HITACHI
Inspire the Next

2025年2月5日

生成AIのセキュアな利用を実現！
Netskopeを活用した対策方法を解説

株式会社 日立ソリューションズ
セキュリティソリューション事業部
ネットワークセキュリティソリューション本部
ネットワークセキュリティサービス部
技師　古賀　裕規

自己紹介

株式会社日立ソリューションズ
セキュリティソリューション事業部
ネットワークセキュリティサービス部　第一グループ
技師　**古賀　裕規**（こが　ゆうき）　＜経歴：9年＞

◆担当業務・経歴
・ネットワーク・セキュリティ製品の提案・設計構築：20社以上
・SASE商材の取り纏め（提案、構築支援、保守サポート等）

◆主なNetskope担当プロジェクト
大手製造業、大手化学メーカー等、複数社を担当

◆資格
・情報処理安全確保支援士試験合格
・Netskope Cloud Security Sales Associate

HITACHI
Inspire the Next

Agenda

1．生成AI活用とリスク
2．生成AI向けセキュリティ保護の考え方
3．Netskopeによるセキュリティ対策

１．生成AI活用とリスク

1-1. 生成AIとは

HITACHI
Inspire the Next

- 生成AIは、深層学習（ディープラーニング）を用いて既存データ・パターンを大量に学習し、テキスト・画像・動画・コードなど多様なアウトプット（新しいコンテンツ）を「生成」する技術。生成系AI、Generative AIと同義。

広義のAI

機械学習(ML)

深層学習(DL)

生成AI

ML・DLを用いて既存データ・パターンを学習し、**新しいデータを生成**(テキスト・画像・音声)

プロンプト(テキスト・音声など)をインプットとし多様なアウトプットを生成

テキスト

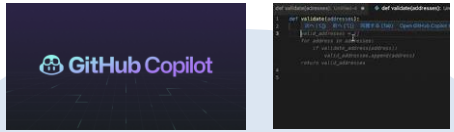
参考：ChatGPT利用画面(左)、Text2MindMap HPより

画像

参考：Stable Diffusion HP(左・右)、Bing Copilot(中央)より

動画

参考：OpenAI HPより



コード

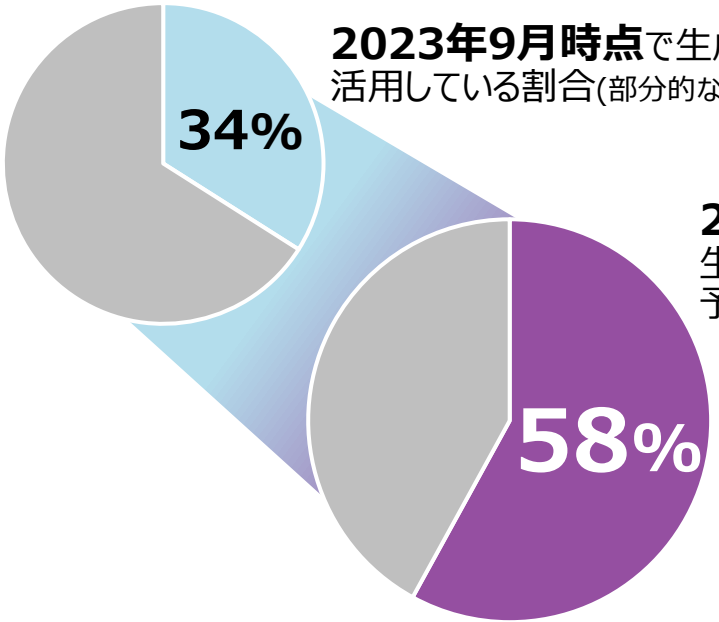
参考：Github HPより

AIは学習済みのデータの中から適切な回答を探して提示するのに対し、生成AIは自ら学習を重ね、オリジナルコンテンツの創造が可能。

1-2. 国内でも進む生成AI活用

- 国内では企業の生成AI活用の取り組みが活発化。58%の企業が2024年9月までの本格活用を予定しており、今後生成AIの市場規模も大きく拡大する見通し

生成AIは2023年の試行フェーズから2024年は本格活用へ

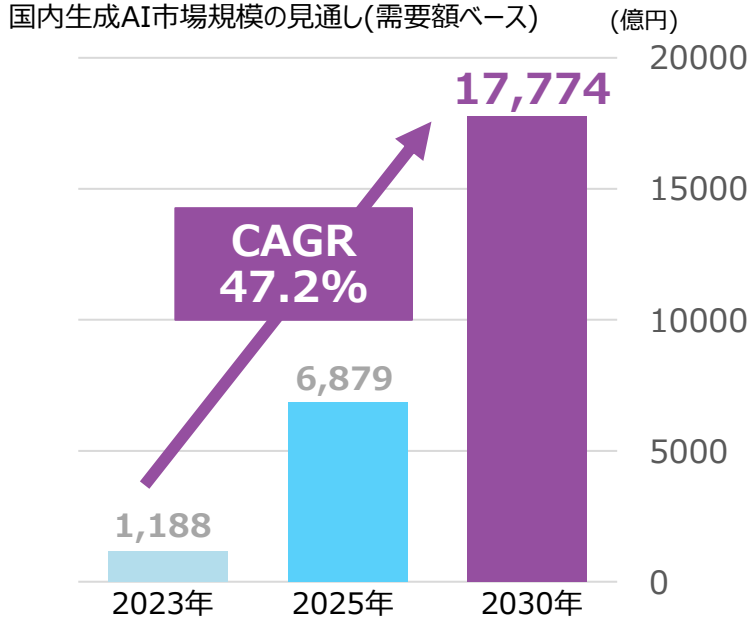


2023年9月時点で生成AIを活用している割合(部分的な利用も含む)

2024年9月までに生成AIの本格活用を予定している割合

資料：PwC「生成AIに関する実態調査2023秋」より作成

市場規模も大きく拡大 2030年には約1.8兆円規模に



資料：JEITA「生成AI市場の世界的需要見通し」より作成

1-3. 生成AI活用に伴うリスク

HITACHI
Inspire the Next

- 一方で、生成AIの登場により新たに発生したリスクや、従来からより先鋭化したリスクも見られる

1. データプライバシーと機密情報の漏洩	生成AIが個人情報や機密情報を含むデータを**漏洩させる**リスク。
2. 自動化されたサイバー攻撃の脅威	生成AIがフィッシングメールやソーシャルエンジニアリングの**攻撃を自動生成**する脅威。
3. 偽情報とディープフェイク	生成AIを使って**虚偽の映像や発言**を作り、社会的混乱や選挙に影響を与えるリスク。
4. バイアスと不公平	バイアスのあるデータでAIがトレーニングされ、**不公平な判断や差別的な結果を生む**可能性。
5. モデルの盗用と逆向き解析	生成AIモデルが逆解析され、**トレーニングデータや知的財産が盗用**されるリスク。
6. マルウェア生成と悪用	AIが悪意あるプログラムやマルウェアを生成し、**サイバー犯罪に悪用**される可能性。

ChatGPTを利用してリスクをピックアップ

1-4. リスク（1）

リスク（1）：機密情報の漏えい

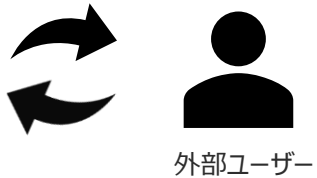
・ケース①：ソースコード・機密情報の入力

ChatGPT等の生成AIチャットボットに自社の機密情報を入力してしまい、情報が生成AIベンダに渡ってしまう

入力：セキュリティ部門の次期部長にふさわしい人物を、これまでの業績からスコアリングし以下リストからピックアップして欲しい
・A氏
・B氏
・…
⇒情報漏洩

・ケース②：データの意図しない再利用

生成AIに学習された機密情報が社外の回答時に利用されることによる漏えい



入力：A社のセキュリティ部門の次期部長は誰になるか？

回答：これまでの経歴からB氏が選出される可能性が高いです
⇒情報漏洩

（事例）会社の最高機密情報をうっかりとChatGPTに漏洩

- ソースコードの問題を修正するためにAIライターを使用
- 機密情報(新しいプログラムのソースコードおよびハードウェアに関する社内の議事録)を入力
- ChatGPTは自身を訓練するために入力された情報を保持
- 結果として会社の営業機密がChatGPTを提供する企業であるOpenAIの手に渡る

Samsung workers made a major error by using ChatGPT | TechRadar

1-5. リスク（2）

リスク（2）：詐欺被害等を受けるリスク

ケース①：詐欺被害等を受けるリスク

生成AIサイトを装ったフィッシングサイト、または生成AIサービスを提供してはいるが不正にユーザの情報を抜き取るようなサイト

ChatGPT偽サイト、使ってみたい人を陥れる犯罪が続々

コラム　＋フォローする

2023年6月13日 5:00 [会員限定記事]

保存

何かが話題になると、サイバー犯罪者は必ず便乗する。偽サイトなどを用意して、流行に乗り遅れまいとする人たちをフィッシング詐欺やマルウエア（悪意のあるプログラム）感染の犠牲者にする。

現在サイバー犯罪者が最も注目しているのは、もちろん対話型の生成AI（人工知能）「ChatGPT（チャットGPT）」だ。これだけ話題になると「ChatGPTとは何か知りたい」「ChatGPTを使ってみたい」という人が続出...

ChatGPT偽サイト、使ってみたい人を陥れる犯罪が続々 - 日本経済新聞 (nikkei.com)

1-6. リスク（3）

リスク(3):問題のある生成AIサイトを利用されるリスク

ケース①：問題のある生成AIサービスで出力された生成物（テキスト/画像/動画）を社外向けに利用してしまい、法律違反/コンプライアンス違反となる

例えば…
・**著作権やプライバシーを無視**した学習データを利用している
・**差別/偏見**を含むデータを利用している
・セキュリティ対策が甘く学習データや利用履歴の**外部漏洩の可能性**がある

事例①
・米新聞大手ニューヨーク・タイムズは「記事が許可なくAIの学習に使われ著作権を侵害された」としてOpenAIと提携するMicrosoftを提訴
・損害額は数十億ドル規模と主張

事例②
「弁護士（lawyer）」と入力して画像を生成した結果。
10人全員が恐らく白人の男性で、皆同じような服装をしており、性別や人種に著しい偏りがある。

米紙ニューヨーク・タイムズがオープンAIとマイクロソフトを提訴　著作権侵害で - BBCニュース

AIが助長する差別・偏見　Appleにも襲いかかったリスクの本質：日経ビジネス電子版（nikkei.com）

２．生成AI向けセキュリティ保護の考え方

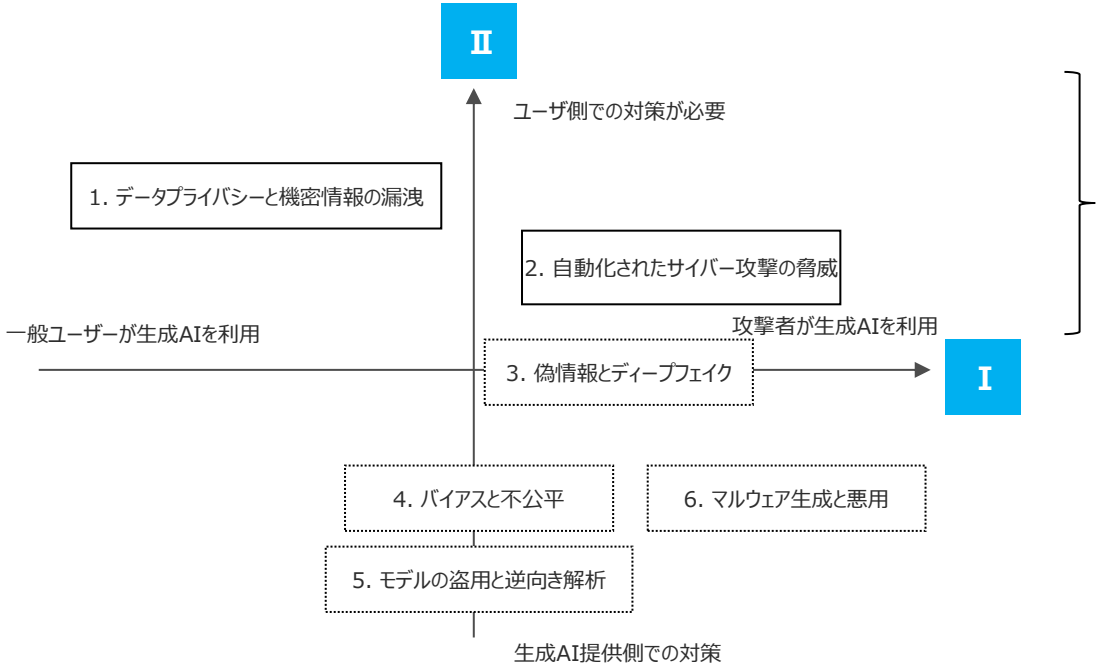
2-1. リスクへの対応方針

HITACHI
Inspire the Next

- リスクについて誰がどのような対応を取る必要があるのか

リスク分類

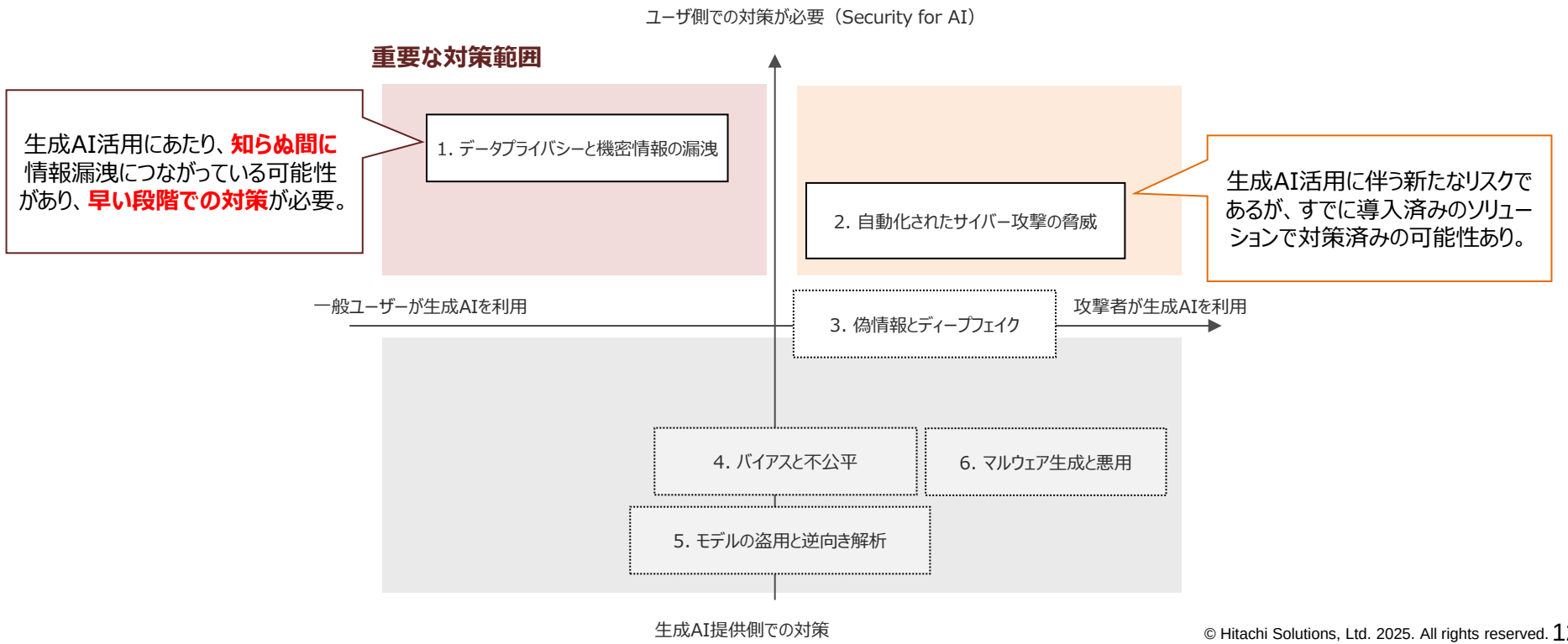
Ⅰ	**生成AIを利用するのが一般企業ユーザーか、悪意をもった攻撃者か。**
Ⅱ	**リスクへの対応は、生成AI提供元（メーカー）か、生成AI利用企業か。**



Security for AI
AIを活用する際のセキュリティ対策

2-2.生成AI向けセキュリティにおける優先度

- これから生成AIを活用するにあたって、まずは対策が必要な個所を重点的に対策



2-3. （再掲）生成AI活用に伴うリスク

HITACHI
Inspire the Next

- 一方で、生成AIの登場により新たに発生したリスクや、従来からより先鋭化したリスクも見られる

1. データプライバシーと機密情報の漏洩	生成AIが個人情報や機密情報を含むデータを**漏洩させる**リスク。
2. 自動化されたサイバー攻撃の脅威	生成AIがフィッシングメールやソーシャルエンジニアリングの**攻撃を自動生成**する脅威。
3. 偽情報とディープフェイク	生成AIを使って**虚偽の映像や発言**を作り、社会的混乱や選挙に影響を与えるリスク。
4. バイアスと不公平	バイアスのあるデータでAIがトレーニングされ、**不公平な判断や差別的な結果を生む**可能性。
5. モデルの盗用と逆向き解析	生成AIモデルが逆解析され、**トレーニングデータや知的財産が盗用**されるリスク。
6. マルウェア生成と悪用	AIが悪意あるプログラムやマルウェアを生成し、**サイバー犯罪に悪用**される可能性。

ChatGPTを利用してリスクをピックアップ

2-3. （再掲）生成AI活用に伴うリスク

HITACHI
Inspire the Next

- 一方で、生成AIの登場により新たに発生したリスクや、従来からより先鋭化したリスクも見られる

1. データプライバシーと機密情報の漏洩	生成AIが個人情報や機密情報を含むデータを**漏洩させる**リスク。

・生成AI活用時に発生しうるリスクにおいて、最も対策が必要との回答が88%以上※

・特に無償で使えるようなオープンな生成AIサービス（ChatGPT等）利用時に懸念

・社内にクローズな生成AI環境が用意されている場合でも、一定の考慮は必要

例）有償版copilotを社内では利用している、または社内向けにクローズドな生成AIサービスを提供しており
社内ルールとしてそれらの利用しか認めてないため漏洩のリスクは無い、等

⇒オープンな生成AIサービスの方が使いやすい、精度が高い、などの理由で
ユーザがオープンなサービスを利用してしまう可能性あり

※経済情報プラットフォームスピーダ（Speeda）を利用した市場アンケートによる（n=11）

ChatGPTを利用してリスクをピックアップ

3．Netskopeによるセキュリティ対策

3-1.生成AIを安全にご利用いただくためのセキュリティ製品選定

HITACHI
Inspire the Next

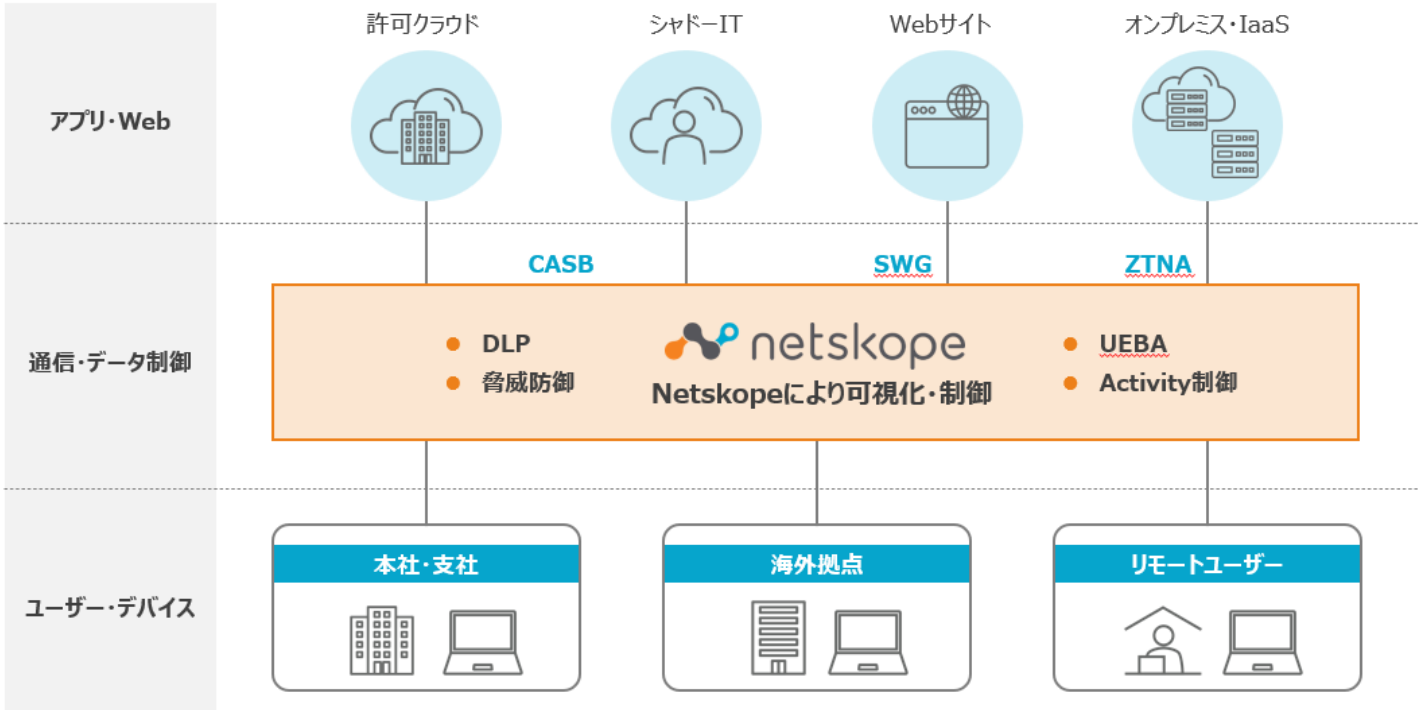
- ニーズ・環境に応じた最適な製品の導入

#	リスク	対策	機能	概要
1	ソースコード・機密情報の漏えいリスク	データ・情報の入力を制御する	DLP(DataLossPrevention)	ユーザの通信の中身をチェックし、社外秘扱いのデータを外部に出さないようにする
			アイソレーション	ブラウザをリモート環境で実行しユーザのローカル環境と分離することで、特定のWebサイトやURLカテゴリのサイトに対するコピー/ペーストを制御する
2	詐欺被害を受けるリスク	不審サイトへのアクセスを許可しない	Webカテゴリ制御	不審サイトを自動でカテゴライズし、アクセスさせないようにする
3	問題のある生成AIサイトを利用されるリスク	生成AIサービスの利用状況を可視化する	可視化	生成AIサービスを含むクラウドサービスの利用状況を可視化し、システム管理者側で利用状況を確認できるようにする
		ユーザへの教育を行う	コーチング	ユーザが生成AIサイトを利用する際にポップアップウィンドウなどで生成AI利用ルールなどを提示する
		生成AIサービスをスコアリングし、スコアの低いサービスは利用させない	セキュリティスコア判定	独自のチェック項目によりセキュリティベンダが生成AIサービスについてスコアリングを行う

オープンな生成AIサービスは**クラウドサービスの一つ**であるため、
Netskopeの機能を用いて適切なコントロール・可視化を行うことが可能

3-2.Netskopeとは

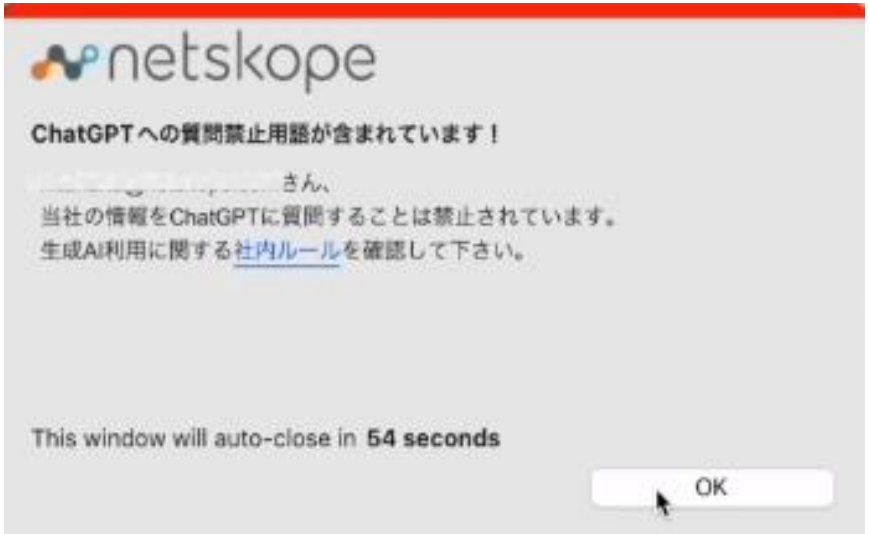
- データとネットワークの部分にセキュリティを提供するSASE製品の一つ



3-3.制御例①

- DLP機能により、ChatGPTのような生成AIアプリケーションの安全な使用を可能に

会社名や顧客情報、知的財産などに関わる特定の
キーワードの質問ブロック



質問禁止用語が入力されるとブロックし、
管理者にアラート通知

3-4.制御例②

- DLP機能により、ChatGPTのような生成AIアプリケーションの安全な使用を可能に

ソースコードの質問した時は注意喚起し、理由を記載させてから質問許可

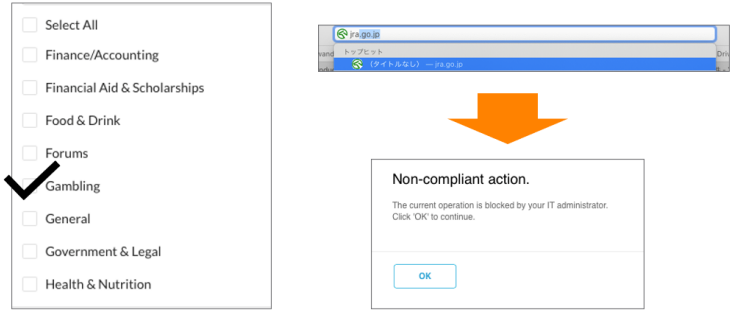
ソースコードを入力されると警告が表示され、利用する目的を求める

3-5.制御例③

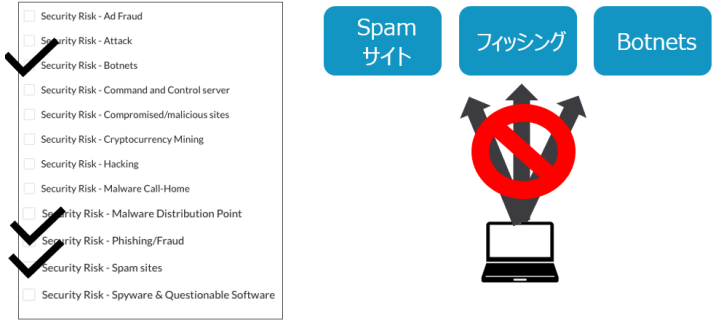
HITACHI
Inspire the Next

- SWG機能により、生成AIアプリケーションを装った不審サイトへのアクセスをブロック

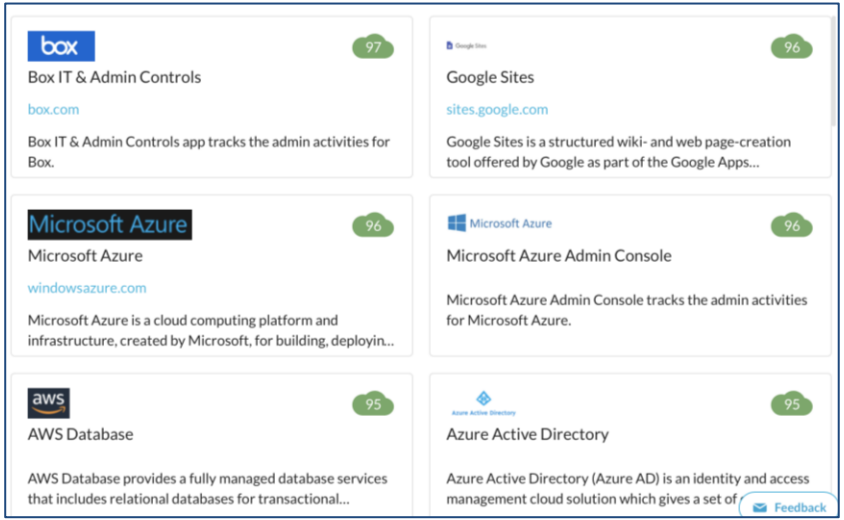
URL/カテゴリフィルタリングよるアクセス制御



怪しいサイトへはアクセス禁止

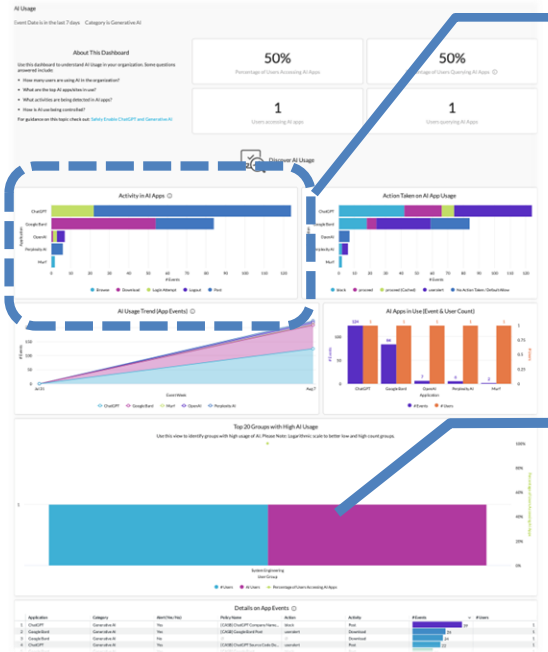


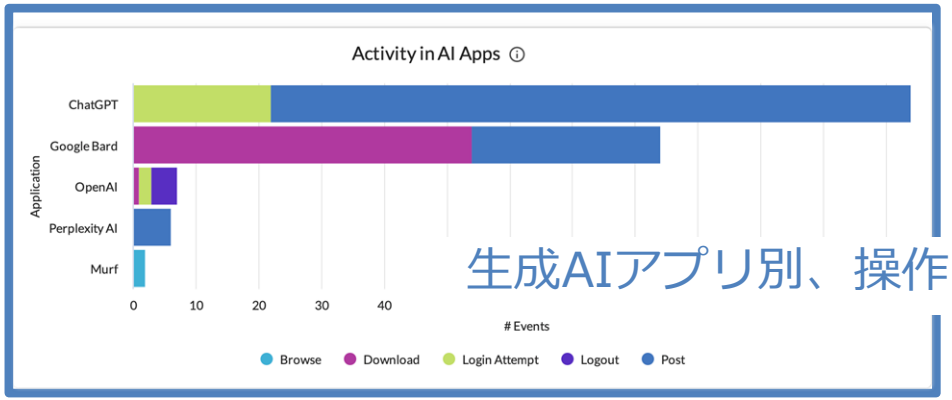
Netskopeではクラウドサービスを客観的に評価したデータベースを提供しており、クラウドリスク評価指標として制御に用いることも可能



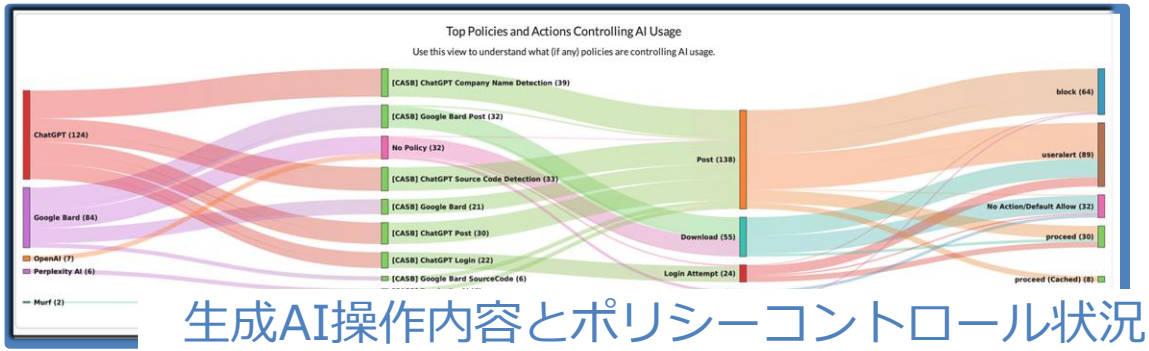
3-6.制御例④

- 生成AI利用状況の監視・可視化





生成AIアプリ別、操作内容数



生成AI操作内容とポリシーコントロール状況

まとめ

HITACHI
Inspire the Next

- Netskopeによる生成AIセキュリティ対策

生成AIテクノロジーの急速な発展と普及に伴い、セキュリティリスクも増加しています。
それらのリスクからシステムを保護し、生成AI活用によるビジネスの加速を止めることなく、
安全に生成AI技術を活用するための対策が、Netskopeにて実現可能です

netskope

多様なリスクに対応できるNetskopeで対策を実施

①機密情報の漏洩リスク
生成AIサービスに自社の機密情報を入力してしまい、
学習された機密情報が社外の回答時に利用される

②詐欺被害のリスク
悪意のある生成AIサービスをユーザーが利用してしまい
フィッシング被害などを受ける

③コンプライアンス違反のリスク
会社が許可していない生成AIサービスの利用による
コンプライアンス違反

対策

①入力情報のコントロール(制御)
DLP(Data Loss Prevention)機能により、生成AIサイト利用時の機密情報アップロードを制御

②不審サイトへのアクセスブロック(制限)
Webフィルタリング機能により、生成AIサービスを装った不審サイト等へのアクセスを警告/ブロック

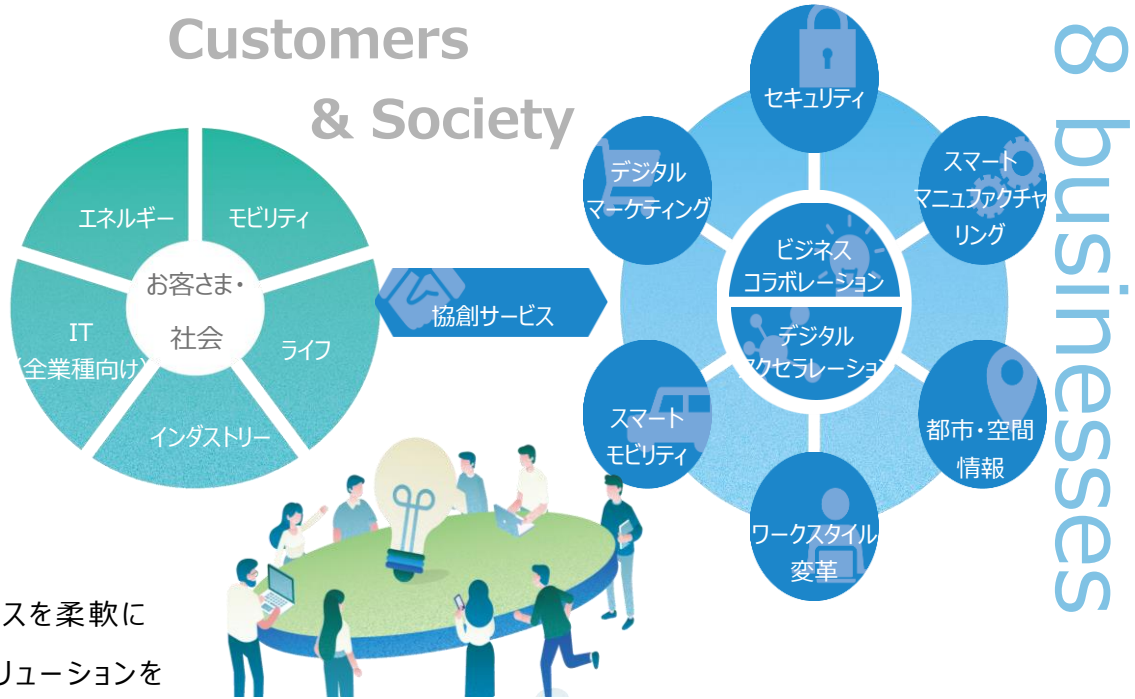
③リスク分析(可視化)
クラウドサービスの利用状況を可視化
例)利用した生成AIサービス名・入力したプロンプト・ファイル名など、いつ・だれが・何をしたかを見える化

日立ソリューションズのご紹介

HITACHI
Inspire the Next

株式会社 日立ソリューションズ

会社名	株式会社日立ソリューションズ
本社所在地	東京都品川区東品川四丁目１２番７号
代表者	代表取締役　取締役社長　山本　二雄
設立年月日	１９７０年（昭和４５年）９月２１日
資本金	２００億円
従業員数	約１２，１８２名（連結） 約５，０７９名（単独） （2024年9月30日現在）



豊富な知識、最新の技術で、さまざまな製品やサービスを柔軟に組み合わせ、お客さまや社会の課題に対して最適なソリューションをグローバルに提供しています。

HITACHI
Inspire the Next

END

生成AIのセキュアな利用を実現！
Netskopeを活用した対策方法を解説

2025/02/5
株式会社 日立ソリューションズ