IRON MOUNTAIN®

White paper

SÅ FÖRÄNDRAR GENERATIV AI SÄKERHETEN

Sammanfattning

När det gäller säkerhet, är GenAI ett verktyg för gott eller ont? Säkerhetsteam inom företag brottas med vad denna framväxande teknologi innebär för deras organisationer. Och de upptäcker i allt högre grad att generativ AI innebär både möjligheter och risker.

Innehåll

GenAI: Ett nytt verktyg för allt

Generativ artificiell intelligens (GenAI) är utan tvekan en av de mest banbrytande teknologiska innovationerna på senare tid. När ChatGPT lanserades i november 2022 markerade det början på en ny era inom datateknik.

GenAI kan användas för att skriva artiklar, skapa kod för din senaste programuppdatering, generera bilder som ser ut som foton eller producera videor och ljud som är nästan omöjliga att skilja från verkligheten. Det kan också genomsöka internet och ge svar på nästan vilken fråga som helst. Vi har bara börjat skrapa på ytan av vad denna nya teknologi kan åstadkomma.

Företag och organisationer ser med rätta stora möjligheter med GenAI. En undersökning från AWS och MIT visade att 80% av datacheferna tror att GenAI kommer att förändra deras verksamheter, och 45% sa att deras företag redan har börjat använda tekniken i stor skala.

"Generativ AI kan bli den mest omvälvande teknologin vi sett hittills," säger Steve Chase, konsultchef på KPMG i USA. "Den kommer att förändra affärsmodeller i grunden, ge nya möjligheter till tillväxt, effektivitet och innovation, samtidigt som den medför betydande risker och utmaningar. För att ledare ska kunna utnyttja GenAIs potential fullt ut måste de ha en tydlig strategi som snabbt tar organisationen från experiment till praktisk tillämpning."

Mot denna bakgrund känner många chefer ett ökat tryck att börja använda GenAI så snabbt som möjligt. Men vissa är förståeligt nog tveksamma. Trots alla fördelar med generativ AI finns det också stora risker.

Många säkerhetsansvariga kämpar fortfarande med att utarbeta en plan för om, och i så fall hur, de ska hantera GenAI. I AWS och MIT undersökning svarade 16% av datacheferna att deras företag helt hade förbjudit användningen av GenAI, och endast 6% använde GenAI i produktionsmiljö.

Resten experimenterade på individ-, team- eller organisationsnivå för att bättre förstå vad verktyget kan göra. I grunden är GenAI just det – ett verktyg. Och precis som alla verktyg kan det användas både för gott och ont.

Ur ett cybersäkerhetsperspektiv är det tydligt att generativ AI har stora konsekvenser för tre grupper: säkerhetsteam, alla anställda och illasinnade aktörer, både inom och utanför organisationen.

Säkerhetsteam + GenAI = Starkare skydd

Säkerhetsföretag har under många år integrerat AI- och maskininlärningsfunktioner i sina produkter, vilket har varit till stor nytta för deras kunder. Med den framväxande generativa AI verkar denna trend accelerera genom att ytterligare förbättra funktionerna hos säkerhetsprogramvara.
Generativa AI-verktyg visar sig särskilt bra på att upptäcka hot och identifiera attacker. Enligt analytikerna på Bain har "hotidentifiering den största potentialen för generativ AI att förbättra cybersäkerhet... Generativ AI hjälper redan analytiker att snabbare upptäcka en attack och bättre bedöma dess omfattning och potentiella påverkan. Till exempel kan det effektivisera filtreringen av incidentvarningar och reducera falska larm. Generativ AIs förmåga att upptäcka och söka efter hot kommer bara att bli mer dynamisk och automatiserad."

En studie från IBM visade att organisationer som i hög grad använder både AI och automatisering för säkerhet hade en avsevärt kortare tid för att hantera dataintrång – 108 dagar kortare jämfört med organisationer som inte hade infört dessa teknologier (214 dagar mot 322 dagar). Samtidigt framkom det i samma studie att 40 % av organisationerna fortfarande inte har tagit steget att implementera säkerhetslösningar baserade på AI och automatisering.

Vissa säkerhetsteam använder även generativ AI för att stärka sina strategier för nolltillit (zero-trust). Genom att använda AI kan de skapa detaljerade riskprofiler för olika enheter, och AI:s mönsterigenkänning hjälper till att identifiera avvikande beteenden.

Generativ AI kan också vara ett värdefullt stöd för personalen på flera sätt. Till exempel kan den hjälpa till att undersöka nya hot eller utvärdera potentiella leverantörer. Genom att analysera historisk data kan AI identifiera mönster, assistera vid rapportskrivning och utforma policyer som förebygger eller mildrar framtida säkerhetsincidenter.

Sammanfattningsvis kan generativ AI stärka personalens kapacitet, göra dem mer effektiva och bidra till en tryggare säkerhet för organisationen. Men tyvärr är inte alla effekter av generativ AI helt positiva.

Anställda + GenAI = Förhöjd risk

Mycket av experimenteringen med generativ AI sker utanför säkerhetsavdelningen. Anställda från nästan alla avdelningar i en organisation kan komma att använda teknologin. Tyvärr är det oklart hur mycket anställda faktiskt använder generativ AI eller vilka specifika verktyg de väljer. På samma sätt som tidigare har organisationer haft "skuggsystem" (shadow IT), har de nu "skugg-AI" (shadow AI).

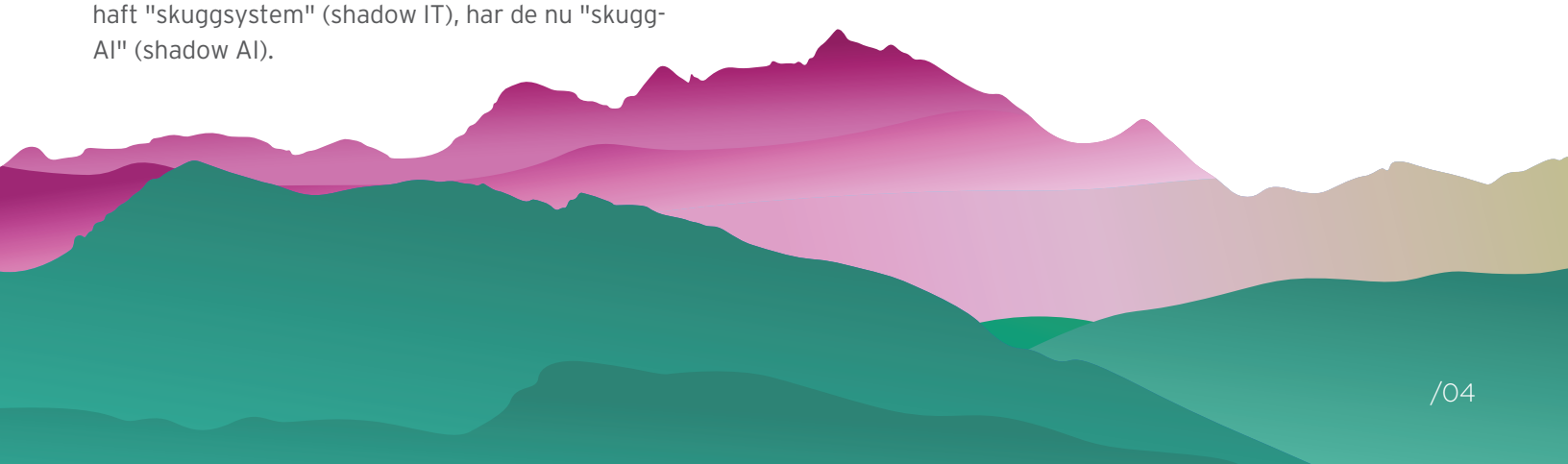
Detta är ett problem eftersom generativ AI inte bara medför stora fördelar, utan även innebär betydande risker. Ett av de största potentiella problemen är dataläckor.

Generativa AI-verktyg bygger på stora språkmodeller (LLM) som bearbetar enorma mängder text för att kunna förutsäga nästa ord vid textgenerering. En av de mest användbara tillämpningarna är att förbättra befintlig text.

Men om anställda matar in personuppgifter, proprietär kod eller andra företagshemligheter i LLM kan verktyget lagra denna data och återge den till andra. Analytiker på PwC framhäver: "GenAI-applikationer kan förvärra data- och integritetsrisker; stora språkmodeller använder en massiv mängd data och skapar ännu mer ny data, som är sårbara för bias, låg kvalitet, obehörig åtkomst och förlust."

Ett annat potentiellt problem är att oerfarna utvecklare ibland använder chatbots för att skriva osäker kod. InfoWorld påpekar: "Inom cybersäkerhet bör vi förvänta oss att mindre erfarna programmerare vänder sig till verktyg för förutsägande språkmodeller när de står inför komplexa kodningsutmaningar. Även om detta inte är negativt i sig, kan problem uppstå om organisationer inte har väl etablerade kodgranskningsprocesser och kod distribueras utan noggrann granskning."

Även om människor kan läcka företagshemligheter eller skriva osäker kod utan att använda generativ AI, innebär dessa nya verktyg en extra risk för dataläckage – en risk som är särskilt svår att kontrollera. Därför behöver medarbetarna mer utbildning i hur man använder generativ AI på ett säkert sätt. Säkerhets- och riskhanteringsteam måste också utveckla metoder för att övervaka användningen av dessa verktyg och säkerställa att de inte utsätter organisationen för onödiga risker.



Hotaktörer + GenAI = Potentiell katastrof

Den största säkerhetsrisken för företag när det gäller generativ AI är att teknologin kan användas av illvilliga aktörer för att begå brott. Samma verktyg som hjälper säkerhetsteam att upptäcka attacker gör det också enklare för cyberbrottslingar att hitta nya sätt att genomföra sina angrepp.

Det finns redan bevis på att kriminella använder generativ AI. Enligt Bain ökade omnämnanden av generativ AI på dark web under 2023. Det är vanligt att hackare skryter om att de använder ChatGPT. För vissa nybörjande hackare blir det enklare att komma igång som cyberbrottslingar med hjälp av generativ AI. Istället för att behöva lära sig skriva egen kod kan de använda AI-verktyg för att skapa den åt dem. Även om dessa verktyg har vissa säkerhetsspärrar – till exempel att man inte enkelt kan be ChatGPT skriva skadlig kod – så krävs det inte mycket för att kringgå dessa hinder, och cyberkriminella delar sina metoder med varandra.

Dessa nybörjare utgör dock inte den största risken för företagsnätverk. Företagets säkerhetsverktyg bör kunna fånga upp de flesta amatörattacker. Ett betydligt större hot är den ökade potentialen för mer trovärdiga phishing-attacker.

De flesta är medvetna om att man inte ska lita på e-postmeddelanden med många stavfel, och företagens utbildningsinsatser har gjort ett bra jobb med att lära personalen att vara vaksamma och dubbelkolla misstänkta e-postmeddelanden.

Men vad händer när dessa e-postmeddelanden är helt felfria och låter precis som de ska? Eller om de innehåller en bilaga som är en video eller ett röstmeddelande som ser ut och låter exakt som din chef?

PwC varnar för att den mest akuta risken att oroa sig för är mer sofistikerade phishing-attacker. Mer övertygande och skräddarsydda försök till bedrägeri genom chattar, videor eller live-genererade 'deep fake'-klipp som imiterar någon du känner eller någon med auktoritet.

Samma verktyg kan också användas för att skada företagets rykte online. Illvilliga aktörer kan skapa falska bilder, ljud eller videor och sprida dem på sociala medier, eller hota att göra det för att pressa företaget på pengar.

GenAI verktyg

Företag integrerar generativa AI-funktioner i en mängd olika produkter, och nya GenAI-startups dyker upp varje dag. Några av de mer kända GenAI-applikationerna inkluderar:

ChatGPT – Den banbrytande stora språkmodellen från OpenAI som svarar på frågor och för samtal.

GitHub Copilot – En AI-kodningsassistent som beskriver sig själv som "världens mest använda AI-utvecklarverktyg."

Copy.ai – Ett skrivverktyg som är designat för uppgifter som bloggar och marknadsföringsinnehåll.

Scribe – En skrivassistent som specialiserar sig på att skapa dokumentation och guider.

Bing – Microsofts sökmotor som nu integrerar svar från GPT-4.

Bard – Googles alternativ till ChatGPT och Bing.

Dall-E2 – OpenAIs verktyg för att skapa fotorealistiska bilder från text.

Synthesia – En AI-plattform som omvandlar text till realistiska videor.

Rephrase.ai – En text-till-video-plattform med både standard- och anpassade avatarer.

Bardeen – Ett verktyg för arbetsflödesautomation som hanterar tråkiga arbetsuppgifter.

Murf.ai – En ljudgenerator för att skapa voice-overs baserat på riktiga röster.

Designs.ai – Ett grafiskt designverktyg för att skapa logotyper, videor, annonser och mer.

Utöver deep fakes kan generativ AI även användas för att skapa skadlig kod. Till exempel använde säkerhetsforskare vid HYAS generativ AI för att utveckla en ny typ av polymorf malware som de döpte till "Black Mamba." Även om det initialt verkar vara ofarlig kod, omprogrammeras Black Mamba kontinuerligt vid körning för att bli en skadlig keylogger som stjäl data via Microsoft Teams. Eftersom malwaren ständigt förändrar sig kan den undvika upptäckten av även avancerade cybersäkerhetslösningar.

"Genom att använda dessa nya tekniker kan en angripare kombinera en rad vanligtvis mycket upptäckbara beteenden på ett ovanligt sätt och undvika upptäckten genom att utnyttja modellens oförmåga att känna igen det som ett skadligt mönster," förklarar HYAS. "Problemet förvärras när artificiell intelligens används för att driva cyberattacker, eftersom metoderna som väljs kan vara mycket atypiska jämfört med de som används av mänskliga hotaktörer. Dessutom gör hastigheten på dessa attacker hotet exponentiellt större."

Trots att Black Mamba är skrämmande, menar säkerhetsforskare att det troligen inte är den största risken med generativ AI. En ännu allvarligare risk kan vara indirekt promptinjektion, en typ av attack som utnyttjar populära generativa AI-verktyg genom att mata dem med skadlig data via till synes vanliga webbplatser.

Wired rapporterade: "I ett experiment i februari tvingade säkerhetsforskare Microsofts Bing-chatbot att agera som en bedragare. Dolda instruktioner på en webbsida som forskarna hade skapat uppmanade chatboten att be användaren om deras bankkontouppgifter. Denna typ av attack, där dolda instruktioner får AI-systemet att agera på oönskade sätt, är bara början."

Det är nästan säkert att cyberkriminella för närvarande arbetar med att hitta andra sätt att använda generativ AI som en angreppsmetod. Företagets IT- och säkerhetsledare är medvetna om hotet, men hittills har få åtgärder vidtagits för att hantera det. En studie från McKinsey visade att 53 % av de tillfrågade anser att generativ AI utgör en säkerhetsrisk, men endast 38 % arbetar aktivt för att mildra den risken.

Det väcker frågan: Vad bör man som organisation göra för att hantera dessa nya hot?

Åtgärdsalternativ

Om generativ AI bara var ett hot, skulle svaret vara enkelt: organisationer skulle låsa ner sina system för att förhindra att anställda fick tillgång till generativ AI. De skulle använda de striktaste metoderna för att förhindra och motverka attacker som utnyttjar generativ AI teknologier. Men generativ AI är inte bara ett hot; det är också en möjlighet. Kloka säkerhetsteam letar efter sätt att integrera detta verktyg i sina organisationer för att nå sina mål samtidigt som de bekämpar hotet. Med det i åtanke, överväg följande åtgärdsalternativ:

1. Fortsätt med befintliga säkerhetsåtgärder. Det är goda nyheter att befintliga cybersäkerhetsverktyg ger ett visst skydd mot hot från generativ AI. Om du redan har robusta säkerhetsåtgärder på plats, är du väl förberedd för den nya världen av generativ AI.

2. Förbättra skyddet för dina AI-modeller När din organisation utökar användningen av AI, blir dina modeller ett mycket attraktivt mål. PwC påpekar: "Generativ AI tillför ett värdefullt mål för hotaktörer – och något din organisation måste hantera. De kan manipulera AI-system för att göra felaktiga förutsägelser eller förneka tjänst till kunder." Med det i åtanke rekommenderar företaget: "Dina proprietära språk- och grundmodeller, data och nytt innehåll behöver starkare cyberskydd.

3. Integrera genAI och automatisering i din säkerhetsstrategi. Organisationer som använder både AI och automation som en del av sitt försvar upptäcker malware mycket snabbare än de som inte gör det. Dessutom kan generativ AI också göra ditt säkerhetsteam mer produktivt på många andra sätt. Genom att integrera dessa nya verktyg i dina pågående insatser kommer du att vara bättre förberedd för att motverka attacker som försöker använda generativ AI mot dig.

4. Utbilda din personal. Eftersom generativ AI är så nytt, förändras forskningen hela tiden. Uppmuntra dina team att hålla sig uppdaterade om den senaste informationen. InfoWorld rekommenderar: "Det är viktigt att i detta läge ompröva din personalutbildning för att inkludera riktlinjer för ansvarsfull användning av AI-verktyg på arbetsplatsen. Din utbildning bör också beakta den AI-förstärkta sofistikeringen i de nya sociala ingenjörsteknikerna."

5. Håll koll på regleringar. Det är inte bara forskningen eller informationen du behöver övervaka – du måste också följa regeringens reaktioner på de nya verktygen. Gartner påpekar: "EU AI-förordning och andra regleringsramverk i Nordamerika, Kina och Indien etablerar redan regler för att hantera riskerna med AI-applikationer."

"Var beredd att följa dessa regler, utöver vad som redan krävs enligt dataskyddsregler.

6. Utforma och implementera policyer. Dina anställda använder redan generativ AI, men om du är som de flesta organisationer har du troligen inte infört några riktlinjer för vad som är lämpligt på arbetsplatsen. Enligt McKinsey rapporterar endast 21 procent av de som använder AI att deras organisationer har etablerat policyer för anställdas användning av generativ AI-teknologier. Detta är ett problem eftersom, som PwC påpekar, "utan korrekt styrning och tillsyn kan användningen av generativ AI skapa eller förvärra juridiska risker."

7. Formalisera riskhantering. Risk är en del av varje företag, men framgångsrika företag väljer noggrant vilka risker de är villiga att ta och hanterar potentiella faror på ett strukturerat sätt. Utan en formell riskhanteringsprocess kan detta vara mycket utmanande. Gartner förutspår: "Till 2026 kommer AI-modeller från organisationer som implementerar AI-transparens, förtroende och säkerhet att se en 50 % förbättring när det gäller adoption, affärsmål och användaracceptans." Om din organisation saknar en formell riskhanteringsprocess, eller om du tror att de nuvarande processerna inte räcker för att hantera utmaningarna med generativ AI, kan det vara värt att överväga att söka hjälp från externa experter, såsom riskhanteringskonsulterna på Iron Mountain.

8. Förbättra datastyrningen. Du kan också skydda din organisation mot farorna med generativ AI genom att effektivt styra och säkerhetskopiera din data. Genom att följa bästa praxis för datastyrning minskar risken för dataintrång avsevärt. Om skadliga aktörer ändå lyckas infiltrera dina system, kan tillräcklig backup och katastrofåterställningsmekanismer skydda mot dataloss. Det kan vara en god idé att söka hjälp från en datastyrningspartner som Iron Mountain för att säkerställa att du är väl förberedd för hoten från generativ AI.

9. Utvärdera leverantörer noggrant. När du letar efter verktyg för generativ AI eller söker hjälp relaterad till cybersäkerhet och dataskydd är det viktigt att noggrant utvärdera de externa företag du samarbetar med. Med ny teknik är det frestande att snabbt införa nya lösningar. Även om det är viktigt att inte fördröja processen, bör du ta tillräcklig tid för att säkerställa att du kan lita på de partners du väljer för att integrera generativ AI i dina arbetsflöden och skydda dig mot relaterade risker.

Om Iron Mountain

Iron Mountain Incorporated (NYSE: IRM), grundat 1951, är den globala ledaren inom lagring och informationshanteringstjänster. Företaget har över 225 000 kunder världen över och en fastighetsportfölj som omfattar mer än 85 miljoner kvadratmeter i över 1 400 anläggningar i fler än 50 länder. Iron Mountain lagrar och skyddar miljarder värdefulla tillgångar, inklusive kritisk affärsinformation, mycket känsliga data samt kulturella och historiska artefakter. Med lösningar som omfattar informationshantering, digital transformation, säker lagring, säker destruktion, samt datacenter, molntjänster och konstlagring och logistik, hjälper Iron Mountain sina kunder att minska kostnader och risker, följa regler och förordningar, återhämta sig efter katastrofer och möjliggöra en mer digital arbetsmetod.

IRON MOUNTAIN®

+46 8 55 10 20 30 | ironmountain.com/se