

## Tre ex ricercatori di OpenAI vogliono rendere l'IA più economica con la loro startup

2026-09-10 07:00:44 di Richard Nieva

URL:<https://forbes.it/2026/09/10/applied-compute-ex-openai-startup-ai-3-25-miliardi/>

Quando **DeepSeek** ha lanciato il suo modello open source R1 nel gennaio 2025, ha sconvolto la **Silicon Valley**. L'idea che un piccolo laboratorio cinese fosse riuscito a creare un modello open source così potente con risorse relativamente inferiori rispetto alle controparti americane lasciò sconcertati i pesi massimi dell'industria tecnologica. **Marc Andreessen** lo definì un "[momento Sputnik](#)". **Alex Wang**, allora ceo di **Scale**, parlò invece di una "[sveglia](#)". All'interno di **OpenAI**, quell'episodio spinse tre giovani ricercatori a rendersi conto del potenziale dei modelli open source. Se fossero riusciti a costruire l'infrastruttura necessaria per ottimizzarli, pensarono, avrebbe potuto trasformarsi in un grande business. "La nostra visione del mondo era semplicemente molto diversa da quella di molti nostri colleghi di OpenAI", racconta a **Forbes** **Yash Patil**, uno dei ricercatori. "Se i modelli open weight diventeranno davvero molto validi, non vediamo OpenAI o Anthropic possedere tutti i dati del mondo necessari ad addestrare questo super-modello. Immaginiamo piuttosto aziende come JPMorgan utilizzare i propri dati per addestrare modelli sulle cose che interessano loro, perché non cederanno mai quei dati". Cinque mesi dopo, Patil, insieme agli altri ricercatori **Rhythm Garg** e **Linden Li**, lasciò OpenAI per fondare **Applied Compute**, una startup che addestra modelli open source personalizzati per clienti aziendali e poi li esegue sul proprio cloud. Da allora è stata una corsa. Nel giugno 2025 ha raccolto un round seed da [20 milioni di dollari](#) e nel giro di un anno aveva già conquistato clienti importanti come **DoorDash**, **Nvidia** e **Microsoft**, il cui ceo **Satya Nadella** a giugno si è seduto con Patil per un'[intervista](#) sul futuro dell'AI nelle imprese. Mercoledì Applied Compute ha poi rivelato a **Forbes** di essere in trattative per raccogliere **350 milioni di dollari** in un finanziamento che porterebbe la sua valutazione a 3,25 miliardi. La società ha precisato che alcune di queste cifre potrebbero cambiare mentre il round viene finalizzato. *The Information* aveva precedentemente [riportato](#) alcuni dettagli dell'ultima raccolta. La nuova valutazione sarebbe quasi il doppio di quella attribuita ad Applied Compute appena quattro mesi fa, con una crescita più rapida persino di quella dell'[unicorno più veloce di Y Combinator](#).

### Un'alternativa più economica ai modelli proprietari

L'angel investor **Elad Gil** è il principale investitore del round, a cui partecipano anche finanziatori già presenti nel capitale, tra cui **Lux Capital** e **Kleiner Perkins**. Patil afferma che il nuovo afflusso di capitale sarà destinato soprattutto all'acquisto di altri chip per ampliare il cloud dell'azienda con una "flotta enorme di capacità di calcolo". La società lavora con grandi aziende per aiutarle a costruire modelli e agenti AI specializzati partendo da modelli open source come Qwen di Alibaba o Kimi di Moonshot, una soluzione molto più conveniente rispetto all'utilizzo di OpenAI o Anthropic. Per perfezionare questi modelli, l'azienda utilizza un metodo chiamato "reinforcement learning", o apprendimento per rinforzo, che fornisce un feedback ai modelli premiandoli quando svolgono bene i propri compiti e li incoraggia così ad adattarsi per ottenere ulteriori ricompense. Applied Compute invia i propri ingegneri presso i clienti aziendali per addestrare i modelli sui loro dati interni, prima di distribuirli e farli funzionare sui propri server. L'idea è offrire un unico interlocutore per addestrare, distribuire e gestire agenti open source personalizzati. La scorsa settimana la società ha lanciato AC2, una piattaforma che permette ai clienti di utilizzare autonomamente gli strumenti interni di Applied Compute per l'addestramento dei modelli: un'opzione più economica rispetto ai servizi di consulenza ad alto livello di assistenza offerti dalla startup. Il round di finanziamento arriva mentre i modelli open source stanno diventando sempre più popolari, con le aziende che cercano di contenere i costi

crescenti legati all'intelligenza artificiale. I principali modelli frontier proprietari, come Fable di Anthropic o Sol di OpenAI, sono estremamente potenti, ma possono costare [quasi 40 volte](#) di più per milione di token rispetto ad alcuni modelli open source. Patil sostiene che, in determinati casi, i loro modelli possono costare dieci volte meno dei modelli frontier. Nel frattempo, con il progressivo aumento delle loro capacità, i modelli open source sono diventati sostituti adeguati per le attività più semplici in alcuni contesti aziendali. "Ho sempre creduto che l'open source avrebbe avuto un ruolo importante", racconta a **Forbes** Gil, principale investitore del round. Secondo lui, il curriculum dei fondatori — tutti di 25 anni o meno — li rende una squadra particolarmente adatta ad affrontare il problema dell'ottimizzazione di questi modelli. "Considerando che sono l'ex team di ricerca di OpenAI che lavorava su questo tipo di problemi, sono davvero all'avanguardia".

## Da DoorDash a Nvidia e Harvey

DoorDash ha già lavorato con Applied Compute per sviluppare un modello che consente ai ristoranti di generare con precisione i propri menu quando entrano sulla piattaforma di delivery. Nvidia ha utilizzato la startup per il post-training dei suoi modelli open weight Nemotron, destinati alla costruzione di agenti complessi. Anche la startup di AI legale Harvey si è affidata ad Applied Compute per creare un agente in grado di analizzare enormi quantità di documenti processuali o esaminare contratti nell'ambito della due diligence. L'intero processo di costruzione dell'agente ha richiesto meno di due mesi, ha raccontato a Forbes Niko Grupen, responsabile della ricerca applicata di Harvey. In precedenza, ha spiegato, sarebbero stati necessari da diversi mesi ad anni e risorse considerevoli, tra cui ricercatori specializzati nel post-training, ingegneri delle infrastrutture di supercalcolo e un'enorme capacità computazionale. "È una differenza di un ordine di grandezza".

## Tre amici di Stanford diventati colleghi in OpenAI

Figlio di immigrati indiani — una pediatra e un ingegnere hardware — **Patil**, 23 anni, è cresciuto ad Austin. Durante la sua infanzia, il padre lavorava per giganti dei chip come Qualcomm e Arm alla progettazione di circuiti e alle architetture energetiche. Patil si è avvicinato alla tecnologia perché voleva essere come il fratello maggiore Neil, oggi ingegnere nella startup biotech Chai Discovery. Durante la pandemia, quando frequentava ancora le superiori, Patil e suo fratello [svilupparono un programma](#) chiamato **Helping Hands**, che metteva in contatto le persone in difficoltà con persone in buona salute disponibili a svolgere attività come fare la spesa o ritirare medicinali. **Garg**, 24 anni, cresciuto a Dallas, scoprì invece la tecnologia partecipando agli hackathon durante le superiori, mentre **Li**, 25 anni, originario della Carolina del Nord, trascorse gli anni formativi delle scuole medie e superiori a Shanghai e si appassionò alla tecnologia dopo aver assistito a una dimostrazione di auto a guida autonoma. I tre fondatori diventarono amici a Stanford. Mentre studiava informatica, Patil incontrò quello che sarebbe diventato il suo futuro capo, il ceo di OpenAI Sam Altman, durante una cena organizzata da quest'ultimo a casa propria per gli studenti di ingegneria di Stanford. "Ero davvero nervoso. Non gli ho parlato. Forse alla fine gli ho detto 'ciao'", racconta Patil. Pochi mesi dopo OpenAI lanciò ChatGpt e, come il resto del mondo, Patil rimase affascinato dalla tecnologia. Scrisse un'email ad Altman dicendogli che gli sarebbe piaciuto lavorare sul chatbot ancora agli inizi e Altman lo invitò a candidarsi al programma di residency dell'azienda. Dopo essere entrato in OpenAI a 20 anni, Patil segnalò Garg per un'assunzione. Garg fece poi lo stesso con Li. Una volta entrati nell'azienda, Patil lavorò sull'infrastruttura di post-training e su quello che sarebbe poi diventato Codex, mentre Garg si occupò dei modelli di ragionamento e Li si concentrò sui sistemi di machine learning. Poiché erano tutti molto giovani rispetto agli altri ricercatori di OpenAI, il loro legame diventò ancora più stretto. Patil e Garg dividevano un appartamento vicino alla sede di OpenAI a San Francisco, mentre Li, che abitava nelle vicinanze, passava spesso da loro dopo il lavoro.

## L'intuizione arrivata sulle piste da sci

Poi arrivò il momento della svolta con **DeepSeek**. Patil, Garg e Li si trovavano casualmente insieme durante una vacanza sugli sci a Lake Tahoe quando venne lanciato il modello R1. Affascinati, trascorsero gran parte

della giornata a parlarne tra una discesa e l'altra. La vicenda accese sui modelli open source un'attenzione senza precedenti. Patil non pensa che i clienti aziendali abbandoneranno i modelli frontier più avanzati di OpenAI e Anthropic per puntare esclusivamente sull'open source. Ritene invece che le aziende utilizzeranno una combinazione di modelli di fascia alta e soluzioni più economiche, a seconda dei casi d'uso. E mentre alcune imprese iniziano a concentrarsi maggiormente sull'efficienza dei costi e le alternative open source diventano opzioni sempre più valide, Applied Compute ritiene che questo sia il suo momento.

## **"Dobbiamo tornare a lavorare"**

Leigh Marie Braswell, partner di Kleiner Perkins che ha guidato l'investimento della prestigiosa società di venture capital in Applied Compute, ha potuto osservare direttamente la determinazione dell'azienda all'inizio di quest'anno. Braswell, appassionata giocatrice di poker, è nota nella Silicon Valley per organizzare serate di poker con le aziende presenti nel portafoglio del fondo. Ma quando ne organizzò una nella sede di Applied Compute a San Francisco, la serata venne interrotta intorno alle 22. "Sono stati l'unica azienda che mi abbia mai cacciata dall'ufficio prima del previsto", racconta ridendo. "Mi hanno detto: 'Dobbiamo tornare a lavorare'".