

Compreender o controle do novo conteúdo de IA generativa e a expansão da cobertura de ferramentas de IA DLP

Contents

[Introdução](#)

[Overview](#)

[Como o DLP pode ajudar a controlar o conteúdo gerado pelo ChatGPT?](#)

[Por que controlar o conteúdo gerado por IA?](#)

[Como posso aplicar a verificação DLP a respostas ChatGPT?](#)

[Qual é a categoria do aplicativo de IA geradora no DLP?](#)

[Uma regra DLP pode ser aplicada a toda a categoria de aplicativo AI geradora?](#)

[Onde posso encontrar documentação relacionada?](#)

[Pretendemos fazer algum anúncio no próximo Cisco Live Amsterdam sobre esses casos de uso de proteção de IA generativa empolgantes?](#)

Introdução

Este documento descreve o novo controle de conteúdo de IA generativo e a expansão da cobertura de ferramentas de IA DLP para o Umbrella.

Overview

Estamos felizes em anunciar a disponibilidade geral do Controle de Conteúdo de IA Gerativa. Esse recurso permite monitorar e, se necessário, bloquear o conteúdo gerado pelo ChatGPT.

Estamos também felizes em compartilhar que expandimos o escopo de nossa cobertura DLP em tempo real para ferramentas de IA geradora. Inicialmente limitado ao ChatGPT, agora oferecemos suporte a todas as 70 ferramentas de IA em nossa categoria de aplicativos Generative AI recentemente lançada. Essa expansão significativa permite que você amplie o caso de uso Uso Seguro da IA, oferecendo uma solução mais abrangente e robusta para a proteção do uso da IA Gerativa.

Como o DLP pode ajudar a controlar o conteúdo gerado pelo ChatGPT?

O DLP pode ajudar as organizações a controlar o conteúdo gerado ao digitalizar respostas ChatGPT usando a política DLP em Tempo Real. Com esta versão, você pode escolher verificar as respostas ChatGPT (ou seja, o tráfego de entrada) para qualquer tipo de conteúdo gerado que você deseja monitorar ou bloquear.

Por que controlar o conteúdo gerado por IA?

O uso de conteúdo gerado por IA apresenta riscos para as empresas por várias razões, incluindo violação de direitos autorais, informações imprecisas, código defeituoso e assim por diante. Por exemplo, você pode querer impedir que os usuários usem o código-fonte gerado pela AI para impedir o uso de código protegido por direitos autorais ou não seguro, enquanto outros podem querer impedir o uso de citações judiciais geradas pela AI por medo de apresentar informações imprecisas.

Como posso aplicar a verificação DLP a respostas ChatGPT?

Geralmente, o DLP em tempo real verifica o tráfego de saída da Web, como os prompts do ChatGPT, para evitar vazamento de dados confidenciais. Com essa versão, estamos introduzindo a capacidade de também verificar o tráfego de entrada, escolhendo a direção do tráfego que o DLP em tempo real verifica, ou seja, o tráfego de entrada, o tráfego de saída ou ambos. Essa capacidade está disponível atualmente apenas para ChatGPT (chatbot e API). Escolher verificar o tráfego de entrada verifica as respostas de ChatGPT.

Destinations

Manage destination lists and vetted applications for this rule.

- ☐ All Destinations
Selecting All Destinations will scan the traffic to any application or website the user is browsing to.
- ☒ Select Destinations Lists and Applications for Inclusion
Scans selected destination lists and vetted applications.

< Destinations / Application Categories / Generative AI

Direction

☒ OpenAI API (Vetted)

Inbound

☒ OpenAI ChatGPT (Vetted)

Outbound

Outbound

Inbound

Outbound & Inbound

2 Selected for Inclusion

REMOVE ALL

Applications Categories

OpenAI API / Generative AI, Inbound

OpenAI ChatGPT / Generative AI, Outbound

23281122679316

Qual é a categoria do aplicativo de IA geradora no DLP?

Antes desta versão, os critérios de destino nas regras de DLP em tempo real incluíam uma lista selecionável finita de cerca de 20 aplicativos. Com esta versão, o DLP em tempo real permite que os clientes escolham qualquer uma de nossas 38 categorias de aplicativos, incluindo IA geradora, ou qualquer um dos ±4.600 aplicativos controláveis disponíveis categorizados dentro deles. A categoria de aplicativos de IA geradora, que foi lançada há apenas alguns meses com 20 aplicativos, agora tem 70 aplicativos, e estamos comprometidos com a atualização contínua

dessa categoria com ferramentas de IA de primeira linha.

Uma regra DLP pode ser aplicada a toda a categoria de aplicativo AI geradora?

Sim, uma regra de DLP em tempo real pode ser aplicada a uma categoria inteira ou a um subconjunto de aplicativos dentro dela.

Onde posso encontrar a documentação relacionada?

- Para aprender como controlar a direção de digitalização para monitorar ou bloquear ChatGPT respostas verificar:
[Adicionar uma Regra em Tempo Real à Política de Prevenção de Perda de Dados](#)
- Para saber como verificar se foi um prompt de chatGPT ou uma resposta de chatGPT bloqueada, verifique a direção de verificação relatada aqui: [Relatório de Prevenção de Perda de Dados](#)
- Para revisar todas as categorias de aplicativos que agora estão disponíveis nas regras de política DLP em Tempo Real, marque aqui: [Categorias de aplicativos](#)

Pretendemos fazer algum anúncio no próximo Cisco Live Amsterdam sobre esses casos de uso de proteção de IA generativa empolgantes?

Sim, realizaremos uma sessão de reunião à parte intitulada [Proteção dos seus dados confidenciais contra o uso de IA geradora](#) no Cisco Live Amsterdam, na terça-feira, 6 de fevereiro, 15h - 16h30 CET.

Por favor, guarde o seu lugar!

Sobre esta tradução

A Cisco traduziu este documento com a ajuda de tecnologias de tradução automática e humana para oferecer conteúdo de suporte aos seus usuários no seu próprio idioma, independentemente da localização.

Observe que mesmo a melhor tradução automática não será tão precisa quanto as realizadas por um tradutor profissional.

A Cisco Systems, Inc. não se responsabiliza pela precisão destas traduções e recomenda que o documento original em inglês ([link fornecido](#)) seja sempre consultado.