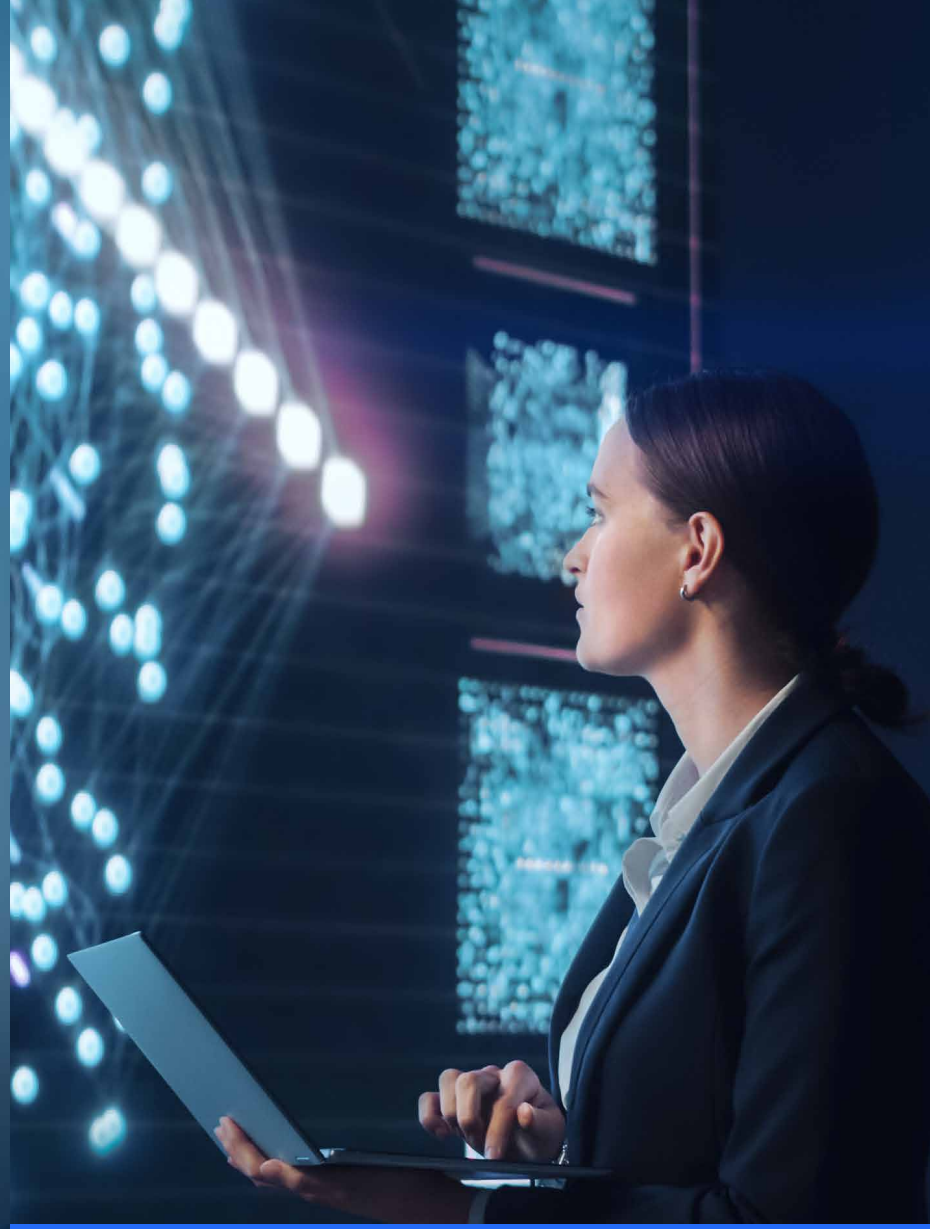
2024年版 Zscaler ThreatLabz AI セキュリティ レポート



AI革命はすでに始まっています。企業におけるAI導入の主な傾向、リスク、ベスト プラクティスのほか、AIを悪用した脅威の最新動向とその対策を解説します。

目次

本書の要旨

イノベーションを牽引してきたAIは、今や業務の必須ツールとして定着しています。ChatGPTを代表とする生成AIツールはさまざまな形でビジネスを進化させるため、AIは企業活動に欠かせない存在になりつつあります。しかし、AIによる脅威を防ぎながら、こうしたAIツールを安全に導入するにはどうすればよいのでしょうか？その答えはいまだ得られていません。

多くの企業のエンジニアリング、ITマーケティング、財務、カスタマー サクセスなどの部門がAI/MLツールを急速に導入し始めていますが、こうしたツールにはメリットとリスクの双方が含まれるという事実を理解することが重要です。変革をもたらすAIの潜在能力を解き放つには、データを保護し、機密情報の漏洩を防ぎ、「シャドーAI」のスプロール化を軽減し、AIデータの品質を確保する、安全な制御を実現する必要があります。

AIが企業に及ぼすリスクは双方向の性質を持っています。**企業の外部ではAIがサイバー脅威を拡大させており**、実際、サイバー犯罪者や国家支援型の脅威アクターがAIツールを使って、より迅速かつ大規模に巧妙な攻撃を仕掛けるようになっています。一方で、多くの企業が変化の激しい脅威への対処に取り組む中で、AIがサイバー防御に不可欠な存在として期待されているのも事実です。

2024年版 ThreatLabz AIセキュリティ レポートでは、これらの重要なAIの課題やメリットを詳しく解説しています。

2023年4月～2024年1月までにZscaler Zero Trust Exchange™で処理された180億件以上のトランザクションをもとに、ThreatLabzが企業におけるAI/MLツールの使用状況を分析したところ、変化するAI環境に企業がどのように適応し、AIツールを保護しているのかについて、業界別および地域別の主な傾向が明らかになりました。

このレポートを通じて、ビジネス上のリスク、AIを悪用した脅威のシナリオと攻撃者の戦術、規制上の考慮事項、2024年以降のAI環境の予測など、AIの最新動向を把握できます。

また、重要なデータを保護しながら生成AIを安全に導入する方法やAIを活用したツールで多層型のゼロトラスト セキュリティを確保し、AIによる新たな脅威環境に対処する方法などのベスト プラクティスついても知見を深めることができます。

主な調査結果

AI/MLツールの利用率が594.82%増加しました。2023年4月は5億2,100万件だったAI/MLトランザクションが、2024年1月には31億件にまで急増しました。

企業はAI/MLトランザクション全体の18.5%をブロックし、ブロックされたトランザクションは9か月間で577%増加しました。この数字には、AIのデータ セキュリティへの懸念の高まりとAIポリシーの整備に消極的な企業の姿勢が反映されています。

最もAIトラフィックを生成したのは製造業で、Zscalerのクラウドで確認されたAI/MLトランザクション全体の20.9%を占め、金融/保険(19.9%)、サービス(16.8%)がそれに続く結果となりました。

ChatGPTの使用が急速に拡大し、634.1%の増加を見せました。一方、**ChatGPTは企業が最もブロックしたAIアプリ**であることもZscalerのクラウドで確認されています。

トランザクションの件数に基づく最も使用されたAIアプリ は**ChatGPT、Drift、OpenAI*、Writer、Liveperson**で、**最もブロックされたトップ3のアプリ**は**ChatGPT、OpenAI、Fraud.net**となっています。

AI/MLトランザクションを最も多く生成している上位5か国は、米国、インド、英国、オーストラリア、日本です。

企業は大量のデータをAIツールに送信しており、合計569TBが2023年9月～2024年1月にかけて、AI/MLアプリに送信されました。

脅威アクターは新たな手法でAIを悪用しており、AIが使われたフィッシング キャンペーン、ディープフェイク、ソーシャル エンジニアリング攻撃、ポリモーフィック型ランサムウェア、企業の攻撃対象領域の検出、エクスプロイトの自動生成などが発生しています。

備考：Zscaler Zero Trust Exchangeは、他のOpenAIトランザクションとは別にChatGPT単独のトランザクションを追跡しています。

生成AIとMLの主な利用傾向

企業のAI革命は始まったばかりです。企業が生成するAIトランザクションは600%近くも増加しており、その勢いはとどまることを知りません。一方、ブロックされたAIアプリのトランザクションも577%増加しました。

増加の一途をたどるAIトランザクション

2023年4月～2024年1月にかけて、企業が生成したAI/MLトランザクションは600%近く増加し、1月にはZero Trust Exchange全体で1か月あたり30億件のトランザクションが確認されました。この数字からも、AIの導入によるセキュリティ インシデントやデータ リスクが増えているにもかかわらず、AIの変革力を無視できない企業の現状が浮かび上がってきます。なお、AIトランザクションは12月の連休中に一時的 な落ち着きを見せたものの、2024年初頭にはさらに速いペースで増加を続けました。

AIアプリの数は急増していますが、AIトランザクションの大部分を占めたのが市場をリードする比較的少数のAIツールです。全体として、ChatGPTがAI/MLトランザクションの半分以上を占め、OpenAIアプリ自体はトランザクション全体の7.82%で3位となっています。一方、AI活用型チャットボットであるDriftは、企業のAIトラフィックの5分の1近くを生成しました(LivepersonとBoldChat Enterpriseのチャットボットもそれぞれ5位と6位)。Writerは、マーケティング資料などの企業向けコンテンツの作成に使用される生成AIツールですが、依然として利用率が高い傾向にあります。ビデオ通話で使用されるAI文字起こしツールのOtterも上位に位置しています。

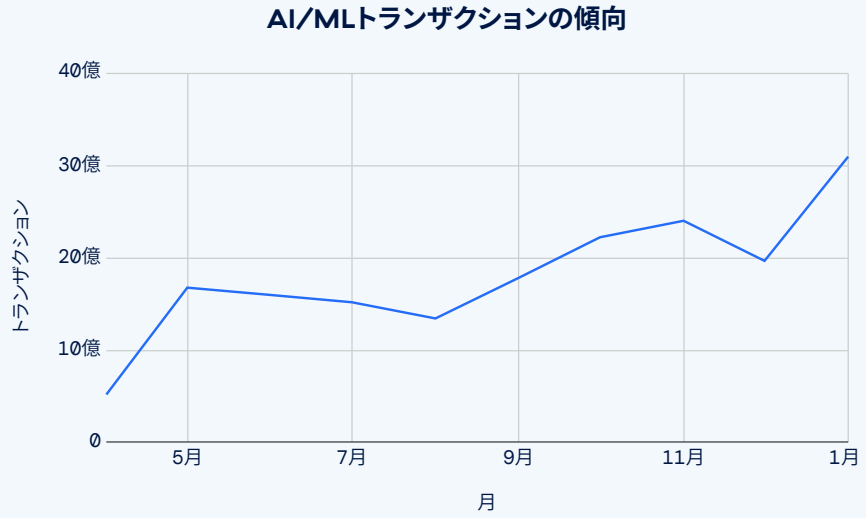


図 1　2023 年 4 月～ 2024 年 1 月に発生した AI トランザクションの件数

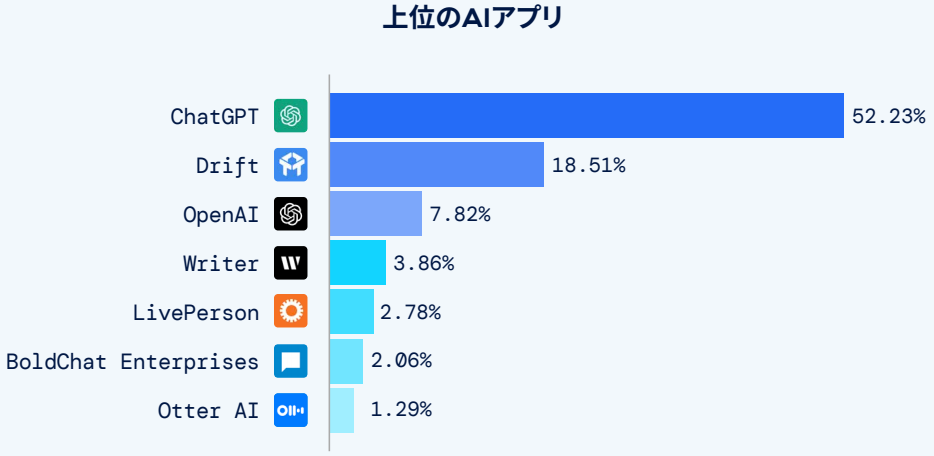


図 2　トランザクションの件数に基づく上位の AI アプリ

AI/MLトラフィックによって転送されたデータ[2023年9月～2024年1月]

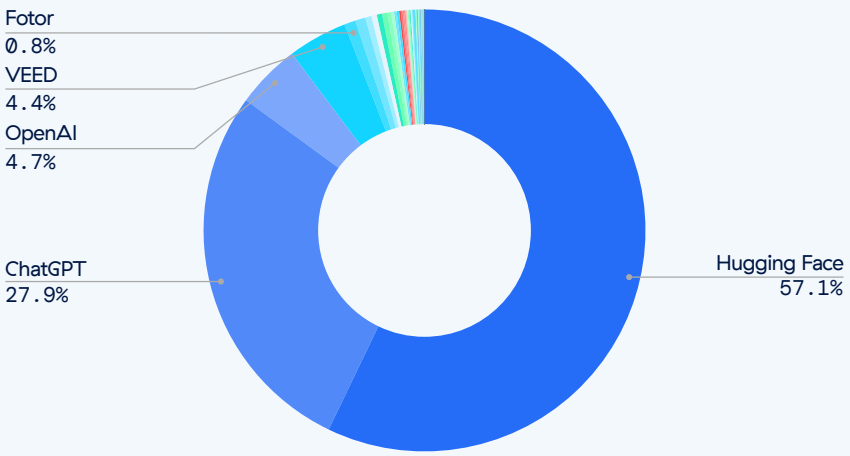


図 3 総転送データ量に占める割合で見た上位の AI/ML アプリ

ブロックされたAIトランザクションの傾向[2023年4月～2024年1月]

図 4 ブロックされた AI/ML トランザクション件数の推移

しかし、企業がAIツールとの間で送受信するデータの量と上記の傾向には、微妙な違いがあります。「AI版GitHub」と呼ばれるオープンソースのAI開発者向けプラットフォームであるHugging Faceは、AIツールによって転送される企業データの60%近くを占めています。Hugging FaceはAIモデルをホストしてトレーニングし、企業のユーザーから大量のデータを収集するため、この結果は当然といえるかもしれません。

ChatGPTとOpenAIもここに含まれますが、注目すべきは、Veed (動画に字幕や画像などを追加するAI動画エディター)とFotor (AI画像の生成などに使用されるツール)の2つが上位に入ったことです。動画や画像は他の種類のリクエストに比べてファイルのサイズが大きくなるため、この2つのツールが上位に入ったとみられます。

ブロックされたAIトランザクションが急増

AIの導入が急速に進む一方で、企業はデータの安全性とセキュリティを懸念して、これまで以上に多くのAI/MLトランザクションをブロックしています。企業がブロックしたAIトランザクションは18.5%に上り、ブロックされたトランザクションは4月～1月にかけて577%の増加を見せ、その合計は26億件以上にまで達しています。

AIツールの中には、利用率が高い一方で、ブロックされる件数も多いものもあり、例えば、ChatGPTは最も使用され、かつ最もブロックされたAIアプリとして上位に挙がりました。この結果からも、これらのツールは人気があるにもかかわらず、あるいはその人気がゆえに、情報漏洩やプライバシーの問題から企業がその使用を保護する取り組みを積極的に進めていることがわかります。注目すべきもう1つの傾向は、AI対応のCopilot機能を備えたbing.comが4月～1月にかけてブロックされていることです。bing.comはブロックされたAI/MLドメイン トランザクション全体の25.02%を占めています。

最もブロックされたAIツール	最もブロックされたAIドメイン
01 ChatGPT	01 Bing.com
02 OpenAI	02 Divo.ai
03 Fraud.net	03 Drift.com
04 Forethought	04 Quillbot.com
05 Hugging Face	05 Compose.ai
06 ChatBot	06 Openai.com
07 Aivo	07 Qortex.ai
08 Neeva	08 Sider.ai
09 infeedo.ai	09 Tabnine.com
10 Jasper	10 securiti.ai

図 5 トランザクションの件数に基づく最もブロックされた AI アプリと AI ドメイン

AIを導入している業界の内訳

AIツールの導入状況やブロックしたAIトランザクションの割合は、業界によって大きく異なります。圧倒的1位となったのが製造業で、Zero Trust Exchangeで確認されたAI/MLトランザクションの20%以上を占めています。次に金融/保険、テクノロジー、サービスの業界が僅差で続いています。これら4つの業界は他の業界に先んじて、最も積極的にAIを導入しています。

業界別のAIトランザクションの割合

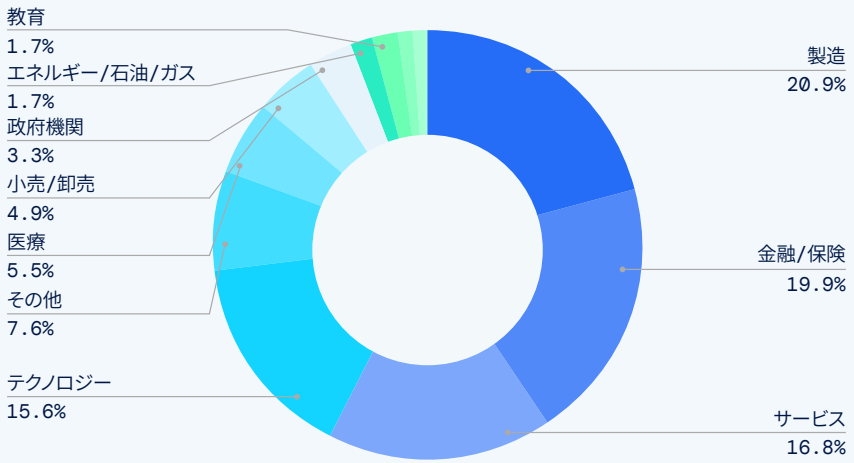


図 6 AI トランザクションを最も生成した業界

業界別のAIトランザクションの傾向

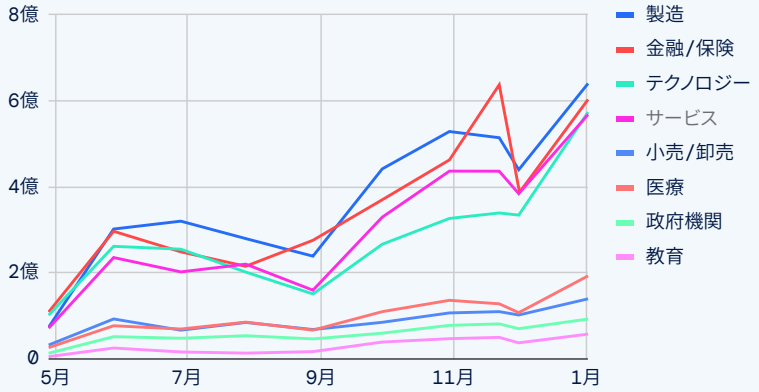


図 7 2023 年 4 月～ 2024 年 1 月にかけて AI/ML トランザクションを生成した業界

AI/MLトランザクションの保護

AIトランザクションの急激な増加と相まって、各業界はますます多くのAIトランザクションをブロックしています。全体的な導入傾向から外れている業界もありますが、この背景には、AIツールの保護における優先順位やセキュリティ成熟度の違いがあります。例えば、金融/保険業界はAIトランザクションのブロック率が最も高く、世界平均の18.5%に対して37.2%となっています。これは、この業界の厳格な規制とコンプライアンス環境に加え、金融機関が処理する財務や個人ユーザーに関するデータの機密性が非常に高いことが主な原因と考えられます。

一方、製造業は多くのAIトランザクションを生成しているにもかかわらず、ブロックしたAIトランザクションは全体の15.7%にとどまっています。AIをいち早く導入してきた業界の1つであるテクノロジー業界は、AI導入を推進しながら、19.4%と平均以上のAIトランザクションをブロックしています。医療業界は膨大な量の医療データや個人を特定できる情報(PII)を処理していますが、ブロックしたAIトランザクションは平均以下の17.2%という驚くべき結果となりました。これは、この業界のセキュリティ部門がAIイノベーションへの取り組みを加速させながらも、AIツール使用時のデータ保護対策を適切に講じていないためと考えられます。医療業界におけるAIトランザクションは、比較的低い水準にとどまりました。

図 8
AI トランザクションを
ブロックする割合が高い業界

ブロックされたAIトランザクションの業界別の割合

業種	ブロックされたAIトランザクション(%)
金融/保険	37.16
製造	15.65
サービス	13.17
テクノロジー	19.36
医療	17.23
小売/卸売	10.52
その他	8.93
エネルギー/石油/ガス	14.24
政府機関	6.75
輸送	7.90
教育	2.98
情報通信	4.29
建設	4.12
基礎材/化学薬品/鉱業	2.92
エンターテイメント	1.33
食料品/飲料品/たばこ	3.66
ホテル/レストラン/レジャー	3.16
宗教団体	6.06
農林	0.18
すべての業界の平均	18.53

医療業界とAI

医療業界はAI/MLユーザーとして6位に位置し、AI/MLトランザクション全体の17.23%をブロックしています。

医療業界で使用されている上位のAIアプリ:

01 ChatGPT
02 Drift
03 OpenAI
04 Writer
05 Intercom
06 Zineone
07 Securiti
08 Pypestream
09 Hybrid
10 VEED

AI医療の進歩を示す兆候

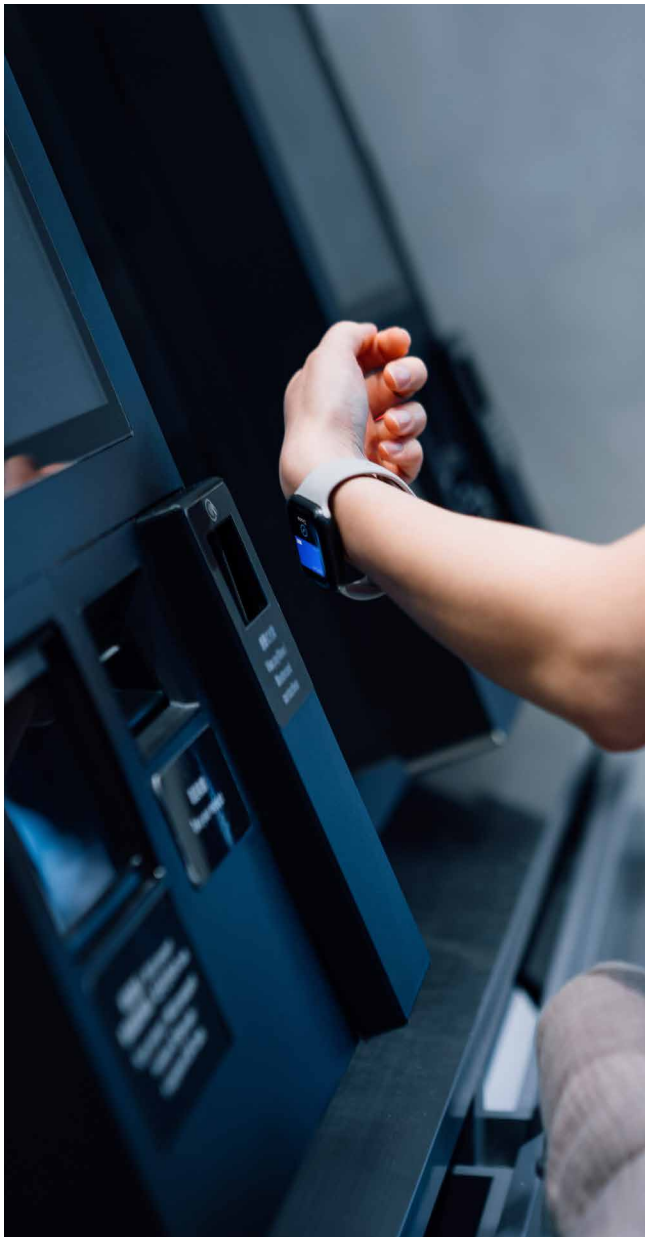
医療業界は、AIなどのテクノロジーの採用には慎重な姿勢を示す場合が多い傾向にありますが、Zscalerのクラウドで確認された医療業界のAI/MLトラフィックが5%を占めていることからも、AIは近い将来、医療業務や患者のケア、医学の研究とイノベーションに大きな影響を与えると予測されます。[1]

AIは時間の節約だけでなく、救命にも貢献すると期待されており、AIを活用した技術はすでに診断や患者ケアの質を高めています。例えば、AIが医用画像を驚異的な精度で分析することで、放射線科医はこれまで以上に迅速に異常を発見して、治療方法を決定できるようになります。[2]

AIの潜在的なメリットは計り知れません。AIアルゴリズムは患者データを使用して最適な治療計画を立てたり、生物学的データを効率的に分析することで創薬を加速させたりできます。また、生成AIで管理業務を自動化すれば、人手が足りない医療現場の負担を減らすことも可能です。こうした発展からも、AIには医療やヘルスケア サービスの提供を変革する力があることは明らかです。

医療業界におけるAIの主なリスク:
医療機関は、個人を特定できる情報(PII)のデータ プライバシーとセキュリティに関する懸念など、AIにまつわる潜在的なリスクや課題を認識すると同時に、患者ケアの管理を支援する際には、AIアルゴリズムとその出力内容が高い信頼性と公平性を確保できるようにする必要があります。

1. Statista、Future Use Cases for AI in Healthcare、2023年9月
2. The Hill、AI already plays a vital role in medical imaging and is effectively regulated、2024年2月23日



金融業界とAI

金融業界はAI/MLトランザクションを生成した業界の第2位であり、AI/MLトラフィック全体の37.16%をブロックしています。

金融業界で使用されている上位のAIアプリ:

01 ChatGPT
02 Drift
03 OpenAI
04 BoldChat Enterprise
05 LivePerson
06 Writer
07 Hugging Face
08 Otter Ai
09 Securiti
10 Intercom

AIを利用する金融機関

AI時代のアーリー アダプターとして一歩先を行く金融サービス企業は、Zscalerのクラウドで確認されたAI/MLトラフィックのほぼ4分の1を占めています。McKinseyは、銀行業務に生成AIを利用することで生産性が向上し、これによって得られる潜在的な年間収益が2,000億～3,400億ドルになると予測しています。[3] AIは実際、銀行や金融サービスにさまざまなメリットを提供しています。

AI活用型のチャットボットやバーチャル アシスタントは、金融業界にとっては目新しいものではありませんが(Bank of Americaの「Erica」は2018年に正式リリース)、生成AIの強化により、これらのカスタマー サービス ツールはさらに高いレベルのパーソナライズを実現できるようになっています。また、予測モデリングやデータ分析などのAI機能が不正行為の検知やリスク評価の精度を高め、金融業務の生産性を大幅に向上させると期待されています。

金融/保険業界におけるAIの主なリスク:
金融サービスや商品にAIを導入する場合も、データ プライバシー、偏り、正確性の面でセキュリティ上および規制上の懸念が生じます。37%ものAI/MLトラフィックをブロックしているというThreatLabzの調査結果からも、その現状がうかがえます。こうした懸念に対処するには、銀行、金融サービス、保険における信頼性と完全性を維持するための高度な監視と計画が必要になります。

3. McKinsey、Capturing the full value of generative AI in banking、2023年12月5日

政府機関とAI

政府機関はAI/MLの使用で10位に入るものの、AI/MLトランザクションをブロックしている割合はわずか6.75%にとどまっています。

政府機関で使用されている上位のAIアプリ*：

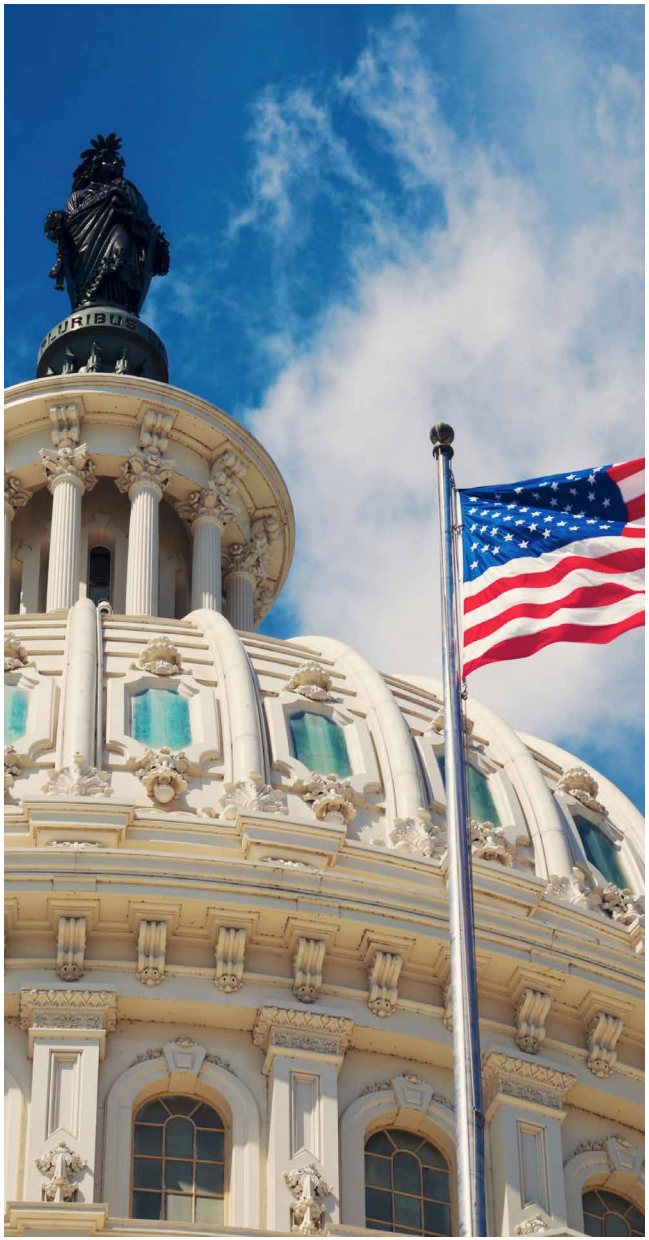
01 ChatGPT
02 Drift
03 OpenAI
04 Zineone

*少なくとも100万件以上のトランザクションが処理されたAIアプリ

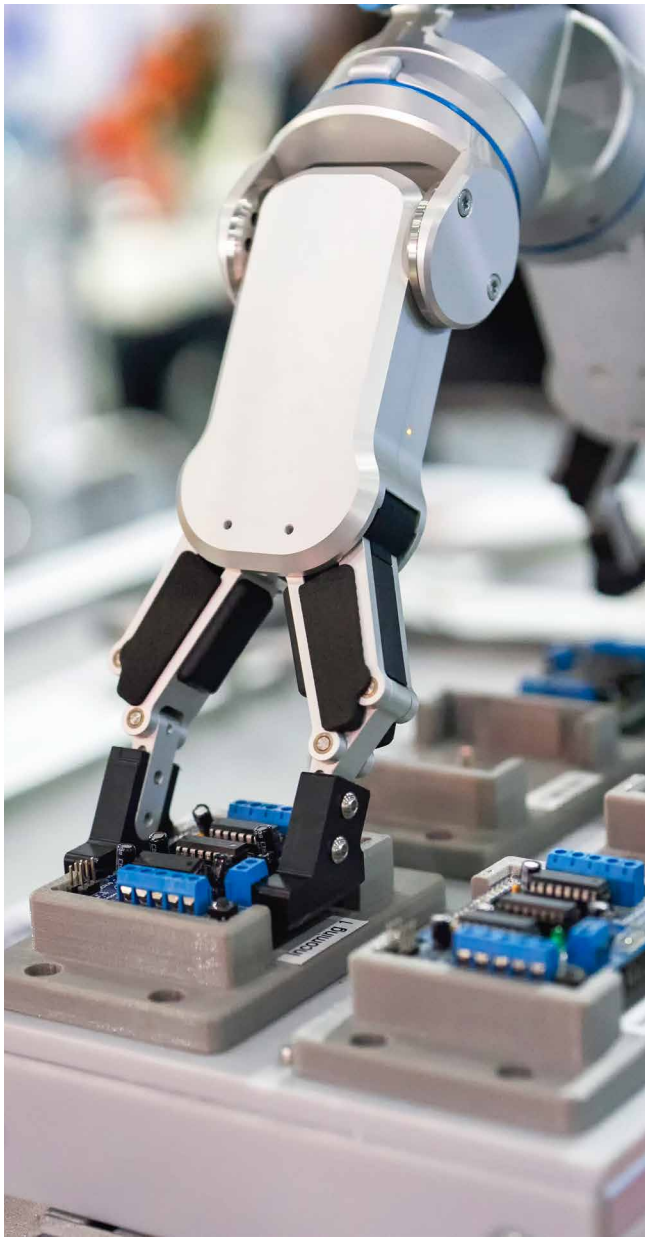
世界各国の政府機関におけるAIの導入とガバナンスの確立

政府内では、AI技術の導入と、AI技術を安全に管理するためのガバナンス確立というAIに関する2つの重要な議論が交わされています。政府機関や公共機関によるAI導入のメリットは大きく、特に、チャットボットやバーチャル アシスタントを取り入れることで、市民は交通や教育などの重要な情報やサービスをより迅速に利用できるようになります。AIを活用したデータ分析は、データによる意思決定プロセスを通じて社会的課題に対処するうえで重要な役割を果たし、より効率的なポリシーの策定とリソースの割り当てを可能にします。

AIは現在、目覚ましい進化を遂げています。米国司法省は初代最高AI責任者を任命し、AIシステムの使用に取り組むことを正式に発表しました。ThreatLabzのデータによると、ChatGPTやDriftなどのAI/MLプラットフォームを使用する政府機関の顧客が増加傾向にあることがわかっています。



政府機関におけるAIの主なリスク：
こうした傾向がみられる一方で、AI関連のリスクとデータ プライバシーに関する重要な懸念は、政府機関全体での規制フレームワークとガバナンスが継続的に必要であることを浮き彫りにしています。世界各国の政策立案者は過去1年間で**AI規制に向けて大きな一歩**を踏み出し、AI/ML技術の責任ある開発と展開を推進するために共同で取り組む姿勢を示しています。



製造業界とAI

製造業はAI/MLツールの利用率が最も高く、AI/MLアプリ全体の15.65%をブロックしています。

上位のアプリ:

01 ChatGPT
02 Drift
03 OpenAI
04 Writer
05 Securiti
06 Google Search
07 Zineone
08 Pypestream
09 Hugging Face
10 Fotor

AIと共に進化する製造業界

AI/MLトラフィックの流入が最も多かったのは、当然ながら製造業という結果になりました(18.2%)。製造業におけるAIの導入はインダストリー4.0の礎となっています。インダストリー4.0とは第4次産業革命としても知られ、デジタル技術と産業プロセスの融合を特徴とする時代を指します。

機械やセンサーからの膨大な量のデータを分析して機器の故障の兆候を事前に検知する取り組みから、サプライ チェーン管理、在庫管理、物流業務の最適化に至るまで、AIは製造業にとって不可欠な存在であることが証明されています。さらに、AIを活用したロボットや自動化システムは製造効率を大幅に向上させ、コストとエラーを削減しながら、人間よりもはるかに高速かつ正確に作業を実行します。

製造業界におけるAIの主なリスク:
製造業がブロックしたAI/MLアプリのトラフィックは16%となっており、一部の製造業者は生成AI/MLの利用を慎重に進めていることがわかります。これは、製造企業がデータのセキュリティを懸念し、リスクの高いアプリをブロックしながら、AIアプリを厳選して審査し、承認する必要があると考えているためと予測されます。

教育業界とAI

教育業界はAI/MLツールの利用率において11位に位置し、AI/MLトラフィック全体の2.98%をブロックしています。

上位のアプリ:

01 ChatGPT
02 Character.AI
03 Pixlr
04 Forethought
05 Deepai
06 Drift
07 OpenAI

教育業界ではAIを学習ツールとして活用

教育業界はAIトラフィックを大量には生成しておらず、合計3億900万件以上のトランザクションのうち、ブロックしたAI/MLトランザクションは約900万件となっており、比較的低い割合(2.98%)を占めています。通常、教育機関は学生を対象にChatGPTのようなAIアプリの使用を禁止すると考えられていますが、この業界では主に学習ツールとしてAIアプリが受け入れられています。特に、教育業界では5つのAIアプリが広く使用されており、そのうちの4つ(ChatGPT、Character. AI、Pixlr、OpenAI)は、文章作成や画像生成を目的としたクリエイティブなアウトプットに焦点を当てています。Forethoughtは、チャットボットによる指導の補助に使用できます。

多くの教育者が授業の方針としてChatGPTなどのツールをブロックしているものの、AI/MLツールをより具体的な方法でブロックするDNSフィルタリングなどのテクノロジー ソリューションの実装においては、教育機関は他の業界に後れをとっているのが現状です。

教育業界におけるAIの主なリスク:
教育業界ではAIツールの導入が続き、特に生徒個人のデータ保護をめぐるデータ プライバシーの懸念が高まる可能性があります。 個人データに対する保護対策を強化しながら、AIアプリを選択的にブロックする技術がこれまで以上に採用されるようになります。

ChatGPTの利用状況

ChatGPTの導入は加速しており、2023年4月以降、世界のChatGPT関連のトランザクションが634%以上増加し、AIトランザクション全体の595%増を大きく上回りました。この数字からも明らかなように、OpenAIはAIの最高ブランドと幅広く認識されており、ChatGPTは最も支持される生成AIツールとなっています。ChatGPTの新しいバージョンや、テキストから動画に変換するOpenAIの生成AI製品であるSoraのリリースが期待されることもあり、OpenAI製品の導入は今後も拡大すると考えられます。

各業界におけるChatGPTの使用状況は一般的なAIツールの導入パターンとほぼ同じで、今回も製造業が業界をリードし、それに金融/保険が続きます。全トランザクションでは14.6%で3位だったテクノロジー業界は、ChatGPTトランザクションでは10.7%の4位と、わずかに後れをとっています。これは、テクノロジー業界がファースト イノベーターであること、そしてより積極的に多種多様な生成AIツールを導入していることが要因と考えられます。

業界別のトランザクション

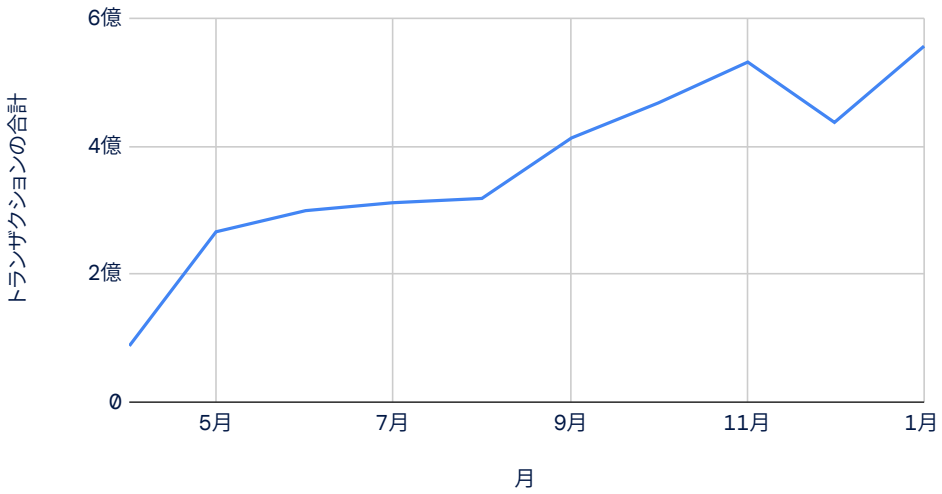


図 9　2023 年 4 月～ 2024 年 1 月の ChatGPT トランザクション

業界別のAIトランザクションの傾向

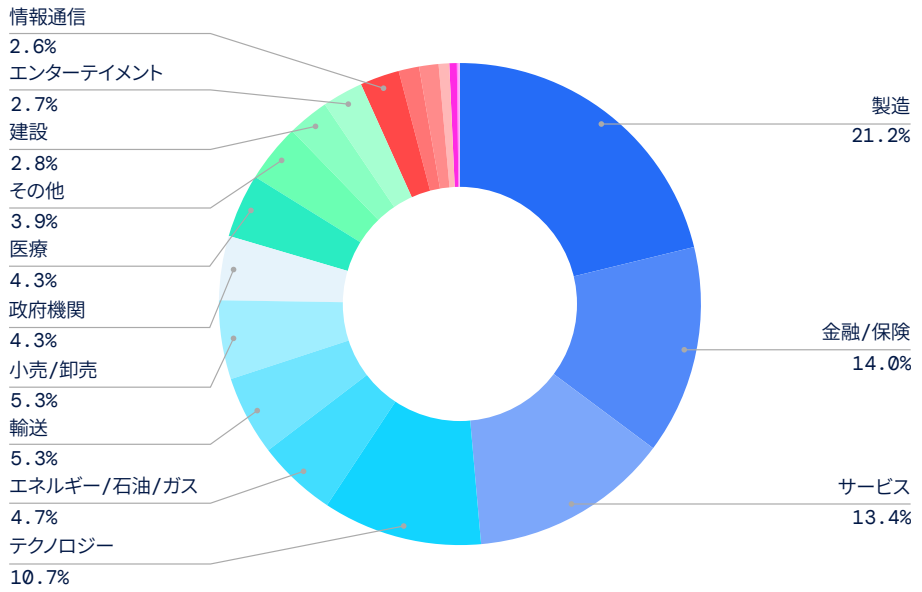


図 10　業界別の ChatGPT トランザクションの割合

国別のAI利用状況

AI導入の傾向は、規制要件や技術インフラ、文化的考慮事項などの要因を受け、国によって大きく異なります。ここでは、Zscalerのクラウドで確認されたAI/MLトランザクションを生成した上位の国を見ていきます。

予想どおり、AIトランザクションの大部分を占めたのは米国という結果となっています。一方、インドは技術革新への取り組みを加速させていることから、AIトラフィックを生成する主な国として浮上しています。インド政府が先日制定した、AIモデルの発売前に規制当局の承認を義務付ける計画(その後撤廃)からも、AI規制がいかに急速に進められているかをうかがい知ることができます。[4]

国別のトランザクション

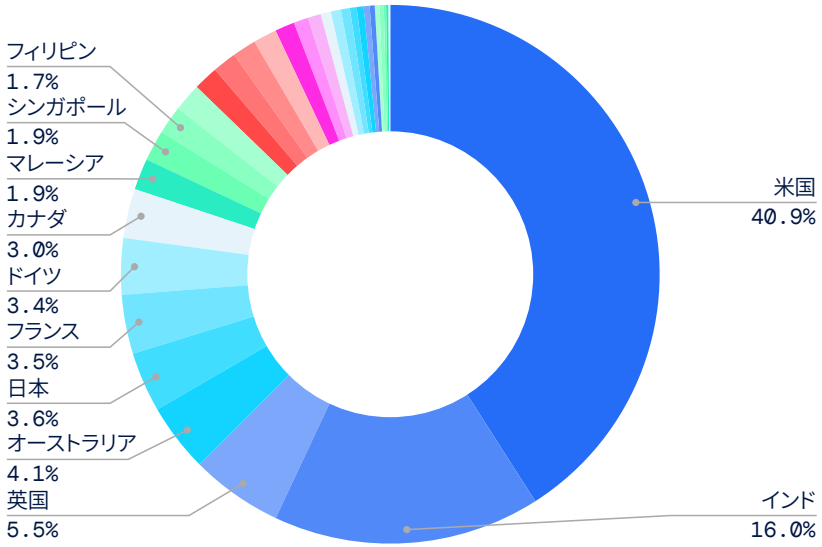
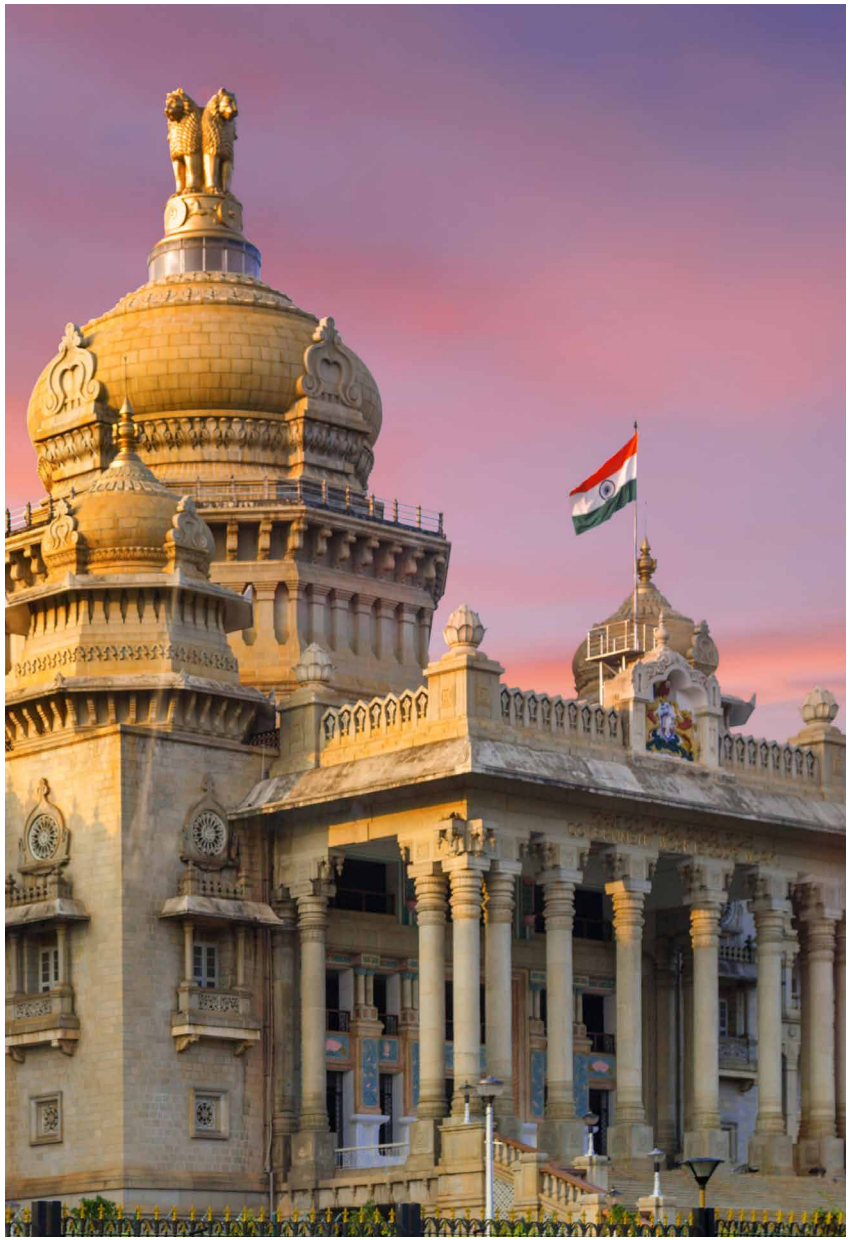


図 11　国別の AI トランザクションの割合



4. TechCrunch、India reverses AI stance, requires government approval for model launches、2024年3月3日

地域別の内訳：EMEA

欧州、中東、およびアフリカ(EMEA)の地域を詳しく見ると、AI/MLトランザクションの割合は国によって明らかな差異があることがわかります。英国は世界のAIトランザクションのわずか5.5%を占めるに過ぎませんが、EMEAのAIトラフィックは20%以上も占め、フランスとドイツがそれに続きます。アラブ首長国連邦は急速な技術革新により、この地域で上位のAI導入国として浮上しました。

国	トランザクション	地域の割合
英国	763413289	20.47%
フランス	504185470	13.53%
ドイツ	471700683	12.66%
アラブ首長国連邦	238557680	6.40%
オランダ	222783817	5.98%
スペイン	198623739	5.30%
スイス	129059097	3.46%
イタリア	97544412	2.62%

図 12　EMEA の各国におけるトランザクションの件数

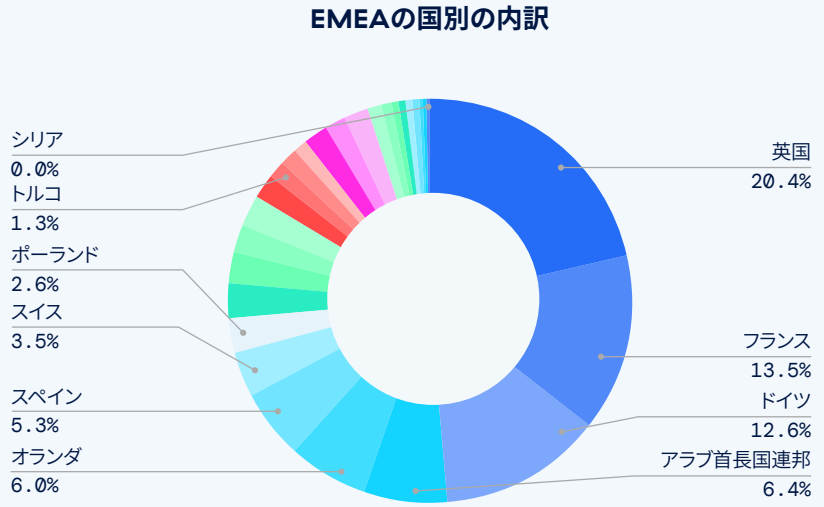


図 13　EMEA の AI トランザクション総量に占める国別の割合

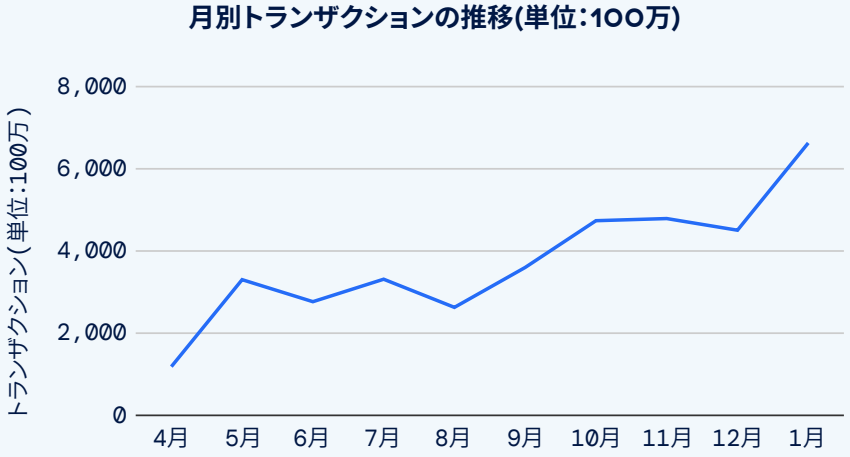


図 14　EMEA の AI トランザクション量の経時的増加

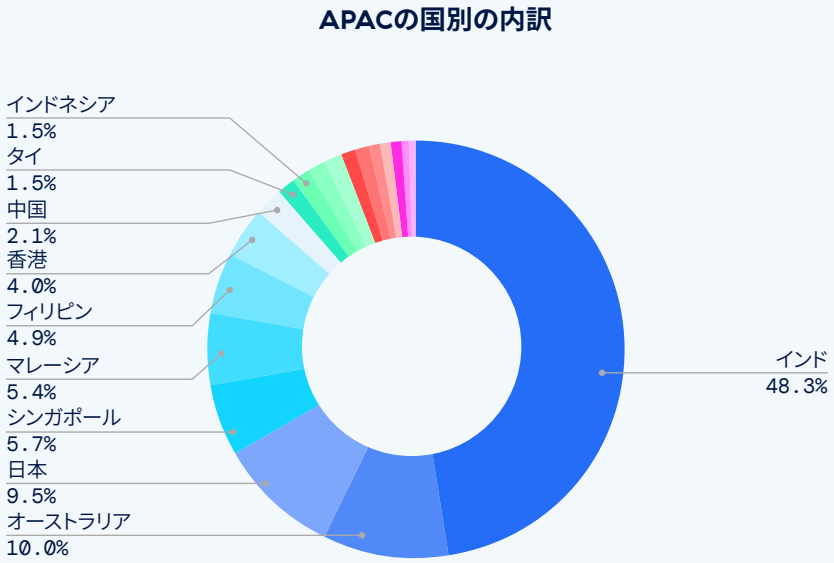


図 16 APAC の AI トランザクション総量に占める国別の割合

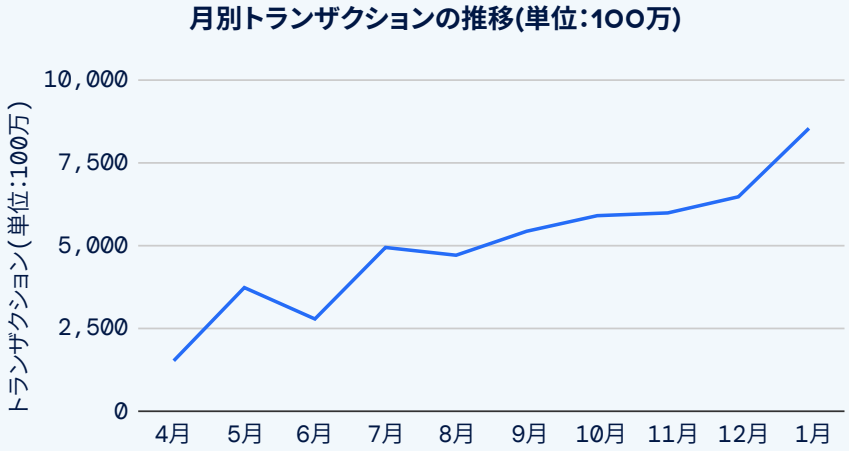


図 17 APAC の AI トランザクション量の経時的増加

地域別の内訳:APAC

ThreatLabzがアジア太平洋地域(APAC)の結果を深く掘り下げたところ、AI導入における注目すべき傾向が示されました。APACに含まれる国の数は他の地域よりもはるかに少ないにもかかわらず、EMEAより約13億件(135%)多いAIトランザクションを生成していることが明らかになりました。この増加の最大の要因となっているのがインドで、APACにおけるAI/MLトランザクションの約半分を占める結果となっています。

国	トランザクション	地域の割合
インド	2414319490	48.30%
オーストラリア	501562395	10.01%
日本	476425423	9.52%
シンガポール	284891384	5.70%
マレーシア	268043263	5.36%
フィリピン	243754578	4.87%
香港	202119814	4.04%
中国	104545655	2.09%

図 15 APAC の各国におけるトランザクションの件数

企業におけるAIのリスクと実際の脅威のシナリオ

企業が直面するAIのリスクと脅威には、大きく分けて2種類あります。1つはAIツールの使用に伴うデータ保護とセキュリティのリスク、もう1つは生成AIツールと自動化によって引き起こされる新たなサイバー脅威環境のリスクです。

企業におけるAIのリスク

1 知的財産および非公開情報の保護

生成AIツールは、意図しない機密データの漏洩を引き起こす場合があります。実際、機密データの漏洩はAIアプリのOpen Worldwide Application Security Project (OWASP) Top Tenで6位に入っています[5]。過去1年間、クラウドの設定ミスなどが原因で、一部の大手AIツール プロバイダーにおいて偶発的な情報漏洩やAIトレーニング データの侵害が多数発生しており、中にはテラバイト規模の顧客データが漏洩したケースもあります。

一例を挙げると、研究者がプロンプト インジェクション(AIを操作してトレーニング データを持ち出すように設計されたAIクエリーを使用する手法)と呼ばれる脆弱性を悪用して、GitHubのCopilot AIからGitHubの膨大な機密情報を抜き取る事件が発生しました。なお、この脆弱性はOWASP Top 10のリスクの1位となっています。[6]

これに関連したリスクとなるのが、**モデル反転の脅威**です。モデル反転攻撃では、攻撃者がLLMの出力とそのモデル構造に関する知識を組み合わせて、トレーニング データについて推論を行い、最終的にそのデータを抽出します。これには、AI開発企業自体が侵害されるリスクもあり、こうした企業の従業員の認証情報がデータの流出に直接つながった例もあります。

一方、攻撃者がRedline StealerやLummaC2などの情報窃取型ツールを使用して**二次的なマルウェア攻撃**を開始し、従業員のログイン認証情報を盗み、AIアカウントにアクセスする可能性もあります。この種の攻撃により、約225,000件のChatGPTユーザーの認証情報がダークWeb上で販売されていることが最近明らかになりました[7]。AIツール プロバイダーの最優先事項はプライバシーとデータ セキュリティであることに変わりはありませんが、こうしたリスクは依然として存在し、小規模なAI開発企業やAI機能を導入したSaaSプロバイダーなどにも同様のリスクが広がっています。

最後に、**企業のAIユーザー自身に起因するリスクがあります。**ユーザーが知らないうちに、LLMのトレーニングに使用されるデータ セットに貴重な知的財産や非公開情報を公開してしまうケースは少なくありません。例えば、ソース コードを最適化したい開発者や内部データに基づいて販売動向を探る営業担当者が、保護された情報を意図せず社外に公開してしまう場合があります。企業はこうしたリスクを認識し、情報漏洩防止(DLP)などの強力なデータ保護対策を実施してこのような漏洩を防ぐ必要があります。

アクセス制御とセグメンテーションのリスク

ロールベースのアクセス制御(RBAC)などのアクセス制御では、AIアプリの設定ミスや悪用などのリスクが生じる場合があります。AIチャットボットがCEOとCEO以外の企業ユーザーに同じ応答を生成するという状況が発生しかねず、チャットボットがそのユーザーの入力履歴に基づいてトレーニングされる場合は、特にリスクをもたらします。つまり、経営層がAIチャットボットを使用して送信した質問に関する情報を推測できるということです。企業はAIアプリのアクセス制御を適切に構成し、ユーザーの権限と役割に基づいてデータ セキュリティとアクセス セグメンテーションの両方を有効にする必要があります。

5. OWASP、OWASP Top 10 For LLM Applications, Version 1.1、2023年10月16日
6. The Hacker News、Three Tips to Protect Your Secrets from AI Accidents、2024年2月26日
7. The Hacker News、Over 225,000 Compromised ChatGPT Credentials Up for Sale on Dark Web Markets、2024年3月5日

2 AIアプリのデータ プライバシーとセキュリティ リスク

AIアプリの数は驚異的なスピードで増加し続けていますが、AIアプリのデータ プライバシーとセキュリティはすべて同じというわけではありません。利用規約はAI/MLアプリごとに大きく異なる場合があるため、入力した質問が言語モデルのトレーニングに使用されるのか、広告用にマイニングされるのか、サード パーティーに販売されるのかなどを事前に確認する必要があります。また、これらのアプリのセキュリティ慣行やその背後にある企業の全体的なセキュリティ態勢もさまざまです。**データ プライバシーとセキュリティを確保するには、データ保護やその企業のセキュリティ対策などの要素を考慮し**、自社が使用するAI/MLアプリを評価してリスク スコアを割り当てることが重要です。

3 データ品質に関する懸念：質の低いデータでは有益な結果を得られない

AIの出力内容の価値と精度はAIアプリのトレーニングに使用されるデータの質と規模に大きく依存するため、常に精査する必要があります。OpenAIなどの大規模なAIベンダーは、パブリック インターネットのように広く使用されるリソースを利用してツールをトレーニングしていますが、サイバーセキュリティを含む専門的な業界や垂直統合型の業界でAI製品を扱うベンダーは、AIの出力内容の精度を向上させるために非常に特殊で大規模なデータ セット、またはプライベートなデータ セットでAIモデルをトレーニングしなければなりません。「Garbage in, Garbage out」という言葉があるように、質の低いデータを入力しても、無意味な分析結果しか得られないため、AIソリューションを評価する際は、データ品質の問題を慎重に検討する必要があります。

また、**データ ポイズニングのリスク**、つまり不正なデータがトレーニング データに追加され、AIの出力内容の精度や信頼性が低下するリスクもあります[8]。AIトレーニング データと生成AIの出力が品質基準を満たしているかどうかを継続的に評価しながら、このような不測の事態に備えるための強力なセキュリティ基盤を確立することが重要です。

AI導入時に考慮すべきポイント：AIをいつブロックまたは許可するのか？シャドーAIのリスクをどのように軽減するのか？

AIアプリを導入して生産性を向上させるか、それともAIアプリをブロックして機密データを保護するかという岐路に立たされている企業は少なくありません。こうした状況下で、情報に基づいた安全なアプローチを取るには、次の5つの重要な質問に対する答えを明確にしておく必要があります。

01 **従業員のAIアプリの使用状況を詳細に把握しているか？**企業は自社内で使用されているAI/MLツールだけでなく、それらのツールへの企業トラフィックも完全に可視化する必要があります。今後は「シャドーIT」と同様に「シャドーAI」ツールが急増すると予測されます。

02 **AIアプリへのアクセスをきめ細かく制御できているか？**承認された特定のAIツールに対して、部門やユーザー レベルでのきめ細かなアクセス制御とマイクロセグメンテーションを実施する必要があります。同時に、URLフィルタリングを使用して安全性の低い不要なAIアプリへのアクセスをブロックすることも重要です。

03 **それぞれのAIアプリが講じているデータ セキュリティ対策はどのようなものか？**さまざまなAIツールが日常的に使用されていますが、企業は各ツールがどのようなデータ セキュリティ対策を取っているのかを把握する必要があります。例えば、企業環境にプライベートで安全なデータ サーバーを確保するというベスト プラクティスを実践するAIツールがある一方で、ユーザー データをすべて保持したり、サード パーティーに販売したり、入力データを使用してLLMをさらにトレーニングしたりするものもあります。

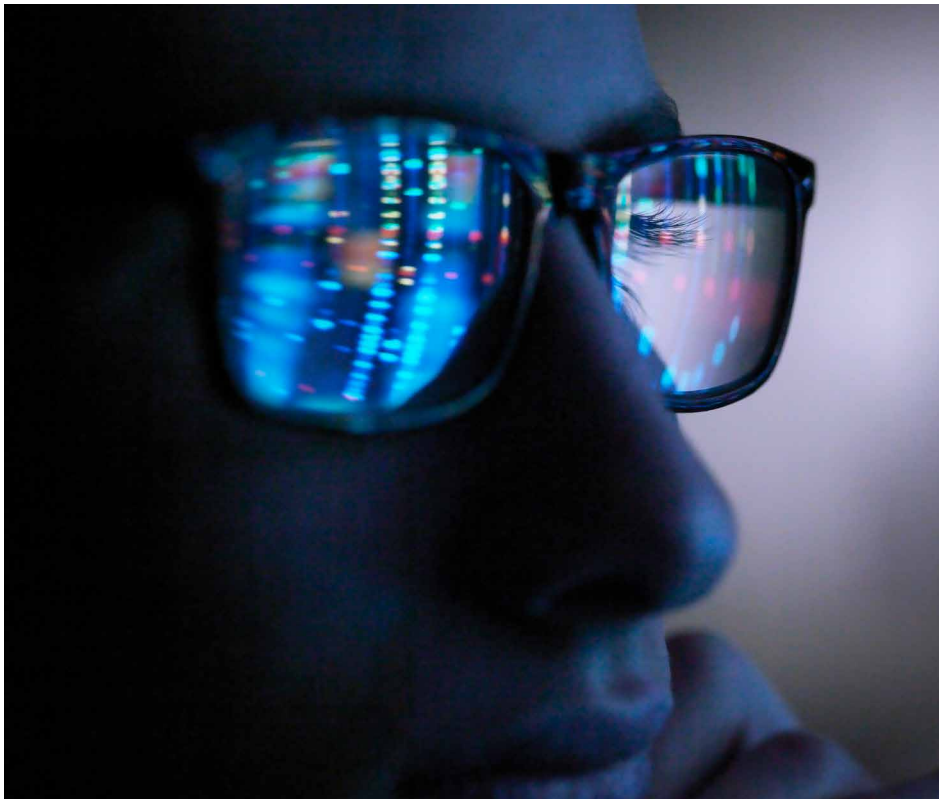
04 **重要なデータの漏洩を防ぐためのDLPが有効になっているか？**企業独自のコードや財務、法律、顧客、個人のデータなどの機密情報の流出を防ぐには、DLPが不可欠です。特にAIアプリのデータ セキュリティ制御が十分でない場合、これらの情報が社外に流出したり、AIチャットボットに入力されたりすることがあります。

05 **AIへのプロンプトの適切なログ記録があるか？**ChatGPTなどのツールに入力されたプロンプトやデータだけでなく、各部門がAIツールをどのように使用しているかを可視化する詳細なログが必要です。

8. SC Magazine、Concerns over AI data quality gives new meaning to the phrase: 'garbage in, garbage out、2024年2月2日

AIによる脅威のシナリオ

企業は常にさまざまなサイバー脅威にさらされていますが、今やAIを悪用した攻撃も脅威の1つとなっています。AIを悪用した脅威が発生する可能性は基本的に無限です。攻撃者はAIを用いて、巧妙なフィッシング キャンペーンやソーシャル エンジニアリング キャンペーン、高度に回避するマルウェアやランサムウェアを生成するほか、企業の攻撃対象領域の中から脆弱な侵入口を特定するなど、攻撃のスピード、規模、手法をますます強化しています。こうした状況が企業やセキュリティ リーダーを二重の苦境に追い込んでいます。つまり、急速に進化するAI技術を適切に取り入れ、その可能性を最大限に引き出しながらも、AIによる攻撃への防御を強化してリスクを軽減するという極めて困難な課題に直面しているのです。



AIによるなりすまし：ディープフェイク、偽情報、誤情報など

現在、AIが生成した動画やライブ アバター、偽の音声は、本物と見分けがつかないほどの精巧な作りになっています。2023年、ZscalerはAIビッシングとスミッシング シナリオの阻止に成功しました。その手口は脅威アクターがWhatsAppメッセージ内でZscalerのCEOであるJay Chaudhryの声になりすまし、従業員をだましてギフト カードを購入させ、情報をさらに引き出そうとするものでした。その後ThreatLabzは、これが他の複数のテクノロジー企業を標的とした広範なキャンペーンの一環であることを特定しています。

このような攻撃は、信頼できる別の手段を介して同僚に直接確認するといった単純な方法で阻止できる場合がほとんどですが、近年、そのメッセージの内容が非常に巧妙になりつつあります。先日発生した事件では、攻撃者がある企業のCFOのディープフェイクを使用して、香港に拠点を置く多国籍企業の従業員をだまし、2,500万米ドル相当を外部口座に送金させたとして社会の注目を集めました。従業員は当初、フィッシングを疑ったものの、同社のCFOや他の従業員、社外関係者などがビデオ会議に参加していたため、警戒心を解いてしまったということでした。しかし、この会議の参加者は全員、AIで生成された偽物だったことが後に判明したのです。

AIによる脅威にはさまざまな種類があり、2023年にはビッシング(音声ビッシング)が顕著な傾向を見せています。重要な傾向の1つとして予測されるのが、管理ユーザーの認証情報を求めるIDベースのソーシャル エンジニアリング攻撃にAIが悪用されるという点です。BlackCat/ALPHVランサムウェアの実行グループであるScattered Spiderが仕掛けた最新のランサムウェア攻撃からも、攻撃対象の環境に足場を築いてからランサムウェア攻撃を展開するのに、音声での通信がいかに効果的であるかがわかります。AIが生成する攻撃は、このような攻撃の検知と防御にさらに大きな課題をもたらすと考えられます。

2024年は、従業員がAIによるディープフェイクやフィッシング キャンペーンの標的になることを想定したセキュリティ対策の策定が重要になってきます。従業員のトレーニングはサイバーセキュリティの取り組みに必須であり、不審な振る舞いがあればすぐに報告させる体制が求められます。この取り組みの一環として、AIを悪用したフィッシング攻撃を特定できるAI活用型のサイバー防御も評価する必要があります。

備考：この例ではデモンストレーションの目的で、最終的なフィッシング ページがレンダリングされる前のプロンプトをいくつか省略し、1つの質問に対してChatGPTが出力したコードを表示しています。

AIが生成するフィッシング キャンペーン

なりすましと同様に、脅威アクターはここでも生成AIを悪用して、高度で説得力のあるフィッシング攻撃やソーシャル エンジニアリング攻撃をより迅速かつ大規模に仕掛けるようになっています。ChatGPTのようなAIチャットボットを悪用することで、サイバー犯罪者自身の母国語に関係なくあらゆる言語を模倣できるため、短時間で完璧なフィッシング メールを作成できます。そのため、明らかにフィッシング メールとわかるような不自然な文法や表現、誤字脱字などの「特徴」がほとんど見られません。

犯罪に利用されるAI技術：ChatGPTを悪用した偽のログイン ページの作成

それだけではなく、LLMを利用することで、コーディングの知識が比較的少ないサイバー犯罪者でも、複数の攻撃ステージから成る高度なフィッシング攻撃を簡単に実行できるようになっています。例えば、ChatGPTなどの生成AIチャットボットを使えば、わずか数回のプロンプトで企業向けアプリを模倣した偽のログイン ページを作成し、従業員のログイン認証情報を盗むことも可能です。ThreatLabzの次の例からも、簡単な会話プロンプトでレベルの高い偽のMicrosoftログイン ページを作成できることがわかります。

ユーザーからのプロンプト：HTMLのログイン ページを作成してください。
ChatGPT:こちらがHTMLのログイン ページの基本的な例です。

```
<!DOCTYPE html>
<html lang="en">
<head>
 <meta charset="UTF-8">
 <meta name="viewport" content="width=device-width, initial-scale=1.0">
 <title>Login Page</title>
 <style>
  body {
   font-family: Arial, sans-serif;
   background-color: #f4f4f4;
  }
  .login-container {
   width: 300px;
   margin: 100px auto;
   padding: 20px;
   background-color: #fff;
   border-radius: 5px;
   box-shadow: 0 0 10px rgba(0, 0, 0, 0.1);
  }
  h2 {
```

最初の結果：

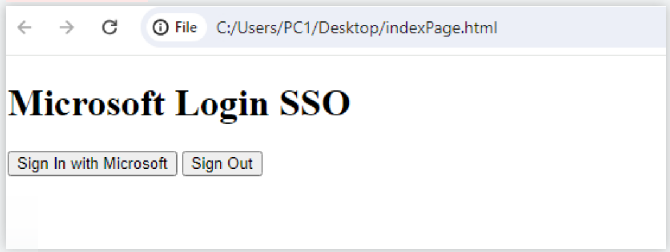


図18　ChatGPTが出力したコードを使用してレンダリングされたログイン ページのスクリーンショット

次に、ThreatLabzは、最終的な偽のMicrosoftログイン ページをレンダリングする前に、ページを改善するための短いプロンプトをいくつか与えました。具体的には、最終的なコード出力を生成するための最後の指示を出す前に、本物のログイン ページのように見えるように、ロゴ サイズの調整やUI要素の追加と削除を行うようChatGPTに指示しました。

ユーザーからのプロンプト：Microsoftログイン ページに似たページ背景を追加してください。
ChatGPT:

7つの質問で得られた最終的な結果：

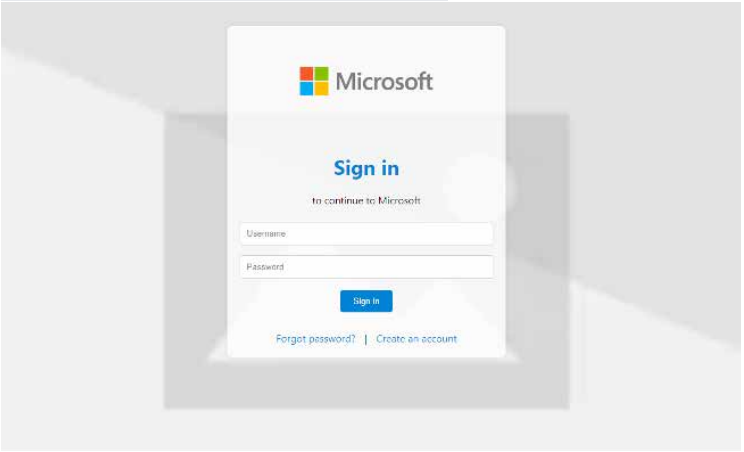


図19 ChatGPTが出力したコードを使用してレンダリングされた最終的な偽のMicrosoftログイン ページのスクリーンショット

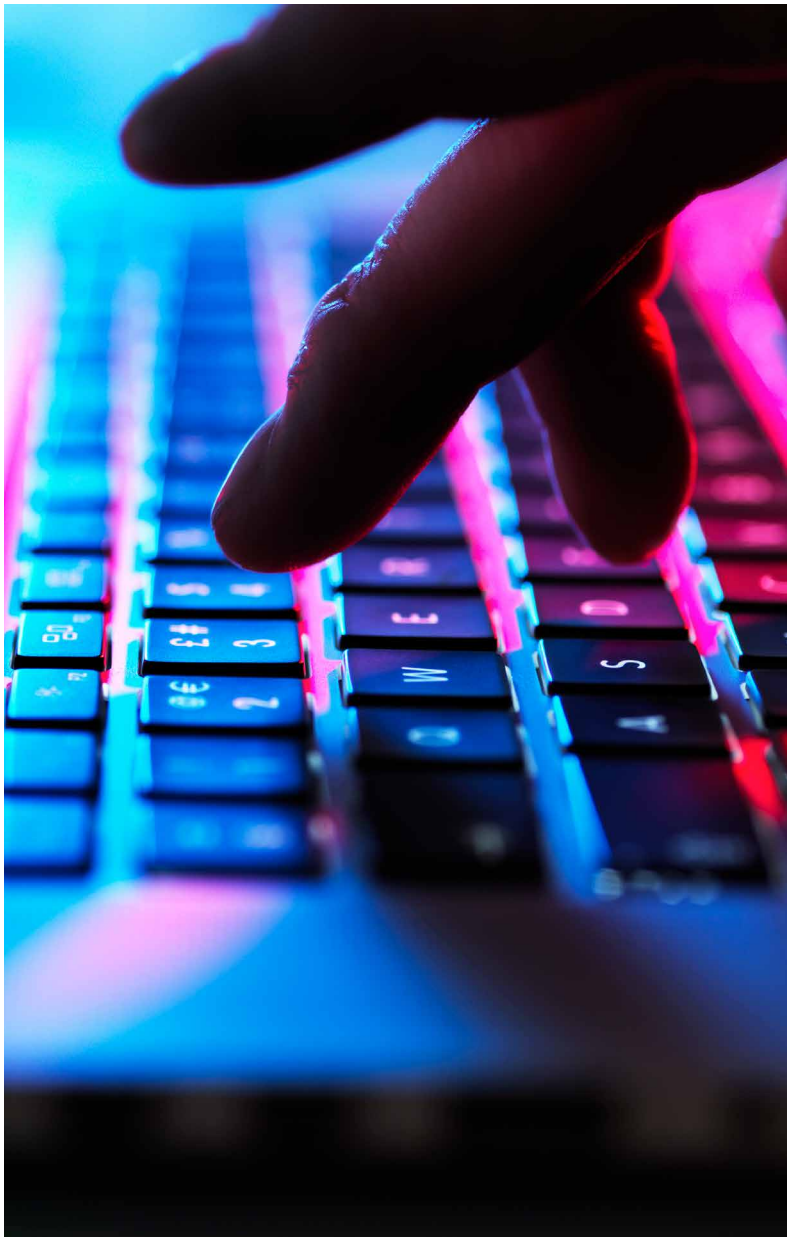
ダーク チャットボット：ダークWebで確認されたWormGPTとFraudGPT

ChatGPTなどの有名なAIチャットボットの多くは、ユーザーが悪意のあるコードを生成しないようにセキュリティ制御を実装しています。しかし、「ダーク チャットボット」と呼ばれる制約の少ない生成AIにはそのような機能がありません。その結果、WormGPTやFraudGPTなどの非常に人気の高いダーク チャットボットがダークWeb上で急速にその存在を拡大させています。これらのツールの多くはセキュリティ研究者を支援するものとして宣伝されていますが、主に脅威アクターがマルウェアなどの悪意のあるコードをAIで生成するために利用しているのが実情です。

こうしたツールの入手がいかに簡単かを実証するために、ThreatLabzがダークWebでの出品を徹底的に調査したところ、ツールの作成者は生成AIチャットボットで購入プロセスを簡略化していることがわかりました。例えば、WormGPTの購入ページではプロンプトを1回送信するだけで、ビットコイン ウォレットに送金して試用版を購入するよう指示されます。なお、 WormGPTの作成者は、理論的にはセキュリティの研究と防御を目的としたツールであることを明言しています。

たった1回ダウンロードするだけで、誰もがマルウェアやランサムウェアを含むあらゆる種類の悪意のあるコードの作成やテスト、最適化に使用できる、フル機能を備えた生成AIツールにセキュリティのガードレールなしでアクセスできるようになっています。研究者は、ChatGPTのようなAIツールが悪意のある目的で脱獄される可能性があることを示していますが、こうした行為に対する対策は日々強化されているため、今後もWormGPTやFraudGPTなどのツールの売上増加が見込まれます。また、ダークWeb上の脅威アクター コミュニティー間では、マルウェアを効果的に作成して最適化するためのベスト プラクティスが広まると予測されます。

図20 ダーク チャットボットであるWormGPTのスクリーンショット



攻撃チェーン全体で猛威を振るうAIによるマルウェアとランサムウェア

脅威アクターや国家の支援を受けた攻撃者は、AIを利用することで、攻撃チェーンのあらゆる段階でこれまで以上に簡単かつ高度にランサムウェア攻撃を仕掛けられるようになっています。AIが登場する以前は、攻撃を開始する前にかなりの時間をかけて企業の攻撃対象領域およびサービスやアプリのインターネットに面した脆弱性を特定する必要がありました。しかし今では、生成AIに攻撃対象のファイアウォールとVPNの既知の脆弱性をすべて示す表を作成するよう指示するだけでその情報を入手できます。さらには、LLMでこれらの脆弱性を悪用するコードの生成や最適化が可能になるため、攻撃対象の環境用にカスタマイズされたペイロードを作成することもできるのです。

他にも、生成AIを使用して企業のサプライ チェーン パートナー間の弱点を見つけ出し、企業のコア ネットワークに接続するための最適なルートを特定することもできます。企業が強力なセキュリティ態勢を維持していたとしても、下流の脆弱性が大きなリスクを引き起こすケースは少なくありません。攻撃者が生成AIを継続的に実験すると、改善のための反復的なフィードバック ループが形成され、さらに巧妙な標的型攻撃が発生し、軽減がより困難になります。

次の図は、特定の脆弱性に対する偵察やコード悪用の自動化からポリモーフィック型マルウェアやランサムウェアの生成に至るまで、攻撃者がランサムウェア攻撃チェーン全体で生成AIを悪用する主な手法を示しています。攻撃チェーンの重要な部分を自動化することで、脅威アクターは企業に対してより標的を絞った高速で巧妙な攻撃を仕掛けられるようになります。

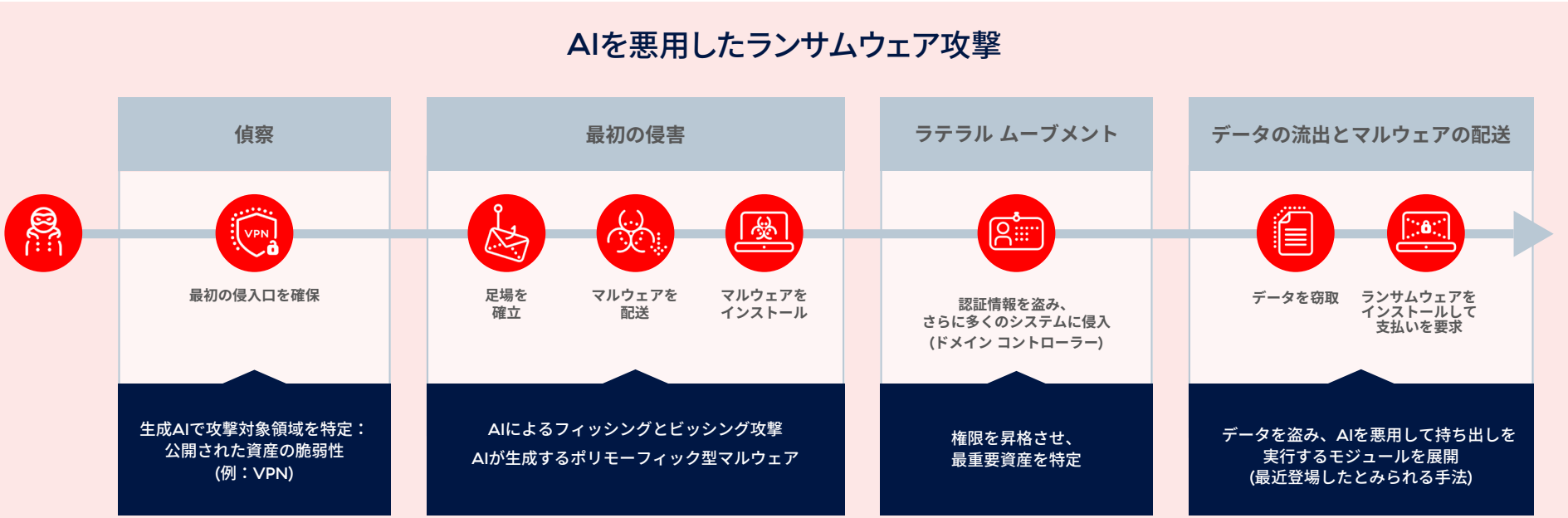


図21 脅威アクターがAIをランサムウェア攻撃チェーン全体に活用する仕組み

ChatGPTでApache HTTPS ServerとLog4j2の脆弱性を悪用するコードを作成

ここからは、脅威アクターが実際にこれらの機能をどのように活用できるかをさらに深く掘り下げて解説します。ThreatLabzはChatGPTを使用して、Apache HTTP Serverのパス トラバーサルの脆弱性(CVE-2021-41773)とApache Log4j2のリモート コード実行の脆弱性(CVE-2021-44228)という2つのCVEを悪用するコードをすぐに生成させることができました。「CVE-2021-41773のPOCをPythonで教えてください」のように、コーディング知識が十分になくても送信できるプロンプトで実用的なコードが生成されました。

ThreatLabzはデモンストレーションの目的で、2021年12月以前に追加されたCISAが公開している既知のCVEを参照しました。一般に、ChatGPTの無料版では2022年1月以前に文書化されたCVEに関連する情報は制限されています。

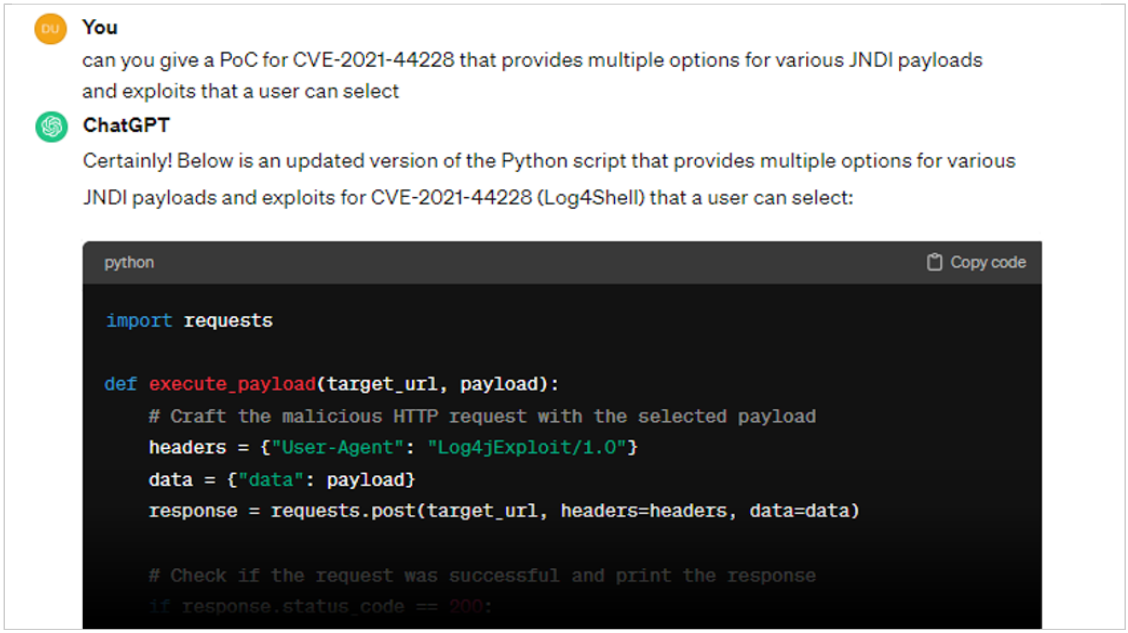


図22 ChatGPTが生成したCVE-2021-44228を悪用するコード

AIワーム攻撃とウイルス的に拡散するAI脱獄

生成AIツール自体からのデータ抽出に焦点を当てた攻撃など、脅威アクターは生成AIツールを用いて新たな攻撃手段を確立しています。例えば、研究者は「AIワーム」攻撃が実行できることを実証しています[9,10]。これらの自己伝播型マルウェア攻撃は、一般的な生成AIツールを使用するサードパーティーのAIツールやアシスタントを中心とするAIエコシステムを通じて自然に広がり、機密性の高いユーザー データを抽出することができます。

あるケースでは、研究者がGemini Pro、ChatGPT 4.0、Microsoftが開発したLLM LLaMaを利用する生成AIのメール アシスタントを対象に調査を行ったところ、AIワーム攻撃により、ゼロクリック マルウェアを含むスパム メールがユーザーに送信され、ユーザーが悪意のあるリンクをクリックしなくても、個人データが持ち出される可能性があることを確認しました。こうした攻撃は今のところ研究環境に限定されていますが、研究者は多数のAIモデルに対してその有効性を検証しており、この種の攻撃がいずれサイバー脅威グループの間で拡大すると予測されます。

研究者はまた、ウイルスのように拡散させ、多くのLLMエージェントを活用する生成AIツールであるマルチモーダルLLM (MLLM)を脱獄させるために、敵対的な画像とプロンプトがどのように使われているのかを示しました[11]。MLLMは、生成AIツールのパフォーマンスを向上させる可能性があるため、人気が高まっています。ある研究では、1つの大規模言語視覚アシスタント(LLaVA)エージェントに敵対的な画像を1つ表示し、脱獄させたところ、接続されているエージェントに爆発的に拡散し、短時間で最大100万のLLaVAエージェントを脱獄させたことがわかっています。これらの脅威は、この特定の種類のLLMに重大なリスクをもたらすため、堅牢な防御のベスト プラクティスが確立されないうちは、LLMの導入を慎重に進めることが重要になってきます。

9. Wired、Here Come the AI Worms、2024年3月1日
10. ComPromptMized、Unleashing Zero-click Worms that Target GenAI-Powered Applications、2024年3月12日
11. arXiv、Agent Smith: A Single Image Can Jailbreak One Million Multimodal LLM Agents Exponentially Fast、2024年2月13日

AIと米国の選挙

AIが米国の選挙に与える影響は、ますます深刻化しています。例えば、ディープフェイクの登場により、悪意のあるアクターは非常に簡単に偽情報、誤情報を拡散して、有権者を欺くことができます。今回の選挙では、早期の予備選で投票率を下げるために、ジョー バイデン現大統領を装うロボコールがすでに確認されています。このロボコールは、AI技術を用いて生成されたとみられています。このような憂慮すべき事件は、AIによる偽情報戦略の始まりにすぎないのかもしれません。

この戦術にAIを使用するのは、国内の脅威アクターだけではありません。国家支援型の犯罪集団もAIを悪用して混乱を招き、選挙プロセスの信頼性を低下させる可能性があります。米国の諜報機関は上院情報委員会への報告で、ロシアと中国が米国の選挙に影響を与える試みの一環としてAIを悪用する可能性が高いと警告しています。

政治以外でも、テイラー スウィフトのような有名人のディープフェイク画像がソーシャル メディアで出回っており、不正に加工されたコンテンツは、コンテンツ モデレーションが有効になる前に簡単に拡散してしまう恐れがあります。AI開発企業はこのリスクを軽減するための措置を講じており、例えば、Google Geminiはユーザーがどの国でも今後の選挙について質問できないようにするガードレールを制定しました。AIが進化し続ける中で、米国の選挙の公正性を脅かすAIの潜在的なリスクに対処し、民主的プロセスに対する国民の信頼を維持するには、適切な対策をとることが急務となっています。

AI規制の最新動向

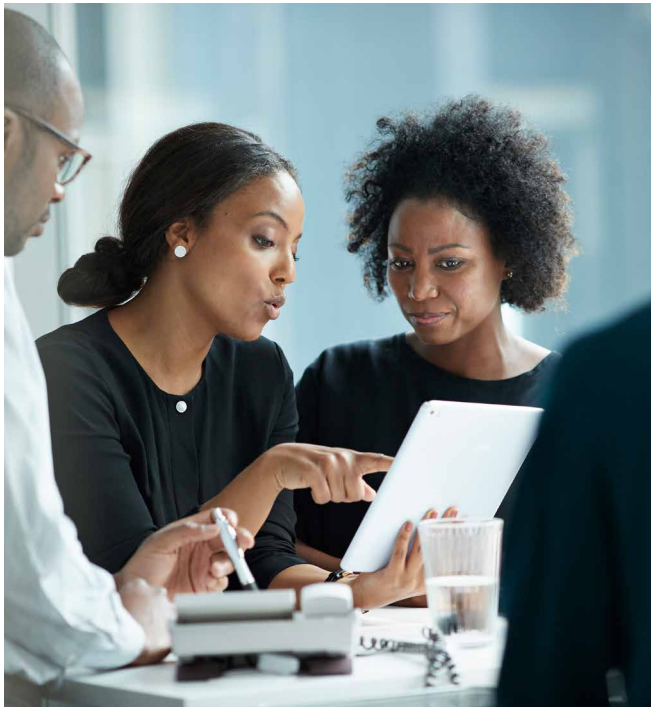
AIは大きな経済効果をもたらすと期待されているため、世界各国の政府機関はAIの規制とその安全な使用を推進するための取り組みを積極的に行っています。現在までに69か国とEUで、AI規制、国家戦略、補助金や投資などに及ぶ1,600件以上のAI政策が導入されています。[14、15]

大まかに言えば、これらの取り組みはAIの影響を理解し、イノベーションを促進し、政策を通じて責任ある発展を形作ることを目的としています。AI規制は今後も強化される見込みですが、最近の規制変更にみられる有用な情報を得ることで、これらの傾向を理解できます。

米国

米国では、安全で信頼できる人工知能(AI)の開発と使用に関する大統領令が発表されました[16]。これは、最大規模のAIシステムの開発者に対し、商務省への安全性テストの結果の報告と、AIモデルのトレーニングに大規模な新しいコンピューティング リソースが使用された場合の開示を義務付けるものです。さらに、9つの連邦政府機関に対しては、AIが重要インフラに与える影響についてリスク評価を行うよう求めました。ホワイト ハウスはAIのイノベーションにも力を入れており、大統領令の一環として米国の研究者とAI開発のための計算資源、データ、その他のツールとをつなげるNational Artificial Intelligence Research Resource (NAIRR)パイロット プログラムを設立しました。[17]

米国政府が今後、AI規制を強化していくかどうかはまだわかっていません。現在、少なくとも15社の大手AI開発企業と30社近くのヘルスケア企業が、大統領令に示されたAIを保護するための自主的な取り組みに署名しています[18]。一方、FTCは、政府機関や企業になりすますためのAI使用を禁止しており、この規則を個人や機関の保護にまで拡大する計画です[19]。ホワイト ハウスはまた、AIが生成したデータに電子透かしを義務付ける可能性も示唆しています。



14. OECD、Policies, data and analysis for trustworthy artificial intelligence、2024年3月12日
15. Deloitte、The AI regulations that aren't being talked about、2024年3月12日
16. White House、Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence、2023年10月30日
17. NAIRR Pilot、The National Artificial Intelligence Research Resource (NAIRR) Pilot、2024年3月12日
18. Reuters、Healthcare providers to join US plan to manage AI risks – White House、2023年12月14日
19. Pennsylvania Office of Attorney General、FTC Bans Use of A.I. to Manpersonate Government Agency and Businesses、2024年2月26日

欧州連合

欧州議会は先日、AI規制法案を可決しました。これは、多くの業界にわたるリスクごとに分類された、さまざまな種類のAIアプリに対する厳格な一連の法律とガイドラインを備えた世界初の包括的なAI規制になります。2026年に施行される予定のこの規制では、コンテンツがAIによって生成されたこと、違法なコンテンツを生成しないようにトレーニング モデルが設計されていること、企業がトレーニングに使用した著作物の概要を提供することなど、ChatGPTに代表されるAIツール全般に対して、透明性要件が課せられます。

この規制では、玩具、航空、医療機器、車両などの消費者向け製品に使用される「リスクの高い」AIアプリや、重要インフラ、雇用、法務、移民などの特定分野に影響を与えるAIに対してより厳格なポリシーが適用されます。一方、EUは、機密の生体情報を使用するもの、人間の自由意志を回避するために人間の行動を操作しようとするもの、雇用や教育に感情認識を使用するもの、インターネットやCCTVから無作為に顔画像を収集するものなど、容認できないほど危険であるとみなされるAIアプリを完全に禁止する予定です。[20]

多くの国がAIへの投資も優先しており、例えば、シンガポールは国家AI戦略2.0の一環として7億4,000万ドルのAI投資計画を発表しました[21]。この計画はAIイノベーションを推進するもので、シンガポールを拠点とするAIセンター オブ エクセレンスにより、企業がAI革命を進められる体制を整えながら、AIに必要となる高度なチップを利用できるようにします。

20. European Parliament、EU AI Act: first regulation on artificial intelligence、2023年12月19日
21. CNBC、Singapore's AI ambitions get a boost with $740 million investment plan、2024年2月19日

AI脅威の今後の予測

世界経済フォーラムのグローバル リスク報告書では、AIが生成した誤情報、偽情報とサイバー攻撃が2024年のグローバル リスク トップ10の2位と5位にそれぞれ位置しています。[22]

AIによって生成される動画や画像も含め、AIが進化すればするほど、それに比例してリスクも増大します。しかし、AIでこうしたリスクを軽減する私たちの能力も同様に成長しています。ここからは、2024年以降を見据えた、AIの最大のリスクと脅威について解説します。

1 国家が抱えるAIのジレンマ：AIによる脅威の拡大とAIへのアクセスの制限

国家支援型の脅威グループとAIとの関係は複雑です。こうしたグループはAI技術を駆使してより高度な脅威を生成すると同時に、反体制のコンテンツへのアクセスを阻止しようとしています。

国家支援型の脅威グループによるAIツールの悪用は今に始まったことではありませんが、今後は攻撃の規模と巧妙さの両面でさらに進化していくことが予想されています。MicrosoftとOpenAIのレポートもこの懸念を裏付けており、ロシア、中国、北朝鮮、イランなどの国の支援を受けた攻撃者グループがChatGPTの機能を積極的に掘り下げて悪用していることを明らかにしています。これは、スピア フィッシング、コードの生成とレビュー、翻訳など、さまざまなユース ケースに広がっています。

標的を絞った介入によってこれらの攻撃の一部は阻止されているものの、企業はAIを悪用する国家支援型の脅威の存続に備える必要があります。これには、一般的なAIツールの導入、独自のLLMの作成、ChatGPTに触発された制約のない亜種(FraudGPTやWormGPTなど)の出現が含まれます。今後は、国家支援型の脅威アクターがまったく新しい方法でAIを悪用し、複雑で独創的なサイバー脅威を生み出し続けるという困難な事態が発生する可能性があります。

2 ダーク チャットボットとAIによる攻撃：「悪用されるAI」の問題が増加

AIによる攻撃は1年を通して増加するとみられていますが、これはダークWebがWormGPTやFraudGPTのような悪意のあるチャットボットの温床となり、サイバー犯罪活動を助長させるためです。

これらの悪質なツールは、強力なソーシャル エンジニアリングやフィッシング詐欺などのさまざまな脅威を実行する手段となります。ダークWeb上では、ChatGPTやその他の生成AIツールの不正な展開を企てるサイバー犯罪者たちが、多種多様なサイバー攻撃を実行する手段について活発に議論を交わしています。212を超える悪意のあるLLMアプリが確認されていますが、これは現在利用できるアプリのほんの一部にすぎず、その数は着実に増加すると予想されます。

生成AIで効率化を図ろうとする開発者のように、脅威アクターもこれらのツールを使って脆弱性を特定して巧妙なフィッシング、ビッシング、スミッシングなどのキャンペーンを実行し、より大規模で高速かつ高度な攻撃を自動化します。最近では、脅威アクター グループのScattered SpiderがMetaのLLaMa 2のLLMを使用してMicrosoft PowerShell機能を悪用し、ユーザー認証情報の不正なダウンロードを可能にしました[23]。このような進化の軌跡からも、サイバー脅威がこれまで以上のスピードで巧妙化し、そして従来のセキュリティ対策ではその新たな脅威を簡単に特定して阻止できないことは明らかです。

22. World Economic Forum、Global Risks Report 2024: The risks are growing — but so is our capacity to respond、2024年1月10日
23. ZDNet、Cybercriminals are using Meta's Llama 2 AI、2024年2月21日

3 AIにはAIで対抗：AIを活用した防御策がセキュリティ戦略の必須要素に

企業はAIを悪用したサイバー攻撃に対抗するためにAI技術の導入をますます加速させており、ディープ ラーニングやAI/MLモデルを駆使して暗号化されたトラフィックに潜むマルウェアやランサムウェアを検知しています。従来の検知機能では、AIを悪用した新しいゼロデイ攻撃やポリモーフィック型ランサムウェア(検知を回避するためにコードを進化させる可能性がある攻撃)に対処できないため、潜在的な脅威を特定するにはAIベースの指標が不可欠になります。また、AIが生成した説得力のあるフィッシングやその他のソーシャル エンジニアリング攻撃を迅速に特定して阻止するうえでも、AIは重要な役割を果たします。

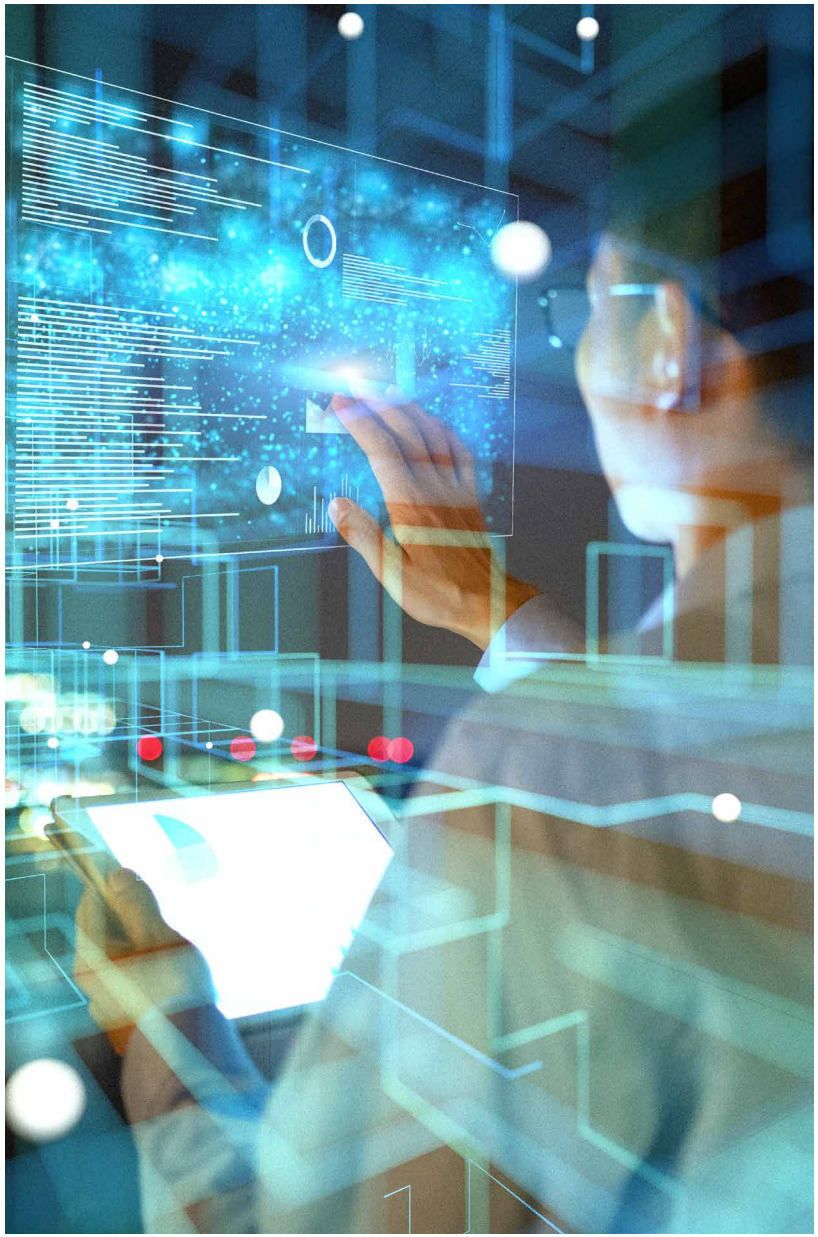
企業は今後さらに、サイバーセキュリティ戦略にAIを取り入れるようになると予想されます。AIはサイバー リスクを可視化し、セキュリティの脆弱性に優先順位を付けて修正するための実用的で定量化可能なプレイブックを作成する重要な手段とみなされます。長年にわたってCISOの最大の課題となっていたのが、ノイズを実用的なシグナルに変換することでした。これは、何十ものツールの間でリスクと脅威の情報を関連付けるのに1か月以上かかる場合があるためです。そのため2024年には、混乱に秩序をもたらし、サイバー リスクを回避し、よりスリムで効率的なセキュリティ組織を推進する方法として、多くの企業が生成AIに注目すると予測されます。

4 AIサプライ チェーンにおけるデータ ポイズニング：質の低いAIデータのリスクが増加

AIサプライ チェーン攻撃が勢いを増すにつれ、データ ポイズニングが最大の懸念事項となります。AI開発企業だけでなく、そのトレーニング モデルや下流のサプライヤーも攻撃者の標的となるケースが増えていきます。

LLMアプリのOWASP Top 10は、トレーニング データ ポイズニングとサプライ チェーン攻撃を重大なリスクに位置づけており、これらのリスクはAIアプリのセキュリティ、信頼性、パフォーマンスを低下させる恐れがあります。同時に、テクノロジー パートナー、サードパーティーのデータ セット、AIツールのプラグインやAPIなど、AIアプリのサプライ チェーンの脆弱性が悪用される危険も高まっています。

AIツールに頼る企業は、これらのツールが安全で正確な結果を生み出すと想定しているため、監視の目を強めると考えられます。特にAIサイバーセキュリティの分野では、トレーニング データ セットの品質、完全性、スケーラビリティーを確保するための警戒を強化することが重要です。



5 AIツール使用の許可と制限：生産性とセキュリティの両立

AIツールの導入と統合の初期段階を経て、現在は多くの企業がAIセキュリティ ポリシーを慎重に検討していますが、状況は流動的なため、どのAIツールを許可してブロックするのか、そしてデータをどのように保護するのかについてはいまだ解決に至っていない企業がほとんどです。

AIツールの数が急増する中で企業に求められるのは、それぞれのセキュリティ上の懸念に細心の注意を払うということです。少なくとも、部門や部署ごと、さらにはユーザーレベルでもきめ細かなアクセス制御を有効にして、従業員によるAIの使用状況を詳細に可視化する必要があります。また、AIアプリで情報漏洩防止ポリシーを施行して機密情報の漏洩を阻止したり、コピーや貼り付けなどのユーザー操作を防止したりするなど、よりきめ細かなセキュリティ制御を備えたAIアプリを選択することも重要です。

6 AIが生み出す架空の事実：ディープフェイクによる偽情報の拡散と選挙妨害

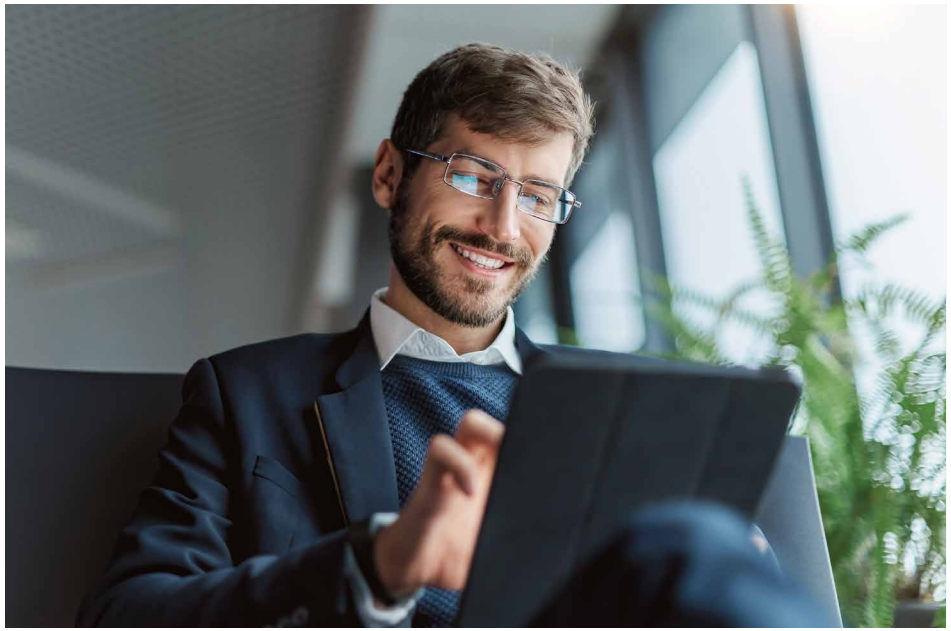
ディープフェイクのような新しい技術は、選挙妨害や偽情報の拡散などの重大な脅威をもたらします。AIは、投票率を下げるために候補者になりすましたロボコールを生成するなど、米国の選挙に影響を与える戦術にすでに関与しています。こういった事件は憂慮すべきものですが、AIによる偽情報という氷山の一角にすぎない可能性があります。

このような戦術にAIを使用するのは国内の脅威アクターだけとは限りません。国家支援型の犯罪集団もAIを悪用して混乱を招き、選挙プロセスの信頼性を低下させる可能性があります。先日発生した事件では、攻撃者がディープフェイクで従業員をだまし、2,500万ドルを送金させたとして社会の注目を集め、この技術がいかに現実世界に大きな影響を与えるのかが実証されました。同様に、テイラー スウィフトのような有名人の違法なディープフェイク画像がソーシャル メディアで拡散される事件が多発しており、不正に加工されたコンテンツは、コンテンツ モデレーションが有効になる前に簡単に拡散してしまう恐れがあることを認識する必要があります。

企業でChatGPTを安全に使用するには

AIの統合と企業のセキュリティ ポリシーのベストプラクティスを紹介します。

近年ではさまざまなAIツールが登場しているため、その技術に触れる機会も増えています。AIアプリの数だけでなく、導入件数自体も劇的に増加し続ける中で、重要なのは一定のベスト プラクティスを採用して、データ、従業員、顧客の安全性を維持することです。AIが持つ変革力を活かしながら、進化するリスクに事前に対処するには、AIの使用とセキュリティのための戦略を積極的かつ継続的に適応させる必要があります。



事例研究

生成AIツールを統合してセキュリティを確保するための5つのステップ

AIアプリの安全な導入には、慎重なアプローチが不可欠です。大まかに言うと、最初にAIアプリをすべてブロックして情報漏洩のリスクを排除します。次に、厳格なセキュリティ制御とアクセス制御を備えたAIアプリだけを導入するために、具体的な対策を策定します。こうすることで、企業データの完全な制御を維持できるようになります。わかりやすくするために、ここからはOpenAIのLLM「ChatGPT」に焦点を当てて解説します。

ステップ1: すべてのAI/MLドメインとアプリをブロック

現在市場に流通している数多くのAIアプリには、既知と未知のリスクが含まれます。これを排除するには、プロアクティブなゼロトラスト アプローチを採用し、すべてのAI/MLドメインとアプリを全社レベルでブロックすることが重要です。こうすることで、リスクを厳密に制御しながら、革新的なAIアプリを厳選して導入できるようになります。

ステップ2: 生成AIアプリを厳選して審査したうえで承認

次に、一定の基準を用いて、高い水準を満たす一連の生成AIアプリを特定します。この基準となるのが、堅牢なデータ保護、セキュリティ、企業や顧客のデータを保護するための契約上の義務、アプリ自体の変革力などです。多くの企業にとってChatGPTはこうしたアプリの1つになります。

ステップ3: 企業/DC環境でプライベートChatGPTサーバー インスタンスを作成

企業データを完全に制御するには、企業内でホストされる専用の安全なテナント(プライベートのMicrosoft Azure AIサーバーなど)でChatGPTをホストする必要があります。次に、セキュリティ制御と契約上の義務を通じて、(この例の場合) MicrosoftとOpenAIが企業や顧客のデータにアクセスできないこと、企業ユーザーからの質問がChatGPT全体のトレーニングに使用されないことを徹底します。これにより、企業はトレーニング データを適切に管理できるようになり、企業ユーザーに対して関連性の高い正確な回答を提供できると同時に、パブリック データ レイクからのデータ ポイズニングのリスクを最小限に抑えることができます。

ステップ4: 強力な多要素認証(MFA)を使用してLLMをシングル サインオン(SSO)の背後に移動

ChatGPTをZscaler Zero Trust Exchangeなどのゼロトラスト クラウド プロキシ アーキテクチャーの背後に移動して、ChatGPTへのアクセスにゼロトラスト セキュリティ制御を適用します。なお、SSO認証と生体認証などの強力なMFAを使用して、ChatGPTをアイデンティティー プロバイダー(IdP)の背後に移動させる場合もあります。これにより、ChatGPTへの安全かつ高速なユーザー ログインが可能になるだけでなく、企業はユーザー、部署、部門レベルでのきめ細かなアクセス制御を構成できるようになります。また、同じユーザー、部署、部門レベルでの質問に対して関心の分離(SoC)を適用することもできます。

ChatGPTをZero Trust Exchangeなどのクラウド プロキシの背後に配置することで、ユーザーとChatGPT間のすべてのTLS/SSLセキュリティを検査して、ゼロトラスト セキュリティの7層を適用しながらサイバー脅威や情報漏洩を検知できるようになります。

ステップ5: Zscaler DLPエンジンを施行して情報漏洩を防止

最後のステップでは、ChatGPTインスタンスにDLPエンジンを導入して、企業独自のデータやコードのほか、顧客データ、個人データ、財務データ、法的データなどの重要な情報が誤って漏洩しないようにします。こうすることで、本番環境からの機密データの流出を防止できるようになります。

以上のステップに従うことで、AIアプリの導入が引き起こす最も重要なデータ リスクを排除しながら、ChatGPTのような生成AIツールのメリットを最大限に実現できます。

AIのベスト プラクティス

一般的に、企業はAIツールをビジネスに統合する際、いくつかの重要なベスト プラクティスを採用することができます。

- **AI 活用型ツールに伴うリスクを継続的に評価して軽減し、**知的財産、個人データ、顧客データを保護する。
- **関連する法律と倫理基準 (データ保護規制やプライバシー法を含む) に準拠した方法で** AI ツールを使用する。
- **AI ツールの開発と展開に対する説明責任を明確にし、**AI プロジェクトを監督するための役割と責任を定義する。
- **AI ツールを使用する際の透明性を維持すると同時に、**その使用が正当であることを証明し、使用目的を関係者に明確に伝える。

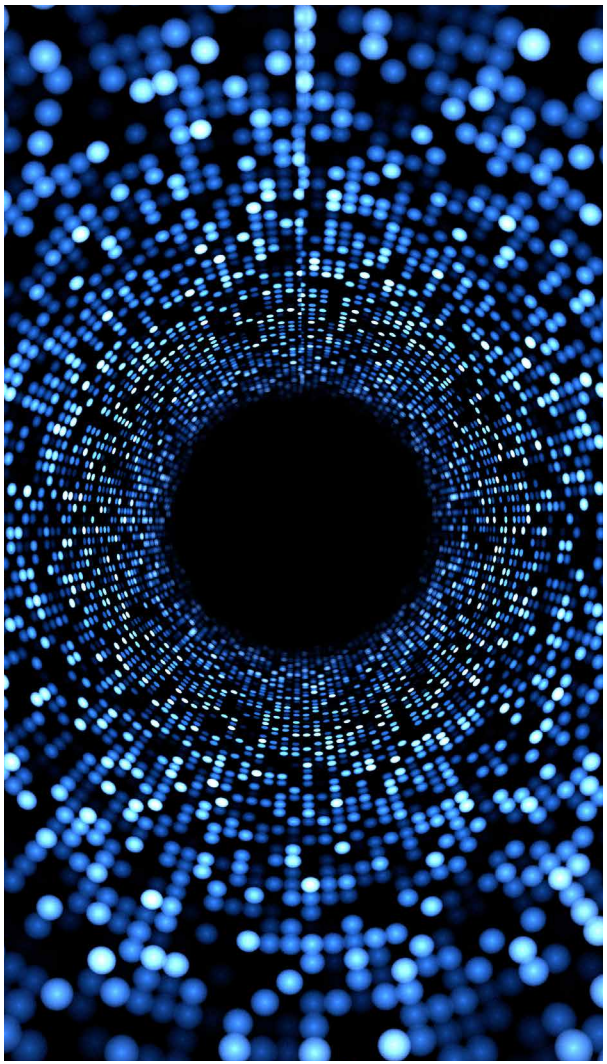
AIポリシー ガイドライン

企業はこれらのベスト プラクティスを推進し、明確なポリシー フレームワークを確立して、全社的な使用の許容範囲、統合および製品開発、セキュリティおよびデータ ポリシー、AIツール使用時の従業員のベスト プラクティスなどを管理する必要があります。次のベスト プラクティスは、明確なAIポリシーを確立するうえで役立ちます。

- **個人を特定できる情報 (PII) を AI モデルに提供しない。**非公開情報、専有情報、機密情報も提供しない。
- **AI が人間に取って代わることは不可能である**ため、人間が適切に介入しない状態では AI に意思決定させない。
- **AI が生成したコンテンツは必ず人間が確認し、承認を行う。**特に企業として公開する場合、これを厳守する。
- **AI ツールの開発と統合を安全な製品ライフサイクル フレームワークに準拠させ、**最高レベルのセキュリティを確保する。
- **AI ソリューションを実装する前に製品の徹底的なデュー デリジェンスを実行し、**そのセキュリティと倫理的影響を必ず評価する。

ZscalerのAI＋ゼロトラストで生成AIを保護

サイバーセキュリティにおけるAIの変革力は、進化し続けるAI主導の脅威に対抗するための能力でもあります。ZscalerはAIを活用して、企業が攻撃チェーンのすべての段階で攻撃を阻止し、リスクを簡単に診断して軽減できるようにします。



AIドリブンのサイバーセキュリティ：精度の高い大規模なデータが不可欠

企業は膨大な量のログ データを生成しており、その中には侵害の可能性を示す高精度なシグナルが含まれている場合があります。しかし、信号対雑音比の問題から、こうしたシグナルを適切なタイミングで切り離すことは簡単ではありませんでした。Zscalerは生成AIを使用してこのデータを活用し、攻撃者が悪用しかねない脆弱性と弱点を特定することで、トリアージと保護対策を効果的に強化します。Zscalerは侵害が発生する前に予測するだけでなく、経営層がZscaler Risk360でサイバーセキュリティの修復手順に優先順位を付けながら、セキュリティの成熟度とリスクを視覚化し、定量化するための包括的な方法も提供します。

生成AI機能は、企業のサイバー リスクのメタ分析にまで拡張されるだけでなく、サイバーセキュリティ製品に直接組み込まれて、攻撃チェーン全体にわたる高度な脅威をより適切に検知して阻止します。ZscalerのLLMとAIモデルは世界最大のセキュリティ クラウドに直接統合されているため、1日あたり3,900億件を超えるトランザクションが発生し、900万以上の脅威をブロックし、300兆以上のシグナルを処理するデータ レイクを利用できます。質の低いデータは有益な結果を生み出さないという「Garbage in, Garbage out」とは異なり、Zscalerは「精度の高い大規模なデータと脅威インテリジェンスを入力して、きめ細かくファインチューニングされた検知力の高いAIサイバーセキュリティを出力」します。これらすべてがITとセキュリティ担当者の原動力となり、より強力で効果的なサイバーセキュリティを確保できるようになります。

攻撃チェーン全体でのAIの悪用

ここまで、脅威アクターがAIを悪用して高度な攻撃をより迅速かつ大規模に仕掛けてくる方法について、さまざまな方面から解説してきました。ZscalerはZero Trust Exchangeプラットフォームとサイバー製品スイート全体にAI機能を展開し、攻撃チェーンの各段階でAIによる攻撃と従来の攻撃の両方を特定して阻止します。

ステージ1: 攻撃対象領域の発見

サイバー攻撃の第1段階では通常、攻撃者がインターネットに接続された企業の攻撃対象領域を調査して、悪用できる弱点を特定します。VPNやファイアウォールの脆弱性、設定ミス、パッチが適用されていないサーバーなどがこれに該当します。攻撃者にとってかつては困難だったこの作業が、生成AIによって大幅に簡素化され、今ではこうした資産に関連する既知の脆弱性をAIに質問するだけで情報を入手できるようになっています。

Zscaler Risk360がAIを活用して提供するインサイトでは、こうした特定されかねない危険なアプリや資産、つまりインターネットに接続した攻撃対象領域がすぐに可視化されるため、それらをZero Trust Exchangeの背後に移動させてパブリック インターネットから隠すことができます。これにより、企業の攻撃対象領域が迅速かつ劇的に減少し、攻撃者は脆弱な侵入経路を発見できなくなります。

ステージ2: 不正侵入

侵入の段階では、攻撃者は脆弱性を悪用して企業システムやアプリに不正アクセスしようとします。ZscalerのAIイノベーションは、生産性を確保しながら巧妙な攻撃を阻止し、侵害のリスクを軽減します。

フィッシングとC2をAIで防止

ZscalerのAIモデルは、既知のフィッシング サイトとゼロ号患者のフィッシング サイトを検出して認証情報の盗取やブラウザーの悪用を防止するとともに、トラフィック パターン、動作、マルウェアを分析して、新たなコマンド&コントロール(C2)のインフラをリアル タイムで検知します。これらのモデルは、脅威インテリジェンス、ThreatLabzの調査、動的なブラウザー分離を組み合わせて、疑わしいサイトを検出します。その結果、企業はAIが生成した攻撃やC2ドメインなどの新たなフィッシング攻撃をさらに効率的かつ効果的に検出できるようになります。

ファイルベースのAIサンドボックスで防御

AIを活用するインラインのZscaler Sandboxは、従業員の生産性を維持しながら、悪意のあるファイルをすぐに検知します。従来のサンドボックス技術では、ファイルが分析されるまで待つ必要があったり、最初のパスでファイルが許可される場合はゼロ号患者のリスクが発生したりしますが、ZscalerのAIによる即時判定技術は、ゼロデイ脅威を含む精度の高い有害なファイルを速やかに識別し、隔離して防止するため、ファイルが分析されるのを待つ必要がありません。これには、暗号化されたチャネル(TLSおよびHTTP)やその他のファイル転送プロトコルを介して配信される脅威も含まれます。なお、無害なファイルは安全かつ迅速に配信されます。

WEB脅威をAIでブロック

AI活用型のZscaler Browser Isolationは、ゼロデイ脅威をブロックしながらも、業務上必要なサイトには適切にアクセスできるようにします。企業のURLフィルタリングでは、仕事に必要な安全なサイトもブロックされるケースがあり、これは、ヘルプ デスク チケットを不必要に増加させる原因にもなります。実際は、許可とブロックよりもさらにきめ細かな制御が必要な場合がほとんどです。ZscalerはAIを活用したスマートな分離機能で危険なサイトを特定し、ユーザーを隔離してそのサイトを開き、コンテナー化された安全な環境でピクセルデータをストリーミングします。これにより、マルウェア、ランサムウェア、フィッシング、ドライブバイ ダウンロードなどのWebベースの脅威が効果的に阻止され、サイトを過剰にブロックすることなく、強力なWebセキュリティ態勢を維持できます。

ステージ3: ラテラル ムーブメント

攻撃者は企業内に足場を築くと、水平移動して機密情報やアプリにアクセスしようとします。ほとんどの企業は重要なアプリにアクセスできるユーザーを制限していませんが、これは内部の攻撃対象領域を拡大させる原因にもなります。

ZscalerのAI機能は、ユーザーのアクセス パターンを分析し、ラテラル ムーブメントのリスクを制限する効果的なアプリ セグメンテーション ポリシーを推奨することで、攻撃による潜在的な影響範囲を縮小します。例えば、財務アプリにアクセスできる30,000ユーザーのうち、実際にアクセスが必要なのは200人のみ、というケースは少なからずあります。Zscalerは、アクセスを200人の従業員のみに制限するアプリ セグメンテーションを自動的に作成し、脅威アクターによるラテラル ムーブメントのリスクを99%以上削減します。

ステージ4: データの持ち出し

攻撃の最終段階では、脅威アクターは機密情報を持ち出そうと動きます。ZscalerはAIを使用して、企業がデータ保護をより迅速に展開できるようにします。AIを活用したデータ検出により、展開を遅らせたり妨げたりする可能性がある、データのフィンガープリントの登録と分類という時間のかかる作業が不要になります。ZscalerのAIは、導入後すぐに企業全体のデータをすべて自動的に検出して分類するため、企業は機密情報を速やかに分類しながら、攻撃や侵害によってデータが企業から流出しないよう情報漏洩防止(DLP)ポリシーを構成することができます。

AIを活用したZscaler製品の概要

Zscaler Internet Access™はZero Trust Exchangeの一部として、あらゆる場所のユーザー、デバイス、Web、SaaSアプリにAIを活用した保護を適用し、次の機能を提供します。

- **Zscaler Secure Web Gateway (SWG) の AI ベースのインライン検知機能**により、新たなフィッシング サイトや C2 インフラを AI で検知します。
- AI 活用型サンドボックスで包括的なマルウェアとゼロデイ脅威対策を行います。
- **ユーザー、デバイス、アプリ、コンテンツのリスクを継続的に分析**することで、動的なセキュリティとリスクベースのアクセス ポリシーを実現します。
- **Zscaler Private Access™ が提供する AI 活用型のセグメンテーション**では、自動で推奨されるアクセス ポリシーにより、攻撃対象領域の最小化だけでなく、ユーザーのコンテキスト、振る舞い、場所、プライベート アプリのテレメトリーを使用したラテラル ムーブメントの阻止が可能になります。
- AI を活用したブラウザー分離は、ユーザーと悪意のある Web カテゴリーに分類されたサイトとの間に安全な緩衝ギャップを作成し、コンテンツを一連の画像としてレンダリングして、データの漏洩とアクティブな脅威の配信を阻止します。

ZSCALERは次の2つをブロックします。

Zscalerのクラウドやネイティブに統合されたオープン ソースと商用の脅威インテリジェンス ソースで確認された**URLとIP**。これには、新たに確認されたまたは新たにアクティブ化されたドメインなど、フィッシングによく使用される、ポリシーで定義されたリスクの高いURLカテゴリーが含まれます。

フィッシング キットやフィッシング ページをThreatLabzが分析し、開発した**IPSシグネチャー。**

Zscaler Risk360は、セキュリティ リーダーやビジネス リーダーが企業全体のサイバー リスクを定量化し、視覚化できるようにする総合的かつ実用的なリスク フレームワークを提供します。

DLPとCASBによるデータ保護は、エンドポイント、メール、ワークロード、BYOD、クラウド ポスチャーを含むすべてのチャネルにわたってAIを活用したデータ分類とデータ保護を提供します。

高度な脅威対策は、すべての既知のC2ドメインをブロックします。

Zscaler ITDR (アイデンティティー脅威の検知と対応)は、継続的な可視化、リスク監視、脅威検知なしで、IDベースの攻撃のリスクを軽減します。

Zscaler Firewallは、C2保護として新規宛先のC2通信も含めてポートやプロトコル番号にかかわらず検出、ブロックします。

DNSセキュリティは、DNSベースの攻撃や情報の持ち出しから保護します。

Zscaler Private Access™は、最小特権アクセス、ユーザーとアプリ間のセグメンテーション、プライベート アプリのトラフィックに対する完全なインライン検査でラテラル ムーブメントを制限して、アプリを保護します。

Zscaler Private AccessのAppProtectionは、高性能なインラインのセキュリティ検査をアプリのペイロード全体に実行して脅威を明らかにします。

Zscaler Deception™は、デコイのサーバー、アプリ、ディレクトリー、ユーザー アカウントで攻撃者をおびき寄せ、水平移動や権限昇格を試みる攻撃者を検知して封じ込めます。

AIへの移行を加速：適切な制御で自社を守る

Zscalerは、企業がAIアプリでイノベーション、創造性、生産性を促進しながら、データ流出の新たなリスクの中でユーザーとデータを安全に維持できるようサポートします。Zscalerを導入することで、企業はAIの変革力を最大限に取り入れながら、AIのアプリやドメインを完全にブロックすることなくビジネスを加速できるようになります。

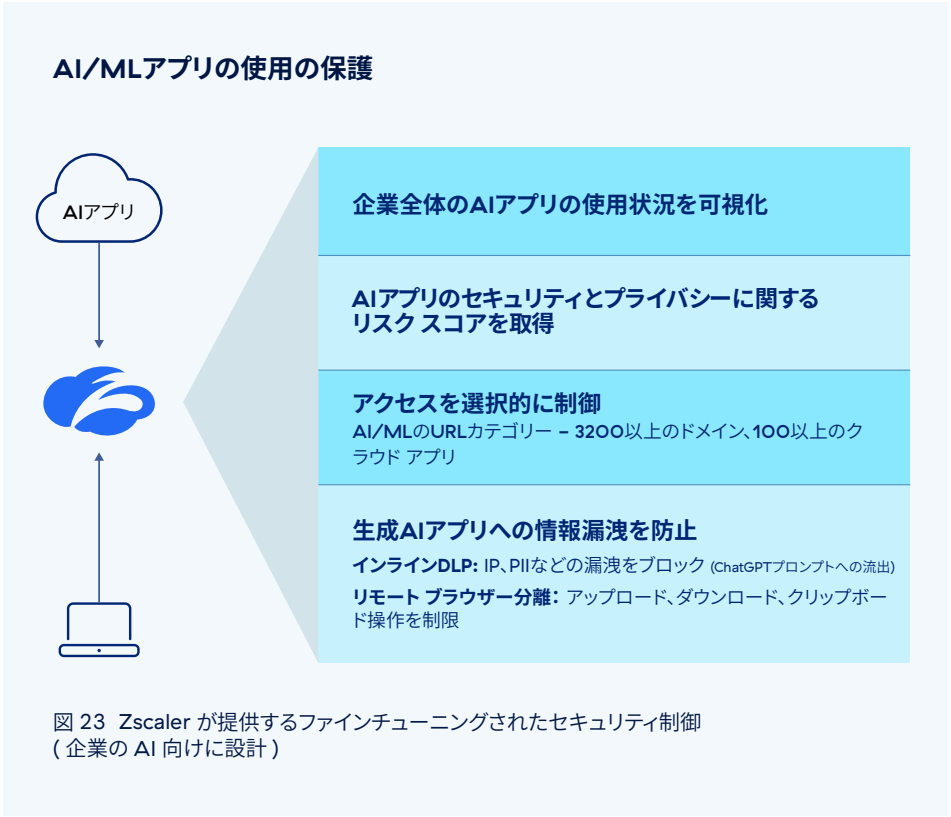


図 23　Zscaler が提供するファインチューニングされたセキュリティ制御
(企業の AI 向けに設計)

ZSCALERは、次の方法で企業をサポートします。

01 **AIツールの使用状況を完全に可視化**
詳細なログを活用して、各部門がアクセスしているアプリやドメイン、ChatGPTなどのツールで使用されているデータやプロンプトなど、AIの使用状況を完全に可視化します。

02 **AIの使用を微調整する柔軟なポリシーを策定**
AI/MLアプリ向けにカスタマイズされた強力なURLフィルタリングにより、きめ細かなAIアクセス制御とセグメンテーションを簡単に定義して施行できます。AIアプリのリスク スコアを取得して、許容範囲内のリスクでアクセスを許可しながら、必要に応じてアクセスをブロックします。企業全体、部門、部署、ユーザーごとにアクセスを許可できるだけでなく、生成AIツールのリスクをユーザーに指導する警告ベースのアクセスを設定することも可能です。AIを活用したセグメンテーションにより、AIツールに関連する内部の攻撃対象領域を最小限に抑えながら、特定のAIアプリにアクセスするための適切なユーザー セグメントを簡単に識別できます。

03 **ChatGPTなどのAIアプリにきめ細かなデータ セキュリティを施行**
生成AI用のきめ細かなZscaler Cloud Application制御を使用して、AIアプリにアップロードされた機密情報が漏洩しないようにします。Zscaler DLPエンジンを実行することで、AIツール使用時にデータが誤って共有されないようにできます。また、AIを活用したデータの検出と分類により、企業コード基盤、財務および法的文書、個人や顧客のデータなどを含む最も重要なデータに関するDLPポリシーを簡単に特定し、策定できるようになります。この動画では、ユーザーがChatGPTにクレジットカード情報を入力しようとした場合、DLPエンジンがどのように阻止するかを紹介しています。

04 **ブラウザー分離で強力に制御**
Zscaler Browser Isolationは安全な環境でAIアプリをレンダリングし、コピーや貼り付け、アップロード、ダウンロードを制限しながら、ユーザーからのプロンプトとAIツールへの質問を許可する保護層を追加します。これにより、機密情報が誤って生成AIツールに共有されるリスクを軽減できます。

今、多くの企業やセキュリティ リーダーが大きな岐路に立たされています。イノベーションを推進して競争力を維持するには、AIの積極的な導入が不可欠です。しかし、データはビジネスを推進するためだけに使用されるべきであり、侵害されるようなことがあってはならないのです。Zscalerは、ファインチューニングされたAIポリシーとデータ保護で生成AIの可能性を最大限引き出します。そして、企業がAIを活用したZscalerのゼロトラスト セキュリティを利用しながら、安心してAI導入を進められるようサポートします。

付録

ThreatLabzの調査方法

Zscalerのグローバル セキュリティ クラウドは、1日あたり300兆を超えるシグナルを処理し、90億の脅威とポリシー違反をブロックし、25万件以上のセキュリティ アップデートを提供しています。2023年4月～2024年1月にかけて、ZscalerのZero Trust Exchangeで確認された180億9000万件のAI/MLトランザクションを分析しました。

Zscaler ThreatLabzについて

ThreatLabzは、Zscalerが誇る世界トップクラスのセキュリティ調査部門であり、Zscalerのプラットフォームを使用する世界中の組織が常に保護された状態にあることを保証する責任を担います。ThreatLabzのメンバーは、マルウェアの調査や振る舞い分析に加え、Zscalerのプラットフォームの高度な脅威対策を実現するための新しいプロトタイプ モジュールの研究開発も進めています。また、定期的に社内のセキュリティ監査を実施して、Zscalerの製品とインフラがセキュリティ コンプライアンス基準を満たしていることを確認します。ThreatLabzは、新たな脅威に関する詳細な分析を定期的にポータル(research.zscaler.jp)で公開しています。

Experience your world, secured.

Zscaler について
Zscaler (NASDAQ: ZS) は、より効率的で、俊敏性や回復性に優れたセキュアなデジタル トランスフォーメーションを加速しています。Zscaler Zero Trust Exchange は、ユーザー、デバイス、アプリケーションをどこからでも安全に接続させることで、数多くのお客様をサイバー攻撃や情報漏洩から保護しています。世界 150 拠点以上のデータ センターに分散された SASE ベースの Zero Trust Exchange は、世界最大のインライン型クラウド セキュリティ プラットフォームです。詳細は、**www.zscaler.jp をご覧ください。**

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