AI Defense

目次

Cisco AI Defense による AI トランスフォーメーションの保護 3

Cisco AI Defense ソリューション 3

Cisco AI Defense の利点 4

AI Defense のコアコンポーネント 4

AI Defense の展開オプション 6

変化する業界標準に対応 7

AI 向けのシスコのセキュリティ 7

次の手順 7

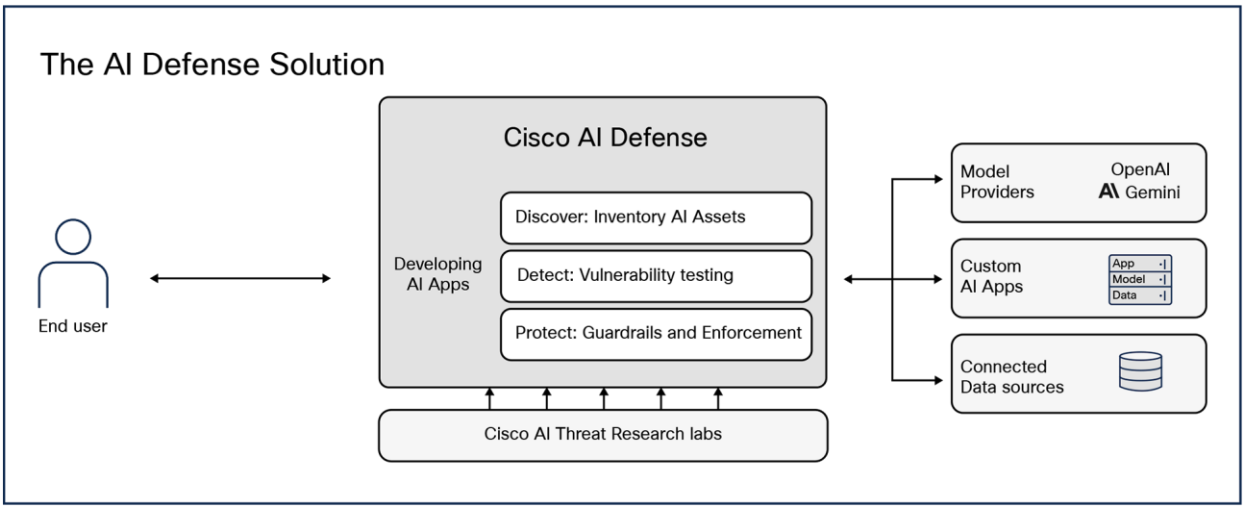
Cisco AI Defense による AI トランスフォーメーションの保護

AI インフラストラクチャを認識していない既存のセキュリティ管理

組織が AI 主導のアプリケーションを作成すると、既存のセキュリティ管理では対応できない重要な課題に直面します。次のようなものがあります。

- **可視性ギャップ：**AI インフラストラクチャはマルチクラウド、マルチモデル、マルチモーダルで構成されるため、カバレッジに大きなギャップが生まれます。
- **AI モデルの脆弱性：**AI モデルは非決定的であり、継続的なアセスメントと検証が必要です。
- **新たな敵対的 AI の脅威：**AI システムは、ランタイムアプリケーションを標的とする新しい攻撃手法（プロンプトインジェクションやジェイルブレイクなど）に対して脆弱です。

これらの脅威から組織を保護するには、新しいアプローチが必要です。



Cisco AI Defense ソリューション

Cisco® AI Defense は、業界をリードする AI およびサイバーセキュリティ テクノロジーを Cisco Security Cloud に組み込むことで、AI の開発、展開、使用に関連するリスクを軽減します。

AI Defense の本質的な機能は、セキュリティチームが脆弱性を検出し、リアルタイムのガードレールを実装し、企業全体の AI 攻撃対象領域にシームレスに統合できるように支援することです。

Cisco AI Defense の利点

AI モデルとアプリケーションの検証により脆弱性とリスクを評価	AI 脅威インテリジェンスで攻撃の脅威に先制対応	ネットワークレイヤにセキュリティを適用
AI アルゴリズムのレッドチーミングにより、セキュリティおよび安全性の脆弱性を自動で評価し、推奨されるガードレール対策を報告します。	シスコのAI脅威調査チームは、最新の攻撃データを継続的に AI Defense に反映させるだけでなく、MITRE の AI システムの敵対的脅威の状況（ATLAS）、Open Web Application Security Project（OWASP）、国立標準技術研究所（NIST）などのフレームワークとの整合性も確保しています。	シスコのネットワーキングに関する専門知識と、AI 攻撃対象領域全体にわたる広範囲な可視性と適用ポイントを活用してください。

AI Defense のコアコンポーネント

AI Defense は、組織全体のリアルタイムな AI リスク管理に向けて、セキュリティ、プライバシー、安全性を一体化した連続的なレイヤを導入します。このソリューションは、次の 3 つのコアコンポーネントを土台としています。

- **AI クラウドの可視性：**クラウド環境内のモデルやエージェントなどの AI アセットを特定し、ギャップを発見して解消します。
- **AI モデルとアプリケーションの検証：**AI アルゴリズムのレッドチーミングを活用して、安全性とセキュリティの脆弱性を迅速に特定します。従来は、手動でのレッドチーミングに数週間を要していました。
- **AI のランタイム保護：**プロンプトインジェクション、DoS 攻撃、ジェイルブレイク、機密データの漏洩など、新しい AI の脅威に自動的に適応する安全性、セキュリティ、プライバシーのガードレール対策により、AI アプリケーションを保護します。

AI クラウドの可視性	
機能	**利点**
• **自動 AI アセット検出：AI** 関連のクラウドトラフィック（インバウンド、アウトバウンド、水平方向トラフィック）全体にわたる継続的な可視性を確保し、Amazon Bedrock などのクラウド環境に存在する AI アセット（モデルやエージェントなど）を認識できます。 • **AI トラフィック フロー マッピング**：企業の AI 攻撃対象領域を完全に把握できます。 • **AI アセットポリシーチェック**：評価された AI モデルを自動的に追跡して優先順位を付け、保護されていないアセットのスキャンと保護を開始します。 • **各モデルの AI セキュリティ管理を確認**：脆弱性テストとランタイム保護を強化するために、関連するセキュリティ管理を迅速に特定します。	• 場所や発生時期にかかわらず、モデルや AI アセットを特定できます。 • 組織の AI セキュリティリスクを自動的にインベントリ化することができる一括管理ビューにより、クラウドに存在する不正な AI アセットの不確実性とリスクを排除できます。 • 新しいアセットまたは承認されていないアセットを特定し、セキュリティルールに準拠させます。 • データ、モデル、およびエージェント間のつながりをマッピングすることにより、セキュリティポリシーに関して情報に基づいた意思決定を行うことができます。

AI モデルとアプリケーションの検証	
機能	**利点**

• **アルゴリズムのレッドチーミング**：シスコ独自の AI を使用して、200 以上の攻撃手法と脅威カテゴリを自動的に実行することで、モデルの脆弱性をテストします。
  ◦ 45 以上のプロンプトインジェクション攻撃手法
  ◦ 30 以上のデータプライバシーカテゴリ
  ◦ 20 以上のセキュリティターゲット
  ◦ 50 以上の安全性カテゴリ
• **脆弱性レポート**：プロンプトインジェクション攻撃手法、データセキュリティ、プライバシー、安全性の脆弱性に関するレポートを提供します。
• **モデル固有のガードレール**：各モデルで見つかった具体的な脆弱性に合わせたガードレールを生成します。
	• 何百もの潜在的な安全性とセキュリティのリスクをリアルタイムで迅速に特定できます。
	• プロンプトインジェクション、データポイズニング、意図しない結果など、悪意のあるアクションに対する脆弱性を特定できます。
	• API またはソフトウェア開発キット（SDK）を通じて、AI 対応アプリケーションスタック全体（モデル、データ、ユーザー、アプリケーションなど）のリスク、脆弱性、および脅威を追跡できます。
	• 全体的な AI セキュリティリスクのレポートを取得して、AI アプリケーションのコンプライアンスとステータスを評価できます。

AI ランタイム保護	
機能	**利点**
• **安全、セキュリティ、プライバシーの管理** ◦ AI ランタイムは、すべてのプロンプトと応答を検査して違反をブロックします。 ◦ **セキュリティ：**プロンプトインジェクション、サービス妨害、悪意のある URL などの AI 攻撃と脅威から保護します。 ◦ **プライバシー：**個人を特定できる情報（PII）、クレジットカードデータ保護基準（PCI）、保護医療情報（PHI）、ソースコード、モデル情報などの機密情報の漏洩を防止します。 ◦ **安全性：**財務、社会、評判、ユーザーに関わるカテゴリで、有害であるか被害を及ぼす可能性のあるコンテンツをブロックします。 • **違反ダッシュボード**：違反を可視化し、違反の種類、プロンプト、対応などに関する情報を提供します。 • **複数の適用ポイント**：AI Defense のガードレールは、API コールや、サービスと AI アプリケーション間のアウトバウンドおよび内部通信など、複数のトラフィックタイプに対応します。	• 機密情報や有害なコンテンツをブロックして、安全な出力を保証します。 • AI 対応アプリケーションのコンプライアンス違反を防ぎます。 • セキュリティチームは、組織の標準やトレランスに合わせてルールを調整できます。 • Splunk などのセキュリティコンソールの統合によって、AI セキュリティ運用やより広範なセキュリティ運用に統合します。

AI Defense の展開オプション

コンポーネント	Validation Essentials	Runtime Essentials	Advantage
AI クラウドの可視性	✓	✓	✓
AI モデルとアプリケーションの検証	✓	。	✓
AI ランタイム保護	。	✓	✓

変化する業界標準に対応

AI 革命の最前線：AI Defense は、NIST の AI リスク管理フレームワーク（AI-RMF）、MITRE の ATLAS、OWASP Top 10 for LLM などの進化する標準に準拠することで、チームが常に先手を打てるように支援します。さらに、AI Defense チームがこれらのフレームワークに積極的に貢献しており、組織が今後の規制や業界要件に対応できるように備えます。

AI 向けのシスコのセキュリティ

シスコは、ネットワーキングとサイバーセキュリティにおける何十年ものリーダーシップを基盤として、以下のような迅速な AI イノベーションと柔軟な AI セキュリティの分野を開拓しています。

- **企業が構築する AI アプリケーション：**Cisco AI Defense は、AI アプリケーションの開発と展開によって生じる安全性とセキュリティのリスクから保護します。
- **サードパーティの生成 AI アプリケーションまたは シャドウ AI：**シスコは、Cisco Secure Access を使用してサードパーティの AI アプリケーションのセキュリティリスクから組織を保護します。許可されていないツールへの従業員のアクセスを制限し、脅威や機密データの損失を防ぎます。
- **AI サプライチェーンとリスク管理：**シスコは、Cisco Secure Access、Cisco Secure Endpoint、Cisco Email Threat Defense によるネットワークベースの制御を通じて、悪意のある AI モデルファイル（例：悪意のあるコードのスキャン、ソフトウェアライセンスのコンプライアンス、地政学的起源のリスクなど）が企業に侵入するのを防ぎます。

次の手順

詳細については、https://www.cisco.com/go/ai-defense をご覧ください。

米国本社
カリフォルニア州サンノゼ

アジア太平洋本社
シンガポール

ヨーロッパ本社
アムステルダム（オランダ）

シスコは世界各国に約 400 のオフィスを開設しています。オフィスの住所、電話番号、FAX 番号は当社の Web サイト（www.cisco.com/jp/go/offices）をご覧ください。

Printed in USA C78-5114794-00 05/25