

Intelligenza artificiale: nei prossimi tre anni in arrivo oltre 200 miliardi di dollari di investimenti

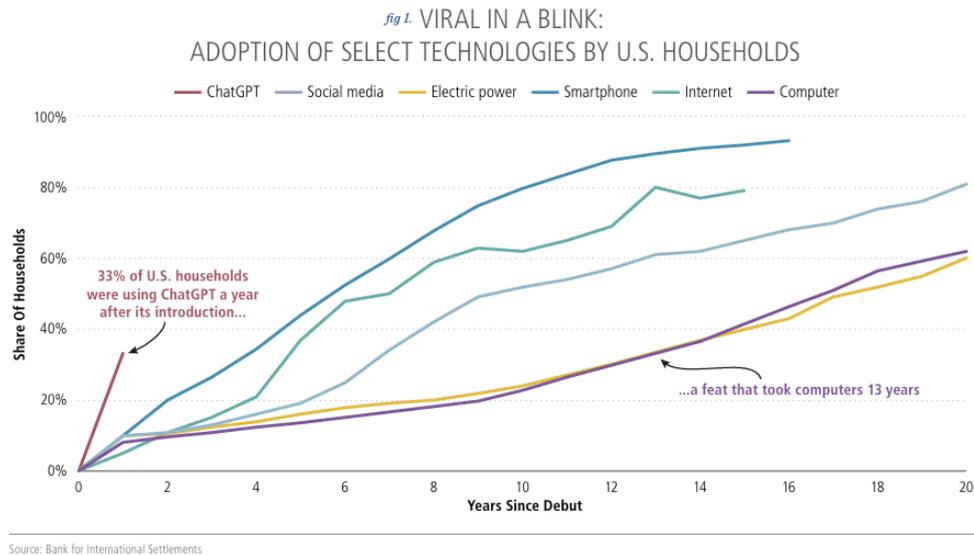
- *L'Intelligenza Artificiale, come tutti i processi industriali, richiede ingenti investimenti in hardware, calcolatori e generatori di energia per progredire*
- *Il settore privato statunitense sta investendo risorse in AI a ritmi a cui non si assisteva dai tempi della corsa allo spazio: negli ultimi 10 anni sono stati spesi ben 335 miliardi di dollari in investimenti, dando vita a oltre 5.500 aziende*
- *Per i prossimi tre anni, gli analisti di Wall Street prevedono oltre 213 miliardi di dollari di investimenti in capitale fisso per l'AI, abbastanza per costruire la capacità equivalente a 12.000 ChatGPT-4 entro il 2026*
- *Gli investimenti privati nella costruzione di data center sono aumentati di 8 miliardi di dollari nel 2023 e di altri 5 miliardi di dollari nei primi sei mesi del 2024, con una crescita cumulativa del 93% in soli due anni*
- *Nonostante gli investimenti elevati, i data center sono ancora in affanno, dal momento che le 9.380 strutture attualmente presenti nel mondo non riescono ancora a immagazzinare più del 15% del totale dei dati creati nel corso del 2023*
- *Con la corsa all'AI, si prevede che la domanda di energia da parte dei data center aumenterà del 160%, arrivando a rappresentare quasi un decimo del consumo energetico mondiale entro il 2030*
- *Generare energia sufficiente per alimentare la rivoluzione AI potrebbe richiedere l'utilizzo di nuove fonti energetiche, incluse le principali fonti rinnovabili e nucleari, e di reti elettriche più moderne, innescando una nuova rivoluzione energetica*
- *Come successo con la nascita di Internet, una crisi dell'AI potrebbe essere inevitabile nel breve termine, eliminando le speculazioni e i modelli di business insostenibili, ma le grandi infrastrutture, insieme ai progressi intellettuali, potrebbero spianare la strada alla prossima ondata di innovazione*

A cura di Jeffrey Cleveland, Chief Economist di Payden & Rygel

21 novembre 2024 – A partire dal debutto di Chat GPT nel novembre 2022, i pronostici sul futuro dell'Intelligenza Artificiale si sono susseguiti senza sosta. Se gli ottimisti sottolineano le sorprendenti potenzialità di un fenomeno in divenire, destinato a crescere ed evolvere con l'arrivo dell'Intelligenza Artificiale Generale (AGI) prima e della Superintelligenza Artificiale (ASI) poi, i pessimisti ne ricordano i fallimenti, convinti come sono che la fase dei chatbot AI sia un'illusione e che, inevitabilmente, si concluderà con l'ennesimo fallimento. Il nostro compito non è ergerci a giudici di questa disputa, ma sottolineare come **l'Intelligenza Artificiale non sia altro che un processo industriale** e che, in quanto tale, richieda ingenti investimenti in hardware, calcolatori e generatori di energia per progredire. Nel frattempo, le aziende continuano a scalare per soddisfare l'attuale domanda di software, facendo scommesse azzardate sull'entità della domanda futura. Ma la "febbre" dell'Intelligenza Artificiale riuscirà a stimolare davvero la crescita economica per il resto del decennio? O l'enorme boom di investimenti finirà per portare il sistema al collasso, come già successo con la bolla di Internet?

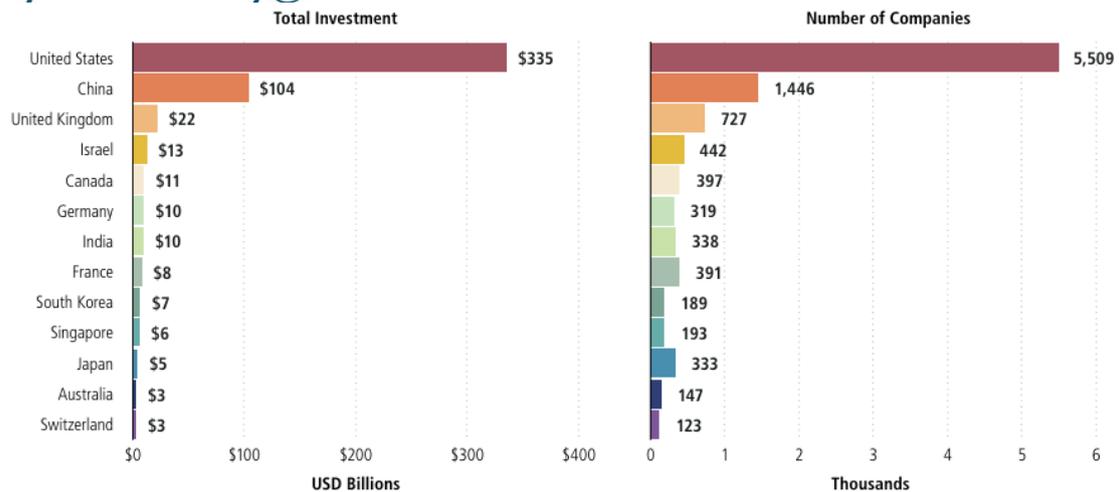
Gran parte dell'Intelligenza Artificiale, Chat GPT inclusa, si basa su Large Language Models, che sono in grado di prevedere una risposta a una specifica domanda da parte di un utente sulla base degli input e dei dati con cui i programmatori li hanno "addestrati". I LLM si sono rivelati estremamente popolari:

la prima versione di ChatGPT, rilasciata nel novembre 2022, nel giro di un anno è stata testata da un terzo delle famiglie americane, un'impresa che ai computer ha richiesto ben tredici anni (vedi Figura 1).



Per quanto molti etichettino i LLM come “semplici macchine” prive di intelligenza che si limitano ad auto completare delle frasi, la verità è che, semplicemente, funzionano. Ad appena un anno dal lancio di GPT-3, il modello è stato perfezionato ed è stato rilasciato GPT-4, mentre Google ha rilasciato Gemini Ultra e Meta ha messo a punto una serie di modelli Llama. **I modelli attuali hanno già iniziato ad avvicinarsi ai benchmark umani nei test MMLU** (Massive Multitask Language Understanding), con performance pari a quelle degli esperti in carne ed ossa. Gran parte del successo si deve alle possibilità di scala: se si hanno abbastanza risorse di calcolo, si può ottenere un risultato utile che inspiegabilmente risponde a ogni possibile domanda in modo sovrapponibile a quello umano. Ovviamente, **la capacità di calcolo richiesta è immensa**: la versione di ChatGPT rilasciata nel 2022 ha richiesto ben 21 milioni di petaFLOPs di calcolo (considerando che un petaFLOP è l'equivalente di un essere umano che aggiunge numeri ogni secondo per 31.688.765 anni). Nessuno può prevedere cosa accadrà in futuro, ma è chiaro che inserire più dati e risorse di calcolo nei LLM li sta migliorando, un semplice fatto di cui aziende e investitori si sono ormai resi conto.

Le aziende statunitensi hanno già speso ben 335 miliardi di dollari in investimenti di AI negli ultimi dieci anni, il 60% dei quali sono stati spesi solo negli ultimi tre anni, dando vita a oltre 5.500 aziende (vedi figura sotto). Per contestualizzare, tra il 1960 e il 1973 lo sbarco sulla Luna è costato complessivamente 280 miliardi di dollari di oggi. In altre parole, il settore privato statunitense sta investendo risorse a ritmi a cui non si assisteva dai tempi della corsa allo spazio. E gli “hyperscaler” – Big Tech come Meta, Google, Amazon e Microsoft – sono soltanto all’inizio: **per i prossimi tre anni, gli analisti di Wall Street prevedono oltre 213 miliardi di dollari di investimenti in capitale fisso per l’AI**, abbastanza per costruire la capacità equivalente a 12.000 ChatGPT-4 entro il 2026. Inoltre, non sono solo le aziende tech a competere per formare i modelli migliori: le earning call per il 1° trimestre 2024 hanno visto 199 società dello S&P 500 menzionare l’AI, contro le sole 60 del 2022. Per far funzionare i modelli è necessario un investimento di capitale: l’Intelligenza Artificiale non fluttua in uno spazio immaginario, ma vive nei data center, il cui funzionamento si serve di semiconduttori (soprattutto GPU, unità di elaborazioni grafiche) per la velocità e la capacità di processazione. Nel corso del 2023, Nvidia ha fornito 3,8 milioni di GPU per data center, un milione in più rispetto all'anno precedente.



Sources: AI Index Report - Stanford HAI, Payden Calculations

Gli investimenti privati nella costruzione di data center sono aumentati di 8 miliardi di dollari nel 2023 e di altri 5 miliardi di dollari nei primi sei mesi del 2024, con una crescita cumulativa del 93% in soli due anni. Nonostante gli investimenti elevati, i data center sono ancora in affanno, dal momento che le 9.380 strutture attualmente presenti nel mondo non riescono ancora a immagazzinare più del 15% del totale dei dati creati nel corso del 2023. I data center pieni di GPU per l'addestramento dei modelli di AI richiedono un utilizzo massiccio di elettricità: in media, una singola richiesta di ChatGPT consuma dieci volte più elettricità di una ricerca su Google, tanto che tra il 2020 e il 2023 Microsoft, i cui data center gestiscono ChatGPT, ha visto più che raddoppiare le proprie bollette energetiche. Con ogni tipo di azienda alle prese con la corsa all'AI, **si prevede che la domanda di energia dei data center aumenterà del 160%**, arrivando a rappresentare quasi un decimo del consumo energetico mondiale entro il 2030, contro l'attuale 1%. Generare energia sufficiente per alimentare la rivoluzione AI potrebbe richiedere l'utilizzo di nuove fonti energetiche, incluse le principali fonti rinnovabili e nucleari, e di reti elettriche più moderne, innescando una nuova rivoluzione energetica. Se la maggior parte delle possibilità di utilizzo di una tecnologia vengono scoperte dopo la sua implementazione, come accaduto con la nascita di Internet, gli investimenti in infrastrutture di telecomunicazione o di connettività (come le fibre ottiche) fatti negli anni '90 rappresentano una delle più grandi basi di capitale mai costruite in un tempo così breve. Gli investimenti in attrezzature e software per l'elaborazione delle informazioni sono cresciuti a un tasso medio trimestrale annualizzato del 21% tra il 1995 e il 1999.

Proprio come i semiconduttori e i data center sono la spina dorsale dell'AI, le apparecchiature per le telecomunicazioni e l'hardware dei computer sono state quella di Internet, tanto che al suo apice, l'investimento in telecomunicazioni sfiorava i 120 miliardi di dollari (224 miliardi di dollari di oggi), con investimenti cumulativi pari a oltre 1.000 miliardi di dollari odierni. Ad ogni modo, l'eccesso di finanziamenti portò a un crollo del mercato finanziario e a una lieve recessione all'inizio degli anni Duemila. Al momento dello scoppio della bolla delle telecomunicazioni, meno del 3% della fibra ottica posata nelle maggiori città degli Stati Uniti trasportava segnali e solo il 10% delle lunghezze d'onda disponibili a livello mondiale era operativa. Improvvisamente, il prezzo della fibra crollò, così come quello dei computer, che tra il 1995 e il 2000 registrarono cali medi annui dei prezzi del 21%. Nello stesso periodo, anche il prezzo dei semiconduttori diminuì in media di quasi il 50%. Ma, nel tempo, **l'abbondanza di infrastrutture in fibra e di computer a basso costo divenne la base per l'ascesa di nuove industrie, incluse quelle degli smartphone**, con la nascita nel 2007 del primo I-Phone, dello

Payden&Rygel

streaming (YouTube nel 2005 e Netflix nel 2007), dei social media (Facebook nel 2004) e del cloud computing (con i servizi web di Amazon e Google nati a partire dal 2006). Le aziende erano più che disposte a pagare per Internet e computer veloci, soprattutto quelle attive nell'industria finanziaria, per le quali la rapidità e l'efficienza delle transazioni si traduce in un vantaggio competitivo.

Ci sono voluti quasi dieci anni dalla firma del Telecommunications Act del 1996 per vedere il traffico Internet correre finalmente attraverso fibre prima inutilizzate. Da lì al 2011 gli Stati Uniti avrebbero posato altri 19,2 milioni di chilometri di fibra per soddisfare la crescente domanda, a ritmi che rispecchiavano quelli della bolla delle Dot-com. Allo stesso modo, una crisi dell'AI potrebbe essere inevitabile nel breve termine, eliminando le speculazioni e i modelli di business insostenibili. Tuttavia, le infrastrutture durature, dai semiconduttori avanzati ai grandi data center, insieme ai progressi intellettuali, dalle invenzioni aerospaziali a quelle energetiche, potrebbero spianare la strada alla prossima ondata di innovazione, mentre la forza lavoro formatasi in questi anni (data scientist, ingegneri AI e sviluppatori) avrà le competenze necessarie per introdurre innovazioni ad oggi inimmaginabili. Anche se l'entusiasmo degli investitori si dovesse raffreddare, **le applicazioni pratiche dell'IA continuerebbero a dare i loro frutti nei più svariati campi**, che sia attraverso aumenti di produttività nel lungo termine o attraverso nuove opportunità di mercato.

In conclusione, non è dato sapere se o quando l'Intelligenza Artificiale manterrà le ambiziose promesse in cui oggi sperano gli investitori: quello che pensiamo è che, proprio come il boom e la crisi delle Dot-com, si tratta di una rivoluzione che lascerà un'eredità di cambiamenti difficili da prevedere.

**

Questo materiale riflette l'opinione di Payden & Rygel aggiornata alla data del presente documento ed è soggetto a modifiche senza preavviso. Le fonti su cui si basa sono ritenute affidabili ma non possono essere garantite. L'articolo ha uno scopo puramente illustrativo e non è da intendersi come consulenza fiscale, legale o finanziaria professionale, né come offerta di acquisto o vendita di titoli. Le performance passate non sono garanzia di risultati futuri. Questo materiale è stato approvato da Payden & Rygel Global Limited, società autorizzata e regolamentata dalla Financial Conduct Authority del Regno Unito, e da Payden Global SIM S.p.A., società di investimento autorizzata e regolamentata dalla CONSOB italiana.

Payden & Rygel

Con oltre 164 miliardi di dollari in gestione, Payden & Rygel è leader nella gestione del risparmio e annovera tra i suoi clienti banche centrali, fondi pensione, imprese di assicurazione, università, banche private e fondazioni di varia natura. Società di gestione indipendente e non quotata, Payden ha sede a Los Angeles con uffici a Boston e hub di gestione a Londra e a Milano.

Per ulteriori informazioni contattare:

BC Communication

Beatrice Cagnoni | Tel: +39 335 5635111 | beatrice.cagnoni@bc-communication.it

Giulia Franzoni | Tel: +39 334 3337 756 | giulia.franzoni@bc-communication.it