LA NUOVA BUSSOLA QUOTIDIANA
DOMENICA
Direttore Riccardo Cascioli

Il caso

Anthropic vs Pentagono: il paradosso dell'IA civile in guerra

ATTUALITà 21_03_2026

Daniele Ciacci

Nella notte tra l'1 e il 2 marzo 2026, mentre le prime ondate di attacchi americani colpivano obiettivi iraniani, un software di intelligenza artificiale elaborava dati, sintetizzava documenti di intelligence e (secondo le fonti citate da CBS News e *Wall Street Journal*

) contribuiva attivamente alle operazioni militari. Quel software non era nulla di estremamente specifico, definito e impiegato dal Pentagono per scenari di guerra. Si trattava di Claude, un modello di IA dell'azienda californiana Anthropic, attualmente impiegato anche in Italia. Con pochi euro fai un abbonamento e lo puoi usare anche tu.

Può sembrare un dettaglio tecnico, ma in realtà è un punto di emersione di una questione morale vastissima: può un servizio civile, impiegato nell'uso quotidiano da milioni di persone, essere impiegato anche in scenari di guerra? E perché questa evenienza non fa parte del dibattito pubblico dell'Occidente?

Facciamo un passo indietro. Per mesi, i vertici del Pentagono, guidati dal sottosegretario Emil Michael, avevano negoziato con Anthropic per ottenere un accesso senza restrizioni ai propri modelli. La posizione del Dipartimento della Guerra (così Washington ha ribattezzato il Dipartimento della Difesa nell'era Trump) era semplice ma ambigua. Dichiarava infatti di voler usare Claude per «tutti gli scopi leciti». Vista anche la nebulosità dell'affermazione, la risposta di Anthropic non poteva che essere estremamente chiara: no, senza garantire almeno due garanzie. La prima: nessun impiego di Claude per armi completamente autonome. La seconda: nessuna sorveglianza di massa sui cittadini americani.

Secondo la ricostruzione pubblicata da *Pirate Wires*, le trattative sarebbero diventate presto lunghe ed estenuanti. Emil Michael descriveva le riunioni al tavolo con i responsabili di Anthropic come sessioni interminabili di narrazione di tutti i possibili «scenari ipotetici», in cui l'azienda concedeva sì eccezioni su eccezioni, ma sempre rivendicando il diritto di valutare caso per caso. «Non posso prevedere ogni eccezione per un dipartimento da tre milioni di persone, né posso prevedere il futuro», ha dichiarato Michael. Il problema è però ben delineato esattamente nella rappresentazione di questa immagine. Un ente governativo che fonda la propria presenza sulla guerra non può affidarsi a un servizio che è invece civile, che basa il proprio utilizzo appunto su "termini e condizioni" che possono essere rappresentati in documenti scritti e per forza ampi. Per il Pentagono, non può che essere inaccettabile la pretesa di Anthropic. Avrebbe infatti significato che una società privata potesse «staccare la spina» a un'operazione militare nel bel mezzo del suo svolgimento per la mancata previsione di tutto lo scibile di scenari ipotetici.

Il punto di rottura sarebbe arrivato dopo la cattura dell'ex presidente venezuelano Nicolas Maduro, un'operazione condotta con il supporto di Palantir, piattaforma che integrava i modelli di Anthropic. Secondo Michael, un alto dirigente di Anthropic avrebbe contattato Palantir per sapere esattamente come il modello Claude fosse stato utilizzato

nell'operazione, lasciando intendere che un uso ritenuto non conforme alle policy avrebbe portato al ritiro del servizio. Anthropic ha definito questa ricostruzione come «falsa», ma la tensione tra le due entità – Pentagono e Anthropic – stava crescendo. Così, Trump ha decretato che tutte le agenzie federali dismettano i prodotti Anthropic entro sei mesi. E così, OpenAI, rivale di Anthropic, ha firmato un accordo con il Pentagono, accettando condizioni simili a quelle che Anthropic aveva rifiutato. Il mercato ha premiato la scelta di Anthropic, e Claude è diventata l'app gratuita più scaricata sull'App Store americano, con un aumento di download del 240% mese su mese.

L'amministratore delegato di Anthropic, Dario Amodei, ha scelto parole significative per la sua risposta pubblica: «Dissentire dal governo è la cosa più americana del mondo. E noi siamo patrioti». Una frase che merita attenzione, perché segna il passo di una posizione peculiare in cui si trovano le maggiori aziende di IA, che devono scegliere tra interessi commerciali enormi, responsabilità etiche proclamate e una pressione geopolitica devastante. E qui emerge il paradosso: può una società privata – in questo caso Anthropic – porre veti su operazioni militari di uno Stato democratico, anche qualora le sue obiezioni siano eticamente fondate e condivise? In breve: chi sceglie chi vive e chi muore, lo Stato o l'economia?

Da una parte, la tradizione del pensiero politico occidentale ha sempre collocato la decisione sull'uso della forza nell'ambito dell'autorità legittima, quella che risponde a un popolo, a una legge, a un ordine riconoscibile. Questo però non significa che il Pentagono abbia ragione. Le preoccupazioni sulle armi autonome sono reali e serie: un sistema di IA che decida autonomamente di colpire un bersaglio senza supervisione umana rappresenta un salto quantico nella storia della guerra, con implicazioni militari e di diritto quantomeno inedite. D'altra parte, queste riflessioni non possono concretizzarsi nelle stanze di una società privata di San Francisco, per quanto ben intenzionata.

Questa vicenda illumina con crudezza un vuoto normativo che tutti fingono di non vedere e rimandano ad un futuro indefinito, ma l'intelligenza artificiale è già nei teatri di guerra, e contribuisce all'identificazione di obiettivi sensibili. Questi tempi oscuri si aprono con una domanda: siamo davvero disposti a delegare la decisione di vita o di morte a un sistema che non conosce coscienza e responsabilità, e che non può essere chiamato a rispondere davanti a nessun tribunale, umano o divino che sia? L'algoritmo dell'IA è l'ultima delega alla responsabilità umana, l'ultimo lascito del libero arbitrio, l'abbandono di noi stessi per qualcosa di altro. Questa guerra non lascerà cadaveri sulla coscienza di nessuno, perché né Claude né ChatGPT ne hanno una.