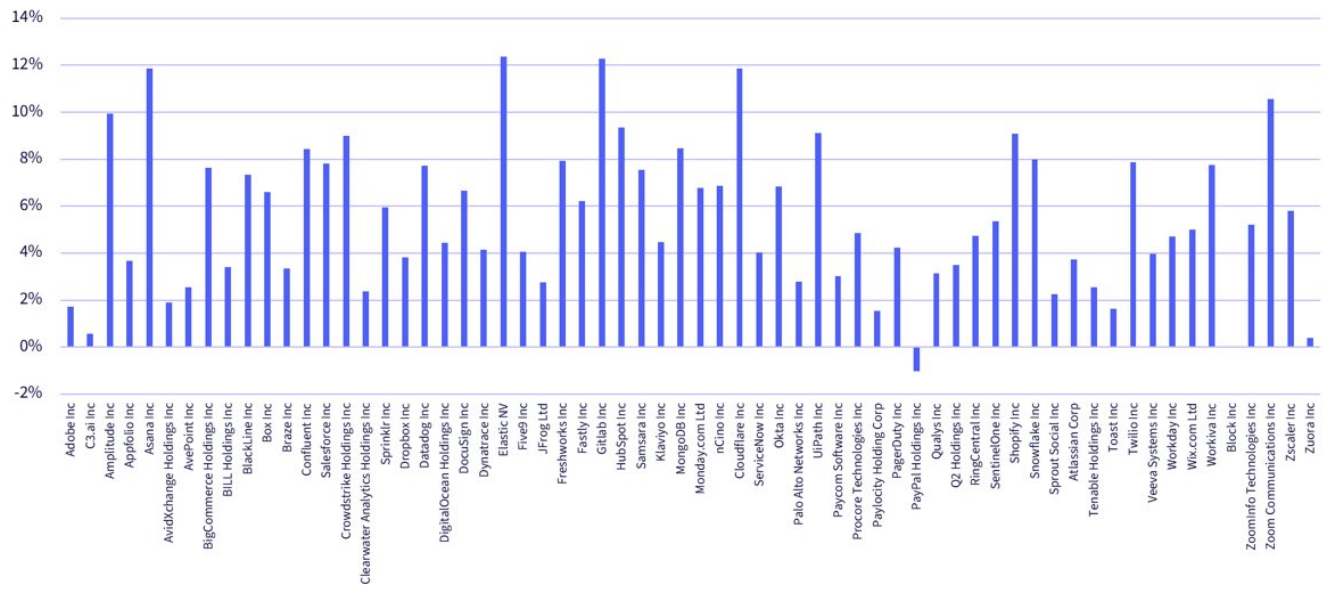
Company	Model	Input Cost Per 1M Tokens ($)	Output Cost Per 1M Tokens ($)	Artificial Analysis Quality Index
DeepSeek	DeepSeek R1	0.55	2.19	89.00
DeepSeek	DeepSeek V3	0.27	1.10	79.00
OpenAI	GPT-o1	15.00	60.00	90.00
OpenAI	GPT-o3 mini	1.10	4.40	89.00
OpenAI	GPT-o1 mini	3.00	12.00	84.00
OpenAI	GPT-4o	2.50	10.00	75.00
Anthropic	Claude 3.5 Sonnet	3.00	15.00	80.00
Google	Gemini 2.0	NA	NA	82.00
Google	Gemini 1.5 Pro	1.25	5.00	80.00

Fonte: Bloomberg, DeepSeek, OpenAI, Anthropic, Google, Artificial Analysis. I prezzi delle API sono forniti dai proprietari dei modelli e possono differire dai prezzi offerti da fornitori terzi. L'Artificial Analysis Quality Index fa riferimento ai risultati medi delle nostre valutazioni, che coprono diverse dimensioni dell'intelligenza dei modelli. Attualmente include MMLU, GPQA, Math e HumanEval. I dati del modello OpenAI o1 sono preliminari e si basano sui dati dichiarati da OpenAI. Dati aggiornati al 31 gennaio 2025. **La performance storica non è indicativa di quella futura e qualsiasi investimento può diminuire di valore.**

Il software cloud può cavalcare l'onda di DeepSeek?

Il BVP Nasdaq Emerging Cloud Index seleziona opportunità pure-play nel panorama del cloud computing, concentrandosi su quelle che superano il mercato più ampio in termini di crescita dei ricavi. Utilizzando questo indice come proxy del sentiment degli investitori, abbiamo valutato le reazioni del mercato di fronte al lancio di DeepSeek il 27-28 gennaio. Escludendo PayPal, tutte le società hanno reagito positivamente, con un rendimento mediano nel periodo del 4,9% (cfr. Figura 3). Dato che in questo lasso di tempo non sono state pubblicate relazioni sugli utili, attribuiamo tale reazione principalmente all'annuncio di DeepSeek. Inoltre, la riunione del Federal Open Market Committee (FOMC) del 28-29 gennaio ha attenuato il sentiment verso la fine della settimana, facendo scendere il rendimento mediano a cinque giorni al 2,7%.

Figura 3: Rendimenti delle società di cloud computing dopo il lancio di DeepSeek



Negli ultimi due anni, si riteneva che le aziende di software cloud non beneficiassero direttamente dell'IA, a differenza degli operatori di infrastrutture e degli sviluppatori di modelli di intelligenza artificiale. Il loro ruolo nella rivoluzione dell'IA sembrava essere a valle, data la necessità di integrare la stessa nei propri prodotti prima di trarne valore. Inoltre, sono emerse preoccupazioni circa l'impatto dei costi legati all'intelligenza artificiale sui margini, soprattutto nel contesto di un potere di determinazione dei prezzi ridotto a causa dei venti macroeconomici e di budget per il software limitati.

L'era della maggiore accessibilità dei modelli di IA, inaugurata dal lancio di DeepSeek, rappresenta un potenziale punto di svolta e offre alle aziende di software cloud opportunità che vanno a vantaggio della diversificazione dei prodotti, dell'efficienza dei costi e di nuovi flussi di entrate. Le recenti tendenze suggeriscono un cambiamento strategico nello spazio del software cloud, con le aziende che si concentrano sempre più sull'IA agenziale (sistemi autonomi che eseguono attività con un intervento umano minimo) e sull'IA verticale (modelli di IA altamente specializzati in un settore o un dominio specifico, ad esempio quello sanitario, legale, finanziario). L'emergere di aziende cloud native dell'IA sottolinea ulteriormente quest’evoluzione, mettendo i fornitori di cloud in condizione di trarre un beneficio più diretto dai progressi della stessa.

Il lancio di DeepSeek, che ha scatenato una nuova ondata di innovazioni nel settore dell'intelligenza artificiale, è destinato ad accelerare l'adozione dell'IA. Man mano che questo si concretizza, le aziende aumenteranno la spesa per il software a fronte di una domanda crescente per sfruttare la tecnologia dell'IA in tutti i settori. Nel breve periodo, una politica monetaria più rigida negli Stati Uniti rappresenta un vento contrario per i fornitori di software cloud, ma i comunicati sugli utili di febbraio e inizio marzo potrebbero offrire una panoramica sui piani delle aziende per capitalizzare la trasformazione del settore guidata dall'IA.

[1] La versione successiva, DeepSeek-R1, applicava l’SFT alla formazione del modello, ma con l'obiettivo di migliorare la leggibilità e la coerenza delle sue risposte. Le capacità di ragionamento e di prendere decisioni del modello sono state perfezionate, principalmente attraverso l'apprendimento per rinforzo (RL).
[2] The Information: https://www.theinformation.com/articles/deepseek-attracts-surge-of-business-users.

Important Risks Related to this Article

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.

Solo per clienti professionali. 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.