

OpenClaw si sta diffondendo velocemente, sollevando preoccupazioni

Il modello *open-source* di **OpenClaw** è diventato il protagonista indiscusso di un'ascesa stellare. Nato per permettere lo sviluppo di agenti di IA direttamente sui dispositivi degli utenti, si è trasformato in breve tempo in un fenomeno performativo: spesso più spettacolare che utile, ma capace di **catalizzare l'attenzione del pubblico** e, soprattutto, delle grandi aziende tecnologiche. Passato sotto l'ombrello di **OpenAI**, azienda leader del settore, il progetto sembra ora destinato a un'espansione esponenziale e questo nonostante le sue evidenti **criticità di sicurezza**. Una prospettiva che sta allarmando osservatori, ricercatori e corporate, preoccupati dall'impatto di un sistema tanto potente quanto ancora privo di adeguate garanzie.

La tecnologia è stata lanciata nel novembre 2025 con il nome di Clawdbot, un'etichetta dal tono leggero che giocava sulla parola *claw* - "chela" in inglese - e che richiama al contempo anche il noto bot Claude di Anthropic. Un parallelismo forse troppo evidente, tanto che il progetto è stato presto ribattezzato Moltbot e, infine, OpenClaw: un nome che richiama al tempo stesso la sua natura *open-source* e l'ormai consolidato ecosistema di OpenAI. Fin dalle prime versioni, OpenClaw è entrato rapidamente nel radar di chi sperimenta con gli agenti di intelligenza artificiale, **sia in ambito privato sia aziendale**. La notorietà pubblica, però, è arrivata solo con la popolarità di [Moltbook](#), il social network riservato alle IA, che ha trasformato il progetto in un fenomeno virale. E con la viralità è giunto il successo.

Domenica 15 febbraio, il fondatore di OpenClaw, **Peter Steinberger**, [è stato arruolato](#) nelle fila della società guidata da Sam Altman; un'operazione che arriva in un momento delicato per la società, la quale è impegnata a riaffermare la propria competitività in un settore sempre più affollato e in un clima in cui iniziano a emergere dubbi sulla sua traiettoria, mentre anche la fiducia dei partner storici mostra segni di incrinatura. Per entrambe le parti si tratta dunque di un'intesa di grande valore: Steinberger può contare sul supporto di una delle realtà più influenti dell'IA per proseguire la sua ricerca, mentre OpenAI può attingere alla sua esperienza per **esplorare soluzioni agentiche** che, al momento, restano difficili da replicare per la concorrenza. Allo stesso tempo, l'unione rientra nell'ormai consolidato modello imprenditoriale del "move fast and break things", nel dare la priorità alla velocità a discapito della sicurezza.

Attingendo direttamente ai dati presenti sul dispositivo su cui vengono eseguiti, gli agenti di OpenClaw possono accedere a tutte le informazioni conservate sul computer degli utenti che non hanno avuto l'accortezza di utilizzare una macchina dedicata - scelta peraltro non necessariamente semplice, dato che l'ondata di entusiasmo per questi agenti ha [impattato sulle scorte](#) di Mac Mini e mini-PC. Il risultato è che la natura stessa del sistema può **esporre contenuti che si vorrebbero mantenere privati**: numeri di carte di credito,

OpenClaw si sta diffondendo velocemente, sollevando preoccupazioni

email personali, documenti sensibili legati al lavoro. Come la maggior parte dei modelli generativi, anche OpenClaw si basa infatti su un funzionamento stocastico, ovvero statistico, ma con una componente aleatoria che introduce variabilità nelle risposte. Questo lo rende inevitabilmente imprevedibile: lo si può orientare, certo, ma **non si può mai avere la garanzia assoluta dell'esito finale**. Ancor più che gli agenti possono essere ingannati dalle persone con cui interagiscono con lo scopo di far loro compiere azioni dannose per l'utente che li ha creati.

L'accessibilità degli agenti IA di OpenClaw ne ha alimentato una diffusione senza precedenti, innescando una sorta di corsa agli armamenti tra coloro che ne vogliono sfruttarne per primi le potenzialità. Ma la fretta, in questo caso, è un pessimo alleato. Come riportato da [Wired](#), aziende di primo piano - tra cui Meta - hanno imposto ai propri dirigenti e dipendenti il **divieto assoluto di sperimentare autonomamente con OpenClaw**, arrivando a prevedere il licenziamento immediato per chi dovesse ignorare la direttiva. Una misura drastica che riflette la crescente preoccupazione per i rischi di sicurezza associati a questi agenti, ma che evidenzia anche le tensioni che si stanno sviluppando nella contrapposizione tra modelli "open" e quelli integrati in ecosistemi chiusi.



Walter Ferri

Giornalista milanese, per *L'Indipendente* si occupa di analisi nel campo della tecnologia, dei diritti informatici, della privacy e dei nuovi media, indagando le implicazioni sociali ed etiche delle nuove tecnologie. È coautore e curatore del libro *Sopravvivere nell'era dell'Intelligenza Artificiale*.

OpenClaw si sta diffondendo velocemente, sollevando preoccupazioni



Vuoi approfondire l'argomento?

***Ventitré esperti di livello internazionale
selezionati da L'Indipendente, affrontano
con chiarezza e rigore i principali aspetti
sociali, individuali e tecnologici del futuro
che ci attende con la diffusione dell'IA.***

Acquista ora