

新しいジェネレーティブAIコンテンツ制御とDLP AIツールカバレッジの拡大について理解する

内容

[はじめに](#)

[概要](#)

[DLPはChatGPTで生成されたコンテンツの制御にどのように役立ちますか。](#)

[AIで生成されたコンテンツを制御する理由](#)

[ChatGPT応答にDLPスキャンを適用するにはどうすればよいですか。](#)

[DLPのジェネレーティブAIアプリケーションカテゴリとは何ですか？](#)

[DLPルールをジェネレーティブAIアプリケーションカテゴリ全体に適用できますか？](#)

[関連ドキュメントはどこで入手できますか。](#)

[これらのエキサイティングなジェネレーティブAI保護のユースケースに関する発表を予定していますか？](#)

はじめに

このドキュメントでは、新しいジェネレーティブAIコンテンツコントロールと、UmbrellaのDLP AIツールカバレッジの拡張について説明します。

概要

ジェネレーティブAIコンテンツコントロールの一般提供を発表します。この機能を使用すると、ChatGPTによって生成されたコンテンツを監視し、必要に応じてブロックすることができます。

また、ジェネレーティブAIツールのリアルタイムDLPカバレッジの範囲が拡大したことも喜ばしく思います。当初はChatGPTに限定されていましたが、最近リリースされたジェネレーティブAIアプリケーションカテゴリでは、70種類のAIツールすべてがサポートされるようになりました。この大きな拡大により、AIの安全な利用のユースケースを拡大し、ジェネレーティブAIの利用保護のためのより包括的で堅牢なソリューションを提供することができます。

DLPはChatGPTで生成されたコンテンツの制御にどのように役立ちますか。

DLPは、リアルタイムDLPポリシーを使用してChatGPT応答をスキャンすることで、生成されたコンテンツの制御を支援します。このリリースでは、モニタまたはブロックする生成済みコンテンツの任意のタイプについて、ChatGPT応答（つまり着信トラフィック）をスキャンすることを選択できます。

AIで生成されたコンテンツを制御する理由

AIで生成されたコンテンツを使用すると、著作権の侵害、不正確な情報、欠陥のあるコードなど、さまざまな理由で組織にリスクが生じます。

例えば、著作権で保護されたコードや安全でないコードの使用を防ぐためにユーザーがAI生成のソースコードを使用しないようにしたい場合や、不正確な情報を提出することを恐れるためにAI生成の裁判所引用の使用を防ぎたい場合があります。

ChatGPT応答にDLPスキャンを適用するにはどうすればよいですか。

一般に、リアルタイムDLPはChatGPTプロンプトなどの発信Webトラフィックをスキャンして、機密データの漏洩を防ぎます。このリリースでは、リアルタイムDLPがスキャンするトラフィック（つまり、着信トラフィック、発信トラフィック、またはその両方）の方向を選択することで、着信トラフィックもスキャンする機能を導入しています。この機能は現在、ChatGPT（チャットボットとAPIの両方）でのみ使用できます。インバウンドトラフィックのスキャンを選択すると、ChatGPT応答がスキャンされます。

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.

The screenshot shows a configuration interface for DLP rules. At the top, there is a search bar containing "OpenAI". Below it, the breadcrumb navigation reads "< Destinations / Application Categories / Generative AI". The "Direction" dropdown is set to "Inbound". Two items are listed with checkboxes: "OpenAI API (Vetted)" and "OpenAI ChatGPT (Vetted)". The "OpenAI ChatGPT (Vetted)" item has a dropdown menu open, showing options: "Outbound", "Inbound", and "Outbound & Inbound". To the right, a box titled "3 Selected for Inclusion" contains a table of selected categories:

Applications Categories	
OpenAI API / Generative AI, Inbound	×
OpenAI ChatGPT / Generative AI, Outbound	×

23281122679316

DLPのジェネレーティブAIアプリケーションカテゴリとは何ですか？

このリリース以前は、リアルタイムDLPルールの宛先基準に、約20のアプリケーションの限定された選択可能なリストが含まれていました。今回のリリースでは、リアルタイムDLPを使用して、ジェネレーティブAIを含む38種類のアプリケーションカテゴリのいずれか、またはカテゴリに分類された±4,600種類の制御可能なアプリケーションのいずれかを選択できます。ほんの数ヶ月前に20個のアプリで立ち上げられたジェネレーティブAIアプリケーションカテゴリは、現在70個のアプリを持っており、私たちは常に頭の上のAIツールでこのカテゴリを更新することに取り組んでいます。

DLPルールをジェネレーティブAIアプリケーションカテゴリ全体に適用できますか？

はい。リアルタイムDLPルールは、カテゴリ全体に適用することも、カテゴリ内のアプリケーションのサブセットに適用することもできます。

関連ドキュメントはどこで入手できますか。

- スキャン方向を制御してChatGPT応答をモニタまたはブロックする方法については、次の点を確認してください。
[データ損失防止ポリシーへのリアルタイムルールの追加](#)
- ブロックされたchatGPTプロンプトまたはchatGPT応答であるかどうかを確認する方法については、次のリンクのスキャン方向レポートを参照してください。[データ損失防止レポート](#)
- リアルタイムDLPポリシールールで現在使用可能なすべてのアプリケーションカテゴリを確認するには、「[アプリケーションカテゴリ](#)」を確認してください。

これらのエキサイティングなジェネレーティブAI保護のユースケースに関する発表を予定していますか？

はい。2月6日（火）午後3時～4時30分(CET)に、Cisco Live Amsterdamで「[Protecting Your Sensitive Data from Generative AI Usage](#)」というブレイクアウトセッションを開催する予定です。

お席を取っておいてください。

翻訳について

シスコは世界中のユーザにそれぞれの言語でサポート コンテンツを提供するために、機械と人による翻訳を組み合わせて、本ドキュメントを翻訳しています。ただし、最高度の機械翻訳であっても、専門家による翻訳のような正確性は確保されません。シスコは、これら翻訳の正確性について法的責任を負いません。原典である英語版（リンクからアクセス可能）もあわせて参照することを推奨します。