Anthropic, bolaget bakom Claude, har publicerat en ovanligt öppen rapport om hur kriminella och stater använder AI:n för dataintrång. Två fall förändrade hur jag ser på säkerhet i Sverige – inte tekniken i sig, utan vilka som nu har råd att attackera dig, skriver Emre Süren, Royal Hacking Lab Manager, KTH.

Anthropic, med vd:n Dario Amodei, publicerade nyligen en rapport om hur kriminella och stater använder AI för dataintrång. "Sluta dimensionera försvar efter frågan 'vem skulle ens vilja attackera oss?'. Utgå från att varje exponerat system är en möjlig måltavla", skriver Emre Süren, Royal Hacking Lab Manager, KTH. FOTO: DON FERIA/TT

Hoten i den nya AI-världen

Det första fallet kopplas till rysk statlig underrättelsetjänst. Målen var väntade: ukrainska myndigheter, europeiska ambassader, försvarsbolag. Nytt var att ett AI-system skötte intrången nästan självt: skrev nätfiskemejl, registrerade falska sajter, hanterade kapade konton.

Mest talande: de testade koden mot verkliga säkerhetsprogram tills den passerade oupptäckt. Inne i nätverket fortsatte AI:n kontrollera om den undgick upptäckt – vid en reaktion skrev den om koden och skickade en ny version.

Den loopen – upptäcka, skriva om, skicka ut igen – har länge varit en av de största kostnaderna i en cyberattack. Nu kan den i hög grad ske utan mänsklig handpåläggning. Hur visste AI:n att koden upptäckts? Inte klassisk command-and-control, där skadlig kod regelbundet "pingar hem" – en svag signal, tystnad kan bero på annat än upptäckt.

Rapporten visar att aktören körde kommandon direkt mot kapade system och rörde sig i nätverken på egen hand. Min tolkning: AI:n använde samma åtkomst för att fråga om en process kördes eller en logg visade en upptäckt – som en administratör, fast automatiserat.

Det är märkligare än ett kontrollanrop: att aktivt fråga offrets egna försvar vad de sett.

Det andra fallet skiljer sig från det första, utom i resultatet. En person använde samma AI för att bygga en sökmotor för stulna personuppgifter: namn, id-nummer, adresser, som sedan såldes på darknet.

"Anthropic kallar det ett tydligt exempel på att en person med AI kan åstadkomma det som tidigare krävde ett företag."

EMRE SÜREN,
ROYAL HACKING LAB MANAGER VID KTH

Anthropic kallar det ett tydligt exempel på att en person med AI kan åstadkomma det som tidigare krävde ett företag. Sverige har redan en laglig variant: en sökruta byggd på offentliga register. Samma typ av produkt – men utan lagstöd, utan offentlig registergrund och utan tillsyn – tar nu bara några veckor.

Ställ fallen bredvid varandra: Anthropics utredare kan inte längre avgöra, utifrån attacken i sig, om den kommer från en stat eller en ensam individ. Med deras egna ord är sofistikering "inte längre en pålitlig signal" om vem som ligger bakom en attack.

Det är avgörande för Sverige: vi har länge dimensionerat försvaret efter motståndarens antagna storlek. Den logiken håller inte längre – att attackera är nu billigt nog för att skalan inte längre säger mycket om vem som bryr sig. Kostnaden hamnar hos offret

Angripare stjäl i allt högre grad AI-kontons inloggningar och använder dem för attacker. Ett stulet konto ger, enligt Anthropic, tre saker: andrahandsvärde, datorkraft betald av någon annan, och skydd – aktiviteten ser ut att komma från den riktiga innehavaren.

Hacktivisten i fall två gjorde just detta: letade efter exponerade AI-nycklar och växlade mellan stulna konton via proxyservrar för att smälta in i trafiken. En hel kampanjmånad drevs enbart på stulna AI-abonnemang.

Det är inte bara en högre nota för kontoinnehavaren – någon annans attack genomförs i ditt namn. Det vänder gammal säkerhetslogik upp och ned: kontoinnehavaren, inte angriparen, riskerar nu kostnaden och skulden.

Vad Sverige faktiskt borde göra? Tre förändringar kräver ingen ny teknik, bara andra prioriteringar.

För det första: sluta dimensionera försvar efter frågan "vem skulle ens vilja attackera oss?". Utgå från att varje exponerat system är en möjlig måltavla.

För det andra: behandla AI-nycklar som bankinloggningar. Ett läckage innebär inte bara en större faktura, utan att ditt namn kan kopplas till någon annans attack.

För det tredje: se vilka som också pekas ut som mål i rapporten – politiska partier, redaktioner, civilsamhällesorganisationer, inte bara företag. Grupper som sällan får frågan "är er sajt säker?" men som borde få den oftare.

Det här kräver ingen panik. Men det kräver att vi överger en gammal föreställning: att angripare måste vara stora för att vara farliga. Det måste de inte längre.

Sverige gick till val i förra veckan. Av allt i rapporten finns inte ett enda fall som rör påverkan mot ett svenskt val – anmärkningsvärt, med tanke på att rapporten dokumenterar påverkan mot flera andra val.

Verklig lucka, eller bara en tidsfråga? Rapportens period slutar i augusti, dagar före vårt val – en operation skulle alltså inte synas här oavsett. Det är inte ett svar, bara en påminnelse om att hålla koll, inte slå sig till ro.

EMRE SÜREN
Royal Hacking Lab Manager, Kungliga Tekniska Högskolan