Cybersicurezza: l'Ai scopre vulnerabilità invisibili. Perchè Mythos di Anthropic cambia tutto di Mariarosaria Taddeo*

Il modello ha individuato una falla rimasta nascosta per 27 anni e ribalta l'equilibrio tra attacco e difesa. Ora il vantaggio è nell'accesso all'AI: l'Europa rischia di restare indietro e deve reagire (Fonte: https://www.corriere.it/ 27 aprile 2026)

Il 7 aprile Anthropic ha annunciato Mythos Preview, un modello che in poche ore ha trovato in OpenBSD — uno dei sistemi operativi più sicuri al mondo: una **vulnerabilità rimasta invisibile per ventisette anni.**

È un progresso tecnologico, ma è soprattutto uno spartiacque concettuale. Per trent'anni, **la cybersicurezza ha compensato il vantaggio strutturale dell'attaccante con le poche certezze del difensore: conoscere i propri sistemi, disporre del codice, avere l'autorità legale di testarli. All'attaccante toccava reinventare tutto dall'esterno, e l'ispezione profonda costava mesi di lavoro umano qualificato. Mythos azzera quel costo** e con esso l'unico vantaggio del difensore. **Da oggi conta chi ha accesso per primo al modello**. Mozilla, che ce l'ha avuto, ha chiuso 271 vulnerabilità in Firefox prima che venissero sfruttate; senza Mythos, sarebbe stato impossibile. Salta l'equilibro che permetteva alle difese informatiche di garantire resilienza dei sistemi, **rischiamo una crisi sistemica della cybersicurezza.**

La risposta di Anthropic è stata duplice: ritardare il rilascio di 90 giorni e creare Glasswing, un consorzio di quaranta organizzazioni con accesso privilegiato al modello. I membri sono tutti americani, salvo qualche presenza britannica, nessun europeo. Il BSI (Ufficio Federale per la

Sicurezza nelle Tecnologie dell'Informazione) tedesco ottiene incontri a San Francisco, non il modello. Qui la tecnologia indossa i panni della geopolitica.

Glasswing è, in sostanza, una Compagnia delle Indie della cybersicurezza.

L'analogia storica è istruttiva. Per un secolo la Compagnia delle Indie batté moneta in Bengala, firmò trattati, mantenne un esercito più grande di quello della Corona, mentre il Parlamento inglese legiferava su come vincolarla. La sovranità, intanto, era già migrata dal Parlamento alla Compagnia. Solo nel 1858, dopo la rivolta dei Sepoys, Londra ricostruì una capacità amministrativa propria e sostituì la Compagnia col Raj.

Lezione: la regolamentazione senza capacità operativa è formalità; quando una tecnologia è troppo potente e lo Stato troppo lento, la funzione sovrana migra verso chi ha la capacità operativa. Nel Seicento erano la potenza navale e il capitale mercantile; oggi sono i modelli di frontiera e i cluster GPU.

Potremmo interrogarci sui criteri con cui Anthropic ha selezionato i membri di Glasswing. Ma il dibattito è vecchio: **l'impatto dell'Ai sulla cybersicurezza era prevedibile almeno dal 2019, e sono domande a cui avremmo dovuto rispondere ieri.** Oggi conta guardare avanti.

Io penso che ci siano tre strade da percorrere. La più urgente da battere: **la Commissione europea deve negoziare un accesso consortile a Mythos per gli operatori di infrastrutture critiche,** sul modello degli acquisti congiunti di vaccini: difesa digitale come salute pubblica.

La strada da costruire nel medio termine: **una mobilitazione europea di compute e capitale paragonabile a quella attivata per l'energia dopo il 2022.** Scalare Mistral non è industrial policy da salotto, è sicurezza nazionale.

A quella più lunga ma di valenza strategica va riconosciuta un'evidenza che preferiamo ignorare**: le aziende di Ai di frontiera svolgono funzioni proprie delle agenzie di intelligence e, come tali, vanno regolate.** Serve una cornice statutaria che preveda designazione formale, nulla osta di sicurezza, oversight parlamentare classificato, obblighi di condivisione con gli alleati. E serve un dossier multilaterale — più ambizioso del G7, meno vincolante del Trattato di non proliferazione — che tratti i modelli di frontiera come questione di sicurezza collettiva, con protocolli di disclosure coordinata sulle vulnerabilità critiche.

Mythos non è un'eccezione: è soltanto la prima goccia di un temporale. I modelli cinesi arriveranno a capacità analoghe entro pochi mesi.

Alla Compagnia delle Indie occorse un secolo di ascesa e una rivolta per essere sostituita. A Mythos sono bastate poche ore per ridisegnare la sicurezza informatica. La storia ha accelerato; deve farlo anche l'Europa.

Mariarosaria Taddeo è docente di Digital Ethics and Defence Technology a Oxford e Defence Science and Technology Fellow all'Alan Turing Institute