Large Language Models: dall’astrazione alla realtà

Pubblicato il 8 ottobre 2024

Christopher Gannatti, CFA
Global Head of Research

Principali insegnamenti

- I Large Language Models (LLM) offrono un impatto reale, permettendo ad aziende come Amazon di risparmiare molti anni di ricerca e sviluppo.
- I modelli di fondazione servono come strato infrastrutturale critico per le future innovazioni dell'IA.
- Gli LLM migliorano la produttività e l'efficienza nello sviluppo di software e nelle operazioni aziendali.
- Il potenziale economico degli LLM si espanderà, dal momento che le principali aziende continueranno a investire fortemente in infrastrutture e talento.
- L'evoluzione dei modelli di IA può portare a interfacce più specializzate e adatte a compiti specifici.
- Prodotti correlati WisdomTree Artificial Intelligence UCITS ETF – USD Acc Scopri di più

Il problema dei Large Language Models (LLM) è che per un ingegnere non informatico è difficile persino immaginare di cosa si tratti. Ciò è reso ancora più difficile se si considera che lo sviluppo e l'addestramento degli LLM di maggiori dimensioni costa centinaia di milioni di dollari.

Eppure, gli amministratori delegati di alcune delle più grandi aziende mondiali hanno dichiarato che nei prossimi anni destineranno, come gruppo, più di 1.000 miliardi di dollari alla costruzione di ulteriori infrastrutture informatiche che permettano a un maggior numero di questi modelli di funzionare in modo più efficiente[1,2].

La nostra attenzione è sempre rivolta a casi d'uso o testimonianze che ci permettano di tradurre l'astrazione di qualcosa come un LLM in un’esperienza di vero e proprio impatto aziendale. Se trovassimo un numero sufficiente di questi casi, potremmo iniziare a osservare i vari impatti nelle statistiche generali sulla produttività economica.

Siamo quindi rimasti entusiasti nel constatare quanto segue, dichiarato da Andy Jassy, CEO di Amazon, in riferimento al sistema Q di Amazon, che è essenzialmente un LLM in grado di generare codice software[3]:

Il tempo medio per aggiornare un'applicazione a Java 17 è precipitato, passando dai tipici 50 giorni di programmazione a poche ore. Secondo le nostre stime, ciò ci consente di risparmiare l'equivalente di 4.500 anni di lavoro nell’ambito dello sviluppo (sì, è un numero assurdo, ma reale).

In meno di sei mesi, siamo stati in grado di aggiornare più del 50% dei nostri sistemi produttivi Java alle versioni più avanzate, in una frazione del tempo e delle risorse normalmente necessari. Inoltre, i nostri sviluppatori hanno consegnato revisioni del codice generate automaticamente senza ulteriori modifiche nel 79% dei casi.

I vantaggi vanno al di là degli sforzi che abbiamo risparmiato agli sviluppatori. Gli aggiornamenti hanno ottimizzato la sicurezza e ridotto i costi dell'infrastruttura, con un guadagno di efficienza annuo stimato in 260 milioni di dollari.

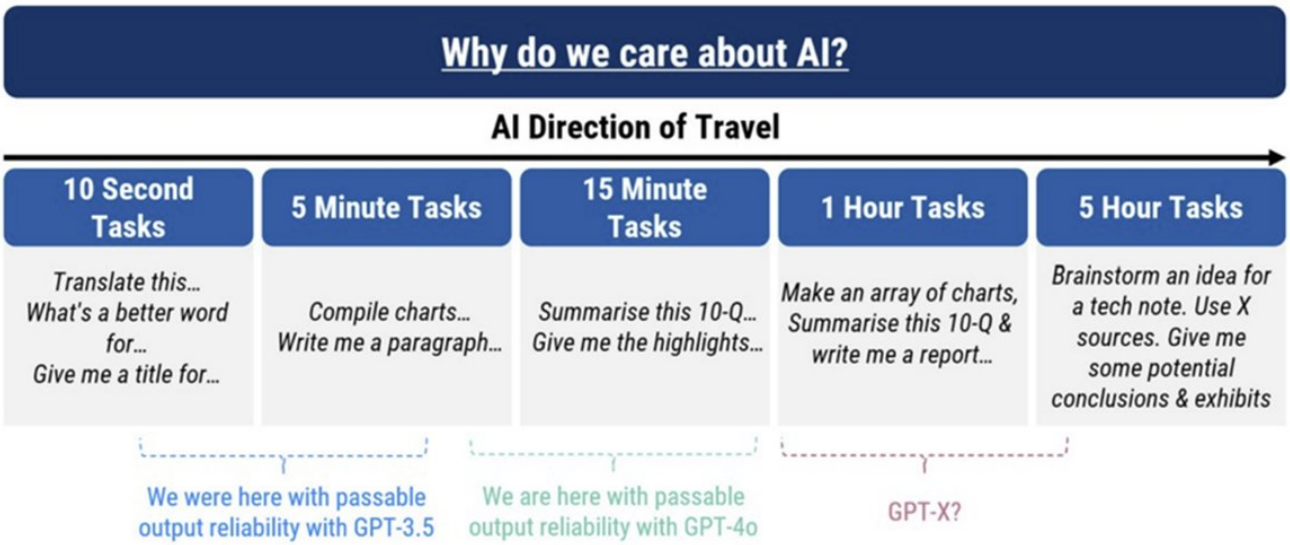
4.500 anni di lavoro nell'ambito dello sviluppo??? 260 milioni di dollari in rendimenti annualizzati in forza dell'efficienza??? Sono numeri importanti. Riconosciamo di essere ancora all'inizio del viaggio della rivoluzione IA, ma forse la concettualizzazione di Jassy riguardante l'impatto dell'IA sugli sviluppatori del suo team sarà d'ispirazione ad altri, così da raccontare esperienze simili e renderle pubbliche.

Tabella di marcia concettuale sullo sviluppo dell'IA[4]

La Figura 1, a nostro avviso, contribuisce alla rapida individuazione di un altro modo di interpretare il senso di tutti questi LLM. Tutti i lavoratori della conoscenza sono consapevoli di come il loro lavoro consista in diversi bucket di compiti e di come ogni bucket possa comportare tempistiche molto diverse.

Le attuali versioni di LLM sono in grado di rispondere a semplici domande o a semplici e-mail, ma non è affatto chiaro come questi sistemi possano formulare, da zero, relazioni completamente nuove o idee originali. Questo non significa che non possano farlo, ma semplicemente che il livello di analisi richiesto per le risposte a semplici domande è molto diverso dal livello di analisi richiesto per una presentazione totalmente inedita, che potrebbe contenere 60 diapositive originali, tutte sviluppate dall'intelligenza artificiale.

Figura 1: Tabella di marcia concettuale sullo sviluppo dell'IA



Fonte: Stanley, Edward et al. "Mapping AI's Rate of Change". Morgan Stanley Research. 4 giugno 2024.

Gli LLM sono software, infrastrutture o entrambe le cose?

Gli LLM sono definiti talvolta come “modelli di fondazione”. Il termine “fondazione” fa riflettere, dal momento che, in molti contesti, una fondazione rappresenta qualcosa su cui si può costruire. Si pensi alla creazione di valore in alcuni ecosistemi:

- Perché i computer hanno un valore? Si potrebbe pensare a tutti i diversi componenti hardware e sommare il valore di ciascuno di essi per determinare il valore di un computer. Tuttavia, se riflettiamo sulla storia della diffusione dei computer, abbiamo assistito a un periodo in cui si pensava che gli utenti volessero assemblare da sé il proprio computer. La semplicità di utilizzo che Microsoft apportò all'esperienza di compiere determinate operazioni con il computer ebbe un valore immenso, così come il fatto che molti utenti adottarono questi pacchetti software e impararono a utilizzarli tutti insieme.
- Perché Internet ha un valore? Si potrebbe dire che Internet permette di inviare informazioni in tutto il mondo con molta più facilità. Eppure, cosa dire di Google, un esempio di azienda il cui obiettivo è stato quello di organizzare le informazioni disponibili su Internet in modo da ottenere molto più rapidamente ciò di cui si ha bisogno? Quale sarebbe la nostra situazione se dovessimo usare Internet senza i motori di ricerca?
- Se consideriamo l'iPhone, il valore risiede nei componenti del telefono e in ciò che la tecnologia è in grado di fare in modo specifico, oppure il valore risiede in tutte le diverse applicazioni sviluppate sul software iOS?

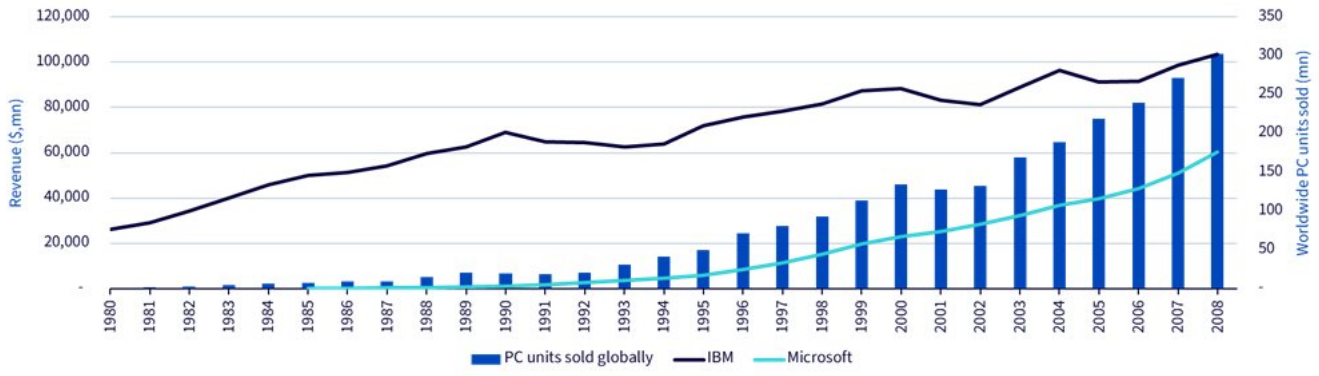
In ognuno di questi casi, la vera risposta è probabilmente che tutto ha un certo valore, ma il motivo per cui spesso ci riferiamo a prodotti e servizi offerti da aziende che superano i 1.000 miliardi di dollari di capitalizzazione di mercato è che diversi sono gli effetti di incidenza, che si moltiplicano in modo esponenziale. Siamo anche in presenza di giganteschi effetti rete: niente genera più utenti e più crescita quanto un'enorme base iniziale di utenti.

Le figure 2a e 2b ci danno la sensazione di avere già assistito a cose simili prima d'ora:

- La Figura 2a mette in relazione il fatturato di IBM e Microsoft con il numero di unità di PC vendute a livello globale. Appare logico che, in un primo momento, IBM venda PC o consenta ad altre aziende di costruire e vendere i cosiddetti cloni IBM. Il numero di PC cresce e dopo un certo periodo, una volta che la base di utenti di PC raggiunge un certo livello, i ricavi di Microsoft decollano, in quanto tutti questi utenti di PC vengono a conoscenza e ricercano i vantaggi dell'utilizzo dei sistemi operativi Microsoft. Non avrebbe senso che il sistema di Microsoft avesse valore prima che i PC siano presenti su quasi tutte le scrivanie.

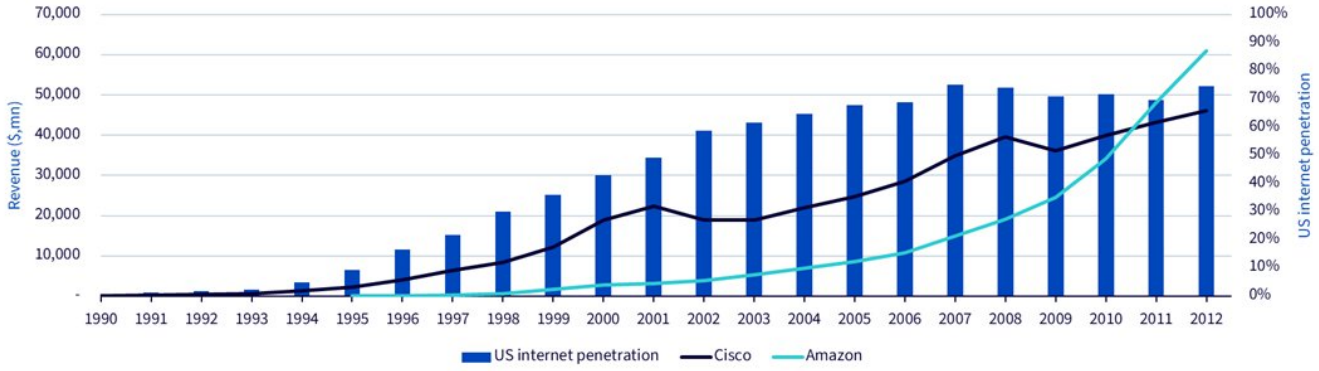
- La Figura 2b mette in relazione la diffusione di Internet negli Stati Uniti con il fatturato di Cisco e Amazon.com. I dispositivi di Cisco consentono a un maggior numero di aziende e individui di utilizzare Internet. Man mano che un numero sempre maggiore di utenti si avvale di Internet, improvvisamente il valore di ciò che Amazon.com sta fornendo - il commercio elettronico - diventa sempre più chiaro. Non avrebbe senso che l'attività di e-commerce di Amazon.com avesse valore prima che una massa critica di utenti adotti Internet.

Figura 2a: Rapporto tra le unità di PC vendute a livello globale e il fatturato di IBM e Microsoft nel tempo



Fonte: WisdomTree, Bloomberg. I dati sulle unità di PC vendute sono tratti da “Total Share: Personal Computer Market Share 1975-2010, Jeremy Reimer e Gartner. **Lo storico dei rendimenti non è indicativo di quelli futuri e gli investimenti potrebbero subire una riduzione di valore.**

Figura 2b: Rapporto tra la diffusione di Internet negli Stati Uniti e il fatturato di Cisco e Amazon.com nel tempo



Fonte: WisdomTree, Bloomberg, Banca Mondiale. **Lo storico dei rendimenti non è indicativo di quelli futuri e gli investimenti potrebbero subire una riduzione di valore.**

Considerazioni finali: Cosa svilupperemo a partire dai modelli di fondazione?

L'aspetto interessante dei modelli di fondazione, almeno nella seconda metà del 2024, è che solo le aziende più grandi e redditizie a livello mondiale hanno le risorse per continuare a svilupparli e farli

avanzare. Sebbene alcune di queste pare facciano parte di società indipendenti, le maggiori aziende mondiali tendono ad assumere partecipazioni finanziarie importanti che consentono di investire in modo appropriato e sempre maggiore nei talenti e nelle infrastrutture di calcolo necessarie.

Non sappiamo esattamente quali saranno i prossimi sviluppi e riconosciamo che si tratta di una domanda da un trilione di dollari. Ma sappiamo che le persone tendono a concentrarsi su attività individuali e discrete e che non sempre hanno bisogno di accedere a un modello in grado di superare tutti i test più importanti che abbiamo sviluppato. A ogni modo, ci piace l'idea di disporre di interfacce più specializzate, basate su compiti più specifici, che possano poi utilizzare parti di modelli più ampi per portare a termine il lavoro.

Comunque si evolva il quadro, riteniamo che le maggiori aziende mondiali collegate a questi modelli di fondazione avranno un ruolo importante da svolgere negli anni a venire.

[1] Goldman Sachs, https://www.datacenterdynamics.com/en/news/goldman-sachs-1tn-to-be-spent-on-ai-data-centers-chips-and-utility-upgrades-with-little-to-show-for-it-so-far/
[2] Nvidia, https://www.linkedin.com/posts/leeps_nvidia-ceo-predicts-1-trillion-will-be-spent-activity-7101349410281836544-hB72/
[3] Fonte: Estratto da un post su Linkedin del CEO di Amazon Andy Jassy, come riportato su https://nextbigteng.substack.com/p/hello-ai-world-evolution-of-developer-economy-in-the-age-of-ai.
[4] Fonte: Stanley, Edward et al. “Mapping AI's Rate of Change”. Morgan Stanley Research. 4 giugno 2024.

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.