L'illusione dell'intelligenza: perché l'AI sta cambiando la cybersecurity più degli hacker

A cura di Francesco Iarlori

Dalla sicurezza perimetrale alla sicurezza cognitiva: come l'Intelligenza Artificiale sta trasformando il concetto stesso di difesa digitale.

Per oltre quarant'anni la sicurezza informatica è stata una disciplina relativamente chiara. Esistevano sistemi da proteggere, reti da monitorare, vulnerabilità da correggere e aggressori da fermare. Le tecnologie cambiavano rapidamente, ma il paradigma rimaneva sostanzialmente invariato: difendere un perimetro digitale.

Oggi, con la diffusione dell'Intelligenza Artificiale generativa, stiamo assistendo a qualcosa di profondamente diverso. Non siamo di fronte a una semplice evoluzione tecnologica, ma a una trasformazione che coinvolge il modo stesso in cui le persone prendono decisioni, producono conoscenza e interagiscono con i sistemi informativi.

Dopo aver vissuto diverse rivoluzioni tecnologiche - dall'era Unix degli anni Ottanta all'esplosione di Internet, dalla telefonia mobile, satellitare ed ai pagamenti digitali – sono convinto che l'AI rappresenti una discontinuità ancora più profonda. Per la prima volta, infatti, non stiamo automatizzando soltanto attività operative, ma stiamo delegando parte dei nostri processi cognitivi.

Ed è proprio qui che emerge la nuova frontiera della cybersecurity.

Dalla protezione dei sistemi alla protezione delle decisioni

Storicamente, la sicurezza informatica si è focalizzata sulla protezione di asset chiaramente definiti, quali server, database, applicazioni, reti e dispositivi. Negli ultimi anni, in numerosi incontri con aziende dei settori Finanza e Telecomunicazioni, è emersa la crescente necessità di salvaguardare le proprie transazioni e gli asset trasmissivi, inclusi cavi sottomarini e satelliti.

Con l'AI generativa, il bersaglio si sposta progressivamente verso la conoscenza e i processi decisionali. Qui i settori sono trasversali, non solo le domande riguardano come introdurre nuove intelligenze in azienda in modo organico, ma anche come proteggerle.

Quando un professionista utilizza un Large Language Model per redigere un documento, sintetizzare un contratto o analizzare dati aziendali, il problema non è più soltanto proteggere il computer che utilizza. Occorre comprendere come vengono generate le informazioni, quali fonti vengono utilizzate, quali errori possono essere introdotti e quali conseguenze possono derivare da una fiducia eccessiva nei risultati prodotti dalla macchina.

La vulnerabilità non risiede più soltanto nel software, ma nel rapporto di fiducia che si instaura tra essere

umano e sistema intelligente.

L'evoluzione dell'attaccante

Anche il profilo dell'aggressore sta cambiando.

Per anni abbiamo associato l'hacker a competenze tecniche elevate: capacità di programmazione, conoscenza delle reti, studio delle vulnerabilità.

L'AI abbassa significativamente questa barriera.

Oggi strumenti avanzati consentono di generare codici, creare campagne di phishing estremamente convincenti, tradurre testi in molte lingue e simulare comunicazioni aziendali con una qualità che fino a pochi anni fa richiedeva competenze specialistiche. Ricordo che negli anni '90 le prime truffe erano abbastanza riconoscibili, mail e siti erano palesemente falsi, anche alla vista di un occhio poco esperto.

Il cybercriminale del futuro potrebbe non essere il miglior programmatore, ma il miglior orchestratore di strumenti intelligenti.

Questo fenomeno democratizza, purtroppo, la capacità offensiva e rende più difficile distinguere una comunicazione autentica da una manipolata.

Il problema della fiducia

Uno degli aspetti meno discussi riguarda il concetto di fiducia.

Nella sicurezza tradizionale il principio *trust but verify* ha sempre rappresentato una buona pratica. Con l'AI questo principio diventa essenziale.

I modelli generativi possono produrre contenuti corretti, plausibili e ben scritti anche quando le informazioni sottostanti sono errate. Possono inventare riferimenti bibliografici, generare dati inesistenti o formulare conclusioni apparentemente convincenti ma prive di fondamento.

La sfida non consiste soltanto nell'individuare contenuti malevoli, ma nel riconoscere quelli che appaiono credibili senza esserlo. Il problema diventa semantico.

Per questo motivo, ritengo che la cybersecurity del prossimo decennio dovrà integrare competenze provenienti da discipline diverse: psicologia cognitiva, comunicazione, gestione del rischio, governance dei dati e intelligenza artificiale.

Dalla sicurezza informatica alla sicurezza cognitiva

Le organizzazioni stanno progressivamente comprendendo che il patrimonio più importante non è rappresentato dall'infrastruttura tecnologica, ma dalla capacità delle persone di prendere decisioni corrette.

In questo contesto emerge il concetto di *sicurezza cognitiva*: l'insieme delle pratiche e delle tecnologie finalizzate a proteggere i processi mentali, informativi e decisionali da manipolazioni, errori e distorsioni.

Deepfake, contenuti sintetici, disinformazione automatizzata e sistemi conversazionali avanzati rappresentano esempi concreti di minacce che non colpiscono direttamente i sistemi, ma influenzano la percezione della realtà.

La domanda non è più soltanto *chi può entrare nei miei sistemi?*, ma *chi può influenzare le mie decisioni?* Sempre più i sistemi aiutano a decidere, ma spesso il poco tempo per decidere e le troppe decisioni da prendere favoriscono l'adozione di agenti che lavorano in autonomia. Da qui, il rischio della perdita di controllo, peraltro noto, che diventa sempre più alto.

Il ruolo della governance

L'errore più comune consiste nel considerare l'AI come una questione esclusivamente tecnica.

In realtà la sicurezza dell'Intelligenza Artificiale è prima di tutto un tema di governance: per questo, alla fine degli anni duemila frequentavo più spesso dipartimenti HR o strategici che dipartimenti tecnici.

Occorre definire regole di utilizzo, criteri di validazione, responsabilità decisionali e meccanismi di controllo. Le organizzazioni che adotteranno l'AI senza costruire adeguati modelli di governance rischiano di introdurre nuove vulnerabilità proprio mentre cercano di aumentare efficienza e produttività.

L'esperienza maturata negli anni nei grandi programmi di trasformazione digitale insegna che la tecnologia, da sola, non risolve i problemi organizzativi. Talvolta li amplifica.

Conclusioni

La cybersecurity sta entrando in una nuova fase storica.

Per decenni abbiamo protetto macchine, reti e applicazioni. Oggi dobbiamo iniziare a proteggere anche **conoscenza**, **fiducia** e, soprattutto, **capacità decisionale**.

L'Intelligenza Artificiale non sostituirà la sicurezza informatica tradizionale, ma ne estenderà il campo d'azione verso territori finora poco esplorati. Le organizzazioni che comprenderanno per prime questa evoluzione saranno meglio preparate ad affrontare un mondo in cui il principale bersaglio non sarà il computer, ma il pensiero umano.

La vera sfida dei prossimi anni non sarà costruire sistemi più intelligenti. Sarà garantire che l'intelligenza, umana e artificiale, rimanga affidabile, verificabile e governabile.