

Prisma Cloud AI セキュリティ体制管理 (AI-SPM)

課題: AIの「ゴールドラッシュ」を受けたセキュリティと
コンプライアンスの確保

人工知能 (AI) と大規模言語モデル (LLM) は現代企業の戦略的な重点分野の一角を占めます。ここ数年の間に、AI と LLM に関する新しいモデルとツールが多数登場しました。マネージドモデル API やオープンソース基盤モデルのほか、運用ツール、プラグイン、アプリケーションから成る新しいエコシステムがその例です。

この状況は、AI を利用したアプリケーションを迅速に構築してデプロイする好機ですが、関連する新たなセキュリティ課題にも事欠きません。

セキュリティをはるかに上回る開発スピード：

AI の価値の具現化に邁進する中で、セキュリティの考慮が後回しになるおそれがあります。AI の迅速なデプロイに対する圧力は大きく、セキュリティ レビューが不十分なまま拙速な実装が行われかねません。

ブラック ボックス システム：

大規模 AI モデルは開発者であっても内部動作を見通せないことが多く、発生しうるセキュリティとコンプライアンスの問題の予測が困難です。従来の試験手法では、検出が容易でない脆弱性や予期しない振る舞いをモデルが示す場合があります。

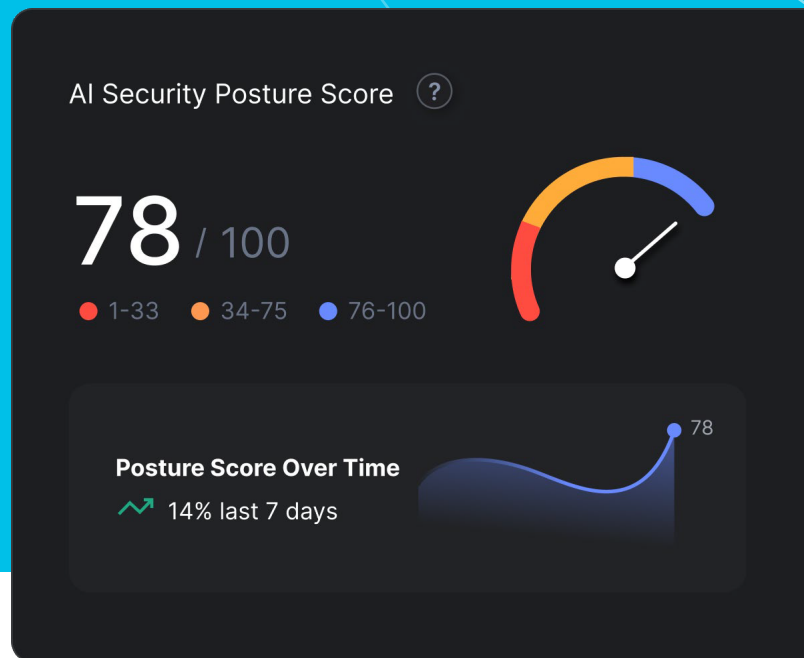
既存のアプローチでは対処困難な新規攻撃ベクトル：

ファイアウォールや体制分析ツール (例：CSPM、DSPM) など既存のセキュリティ手段では、AI 特有の攻撃 (例：データ ポイズニング、モデル反転、敵対的試行) に対処できません。

高まるコンプライアンス リスク：

今年後半に施行される見込みの EU AI 法は、データ プライバシー、アルゴリズム バイアス、説明可能な AI に関する新たな要件を課す内容です。違反に対する罰金も GDPR の約 2 倍に強化されます。また、米国など他地域でも同様の法律が制定される見込みです。

解決策： AI-SPM の可視性、 コントロール、 ガバナンス



Prisma Cloud AI-SPM は、データの公開、悪用、モデルの脆弱性など AI、機械学習 (ML)、生成 AI モデルに関する特有のリスクから組織を保護する各種機能から構成されます。より広範な [Code-to-Cloud™ アプローチ](#) の構成要素として、Prisma Cloud セキュリティ プラットフォームとシームレスに統合されているほか、データ セキュリティ体制管理 (DSPM)、クラウドセキュリティ体制管理 (CSPM)、クラウドネイティブ アプリケーション保護 (CNAPP) といった既存機能を土台とします。

Prisma Cloud AI-SPM を利用すると、AI モデルのライフサイクルをデータの取り込みから訓練、デプロイまで可視化できます。モデルの振る舞い、データ フロー、システムとのやり取りを分析することで、従来のリスク分析・検出ツールでは見逃す可能性があるセキュリティ リスクとコンプライアンス リスクを特定します。

AI-SPM がもたらす見識の用途：

- ・ポリシーとベストプラクティスを適用し、安全かつ法令に準拠した形で AI システムをデプロイ。
- ・データ ポイズニング、モデル反転、敵対的攻撃など AI 特有の脅威を検出。
- ・AI 開発ライフサイクルにプライバシーと利用規定に関する検討事項を組み込み、厳しさを増すコンプライアンス要件に先手に対応。

Prisma Cloud AI-SPM と
AI ランタイム セキュリティを統合し、
AI アプリケーション エコシステムの
効果的なランタイム保護を支援。

機能とメリット

AI モデルの検出とインベントリ作成

マネージド / セミマネージド / アンマネージド AI モデルの急増により、監視が難しくなる可能性があります。AI のデプロイを可視化することで、利用を効果的に監視し下流工程のリスクを予防することがセキュリティチームには必要です。

モデルの無秩序な増加とシャドー AI の管理 : モデル API、オープンソースモデル、仮想マシンにデプロイされたモデルのインベントリを閲覧できます。

モデルの悪用防止 : 誰が、どのモデルを使用しているかを特定し、モデルの未許可利用と不正ユースケースを防止します。

ガバナンス強化 : 新規モデルのデプロイ時にアラートを受信することで、適切なコントロールを確実に実施します。

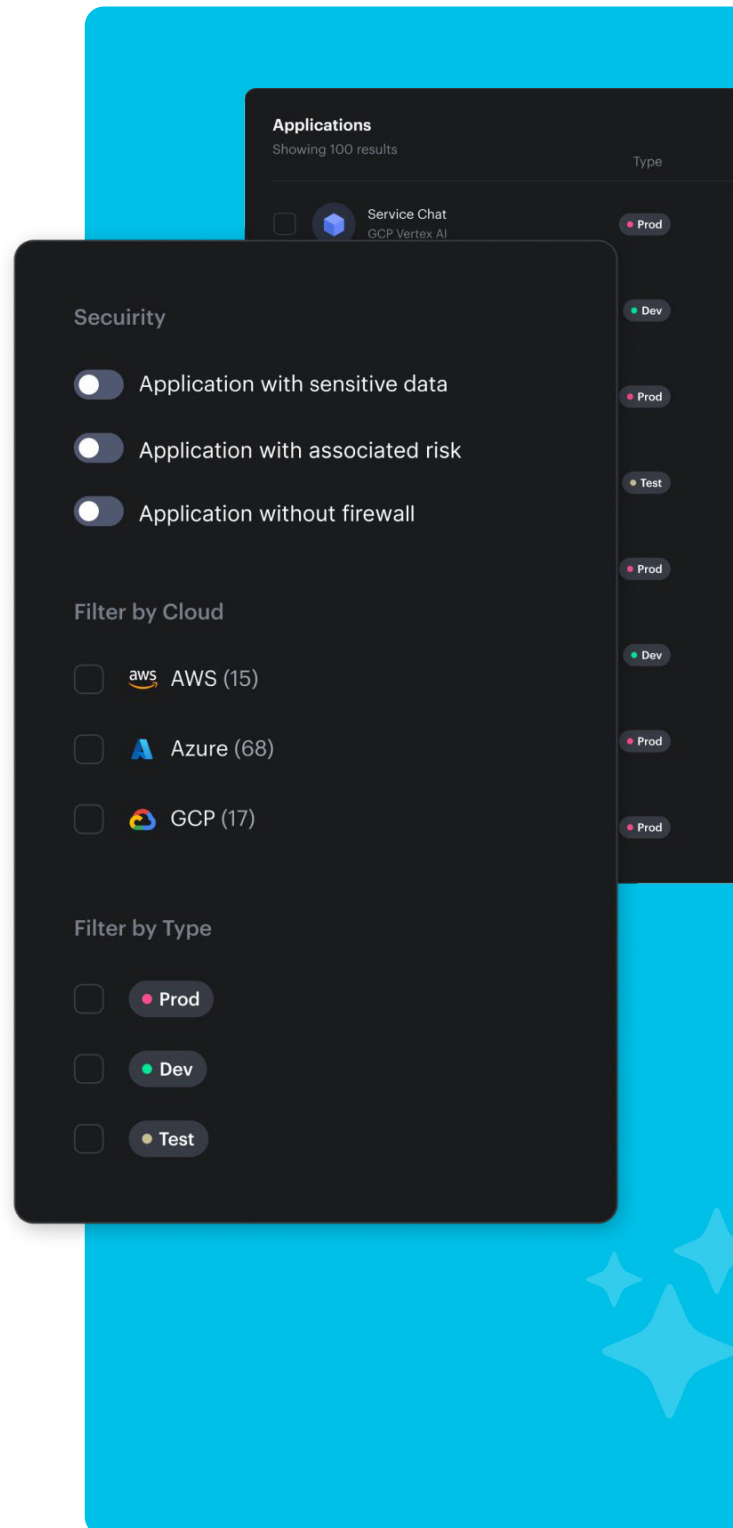
データ公開防止

AI モデルの訓練には膨大なデータを使用しますが、このデータには個人を識別可能な情報 (PII) や営業秘密などの機密データや規制対象データが含まれる可能性があります。加えて、不注意や敵対的攻撃によってこれらのデータが漏洩するおそれもあります。Prisma Cloud AI-SPM は、デプロイ済みの各モデルを通じてアクセス可能な内部データの把握に役立ちます。

訓練データセットの検出と分類 : データポイズニングを防止するとともに、モデルの訓練や微調整に機密データが使用されているか否かをデプロイ前に解明します。

検索拡張生成 (RAG) と推論データ監視の実施 : 検索に使用されるデータセットとデータフローを把握し、データへの効果的なアクセスに与える影響を明らかにします。

モデルとのやり取りを分析 : プロンプトと出力ログをスキャンし、モデルの悪用や意図しないデータ公開のエビデンスを取得します。



体制およびリスクの分析

データ パイプライン、訓練環境、デプロイ インフラに含まれる設定ミスとアクセス制御の問題は重大なセキュリティ リスクとコンプライアンスリスクを招きかねません。Prisma Cloud AI-SPM は AI のデプロイをエンドツーエンドでスキャンして問題を検出した上で、重大度に応じて是正措置に優先順位を付けます。

設定ミスの優先順位付けと対処 : AI 関連のアプリケーション、データ、パイプラインの攻撃経路全体を分析して脆弱性を検出します。Prisma Cloud の DSPM 機能と CSPM 機能から得たコンテキストに関する知見に基づく機能です。

モデルへのアクセスを統制 : デプロイ済みの AI モデルと関連リソース (コンピューター、データ、アプリケーション) に誰がアクセスできるかを視覚的にマッピングします。

アプリケーションとデータに対する権限を適正化 : 内部モデルに対する過剰な権限や、モデルへの望ましくないパブリックアクセスを解消します。

Top risks

	Training dataset publicly writable First Discovered on Dec 19, 2023, 9:01 AM	Critical
	Public inference dataset First Discovered on Dec 19, 2023, 9:01 AM	High
	AI deployment without content filtering First Discovered on Dec 19, 2023, 9:01 AM	High
	Model serving misconfigured app First Discovered on Dec 19, 2023, 9:01 AM	Medium

リスク シナリオの例

Prisma Cloud AI-SPM は単に修正すべき問題の一覧を表示するだけでなく、データとアプリケーションに関する充実したコンテキスト情報を提供します。優先順位付けに基づく効果的な是正措置に必要なツールとして、多忙なセキュリティ チームを支援します。

下表は Prisma Cloud AI-SPM が検出可能なリスクの一例です。クラウド環境のエージェントレス スキャンから入手可能な追加コンテキストも掲載しています。

リスク	優先順位付けのための追加コンテキスト
匿名アクセスが可能な AI デプロイ	デプロイに使用されているモデルはどれか。訓練データセットや推論データセットに機密データが含まれるか
パブリック書き込み可能な訓練データセット	当該データセットを基に訓練されたモデルはどれか。誰がモデルにアクセス可能か
プロンプト履歴を含む、設定ミスのある資産	設定ミスの深刻度。プロンプト履歴のスコープ (パブリック / 内部)
パブリック書き込み可能な訓練データセット	当該デプロイへのアクセス権を持つのは誰か。モデルの訓練に使用された内部データはどれか
本番環境の AI デプロイをフィルタリングするコンテキストが不足	コンプライアンスへの潜在的な影響と、それが環境に与える影響
本番環境外からアクセス可能な AI デプロイ	意図しない公開のおそれがある機密データはどれか。違反状態の関連コンプライアンス ポリシーはどれか

エージェントレス デプロイは 迅速かつ容易



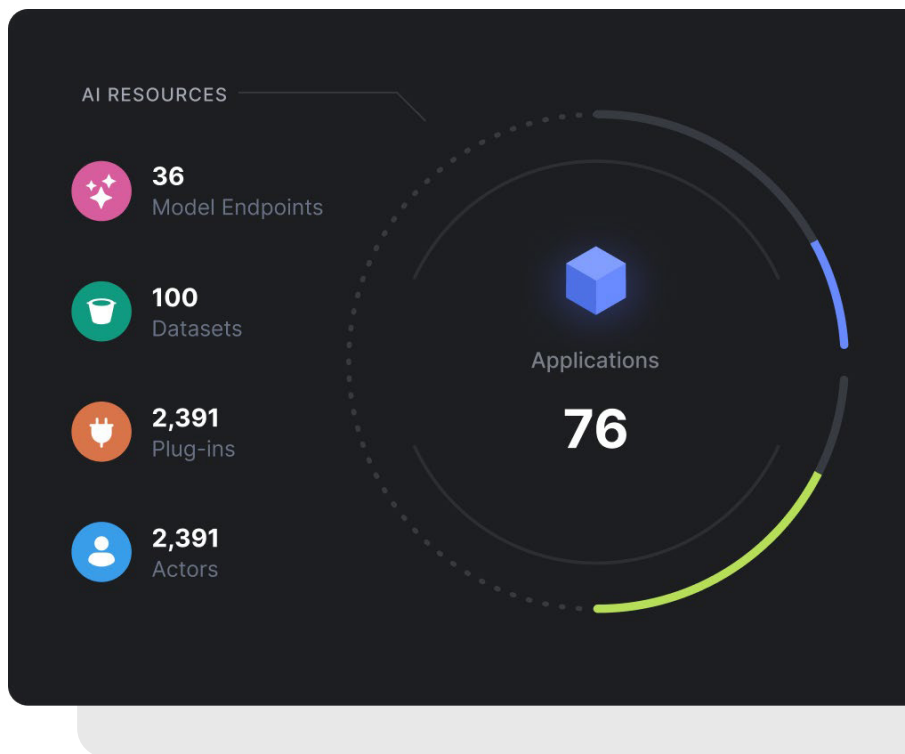
インストールは数分で完了。
クラウド データへのアクセスに
エージェントもコネクタも不要。



データをアウトオブバンドで
スキャンし、本番環境への
影響を最小化。



クラウド環境からの
機密データの流出を防止。



Prisma Cloud について

Prisma Cloud は業界で最も包括的なクラウドネイティブアプリケーション保護プラットフォーム (CNAPP) です。アプリケーション、データ、クラウドネイティブテクノロジースタック全体をカバーする最も広範なセキュリティとコンプライアンスを開発ライフサイクル全体を通して提供します。加えて、マルチクラウド環境とハイブリッドクラウド環境の両方に対応しています。

統合型アプローチの採用により、セキュリティ運用チームと DevOps チームの俊敏性維持、効果的な連携、安全で迅速なクラウドネイティブアプリケーション開発を可能にする製品です。詳しくは、Web サイト (prismacloud.io) をご覧ください。



〒100-0011
東京都千代田区幸町2丁目1番6号
日比谷パークフロント 15 階
電話番号: 03-6205-8061
www.paloaltonetworks.jp

© 2024 Palo Alto Networks, Inc. 米国およびその他の司法管轄権域内における商標の一覧は、<https://www.paloaltonetworks.com/company/trademarks.html> をご覧ください。本書に記述されているその他の商標はすべて、各社の商標である場合があります。
prisma_ds_ai-security-posture-management_052324