DELL Technologies

5

Rekommendationer för att maximera GenAI på ett säkert sätt

1	2	3	4	5
Säkra lagren i ett GenAI-system	**Använd nollförtro-endeprinciper**	**Upprätthåll styrning och mänsklig tillsyn**	**Dra nytta av GenAI-säker-hetsverktyg när de blir tillgängliga**	**Innovera med gott självförtroende**
Infrastruktur	Lita aldrig på något, verifiera alltid	Engagera viktiga intressenter	Innehåll	Sträva efter att cybersäkerhet ska underlätta uppdraget, inte hämma det
OS och Kubernetes	Minsta möjliga åtkomstbehörighet	Upprätta policyer för etik- och regelefterlevnad, datahantering	Förutsägelse av risker	Låt cybersäkerhets-mognad bygga upp organisationens självförtroende för innovation
GenAI-program	Systemhärdning	Övervaka och utkräv ansvar	Kunskap och automatisering	
Data	Identitetshantering	Övning och utbildning		
	Segmentering			
	Loggning, övervakning och granskning			

Generativ AI-teknik utlovar transformativ kapacitet men medför unika säkerhetsutmaningar.

Generativ AI revolutionerar företag som aldrig förr, driver på innovation och erbjuder oöverträffade fördelar som ger en konkurrensfördel. Även om den här tekniken har potential till stora omdaningar har den också egna säkerhetsutmaningar.

Dells ämnesexperter Steve Brodson, Services Product Manager, och Eitan Lederman, Cybersecurity Consultant, slog sig ihop med Chris Cicotte från APEX- och AI-marknadsföringsteamet för att ta itu med dessa problem och diskutera hur man kan maximera GenAI på ett säkert sätt. Läs vidare för en sammanfattning av samtalet och ytterligare insikter om ämnet och se hela diskussionen på dell.com/cybersecuritymonth.

> Det handlar om att utbilda människor. Människor måste veta hur man använder GenAI-systemet. Vad man kan göra, men också vad man inte ska göra.”
>
> ***Eitan Lederman***
> *Dell Cybersecurity Consultant*

Säkra lagren i ett GenAI-system

Även om GenAI är en relativt ny teknik är de flesta säkerhetsprotokoll samma etablerade cybersäkerhetstekniker som används för att skydda andra arbetsbelastningar.

Infrastruktur – fokusera på att minimera attackytan:

- Sårbarhets- och intrångstester
- Korrigering
- Härdning
- Identitetshantering, inklusive starka lösenord, flerfaktorautentisering (MFA)
- Övervakning och granskning
- Säkerställa att leverantörskedjan från tredje part är säker

OS och Kubernetes – även fokus på minskning av attackytan, inklusive:

- Sårbarhetsavsökning
- Regelbunden korrigering
- Uppdatering av Kubernetes-komponenter
- Begränsa åtkomstkontroll baserat på identitetshantering, rollbaserad åtkomst (RBAC) och åtkomst med lägsta behörighet
- Skydda kontrollplanet, inklusive API-servern, hemligheter, kubelet och andra komponenter
- Använda namnrymder

GenAI-program – implementera säkerhetsåtgärder riktade mot de nya attackytorna som skapas av GenAI:

- Identitetshantering för att hantera promptinmatning, avslöjande av känslig information, modellstöld, förgiftning av övningsdata
- Validering av datakälla för att skydda mot förgiftning av övningsdata, modellpartiskhet
- Övervakning och granskning för att identifiera och förhindra DOS (Denial of Service) i modellen, modellstöld, avslöjande av känslig information, avvikelsedetektering, kriminalteknik

Data – införliva starka dataskyddsåtgärder för att skydda data i språkmodeller och program:

- Avgränsat cybervalv
- Kryptering
- Incidentresponsplan
- Övervakning och granskning av övningsdata och utdata

Se till att dataskyddsprinciper tillämpas på alla data, inklusive inlärningsdata, modellers utdata och alla data som ingår i Retrieval Augmented Generation (RAG), om de används. Säkerställ dessutom kontinuerlig överensstämmelse med alla tillämpliga dataskyddsregelverk.

Använd nollförtroendeprinciper

Rollen för flera nollförtroendeprinciper, till exempel identitetshantering, åtkomst med lägsta behörighet, systemhärdning och korrigering, har redan nämnts, vilket visar värdet av nollförtroendeprinciper för att skydda en GenAI-arbetsbelastning. Nollförtroendearkitekturer kräver även kontinuerlig loggning, övervakning och granskning av nätverksaktivitet, vilket kan förhindra GenAI-specifika risker som resultatmanipulering och dataförgiftning.

DELL Technologies

Dessutom uppmuntrar nollförtroende till mikrosegmentering, vilket minskar effekten av ett intrång. Det kräver också datakryptering, både under överföring och i vila, vilket är en viktig del av den övergripande dataskyddsstrategin.

Även om det här bara är några av de sätt som nollförtroende kan skydda en GenAI-arbetsbelastning på bör det anses vara bästa praxis att använda nollförtroendeprinciper.

Upprätthåll styrning och mänsklig tillsyn

En stor del av värdet i GenAI ligger i att automatisera uppgifter som människor normalt skulle utföra, men mänsklig styrning är avgörande för att säkerställa säkerhet och korrekt funktion hos programmen. En styrningsmodell involverar vanligtvis viktiga intressenter i hela organisationen som fastställer riktlinjer och krav för etisk efterlevnad och regelefterlevnad, principer och procedurer för datahantering och i som slutändan utkräver ansvar.

Lämplig styrning och tillsyn kan hjälpa till att lösa problem som övertro på modellen, partiskhet, resultatmanipulering, avslöjande av känslig information och dataförgiftning.

Lederman påpekade också vikten av utbildning: ”Det handlar om att utbilda människor. Människor måste veta hur man använder GenAI-systemet – vad man kan göra, men också vad man inte ska göra.”

Utöver den risk som en organisations GenAI-program utgör ökar även förekomsten av cyberattacker som möjliggjorts av GenAI, och som ofta kräver mänsklig inblandning. Några exempel är illvilliga aktörer som använder deepfakes för att manipulera mänskligt beteende och nätfiskeattacker som görs mycket effektivare genom att mer exakt efterlikna en människas skriv- eller talstil. Fortlöpande övning och utbildning är några av de mest effektiva sätten att ta itu med dessa risker, vilket återigen stärker den mänskliga faktorn.

Dra nytta av GenAI i säkerhetsverktyg när de blir tillgängliga

Även om mycket fokus ligger på risker har GenAI också potential att stärka säkerhetsarbetet. Dessa funktioner är ännu i sin linda men de erbjuder fördelar inom tre viktiga områden:

- **Innehåll: Generera säkerhetspolicy, anpassad utbildning, dataklassificering och rapportering**
- **Förutsägelse:** av risker och attackaktiviteter, förslag på åtgärder
- **Kunskap:** Ställa frågor till miljön (prata med systemet), kriminalteknik, automatisering

GenAI:s bidrag till säkerhetsverktyg kan hjälpa till att maximera säkerhetsteamens kapacitet, sänka kostnaderna och förbättra försvaret. Dra nytta av dessa lösningar när de utökas och mognar.

Innovera med gott självförtroende

Viktigast av allt är att inte låta säkerhetsrisker hindra dig från att utnyttja potentiellt revolutionerande teknik. Effektivitet, automatisering, kostnadsreducering, problemlösning och ökad kreativitet är bara några av de sätt som GenAI kan omvandla verksamheten på.

Även om GenAI kräver robusta och ibland nya cybersäkerhetsåtgärder bör målet vara att underlätta organisationens uppdrag, inte hämma det. Att utveckla rätt cybersäkerhetsstrategi bör ge organisationer självförtroende att växa och förnya sig.

Lär dig att hantera några av dagens främsta cybersäkerhetsutmaningar på **dell.com/cybersecuritymonth**

DELL Technologies