# AIセキュリティ ギャップ：セキュリティ部門向けゼロトラスト導入ガイド

シャドーAIの検出からエージェント型AIのランタイム保護まで、AIライフサイクル全体を保護するための実践ガイドです。ゼロトラストの原則、導入に向けた30、60、90日のロードマップ、経営層への報告に使用するKPIについて解説します。

EBOOK



©2026 Zscaler, Inc.All rights reserved.

# はじめに

セキュリティ部門は、AIに関する問題を抱えていることは認識していますが、その問題がどれほど深刻なのかまだ把握できていない担当者がほとんどです。自社の環境ではどのような人工知能(AI)アプリが現在稼働していますか？開発者は統合開発環境(IDE)のプラグインやコマンドライン インターフェイス(CLI)ツールを通じて、どのモデルにクエリーを送信していますか？プロンプトを通じてどのようなデータがやり取りされていますか？また、その中にソース コード、顧客の個人を特定できる情報(PII)、あるいは規制に基づく開示ルールの対象となるコンテンツは含まれていませんか？ユーザーに代わって自律的に動作しているエージェントはどれで、どのシステムにアクセスできますか？

**AIが組織環境内で実際に行っていることと組織が把握して管理できていることの間にあるギャップこそが、実際のリスクの温床です。**リスクは、理論上想定される攻撃シナリオにあるのではありません。セキュリティ部門が承認していないにもかかわらず誰も確認しようと思わなかったために見過ごされている、日常的なAIツールの無許可使用や悪意のない使用に潜んでいます。このギャップを生み出す主な要因は、シャドーAIとすでに承認済みのSoftware as a Service (SaaS)プラットフォームに組み込まれたAI機能です。IT部門が特定の目的で使用を承認したツールにAIアシスタントがデフォルトで搭載され、その結果、セキュリティ審査を受けないまま社内データを処理するようになっているのです。

---

本ガイドに掲載されているデータと統計は、*Zscaler*が*2026*年に公開した*2026*年版 *ThreatLabz AI*セキュリティ レポートに基づいています。同レポートでは、*2025*年*1*月から*12*月までの期間に*Zscaler Zero Trust Exchange*上で確認された*9,893*億件の*AI/ML*トランザクションを分析しています。特に明記されていない限り、すべての数値はこの調査期間のデータを反映しています。


©2026 Zscaler, Inc.All rights reserved.

## 原則に基づく対応を提供するゼロトラスト：

アプリ、ユーザー、プロンプト、コネクターのいずれも、最初から信頼できるものとみなしてはなりません。すべてのレイヤーで、アイデンティティーとコンテキストに基づくアクセス制御を施行します。継続的に検証し、問題が発生した場合の影響範囲を最小限に抑えます。課題は、導入の順序です。ポリシーの施行に先立って、可視性を確保する必要があります。自動化を追加する前に、アクセス制御を安定させなければなりません。この順序を誤ると、正当な業務を妨げるか、実際のリスクを阻止できないかのいずれかの制御になってしまいます。

| 83% | 18,033 TB | 39% | 93% |
| --- | --- | --- | --- |
| 組織におけるAI/MLトランザクションの前年比増加率 ── 導入規模はすでにインフラレベルに到達 | 2025年にAIアプリを通じて転送されたデータ量 ── 多くのプログラムでは追跡しきれない規模 | 積極的にブロックされたAIトランザクションの割合 ── ガバナンスへの対応がすでに求められている状況 | AIによるデータ転送量の前年比増加率 ── セキュリティ プログラムの対応能力を上回るペースで増加 |

出典：2026年版 ThreatLabzセキュリティ レポート、Zscaler (pp. 5〜6)

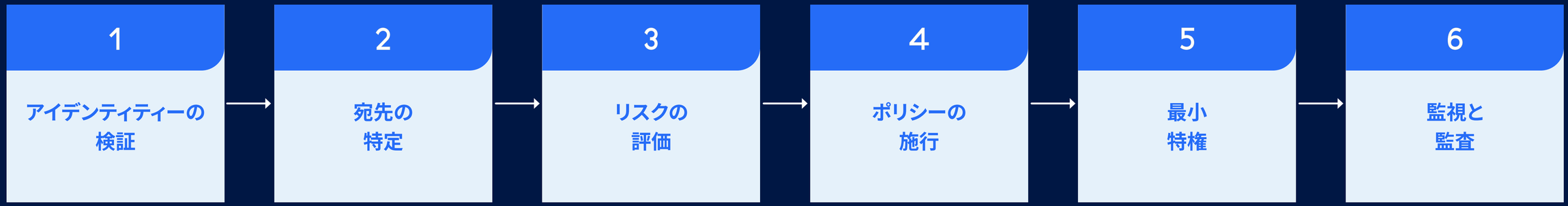
**このガイドの内容**

1. AIシステム、エージェント、データ フローに特化した7つのゼロトラストの原則
2. 依存関係に基づいて構築された段階的な30/60/90日の導入ロードマップ
3. 新たなAIリスク領域への対応範囲：シャドーAI、プロンプト インジェクション、エージェント型システム、モデル コンテキスト プロトコル(MCP)/エージェント間(A2A)通信、SaaSに組み込まれたAI、AIインフラのセキュリティ態勢
4. AIセキュリティ プログラムの進捗を経営層に報告するためのKPIモデル


©2026 Zscaler, Inc.All rights reserved.

# AIに適用されるゼロトラストの原則

以下の7つの原則は、導入ロードマップ全体の意思決定フレームワークとなります。それぞれの原則は、制御を評価するための判断基準です。各原則の順序は、セキュリティ部門が実際の運用で適用する手順に沿っています。




順番に適用します。各原則は、その前の原則を基盤として成り立っています。


©2026 Zscaler, Inc.All rights reserved.

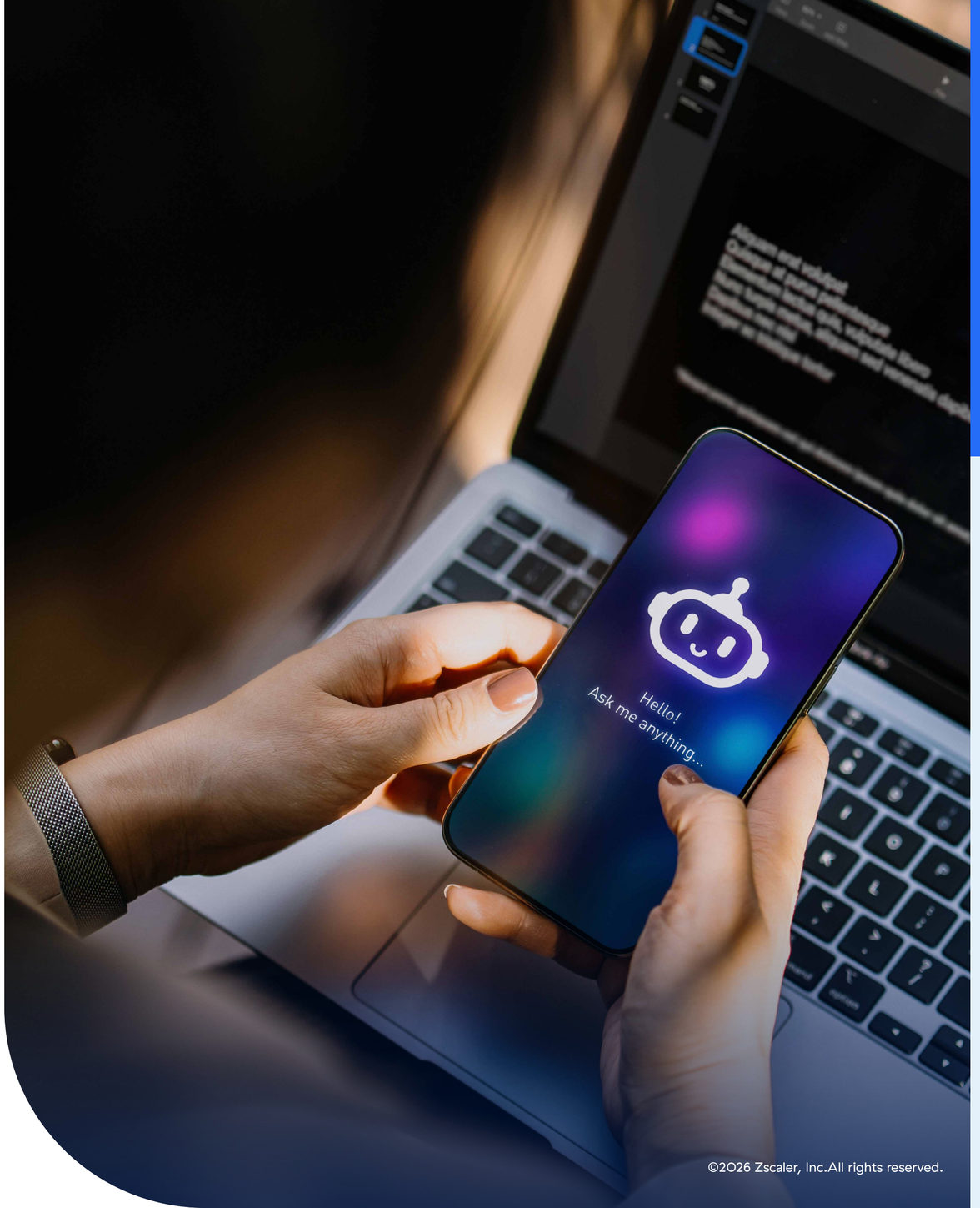
## 原則1:
### ユーザー、開発者、サービスのアイデンティティーを検証

AIへのアクセスは、アイデンティティーの検証から始まります。生成AIのSaaSツールにアクセスするすべてのユーザー、IDEプラグインからモデルのエンドポイントにクエリーを送信するすべての開発者、AIインフラのコンポーネントに接続するすべてのサービス アカウントは、一貫したアイデンティティー制御プレーンを通じて認証される必要があります。シングル サインオン(SSO)と多要素認証(MFA)は基本要件ですが、細部も重要です。

従業員と開発者ではリスク プロファイルが異なるため、それぞれに応じたポリシーを適用する必要があります。マーケティング アナリストがコンテンツの下書き作成にChatGPTを使用する場合と、ソフトウェア エンジニアが社内リポジトリーにアクセスできるMCPサーバーに接続されたGitHub Copilotプラグインを介してIDEでコード補完を実行する場合とでは、データ漏洩のリスクが異なります。AIエージェント向けのサービス間認証は、新たに浮上している課題ですが、ほとんどの組織でまだ対応できる準備が整っていません。

**主な制御**

- AI SaaSと開発者向けAIツールにおいてSSOとMFAを施行する
- 従業員による生成AIの使用と開発者向けAIツール(IDE、CLI、MCP統合)の使用とで、アイデンティティーポリシーを分ける
- **AIエージェント向けのサービス アカウント制御:** 最小特権に基づくアクセス範囲の設定、短期間のみ有効な認証情報、監査ログの記録
- **アイデンティティーのシグナルとしてのデバイスのセキュリティ態勢:** 管理対象のエンドポイントには、個人所有デバイスの持ち込み(BYOD)とは異なるアクセス権限を付与する





©2026 Zscaler, Inc.All rights reserved.

## 原則2:
### 承認されたAIと未承認のAIの宛先を特定

すべてのAIアプリが同じというわけではありません。また、すべてを一律にブロックする方法は運用上現実的ではなく、セキュリティ対策としても効果的ではありません。**目指すべきは、宛先に応じた段階的なポリシーの適用です。**承認されたツールにはインライン検査を実施したうえで完全なアクセスを許可し、許容対象のツールには使用上の制限を設けて監視下でのアクセスを許可します。未承認または高リスクのツールはブロックするか、隔離します。こうした対策の対象となるAIの使用先は、多くの組織が想定するよりも広範囲に及びます。既存のプラットフォームに新たなAI機能が組み込まれるにつれ、その数は1年で4倍に増加しています。

**主な制御**

- 段階的な宛先ポリシーにより、アプリとユーザー グループに応じて許可、警告、隔離、ブロックを使い分ける
- トラフィック内に新たなAIツールが出現した際に、アプリを動的に検出する
- 生成AIのSaaS、組織向けSaaSに組み込まれたAI、開発者向けAIツールについて、それぞれ個別のポリシーを適用する
- 新たなAIツールの使用申請に対する承認ワークフローを整備し、ユーザーが制御を回避するのではなく遂行する

# 3,400以上

**組織のトラフィック全体で追跡されたAIアプリケーションの数**
**1年で4倍に増加し、月ごとに増加**

*2026年版 ThreatLabz AIセキュリティ レポート、Zscaler (p. 5)*




©2026 Zscaler, Inc.All rights reserved.

## 原則3:
**コンテキストに応じた意思決定でリスクを評価**

AIへのアクセス権限を判断する際、二択で考えるべきではありません。同じユーザーが同じAIツールにアクセスする場合でも、デバイス、データの機密性、操作などのコンテキストによってリスクは異なります。効果的なリスク評価を行うには、保護対象とする機密データをあらかじめ定義する必要があります。具体的には、ソース コード、PII、ペイメント カード業界(PCI)のデータ、保護対象医療情報(PHI)、未公表の財務業績、合併と買収(M&A)に関する詳細情報などです。組織のトラフィックから明らかになった主な違反カテゴリーを見ると、従業員が何を入力しているのかが具体的にわかります。入力内容には、氏名や国民識別番号のデータ、ソース コード、医療記録が該当します。

### 主な制御

- ソース コード、PII/PCI/PHI、機密情報、未公開の計画を「保護対象の機密データ」のカテゴリーとして明確に定義する
- ユーザー グループ、部門、デバイスのセキュリティ態勢、データの機密性に基づくリスク スコアリングを実施する
- 静的な許可/ブロック ルールだけでなく、コンテキストに応じたシグナルに基づいてポリシーを動的に調整する
- ファイルのアップロードを伴うAIとのやり取りとテキストのみのプロンプトによるやり取りとで、異なるリスク対策を適用する

# 4億1,000件以上

**CHATGPT単体に関連する情報漏洩防止(DLP)違反の件数 主なタイプ：氏名、SSN、ソース コード、医療データ**2026年版 *ThreatLabz AIセキュリティ レポート、Zscaler (pp. 6、15)*

*2026年版 ThreatLabzセキュリティ レポート、Zscaler (pp. 6、15)*




©2026 Zscaler, Inc.All rights reserved.

## 原則4:
### セッションごとにポリシーをインラインで施行

ログからは何が起きたかを把握できます。開発者が未承認のモデルに独自のソース コードを送信した記録を見つけた時点で、すでにデータは流出しています。多くの組織はAIツールへのファイル アップロードにDLPを導入する一方で、プロンプトのテキストを検査していないため、この問題がさらに深刻化しています。開発者がプロンプト入力欄にソース コードを直接入力すると、アラートが一度も発生しないまま、こうした制御をすり抜けてしまいます。データ転送量の増加は、こうした保護されていない経路を通じて移動しているデータの規模を示しています。

**主な制御**

- URLやカテゴリーのレベルだけでなく、プロンプト、応答、ファイル アップロードに対してもインライン DLPを適用する
- ポリシー施行ポイントで、許可、警告、マスキング、ブロック、隔離の各アクションを実行できるようにする
- 主要な生成AIアプリを対象に、コンテキスト認識型のプロンプト分類を実施する
- 不適切な内容、制限対象、競争上機微なやり取りに対するコンテンツ モデレーションを実施する
- 許可されているもののリスクが高いAIツールには、ブラウザー分離を適用し、全面的に信頼することなくアクセスを許可する

# 93%

**組織におけるAIデータ転送量の前年比増加率。ほとんどのセキュリティ プログラムで追跡および管理できる範囲を上回るペースで増加**

*2026年版 ThreatLabz AIセキュリティ レポート、Zscaler (pp. 5、14)*




©2026 Zscaler, Inc.All rights reserved.

## 原則5:
### アクセスと漏洩を最小化

AIシステムは、従来のアイデンティティーやネットワーク モデルでは把握しきれないアクセス経路を生み出します。検索拡張生成(RAG)パイプラインによって、モデルがドキュメント コーパスにアクセスできる場合があります。エージェントが社内ツール、データセット、ベクトル ストア、MCPサーバーに接続する場合もあります。こうした権限は、標準的なアクセス レビューの対象に含まれていなくても、ユーザーのアクセス権限と同様に実質的な意味を持ちます。

アクセスと漏洩を最小限に抑えるには、AIシステムがアクセスできる範囲を縮小し、AIコンポーネント間の通信を制限する必要があります。エージェント、データ パイプライン、RAGコンポーネント、バックエンドのデータ ソースは、明示的に承認された経路を通じてのみ接続してください。最小特権に基づくアクセス範囲の設定とセグメンテーションを行わなければ、過剰な権限を付与されたエージェントや侵害されたAIコンポーネントが、ラテラル ムーブメントの経路となる恐れがあります。

**主な制御**

- AIアプリ、エージェント、モデル コネクター、データセット、ベクトル ストア、AIインフラの各コンポーネントに対して、ロールベースのアクセス範囲を設定する
- 機密データセットやAI接続システムに対する過剰なアクセス権限を定期的に見直し、削減する
- AIエージェントに付与する権限を最小限に抑え、アクセス可能な対象を明確に定義し、それ以外へのアクセスを制限する
- AIサービス、エージェント プラットフォーム、データ パイプライン、RAGコンポーネント、バックエンドのデータ ソースをセグメント化する
- AIコンポーネント間の東西方向のアクセスを、明示的に定義された経路に制限する
- MCPとA2Aトラフィックに対してプロトコルを考慮して検査し、ログを記録する

## AIのすべてのアクセス経路で被害の影響範囲を最小化

**すべてのAIコネクター、エージェントの権限、RAGパイプライン、ベクトル ストア、MCPサーバー、コンポーネント間の接続を、漏洩につながる可能性のある箇所として扱います。アクセス可能な範囲を限定し、通信可能な範囲をセグメント化し、トラフィックを検査するとともに、アクティビティーをログに記録します。**




©2026 Zscaler, Inc.All rights reserved.

## 原則6:
### 継続的な監視と監査可能性

継続的な検証は、ある時点で達成して終わりにする目標ではなく日々の運用モデルです。AIシステムにおいては、プロンプトと応答のログ記録、ポリシー違反のリアルタイム アラート、使用状況のダッシュボード、そして5つの異なるツールのログを手作業で突き合わせることなく調査を進められるワークフローが必要です。**その背景には、現実的な規制上の要請があります。**欧州連合人工知能法(EU AI法)、米国国立標準技術研究所の人工知能リスク管理フレームワーク(NIST AI RMF)をはじめとする新たなAIガバナンスのフレームワークでは、制御が実際に機能していることを示す文書化された証拠が求められます。

**主な制御**

- インラインでのAIとのやり取りにおいてプロンプトと応答のログを記録する
- リスク シグナルを重ねて表示する、ユーザー、部門、AIアプリ別の使用状況ダッシュボード
- AI関連のポリシー違反や異常を検知した際のアラート通知と調査ワークフロー
- 監査に対応できるレポート：使用中のAI、適用されている制御、発生した違反
- NIST AI RMF、EU AI法、その他の適用対象となるフレームワークに対して、ガバナンス態勢をマッピングする




©2026 Zscaler, Inc.All rights reserved.

| 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|
| **事後対応** | **可視化** | **ガバナンス** | **自動化** | **継続** |
| インベントリーなし<br>場当たり的なブロック | AIの利用状況を把握<br>基本ポリシーを整備 | ロールベースの<br>ポリシー ガバナンス<br>ワークフロー | 教育の自動化 継続的<br>なコンプライアンス | レッド チーム演<br>習＋ランタイム<br>ガードレール |

プログラムの成熟度 →

## 1〜30日目
可視性の確立

可視化できていないAIを保護することはできません。最初の30日はすべてを厳しく制限するのではなく、十分なインベントリーを整備し、基本的な制御を確立することで情報に基づいた意思決定を行える状態を目指します。

| 1〜30日目<br>可視性の確立 | 31〜60日目<br>ポリシーの成熟度の構築 | 61〜90日目<br>自動化と検証 |
|---|---|---|
| • AIアプリの検出を導入<br>• 保護対象の機密データ カテゴリーを定義<br>• 基本的なアクセス ポリシー<br>• 最初の経営層向けダッシュボード | • ロールベースのアクセス ポリシー<br>• ガバナンス ワークフローの運用を開始<br>• AI-SPMを展開<br>• プライベートAIを保護 | • ユーザーへの教育を自動化<br>• 継続的なコンプライアンス<br>• レッド チーム演習の実施頻度<br>• クローズドループを備えたガードレール |

30/60/90日の導入ロードマップ概要





©2026 Zscaler, Inc.All rights reserved.

# 30日の導入チェックリスト

- **AIツールの定期的なレビュー サイクルを確立:** トラフィック内で新たにツールを検出し、例外申請を受け付け、ポリシー変更を実施する
- 経営層向けAIセキュリティ ダッシュボードを初めて構築:AIアプリのフットプリント(承認済みとシャドーAI)、カテゴリー別のDLPイベント、ポリシー アクション、推移を示すグラフ
- AI関連で最も一般的なセキュリティ イベントに対応するインシデント対応手順を策定:プロンプトに機密データを入力し、ブロック対象の宛先にアクセスするとともに、AIツールへの大容量ファイルをアップロードする
- **AIセキュリティ ログを一元管理:** ユーザーのアクティビティー、アプリの使用状況、プロンプト/応答のイベント、ポリシー アクション、違反を1か所で確認できるダッシュボード

## 運用化

- **最初の2週間で収集したテレメトリーに基づいてDLPポリシーを見直して調整:** 誤検知を減らし、確認されたデータ漏洩のパターンに対する検出精度を改善する
- **許容される使用方法を徹底するため、コンテンツ モデレーションを活用:** 不適切なやり取り、制限対象のトピック、競争上機微なコンテンツ、不適切な出力
- **許可されているもののリスクが高いAIツールにブラウザー分離を追加:** ユーザーはツールにアクセスできるが、ブラウザーのレイヤーでアップロード、ダウンロード、コピー&ペーストを制限する
- **DLPの対応範囲を保護対象の機密データ全体に拡大:** ソース コード(言語とパターンを追加)、PII (明確な形式のもの以外も対象)、PCIに該当するカード会員データ、PHI、機密性の高い業務コンテンツ

## データ保護の拡大

- **AIの許容される使用方法を示す信号機方式のガイドを公開:** 緑(承認済み)、黄(制限付きで使用可能)、赤(理由を明示してブロック)。例外申請の手続きも可視化する
- プロンプトに対するインラインDLPを有効にし、保護対象の機密データ カテゴリーについて検出精度の高いルールから適用を開始:ソース コードのパターン、認証情報の形式、PIIのスキーマ
- **未承認のツールでリスクの最も高い操作を制限:** ファイルをアップロードし、大量にコピー&ペーストを行い、管理対象外のストレージにAI出力をダウンロードする
- **宛先ポリシーを施行:** 明らかに未承認または高リスクのAIの宛先について、ユーザー グループとアプリに応じて警告またはブロックを行う

## 基本的な制御の適用

- **AIリスクのベースラインとなるインベントリーを整備:** どのアプリケーションがどの種類のデータにアクセスしているか、どのユーザー グループがAIを最も頻繁に使用しているかを把握する
- **保護対象の機密データ カテゴリーを明確に定義:** ソース コード、PII、PCIデータ、PHI、機密情報と認証情報、未公表の財務データ、機密性の高いM&A情報や戦略計画

## 保護対象の定義

- トラフィックの最も多いAIアプリを対象にプロンプトの可視化を活用し、プロンプトや応答を通じてどのようなデータがやり取りされているかを把握する
- どのAIツールが社内システムと連携し、どの種類のデータにアクセスしているかを明らかにする
- **シャドーAIを可視化:** 正式な承認を受けていないなかで実際に使用されているツール
- 環境全体で使用されている生成AIのSaaS、組織向けプラットフォームに組み込まれたAI、開発者向けAIツールを網羅的に把握するため、AIアプリの検出機能を導入する

**考慮すべきトレードオフ**

**最初から厳格にブロックするのではなく警告ポリシーから始めることで、摩擦を抑えながら、DLPルールによって正当な業務がブロックされるようになる前に誤検知を洗い出します。その代償として、警告期間中はデータ漏洩の発生する可能性が残ります。この点を、保護対象の機密データに関するリスク許容度と照らし合わせて判断してください。ソース コードの漏洩が重大なリスクである場合は、導入初期に誤検知が発生する負担を受け入れてでも、コードパターンの検出時に厳格にブロックする価値があります。**

### 30日の成果物

- リスク レベル(承認済み、許容対象、禁止)に分類したAIアプリのインベントリー
- リスクの最も高いAIアプリとユーザー グループに対して運用する初期アクセス ポリシー
- 保護対象の機密データ カテゴリーに対する初期調整を完了した有効なプロンプトDLP
- AIフットプリントとリスク指標を示す初の経営層向けダッシュボード
- AIの許容される使用方法に関するガイドと例外申請の手続きの公開

## 31～60日目
ポリシーの成熟度の構築

31～60日目は、可視化と初期制御を基盤として成熟したポリシー モデル、機能するガバナンス ワークフロー、AIインフラのインベントリーを確立します。




©2026 Zscaler, Inc.All rights reserved.

# 60日の導入チェックリスト

## アクセス ポリシー モデルの成熟化

- **部門ごとにAIアクセス ポリシーを構築：**人事、財務、エンジニアリング、営業では、必要とするAIツールもデータ漏洩のリスク プロファイルも異なる
- **条件に応じたポリシー制御を追加：**管理対象デバイスとBYOD別のアクセス レベル、地理的なアクセス情報、デバイスのセキュリティ態勢に基づくリスク スコアリング
- **AIアプリのアクセス レベルを3段階に正式に設定：**承認済み(インライン検査による完全なアクセス)、許容対象(使用制限を設けた監視下でのアクセス)、禁止(例外申請の手続きを伴うブロック)
- 初期30日に受け付けた例外申請を成熟度が高まった現在のポリシー モデルに照らして見直し、終了する

## ガバナンス プロセスの確立

- **AIガバナンス要件を社内ポリシーと適用対象となるフレームワークにマッピング：**NIST AI RMF、EU AI法、PHIが対象となる場合のHIPAA
- **AIツールの承認ワークフローを正式に整備：**申請、セキュリティ レビュー、リスク分類、条件付き承認、定期的な再検証を実施する
- AIの使用に関連するデータの種類について、標準的な機密性ラベルを定め、各ラベルに対応する明確な取り扱いルールを策定する
- **AIセキュリティ インシデントのエスカレーション経路を定義：**対応責任者、経営層への報告が必要となるインシデントの基準、保全すべき証拠

## AIインフラのインベントリー

- AIモデル、エージェント、サービス、データセット、ベクトル ストア、パイプラインをインベントリー化し、データセットからランタイムの使用に至るまでのリネージを追跡できるAI-BOMを構築する
- **AI-SPMの評価を実施：**AIインフラの各コンポーネントにおける設定ミス、過剰な権限付与、機密データの漏洩、脆弱性を特定する
- **AI-SPMで検出された高リスクの修復を優先：**過剰な権限付与、機密データの漏洩、アクセス経路が制限されていないAIコンポーネント
- **主要なAIプラットフォーム全体で対応範囲を検証：**クラウドでホストされるモデル、組織に導入されたエージェント プラットフォーム、MCPサーバーとの連携、管理対象外のAIインフラ

## プライベートAIアプリケーションの保護

- **プライベートAIアプリの制御にプロンプト インジェクション検出を追加：**モデルの指示を上書きしたり、機密情報を引き出したりすることを目的に、敵対的に細工された入力を検出する
- **RAGパイプラインにデータ ポイズニングの監視を導入：**モデルの出力に偏りを生じさせる恐れのある、ドキュメント コーパスへの不正な変更を検出する
- **異常検知機能を備えた、プライベートAIシステムのプロンプト/応答のログ記録を活用：**通常とは異なるクエリー パターン、想定外の出力内容、不適切なやり取りを検知する
- 機密性の高いプライベートAIアプリケーションを対象に、モデルとのやり取りをレビューするワークフローを確立する


©2026 Zscaler, Inc.All rights reserved.

**60日の成果物**

- 条件に応じて施行する、成熟したロールベースのアクセス ポリシー モデル
- AIガバナンス ワークフローを運用：ツールの承認、機密性ラベルの付与、エスカレーション経路の明確化
- AI-SPMの可視化に向けたベースラインの整備完了と、優先順位を付けた修復対応項目のバックログ
- 主要なAIインフラ コンポーネントを網羅したAI-BOM
- 優先度の最も高いプライベートAIアプリに対する保護
- コンプライアンスの報告頻度の定義と初回のレポート作成

## 61～90日目
自動化と検証

**61～90日目はプログラムを自律的に継続できる体制を構築：**施行の自動化、継続的なコンプライアンス監視、レッド チーム演習による敵対的な検証を実施します。




©2026 Zscaler, Inc.All rights reserved.

# 90日の導入チェックリスト

## 施行の自動化

- **ポリシー違反に対するユーザー向けガイダンスを自動化：**ブロックに移行する前に、コンテキストに応じたガイダンスとともに警告する
- **違反が繰り返された場合に備え、ITSMまたはSOARのチケット管理と統合：**同じ制御に繰り返し抵触するユーザーについて、調査チケットを自動的に作成する
- **高リスクの振る舞いを検知した際に、自動的に隔離するトリガーを設定：**機密データの大量送信、禁止対象のAIの宛先へのアクセス、異常なプロンプト パターン
- **フィードバック ループを構築：**ポリシー違反の傾向を時系列で追跡し、施行に失敗しているのではなくポリシーの調整が必要であることを示す行動パターンを特定する

## コンプライアンス報告の定着

- **AIのコンプライアンス態勢を常時監視するよう設定：**有効な制御、保護対象となっている資産、対応範囲にギャップがある箇所を把握する
- **制御を適用対象となるフレームワークにマッピング：**NIST AI RMF (生成AIリスクに関するNIST AI 600-1など)、EU AI法の義務、GDPR、PHIが対象となる場合のHIPAA
- **監査に対応できるレポート一式を作成：**AIアプリの使用概要、プロンプトDLPのアクションとその結果、承認済みと期限切れのポリシー例外、AI-SPMで検出された問題の修復状況
- **AIセキュリティ プログラムを四半期ごとにレビューする体制を確立：**ポリシーの有効性、対応範囲の不足、新たなリスク領域、ロードマップの見直し

## レッド チーム演習と強化

- **社内開発のAIアプリケーションを対象とするAIレッド チーム演習の定期的な実施体制を確立：**既成の攻撃プローブ ライブラリーに加え、各アプリケーション固有のコンテキストを狙ったカスタム テスト ケースを用意する
- **AIの攻撃対象領域全体を網羅するテストを実施：**プロンプト インジェクション、ジェイルブレイク、コンテキスト ポイズニング、PIIの抽出、ソース コードの漏洩、不適切な動作、モデル ベンチマークの回帰
- **本番環境のAIシステムにランタイム保護ポリシーを導入：**プロンプト インジェクション、データ漏洩(PIIとソース コード)、ジェイルブレイクの試み、コンテンツ モデレーション ポリシー違反の検出機能
- **レッド チーム演習で得られた結果をランタイム ガードレールに自動的に反映：**テストと本番環境での施行の間にあるギャップこそ、多くのAIアプリケーションのセキュリティ プログラムが機能不全に陥る原因となっている

| 16分 | 72% | 100% |
|---|---|---|
| AIレッド チーム演習で重大な不具合が初めて発生するまでの時間の中央値<br><br>25以上の環境で222,000件以上の敵対的攻撃を実行 | 初回テストで重大な脆弱性が明らかになった組織の割合<br><br>成熟した環境であっても、導入初日から重大なリスクが存在 | テストで少なくとも1つの重大な脆弱性が確認されたAIシステムの割合<br><br>初期状態で安全なAIシステムはなく、継続的なテストが不可欠 |


©2026 Zscaler, Inc.All rights reserved.

### アーキテクチャーの整合性検証

- **すべてのAIアクセス経路を6つのゼロトラストの原則に照らしてマッピング：**従業員、開発者、AIインフラの各アクセス領域で、各原則がそれぞれ適用されていることを確認する
- **制御の境界に存在するポリシーのギャップを特定および解消：**ある制御レイヤーの適用範囲が終わり、別のレイヤーの適用範囲が始まる境界で一貫した保護が確保されていない箇所
- アイデンティティーの検証、宛先制御、リスク評価、インラインでの施行、アクセス権限と漏洩の最小化、監視を網羅する統一されたポリシー モデルを通じて、施行を標準化する

### 90日の成果物

- 違反への自動対応とユーザーへの教育ワークフローの運用
- フレームワークとの整合性を文書化したうえで宇運用する継続的なコンプライアンス監視
- 経営層への報告と監査に使用するガバナンス レポート一式
- AIレッド チーム演習の定期的な実施体制の確立と、ランタイム保護ポリシーの適用
- レッド チーム演習の結果を本番環境のガードレールに反映するクローズドループ処理
- プログラムの評価指標と継続的なガバナンスに向けた四半期ごとのロードマップ





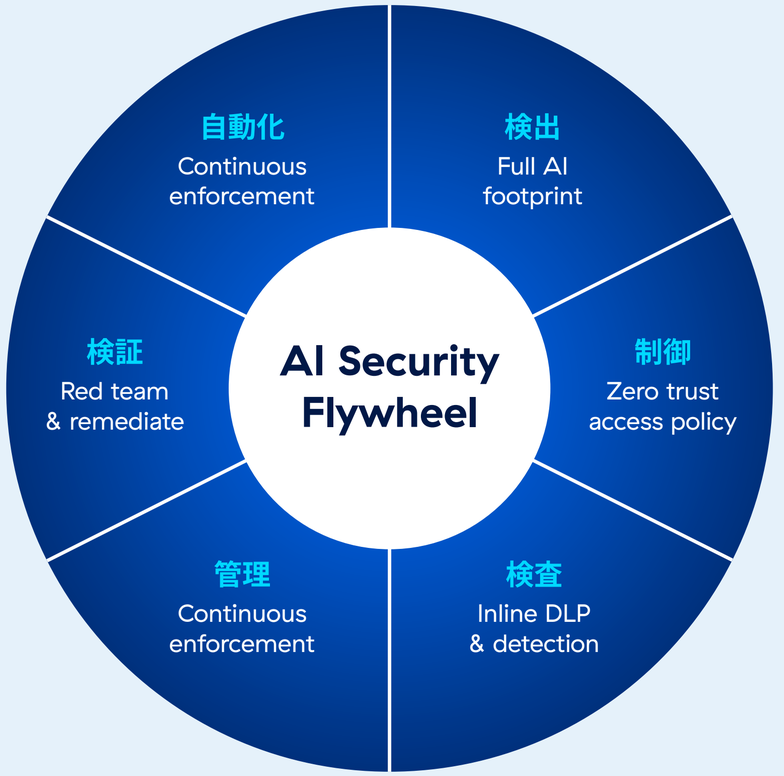
©2026 Zscaler, Inc.All rights reserved.

# Zscalerが安全なAI導入を可能にする仕組み

Zero Trust Exchange™プラットフォームは、組織規模でAIを管理および保護するために構築されました。世界中の数千社の組織ですでにユーザー、ワークロード、ネットワークのアクセスを管理している同じプラットフォームを拡張したものであり、従来のセキュリティ アーキテクチャーに後付けしたものではありません。

右図の各段階は、プラットフォームの具体的な機能に対応しています。以下の表は、Zscalerが各レイヤーで提供する機能を示しています。

| 検出<br>AIフットプリント全体を把握 | 制御<br>AI向けゼロトラスト アクセスの施行 | 保護<br>AIアプリケーションの強化と防衛 |
|---|---|---|
| **AI資産管理**<br>Zscalerは、IT部門が承認していないツールも含め、環境全体に存在するすべてのAIアプリケーション、モデル、エージェント、MCPサーバー、SaaSに組み込まれたAI機能へのアクセスを把握します。 | **AIアクセス セキュリティ**<br>AIのSaaS、組織向けプラットフォームに組み込まれたAI、AI開発環境に対してゼロトラストのアクセス制御を施行し、すべてのレイヤーでインライン検査を実施します。 | **AIレッド チーム演習とAIガードレール**<br>社内開発のAIアプリケーションを実際の敵対的な状況を想定してテストし、その結果をランタイム保護ポリシーに自動的に反映します。 |
| • すべてのユーザー層を対象としたシャドーAIの検出と使用状況の可視化<br>• **AI部品表 (AI-BOM):** データセットからモデル、エージェント、ランタイムの使用に至るまでのリネージ<br>• **AI-SPM:** 設定ミス、過剰な権限、機密データの漏洩に対する継続的なセキュリティ ポスチャー管理<br>• 組織に導入されたエージェントとSaaSに組み込まれたエージェントを対象としたAIエージェントの検出 | • ユーザー グループ、アプリ、データの種類に応じて、許可、警告、隔離、ブロックをきめ細かく使い分けるポリシー<br>• すべてのAIトラフィックを対象とした、プロンプト、応答、ファイル アップロードに対するインラインDLP<br>• 管理対象外のデバイスやBYODからAIツールにアクセスする際のゼロトラスト ブラウザー<br>• 主要な生成AIアプリを対象としたプロンプトの抽出と分類 | • **自動化された継続的なレッド チーム演習:** プロンプト インジェクション、ジェイルブレイク、コンテキスト ポイズニング、データ漏洩、動作のドリフト<br>• プロンプトの強化とモデルのベンチマーク評価、NIST AI RMF、EU AI法、OWASP LLM Top 10へのガバナンスのマッピング<br>• **ランタイム ガードレール:** ジェイルブレイク、PII/ソース コードの漏洩、コンテンツ モデレーション違反、不適切な動作を検出する機能<br>• **クローズドループ ポリシーの生成:** レッド チーム演習の結果を本番環境の検出機能に自動的に反映 |




**プラットフォームの優位性**

**多くのセキュリティ ベンダーは、AIセキュリティ課題の一部分にしか対応していません。Zscalerは、単一のプラットフォームでAIライフサイクル全体をカバーします。環境全体のAIを検出し、ゼロトラスト ポリシーによって誰がどのAIにアクセスできるかを制御し、構築からランタイムまでAIアプリとインフラを保護します。レッド チーム演習の結果を本番環境のランタイム ガードレールに反映するクローズドループは、ポイントソリューションでは再現できない機能です。**


©2026 Zscaler, Inc.All rights reserved.

# よくある質問

**「ゼロトラストAI」とは、実際の運用においてどのような意味を持ちますか？**
ゼロトラストAIとは、暗黙の信頼を前提とせず、継続的に検証するというゼロトラストの基本原則を環境内のあらゆるAIとのやり取りに適用することです。具体的には、AIアプリ、ユーザー、デバイス、プロンプト、コネクターのいずれも、最初から信頼できるものとはみなしません。検証済みのアイデンティティー、デバイスのセキュリティ態勢、データの機密性、セッションのコンテキストに基づいてアクセスを許可します。制御はインラインでリアルタイムに適用されます。すべてのアクセス判断をログに記録し、監査できるようにします。

**シャドーAIのリスクを軽減するために、最初の30日で何をすべきですか？**
まず、AIの使用状況を把握することから始めます。可視化できていないAIは管理できません。ThreatLabzは、組織のトラフィック内で3,400種類を超えるAIアプリケーションを追跡しています。その数は1年で4倍に増加しており、その主な要因は、IT部門がすでに承認しているツールにAI機能が組み込まれるようになったことです。使用状況を可視化したら、リスクの最も高いアプリを分類し、基本的な宛先ポリシーを設定します。リスクの最も高いデータ カテゴリーを対象に、プロンプトに対するDLPを導入します。許容される使用方法に関するガイドと例外申請の手続きを明確に公開します。これにより、ユーザーが制御を回避するのではなく、コンプライアンスを順守できるようにします。

**AIのアクセス制御とインラインDLPにより、プロンプトやアップロードはどのように保護されますか？**
アクセス制御は、どのユーザーがどのデバイスや場所から、どのような条件下でどのAIツールにアクセスできるかを決定します。インラインDLPはコンテンツ レイヤーで機能し、ユーザーが実際に送信するプロンプトのテキスト、ファイルのアップロード、コピー&ペーストによるやり取りをリアルタイムで検査します。この2つの制御は相互に補完し合います。アクセス制御はどのユーザーがどのAIにアクセスできるかを制限し、DLPは許可された経路を通じてどのようなデータが移動するかを制御します。どちらか一方でもギャップであれば漏洩が生じます。そのため、アップロードに対してDLPを導入していても、プロンプト検査を行っていない組織では依然として重大な漏洩リスクが残ります。

**AI-SPMとは何ですか？また、AI導入において重要なのはなぜですか？**
AI-SPMとは、AIインフラの設定、権限、漏洩リスクを継続的に評価するもので、クラウド インフラにおけるCSPMに相当します。AI資産を検出し、セキュリティのベスト プラクティスに照らして設定を評価するとともに、過剰な権限付与や機密データの漏洩を特定し、時間の経過に伴うセキュリティ態勢の変化を追跡します。AIの実験段階から本番環境への導入に移行する多くの組織では、AIインフラにおける設定ミスの発生範囲が広く、その実態も十分に把握されていません。そのため、AI-SPMは重要です。

**AIリスクが低下し、生産性も損なわれていないことを経営層に示すには、どのKPIが最も効果的ですか？**
最も効果的なのは、リスク軽減を示す指標とAIの使用状況を示す指標を組み合わせ、両方の推移を経営層が確認できるようにすることです。シャドーAIアプリの数が減少する一方で承認済みAIの使用が増加していれば、ガバナンスが機能していることを明確に示せます。プロンプトDLPのアクション(ブロック、警告、許可の割合)からは、施行がどのように調整されているかがわかります。例外申請の処理時間やAIツールの承認にかかる時間は、ガバナンスの対応力を示します。


ZSCALER AI SECURITY

- Zero Trust Exchangeプラットフォーム
- AI資産管理
- AIアクセス セキュリティ
- AIを活用したレッド チーム演習
- AIガードレール

Zscaler.com/jp/ai-security





©2026 Zscaler, Inc.All rights reserved.

早く行動し、安全を確保しましょう。

## Zscalerについて

Zscaler (NASDAQ: ZS)は、ゼロトラスト セキュリティの先駆者であり、世界的なリーダーです。世界中の大企業や重要インフラ機関、政府機関がZscalerを採用し、ユーザー、拠点、アプリケーション、データ、デバイスを保護しつつ、デジタル トランスフォーメーションを加速させています。Zscaler Zero Trust Exchange™プラットフォームは世界の160か所以上のデータセンターに分散され、高度なAIを活用して、毎日数十億件のサイバー脅威やポリシー違反を防ぎ、コストや複雑さを軽減しながら、現代企業の生産性向上を支援します。

© 2026 Zscaler, Inc. All rights reserved. Zscaler™および**zscaler.com/jp/legal/trademarks**に記載されたその他の商標は、米国および/または各国のZscaler, Inc.における(i)登録商標またはサービス マーク、または(ii)商標またはサービス マークです。その他の商標はすべて、それぞれの所有者に帰属します。

**+1 408.533.0288** **Zscaler, Inc. (HQ) • 120 Holger Way • San Jose, CA 95134** **zscaler.com/jp**