AIセキュリティ：AI競争における貴社の競争優位性

リスク管理を犠牲にすることなくAI導入を加速させるための経営幹部向けガイド

はじめに

多くの企業がAIの変革力を活用しようとそのスピードを速めている一方で、セキュリティ対策が遅れを取っています。正式に承認されたAI導入プロジェクトの裏で、個人やチームが隠れて使用するシャドーAIの事例は数え切れないほど存在します。実際、IT意思決定者の85%が、「ITチームによるAIツールの評価を待たずに従業員がAIツールを導入している」と答え[1]、従業員の93%が、「承認なしにAIツールに情報を入力している」ことを認めています[1]。

同時に、従来のセキュリティツールで対応が難しいAIネイティブな攻撃ベクトル（AIを悪用した攻撃手法）は急増し、前年比で47%増[2]となっています。新しい形態のデータ流出やコンプライアンスのギャップは、既存のガバナンス慣行に課題をもたらします。経営幹部は、「AIがもたらす革新を妨げずに、いかに新たなリスクを管理するか?」という重要な問題に直面しています。

そのために求められる要件は明確です。組織は次のような環境を整える必要があります：

- すべてのAI利用状況が把握できる環境
- すべてのAIとのやり取りが安全に保護されている環境
- すべてのAIポリシーが適用・管理できる環境
- すべてのAIモデルを不正利用から守れる環境

Cloudflare AI Security Suiteは、AI導入に伴うリスクに関する不確実性を取り除くことで、組織が安心してAI活用を加速できるよう支援します。当社の統合プラットフォームは、AIの利用状況を可視化し、ゼロトラストアクセスを適用し、そしてプロアクティブな脅威防御でWebとAPIエンドポイントを保護することで、お客様のAIライフサイクル全体を保護します。さらにデータガバナンス機能が統合されているため、チームはセキュリティを損なうことなく自由にイノベーションを進めることができます。

85%

IT意思決定者のうち、「ITチームによるAIツールの評価を待たずに従業員がAIツールを導入している」と答えた人の割合[1]。

AIセキュリティの課題

AI導入に伴うリスクの高まりにより、AIセキュリティソリューション市場は急速に拡大しています。たとえば、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）などは、AI開発のワークフローを対象とした保護に重点を置いています。こうした初期のソリューションは、確かに重要な防御手段のひとつですが、それだけでは不十分です。企業は、本番稼働後のAIシステムの保護も含め、AIライフサイクル全体の安全性を確保する必要があります。

現在、AIセキュリティの課題はますます明確になっています。現在、開発者はAI機能を構築し、従業員は外部のAIツールを使用し、顧客はAIを活用したAI搭載アプリケーションとやり取りしています。これらの異なる環境をすべて手動で保護することは複雑であり、これらすべてで一貫したセキュリティを維持することはさらに困難です。

従業員によるAIの不適切な利用を防ぐ

組織の従業員が、承認されたAIツールだけでなく、非承認のAIツールも使うようになるにつれて、機密データや業務運用がリスクにさらされる可能性が高くなります。セキュリティチームは、次のような問題に対処する必要があります：

- **AI利用状況の可視性の欠如：**リーダーは、従業員が使用しているAIツール、機密データが処理されている場所、およびこれらのAIシステムがどの業務アプリと連携しているのかを正確に把握できていないことが多くあります。この「見えない領域」が、大きなリスクの温床になります。
- **データセキュリティとコンプライアンスのリスク：**AIの導入によって、組織内のデータの流れが大きく変化します。その結果、個人情報や社内の機密データ、顧客情報などが、意図せずAIシステムに入力・保存され、法令違反や競合への情報漏えいを引き起こす恐れがあります。
- **エージェント型AIのアクセス制御：** 管理すべきは人間によるAIツールの利用だけではありません。エージェント型AIが行うMCPサーバーやその他の重要システムに対するアクセスも管理する必要があります。これには、従来とは異なる新しいアイデンティティ管理とアクセス制御のアプローチが求められます。

公開AIアプリやモデルの保護

社内開発・外部提供を問わず、AIシステムは顧客体験やユーザー体験の重要な柱になっています。そのため、他の重要システムと同様に適切な保護が必要です。「OWASP Top 10 for LLM Applications」では、従来のツールでは対応できない脅威が指摘されています：

- **Unbound型消費攻撃（Unbound consumption attacks）：**従来のDoS攻撃と同様に、Unbound型消費攻撃はリソースを大量に消費するリクエストでLLMを圧倒しようとします。クラウドの「従量課金制」を利用している場合、この攻撃によってコストが急激に増大する可能性があり、正当なユーザーはサービス品質の低下に直面する可能性があります。
- **モデル汚染（Model Poisoning）：**攻撃者は、公開データセットやリポジトリに破損したデータを挿入し、それを開発者がモデルのトレーニングに使用することで、LLMにバックドア、バイアス、脆弱性を埋め込みます。汚染されたモデルは一見正常に動作しますが、特定の条件を満たすと悪意のある動作を引き起こします。
- **プロンプトインジェクション攻撃（Prompt Injection Attacks）：**プロンプトインジェクションとは、AIとのやりとりを介してデータを流出させる手口であり、悪意のある指示をユーザープロンプトまたは外部コンテンツに埋め込むことで、LLMへの入力を操作します。これにより、モデルは本来の指示を無視し、攻撃者の命令を実行してしまいます。
- **ジェイルブレイク（Jailbreaking）：**巧妙に作られたプロンプト（ロールプレイ形式の指示、指示命令の上書き、複数ターンの誘導）を用いて、LLMの安全ガードレールを回避して、禁止されているコンテンツを生成したり、機密情報の抽出を行います。

組織にリスクを与えることなく、AIによる迅速なイノベーションを実現するためには、次の3つの領域をすべてカバーする包括的なAIセキュリティ対策を導入することが重要です：

- 従業員による生成AI利用の保護
- AI搭載アプリとワークロードの保護
- 設計段階からセキュリティを考慮したAI搭載アプリを構築する

AI開発およびトレーニングの作業を安全に守る

AIプロジェクトでは、膨大なデータセット、高コストのリソース、そして反復的な実験が必要となり、これらすべてが新たな攻撃対象領域と脆弱性を生み出します。セキュリティ担当者は次のような課題に対応する必要があります：

- **トレーニングデータのセキュリティと完全性：**トレーニングデータの安全性が損なわれると、バイアス、バックドア、または本番モデルに永続的な脆弱性が生じる可能性があります。組織は、トレーニングデータセットへのアクセスを厳重に管理し、不正な変更を防ぎ、モデルのライフサイクル全体を通してデータの来歴を保証する必要があります。
- **資格情報とシークレットの管理：**AI開発業務では、データセット用のクラウドストレージ、トレーニング用のコンピューティングクラスター、モデルレジストリ、およびサードパーティサービスなどのシステムへのアクセスが必要です。秘密鍵またはAPIキーはの保護が不十分な場合、機密性の高い認証情報の漏洩、独自のモデル、トレーニングデータ、または本番システムへの不正アクセスにつながる恐れがあります。
- **開発環境のアクセス制御：**AIエンジニアは、モデルの実験や機密データの利用のために高い権限を持つことがあります。アクセス制御が不十分だと、内部脅威、偶発的なデータ漏洩、モデルの不正持ち出しにつながる可能性があります。

Cloudflareでセキュリティを統合

AIを効果的に導入するには、「制限する」ことではなく「生産性を高める」ことに焦点を当てる必要があります。適切な安全対策が整っていれば、チームは安心してAIを活用でき、イノベーションが加速され、より意欲的なプロジェクトに取り組むことが可能になります。

Cloudflare AI Security Suiteは、AIイノベーションのための安全な環境を構築します。セキュアアクセスサービスエッジ（SASE）とWebアプリケーションセキュリティの機能を統合し、経営層（CxO）が次の2つの領域を安全に接続・保護できるようにします：

- 外部向け（パブリック）のAI搭載アプリケーション
- 社内向け（プライベート）のAIシステムやワークロード

Cloudflare AI Security Suiteは、AIのライフサイクル全体を対象とし、検出とリスク管理からデータ保護、ユーザーアクセス保護、そしてAI搭載アプリケーションと開発ワークフローの保護まで、セキュリティニーズに対応します。さらに、Cloudflareのグローバルネットワークはすべてのインタラクションをリアルタイムで検査およびフィルタリングし、ユーザーとアプリケーション間のデータを安全に保護します。

Cloudflareは問題が発生してから検出するのではなく、AIモデルに到達する前に問題や脅威を未然に防ぐ仕組みを提供します。これにより、セキュリティチームは新たな脅威に対して先手を打つために必要な可視性を得ることができます。

Cloudflareの主要機能

Cloudflare AI Security Suiteは、単一のプラットフォームに、包括的な監視、リアルタイム保護、およびリスクの事前管理を統合した、AIセキュリティに対する包括的なアプローチを提供します。

包括的なAIの検出と可視性

効果的なAIセキュリティには、承認済み・未承認を問わず組織内で使用されているAIリソースを完全かつリアルタイムに把握できるインベントリが不可欠です。Cloudflare AI Security Suiteは、AIモデル、アシスタント、エージェント、そしてシャドーAIまで、あらゆる環境（公開環境・社内環境・プライベート環境）を対象に継続的な監視と自動検出を行います。

事前予防的なAIのリスク管理

Cloudflare AI Security Suiteは、AI特有の脆弱性、設定ミス、攻撃経路（OWASP LLM Top 10に含まれるもの）を検出・軽減することで、組織を攻撃から保護します。また、アプリケーション信頼度スコアリングは、チームが最も重大なリスクに最初に対処できるように、修復の優先順位付けに役立ちます。

AIを活用したアプリケーションのセキュリティ

SecOpsチームが最新のAI脅威に対応できるよう、Cloudflare AI Security SuiteはAI固有の脆弱性、設定ミス、AIパイプライン内の攻撃経路に対するプロアクティブな脅威検出と軽減機能を備えており、これにはプロンプトインジェクション、データポイズニング、モデルの悪用に対する保護が含まれています。

専用のAIファイアウォールは、生成AIやエージェント型AI、APIエンドポイントを検出してラベル付けを行い、PIIの流出を試みる行為を検出し、悪意のあるプロンプトがAIモデルのパフォーマンスに影響を与えたり、有害なコンテンツや誤った情報でモデルを汚染する前に、それらをブロックします。

生成AIとエージェント型AIワークフローのゼロトラストアクセス

最小特権などのゼロトラスト原則は、人間のユーザーだけでなくAIエージェントにも適用されます。Cloudflare AI Security Suiteは、人間とAI間、AI同士の通信の両方に対してゼロトラストネットワークアクセス（ZTNA）ポリシーを適用できます。さらに、MCPサーバーのアクセスログや制御を一元管理することで、エージェント型AIが許可されたもの（ジャストインタイムワークフローを含む）のみにアクセスできるようにします。

AIに対応したデータ保護

効果的なAIセキュリティには、複数の言語モデルを活用してプロンプトの内容とその背後にある意図を理解するデータ損失防止機能が必要です。Cloudflare AI Security Suiteには、トレーニング、プロンプト、応答全体にデータ損失防止（DLP）機能が組み込まれており、AIモデルとパイプライン内でのPIIの露出、データ漏洩、不正アクセスを防止します。インラインでデプロイされるAPI中心のランタイムセキュリティは、CNAPPによって実現されるシフトレフトアプローチを補完する、高速かつシンプルな第一の防御層として機能します。

データローカライゼーション

LLMおよびAIアプリケーションも、他の種類のデータ環境と同様に規制の対象となります。Cloudflare AI Security Suiteは、承認された地域内でトレーニングデータと推論リクエストを保持するポリシーにより、AIワークロードが地理的および管轄区域の境界を確実に遵守できるように支援します。

AI開発におけるセキュリティ・バイ・デザイン

あらゆる種類のソフトウェアと同様に、AIアプリのセキュリティも開発サイクルの最初から組み込む必要があり、後付けで追加するものではありません。Cloudflare AI Security Suiteは、開発者が設計段階からセキュリティを考慮したAI搭載アプリを構築するためのツールとフレームワークを提供します。

Cloudflare AI Security Suiteが与えるビジネスへの影響

AIの台頭は、単なる進化ではなく、産業革命やコンピューター化に匹敵する根本的な変革です。そのため、AIを迅速かつ効果的に取り入れることは、単に組織の成長を促すだけでなく、組織の存続に関わる重要な問題となります。導入の遅延や安全でない導入は、今や業界やセクター全体の組織にとって存続に関わる脅威となっています。

Cloudflare AI Security Suiteは、競争の激しい環境下で、高まる顧客の期待と市場のニーズに応えるために、安全かつ管理された効率的な変革を可能にします。

- **より迅速なAIイノベーション：**セキュリティによって導入が妨げられずに安全な利用が可能になることで、従業員やチームは新しいAIアプリを積極的に活用し、日々の業務効率を向上させることができます。適切なセキュリティフレームワークにより、開発者が機密データやシステムを危険にさらすことなく、安心して意欲的なAI機能の開発に取り組むことを可能にします。
- **AI関連リスクの低減：**AIを活用したWebアプリケーションに固有のAI関連リスクを含め、AIの導入に伴うリスクの全容を管理することを可能にします。開発中のAIアプリも従業員が機密データを漏洩したり、AIトレーニングセットで機密情報を公開したりしないように、セキュリティを確保します。SecOpsは、AI特有の脅威と脆弱性を事前に特定および軽減し、攻撃対象領域を最小限に抑え、重要なデータとモデルを保護することが可能になります。
- **セキュリティ運用を合理化：**AIセキュリティの状況を一元的に把握、管理できるため、運用が簡素化され、管理が簡素化され、インシデント対応が効率化されます。SecOpsチームは、AI関連のインシデント対応に追われることなく、戦略的イニシアチブに集中できるようになります。
- **堅牢なデータガバナンスとコンプライアンス：**AIに特化したデータ保護制御により、機密情報を保護し、AIライフサイクル全体で変化する規制要件にも対応できます。
- **総所有コスト（TCO）の低減：**既に投資した既存のセキュリティをまとめた統合プラットフォームを活用する方が、AIセキュリティの課題ごとに個別のポイントソリューションを実装するよりも経済的です。

Cloudflare AI Security Suiteの活用事例

Cloudflare AI Security Suiteは、AI導入における不可欠な要件への対応に最適です。

- **従業員によるAIツール利用を安全に管理：**ChatGPTのような一般公開されている生成AIツールと、社内で開発されたAI搭載アプリの両方に対する従業員のアクセスに、ゼロトラストポリシーを適用します。
- **公開されているAI搭載アプリの保護：**チャットボットやレコメンデーションエンジンなど、AIモデルを組み込んだWebアプリケーションやAPIを、機密データの流出やモデルの悪用につながる可能性のある攻撃から保護します。
- **シャドーAIの管理：**組織全体にわたる未承認のAIツールを自動的に検出し、適切な制御を適用することで、管理されたリスクの範囲内でのイノベーションを促進します。
- **AIインタラクションのためのAI活用型DLP：**AIプロンプトまたは応答で機密データが公開されるのを防ぎ、PIIおよび機密情報を確実に保護します。
- **AI開発のセキュリティ：**セキュアな開発をデフォルトのアプローチにし、セキュリティを損なうことなくAI機能を迅速に構築できるよう、エンジニアリングチームにフレームワークを提供します。

実装に関する考慮事項

AIセキュリティは、基本的なセキュリティ戦略であるため、慎重に検討することが重要です。

現在のインフラを活かす	最も効果的なAIセキュリティ導入は、既存のSASEやアプリケーションセキュリティツールを置き換えるのではなく、それらを基盤として活用する方法です。このアプローチは、現在の投資を活用しつつ、AI特有のリスクにも対応範囲を広げることができます。
統合インライン保護を展開する	AIセキュリティにとって重要なのは、ネットワークエッジで悪意のある活動をリアルタイムで阻止できることです。リアルタイムのインライン制御を実装し、APIベースのモニタリングで補完します。
あらゆる領域をカバーする	AIセキュリティ戦略では、従業員によるGenAIツールの利用、AI搭載アプリと業務フローの保護、エージェント型AIワークフローの保護、そしてAI開発ワークフローのセキュリティといった、全体的な要件をカバーする必要があります。
企業規模での拡張を想定する	AI活用の進みに合わせて拡張できるソリューションを選択することが重要です。あるパイロットプログラムで有効な対策も、AIの利用拡大に合わせて組織全体に適用できるものである必要があります。
成功基準を検証する	本格導入の前に、現実のシナリオでソリューションを検証することが重要です。エンタープライズ機能を無料かつ自由に試せるプラットフォームを選択することで、小規模（単一のチームやアプリなど）からセキュリティスイート全体をテストし、その価値を迅速に証明し、成功基準を満たしていることを確認できます。

Cloudflare AI Security Suiteで、次のステップへ

世界的にAIの導入が加速する中、AIを安全に拡張できる組織こそが、大きな競争優位性を手にすることができます。重要なのは、AIセキュリティがAIの利用を制限することではなく、より賢いAI利用を可能にするものであると認識することです。

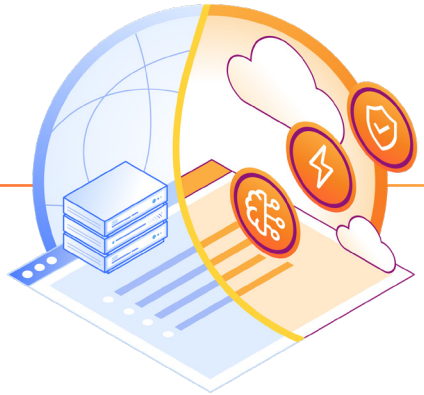
Cloudflare AI Security Suiteは、組織が安心してAIのイノベーションを推進できるよう支援します。広く普及しているCloudflareのSASEおよびアプリケーションセキュリティプラットフォームを基盤とし、それを拡張した当社のAI Security Suiteは、統合されたAI検出、ゼロトラストアクセス制御、インテリジェンス主導の事前的な脅威防御、および堅牢なデータガバナンスを提供します。AIの利用とAIモデルをあらゆる角度から保護することで、企業は、開発者の開発速度アップを図り、従業員の生産性を高めながら、利用者の利便性も損なうことなくAIを活用を進められるよう支援できます。

相談を予約

Cloudflare AI Security Suiteが、安全なAI導入に向けた組織のアプローチをどのように変革できるかをご紹介します。

enterprise@cloudflare.com

www.cloudflare.com/ja-jp



1. ManageEngine。企業におけるシャドーAIの急増：米国およびカナダからの洞察
2. Check Point Research。Check Point Softwareによる2025年第1四半期世界サイバー攻撃レポート：世界全体でサイバー攻撃が約50%急増、ランサムウェア攻撃は126%増加