DELL Technologies

5 Recomendações para maximizar a IA generativa com segurança

1	2	3	4	5
Proteja as camadas de um sistema de IA generativa	**Utilize os princípios do Zero Trust**	**Mantenha a governança e a supervisão humana**	**Aproveite as ferramentas de segurança da IA generativa à medida que elas se tornam disponíveis**	**Inove com segurança**
Infraestrutura	Nunca confie, sempre verifique	Envolva as principais partes interessadas	Sumário	Busque a segurança cibernética para viabilizar a missão, não impedi-la
Sistema operacional e Kubernetes	Acesso com privilégios mínimos	Defina políticas para conformidade com normas e ética e gerenciamento de dados	Previsão de riscos	Deixe que a maturidade da segurança cibernética crie confiança organizacional para inovar
Aplicativos de IA generativa	Fortalecimento do sistema	Monitore e reforce a responsabilidade	Conhecimento e automação	
Dados	Gerenciamento de identidade	Treinamento e educação		
	Segmentação			
	Registro, monitoramento e auditoria			

A tecnologia de IA generativa promete recursos transformadores, mas traz desafios de segurança únicos.

A IA generativa está revolucionando os negócios como nunca antes, impulsionando a inovação e oferecendo vantagens inigualáveis que fornecem uma vantagem competitiva. Embora essa tecnologia tenha potencial transformador, ela também vem com seu próprio conjunto de desafios de segurança.

Os especialistas da Dell, Steve Brodson, gerente de produtos de serviços e Eitan Lederman, consultor de segurança cibernética, juntaram-se a Chris Cicotte, da equipe de marketing de IA e do APEX, para abordar essas preocupações e discutir maneiras de maximizar a IA generativa com segurança. Leia a seguir um resumo da conversa e insights adicionais sobre o tema e assista à discussão completa em dell.com/cybersecuritymonth.

> "Trata-se de treinar as pessoas. Elas precisam saber como usar o sistema de IA generativa. O que fazer, mas também o que não fazer."
>
> ***Eitan Lederman***
> *Consultor de segurança cibernética da Dell*

Proteja as camadas de um sistema de IA generativa

Embora a IA generativa seja uma tecnologia relativamente nova, a maioria dos protocolos de segurança são as mesmas técnicas estabelecidas de segurança cibernética que são usadas para proteger outras cargas de trabalho.

Infraestrutura — concentre-se em minimizar a superfície de ataque:

- Testes de vulnerabilidade e penetração
- Aplicação de patches
- Fortalecimento
- Gerenciamento de identidade, incluindo senhas fortes, autenticação baseada em vários fatores (MFA)
- Monitoramento e auditoria
- Garantia de que a cadeia de suprimentos terceirizada esteja segura

Sistema operacional e Kubernetes — também um enfoque na redução da superfície de ataque, incluindo:

- Verificação de vulnerabilidades
- Aplicação regular de patches
- Atualização de componentes do Kubernetes
- Limitação do controle de acesso com base no gerenciamento de identidade, no acesso baseado em funções (RBAC) e no acesso com privilégios mínimos
- Proteção do plano de controle, incluindo o servidor de API, segredos, kubelet e outros componentes
- Uso de namespaces

Aplicativos de IA generativa — implemente ações de segurança direcionadas às novas superfícies de ataque criadas pela IA generativa:

- Gerenciamento de identidade para lidar com injeção imediata, divulgação de informações confidenciais, roubo de modelo, treinamento de envenenamento de dados
- Validação da fonte de dados para proteção contra envenenamento de dados de treinamento, viés de modelo
- Monitoramento e auditoria para identificar e prevenir DOS de modelo, roubo de modelo, divulgação de informações confidenciais, detecção de anomalias, perícia

Dados — incorpore fortes medidas de proteção de dados para proteger os dados no modelo de linguagem e no aplicativo:

- Cofre cibernético com air gap
- Criptografia
- Plano de resposta a incidentes
- Monitoramento e auditoria dos dados e resultados do treinamento

Garantir que os princípios de proteção de dados sejam aplicados a todos os dados, inclusive entradas de treinamento, saídas de modelo e quaisquer dados envolvidos na Geração aumentada de recuperação (RAG), se usados. Além disso, garanta a conformidade contínua com todas as normas de proteção de dados aplicáveis.

Utilize os princípios do Zero Trust

A função de vários princípios de Zero Trust, como gerenciamento de identidade, acesso com privilégios mínimos, fortalecimento do sistema e aplicação de patches, já foi mencionada, indicando o valor dos princípios de Zero Trust na proteção de uma carga de trabalho de IA generativa. As arquiteturas Zero Trust também exigem registro, monitoramento e auditoria contínuos da atividade da rede, o que pode prevenir riscos específicos da IA generativa, como manipulação de resultados e envenenamento de dados.

DELL Technologies

Além disso, o Zero Trust também incentiva a microssegmentação, o que reduz o impacto de uma violação. Ele também requer criptografia de dados, em trânsito e em repouso, que é uma parte importante da estratégia geral de proteção de dados.

Embora essas sejam apenas algumas das maneiras pelas quais o Zero Trust pode proteger uma carga de trabalho de IA generativa, a adoção de princípios de Zero Trust deve ser considerada uma prática recomendada.

Mantenha a governança e a supervisão humana

Grande parte do valor da IA generativa está em automatizar tarefas que os humanos normalmente executariam, mas a governança humana é essencial para garantir a segurança e o funcionamento adequado dos aplicativos. Um modelo de governança normalmente envolve as principais partes interessadas em toda a organização, que definem diretrizes e requisitos para conformidade ética e com normas, políticas e procedimentos de gerenciamento de dados e, em última análise, reforçam a responsabilidade.

A governança e a supervisão adequadas podem ajudar a resolver problemas como excesso de confiança do modelo, viés, manipulação de resultados, divulgação de informações confidenciais e envenenamento de dados.

Lederman também ressaltou a importância do treinamento: "Trata-se de treinar as pessoas. Elas precisam saber como usar o sistema de IA generativa — o que fazer, mas também o que não fazer".

Além do risco representado pelos aplicativos IA generativa de uma organização, há também a proliferação de ataques cibernéticos habilitados pela IA generativa, que geralmente exigem intervenção humana. Exemplos incluem agentes mal-intencionados que usam deepfakes para direcionar o comportamento humano e ataques de phishing que se tornam muito mais eficazes ao imitar com mais precisão o estilo de escrever ou falar de um humano. A formação e a educação contínuas são algumas das formas mais eficazes de enfrentar estes riscos, reforçando mais uma vez o elemento humano.

Aproveite a IA generativa nas ferramentas de segurança à medida que elas se tornam disponíveis

Embora grande parte do foco esteja no risco, a IA generativa também tem o potencial de reforçar os esforços de segurança. Embora esses recursos estejam em sua fase inicial, eles oferecerão benefícios em três áreas principais:

- **Conteúdo: geração de política de segurança, treinamento personalizado, classificação de dados e geração de relatórios**
- **Previsão:** de risco e atividade de ataque, sugestão de ações de correção
- **Conhecimento:** consulta do ambiente (conversando com o sistema), perícia, automação

A contribuição da IA generativa para as ferramentas de segurança pode ajudar a maximizar a capacidade das equipes de segurança, reduzir custos e aprimorar as defesas. Aproveite essas soluções à medida que elas crescem e amadurecem.

Inove com segurança

Mais importante, não deixe que os riscos de segurança o impeçam de aproveitar uma tecnologia potencialmente revolucionária. Eficiência, automação, redução de custos, solução de problemas e estímulo à criatividade são apenas algumas das maneiras pelas quais a IA generativa pode transformar os negócios.

Embora a IA generativa exija medidas robustas e, às vezes, novas de segurança cibernética, o objetivo deve ser viabilizar a missão da organização, não impedi-la. Desenvolver a estratégia certa de segurança cibernética deve dar às organizações confiança para crescer e inovar.

Saiba como lidar com alguns dos principais desafios de segurança cibernética atuais em **dell.com/cybersecuritymonth**

DELL Technologies