Segurança de IA: sua vantagem competitiva na corrida da IA

Um guia para CxOs para acelerar a adoção da IA e gerenciar riscos

Introdução

À medida que as organizações se apressam para aproveitar o poder transformador da IA, a adoção frequentemente supera a segurança. Para cada iniciativa aprovada, existem inúmeras instâncias de IA oculta criadas por indivíduos e equipes. Na verdade, 85% dos tomadores de decisão de TI dizem que os funcionários estão adotando ferramentas de IA antes mesmo que suas equipes de TI possam avaliá-las.[1] E 93% dos funcionários admitem inserir informações em ferramentas de IA sem aprovação.[1]

Enquanto isso, os vetores de ataque nativos de IA estão aumentando, 47% a mais no último ano,[2] enquanto as ferramentas de segurança tradicionais têm dificuldade para acompanhar. Novas formas de exposição de dados e lacunas de conformidade desafiam as práticas de governança existentes. Os líderes executivos enfrentam uma pergunta crítica: como gerenciar os riscos introduzidos pela IA sem bloquear a inovação que ela possibilita?

Os requisitos são claros. As organizações precisam criar um ambiente no qual:

- Todo o uso de IA é conhecido
- Todas as comunicações de IA podem ser protegidas
- Todas as políticas de IA podem ser implementadas
- Todos os modelos de IA podem ser protegidos contra o uso indevido

O **Cloudflare AI Security Suite** oferece às organizações a confiança para avançar mais rapidamente com a IA, eliminando a incerteza em torno do risco. Nossa plataforma unificada protege todo o ciclo de vida da sua IA ao descobrir o uso de IA, aplicar o acesso Zero Trust e proteger endpoints de APIs e da web com defesa proativa contra ameaças. Com a governança de dados integrada, as equipes podem inovar livremente sem comprometer a segurança.

85%
dos tomadores de decisão de TI dizem que os funcionários estão adotando ferramentas de IA antes mesmo que suas equipes de TI possam avaliá-las.[1]

O desafio de segurança da IA

Os riscos da adoção da IA estão impulsionando um mercado em rápido crescimento para soluções de segurança de IA. Por exemplo, as plataformas de proteção de aplicativos nativas da nuvem (CNAPPs) se concentraram nos fluxos de trabalho de desenvolvimento de IA. Ofertas iniciais como estas são uma linha de defesa essencial, mas não são suficientes. As empresas precisam proteger todo o ciclo de vida da IA, incluindo a proteção dos sistemas de IA quando estes estiverem em execução na produção.

A dimensão completa do desafio de segurança da IA está se tornando cada vez mais evidente. Atualmente, os desenvolvedores estão criando recursos de IA, os funcionários estão utilizando ferramentas externas de IA e os clientes estão interagindo com aplicativos alimentados por IA. Proteger manualmente todos esses ambientes diferentes é complexo, e fazer isso de forma consistente é ainda mais difícil.

Prevenção do uso indevido de IA pela força de trabalho

À medida que a força de trabalho organizacional adota ferramentas de IA autorizadas e não autorizadas, dados e operações confidenciais ficam em risco. As equipes de segurança têm que lidar com:

- **Falta de visibilidade do uso de IA:** os líderes geralmente não têm uma visão completa de quais ferramentas de IA seus funcionários estão utilizando, onde dados confidenciais estão sendo processados e como esses sistemas de IA se conectam aos aplicativos da empresa. Esse ponto cego cria uma exposição significativa ao risco.
- **Riscos de segurança e conformidade de dados:** a IA altera a forma como os dados fluem em sua organização. Informações pessoais, dados proprietários e registros de clientes podem acabar em sistemas de IA de maneiras que desencadeiam violações de conformidade ou expõem a inteligência competitiva.
- **Controles de acesso para fluxos de trabalho de IA agêntica:** não é apenas o acesso humano às ferramentas de IA que deve ser gerenciado. O acesso da IA agêntica aos servidores MCP e a outros sistemas críticos também deve ser gerenciado. Isto requer novas abordagens para o gerenciamento de identidade e controles de acesso.

Proteção de aplicativos e modelos de IA voltados para o público

Quer sejam desenvolvidos internamente ou consumidos por meio de terceiros, os sistemas de IA estão se tornando um pilar essencial da experiência do cliente e do usuário, e, como tal, precisam ser protegidos. Os Top 10 do OWASP para aplicativos de LLM destacam ameaças que as ferramentas tradicionais não conseguem resolver:

- **Ataques de consumo ilimitado:** semelhantes aos ataques de DoS tradicionais, os ataques de consumo ilimitado buscam sobrecarregar os LLMs com solicitações que consomem muitos recursos. Os modelos de preços de nuvem com pagamento por uso podem aumentar o impacto financeiro desses ataques, enquanto usuários legítimos enfrentam degradação da qualidade do serviço.
- **Envenenamento de modelo:** invasores implantam backdoors, vieses ou vulnerabilidades em LLMs ao injetar dados corrompidos em conjuntos de dados ou repositórios públicos usados por desenvolvedores para treinamento de modelos. Um modelo envenenado dessa forma se comporta normalmente até que gatilhos específicos ativem comportamentos prejudiciais.
- **Ataques de injeção de prompts:** uma tática popular para exfiltração de dados por meio de interfaces de IA. A injeção de prompts manipula entradas de LLMs ao incorporar instruções maliciosas em prompts de usuários ou conteúdo externo, fazendo com que os modelos ignorem as instruções originais e executem comandos de invasores.
- **Jailbreaking:** prompts cuidadosamente elaborados, como cenários de role-playing, comandos de substituição de instruções e estratégias de várias rodadas, podem burlar as proteções de segurança dos LLMs para gerar conteúdo proibido ou extrair informações confidenciais.

Para permitir a rápida inovação em IA sem colocar a organização em risco, recomendamos que os líderes adotem uma abordagem de segurança de IA que abranja todo o espectro de requisitos:

- Proteção do uso da GenAI pela força de trabalho
- Proteção de aplicativos e cargas de trabalho com tecnologia de IA
- Criação de aplicativos com tecnologia de IA que são seguros por design

Proteger fluxos de trabalho de desenvolvimento e treinamento de IA

Projetos de IA envolvem conjuntos de dados gigantescos, recursos dispendiosos e experimentação iterativa, tudo isso cria novas superfícies de ataque e vulnerabilidades. As equipes de segurança devem abordar:

- **Segurança e integridade dos dados de treinamento:** dados de treinamento comprometidos podem introduzir vieses, backdoors ou vulnerabilidades que persistem nos modelos de produção. As organizações devem proteger o acesso aos conjuntos de dados de treinamento, evitar modificações não autorizadas e garantir a proveniência dos dados durante todo o ciclo de vida do modelo.
- **Gerenciamento de credenciais e segredos:** os fluxos de trabalho de desenvolvimento de IA exigem acesso a sistemas, incluindo armazenamento em nuvem para conjuntos de dados, clusters de computação para treinamento, registros de modelos e serviços de terceiros. Segredos ou chaves de API inadequadamente protegidos podem expor credenciais confidenciais, permitindo o acesso não autorizado a modelos proprietários, dados de treinamento ou sistemas de produção.
- **Controles de acesso ao ambiente de desenvolvimento:** os engenheiros de IA geralmente precisam de privilégios elevados para experimentar modelos e acessar dados confidenciais. Sem controles de acesso adequados, isso pode levar a ameaças internas, exposição acidental de dados ou extração não autorizada de modelos.

Unificar a segurança com a Cloudflare

A adoção bem-sucedida da IA se concentra em promover a produtividade, não em restringi-la. Quando as equipes podem usar a IA com confiança, sabendo que as salvaguardas adequadas estão implementadas, elas inovam mais rapidamente e assumem projetos mais ambiciosos.

O Cloudflare AI Security Suite cria um ambiente seguro para a inovação em IA. Ao unificar os recursos de segurança de aplicativos web e de serviço de acesso seguro de borda (SASE), os CxOs podem conectar e proteger dois domínios fundamentais:

- Aplicativos externos voltados para o público e habilitados para IA
- Sistemas e cargas de trabalho de IA internos e privados

Com foco em todo o ciclo de vida da IA, o Cloudflare AI Security Suite atende às necessidades de segurança, desde a descoberta e gerenciamento de riscos até a proteção de dados, protegendo o acesso para usuários e aplicativos habilitados por IA, bem como os fluxos de trabalho de desenvolvimento. Para fornecer uma camada essencial de segurança de produção em tempo real, a rede global da Cloudflare atua em linha para inspecionar e filtrar cada interação de IA, protegendo os dados de todos os usuários e aplicativos.

Em vez de apenas detectar problemas depois que eles acontecem, a Cloudflare evita os problemas e bloqueia ameaças antes que elas atinjam os modelos de IA. As equipes de segurança obtêm a visibilidade de que precisam para ficar à frente das ameaças emergentes.

Recursos básicos da Cloudflare

O Cloudflare AI Security Suite oferece monitoramento abrangente, proteção em tempo real e gerenciamento proativo de riscos, tudo em uma única plataforma, proporcionando uma abordagem holística para a segurança da IA.

Descoberta e visibilidade de IA abrangentes

A segurança eficaz da IA depende de um inventário completo e em tempo real de recursos e uso de IA, tanto autorizados quanto não autorizados. A base do Cloudflare AI Security Suite é o monitoramento contínuo e a descoberta automatizada para identificar todos os modelos de IA, assistentes, agentes e implantações de IA oculta em todos os tipos de ambiente: público, privado ou interno.

Gerenciamento proativo de riscos de IA

O Cloudflare AI Security Suite ajuda as organizações a evitar ataques, detectando e mitigando vulnerabilidades, configurações incorretas e vetores de ataque específicos de IA, incluindo aqueles listados nos Top 10 do OWASP para LLMs. A pontuação de confiança de aplicativos ajuda a priorizar a correção, para que as equipes abordem primeiro os riscos mais significativos.

Segurança de aplicativos para aplicativos com tecnologia de IA

Para ajudar os SecOps a se manterem atualizados sobre os vetores de ameaças de IA mais recentes, o Cloudflare AI Security Suite inclui detecção e mitigação proativas de ameaças para vulnerabilidades específicas de IA, configurações incorretas e vetores de ataque dentro dos pipelines de IA, incluindo proteção contra injeção de prompts, envenenamento de dados e violação de modelos.

Firewalls de IA especializados descobrem e rotulam IA generativa ou agêntica e endpoints de API, detectam tentativas de exfiltrar informações de identificação pessoal e bloqueiam prompts maliciosos antes que impactem o desempenho do modelo de IA ou contaminem o modelo com conteúdo tóxico ou desinformação.

Acesso Zero Trust para fluxos de trabalho de GenAI e IA agêntica.

Os princípios de Zero Trust, como o de menor privilégio, aplicam-se tanto à força de trabalho quanto aos agentes de IA. O Cloudflare AI Security Suite pode aplicar políticas de acesso à rede Zero Trust (ZTNA) para interações entre humanos e IA e entre IA e IA. O registro e os controles centralizados para servidores MCP garantem que a IA agêntica acesse apenas o que foi autorizado, incluindo fluxos de trabalho just-in-time.

Proteção de dados com reconhecimento de IA

Uma segurança de IA eficaz requer um recurso de prevenção contra perda de dados, um que utilize vários modelos de linguagem para entender o conteúdo de um prompt, bem como a intenção por trás dele. O Cloudflare AI Security Suite incorpora recursos de prevenção contra perda de dados (DLP) durante o treinamento, prompts e respostas para evitar a exposição de informações de identificação pessoal, vazamento de dados e acesso não autorizado em modelos e pipelines de IA. A segurança de tempo de execução implantada em linha e centrada em APIs atua como uma primeira camada de defesa rápida e simples, complementando a abordagem shift-left habilitada por uma CNAPP.

Localização de dados

Os LLMs e aplicativos de IA estão sujeitos às mesmas regulamentações que qualquer outro tipo de ambiente de dados. O Cloudflare AI Security Suite ajuda as organizações a garantir que as cargas de trabalho de IA respeitem os limites geográficos e jurisdicionais, com políticas que mantêm os dados de treinamento e as solicitações de inferência dentro das regiões aprovadas.

Segurança por design para o desenvolvimento de IA

Como em todos os tipos de software, a segurança para aplicativos de IA deve ser integrada desde o início do ciclo de desenvolvimento, e não adicionada posteriormente. O Cloudflare AI Security Suite fornece aos desenvolvedores ferramentas e estruturas para criar aplicativos com tecnologia de IA que são seguros por design.

O impacto do Cloudflare AI Security Suite nas empresas

A ascensão da IA não é apenas uma evolução, mas sim uma disrupção fundamental na mesma magnitude da industrialização ou da informatização. Isso faz com que a adoção rápida e eficaz seja uma questão não apenas de impulsionar o crescimento, mas também de viabilidade organizacional. A adoção lenta ou insegura agora representa uma ameaça existencial para organizações em todos os setores e indústrias.

O Cloudflare AI Security Suite possibilita uma transformação segura, controlada e eficiente para atender às crescentes expectativas dos clientes e às exigências do mercado em ambientes altamente competitivos.

- **Inovação de IA mais rápida:** quando a segurança permite o uso seguro, em vez de bloquear a adoção, os funcionários e as equipes podem explorar novos aplicativos de IA para se tornarem mais produtivos em seu trabalho diário. Frameworks de segurança adequados dão aos desenvolvedores a confiança necessária para criar recursos ambiciosos de IA sem colocar em risco dados ou sistemas confidenciais.
- **Risco relacionado à IA reduzido:** as organizações podem gerenciar todo o espectro de riscos que acompanham a adoção da IA, incluindo os riscos relacionados à IA inerentes a aplicativos web habilitados para IA. Aplicativos de IA em desenvolvimento podem ser protegidos para garantir que os funcionários não vazem dados confidenciais ou os exponham em conjuntos de treinamento de IA. O SecOps pode identificar e mitigar proativamente ameaças e vulnerabilidades específicas de IA, minimizando a superfície de ataque e protegendo dados e modelos críticos.
- **Operações de segurança otimizadas:** a visibilidade e o controle centralizados sobre sua postura de segurança de IA simplificam o gerenciamento e agilizam a resposta a incidentes. Sua equipe de SecOps pode se concentrar em iniciativas estratégicas, em vez de combater constantemente incidentes relacionados à IA.
- **Governança e conformidade de dados robustas:** os controles de proteção de dados específicos para IA ajudam você a proteger informações confidenciais e a atender aos requisitos regulamentares em constante mudança durante todo o ciclo de vida da IA.
- **Menor custo total de propriedade (TCO):** aproveitar uma plataforma consolidada que integra os investimentos existentes em segurança é mais econômico do que implementar soluções pontuais separadas para cada desafio de segurança de IA.

Casos de uso do Cloudflare AI Security Suite

O Cloudflare AI Security Suite é ideal para atender aos requisitos essenciais relacionados à adoção de IA.

- **Proteção do uso de ferramentas de IA pela força de trabalho:** aplicar políticas de Zero Trust para o acesso da força de trabalho a ferramentas públicas de IA generativa, como o ChatGPT, e aplicativos desenvolvidos internamente com tecnologia de IA.
- **Proteção de aplicativos habilitados por IA voltados para o público:** proteger aplicativos web e APIs que incorporam modelos de IA, como chatbots e mecanismos de recomendação, contra ataques que podem expor dados confidenciais ou violar o modelo.
- **Gerenciamento de IA oculta:** descobrir automaticamente ferramentas de IA não autorizadas em sua organização e aplicar controles adequados para permitir a inovação contínua dentro dos limites de risco gerenciados.
- **DLP com tecnologia de IA para interações de IA:** evitar a exposição de dados confidenciais em prompts ou respostas de IA, garantindo que as informações de identificação pessoal e confidenciais permaneçam protegidas.
- **Segurança no desenvolvimento com IA:** fornecer às equipes de engenharia estruturas que tornem o desenvolvimento seguro a abordagem padrão, permitindo que elas criem recursos de IA rapidamente sem comprometer a segurança.

Considerações sobre a implementação

Como uma estratégia de segurança fundamental, a segurança da IA deve ser abordada de forma ponderada.

Comece com sua infraestrutura atual	As implementações de segurança de IA mais bem-sucedidas se baseiam em ferramentas de segurança de aplicativos e SASE existentes, em vez de substituí-las. Essa abordagem aproveita os investimentos atuais ao mesmo tempo em que estende a proteção para abranger riscos específicos da IA.
Implante proteções em linha unificadas	A capacidade de bloquear atividades maliciosas na borda da rede, conforme ocorrem, é essencial para a segurança da IA. Implante controles em linha em tempo real e complemente-os com monitoramento baseado em APIs.
Assegure cobertura completa	Sua estratégia de segurança de IA deve abordar todo o espectro de requisitos: uso de ferramentas de GenAI pela força de trabalho, proteção para aplicativos e fluxos de trabalho com tecnologia de IA, proteção de fluxos de trabalho de IA agêntica e segurança para fluxos de trabalho de desenvolvimento com IA.
Planeje para escala empresarial	Escolha soluções que possam crescer com a sua adoção de IA. O que funciona para um programa piloto deve ser escalável para toda a organização à medida que o uso da IA se expande.
Valide seus critérios de sucesso	Antes de se comprometer totalmente, valide a solução em cenários do mundo real. Escolha uma plataforma que permita a ativação gratuita e por autoatendimento de seus recursos empresariais. Isso permite testar o pacote de segurança completo em pequena escala, como para uma única equipe ou aplicativo, para comprovar rapidamente seu valor e garantir que ele atenda aos seus critérios de sucesso.

Dar o próximo passo com o Cloudflare AI Security Suite

À medida que a adoção da IA acelera mundialmente, as organizações que conseguirem escalar a IA de forma segura terão uma vantagem competitiva significativa. A chave é reconhecer que a segurança da IA não é sobre impedir o uso da IA, mas sim sobre habilitá-la de forma inteligente.

O **Cloudflare AI Security Suite** permite que as organizações inovem com confiança. Construindo e ampliando nossas plataformas SASE e de segurança de aplicativos amplamente adotadas, o Cloudflare AI Security Suite fornece descoberta de IA integrada, controles de acesso Zero Trust, defesa proativa contra ameaças orientada por inteligência e governança de dados robusta. Ao proteger o uso e os modelos de IA de todos os ângulos, as organizações podem capacitar os desenvolvedores a criar mais rapidamente e permitir que os funcionários sejam mais produtivos, tudo isso sem sacrificar a experiência do usuário final.

Agende um atendimento

Explore como o Cloudflare AI Security Suite pode transformar a abordagem de sua organização para a adoção segura de IA.

+55 (11) 3230 4523

enterprise@cloudflare.com

www.cloudflare.com

1. ManageEngine. The Shadow AI Surge in Enterprises: Insights from the US & Canada
2. Check Point Research. Q1 2025 Global Cyber Attack Report from Check Point Software: An Almost 50% Surge in Cyber Threats Worldwide, with a Rise of 126% in Ransomware Attacks