# Anthropic ha bloccato i tentativi di alcuni scienziati di usare l'Ai per realizzare armi biologiche di Roberto Cosentino

**Secondo un recente rapporto della compagnia di intelligenza artificiale, sarebbero state effettuate delle ricerche che avrebbero potuto portare allo sviluppo di armi biologiche (Fonte: https://www.corriere.it/ 10 settembre 2026)**


In un rapporto condiviso di recente, **Anthropic** ha affermato di aver bloccato la possibilità che fossero costruite armi biologiche attraverso i propri modelli di intelligenza artificiale. Sarebbero stati diversi i tentativi, da parte di alcuni scienziati, di usare i prodotti di Anthropic per sviluppare armi simili. Un report che arriva a poche ore di distanza dall'allarme di un ex dipendente della compagnia di AI, Jacob Coxon, il quale ha causato un ennesimo (giustificato?) terremoto mediatico attorno alla possibilità che l'AI possa un giorno porre fine alla razza umana.

**La linea sottile**

Sembrerebbe esistere una linea **estremamente sottile** tra l'utilizzo dell'AI per il bene dell'umanità o per il suo male. Come riferito dal responsabile dell'intelligence sulle minacce presso Anthropic, **Jacob Klein**, «Non si tratta di un personaggio dei fumetti che dice: "Ehi, voglio costruire un'arma biologica per uccidere tutti"». Secondo Anthropic, non è stato possibile stabilire se lo scopo della ricerca fosse legittimo oppure no, proprio perché un'indagine biologica rigorosa può portare sia a scoperte degne di nota (nuovi farmaci salvavita, appunto), sia a patogeni pericolosi. Nel rapporto condiviso dalla compagnia degli Amodei, vi sono una sfilza di abusi degli ultimi 8 mesi a cui i propri modelli, tra cui **Claude**, sono stati sottoposti. Non è una situazione del tutto nuova.

Già in passato Anthropic aveva affermato che vi fosse la possibilità di utilizzi riconducibili ad affiliati al governo cinese e a quello iraniano per scopi di sorveglianza verso comunità sensibili. Un altro utilizzo improprio è riconducibile invece a **media statali russi**. Anche in questo caso Claude sarebbe stato utilizzato per creare propaganda sotto mentite spoglie di giornalismo indipendente, tra cui fake news relative alle elezioni moldave. E ancora, Anthropic ha assistito a tentativi di utilizzo per lo sviluppo di software per **progettare armi da fuoco, missili, droni armati e bombe**. I casi, ancora una volta, arrivavano da **Cina**, **Russia** e **Yemen**.
Come riporta il New York Times, la compagnia dei fratelli Amodei non ha riferito quali fossero le cellule che avrebbero cercato di utilizzare a proprio vantaggio i sistemi di Anthropic, ma si presume che, nel caso dello Yemen, possa essersi trattato della milizia degli **Houthi**, sostenuta dall'Iran. Ma anche questa non è del tutto una novità. Uno studio pubblicato da Cambridge lo scorso luglio e a cura della dr.ssa Antonia Juelich, aveva documentato, attraverso interviste a 27 ex membri, l'utilizzo dell'AI da parte delle due principali fazioni di **Boko Haram** per diversi scopi, tra cui la pianificazione di attacchi, la risoluzione di problemi legati alle armi e la **progettazione di ordigni esplosivi**. dubbia provenienza.

**Poco spazio ai rischi biologici**

Negli ultimi tempi, tuttavia, si è dato molto spazio agli incidenti di natura informatica, mentre meno ne è stato riservato ai rischi di natura biologica menzionati invece nel rapporto appena pubblicato da Anthropic che, anche in questo caso, tutela la privacy dei ricercatori. Secondo quanto riferito nel documento, questi avrebbero aggirato i controlli posti dalla compagnia, **utili a impedire l'accesso ai propri sistemi agli utenti provenienti da aree geografiche soggette a restrizioni**.
Ciò non toglie che, benché **non sia stato possibile apprendere la vera natura della loro ricerca,** abbiano tentato di «offuscare lo scopo della loro ricerca per eludere le nostre misure di sicurezza». Tra le ricerche più inquietanti e che la stessa compagnia di AI **ha ritenuto preoccupanti**, vi è quella di uno scienziato che lo scorso maggio ha chiesto a Claude di **scrivere una richiesta di finanziamento** utile a condurre una ricerca per fare esperimenti sul virus **Chikungunya**, malattia trasmessa dalle zanzare. Una ricerca che, come già detto, **potrebbe avere una doppia valenza**: sviluppare un vaccino oppure creare versioni del virus più pericolose. Lo scopo dello scienziato era indurre mutazioni nel virus **per renderlo più dannoso attraverso infezioni ripetute su animali vivi**. Anthropic ha inoltre affermato di aver rilevato che il lavoro era destinato a essere svolto **presso un istituto di ricerca militare**. «Non sappiamo se la ricerca fosse finalizzata alla creazione di armi», ha affermato Klein. «Ma il fatto che un'istituzione militare conduca ricerche volte a ottenere un aumento di funzionalità è motivo di preoccupazione».
All'indomani dell'allarme condiviso dall'ex ricercatore di Anthropic, su X è emerso un dibattito, peraltro sostenuto anche da **Elon Musk**, secondo cui le rivelazioni dei tre ricercatori della compagnia potessero essere parte di **"Psyops"**, ovvero tentativi nell'ambito della guerra

psicologica che possono servire per **influenzare opinioni, paure, sentimenti e comportamenti** a seconda dello scopo di qualcuno che ne abbia interesse. Altri invece sostengono che si sia trattato di una semplice **operazione di marketing** volta ad aumentare il valore di una società. Ma perché una società dovrebbe cercare un autogol simile, specie a ridosso di un ingresso in borsa?
I modelli più vecchi non erano ancora in grado di portare a termine compiti complessi, ma gli ultimi, più recenti**, potrebbero essere in grado di fornire un supporto strategico anche nella conduzione di ricerche biologiche pericolose**. Come già detto in precedenza, condurre una ricerca simile **potrebbe essere un'arma a doppio taglio**, che potrebbe sì apportare benefici, ma anche l'esatto contrario. **Un'incertezza per cui Anthropic ha deciso di inibire il prosieguo di ricerche simili** di dubbia provenienza.