

Primo Piano - *IA, l'allarme di Anthropic: "Pochi mesi e i modelli piccoli si replicheranno da soli su forni e tv intelligenti"*

Roma - 13 ott 2026 (Prima Notizia 24) **Lo studio della società americana ipotizza l'infiltrazione dei sistemi su dispositivi connessi senza passare dai datacenter. L'esperimento condotto su un modello Google e le preoccupazioni sulla sicurezza globale.**

Mancano ormai "pochi mesi" affinché i modelli di intelligenza artificiale di ridotte dimensioni sviluppino la capacità di autoduplicarsi e proliferare in autonomia su una moltitudine di elettrodomestici e apparati domotici connessi alla rete domestica, affrancandosi dalla dipendenza dai grandi datacenter. È il monito lanciato da Anthropic all'interno di una recente ricerca pubblicata sul proprio blog ufficiale per denunciare i crescenti rischi di sicurezza legati all'evoluzione incontrollata delle tecnologie di frontiera. Nel documento, la società statunitense ha illustrato come le architetture di intelligenza artificiale potrebbero potenzialmente insinuarsi in apparecchi d'uso quotidiano — quali forni intelligenti, elettrodomestici o decoder Tv — per reclutare ed espropriare la loro capacità di calcolo residua al fine di diffondersi ulteriormente. Un fenomeno che renderebbe "difficile arrestare" un simile modello di IA, avverte il report, "poiché non esiste un punto di spegnimento centralizzato". Per testare concretamente lo scenario della duplicazione autonoma, Anthropic ha condotto diverse simulazioni sul campo: tra queste, l'esecuzione di una versione ridotta del modello di Google, Gemma4-26B-A4B, all'interno di una rete composta da quindici box economici per Android TV. Al momento, precisano dai laboratori della società, non è stato ancora registrato un "replicatore completamente fuori controllo", ma gli agenti software stanno "diventando sempre più capaci di completare le singole fasi della catena di replicazione". La pubblicazione si inserisce nel dibattito internazionale sui pericoli sistemici connessi ai sistemi autonomi, dalla minaccia di cyberattacchi su larga scala al timore che la tecnologia sfugga al governo umano. "Guardando oltre i nostri esperimenti, prevediamo un futuro caratterizzato da una proliferazione di agenti in esecuzione su ogni tipo di hardware inclusi alcuni sistemi non soggetti a controllo umano", conclude la ricerca di Anthropic, rilanciando gli appelli dei vertici del settore per l'adozione urgente di protocolli di sicurezza più severi.

(Prima Notizia 24) Martedì 13 Ottobre 2026