

了解新的生成AI内容控制和DLP AI工具覆盖范围的扩展

目录

[简介](#)

[概述](#)

[DLP如何帮助控制ChatGPT生成的内容？](#)

[为什么要控制AI生成的内容？](#)

[如何对ChatGPT响应应用DLP扫描？](#)

[DLP中的生成AI应用类别是什么？](#)

[能否将DLP规则应用于整个生成AI应用类别？](#)

[在哪里可以找到相关文档？](#)

[我们是否计划在即将举行的Cisco Live Amsterdam中发布有关这些激动人心的生成AI保护使用案例的任何公告？](#)

简介

本文档介绍新的生成AI内容控制和扩展适用于Umbrella的DLP AI工具覆盖。

概述

我们很高兴地宣布，生成AI内容控制已全面推出。
此功能使您能够监控ChatGPT生成的内容，并在必要时阻止这些内容。

我们还很高兴与您分享我们扩展了生成式AI工具的实时DLP覆盖范围。最初仅限于ChatGPT，我们现在支持我们最近发布的生成AI应用类别中的全部70个AI工具。这一显著的扩展使您能够拓宽AI安全使用案例，为生成式AI使用保护提供更加全面且稳健的解决方案。

DLP如何帮助控制ChatGPT生成的内容？

DLP可以通过使用实时DLP策略扫描ChatGPT响应来协助组织控制生成的内容。在此版本中，您可以选择扫描ChatGPT响应（即入站流量），以查找要监控或阻止的任何类型的已生成内容。

为什么要控制AI生成的内容？

使用人工智能生成的内容会因为各种原因给组织带来风险，包括版权侵权、不准确的信息、错误的代码等。

例如，您可能希望阻止用户使用AI生成的源代码来防止使用受版权保护或不安全的代码，而其他人可能希望阻止使用AI生成的法院引文，以免提交不准确的信息。

如何对ChatGPT响应应用DLP扫描？

通常，实时DLP会扫描出站Web流量（如ChatGPT提示），以防止敏感数据泄露。在此版本中，我们引入了通过选择实时DLP扫描的流量方向（即入站流量、出站流量或两者兼有）来扫描入站流量的功能。此功能当前仅适用于ChatGPT（聊天机器人和API）。选择扫描入站流量扫描ChatGPT响应。

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中的创成AI应用类别是什么？

在此版本之前，实时DLP规则中的目标条件包含一个可供选择的约20个应用的有限列表。在此版本中，实时DLP允许客户选择我们的38个应用类别中的任意一个（包括创成AI），或选择其中分类的±4,600个可用可控应用中的任意一个。创成式AI应用类别，几个月前刚刚推出，包含20个应用，现在已有70个应用，我们致力于不断更新这一类别，使用最顶尖的AI工具。

能否将DLP规则应用于整个创成AI应用类别？

是，实时DLP规则可以应用于整个类别或其中的应用程序子集。

在哪里可以找到相关文档？

- 若要了解如何控制扫描方向以监控或阻止ChatGPT响应检查，请执行以下操作：
[向数据丢失保护策略添加实时规则](#)
- 若要了解如何检查被阻止的chatGPT提示还是chatGPT响应，请在此处检查扫描方向报告：[防数据丢失报告](#)
- 要查看实时DLP策略规则中现在可用的所有应用类别，请在此处进行检查：[应用类别](#)

我们是否计划在即将举行的Cisco Live Amsterdam中发布有关这些激动人心的创成AI保护使用案例的任何公告？

是的，我们将在Cisco Live Amsterdam举办一个名为[保护您的敏感数据不受生成AI使用影响的分组会](#)，时间是2月6日星期二下午3:00 - CET下午4:30。

请保住您的座位！

关于此翻译

思科采用人工翻译与机器翻译相结合的方式将此文档翻译成不同语言，希望全球的用户都能通过各自的语言得到支持性的内容。

请注意：即使是最好的机器翻译，其准确度也不及专业翻译人员的水平。

Cisco Systems, Inc. 对于翻译的准确性不承担任何责任，并建议您总是参考英文原始文档（已提供链接）。