CLOUDFLARE

E-BOOK

Garantir práticas seguras de IA

Um guia para CISOs sobre como criar uma estratégia de IA escalável

Índice

Resumo executivo

Boas-vindas, CISO

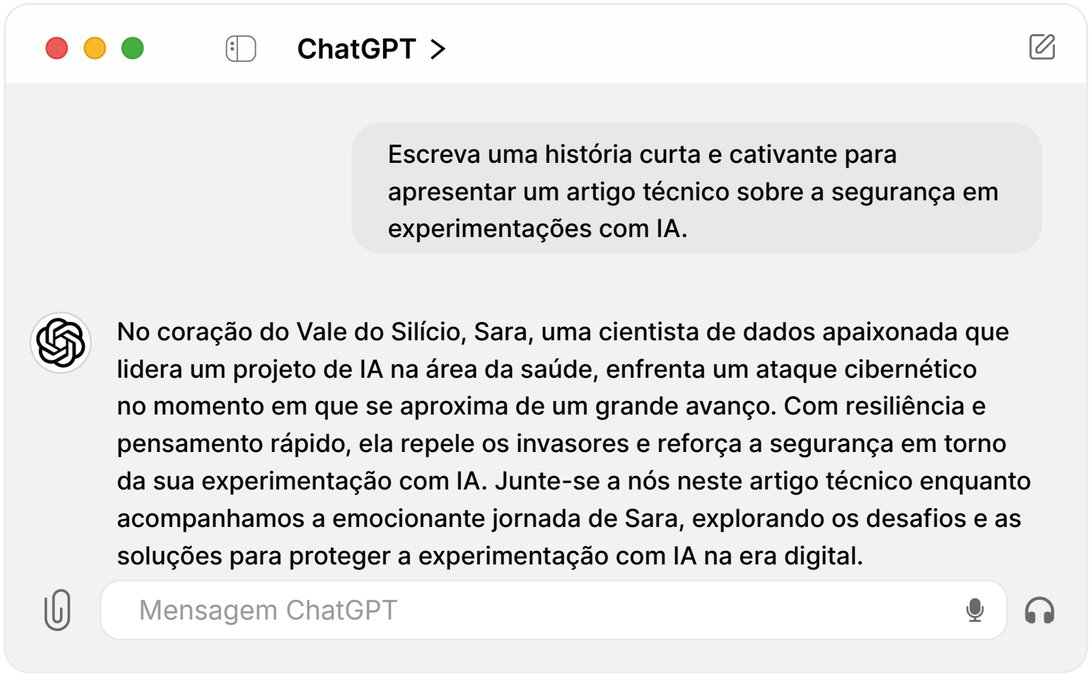
IA é possivelmente a palavra mais comentada atualmente e também é uma das questões mais urgentes para a comunidade de segurança. Sua influência exige nossa atenção, e é por isso que nós, da Cloudflare, escrevemos este guia para ajudar você a pensar na experimentação segura da inteligência artificial generativa (GenAI) em sua organização.

As ferramentas de IA estão rapidamente se tornando mais poderosas e acessíveis, desbloqueando oportunidades de inovações em todos os setores. No entanto, como acontece com outras mudanças de paradigma, a GenAI apresenta desafios específicos de segurança, privacidade e conformidade. A adoção generalizada da GenAI pode desencadear picos de uso imprevistos, casos de abuso de usuários, comportamentos maliciosos e práticas perigosas de TI invisível, tudo isso aumentando o risco de violações de dados e vazamento de informações confidenciais.

À medida que a adoção se expande em seu local de trabalho, você precisa se preparar com um plano de GenAI que informe como usar, criar e proteger em escala. Vamos discutir os riscos e analisar as dicas que sua equipe pode usar para proteger a GenAI com base nos níveis de maturidade e no uso. Com essas estratégias, sua organização pode criar uma estratégia de GenAI que atenda às necessidades da empresa e, ao mesmo tempo, proteja seus dados e garanta a conformidade.

- Dawn Parzych, Director of Product Marketing, Cloudflare

Segurança em experimentações com GenAI



Lamento informar, mas a história da Sara termina aí. Enquanto dizemos adeus ao nosso personagem fictício, à medida que a IA preditiva e a GenAI se expandem, haverá inúmeras "Saras" na vida real, cada uma atuando como uma heroína nas equipes de TI e de desenvolvedores, como tecnólogos de negócios e funcionários individuais.

A IA encantou os tecnólogos e os usuários comuns, despertando curiosidade e experimentação. Essa experimentação é necessária enquanto trabalhamos para desbloquear todo o potencial da IA. Mas sem cautela e proteções, também pode comprometer a segurança ou resultar em não conformidade.

Para alcançar o equilíbrio e entender e gerenciar as iniciativas de IA de forma mais eficaz, as organizações devem considerar três áreas principais:

1 Usar a IA

Usar tecnologias de IA (por exemplo ChatGPT, Bard e GitHub Copilot) oferecidas por fornecedores terceirizados enquanto protege ativos (por exemplo, dados confidenciais, propriedade intelectual, código-fonte, etc.) e mitiga possíveis riscos com base no caso de uso

2 Desenvolver com IA

O desenvolvimento de soluções de IA personalizadas para as necessidades específicas de uma organização (por exemplo, algoritmos proprietários para analytics preditiva, copilotos ou chatbots voltados para o cliente e sistema de detecção de ameaças orientado por IA)

3 Proteger a IA

Proteger aplicativos e sistemas de IA de agentes ruins que os manipulam para se comportar de forma imprevisível

Segurança em experimentações com GenAI

Transformação da GenAI: hoje e no futuro

O apelo da GenAI aos consumidores e organizações a colocou em uma trajetória de adoção sem precedentes. Um pequeno grupo de usuários avançados cresceu rapidamente graças, em parte, a uma comunidade ativa de código aberto e à experimentação de aplicativos como ChatGPT e Stable Diffusion, voltada para o consumidor.

O que os usuários descobriram com tudo isso é que os robôs não vão, de fato, "nos substituir".

A GenAI coloca os humanos na posição de refinar e aumentar, em vez de criar tudo do zero, e pode ajudar as empresas a ampliar a eficiência de sua força de trabalho. A IA preditiva oferece benefícios semelhantes ao facilitar o acesso a dados para melhorar a tomada de decisões, criar produtos mais inteligentes e personalizar experiências do cliente, entre uma série de iniciativas.

Hoje, **59% dos desenvolvedores** estão usando IA em seus fluxos de trabalho de desenvolvimento[1]

Até 2026, mais de **80% das empresas** vão usar APIs, modelos e/ou aplicativos habilitados por GenAI implantados em ambientes de produção (acima dos 5% de hoje)[2]

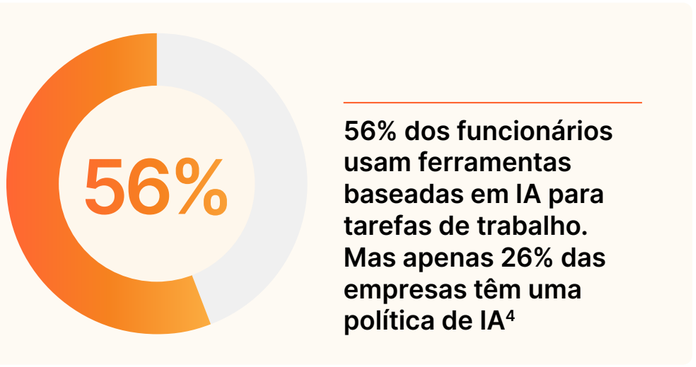
Até 2030, a GenAI aumentará **50% das tarefas dos trabalhadores do conhecimento** para aumentar a produtividade ou aumentar a qualidade média do trabalho (acima de menos de 1% hoje)[3]

1. SlashData, "How developers interact with AI technologies," maio de 2024
2. Gartner, "A CTO's Guide to the Generative AI Technology Landscape", setembro de 2023
3. Gartner, "Emerging Tech: The Key Technology Approaches That Define Generative AI", setembro de 2023

Usar a GenAI com segurança

A experimentação com IA abrange um espectro, desde o uso de ferramentas e serviços de IA pré-criados até a criação de soluções de IA personalizadas do zero. Embora algumas organizações possam avançar na criação de seus próprios modelos e aplicativos de IA, muitas continuarão a consumir ferramentas de IA de terceiros.

Nestes casos, as ferramentas de IA de terceiros criam novos riscos porque as organizações têm apenas controles diretos limitados sobre suas configurações de segurança e privacidade.



56% dos funcionários usam ferramentas baseadas em IA para tarefas de trabalho. Mas apenas 26% das empresas têm uma política de IA[4]

Os funcionários provavelmente estão usando ferramentas de IA prontas para o trabalho agora por meio de pacotes SaaS como o Microsoft 365, chatbots integrados a mecanismos de busca ou aplicativos públicos e até mesmo APIs.

As organizações devem tomar providências para minimizar riscos, incluindo:

- **Avaliar** o risco de segurança de ferramentas de terceiros
- **Abordar** as preocupações com a privacidade de dados
- **Gerenciar** a dependência (ou excesso de dependência) de APIs externas
- **Monitorar** possíveis vulnerabilidades

Um exemplo disso seria quando os funcionários usam aplicativos web públicos como o ChatGPT. Cada entrada inserida em um prompt torna-se um dado que sai do controle de uma organização. Os usuários podem compartilhar informações sensíveis, confidenciais ou regulamentadas, como informações de identificação pessoal (PII), dados financeiros, propriedade intelectual e código-fonte. E mesmo que não compartilhem informações confidenciais explícitas, é possível juntar o contexto às entradas para inferir dados confidenciais.

Para garantir, os funcionários podem alternar uma configuração para evitar que suas entradas treinem ainda mais o modelo, mas devem fazer isso manualmente. Para garantir a segurança, as organizações precisam de formas de impedir que as pessoas insiram dados privados.

Prepare-se para as implicações de segurança da IA

Exposição de dados

Até que ponto os usuários estão compartilhando indevidamente dados confidenciais com serviços externos de IA? As técnicas de anonimização/pseudonimização são suficientes?

Riscos para a API

Como você vai abordar as vulnerabilidades das APIs de terceiros que poderiam ser possíveis portas de entrada para invasores?

Sistemas de caixa preta

Quais são os processos de tomada de decisão dos modelos externos de IA que podem criar riscos inesperados?

Gerenciamento de riscos de fornecedores

O que você sabe sobre as práticas de segurança de seus provedores de IA terceirizados? Mais importante ainda, o que você não sabe?

4. The Conference Board, setembro de 2023

Etapas para defender o consumo de IA

1 Gerenciar a governança e o risco

- Desenvolver políticas sobre como e quando usar a IA, incluindo quais informações a organização permite que os usuários compartilhem com a GenAI, diretrizes de controle de acesso, requisitos de conformidade e como denunciar violações.
- Realizar uma avaliação de impacto para coletar informações, identificar e quantificar os benefícios e os riscos do uso da IA.

2 Aumentar a visibilidade e os controles de segurança e privacidade

- Registrar todas as conexões, inclusive com aplicativos de IA, para monitorar continuamente as atividades dos usuários, o uso de ferramentas de IA e os padrões de acesso a dados para detectar anomalias.
- Descobrir TI invisível existente (incluindo ferramentas de IA) e tomar decisões para aprovar, bloquear ou adicionar controles adicionais.
- Analisar as configurações de aplicativo SaaS em busca de possíveis riscos de segurança (por exemplo, permissões de OAuth concedidas por aplicativos aprovados a aplicativos habilitados por IA não autorizados, arriscando a exposição de dados).

3 Examinar quais dados entram e saem das ferramentas de IA e filtrar qualquer coisa que possa comprometer o IP, afetar a confidencialidade ou violar restrições de direitos autorais

- Aplicar controles de segurança para a forma como os usuários podem interagir com as ferramentas de IA (por exemplo, parar uploads, impedir copiar/colar e procurar e bloquear entradas de dados confidenciais/proprietários).
- Implementar salvaguardas para impedir que bots de IA raspem seu site.
- Bloquear ferramentas de IA completamente somente se nenhum outro controle for possível. Como sabemos, os usuários encontrarão soluções alternativas, o que coloca a segurança fora de seu controle.

4 Controlar o acesso a aplicativos e infraestrutura de IA

- Garantir que cada usuário e dispositivo que acessa as ferramentas de IA passe por uma verificação de identidade rigorosa para definir quem pode usar as ferramentas de IA.
- Implementar controles de acesso Zero Trust baseados em identidade. Aplicar o mínimo de privilégios para limitar possíveis danos causados por contas comprometidas ou ameaças internas.

5 Simplificar custos e eficiência operacional

- Entender como as pessoas estão usando aplicativos de IA com analytics e logging para que você tenha controle sobre a limitação de taxa, o armazenamento em cache, bem como novas tentativas de solicitação e fallback de modelo à medida que o uso escala.

Proteja o que você cria

Treine seu modelo de IA

Os pipelines de IA estão ampliando o espectro de vulnerabilidades. Mas com a experiência adquirida no início e durante todo o processo de desenvolvimento, temos insights sobre o que leva ao sucesso. Para segurança de IA, o lugar natural para começar é em seu modelo.

Como base para aplicativos de IA, tudo o que é usado para treinar seu modelo de IA flui para seus resultados. Considere como você vai proteger esses dados inicialmente para evitar repercussões negativas mais tarde. Se ficar sem proteção, você corre o risco de expandir sua superfície de ataque e criar problemas de aplicativos no futuro.

Uma segurança que garante a integridade dos dados é fundamental para mitigar o comprometimento de dados deliberado e acidental. Os riscos de segurança no pipeline de IA podem incluir:

- **Envenenamento de dados:** conjuntos de dados maliciosos influenciam os resultados e criam vieses
- **Abuso de alucinações:** os agentes de ameaças legitimam as alucinações de IA, a invenção de informações para gerar respostas, para que conjuntos de dados maliciosos e ilegítimos informem os resultados

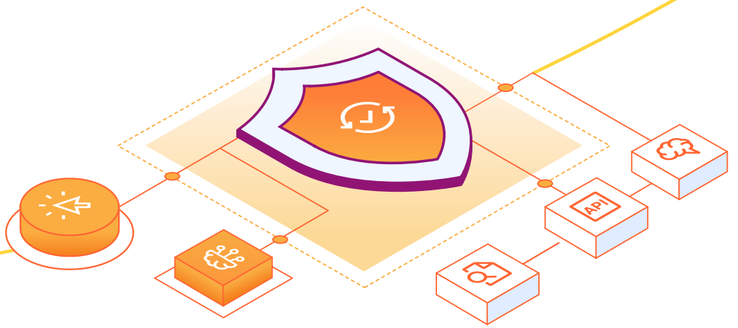
Por outro lado, se você não estiver treinando modelos, sua IA interna começa selecionando um modelo para executar tarefas. Nesses casos, você vai querer explorar como os criadores fizeram e protegeram o modelo, pois ele desempenha um papel na inferência.

A inferência é o processo que segue o treinamento de IA. Quanto melhor treinado e ajustado for um modelo, melhores serão as inferências, embora nunca haja garantia de que sejam perfeitas. Mesmo modelos altamente treinados podem ter alucinações.

Segurança pós-implantação

Depois de criar e implantar sua IA interna, você precisa proteger seus dados privados e garantir o acesso a eles. Além das recomendações que já fizemos neste artigo, incluindo a aplicação de tokens para cada usuário e limitação de taxa, você deve considerar também o seguinte:

- **Gerenciamento de cotas:** usa limites para proteger contra o comprometimento e o compartilhamento das chaves de API dos usuários
- **Bloqueio de determinados números de sistemas autônomos (ASNs):** impede que invasores enviem quantidades enormes de tráfego para aplicativos
- **Habilitar salas de espera ou desafiar usuários:** torna as solicitações mais difíceis ou demoradas, arruinando a economia dos invasores
- **Criar e validar um esquema de API :** descreve o uso pretendido ao identificar e catalogar todos os endpoints de API e, em seguida, lista todos os parâmetros específicos e limites de tipo
- **Analisar a profundidade e a complexidade das consultas:** ajuda a proteger contra ataques DoS e erros de desenvolvedores, mantendo sua origem íntegra e atendendo às solicitações de seus usuários conforme o esperado
- **Criar disciplina em torno do acesso baseado em tokens:** protege contra acesso comprometido quando os tokens são validados na camada de middleware ou no API Gateway



Proteção robusta contra ameaças em toda a sua experimentação com GenAI

Da adoção à implementação, cada estágio do espectro de experimentação com GenAI deve progredir com risco mínimo ou tolerado. Com o conhecimento adquirido neste documento, independentemente de sua organização usar, desenvolver ou planejar a IA, de alguma forma no futuro, você terá o poder de controlar seu ambiente digital.

Embora seja natural sentir-se hesitante ao adotar novos recursos, existem meios para lhe dar a confiança necessária para experimentar a IA com segurança. Desses recursos, o que as organizações mais precisam hoje é um tecido conjuntivo para tudo que é TI e segurança. Um que atue como um fio comum que reduz a complexidade ao trabalhar com tudo no ambiente, esteja disponível em todos os lugares e desempenhe as funções necessárias de segurança, rede e desenvolvimento.

Com o tecido conjuntivo, você confia em vários casos de uso, incluindo:

- Cumprir as regulamentações com a capacidade de detectar e controlar a movimentação de dados regulamentados.
- Recuperar a visibilidade e o controle de dados confidenciais em aplicativos SaaS, TI invisível e ferramentas emergentes de IA.
- Proteger o código do desenvolvedor detectando e bloqueando o código-fonte em uploads e downloads. Além disso, evitar, encontrar e corrigir configurações incorretas em aplicativos SaaS e serviços em nuvem, incluindo repositórios de código

À medida que a IA continua a evoluir, a incerteza é certa. É por isso que contar com uma força estabilizadora como a da Cloudflare é tão benéfico.

Proteja-se contra os riscos da IA em três tipos de LLMs

Dependendo do uso, o nível de exposição ao risco que a IA cria para uma organização varia. É fundamental entender os vários riscos associados ao uso e desenvolvimento de modelos de linguagem grande (LLMs) e, em seguida, estar ativamente envolvido em qualquer implantação de LLMs.

Tipo de LLM	Principal risco
Interno	Acesso a dados confidenciais e propriedade intelectual
Produto	Risco de reputação
Público	Vazamento de dados confidenciais

Escala, facilidade de uso e integração perfeita

A nuvem de conectividade da Cloudflare coloca o controle em suas mãos e melhora a visibilidade e a segurança, tornando a experimentação com IA segura e escalável. Melhor ainda, nossos serviços fortalecem tudo, garantindo que não haja comprometimento entre a experiência do usuário e a segurança.

Considerando que a maioria das organizações vai apenas usar a IA ou vai usar e desenvolver, utilizar a Cloudflare significa nunca parar em projetos de IA.

- Nossa **rede global** permite que você dimensione e imponha controles com velocidade, sempre que precisar
- Nossa **facilidade de uso** simplifica a implantação e o gerenciamento de políticas de como seus usuários consomem IA
- Uma **arquitetura programável** permite que você adicione segurança em camadas aos aplicativos que está criando, sem interromper a forma como seus usuários consomem a IA

A nuvem de conectividade da Cloudflare protege todas as facetas de sua experimentação com IA, especificamente:

- Nossos serviços **serviço de acesso seguro de borda (SASE) e Zero Trust** ajudam a mitigar o risco na forma como sua força de trabalho **usa** ferramentas de IA de terceiros
- Nossa **plataforma para desenvolvedores** ajuda sua organização a **criar** suas próprias ferramentas e modelos de IA com segurança e eficiência
- Para **proteger com IA**, nossa plataforma utiliza técnicas de IA e aprendizado de máquina para criar uma inteligência contra ameaças que é usada para proteger as organizações em suas experiências com IA

	Como você usa a IA	Como você desenvolve a IA
A escala da nossa rede global	Escale e imponha controles em qualquer lugar com consistência	Acelere a inferência, a consulta e o armazenamento em cache
Nossa simplicidade de gerenciamento	Um plano de controle com implantação e políticas simples	Modelos para integração rápida
Nossa arquitetura de rede unificada e programável	Crie uma nova camada de segurança sem interromper o modo como você usa a IA	Privacidade e conformidade integradas

Próximas etapas

Da proteção de como sua organização usa IA à defesa dos aplicativos de IA que você cria, o Cloudflare para IA tem tudo o que você precisa. Com nossos serviços, você pode adotar novos recursos em qualquer ordem, com interoperabilidade ilimitada e integrações flexíveis.

Fale com um especialista

Para obter mais informações, acesse cloudflare.com

+55 (11) 3230 4523 | enterprise@cloudflare.com | www.cloudflare.com/pt-br/

REV:BDES-6307.2024AUG19