



Anthropic sfida il Pentagono: «L'IA deve avere limiti etici»

Il segretario alla Guerra dell'amministrazione Trump, Pete Hegseth, ha minacciato di togliere le commesse del Pentagono all'azienda californiana di IA Anthropic.

© EPA/SHAWN THEW

IL CASO / L'azienda di San Francisco non intende concedere all'Esercito statunitense l'accesso completo al sistema Claude per il pericolo che esso sia utilizzato per la sorveglianza di massa o per armi letali senza il controllo umano - Scaduto ieri l'ultimatum del segretario alla Guerra Pete Hegseth

Dario Campione

Esiste, o può esistere, un'etica dell'intelligenza artificiale (IA) applicata alla guerra? Chi dovrebbe deciderne i principi e le regole? E chi dovrebbe sanzionarne le eventuali violazioni? Domande che qualcuno potrebbe giudicare astratte sono, in realtà, molto concrete. E lo dimostra quanto sta accadendo in queste ore tra Washington e San Francisco: il «caso Anthropic», svelato tre giorni fa dal portale d'informazione politica Axios e diventato, in poche ore, uno degli argomenti di maggiore discussione sui media occidentali.

L'incontro di martedì

Martedì scorso, i vertici del Pentagono - guidati dal segretario alla Difesa dell'amministrazione repubblicana Pete Hegseth - hanno incontrato i dirigenti della società californiana Anthropic per dirimere una questione giudicata essenziale dai militari: i limiti dell'utilizzo di Claude, il potente modello di IA che l'azienda di San Francisco fornisce all'Esercito USA.

Il Pentagono spinge da tempo per un accesso illimitato alle capacità di Claude, mentre

Anthropic non ha mai permesso che il suo prodotto venisse utilizzato per sistemi di sorveglianza di massa o per sistemi d'arma autonomi, gli stessi in grado di usare l'intelligenza artificiale per uccidere persone senza l'intervento umano.

L'incontro fra Anthropic e il Pentagono è avvenuto un mese dopo che l'Esercito statunitense ha usato Claude nell'operazione che ha portato alla cattura del leader venezuelano Nicolás Maduro.

In gioco, nelle trattative, c'è soprattutto un principio, ovvero se l'industria dell'IA possa o meno opporsi alla pretesa governativa di utilizzare liberamente gli algoritmi in campo militare, «una questione

ne da tempo controversa tra ricercatori e sostenitori dell'IA etica», ha commentato ieri il Guardian.

Secondo le rivelazioni di Axios, Pete Hegseth avrebbe dato al CEO di Anthropic Dario Amodei tre giorni di tempo per accettare le richieste del Pentagono o, in alternativa, affrontare azioni punitive: perdere i giganteschi finanziamenti garantiti dalle commesse dell'Esercito e, soprattutto, vedersi classificare come un «rischio nella catena di ap-

provvisionamento». Quest'ultima eventualità sarebbe un durissimo colpo per l'azienda di San Francisco. La designazione di «rischio», infatti, più comunemente pensata per le imprese straniere, vieterebbe anche ad altri fornitori che fanno affari con l'Esercito statunitense di utilizzare i prodotti di Anthropic.

Emil Michael, oggi sottosegretario alla Difesa degli Stati Uniti ma in passato vicepresidente senior di Uber e direttore operativo di Klout, ha attaccato personalmente Amodei l'altro ieri scrivendo su X che l'Esecutivo americano «non desidera altro che tentare di controllare in modo diretto le proprie forze armate», mentre il CEO dell'azienda californiana, ha aggiunto Michael, «è disposto a mettere a rischio la sicurezza della nostra nazione».

La risposta alla politica

Un'accusa che Amodei ha smontato pezzo dopo pezzo in una lunga nota pubblicata giovedì sul blog di Anthropic e intitolata semplicemente «Dichiarazione sulle nostre discussioni con il Dipartimento della Guerra». Nota che segna un passaggio decisivo nel rapporto tra le Big Tech, il Pentagono e la Casa Bianca. Un colosso digitale può ancora, se

vuole, dire no alle richieste della politica.

«Il Dipartimento della Guerra ha dichiarato che avrà contratti soltanto con azien-

de di IA che accettino "qualsiasi uso lecito" dei propri algoritmi «e rimuovano» quindi ogni ostacolo e barriera, scrive Amodei. «Hanno minacciato di rimuoverci dai loro sistemi se manteniamo queste barriere; hanno anche minacciato di designarci come "rischio della catena di approvvigionamento", un'etichetta riservata ai nemici degli Stati Uniti, mai applicata prima a un'azienda americana, e di invocare il *Defense Production Act* (la legge che conferisce a un presidente degli Stati Uniti l'autorità di ritenere una determinata azienda o il suo prodotto strategici per le esigenze di difesa del Paese, ndr) per forzare la rimozione degli ostacoli. Minacce, queste ultime, intrinsecamente contraddittorie: una ci definisce un rischio per la sicurezza; l'altro etichetta *Claude* come essenziale per la sicurezza nazionale. Minacce, in ogni caso, che non cambiano la nostra posizione: non possiamo in coscienza acconsentire alla loro richiesta».

Vale la pena leggere quasi integralmente quanto sottolineato da Amodei nella sua ri-

sposta al Pentagono. La lucidità e la chiarezza delle argomentazioni non hanno in pratica bisogno di commento.

«In un insieme ristretto di casi, crediamo che l'IA possa minare, invece che difendere, i valori democratici. Alcuni usi sono anche semplicemente al di fuori dei limiti di ciò che la tecnologia odierna può fare in modo sicuro e affidabile. Due casi d'uso di questo tipo non sono mai stati inclusi nei nostri contratti con il Dipartimento della Guerra, e riteniamo che ora non debbano esserlo - scrive Dario Amodei - Il primo è la sorveglianza interna di massa. Sosteniamo l'uso dell'IA per missioni di intelligence estera e controspionaggio legali. Ma l'uso di questi sistemi per la sorveglianza interna di massa è incompatibile con i valori democratici. La sorveglianza di massa guidata dall'IA presenta rischi seri e

nuovi per le nostre libertà fondamentali. [...] Secondo la legge attuale, il governo può acquisire registri dettagliati dei movimenti, della navigazione Web e delle associazioni degli americani da fonti pubbliche senza ottenere un mandato, una pratica che la comunità dell'intelligence ha riconosciuto sollevare preoccupazioni sulla privacy e che ha generato opposizioni bipartisan al Congresso. Un'IA potente rende possibile assemblare questi dati sparsi e individualmente innocui in un quadro completo della vita di qualsiasi persona, automaticamente e su larga scala».

Il secondo caso sollevato da Amodei riguarda invece le «armi completamente autonome. Le armi parzialmente autonome, come quelle usate oggi in Ucraina, sono vitali per la difesa della democrazia. Anche armi completamente autonome (quelle che escludono completamente gli esseri umani e automatizzano la selezione e l'ingaggio dei bersagli) potrebbero rivelarsi fon-

damentali per la nostra difesa nazionale. Ma oggi - spiega il CEO di *Anthropic* - i sistemi di IA di frontiera semplicemente non sono abbastanza affidabili per alimentare armi completamente autonome. Non forniremo consapevolmente un prodotto che metta a rischio i combattenti e i civili americani. Abbiamo offerto di collaborare direttamente con il Dipartimento della Guerra in ricerca e sviluppo per migliorare l'affidabilità di questi sistemi, ma non hanno accettato questa offerta».

Uno scenario del tutto nuovo

La filosofa Mariarosaria Taddeo, docente di *Digital Ethics and Defence Technologies* all'*Oxford Internet Institute* e autrice di *Codice di guerra. Etica dell'intelligenza artificiale nella difesa* (Raffaello Cortina, 2025) ha spiegato come «l'IA sta ridefinendo non soltanto il modo in cui vengono condotte le guerre, ma anche il funzionamento delle organizzazioni della difesa e i processi decisionali e operativi, ovvero i modi di acquisizione di dati e informazioni e le tattiche e strategie di guerra».

Lo scenario, sottolinea la

studiosa italiana dell'Università di Oxford, è totalmente cambiato. Ed è potenzialmente sempre più distruttivo proprio quando si consideri «l'uso dell'IA nella difesa e il rischio che gli esseri umani possano basarsi su queste tecnologie per perpetrare crudeltà e ingiustizia nella condotta bellica. La storia delle guerre del Novecento ci ricorda l'incredibile livello di atrocità che siamo capaci di infliggerci gli uni gli altri e come abbiamo saputo far leva sulla tecnologia a questo fine. Considerando che l'intelligenza artificiale è progettata per massimizzare (e persino superare) le capacità umane, l'urgenza e la necessità di un'analisi etica che ne guidi l'uso nella difesa dovrebbero apparire evidenti a tutti».



In un insieme ristretto di casi crediamo che l'IA possa minare invece che difendere i valori democratici

Dario Amodei
CEO di *Anthropic*

La scheda

Il valore stimato è salito in cinque anni fino a 380 miliardi

La fondazione nel 2021

Anthropic PBC è un'azienda americana di IA con sede a San Francisco, in California. Fondata nel 2021, ha sviluppato una famiglia di grandi modelli linguistici (LLM) chiamati *Claude* e afferma di operare come società di utilità pubblica, che studia e sviluppa l'IA per distribuire modelli sicuri per il pubblico.

I contratti con le Big Tech

Anthropic vanta contratti di collaborazione con *Amazon*, *Google*, *Nvidia* e *Microsoft* e commesse pubbliche anche del Pentagono. A febbraio 2026, il suo valore stimato era di 380 miliardi di dollari.

Il padre italo-americano

Dario Amodei è nato a San Francisco nel 1983. Il padre Riccardo era un artigiano italo-americano originario di Massa Marittima; la madre, Elena Engel, è un'americana ebrea nativa di Chicago, in Illinois.

Un patrimonio di 7 miliardi

Amodei si è diplomato alla Lowell High School di San Francisco e ha fatto parte della squadra USA alle Olimpiadi di Fisica nel 2000. Si è laureato a Stanford in Fisica e ha conseguito un dottorato in Biofisica a Princeton. Dopo aver lavorato nella cinese *Baidu* tra il 2014 e il 2015, è approdato prima a *Google* e poi, nel 2016, in *OpenAI*. Con la sorella Daniela e altri ex membri senior di *OpenAI* ha fondato *Anthropic*. Nel 2025, la rivista *Time* lo ha inserito tra le 100 persone più influenti al mondo. *Forbes* stima il patrimonio netto di Amodei in 7 miliardi di dollari.