Forbes
ITALIA

La Bce convoca le banche: il modello IA Claude Mythos minaccia la sicurezza

2026-05-25 15:02:17 di Forbes.it

URL:https://redazione.forbes.it/2026/05/25/la-bce-convoca-le-banche-il-modello-ia-claude-mythos-minaccia-la-sicurezza/

La Banca centrale europea ha convocato per domani, martedì 26 maggio, un vertice d'urgenza con i responsabili dei principali istituti di credito dell'area euro per affrontare i rischi di cybersicurezza legati a **Claude Mythos**. Secondo la vigilanza dell'Eurotower, la potenza del nuovo modello avanzato di intelligenza artificiale sviluppato da **Anthropic** è tale da poter compromettere gli attuali standard di protezione dei sistemi informatici bancari. ***LEGGI ANCHE:*** *Anthropic potrebbe presto valere più di OpenAI*

Cosa preoccupa la Bce

Il Financial Times ha riferito che l'obiettivo è sottolineare la gravità della minaccia posta dalla versione Preview di Claude Mythos e da analoghi modelli di IA. Al contempo, la Bce vuole esortare i gruppi bancari statunitensi – che già utilizzano queste tecnologie di ultima generazione – a condividere le informazioni sensibili con le controparti europee, che scontano un ritardo nell'accesso ai nuovi strumenti. La asimmetria nell'adozione della tecnologia dipende dalle modalità di rilascio scelte dalla stessa Anthropic, che finora ha concesso l'accesso al modello a un numero limitato di organizzazioni non europee, prevalentemente statunitensi, nell'ambito di una fase di test denominata Project Glasswing. Intervistato dal *Financial Times*, **Frank Elderson**, membro del board della Bce e vicepresidente del Consiglio di vigilanza, ha chiarito la posizione di Francoforte: data la velocità con cui si evolvono le **minacce cyber legate all'intelligenza artificiale**, queste "devono essere affrontate più rapidamente". Per questo, ha aggiunto Elderson, nell'incontro di domani l'Eurotower intende "raccogliere le valutazioni delle banche, favorire la condivisione delle loro esperienze e sottolineare l'urgenza del problema". A preoccupare i regolatori è soprattutto l'accelerazione dei tempi di attacco consentita dai nuovi algoritmi. "Se uno dei grandi fornitori di software rilascia una patch di sicurezza — ha spiegato Elderson —, oggi sembra possibile effettuare il reverse engineering della vulnerabilità che la patch dovrebbe correggere non più in settimane, ma nell'arco di appena trenta minuti".

Il contesto

Anthropic è stata fondata nel 2021 da un gruppo di ex ricercatori di OpenAI, tra cui **Dario e Daniela Amodei**, i cui patrimoni personali sono stimati da *Forbes* intorno ai **7 miliardi di dollari**. In pochi anni la società è diventata un punto di riferimento nel settore dell'intelligenza artificiale attraverso lo sviluppo dei modelli della famiglia Claude, impiegati da aziende e sviluppatori per l'automazione dei processi, la programmazione e la cybersicurezza. L'ultimo modello sviluppato, Claude Mythos, presenta capacità di analisi informatica tali da aver indotto l'azienda a escluderne il rilascio pubblico. Durante i test, il sistema ha individuato autonomamente migliaia di vulnerabilità "zero-day" nei principali sistemi operativi e browser, trovando anche falle rimaste irrisolte per decenni in software come OpenBSD e FFmpeg. Poiché il modello è in grado di generare exploit efficaci in circa dodici ore e con costi inferiori a mille dollari, Anthropic ne ha limitato l'utilizzo per motivi di sicurezza. A questo scopo è stato avviato il Project Glasswing, un'iniziativa che consente l'accesso a Mythos esclusivamente a grandi aziende tecnologiche (come Apple, Google e Microsoft) e a organizzazioni che gestiscono infrastrutture critiche, al fine di correggere i difetti dei software prima che possano essere sfruttati per attacchi informatici.