

DeepSeek ha messo in crisi il metodo Usa su larga scala

Intelligenza artificiale

Daron Acemoglu

Dopo che il 20 gennaio il lancio di DeepSeek-R1 ha provocato un forte calo del prezzo delle azioni del chipmaker Nvidia e una forte diminuzione delle valutazioni di altre aziende tecnologiche, alcuni hanno dichiarato che si trattava di un "momento Sputnik" nella corsa sino-americana per la supremazia nell'intelligenza artificiale. Sebbene l'industria americana dell'intelligenza artificiale avesse bisogno di una scossa, l'episodio solleva alcune domande difficili. Gli investimenti dell'industria tecnologica statunitense nell'Intelligenza artificiale sono stati massicci, e Goldman Sachs ha stimato che «le grandi aziende tecnologiche, le società e i servizi pubblici sono destinati a spendere circa 1.000 miliardi di dollari in spese di capitale nei prossimi anni per sostenere l'Intelligenza artificiale». Tuttavia, per molto tempo, molti osservatori, compreso, hanno messo in dubbio la direzione degli investimenti e dello sviluppo dell'Intelligenza artificiale negli Stati Uniti.

Con tutte le aziende leader che seguono essenzialmente lo stesso schema (anche se Meta si è leggermente differenziata con un modello parzialmente *open source*), il settore sembra aver puntato tutto su una carta sola. Senza eccezioni, le aziende tecnologiche statunitensi sono ossessionate dalla scala. Citando "leggi di scala" ancora da dimostrare, presumono che l'immissione di sempre più dati e potenza di calcolo nei loro modelli sia la chiave per sbloccare capacità sempre maggiori. Alcuni affermano addirittura che «la scala è tutto ciò che serve». Prima del 20 gennaio, le aziende statunitensi non erano disposte a prendere in considerazione alternative ai modelli di fondazione pre-addestrati su enormi serie di dati per prevedere la parola successiva in una sequenza. Date le loro priorità, si sono concentrate quasi esclusivamente sui modelli di diffusione e sui chatbot volti a svolgere compiti umani (o simili a quelli umani). Sebbene l'approccio di DeepSeek sia in linea di massima lo stesso, sembra che si sia affidato maggiormente all'apprendimento per rinforzo, ai metodi di miscela di esperti (utilizzando molti modelli più piccoli ed efficienti), alla distillazione e a un raffinato ragionamento a catena. Questa strategia avrebbe permesso di produrre un modello competitivo a una frazione del costo.

Anche se c'è una certa controversia sul fatto che DeepSeek ci abbia raccontato tutta la storia, questo episodio ha messo in luce il "pensiero di gruppo" all'interno dell'industria dell'Intelligenza artificiale statunitense. La sua cecità nei confronti di approcci alternativi, più economici e più promettenti, unita all'hype, è esattamente ciò che Simon Johnson e io avevamo previsto in *Potere e progresso* (il Saggiatore, 2023), scritto poco prima dell'inizio dell'era dell'Intelligenza generativa. La domanda che ci si pone ora è se l'industria statunitense abbia altri punti ciechi, ancora più pericolosi. Ad esempio, le principali aziende tecnologiche statunitensi stanno perdendo l'opportunità di portare i loro modelli in una direzione più "pro-umana"? Sospetto che la risposta sia sì, ma solo il tempo potrà dirlo. C'è poi da chiedersi se la Cina stia superando gli Stati Uniti. Se è così, significa che le strutture autoritarie verticistiche *top-down* (quelle che James A. Robinson e io abbiamo chiamato "istituzioni estrattive") possono eguagliare o addirittura superare modalità *bottom up* nel guidare l'innovazione? Sono propenso a pensare che il controllo dall'alto verso il basso ostacoli l'innovazione, come Robinson e io abbiamo sostenuto in *Perché le nazioni falliscono* (il Saggiatore, 2013). Sebbene il successo di DeepSeek sembri sfidare questa affermazione, è ben lungi dall'essere una prova definitiva che l'innovazione sotto istituzioni estrattive possa essere altrettanto potente o duratura di quella sotto istituzioni inclusive. Dopo tutto, DeepSeek si basa su anni di progressi negli Stati Uniti (e in parte in Europa). Tutti i suoi metodi di base sono stati sperimentati negli Stati Uniti. I modelli di miscela di esperti e l'apprendimento per rinforzo sono stati sviluppati negli istituti di ricerca accademici decenni fa; e sono state le aziende statunitensi di Big Tech a

introdurre i modelli di trasformazione, il ragionamento a catena e la distillazione. Quello che DeepSeek ha fatto è stato dimostrare il successo dell'ingegneria: combinare gli stessi metodi in modo più efficace di quanto abbiano fatto le aziende statunitensi. Resta da vedere se le aziende e gli istituti di ricerca cinesi saranno in grado di fare il passo successivo, proponendo tecniche, prodotti e approcci innovativi. Inoltre, DeepSeek sembra essere diversa dalla maggior parte delle altre aziende cinesi di IA, che generalmente producono tecnologie per il governo o con finanziamenti statali. Se l'azienda (nata da un hedge fund) operava sottotraccia, la sua creatività e il suo dinamismo continueranno anche ora che è sotto i riflettori? In ogni caso, i risultati ottenuti da un'azienda non possono essere considerati una prova definitiva del fatto che la Cina può battere le società più aperte sul piano dell'innovazione.

Un'altra domanda riguarda la geopolitica. La saga di DeepSeek significa che i controlli sulle esportazioni degli Stati Uniti e le altre misure per frenare la ricerca cinese sull'Intelligenza artificiale hanno fallito? Anche qui la risposta non è chiara. Sebbene DeepSeek abbia addestrato i suoi ultimi modelli (V3 e R1) su chip più vecchi e meno potenti, potrebbe ancora aver bisogno dei chip più potenti per ottenere ulteriori progressi e scalare.

Tuttavia, è chiaro che l'approccio a somma zero dell'America era impraticabile e mal consigliato. Una strategia del genere ha senso solo se si crede che ci stiamo dirigendo verso l'AgI (Artificial general intelligence – modelli in grado di eguagliare gli esseri umani in qualsiasi compito cognitivo) e che chi arriverà per primo all'AgI avrà un enorme vantaggio geopolitico. Aggrappandoci a questi presupposti – nessuno dei quali è necessariamente giustificato – abbiamo impedito una proficua collaborazione con la Cina in molti settori. Ad esempio, se un Paese produce modelli che aumentano la produttività umana o ci aiutano a regolare meglio l'energia, tale innovazione sarebbe vantaggiosa per entrambi i Paesi, soprattutto se ampiamente utilizzata. Come i suoi cugini americani, DeepSeek aspira a sviluppare l'AgI, e la creazione di un modello che sia significativamente meno costoso da addestrare potrebbe cambiare le carte in tavola. Ma ridurre i costi di sviluppo con metodi noti non ci porterà miracolosamente all'Intelligenza artificiale generale nei prossimi anni. Se l'AgI sia realizzabile a breve termine rimane una questione aperta (e se sia auspicabile è ancora più discutibile). Anche se non conosciamo ancora tutti i dettagli su come DeepSeek abbia sviluppato i suoi modelli o su cosa significhi il suo apparente successo per il futuro dell'industria dell'Intelligenza artificiale, una cosa sembra chiara: un'azienda cinese ha messo a dura prova l'ossessione dell'industria tecnologica per la scala e potrebbe persino averla scossa dal suo compiacimento.

Premio Nobel per l'economia 2024 e professore di economia al MIT
È coautore (con Simon Johnson) di «Power and Progress» (Public Affairs, 2023),
tradotto in Italia come «Potere e progresso» (il Saggiatore, 2023)

© PROQUEST SYNDICATE 2025

1000

MILIARDI DI DOLLARI

Goldman Sachs ha stimato che le grandi aziende tecnologiche, le società e i servizi pubblici statunitensi sono destinati a spendere circa 1.000

miliardi di dollari in spese di capitale nei prossimi anni per sostenere l'ia. Tutte le aziende leader seguono essenzialmente lo stesso schema: sono ossessionate dalla scala.



Concorrenti. I loghi di DeepSeek e di chatbot ChatGPT