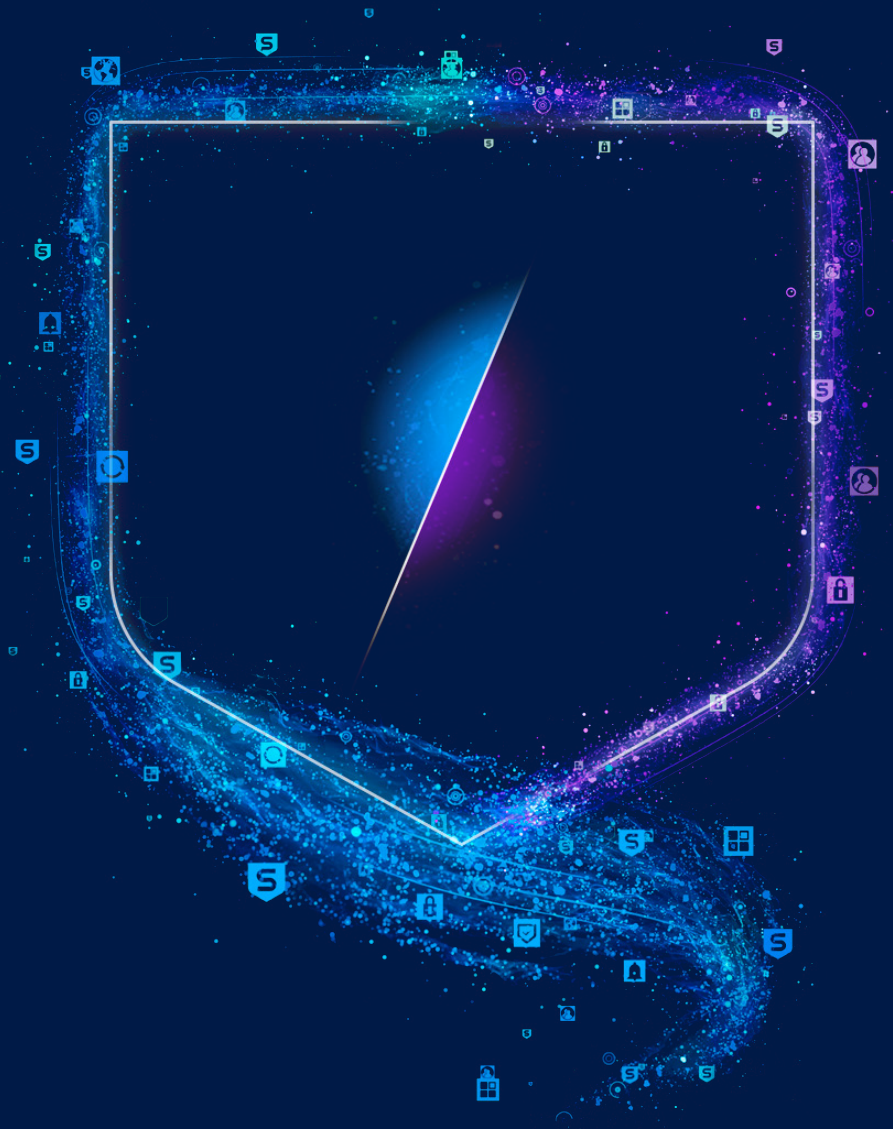
SOPHOS

サイバーセキュリティにおけるAIへの過剰な期待や誤解を解く

組織のサイバー防御を強化するために
AIを安全かつ確実に活用する方法



目次

はじめに

サイバーセキュリティ分野での AI の活用では過剰な期待や誤解が溢れています。AI を活用してサイバーセキュリティを変革すれば、保護の強化、コスト削減、専門家の人員削減といった魅力的な効果が約束されるといった過剰な期待や、AI によってサイバー攻撃の新時代が到来し悲惨な結果となるという誤解も広がっています。

このガイドは、サイバーセキュリティにおける AI をめぐる過剰な期待や誤解を紐解くために作成されています。組織のサイバー防御を強化するために AI が出来ることと出来ないことを明確にし、AI がもたらすサイバーセキュリティと運用上のリスクについて説明します。このガイドでは、これらのリスクを軽減し、サイバー攻撃への防御と投資収益率 (ROI) を向上させるために AI を安全かつ確実に活用する方法についても助言します。

このガイドでは、2024 年後半に IT およびサイバーセキュリティのリーダー 400 人を対象に独立した調査会社が実施した調査結果に基づき、AI の利用実態、期待、懸念に関する洞察を紹介します。サイバーセキュリティの最前線にいるリーダーの視点は、貴重な知見を提供し、AI をどのように活用していくべきか模索している組織にとって有益な比較基準となるはずです。調査結果の詳細については、「過剰な期待や誤解の解消：サイバーセキュリティの AI で企業が直面している現実」を参照してください。

AI を活用するかどうかにかかわらず、サイバーセキュリティの最終的な目標が変わることはありません。究極の目標は、総支出を最小限に抑えながら、組織がビジネスを成功させるために求められるレベルのサイバーレジリエンスを提供することです。言い換えれば、ビジネスをサポートするために、限られているサイバーセキュリティの予算を最大限に活用することが重要です。このガイドは、AI 時代にこの目標に向かって正しく進むために役立つはずです。

サイバーセキュリティにおける AI の利点

AI は、サイバーセキュリティをさまざまな方法で支援し、加速させることができるさまざまな機能を網羅しています。良いニュースは、AI はサイバー攻撃者よりも、防御側に長期的かつ継続的な利点をもたらすことです。サイバーセキュリティでは、ディープラーニングモデルと生成 AI の 2 つのアプローチが一般的に使用されます。

ディープラーニング

ディープラーニング (DL) モデルは、学習した内容を応用してタスクを実行します。AI は、人間の能力をはるかに超えて、知識の応用を加速させることができます。例えば、適切に訓練された DL モデルであれば、初めて見るファイルであっても、瞬時に悪意あるファイルか良性のファイルかを識別できます。

DL は繰り返し作業を大規模に実行するのに適しています。DL は、非常に大規模なトレーニングデータセットから習得した知識に基づいて、新しいアイテムを評価する「統計的」なモデルを作成します。例えば、DL モデルは、マルウェアが含まれるかどうかを判断するために、何百万ものファイルサンプルを一貫性のある方法で正確に評価できます。そのため、DL はサイバーセキュリティ製品の保護能力を強化するために広く利用されています。

DL モデルにより、防御側は、自動化やサービスとしてのサイバー犯罪を利用して攻撃者が生み出す膨大な量の脅威に効果的に対処できるようになります。DL モデルはまた、攻撃の進化に合わせて更新および適合させることができ、最新の脅威環境に常に対応できる状態を維持できます。

生成 AI

生成 AI モデルは、入力を取り込んで、新しいコンテンツを創出します。生成 AI の応用例として、以下があります。

- これまでの脅威活動を自然言語で要約し、アナリストに推奨される次の対策を作成する。
- 検出に関連しているコマンドを分析し、攻撃者の行動に関する知見を提供する。
- アナリストがコードベースのクエリではなく、自然言語検索によって攻撃が疑われる検出結果を調査できるようにする。
- 脆弱性が悪用される可能性に基づいて、パッチ適用の優先度を決定する。

生成 AI はセキュリティ業務を加速させる強力なツールとなります。労働集約型の膨大なデータ処理を生成 AI が引き受けることで、アナリストは迅速に優れた意思決定を下せるようになり、最も影響を与える可能性がある脅威への対応に集中できます。生成 AI は、このような支援によって、アナリストが抱えているプレッシャーや労力を軽減し、燃え尽き症候群や離職のリスクを軽減できます。生成 AI はまた、セキュリティ業務に対する技術的な障壁を低くし、経験の浅いアナリストであっても迅速に貢献できるようになり、スキルの開発を加速させることが可能になります。

生成 AI 導入のアプローチ

現在の生成 AI の基盤はトランスフォーマーです。これは入力 (例えば、文中の単語) 間の文脈や関係を学習するディープラーニングニューラルネットワークで、この学習を活用して関連する出力を生成します。トランスフォーマーは、テキストの翻訳や質問に対する回答など、自然言語処理 (NLP) タスクで多く使用されています。実際、ChatGPT の「T」はトランスフォーマーの略です。

トランスフォーマーは生成 AI で広く使用されていますが、トランスフォーマーを基盤にしたすべてのモデルが「新しいコンテンツ」を生成する能力があるわけではありません。例えば、BERT (Bidirectional Encoder Representations from Transformers) は、入力テキストを双方向で読み取ることができる自然言語処理 (NLP) のためのオープンソースの機械学習フレームワークであり、双方向的に文脈を深く理解できます。このアプローチにより、ラベルのないテキストの文脈の理解を大幅に向上できます。ソフォスは、何年にもわたって BERT を使用してビジネスメール詐欺攻撃を特定して、その攻撃から組織を保護しています。

万能型の AI は存在しない

AI モデルの規模は多岐にわたります。Microsoft Copilot や Google Gemini のような**大規模モデル**は、非常に広範なデータセットで訓練された大規模言語モデル (LLM) であり、幅広いタスクを実行できます。対照的に、小規模なモデルは通常、非常に特定されたデータセットで訓練され、悪意のある URL や実行ファイルの検出などの単一のタスクを実行するように設計されています。**小規模なモデル**は、対象範囲は限定されていますが、大規模モデルよりもコスト、スピード、パフォーマンスに優れています。

AI の制限

AI だけでは解決できない問題があり、少なくとも近い将来は、AI だけに依存することはできません。AI は人間の専門知識を補完しますが、完全に取って代わるものではありません。脅威は非常に複雑であり、効果的にセキュリティ業務を運用するには、技術的なスキルと、組織の状況に合わせて洞察を適用する能力の両方が必要となります。AI だけでは、今日の高度な技術力があり豊富な資金力のあるサイバー犯罪組織の攻撃を防ぐことはできません。

AI の導入状況

AI はすでに多くの組織のサイバーセキュリティインフラに広く組み込まれています。

- 73% の組織が、自社のサイバーセキュリティソリューションにディープラーニングモデルが含まれていると回答
- 65% の組織が、自社のサイバーセキュリティソリューションに生成 AI 機能が含まれていると回答

サイバーセキュリティへの AI の応用は、外部のセキュリティベンダーだけが取り入れているわけではありません。34% の組織がすでに社内でも生成 AI を使用しており、フィッシングテストのメールの生成を支援するなど、サイバーセキュリティの向上に役立てています。

近い将来、ほぼすべての組織が AI を導入する可能性が高く、現在、サイバーセキュリティプラットフォームを選定する際に 99% (四捨五入) 以上の組織が AI 機能を要件として挙げています。

- 57% の組織が AI 機能が「必須」または「極めて重要」と回答
- 41% の組織が AI 機能を「重要」と回答

このような導入状況と今後の利用も拡大することを踏まえ、サイバーセキュリティにおける AI のリスクとそれに伴う対策を理解することは、あらゆる規模や業種の組織にとって優先事項となっています。

73%

ディープラーニングモデルを実装したサイバーセキュリティツールを使用する

65%

生成 AI 機能を実装するサイバーセキュリティツールを使用する

99%

サイバーセキュリティプラットフォームを選択するときに AI 機能が搭載されていることを要件とする

生成 AI：大きな期待

生成 AI を取り巻く誇大広告は、このテクノロジーがサイバーセキュリティの成果の向上について大きな期待を抱かせる結果になりました。調査結果から、組織がサイバーセキュリティツールにおける生成 AI の機能に期待する最大の利点が明らかになりました。これらの期待を下記の表に示します。

生成 AI に最も期待する利点 最も多かった回答から順に表示	
1=	サイバー脅威への保護対策の強化 (20%)
1=	サイバーセキュリティに対する投資利益率 (ROI) の向上 (20%)
3	IT アナリストの効率性と成果の向上 (17%)
4	最先端のサイバーセキュリティ技術を利用している確信を得ること (15%)
5=	自社がサイバー攻撃から守られているという安心感を得ること (14%)
5=	従業員の燃え尽き症候群の軽減 (サイバーセキュリティ従業員の時間を確保するためのタスクの自動化) (14%)

サイバーセキュリティツールに生成 AI 機能を取り入れることによって、どのような利点を得たいと考えていますか？最も多かった回答から順に表示 (回答者数 =400)

さまざまな回答が寄せられており、サイバーセキュリティにおける生成 AI の利用に期待している利点は 1 つではないことが明らかになりました。同時に、最も広く求められている成果は、サイバー保護の強化や財務面および業務面におけるビジネスパフォーマンスの向上に関連しています。また、今回の調査データは、サイバーセキュリティソリューションに生成 AI 機能を組み込むことで、最新の保護機能に追いついているという安心感と自信をもたらすことを示しています。

従業員の燃え尽き症候群の軽減がランキングの最下位に位置していることは、生成 AI がユーザーを支援できる可能性について、組織が十分に認識していない、またはそれほど懸念していないことを示しています。サイバーセキュリティの人材不足が慢性化する中で、離職への対策は重要な課題となっており、AI が支援できる分野です。

保護機能の強化と **ROI** の向上 は、
組織が生成 AI に求めている最大の利点

サイバーセキュリティにおける AI のリスク

サイバーセキュリティにおける AI の活用は、ポジティブな面とネガティブな面の両方があります。AI はサイバー攻撃者との戦いにおいて防衛側に多くの利益をもたらす一方で、多くのリスクが発生する恐れもあります。

1. **脅威のリスク：**サイバー攻撃における AI の悪用
2. **防衛に関するリスク：**品質の低い AI、実装が不十分な AI
3. **運用に関するリスク：**AI への過度の依存
4. **財務に関するリスク：**AI 投資の効果が薄い
5. **AI が乗っ取られるリスク：**攻撃者による公開 AI モデルの侵害

1. 脅威のリスク：サイバー攻撃における AI の悪用

AI がまったく新たな脅威の領域を作り出しているという誇張も多く見られますが、実際にはそれほど劇的な変化は起こっていません。サイバー犯罪フォーラムでは AI についての議論は多くなく、多くのサイバー攻撃者は AI に懐疑的になっています。観察されている範囲では、AI を使用してマルウェア、攻撃ツール、エクスプロイトを開発しようとする多くの試みは、初歩的でありその品質も高くありません。

合法的な企業と同じように、サイバー攻撃グループも主に、コンテンツの品質と業務 (攻撃) の効率性を向上させるために AI を利用しています。最新の脅威環境や AI を利用した攻撃の詳細については、ソフォスのブログをご覧ください。

コンテンツの品質の向上

サイバー攻撃で AI を最もすばやく簡単に利用しやすい領域の一つは、フィッシングメールや詐欺をよりもっともらしく見せて、多くの被害者を信頼させて攻撃の成功率を高めることです。

稚拙な文法、スペルミス、プロフェッショナルではないフォーマットなどの、従来のフィッシングに見られる「詐欺の兆候」は、AI ツールを使用して簡単に排除できます。フィッシングキャンペーンに使用される精巧なメールは、一般の LLM でも 1 分もかからずに作成できます。同様に、メールを受信したユーザーを騙してリンクをクリックさせたり、個人情報を共有させたりすることを目的とした、説得力のある文章やソーシャルメディアのメッセージは、どのような言語でも簡単に作成できます。LLM を使用することで攻撃者は、タイムリーな情報を攻撃に簡単に組み込むことが可能になり、ユーザーが詐欺に引っかかる傾向をさらに高めます。

生成 AI ツールによって、経営幹部になりすまして警戒心の薄い被害者を騙して金銭を振り込ませる新しい詐欺も生まれています。音声クローンテクノロジーは、十分に訓練すれば、第三者になりすまし、ユーザーを騙して本人と話していると信じ込ませることができるところまで進化しています。このようなボイスフィッシング (ビッシング) 攻撃では、サイバー攻撃者が経営幹部になりすまして従業員に電話をかけ、不正なギフトカードの購入、銀行振込、ファイル転送などを「要請」することが多くあります。

攻撃者はまた、AI を活用したディープフェイクテクノロジーを使用して、攻撃で視覚的に第三者になりすます場合があります。ディープフェイク動画は、このようななりすましを疑うことのない従業員を騙して多額の送金をさせたり、ローン申請や銀行口座登録で使用される顔認識プログラムを欺くために使用されています。

業務効率の改善

多くの企業がユーザーエクスペリエンスを向上させるために AI チャットボットを使用しているように、攻撃者も効率性を高めるために AI を利用しています。LLM を利用して、チャットボットや自動応答を作成し、頻繁に利用するフォーラムを強化しているサイバー攻撃者もいます。Sophos X-Ops が調査した例では、XSS に関連するフォーラムで、ユーザー (攻撃者) からの質問に対応するための専用チャットボットが作成されていました。フォーラム管理者のコメント（ロシア語からの翻訳）：

「このセクションでは、AI (人工知能) とチャットができます。質問してください。AI ボットがお答えします。このセクションと AI ボットは、技術的に簡単な問題を解決することを目的としており、ユーザーの技術的なエンターテイメントのため、またユーザーに AI の可能性を理解してもらうために設計されています。

独自のモデルを構築して訓練するためには、高度な AI の専門知識が必要であり、高コストであり、かつ簡単に得られるものではありません。AI の専門知識を有しており専門のチームを確立しているサイバー犯罪組織も存在していますが、サイバー攻撃者は通常、独自に LLM を構築しているわけではなく、既存の LLM を攻撃に転用しています。

攻撃者による AI の利用

AI を使った攻撃者の行動がどのようにサイバーセキュリティに影響を与えるか、AI の利用がどの程度進んでいるのか、またそれがどのような目的で行われているのかを理解することが重要です。AI は攻撃者のツールキットに含まれる数多くのツールの一つに過ぎません。サイバー攻撃者は数年前から、攻撃の規模を拡大し、頻度を増やすために、自動化と「サービスとしてのサイバー犯罪」のモデルを採用しています。多くの組織にとって、AI 自体がリスク管理に与える影響は限られており、他のテクノロジーや戦略の方がより大きな効果をもたらす可能性があります。

2. 防衛に関するリスク：品質の低い AI、実装が不十分な AI

これまで見てきたように、AI モデルはすでに組織のサイバー防衛に広く組み込まれています。その目的は正しいのですが、品質の低い AI や実装が不十分な AI モデルは、大きなサイバーセキュリティリスクを引き起こす恐れがあります。AI モデルがリスクを引き起こす傾向は、以下に示すいくつかの要因に左右されます。

- **モデルを訓練するデータの品質**。「ダメなデータからはダメな結果しか生まれない」という格言がありますが、これは特に AI に当てはまります。品質の低いデータを使用してモデルを訓練すると、誤りを引き起こすリスクがあります。一方で、偏ったデータセットを使用すると、特定の変数が過剰または過小に表出され、出力に偏りが生じる恐れがあります。品質の高いトレーニングデータの量が多ければ多いほど、得られる出力は向上します。
- **モデルを作成するチームの専門知識**。サイバーセキュリティのための効果的な AI モデルを構築するには、別々でありながら補完的な以下の 2 つの分野についての広範な理解が求められます。
  - **脅威：**AI モデルに何をさせるのかを特定するためには、まずマルウェアがどのように動作し、サイバー攻撃者がどのように操作するかを理解する必要があります。
  - **AI：**AI に何をさせるべきかが分かったら、次にその目標を達成するために適切なモデルを特定し、構築していく必要があります。

  サイバーセキュリティで大きな成果を発揮する効果的な AI モデルを構築するには、この 2 つのスキルセットを緊密に連携させ、両方の専門知識を活用することが不可欠です。
- **製品開発と展開のプロセスの品質**。2024 年半ば、あるサイバーセキュリティ製品で欠陥のあるコンテンツアップデートが実施され、世界中の企業で混乱が生じました。十分にテストされておらず、品質評価もされていない AI 機能は、さらに大きな被害を引き起こす恐れがあります。また、問題が簡単に識別や修正されないというリスクも伴います。

サイバーセキュリティへの間違った意識

多くの組織は、十分に開発されていないサイバーセキュリティソリューションの AI が引き起こすリスクに警戒しています。調査対象となった IT およびサイバーセキュリティプロフェッショナルの大多数 (89%) は、サイバーセキュリティツールの生成 AI 機能に欠陥がある場合、組織に悪影響を及ぼす可能性があることを懸念していると回答しており、その内訳は、「非常に懸念している」が 43%、「やや懸念している」が 46% でした。

この調査結果からも、サイバーセキュリティソリューションの生成 AI 機能を評価するときに、以下のように 99% (四捨五入) の組織が生成 AI の開発において使用されるサイバーセキュリティプロセスと管理の品質を評価すると述べていることは驚くべきことではありません。

- 73% がサイバーセキュリティのプロセスと管理レベルを詳細に評価していると回答。
- 27% がサイバーセキュリティプロセスと管理レベルの一部を評価していると回答。

詳細な評価を実施していると報告する割合が高くなっており、一見するとこれは心強く思えるかもしれませんが、実際には多くの組織がこの評価においては重大な盲点を抱えています。

生成 AI の能力を開発するために使用されるプロセスと管理レベルを評価するには、ベンダーの透明性が求められ、評価者には AI に関する相応の知識が必要となります。残念ながら、どちらも不足しているのが現状です。ソリューションプロバイダーが生成 AI 開発のすべてのプロセスをそのまま公開することは稀であり、IT チームは AI 開発のベストプラクティスについて限定された知見しか持っていないことが多くあります。多くの組織は、見逃しているリスクや問題に気づいていない可能性があります。

3. 運用に関するリスク：AI への過度の依存

AI は、スーパーマーケットへの最適な経路を見つけたり、テレビ番組をお勧めしたりするなど、私たちの日常生活のあらゆる分野に深く入り込んでいます。AI が広く普及したことで、安易に AI に依存し、AI が人間よりも優れた仕事をすることが当然だと考える傾向が高まっています。多くの組織が AI への過度の依存がサイバーセキュリティに及ぼす影響を認識し懸念していることは評価できます。

- 84% が、サイバーセキュリティプロフェッショナルの人員削減を迫られることを懸念している。
- 87% が、サイバーセキュリティの責任の所在が明確でないことを懸念している。

こうしたリスクに注意を払うことが、リスクを軽減する第一歩となります。AI は組織のサイバー防御に利用するツールの一つに過ぎないことを忘れないでください。AI はセキュリティスタックの重要な一部ですが、常に最適なアプローチとなるとは限らず、完全な解決策になることはほとんどありません。各組織の環境やニーズは異なります。AI の使用はその組織の全体的なビジネス状況やニーズに応じて適切に行うべきです。

4. 財務に関するリスク：AI 投資の効果が薄い

サイバーセキュリティソリューションに高品質な生成 AI 機能を組み入れようとする場合、開発と維持に多額の費用がかかります。IT とサイバーセキュリティ部門のリーダーは、このような支出に警戒しており、80% の回答者が生成 AI はサイバーセキュリティ製品のコストを大幅に増加させると述べています。

このようなコスト上昇が予想されているにもかかわらず、多くの組織は生成 AI によってサイバーセキュリティ全体の支出を削減できると考えており、回答者の 87% は、サイバーセキュリティツールの生成 AI のコストは、AI がもたらす費用削減効果によって完全に相殺されると確信していました。

同時に、組織はこれらのコストを定量化することが課題であることを認識しています。生成 AI の費用は通常、サイバーセキュリティ製品およびサービスの全体的な価格に組み込まれているため、サイバーセキュリティの生成 AI にどれだけ費用をかけているのかを特定することは困難です。このように十分にコストが可視化されていないことから、75% の回答者がこれらのコストを査定することは難しいと考えています (39% が「強く同意する」と回答し、36% が「やや同意する」と回答)。

投資の成果や進捗を明確かつ正確に伝える報告書やデータがなければ、組織はサイバーセキュリティの AI への投資によって期待されるリターンを確認できない可能性があり、さらに、AI よりも他の分野に投資したほうがより優れた効果が得られる可能性も把握できないままとなります。

5. AI が乗っ取られるリスク大規模言語モデル (LLM) の侵害

AI のサイバーセキュリティリスクがもたらす影響は、サイバーセキュリティツールやアプリケーションだけにとどまりません。LLM が公開され、その利用が世界的に急速かつ広範に拡大したことで、十分な資金力のある狡猾な人物や組織が、自らの目標を達成するために AI モデル自体を侵害する可能性もあります。AI モデルを侵害する場合、以下のいくつかの方法が考えられます。

- **データポイズニング**。2023 年に公開された論文「Poisoning Web-Scale Training Datasets is Practical」では、Carlini 氏らが、データポイズニング (訓練用のデータを操作して、AI モデルの出力に影響を与えること) が実行可能な脅威であることを実証しました。
- **国家主導の攻撃者が仕掛けるバックドア**。多くの国家には、強力な LLM を生み出すための資源があります。秘密のバックドアを AI モデルに追加しておき、その後で公開することによって、国家が支援する攻撃者は必要に応じて生成 AI (LLM) を自らの有利になるように意図的に操作できます。
- **LLM スプーフィング**。サイバー攻撃者は、バックドアを追加するなどの方法で、正規の LLM を侵害し、その変更を「機能強化」として宣伝する場合があります。ユーザーを騙して侵害されたツールを使わせるために、信頼できるプロバイダーを偽装する場合があります。その場合、プロバイダーの名前の 1 文字を省略したり、文字「O」を数字の「0」に置き換えたりしています。

LLM 侵害の詳細については、ソフォスの AI チームによる最新の研究を参照してください。

AI の過剰な期待や誤解を解消するための実践的なステップ

AI はリスクをもたらしますが、慎重なアプローチを採用することでリスクを回避し、サイバー防御を強化するために AI を安全かつ活用できます。これらの推奨事項の多くは、サイバーセキュリティ以外の分野で AI の導入を成功させるためにも役立てることができます。

脅威のリスク：AI 時代におけるサイバー防御のレベルアップ

AI を悪用する脅威へのレジリエンスを高めることに重点を置く必要があります。サイバー攻撃者が、フィッシングメールや詐欺の精度と信頼性を高めることを主眼として AI を利用していることを考えると、これらの分野への対策に力点を置くことは合理的です。ソフォスの提言：

- **メール保護をレベルアップする。**AI が生成したフィッシングメールや詐欺メールを検知し、ユーザーの受信トレイに配信されるのを防ぐことができるソリューションを探してください。
- **ビジネスメール詐欺を防ぎ、VIP を保護する機能を展開する。**例えば、詐欺を検出するためにコンテンツのトーンやスタイルをスキャンでき、ビジネスメール詐欺を防止し、VIP を保護する機能を実装するメールセキュリティソリューションを選択してください。
- **ソーシャルメディアに特に注意する。**SNS を参照しているときにユーザーは、十分な関心を払っていないことが多く、詐欺に遭いやすくなります。
- **音声クローンのリスクを軽減するプロセスを導入する。**予期しない支払いやデータ共有の要求があった場合に従うべき手続きを策定しておきましょう。次のような対策も検討してください。
  - 依頼元に信頼できる方法で連絡して、依頼の内容を確認する。
  - パスコードやパスフレーズを取り入れる。

防衛に関するリスク：サイバーセキュリティ製品に使用されている AI の品質を評価する

セキュリティに投資するときには、品質の低い AI がもたらすリスクと影響に注意してください。ベンダーへの質問事項：

- **訓練データ。**モデルの訓練に使用するデータの品質、量、ソースは何か？優れた入力が優れた出力につながります。
- **開発チーム。**モデルを背後で支援しているチームについて理解してください。そのチームがどの程度の AI の専門知識を有しているのか、脅威、攻撃者の行動、セキュリティ運用についてどれだけの知識があるのかを確認しましょう。
- **製品エンジニアリングとロールアウトプロセス。**ベンダーは自社のソリューション向けの AI 機能を開発して導入する際に、どのようなステップを踏んでいるのかを理解してください。どのような検証と管理が行われているかを確認しましょう。

最終的に、以下を確認しましょう。ベンダーが AI を適切に取り入れており、求められる厳格な品質管理とデプロイ管理を行っていることをどの程度信頼できるのかを確認してください。

運用に関するリスク：AI を人間中心の視点で見る

自社でセキュリティ侵害が発生しても、AI は気にすることはありませんが、自社の従業員には大きな影響が及びます。また、最悪の事態が発生し、セキュリティが侵害された場合、自社のビジネスの状況を理解し、問題による影響を修復できる経験豊富なチームメンバーが必要となります。

- **客観性を維持する。**AI は防御側が利用できるツールキットの 1 つに過ぎません。もちろん利用すべきですが、サイバーセキュリティの説明の所在は、最終的には人間にあることを明確にしておくことが重要です。
- **AI は人材を減らすためのものではなく、人の能力を強化するもの。**熟練したサイバーセキュリティプロフェッショナルが世界的に不足していることは、広く知られています。燃え尽き症候群は大きな課題であり、人材不足に拍車をかけています。 人員削減を目的として AI に注目するのではなく、AI が従業員をどのように支援できるかに最初に焦点を当ててください。AI は簡単で反復的なセキュリティ運用タスクの多くを引き受け、実用的な知見を提供し、以下を実現します。
  - より付加価値の高く、ビジネスに影響を与える仕事により多くの時間を割くことができる。
  - アラート処理にかかる負担を減らし、疲労を軽減できる。
  - 熟練したアナリストの専門的な能力開発を加速する。
  - 経験の浅いアナリストでもセキュリティ業務を担うことができるようにし、リソースパイプラインを構築する。

財務に関するリスク：AI 投資でも、ビジネスの観点から慎重かつ厳密な分析や評価を行う

これは、組織が最も対策を取りやすい領域の 1 つであり、多くの要因を完全に自社でコントロールできます。

- **目標を設定する。**AI に求める成果を、明確に、具体的に、詳細に説明できるようにします。
  - 何が必要なのかを特定します。現在のギャップ (問題) は何か？ AI はどのようにその問題を解決できるか？
  - 経済、時間、セキュリティ保護にもたらされる利益を考えてください。
- **利益を定量化する。**AI 投資によってどれほどの違いがもたらされるのかを理解します。
  - 目標がサイバーセキュリティの全体的なコスト (TCO) を削減することであれば、AI 導入によるコスト削減の効果を定量化してください。
  - AI によって IT/ サイバーセキュリティ部門の人員を削減したいと考えているのであれば、AI ツールがチームに具体的にどのような影響を与えるのかを明確に検討してください。AI によってどのような業務の負担が削除されるのか、従業員の作業時間がどれほど解放されるのかを確認しましょう。
- **投資先の優先順位を決定する。**AI はさまざまな形で組織を支援できますが、投資の効果がより高い分野もあります。自社にとって重要な指標 (財務的な節約、従業員の離職率の低下、リスクの軽減など) を特定し、さまざまな選択肢を比較して、優先順位を決定してください。
- **影響を測定する。**投資には目的があるものですが、実際のパフォーマンスが元々の期待を満たしているかを確認することが重要です。期待していた利点が得られているか？予期しない利点はあったか？そして、期待した結果が得られていない領域はなかったか？など、得られた洞察を活用して、必要な調整を行ってください。

目標達成のために AI が最適な方法であるのか、また、別のテクノロジーやアプローチがより大きな影響を与えることができるのかを検証してください。

AI が乗っ取られるリスク：危険への警戒を怠らないこと

これは、組織が軽減することが最も難しいリスクです。しかし、このリスクについて警戒するだけで、影響を軽減することができます。しかし、公開されている LLM を選ぶときには、以下の点に注意してください。

- **実績がある評判の良いプロバイダーのモデルを利用する。**データポイズニング攻撃を防ぐことは困難ですが、評判の良いプロバイダーは、出力されるデータに問題がある場合、問題を公表して共有している場合が多くあります。
- **プロバイダー名を確認する。**攻撃者は、評判の良いプロバイダーになりすまし、侵害したモデルを正規のモデルのように偽って提供しています。

サイバーセキュリティの AI 分野の専門家は、このようなリスクを軽減するための対策に積極的に乗り出しています。

まとめ

AI はサイバーセキュリティに多大な恩恵をもたらします。AI への過剰な期待や誤解を解消し、サイバーセキュリティの向上という成果につながる慎重なアプローチを AI に採用することで、AI テクノロジーを活用してサイバー防御を強化し、IT およびサイバーセキュリティプロフェッショナルのスキルを強化し、労働環境を改善することが可能になります。

調査について

出典：「過剰な期待や誤解の解消：サイバーセキュリティの AI で企業が直面している現実」

ソフォスは、独立したリサーチ専門の企業である Vanson Bourne 社に依頼し、従業員数 50 人から 3,000 人の組織の IT およびサイバーセキュリティのリーダー 400 人を対象に調査を実施しました。調査は 2024 年 11 月に実施され、13 の業界のリーダーから回答を得ることができました。さまざまな業界の意見を取り入れるために、独立したリサーチ企業が調査を実施しています。回答者の組織は 19 の異なるベンダーのエンドポイントセキュリティソリューションを使用していました。

ソフォスについて

ソフォスは、ファイアウォール、エンドポイントプロテクション、EDR/XDR ツールから、MDR、インシデント対応 (IR) サービスなど、受賞歴のある幅広いサイバーセキュリティ製品とサービスのポートフォリオを提供しているサイバーセキュリティにおける世界的なリーダー企業です。

ソフォスは 2017 年以来、AI を活用したサイバーセキュリティ機能の強化に取り組んでおり、AI テクノロジーと人間の専門知識を組み合わせ、あらゆる場所で発生する幅広い脅威を防いでいます。ディープラーニングと生成 AI の機能は、ソフォスの製品やサービスに組み込まれており、業界最大の AI ネイティブセキュリティプラットフォームを通じて提供され、お客様の最も重要な課題を解決します。ソフォスの適応型 AI ネイティブプラットフォームは、60 万社以上のさまざまな顧客環境における攻撃データに基づいて訓練されており、高度な脅威から組織を保護する比類のない防御機能を提供し、防御力を強化します。

ソフォスのソリューションの詳細については、https://www.sophos.com/ja-jp をご覧ください。

2025-01-15 (WP-MP)

SOPHOS