

理解新的生成式AI內容控制和DLP AI工具覆蓋範圍的擴展

目錄

[簡介](#)

[概觀](#)

[DLP如何幫助控制ChatGPT生成的內容？](#)

[為什麼要控制AI生成的內容？](#)

[如何對ChatGPT響應應用DLP掃描？](#)

[DLP中的創成AI應用程式類別是什麼？](#)

[DLP規則能否應用於整個創成AI應用類別？](#)

[在哪裡可以找到相關文檔？](#)

[我們是否計畫在即將舉行的Cisco Live Amsterdam中發佈有關這些激動人心的創意AI保護使用案例的任何公告？](#)

簡介

本文檔介紹新的創成AI內容控制和擴展用於Umbrella的DLP AI工具覆蓋範圍。

概觀

我們很高興地宣佈Generated AI Content Control已全面推出。此功能使您能夠監控ChatGPT生成的內容，並在必要時阻止這些內容。

我們還很高興分享我們擴展了生成式AI工具的即時DLP覆蓋範圍。最初僅限於ChatGPT，我們現在支援我們最近發佈的創成AI應用類別中的所有70個AI工具。這一顯著的擴展使您能夠拓寬AI安全使用案例，為生成式AI使用保護提供更加全面且穩健的解決方案。

DLP如何幫助控制ChatGPT生成的內容？

DLP可以通過使用即時DLP策略掃描ChatGPT響應來協助組織控制生成的內容。在此版本中，您可以選擇掃描ChatGPT響應（即入站流量），尋找您要監控或阻止的任何型別的已生成內容。

為什麼要控制AI生成的內容？

使用人工智慧生成的內容會因為各種原因給組織帶來風險，包括侵犯版權、資訊不準確、代碼錯誤等。

例如，您可能希望阻止使用者使用由AI生成的原始碼來防止使用受版權保護或不安全的代碼，而其他人可能希望防止使用由AI生成的法庭引文，以擴音交不準確的資訊。

如何對ChatGPT響應應用DLP掃描？

通常，即時DLP會掃描出站Web流量（如ChatGPT提示），以防止敏感資料洩露。在此版本中，我們引入了通過選擇即時DLP掃描的流量方向（即入站流量、出站流量或兩者兼而有之）來掃描入站流量的功能。此功能目前僅適用於ChatGPT（chatbot和API）。選擇掃描入站流量掃描ChatGPT響應。

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

DLP中的創成AI應用程式類別是什麼？

在此版本之前，即時DLP規則中的目標標準包含一個可供選擇的約20個應用程式清單。在此版本中，即時DLP允許客戶選擇我們的38個應用類別中的任何一個（包括創成AI），或選擇其中分類的±4,600個可用可控應用中的任何一個。創成AI應用類別，幾個月前剛剛推出，現在有70個應用，我們致力於不斷更新這一類別，配備最先進的AI工具。

DLP規則能否應用於整個創成AI應用類別？

可以，即時DLP規則可應用於整個類別或其中的應用程式子集。

在哪裡可以找到相關文檔？

- 要瞭解如何控制掃描方向以監視或阻止ChatGPT響應檢查，請執行以下操作：
[向資料丟失保護策略新增即時規則](#)
- 要瞭解如何檢查被阻止的chatGPT提示還是chatGPT響應，請在此處檢查掃描方向報告：
[防資料丟失報告](#)
- 要檢視即時DLP策略規則中現在可用的所有應用程式類別，請在此處檢視：
[應用類別](#)

我們是否計畫在即將舉行的Cisco Live Amsterdam中發佈有關這些激動人心的創意AI保護使用案例的任何公告？

是的，我們將在Cisco Live Amsterdam舉辦一場名為[保護您的敏感資料免受生成AI使用情況的專題會議](#)，時間是2月6日星期二下午3:00 - CET下午4:30。

請保住您的座位！

關於此翻譯

思科已使用電腦和人工技術翻譯本文件，讓全世界的使用者能夠以自己的語言理解支援內容。請注意，即使是最佳機器翻譯，也不如專業譯者翻譯的內容準確。Cisco Systems, Inc. 對這些翻譯的準確度概不負責，並建議一律查看原始英文文件（提供連結）。