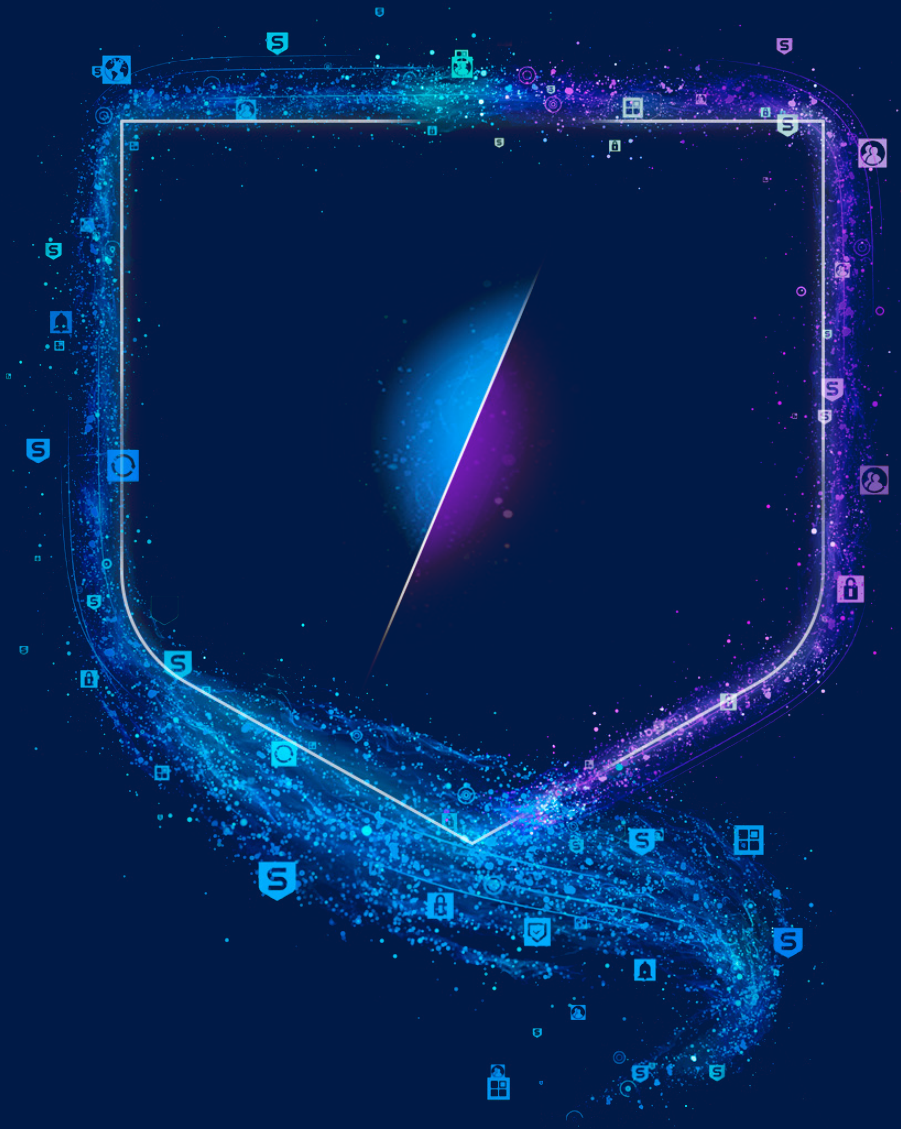
SOPHOS

Navegando pelo hype da IA na segurança cibernética

Como aproveitar a inteligência artificial com segurança para reforçar as defesas cibernéticas da sua organização



Índice

Introdução	3
Os benefícios da IA para a segurança cibernética	4
Taxas de adoção da IA	6
GenAI: Grandes expectativas	7
Os riscos da IA para a segurança cibernética	8
Etapas práticas para navegar pelo hype da IA	11
Conclusão	13
Sobre a pesquisa	13
Sobre a Sophos	13

Introdução

A segurança cibernética foi tomada pelo hype da IA. As organizações estão sendo bombardeadas com promessas sedutoras de transformação na segurança cibernética através do uso da inteligência artificial, como aumento da proteção, diminuição de custos, redução no quadro de funcionários especializados — além de previsões assustadoras de que a IA está nos levando a uma nova era de ataques cibernéticos.

Este guia foi criado para ajudar as organizações a navegarem por esse mar de ideias equivocadas sobre a inteligência artificial na segurança cibernética. Nele, esclarecemos o que a IA pode (e não pode) fazer para elevar as defesas cibernéticas corporativas e exploramos a segurança cibernética e os riscos operacionais que a IA pode impor. O guia também oferece orientações sobre como mitigar esses riscos e beneficiar-se da inteligência artificial de forma segura para aumentar a proteção virtual e o retorno sobre o investimento.

No decorrer, esse guia expõe dados e fatos sobre o verdadeiro uso da IA, suas expectativas e considerações com base nos resultados de uma pesquisa independente com 400 líderes de TI e segurança cibernética realizada no fim de 2024. Essas perspectivas da linha de frente oferecem um contexto valioso e atuam como um ponto de comparação de uso prático para as organizações explorarem sua posição frente à inteligência artificial. Para conhecer os resultados completos, consulte Beyond the hype: The business reality of AI for cybersecurity

No final, com ou sem a inteligência artificial, o objetivo permanece o mesmo: encontrar o nível ideal de resiliência necessária para a sua organização atingir o sucesso nos negócios minimizando as despesas. Ou, em outras palavras, fazer o melhor uso do (invariavelmente limitado) orçamento de segurança cibernética para dar suporte aos negócios. Este guia irá ajudá-lo a operar na era da IA.

Os benefícios da IA para a segurança cibernética

IA é um acrônimo que abrange uma série de recursos que dão suporte e aceleram a segurança cibernética de formas variadas. Um fato de apreço é que a IA oferece mais vantagens às defesas do que aos adversários. Duas abordagens comuns da inteligência artificial usadas na segurança cibernética são os modelos de deep learning e a IA generativa.

Deep learning

Os modelos de deep learning (DL) APLICAM aprendizados para desempenhar tarefas. Eles podem acelerar a aplicação de conhecimentos muito além do que os humanos podem atingir. Por exemplo, modelos DL devidamente treinados podem identificar se um arquivo é maligno ou benigno em uma fração de segundo sem nunca ter visto o arquivo antes.

O DL é ideal para o desempenho de tarefas repetitivas em grande escala. Ele cria um modelo *estatístico* que visualiza novos itens com base na distribuição e tudo aquilo que aprendeu de seu imenso conjunto de dados de treinamento. Por exemplo, os modelos DL podem avaliar milhões de amostras de arquivos sem hesitar ao identificar se contêm um malware. Consequentemente, o DL é amplamente utilizado para elevar a capacidade de proteção em produtos de segurança cibernética.

Os modelos DL permitem que a defesa manipule com sucesso os grandes volumes de ameaças criadas pelos usuários usando a automação e o crime cibernético "as-a-service". Os modelos DL também podem ser atualizados e adaptados para acompanhar a evolução dos ataques, mantendo-os em dia com o ambiente de ameaças.

IA generativa

Os modelos de IA generativa (GenAI) assimilam as entradas e as utilizam para CRIAR um novo conteúdo. Exemplos de aplicações incluem:

- Criar um resumo em linguagem natural da atividade da ameaça até o momento e recomendar os próximos passos para o analista seguir
- Gerar insights sobre o comportamento do invasor examinando os comandos que criam detecções
- Habilitar os analistas para usar a pesquisa em linguagem natural em vez de consultas baseadas em código para investigar detecções suspeitas
- Priorizar a aplicação de patches com base na propensão de uma vulnerabilidade ser explorada

A GenAI é uma ferramenta poderosa para acelerar as operações de segurança. Ao fazer o tratamento de grande parte dos dados mais pesados, ela dá aos analistas o poder de tomarem decisões mais rápidas e perspicazes, permitindo que concentrem seu tempo naquilo que possa gerar um maior impacto. Dessa forma, a GenAI pode aliviar parte da pressão exercida nos analistas, reduzindo o risco de esgotamento e rotatividade dos funcionários. A GenAI também pode ajudar a superar a barreira tecnológica das operações de segurança, permitindo que analistas menos experientes possam contribuir positivamente e acelerar o desenvolvimento de suas habilidades.

O percurso à GenAI

A base da GenAI moderna é o transformer, uma rede neural de deep learning que aprende o contexto e a relação entre entradas (por exemplo, as palavras em uma frase) e utiliza esse aprendizado para criar saídas relevantes. Geralmente, os transformers são usados nas tarefas de processamento de linguagem natural (NLP), como, por exemplo, traduzindo textos e respondendo a perguntas. De fato, o T em ChatGPT significa Transformer.

Ainda que os transformers sejam amplamente utilizados na GenAI, nem todos os transformers são generativos. Por exemplo, BERT (Bidirectional Encoder Representations from Transformers) é uma estrutura de Machine Learning de código aberto para NLP que pode ler a entrada de texto de modo bidirecional, ou seja, da esquerda para a direita e da direita para a esquerda. Essa abordagem permite uma melhora contextual significativa no entendimento do texto sem rótulos. Temos usado o BERT na Sophos há muitos anos na identificação e defesa contra ataques de comprometimento de e-mail corporativo.

O que me serve nem sempre lhe serve

Os modelos de IA variam imensamente em tamanho. **Modelos imensos**, como o Microsoft Copilot e o Google Gemini, são grandes modelos de linguagem (LLMs) treinados em um conjunto de dados bastante extenso que pode desempenhar uma ampla gama de tarefas. Em contrapartida, normalmente os pequenos modelos são designados e treinados com dados bastante específicos para desempenhar uma tarefa única, como detectar URLs maliciosas ou executáveis. Ainda que sejam mais limitados em termos de escopo, os **pequenos modelos** apresentam vantagens em custo, velocidade e desempenho sobre os grandes modelos.

Limitações da IA

A inteligência artificial apenas não é a resposta, pelo menos quando se trata de um futuro iminente. A IA complementa, mas não substitui completamente a perícia humana. As ameaças são imensamente complexas, e a execução de operações de segurança eficientes exigem habilidade técnica e capacidade de aplicar insights no contexto organizacional. A inteligência artificial apenas não vai conseguir equiparar as empresas com as organizações criminosas altamente financiadas e competentes de hoje.

TIPO	**IA com Deep Learning** *Aplicar* Usa redes neurais artificiais para reconhecer padrões e tomar decisões de forma a imitar o cérebro humano. APLICA aprendizados para desempenhar tarefas. Exemplo: **Detectar URLs maliciosas** O modelo de IA é treinado para identificar sites maliciosos, permitindo que produtos de segurança bloqueiem o acesso a eles	**IA generativa** *Criar* Use a estrutura e o padrão dos dados existentes para CRIAR (gerar) conteúdo totalmente novo. Exemplo: **Resumo do caso de ameaça** O modelo de IA cria um resumo da atividade da ameaça e dá aos analistas as próximas etapas recomendadas
TAMANHO	**Imensos modelos de IA** Ferramentas multiuso que são treinadas com vastas quantidades de dados públicos disponíveis e que podem auxiliar em uma ampla variedade de tarefas. Exemplo: **Microsoft Copilot, Google Gemini**	**Pequenos modelos de IA** Modelos focados na saída são projetados, treinados e compilados para casos de uso específicos. Exemplo: **Modelo de detecção de malware do Android**

Taxas de adoção da IA

A inteligência artificial já está altamente incorporada à infraestrutura de segurança cibernética da maioria das organizações:

- 73% dizem que suas soluções de segurança cibernética incluem modelos de deep learning
- 65% dizem que suas soluções de segurança cibernética incluem recursos de IA generativa

As aplicações da IA na segurança cibernética não se limitam a fornecedores externos, e 34% das organizações já utilizam a GenAI internamente para elevar sua segurança cibernética para, por exemplo, ajudar a gerar e-mails de teste de phishing.

Em um curto espaço de tempo, a adoção da inteligência artificial será algo quase universal, com seus recursos despontando, hoje, na lista de requisitos de 99% (arredondados) das organizações quando selecionam uma plataforma de segurança cibernética:

- 57% dizem que os recursos de IA são essenciais/extremamente importantes
- 41% dizem que os recursos são importantes

Tomando por base esse nível de adoção e uso futuro, entender os riscos e as mitigações associadas da IA na segurança cibernética é uma prioridade para organizações de todos os tamanhos e foco comercial.

73%
Usam ferramentas de segurança cibernética com modelos de Deep Learning

65%
Usam ferramentas de segurança cibernética com recursos de GenAI

99%
Exigem recursos de IA ao escolher uma plataforma de segurança cibernética

GenAI: Grandes expectativas

O hype em torno da inteligência artificial generativa resultou em altas expectativas sobre como essa tecnologia pode melhorar a resposta da segurança cibernética. A pesquisa revelou qual o maior benefício que as organizações desejam que os recursos de GenAI nas ferramentas de segurança cibernética proporcionem, como mostra a tabela abaixo.

Benefício mais desejado da IA generativa
Respostas classificadas em primeiro lugar

1=	Proteção melhorada contra ameaças cibernéticas (20%)
1=	Melhor retorno do gasto com a segurança cibernética (ROI) (20%)
3	Impacto e eficiência aumentados dos analistas de TI (17%)
4	Confiança de que estamos acompanhando as inovações em segurança cibernética (15%)
5=	Maior tranquilidade de saber que a nossa organização está bem protegida contra ataques (14%)
5=	Redução no esgotamento dos funcionários; ou seja, automação de tarefas para liberar o tempo do funcionário de segurança cibernética (14%)

Quais são os benefícios, se algum, que você deseja que os recursos da IA generativa proporcionem nas ferramentas de segurança cibernética? Respostas classificadas em primeiro lugar (n=400)

A diversidade de respostas revela que não existe um desejo que se sobressaia entre os benefícios esperados da GenAI na segurança cibernética. Mesmo assim, o ganho mais almejado se relaciona à proteção cibernética melhorada e ao desempenho dos negócios (tanto financeiro quanto operacional). Os dados sugerem que a inclusão dos recursos de GenAI nas soluções de segurança cibernética oferecem maior tranquilidade e confiança de que a organização está em dia com os mais recentes recursos de proteção.

A classificação de redução no esgotamento dos funcionários em último lugar sugere que as organizações estão menos atentas ou menos preocupadas com o potencial da GenAI em dar apoio aos usuários. Com a escassez de funcionários de segurança cibernética, diminuir a rotatividade é uma área importante de foco e que a IA pode ajudar.

Proteção **melhorada** e aumento do **ROI** são os principais benefícios que as organizações almejam da GenAI

Os riscos da IA para a segurança cibernética

O uso da IA na segurança cibernética reflete dois lados de uma mesma moeda. Enquanto a IA oferece benefícios incríveis para a linha de defesa no combate aos adversários, ela também inflige uma série de riscos:

1. **Risco de ameaça:** o uso da IA nos ataques cibernéticos
2. **Risco de defesa:** baixa qualidade e má implementação da IA
3. **Risco operacional:** confiança excessiva na IA
4. **Risco financeiro:** baixo retorno sobre o investimento na IA
5. **Risco de sequestro:** o comprometimento dos modelos públicos de IA pelos adversários

1. Risco de ameaça: o uso da IA nos ataques cibernéticos

Mesmo com o frenesi sobre como a inteligência artificial está traçando um cenário de ameaças totalmente novo, a realidade é menos dramática. As conversas sobre a IA nos fóruns de crimes cibernéticos não são tantas assim e muitos dos agentes de ameaças se mantêm céticos sobre a inteligência artificial. Quando e onde observadas, as tentativas de desenvolver malwares, ferramentas de ataque e exploits usando a IA são normalmente primitivas e de baixa qualidade.

Da mesma forma como as organizações legítimas, os adversários estão, essencialmente, se aproveitando da IA para melhorar a qualidade de seu conteúdo e a eficiência de suas operações, apesar da diferença de objetivos. Para ver mais detalhes sobre o recente cenário de ameaças e os ataques via IA, leia o blog da Sophos.

Melhorar a qualidade do conteúdo

Um das aplicações mais rápidas, fáceis e acessíveis da IA nos ataques cibernéticos é aumentar a qualidade e a credibilidade dos e-mails de phishing e scams, fazendo com que mais vítimas caiam nas armadilhas de ataque.

A péssima gramática, os erros ortográficos e a má formatação que "entregam" o phishing clássico são facilmente eliminados com as ferramentas de IA. Um e-mail bem escrito para ser usado em campanhas de phishing pode ser criado por LLMs públicos em menos de um minuto. Da mesma forma, textos e mensagens de redes sociais elegantes e convincentes para levar os destinatários a clicar em links ou compartilhar informações pessoais estão acessíveis em qualquer idioma. Os LLMs também ajudam os invasores a incorporar informações distintas em seus ataques, aumentando ainda mais a propensão de as vítimas caírem em um golpe.

As ferramentas de IA generativa também abriram as portas para uma nova era de golpes que clonam funcionários de alto escalão para levar vítimas desavisadas a fazer transferências financeiras. A tecnologia de clonagem de voz avançou a ponto de — com treinamento suficiente — levar os adversários a enganarem as pessoas fazendo-as acreditar que estão falando com alguém de verdade. Nesses ataques de phishing de voz, ou "vishing", o vetor se fará passar por um líder sênior e telefonará para um funcionário para "pedir" que compre um cartão-presente, faça um pagamento bancário ou transfira um arquivo de modo ilícito.

Os adversários também estão usando a tecnologia deepfake alimentada por IA para clonar visualmente as pessoas em seus ataques. Vídeos deepfake têm sido usados para enganar funcionários desavisados, fazendo com que realizem pagamentos substanciais, e para ludibriar programas de reconhecimento facial, para que efetuem pedidos de empréstimo e abertura de contas bancárias.

Melhorar a eficiência operacional

Assim como muitas empresas usam os chatbots por IA para melhorar a experiência do usuário, os invasores também se aproveitam deles. Alguns agentes de ameaças usam os LLMs para melhorar os fóruns que frequentam criando chatbots e respostas automáticas. Em um exemplo compartilhado pelo Sophos X-Ops, o fórum XSS criou um chatbot dedicado a responder às perguntas dos usuários. O administrador anunciava (traduzido originalmente do russo):

"Nesta seção, você pode conversar com a IA (inteligência artificial). Faça uma pergunta e nosso robô de IA a responderá... Esta seção e o robô de IA são designados para solucionar problemas técnicos simples, para o deleite de nossos usuários [e] para familiarizá-los com as possibilidades da IA."

Criar e treinar modelos personalizados requer grande perícia em IA, que é cara e escassa. Algumas quadrilhas cibernéticas têm pessoal interno especializado em IA, já os agentes de ameaças normalmente utilizam os LLMs existentes em seus ataques em vez de criar os seus próprios modelos.

Nivelar o invasor

É importante contextualizar o uso da IA pelos adversários. A IA é apenas uma das muitas ferramentas nas mãos dos invasores. Os agentes de ameaças utilizam a automação e os modelos de cybercrime-as-a-service para aumentar a escala e a frequência de seus ataques já há muitos anos. Para muitas organizações, esses recursos terão um impacto bem maior na exposição ao risco do que a IA.

2. Risco de defesa: baixa qualidade e má implementação da IA

Como vimos, os modelos de IA já estão amplamente incorporados nas defesas cibernéticas das organizações. Ainda que suas intenções sejam indubitavelmente boas, a baixa qualidade e a má implementação dos modelos de IA podem introduzir inadvertidamente consideráveis riscos à segurança cibernética. A propensão dos modelos de IA introduzirem riscos depende de diversos fatores, como:

- **Qualidade dos dados com os quais os modelos são treinados**. A máxima "garbage in, garbage out" é particularmente relevante no caso da IA. Usar dados de baixa qualidade para treinar modelos abre a porta para a entrada de erros, e o uso de conjuntos de dados desbalanceados tem o potencial de distorcer a saída devido ao excesso ou escassez de representação de certas variáveis. Quanto maior a quantidade de dados de alta qualidade no treinamento, melhor a saída resultante.
- **Experiência das equipes que criam os modelos**. Compilar modelos eficientes de IA para a segurança cibernética exige grande compreensão de duas áreas independentes, mas complementares:
  - **Ameaças:** para identificar o que você quer que o modelo de IA faça, primeiro é preciso entender como os malwares e os adversários operam.
  - **IA:** uma vez identificado o que você quer que a IA faça, é necessário identificar e construir o modelo certo para atingir o objetivo.

  Para construir modelos eficientes de IA e que tenham um impacto substancial na segurança cibernética, é essencial que esses dois grupos trabalhem em estreita colaboração, contrabalançando suas experiências mutuamente.
- **Qualidade de desenvolvimento de produto e processo de implantação**. Em meados de 2024, a implantação de uma atualização de conteúdo defeituosa em um produto de segurança cibernética levou as empresas em todo o mundo a uma parada generalizada. Funcionalidades de IA testadas, avaliadas e implantadas precariamente têm o potencial de causar sérios danos, com o agravante de o problema ser difícil de identificar ou retificar.

A falsa sensação de segurança (cibernética)

As organizações estão cientes do risco do desenvolvimento e implantação precários da inteligência artificial nas soluções de segurança cibernética. A grande maioria (89%) dos profissionais de TI e segurança cibernética entrevistados disse estar preocupada com o potencial da falha dos recursos de IA generativa nas ferramentas de segurança cibernética prejudicar suas organizações, com 43% se dizendo extremamente preocupados e 46% relativamente preocupados.

Portanto, não surpreende que 99% (arredondados) das organizações digam que, ao avaliar os recursos de GenAI nas soluções de segurança cibernética, elas também avaliem a capacidade dos processos e controles da segurança cibernética usados no desenvolvimento da GenAI:

- 73% dizem que avaliam completamente a capacidade dos processos e controles da segurança cibernética
- 27% dizem que avaliam parcialmente a capacidade dos processos e controles da segurança cibernética

A alta porcentagem relatada daqueles que realizam uma avaliação completa pode parecer encorajadora, mas, na verdade, isso sugere que muitas organizações têm um ponto cego de grande peso nessa área.

Avaliar os processos e controles usados para desenvolver os recursos de GenAI exige transparência do fornecedor e um grau razoável de conhecimento em IA pelo avaliador. Infelizmente, ambos requisitos estão bastante escassos. Os provedores de soluções raramente disponibilizam seus processos finais de implantação de GenAI completamente, e as equipes de TI geralmente têm insights limitados sobre as boas práticas de desenvolvimento. Para muitas organizações, essa descoberta sugere que "eles não sabem o que não sabem".

3. Risco operacional: confiança excessiva na IA

A inteligência artificial toca praticamente todos os aspectos de nossas vidas cotidianas — desde encontrar o melhor caminho ao supermercado até recomendar filmes e programas de TV. Sua natureza pervasiva facilita criar-se certa dependência da IA com o pressuposto de que a inteligência artificial pode realizar determinadas tarefas melhor do que os humanos. Felizmente, a maioria das organizações está ciente disso e preocupa-se com as consequências do excesso de confiança na IA:

- 84% estão preocupados com a pressão resultante para reduzir o quadro de profissionais de segurança cibernética
- 87% estão preocupados com a falta de responsabilidade resultante pela segurança cibernética

Estar atento a esses riscos é o primeiro passo para mitigá-los. É importante lembrar que a IA é apenas uma ferramenta nas defesas cibernéticas de uma organização — ainda que seja uma parte valiosa do arsenal de segurança, ela nem sempre constitui a abordagem certa e raramente representa uma solução completa. Cada organização é diferente, e o uso da IA deve ser contextualizado à totalidade dos negócios e suas necessidades.

4. Risco financeiro: baixo retorno sobre o investimento na IA

Capacidade de GenAI de alto gabarito nas soluções de segurança cibernética são caras para desenvolver e manter. Os líderes de TI e segurança cibernética estão atentos às consequências desses gastos, com 80% dizendo que acreditam que a GenAI aumentará significativamente o custo de seus produtos de segurança cibernética.

Apesar dessas expectativas de aumento de preços, a maioria das organizações vê a GenAI como um caminho para reduzir as despesas gerais com a segurança cibernética, com 87% dos entrevistados dizendo que estão confiantes de que os custos da GenAI nas ferramentas de segurança cibernética serão totalmente compensados pela economia que ela proporciona.

Ao mesmo tempo, as organizações reconhecem o desafio que há em quantificar esses custos. Normalmente, as despesas com GenAI são incorporadas no preço total dos produtos e serviços de segurança cibernética, dificultando identificar o quanto que as organizações estão gastando com a GenAI na segurança cibernética. Como reflexo dessa falta de visibilidade, 75% concordam que esses custos são difíceis de mensurar (39% concordam totalmente, 36% concordam relativamente).

Sem um relatório eficaz, as organizações correm o risco de não ver o retorno desejado de seus investimentos em inteligência artificial na segurança cibernética ou, pior ainda, direcionar investimentos para a área de IA que poderiam ter sido melhor aproveitados nalgum outro lugar.

5. Risco de sequestro: comprometimento de grandes modelos de linguagem (LLMs)

Os riscos da IA para a segurança cibernética se estendem além das ferramentas de segurança cibernética e aplicações. A rápida expansão global do uso do LLM público abriu as portas para agentes sofisticados comprometerem os próprios modelos para ajudar a atingir seus objetivos. Isso tem o potencial de acabar com o jogo de diferentes maneiras, incluindo:

- **envenenamento de dados**. Em seu estudo de 2023, Poisoning Web-Scale Training Datasets is Practical, Carlini et. al. mostraram que o envenenamento de dados (ou seja, a manipulação de dados em que o modelo é treinado para influenciar suas saídas) é um risco de ameaça viável.
- **Backdoors para estado-nação**. Muitos estados-nação têm condições para criar poderosos LLMs. Ao adicionar backdoors secretos e tornar seus modelos facilmente disponíveis para o uso pelo público em geral, os hackers de estado-nação podem manipular o LLM em seu benefício, se necessário.
- **Spoofing de LLM**. Hackers mal-intencionados podem comprometer LLMs legítimos (por exemplo, adicionando backdoors) e promover as mudanças como "melhorias". Para ludibriar as pessoas a usar a ferramenta comprometida, eles falsificam o nome de um provedor respeitado — por exemplo, omitindo uma letra do nome ou trocando a letra O pelo número 0.

Para saber mais sobre o comprometimento de um LLM, leia a última pesquisa da equipe Sophos AI.

Etapas práticas para navegar pelo hype da IA

Embora a IA traga riscos, seguindo uma abordagem discernente, as organizações podem se beneficiar da inteligência artificial de forma segura, aproveitando-se das vantagens da IA para avançar suas defesas cibernéticas. Muitos dessas recomendações também podem ser usadas para auxiliar na implementação de sucesso da IA em outras áreas.

Risco de ameaça: elevar as defesas cibernéticas para nivelá-las com a era da IA

Um ponto-chave de foco deve ser a melhoria da resiliência contra as ameaças alimentadas por IA. Dado que os adversários estão essencialmente utilizando-se da IA para aumentar a qualidade e a credibilidade dos e-mails de phishing e scams, faz sentido focar nestas áreas. Nossas sugestões incluem:

- **Elevar a proteção de e-mail**. Busque soluções que possam detectar e-mails de phishing e scams gerados por IA, evitando que atinjam a caixa de entrada dos usuários.
- **Implantar uma proteção contra Comprometimento de e-mail corporativo, proteção VIP**. Escolha soluções de segurança de e-mail que incluam proteção BEC e VIP, por exemplo, fazendo a varredura do conteúdo em busca de estilos para detectar golpes.
- **Ficar especialmente atento às redes sociais** – geralmente, os usuários não estão totalmente engajados quando navegam pelos canais sociais, deixando-os mais suscetíveis a cair em golpes.
- **Colocar em vigor processos para mitigar o risco da clonagem de voz**, tais como, procedimentos a seguir se um pagamento inesperado ou compartilhamento de arquivo lhe for solicitado. As opções incluem:
  - Ligar de volta para o solicitante para confirmar o pedido
  - Utilizar códigos de acesso ou frases de segurança

Risco de defesa: avaliar a qualidade da IA usada nos produtos de segurança cibernética

Fique atento aos riscos e impacto da IA de má qualidade a seus investimentos em segurança.
Pergunte aos fornecedores sobre:

- **Treinamento dos dados**. Qual é a qualidade, a quantidade e a fonte dos dados com os quais os modelos são treinados? Melhores entradas levam a melhores saídas.
- **Equipe de desenvolvimento**. Informe-se sobre as pessoas por trás dos modelos. Qual é o nível de especialização que têm em inteligência artificial? Qual é o nível de conhecimento que têm sobre ameaças, comportamento de adversários e operações de segurança?
- **Engenharia de produto e processo de implantação**. Quais são as etapas que o fornecedor segue ao desenvolver e implantar recursos de IA em suas soluções? Quais verificações e controles estão em vigor?

Por fim, pergunte-se: Qual o meu nível de confiança de que essa organização está fazendo um bom trabalho de IA e que aplica rigorosos controles de qualidade necessários à implantação?

Risco operacional: fitar a IA com um olhar humano

A inteligência artificial não se importa se você sofrer uma violação, mas o seu pessoal vai se importar. E se o pior acontecer e você for comprometido, você vai precisar de uma equipe experiente que possa entender e remediar a situação no contexto do seu negócio.

- **Manter as perspectivas**. A IA é apenas um item na caixa de ferramentas da defesa. Utilize-a, mas deixe claro que a responsabilidade pela segurança cibernética está nas mãos de humanos.
- **Não substitua, acelere**. A contínua escassez global de profissionais capacitados em segurança cibernética é bem conhecida. Sérios problemas de desgaste exacerbam o desafio. Em vez de buscar na IA uma forma de reduzir o quadro de funcionários, concentre-se em como a IA pode apoiar os seus trabalhadores. Ao cuidar das muitas operações de segurança repetitivas de níveis mais baixos e fornecer insights guiados, a IA pode:
  - Liberar o tempo para trabalhos mais valiosos e de maior impacto comercial
  - Reduzir a sobrecarga de alertas, ajudando a diminuir a fatiga
  - Acelerar o desenvolvimento profissional de analistas especializados
  - Permitir que analistas menos experientes desempenhem operações de segurança e desenvolvam um pipeline de recursos

Risco financeiro: aplicar rigor comercial às decisões de investimento em IA

Esta é uma da áreas mais fáceis para as organizações mitigarem, com vários fatores totalmente sob controle.

- **Definir metas**. Seja claro, específico e granular sobre quais os resultados que você deseja da IA.
  - Identifique o que você precisa. Quais são as lacunas? Em que a IA pode ajudar?
  - Considere os ganhos em tempo, proteção e financeiros.
- **Quantificar benefícios**. Entenda quanta diferença os investimentos em IA farão.
  - Se uma meta for diminuir os custos totais com a segurança cibernética, quantifique a economia resultante que você fará.
  - Se você deseja reduzir a rotatividade na TI e segurança cibernética, seja claro sobre como, exatamente, a ferramenta de IA vai impactar a equipe. Quais as tarefas que serão eliminadas da fila? Quantas horas serão liberadas?
- **Priorizar investimentos**. A inteligência artificial pode ajudar de muitas formas, e algumas terão maior impacto do que outras. Identifique as métricas importantes para a sua organização – economia financeira, impacto na rotatividade dos funcionários, redução da exposição, etc. – e compare como as diferentes opções se classificam.
- **Medir o impacto**. As decisões sobre investimentos são feitas com boas intenções. Certifique-se de analisar como o desempenho atual se relaciona com as expectativas iniciais. Você consegue ver as vantagens que espera? Você vê ganhos inesperados? E há áreas em que você não vislumbra os resultados esperados? Use esses insights para fazer os ajustes que sejam necessários.

Pergunte-se se a inteligência artificial é a melhor maneira de ajudar a atingir a sua meta, ou se outra tecnologia ou uma abordagem diferente teriam um impacto maior.

Risco de sequestro: mantenha-se alerta ao perigo

Este é o risco mais difícil para as organizações mitigarem. Ficar atento a esse risco já ajuda a reduzir o impacto. Dito isso, ao optar por LLMs públicos, busque:

- **Modelos de provedores conhecidos e de boa reputação**. Mesmo que não estejam imunes aos ataques de envenenamento de dados, os problemas com a saída de dados são mais prováveis de serem publicados e compartilhados.
- **Nome do provedor correto**. Os invasores falsificam o nome de provedores de reputação para que as pessoas pensem que seus modelos são legítimos.

Os especialistas em IA para segurança cibernética estão trabalhando ativamente para encontrar formas de neutralizar esse risco.

Conclusão

A inteligência artificial oferece grandes benefícios para a segurança cibernética. Ao evitar o hype da IA e adotar uma abordagem discernente focada em resultados, as organizações podem aproveitar essa tecnologia para enriquecer suas defesas cibernéticas e capacitar seus valiosos profissionais de TI e segurança cibernética.

Sobre a pesquisa

Fonte: Beyond the hype: The business reality of AI for cybersecurity

A Sophos contratou a Vanson Bourne, especializada em pesquisas independentes, para realizar um estudo com 400 líderes de TI e segurança cibernética em organizações com 50 a 3.000 funcionários. O estudo foi realizado durante o mês de novembro de 2024 e os entrevistados estavam distribuídos entre 13 setores diferentes. Para garantir uma ampla representação da indústria, a pesquisa foi totalmente desvinculada de fornecedores e as organizações dos entrevistados usavam soluções de segurança de endpoint de 19 fornecedores diferentes.

Sobre a Sophos

A Sophos é líder global em segurança cibernética e oferece um portfólio de produtos e serviços premiados de segurança cibernética, desde firewalls, proteção de endpoint e ferramentas EDR/XDR até serviços de detecção e resposta gerenciadas (MDR) e de resposta a incidentes (IR).

A Sophos tem enriquecido a segurança cibernética com inteligência artificial desde 2017, unindo as tecnologias IA e perícia humana para interromper uma amplitude cada vez maior de ameaças, onde quer que estejam. As funcionalidades de deep learning e IA generativa que resolvem os problemas mais críticos de clientes estão incorporadas em nossos produtos e serviços e trabalham através da maior plataforma de segurança de IA nativa do setor. Treinada com dados de ataques em mais de 600.000 ambientes de clientes diversificados, a nossa plataforma de IA adaptável proporciona proteção sem igual contra ameaças avançadas e melhora o poder das equipes de defesa.

Para saber mais e explorar as soluções Sophos, acesse www.sophos.com

2025-01-15 (WP-MP)

SOPHOS