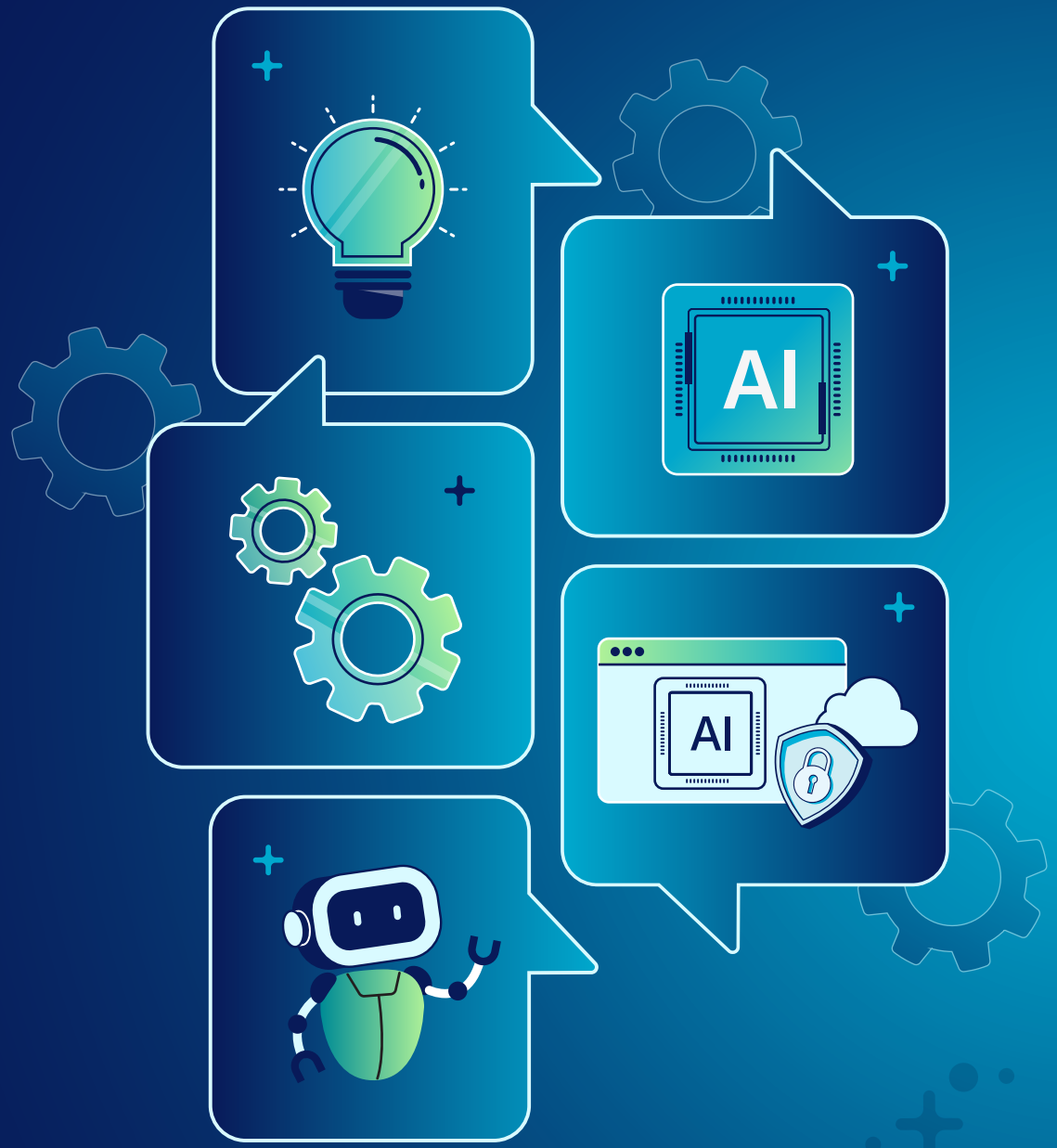
netskope

AI のセキュリティ：CISOのための5つの重要な会話



目次

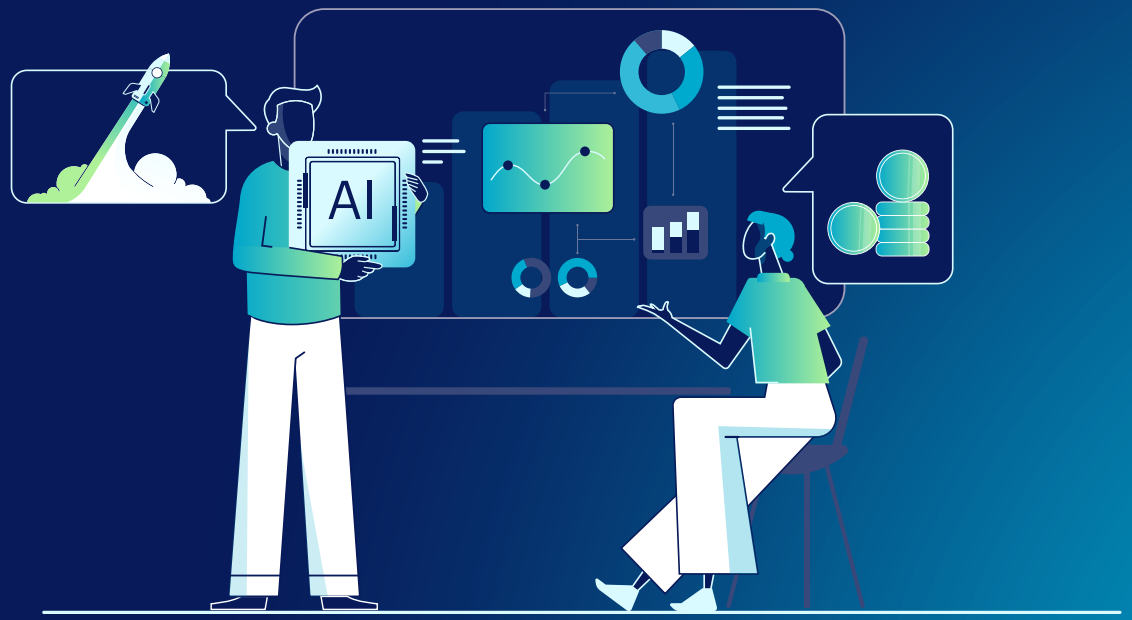
はじめに：AI に関わる2つの相反する役割

現代の組織においてテクノロジーが大きな注目を集める中、IT部門はかつてないほど脚光を浴びています。その結果、IT部門統括者がテクノロジーを効果的に使ってビジネスの成功に最大限貢献する方法を模索する過程で、CEOや取締役会、そして自らと肩を並べる経営陣メンバーと一連の重要な対話を行うようになっています。そして、そうした議論において、AI ほど重要なトピックはありません。

AI は、CIOやCISO、そしてそのチームにとって、特に難しい課題です。Netskopeのレポート『 CEOとの極めて重要な対話』[1]では、CEOがIT責任者に対して、2つの相反する役割を求めていることが明らかになりました。それは、AI の試験的な導入を促し、測定可能なビジネスの価値を生み出すこと、同時にコスト削減を図り、過剰な支出を防ぎ、過度な騒ぎに惑わされず、潜在的なデータ漏洩やセキュリティ侵害から組織を守ることです。

要するに、ITリーダーは AI を活用して劇的なイノベーションを実現すると同時に、それがもたらすリスクから会社・組織を守らなければなりません。この二面性のある役割を果たすべきという使命が、ITリーダーの肩に重くのしかかっています。

AI の成熟度は組織毎に異なります。AI の効果を発揮できるユースケースを模索している組織もあれば、独自のAIアプリを開発し、従業員にこれらのツールを広く活用するよう促しながら、着実に前進している組織もあります。どの組織も、AI を活用して成長を加速させようとしていますが、その出発点もスピードも企業によってさまざまです。



[1] https://www.netskope.com/crucial-conversations

AI 導入に向けた5つのステップ

企業がAI 導入のどの段階にあるかに関わらず、必ず考慮すべき重要なセキュリティ上の課題がいくつかあります。重要なのは、各段階におけるリスクを十分に理解した上で、セキュリティ戦略を策定することです。

1. シャドー AI のリスクを管理しつつ、AI ツールを試験的に運用することは可能か？
2. 承認されていないデータ共有を許すことなく、SaaSプラットフォームに組み込まれた AI を活用できるのか？
3. スタンドアロンの AI アプリをどのように管理して、データ漏洩を防ぐのか？
4. プライベート AI アプリを構築するにあたり、AI モデルによる有害または偏見ある応答、さらにアプリケーションの一般的な脆弱性をどのように回避するのか？
5. 自律型エージェントを展開する際、過度なアクセス権限の付与をいかに防ぐのか？

試験的導入

SaaSプラットフォームに組み込まれた AI

スタンドアロン AI アプリの許可制限

プライベート AI アプリ

自律型エージェント

経営陣とのより建設的な対話に向けて

AI は、今日のIT業界で話題の中心となっているにとどまらず、経営陣や取締役会にとっても最重要課題の一つです。CEOを対象とした弊社の調査[1]から、CEOは AI の可能性に大きな期待を寄せており、IT部門統括者には業界の過度な熱狂に惑わされることなく、適切な場面での導入と統合に向けた道筋を築くことを求めていることが分かっています。

IT実務者、特にセキュリティ部門にとっての課題は、まずパフォーマンスとセキュリティのいずれも妥協せずに AI を導入すること、そしてコストと複雑さを抑制しながらコンプライアンスを守ることです。同様に、導入計画がより詳細になるにつれ、ビジネス上のメリットとリスクを常に明確化しておくことが極めて重要になってきます。

本eBook『AI のセキュリティ：CISOのための5つの重要な対話』を通じて、上記の条件を満たした AI 導入に貢献できることを願っております。本eBookは、セキュリティチームがAI導入の道のりを自信を持って進み、AI に関わるより建設的な対話を促進する一助となることを目的としています。究極の目的は、組織において AI の原則をセキュリティ戦略に、セキュリティの原則を AI 戦略に組み込むことを支援すること、さらにはセキュリティを「事業の阻害要因」ではなく「成長の原動力」にすることなのです。

AI による混乱とリスクに対する防御：優先すべき三原則

AI導入のどの段階であっても、新たな技術のポテンシャルを安全かつ最大限に活用するために適用すべき三原則があります。

1. **可視性：**セキュリティチームが AI 環境の全体像を完全に把握し、どの AI ツールがどのように使用されているかを理解すること。
2. **保護：**実務担当者はコンテキストベースの保護を実行することで、ダイナミックかつ適応性の高いセキュリティ対策がイノベーションを阻害することなくビジネスを保護できること。
3. **AI 導入の準備度：**セキュリティ担当者は、積極的にデータやアプリケーションについての理解を深めることで、組織の成功につながる AI 導入準備をすること。

[1] https://www.netskope.com/crucial-conversations

ステップ1：試験的導入

シャドー AI のリスクを管理しつつ、AI ツールを試験的に運用することは可能か？

2022年11月にChatGPTがリリースされると、世界中で大旋風を巻き起こし、企業を驚かせました。ほぼ即座に、従業員たちは業務の効率化や課題解決のため、個人アカウントの AI チャットボットを使い始めました。今日においてもシャドー AI は多くの組織にとって継続的な問題です。Netskope Threat Labsの調査[1]で明らかになったのは、2025年になんと 72 %もの企業ユーザーがいまだ職場において、個人アカウントを使って ChatGPT や Google Gemini、その他の人気の生成 AI アプリにアクセスしている、ということでした。

シャドー AI の問題は複雑さを増すばかりです。既存の SaaS アプリケーションのほぼすべてにAI 機能が組み込まれており、AI モデル同士が直接やりとりをするようになり、自然言語でエージェントを作成できることから AI はもはや技術に精通した人だけのものではなくなりました。さらに、これら AI インスタンスのすべては、人間の処理量をはるかに超えた多くのデータやアプリケーションとやり取りを行っています。その結果、「シャドー AI」がかつてない速さで拡大しています。

企業ユーザーの 72%はいまだ職場において、個人アカウントを使って生成AIアプリにアクセスしています。

Netskope『クラウドと脅威レポート：ジェネレーティブ AI 2025』

[1] https://www.netskope.com/resources/reports-guides/cloud-and-threat-report-generative-ai-2025

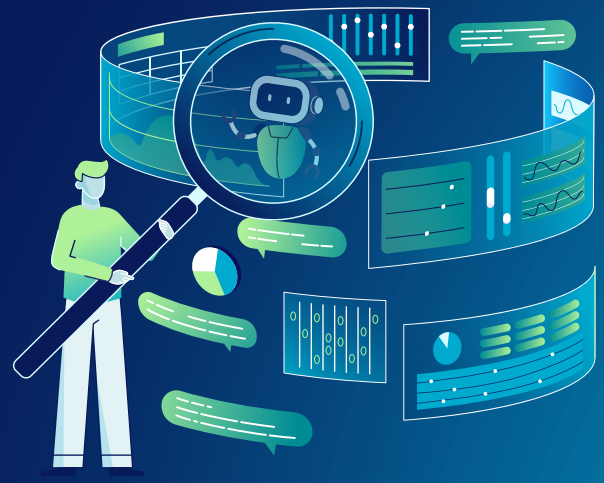
自社のAI環境について幅広くかつ深く理解することはセキュリティチームにとって喫緊の課題です。管理対象外のアプリや個人用インスタンスを含め、組織内で使用されているあらゆる AI ツールについて、可視性を担保しなければなりません。さらには、ユーザーやエージェントが各やりとりのなかで何を行っているのかを見極め、理解しなければなりません。

実態をそれほどにまで深掘りしてはじめて、セキュリティチームはいわゆる「根拠のない信頼」から脱却し、組織内の AI 活動に対して戦略的な統制力を発揮できるようになるのです。

見えないものは守れない。それが、厳しい現実です。

Netskopeのソリューション

Netskope Oneのクラウドアクセスセキュリティブローカ（CASB）および次世代セキュアウェブゲートウェイ（NG-SWG）を活用することで、企業は自社環境内における AI の活動状況を詳細に把握することができます。弊社の AI ダッシュボードでは、どのユーザーがどのアプリケーションにアクセスしているか、またどのような操作を行っているかといった詳細情報を確認できます。また、ユーザー／アプリ間のやりとりを含む公開型 LLM とのやり取り（移動中のデータ）すべてについて、インラインでの検出および検査を企業が実施できるよう支援します。



ステップ2：SaaSプラットフォームに組み込まれたAI

承認されていないデータ共有を許すことなく、SaaSプラットフォームに組み込まれたAIを活用できるのか？

AIの観点でいうと、LLMやAI専用アプリだけがリスク要因だった時代は終わりました。技術の進化に伴い、ますます多くのSaaSアプリケーションにAI機能が組み込まれるようになっており、そのアプリケーションもビデオ通話プラットフォームから生産性向上ツール、営業管理システムと多岐にわたります。

こうしたSaaSツールは多くの場合、現代の企業に深く浸透しており、日々の重要な業務機能に不可欠な存在となっています。そのため、これらを遮断したり削除したりすることはほぼ不可能な上、AI機能は生産性を飛躍的に向上させてくれることが一般的であるため、その効果を妨げようと考える企業はまずないでしょう。

新しいAI機能は、ほとんど手間や不便を感じることなく追加されることが通常です。例えば、一般的なアップデートに含んだ形で導入され、データ利用規約に関する情報がほとんど提供されないこともあります。ビデオ通話アプリを例にとれば、デフォルトでAI議事メモ機能が有効になっていて、企業の機密情報を記録・保存してしまう可能性があります。このため、セキュリティチームの盲点となる恐れがあります。

既存のSaaSアプリに新たなAI機能が導入され、その上、自動的に有効化されてしまう可能性があり、セキュリティチームの盲点となる恐れがあります。

セキュリティ担当者は、SaaSアプリケーションの状況を常に把握し、AI 機能やその動作、およびデータガバナンスに関連する契約条件を明確に理解しておく必要があります。把握すべき内容には、各アプリケーションが AI をどのように活用するのか、AI モデルのトレーニングにユーザーのデータを使用するのか、主要な規制に準拠しているのか、また AI 機能を無効化できるかといった点が含まれます。

また、組織として機密性の高いデータの分類や格付けを行うことも強く推奨します。それにより、例えば企業の知的財産や規制対象データを保護するためには具体的なポリシーを厳格に適用しつつ、機密ではない情報についてはよりルールを緩和する、といった運用が可能となります。

Netskopeのソリューション

Netskopeの Cloud Confidence Index（CCI）は、85,000を超えるSaaSアプリケーションを網羅したデータベースで、詳細なリスク情報を提供しています。セキュリティチームはこれらの情報により、どの AI 対応アプリを許可、制限、またはブロックするかという判断ができるようになります。

ステップ3：スタンドアロン AI アプリの許可制限

スタンドアロンの AI アプリをどのように管理し、データ漏洩を防ぐのか？

OpenAI のChatGPT、MicrosoftのCopilot、GoogleのGemini、AnthropicのClaudeなど、多くの企業が自社に適したAIツールをすでに選定しています。全社的に1つのシステムで標準化することには、エンタープライズ向けの機能、学習の定着、セキュリティ保護の面で明らかなメリットがあります。さらに、それ以外の AI システムの利用をブロックすれば、潜在的な攻撃対象領域も縮小できます。

しかし、このアプローチでリスクを完全に排除することはできません。企業向け AI ツールがその真の価値を発揮するためには、社内の他の文書や情報源と連携させる必要があります。そのため、個々のユーザーがアクセス権限のない社内文書からデータを抽出してしまう、つまり組織内でのデータ漏洩を引き起こす可能性があるのです。

エンタープライズAI に過剰に権限が付与されていると、例えばマーケティング部門の社員が製品ロードマップ上の将来の製品機能についてAIに尋ねた場合、その社員が本来アクセスできない機密文書から抽出された情報がAI 経由で提供されてしまう可能性があります。

Netskopeのソリューション

Netskopeは、ランタイム中に発生するAI 特有の脅威から積極的にガードします。ユーザーが機密情報を入力しようとした場合、Netskope Oneのデータ損失防止（DLP）機能が即座に介入し、個人識別情報（PII）、ソースコード、企業秘密などがAI モデルへ入力されるのをブロックします。さらに、ユーザー教育のため、コーチング目的のポップアップ画面を表示します。

同時に、Netskope One AI Guardrailsは、あらゆるやり取りに対してリアルタイムのコンテンツモデレーションを実行します。プロンプトや応答の背後にある意図を分析し、プロンプトインジェクションやジェイルブレイク（脱獄）などの非常に高度な悪意ある攻撃を自動的にブロックします。さらに、Guardrailsは、有害または差別的なコンテンツをフィルタリングし、著作権保護された素材の配信をブロックすることで、責任ある AI の利用を徹底します。これらDLPとGuardrailsの機能を組み合わせることで、企業はユーザーを積極的に指導しつつ、データ漏洩や新たな脅威から AI エコシステム全体を守ることができます。

ステップ4：プライベート AI アプリ

プライベート AI アプリを構築するにあたり、AI モデルによる有害または偏見ある応答、さらにアプリケーションの一般的な脆弱性をどのように回避するのか？

規制の厳しい業界（医療、金融サービス、政府機関など）の組織が、プライベート AI アプリケーションの開発を牽引しています。AI に対する信頼が高まるにつれ、多くの組織は、自社データでトレーニングしたオンプレミス型 AI モデルを採用することで、データの保管場所、プライバシー保護、コンプライアンス、第三者への情報開示に関するリスクを低減しつつ、アプリの妥当性と信頼性の向上を図っています。

すでに Azure経由でOpenAI のサービスを利用している企業は3分の1、Amazon Bedrockを利用している企業が27%、Google Vertex AIを活用している企業が10%にのぼります[1]。これらのエンタープライズ向けプラットフォームはいずれも、一般バージョンよりも強力なプライバシー制御機能と高度な統合オプションを備えた、安全なクラウドベースの AI サービスを提供しています。

その反面、プライベートAI を構築することは、セキュリティ上の責任を組織側に移管することも意味します。AI 特有の脅威や従業員の誤運用に対するランタイムの保護に加え、これらのシステムの設計や導入に使用されるツールにも新たな攻撃対象領域が存在します。これらのツールに組み込まれた安全対策が不十分である可能性があるからです。

オンプレミスで AI モデルを運用する際に考慮すべき重要な点は、そのモデルが脆弱性の影響を受けやすいか否かです。例えば、組織がオープンソースのモデルをカスタマイズする場合でも、セキュリティチームはコードを徹底的にテストし、悪意のあるコンポーネントが混入していないことを確認する必要があります。悪意あるコンポーネントとは具体的には、プロンプトを取り込んだり外部へ送信したりする可能性のあるコードなどが該当します。

もう1点、組織のトレーニングデータに望ましくない情報が含まれていないか確認することも重要です。多くの場合非常に大規模なデータセットを利用するため、その中に偏見のある情報、機密情報、または有害な情報が含まれていないかを確認する必要があります。

Netskopeのソリューション

Netskope One AI Gatewayはプライベートアプリと LLM 間の認証、トラフィック管理、コンテンツ検査を一元化することで、自律的なエージェント型データフローを適切かつ安全な状態に維持します。さらに、Netskope One AI Red Teamingは、CI/CDパイプライン内で自動的に侵害・攻撃シミュレーションを実施することで、カスタマイズしたモデルに対して積極的にストレステストを実施し、プロンプトインジェクションなどの脆弱性を発見します。

Netskope One AI Guardrailsは、すべてのトラフィックをリアルタイムで分析することで、プロンプトインジェクションや脱獄の試みなどの高度な攻撃から防御します。また、コンテンツモデレーターとしての機能も備えており、人間とエージェント双方のやり取りにおいて、有害または差別的なコンテンツを特定し、管理します。

さらに、Netskope One DSPM を利用することで、セキュリティ担当者は、データの場所に関わらず、その可視化と管理が可能になります。これにより、例えば AI モデルのトレーニングに使用される可能性のある機密情報を特定・分類することができるようになります。

[1] Netskope Threat Labs『Netskope クラウドと脅威レポート 2026』

ステップ5：自律型エージェント

自律型エージェントを展開する際、過度なアクセス権限の付与をいかに防ぐのか？

「エージェント型AI」は、AIブームの中で今最も注目されている存在です。事実、多くの評論家がこれをエンタープライズ技術の未来における重要な要素であると主張しており、調査会社Gartner®は、日常的なビジネス上の意思決定のうち、エージェント型 AIが自律的に行っていたものは2024年には0%だったのが、2028年までに少なくとも15%にも上るであろうと予測しています[1]。

本技術の導入はまだ初期段階にあるものの、2025年8月以降のNetskope Threat Labsによる調査では、すでに多くの組織において、AI エージェントを構築しているユーザーやSaaSソリューションの AI 機能を積極的に活用しているユーザーが一定数存在することが判明しています。

例えば、GitHub Copilotを利用している組織は現在 39%、主要な AI エージェントフレームワークから生成したエージェントをオンプレミスで実行しているユーザーがいる組織は5.5%です。調査によると、66%の組織で社内ユーザーがapi.openai.comに対してAPI 呼び出しを行っており、13%は api.anthropic.comに対してAPI呼び出しを行っています[2]。

組織の39%がGitHub Copilotを利用しており、5.5%は一般的なフレームワークから生成したAIエージェントをオンプレミスで運用しています。

Netskope
『クラウドと脅威レポート：シャドー AI とエージェントAI 2025』

[1] Gartner社プレスリリース『Gartner社、2025年の戦略的テクノロジートレンドトップ10を発表』2024年10月21日

[2] https://www.netskope.com/resources/cloud-and-threat-reports/cloud-and-threat-report-shadow-ai-and-agentic-ai-2025

現在、多くの企業は自社のエージェント型 AI 資産の全体像を正確に把握できていません。この分野は急速に進化し、常に新しい機能が追加されているため、シャドーAI 問題全体のなかでもシャドーエージェント型AIという側面はますます重要度を増しています。

AI エージェントの導入が進み、組織全体でのその機能範囲が広がるにつれ、それに伴うセキュリティリスクも倍増していくことでしょう。ITチームが各エージェントが実行するアクションを把握し、権限や実行する活動を管理するための適切な対策やポリシーを整備することが不可欠です。

AI を活用したアプリケーションは、社内アプリケーション、自律型エージェント、およびプライベート環境でホストするLLM間の認証済コミュニケーションに依存します。その過程でモデルコンテキストプロトコル（MCP）やAPIが使用されますが、プロトコル自体は安全な通信経路だったとしても、このような人間を介さない相互通信は、重大なセキュリティ上の死角を生み出します。APIとMCPが機密データやツールとの直接やり取りを許可し、AI エージェントが人間を中心とした従来型のセキュリティを迂回することになります。このような死角は、自律的なやり取りが監視の目をくぐりぬけてしまうリスクを生じ、認証情報の漏洩、悪意のあるツールの改ざん、および不正なデータ流出につながってしまいます。

Netskopeのソリューション

Netskope One AI Gatewayはソフトウェア定義のゲートウェイとして機能し、社内アプリケーション、自律型エージェント、およびプライベート環境でホストするLLM間のAPIトラフィックを傍受・管理し、すべてのリクエストに対してAI Gatewayが生成した有効なトークンを要求することで、認証済みのエージェントのみがLLMと通信できるようになっています。

Netskope One Agentic Brokerは、AIエージェントとデータソース間のMCPトラフィックを解析・保護することでAIコードエディタ、チャットインターフェース、開発者ツールなどのMCP対応アプリケーションを統合的に可視化し、リアルタイムに保護します。これにより、人間とLLM間のやり取りと、マシン間AI ワークフローとのギャップを埋めます。このため、機密性の高い企業データを保護しつつ、エージェント型自動化のスピードと規模を維持し、一貫したセキュリティを確実なものとします。

結論：トレードオフのないAI関連リスクの管理

AIは、CIOやCISO、そのITチームに、現代特有のチャンスをもたらすとともに課題も提起しています。AIはかつてなかったほどビジネス戦略や成長と密接に結びつき、その影響力は拡大しつづけています。その反面、組織のデータ、収益、評判に関わる重大なリスクももたらし、リスクも急速に変容しているため、あらゆる情報漏洩やハッキングによる被害の深刻さを増大させています。

この状況に対応するためには、ITリーダーにとって必要なのは、AIがもたらす破壊的な可能性と防御の要件を理解するための明確な枠組みです。この枠組みがあれば、テクノロジーに詳しくない同僚ともAIについて自信を持って話し合うことが可能になり、AIがもたらす潜在的なリスクを管理し、厳格なコンプライアンス基準を遵守することができるのです。

今日のITおよびセキュリティ実務者にとって、エンド・ツー・エンドでAI導入を保護し、安全なイノベーションを可能にしながら、同時に事業運営のレジリエンスを維持することが鍵となります。

Netskope Oneはひとつの統合プラットフォーム上で、パフォーマンスとユーザーエクスペリエンスのどちらも妥協することなくAI関連リスクを管理し、同時に複雑さを軽減しつつコンプライアンスを確保できます。

初期の実験的導入からAIエージェント導入へと、AI導入の道をつき進む企業がますます増える中、ITリーダーはビジネスイノベーションの促進と実現という重要な役割を担うことができます。安全を確保しながらAIを活用することで、今日のCIOやCISOは事業を推し進め、組織を新たなレベルへと引き上げることができるのです。

Netskope One AI Securityは、AIエコシステムを管理し、データを保護する統合ソリューションです。パブリックSaaS、プライベートAIツール、エージェント型ワークフロー上のユーザーと自動化エージェントを保護します。Netskopeは高性能とコンテキストベースのゼロトラスト制御を組み合わせることで、組織がAIの実験的導入段階のその先に進み、AIの真の価値を引き出すフェーズへと移行することをお手伝いいたします。

CEOがITリーダーに何を求めているのか。詳しくは、こちらのNetskopeのレポート『AI時代にCIOとCEOの連携を実現する方法』をご覧ください。

Netskopeについて

Netskopeは、クラウドおよびAI時代における最新のセキュリティ、ネットワーク、分析分野をリードしている企業です。Netskope Oneプラットフォームのユニークなアーキテクチャにより、ユーザー、デバイス、データの場所を問わず、コンテキストベースのリアルタイム セキュリティを実現し、トレードオフや妥協をすることなくネットワークパフォーマンスを最適化します。Netskope Oneプラットフォームと弊社特許取得済みのゼロトラストエンジン、そしてパワフルなNewEdge Networkは何千社ものお客様やパートナーから信頼されています。これらのソリューションによりリスクの低減、集中型インフラストラクチャの簡素化、そしてクラウド、AI、SaaS、Web、プライベートアプリケーションのアクティビティに対する完全な可視性と制御性を実現しています。

さらにご興味をお持ちですか？

デモをリクエストする

netskope