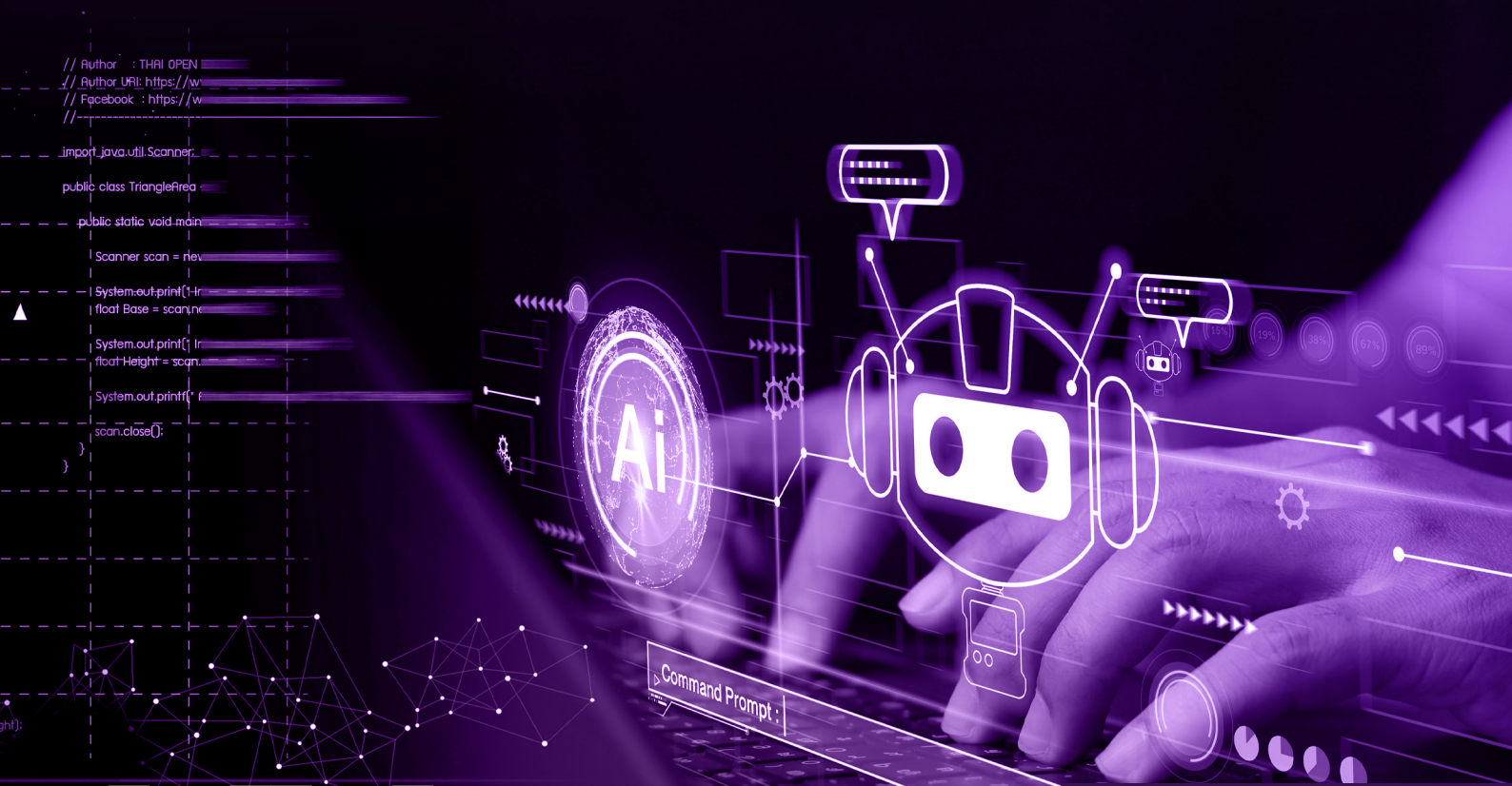
MENLO SECURITY

従業員による生成AIの使用がセキュリティ体制に与える影響とは

ChatGPTに代表される生成AIプラットフォームは、働き方を変革すると同時に、組織に新たなリスクをもたらしています



SNAPSHOT

生成AIは組織の生産性を向上させますが、それと同時にリスクももたらします

2022年11月に公開されたChatGPTは、IT史上最も急速に成長したプラットフォームの1つとなり、わずか2か月の間にユーザー数は1億人を超えました。同じ数のユーザーを獲得するのに、TikTokは9か月、Instagramは2年半かかりました。ChatGPTは世界中の人々の注目を集めましたが、これは日々利用されている多くの生成AIサイトのうちの1つにすぎません。これらの生成AIを活用することで、従業員が新しいアイデアを生み出し、メールを修正し、コンテンツを作成し、スペルや文法の間違いをチェックするなどができるようになり、生産性と革新性が向上しました。

生成AIの利用により大きなメリットが得られましたが、その一方で、特にサイバーセキュリティに関しては、大きな懸念をもたらしてもいます。ChatGPTは、いくつかの重要な側面でサイバーセキュリティの様相を変えました。多くの人々は、攻撃者がこれらのツールを使うことで、これまでに無い規模で高い回避性を持つ脅威を開発できるようになるのではないかと考えています。さらに、ChatGPTのようなプラットフォームにより、ハッカーがより高度で効果的なフィッシング攻撃を行う際の障壁が低くなったことも指摘されています。ChatGPTが世界中の組織や個人にもたらすこれらのリスクについて、サイバーセキュリティの専門家が世界へ向けて警告を発するのは当然のことです。しかしその一方で、これらの生成AIプラットフォームやチャットボットに潜む差し迫ったマイナスの側面、つまり機密データやその他の知的財産（IP）が流出する可能性を見逃してはなりません。

これは、従業員がChatGPTやBardなどの生成AIツールを利用する際に、企業の機密データを外部と共有したり、公開したりしてしまう可能性があるということです。これには、顧客データ、企業秘密、機密情報、さらには知的財産が含まれます。生成AIを利用した結果、他の一般的なデータ漏洩の場合よりもはるかに幅広いユーザーに対して、機密情報がばらまかれる可能性があるのです。ChatGPTなどの生成AIプラットフォームは、ユーザーとのチャット履歴などのデータを保存して、モデルをトレーニングしたり改善したりします。ユーザーが入力したデータがモデルのトレーニングに使用されるということは、後にその情報が他のユーザーに提供される可能性があるということを意味します。

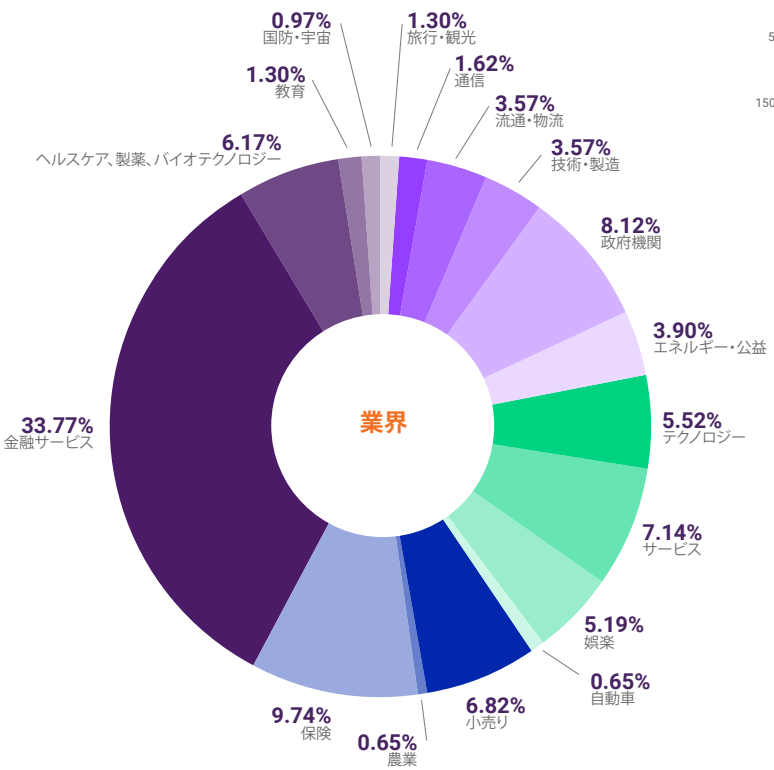
現実世界の例：

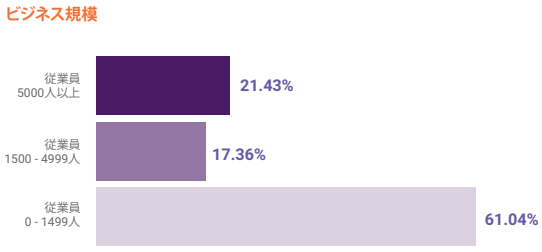
Samsungの半導体グループのエンジニアが、同社が開発中の新機能のコードをさらに効率化できないかを確認するために、ソースコードをChatGPTに入力したことが最近報告されました。ChatGPTやその他の生成AIツールは、入力データを自身の学習に使うため、入力されたSamsungのソースコードはAIの一部となり、他のユーザーからのリクエストに対する回答に含まれるかも知れません。このようなユーザーには、企業の脆弱性を探している攻撃者や、機密情報を探す競合他社が含まれます。

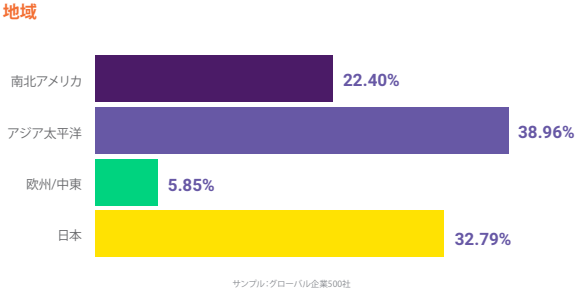
組織を保護するために、生成AIの利用を完全に禁止した企業もあります。イタリアはデータプライバシーへの懸念を理由に、一時的にChatGPTを全国で禁止しましたが、サービスは約1カ月後に再度利用可能になりました。生成AIサービスへのアクセスを禁止することは、潜在的なセキュリティリスクに対する解決策のように見えるかもしれませんが、それは長期的な解決策ではなく、その場しのぎの応急処置にすぎません。ChatGPTやその他の無数の生成AIプラットフォームは、ビジネスプロセスを合理化し、退屈なタスクを自動化し、執筆やデザイン、コーディングプロジェクトを有利に進めことができる、強力なビジネスツールです。これらのサイトを利用禁止にすると、ビジネスの生産性や俊敏性も妨げられてしまいます。

調査の概要

Menlo Securityは、生成AIの世界とそれがサイバーセキュリティに与える影響についての洞察を提供するために、**500社に上るグローバル企業からサンプルを収集**し、生成AIとどのようなインタラクションを行っているのかを分析しました。従業員が生成AIを利用する頻度と、生成AIがデータ漏洩に及ぼしている実際の影響に焦点を当て、このスナップショットを作成しました。







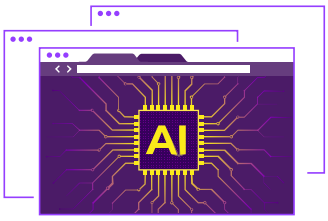
第1の洞察：

従業員による生成AIの利用は指数関数的な速度で増加している

生成AIを使用する従業員が増えるにつれて、組織が直面する潜在的なリスクも増加します。そして今、まさにそれが起きているのです。2022年11月から2023年5月にかけて生成AIの利用は1200%増加し、その値は日々増加しています。

そして、従業員が生成AIサイトを利用するたびに、データ漏洩の可能性が生じているのです。

過去30日間のデータ：

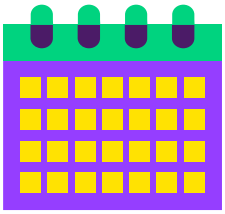


250万回

500社から生成AIサイトへアクセスした回数

78,825人

生成AIサイトを利用したこれらの企業の従業員の人数



32回

ユーザーが1ヶ月間にアクセスした回数

ユーザーはどのようなサイトを訪問しているのでしょうか?

現在利用できる生成AIのサイトは膨大な数に上りますが、このスナップショットでは、上位6つのAIアプリケーションに焦点を当てることにしました：
chat.openai.com、bing.com、bard.google.com、writesonic.com、copy.ai、jasper.ai

生成AIサイトのアクセス数トップ3は以下のとおりです：

OpenAI (ChatGPT)	1977607 アクセス
Microsoft (Bing)	410850 アクセス
Google (Bard)	96181 アクセス

過去30日間のデータによると、OpenAI は
生成AIサイトへのアクセスの50%以上を占めています

第2の洞察：

データ漏洩の経路は1つではない

従業員は、さまざまな方法で生成AIを利用しており、それらをすべて考慮に入れることが重要です。彼らは質問や疑問を検索バーに入力していませんか？ 別のソースから情報をコピーして貼り付けていませんか？ 質問と一緒にファイルをアップロードしていませんか？

ほとんどのユーザーは質問や疑問を入力するだけですが、他の2つのデータ漏洩経路は、大きな問題になる可能性があります。ファイルのアップロードやコピー&ペーストは、大量の機密データを以前よりもはるかに高速に公開してしまう可能性があります。例としては以下が挙げられます：

- ソースコード、顧客リスト、または計画ロードマップのコピー&ペースト
- 数百の列を含むスプレッドシートのアップロード

ファイルアップロード：10190 イベント（過去30日間）

コピー/ペースト：3394 イベント（過去30日間）

イベント数自体は生成AIへのアクセス数に比べて少ないものの、潜在的なデータ漏洩という観点からは、非常に大きな影響を持っています。

（一部の生成AIサイトにはネイティブファイルのアップロードオプションがないことに注意してください。しかしユーザーは、プラグインを使えばアップロードできます。ここでは、失敗した可能性のあるファイルのアップロード試行も含んでいます。）

過去30日間にアップロードされたファイルタイプ		
不明（テキストファイルなど）	1237	従業員は、さまざまな種類のファイルを生成AIプラットフォームにアップロードしています：「不明」の次に多いファイルタイプは「PDF」です。 さらに、Excelなどのファイルタイプでは、意図しないデータ漏洩が発生する可能性があります。本人はファイルは安全であると考えているかもしれませんが、非表示の行や列に機密データが含まれている可能性があります。
PDF	133	
Word	48	
Excel	22	
PowerPoint	21	
スクリプト	13	
WinEXE	3	
ZIP	2	

第3の洞察：

生成AIへのデータ漏洩の脅威は現実となっている

データ漏洩防止（DLP）という課題は、サイバーセキュリティの世界にとって新しい概念やテクノロジーではありません。データが組織の制御外に移動することは、潜在的な影響を引き起こす可能性を示しているのです。そのため、あらゆる業種および規模の組織にとって、データ漏洩を制限および/またはブロックすることは常に大きな課題であり、継続的に取り組む必要があります。生成AIの登場により、懸念すべき事項がさらに増えました。

実際に生成AIサイトに送信されている機密データの量を見れば、ソースコード問題は公表されているものをはるかに超えていることがわかります。

以下の表は、潜在的なデータ漏洩インシデントが組織で定期的に発生していることを示しています。従業員は生成AIを使用するために、意図的か否かに関わらず、機密データや社内データを入力またはアップロードしています。

従業員が生成AIに入力しているデータの種類		
PCI	5.4%	私たちは、従業員が機密情報を生成AIプラットフォームに入力しようとする頻度を分析しました。過去30日間に、これらのカテゴリに関連するデータを含むDLP関連のイベントがありました： 漏洩の可能性が最も大きかったのは、個人を特定できる情報（PII：Personally Identifiable Information）でした。 これらの組織には、これらのインスタンスをブロックするポリシーが適用されています。
PII	50.4%	
機密のドキュメント	24.6%	
医療情報	2.2%	
制限された情報	1.5%	
その他	15.9%	

ソリューション

生成AIを安全に利用してデータ漏洩のリスクを適切に管理する

この新しいテクノロジーには注意深い対応が必要になるため、既存のデータ漏洩防止（DLP）やCloud Access Security Broker（CASB）、およびその他の内部脅威防止ソリューションでは残念ながら不十分です。これらのソリューションは検知と対応のアプローチを使用して、組織の外部へ向けた膨大な量のトラフィックの中からキーワードや文言を探します。これらはセキュリティ専門家やプロダクトオーナーが設定する必要がありますが、そのプロセスの多くは手作業となるため、すべてを見つけ出すことはほとんど不可能です。たとえソリューションがデータ漏洩を検知できたとしても、その時にはすでに手遅れになっている可能性もあります。情報が出て行ってしまった後にそれを元に戻す「元に戻す」ボタンは存在しないのです。入力された情報は生成AIプラットフォーム内に永久に組み込まれ、問い合わせへの回答のために情報を提供し続けます。

組織は、これらの生成AIプラットフォームやチャットボットに機密情報が入力されることを防ぐ必要がありますが、それは従業員によるこれらの便利なツールの使用を妨げない方法で行う必要があります。DLPを使えば簡単ですが、組織には一括管理のソリューションではなく、多階層的なアプローチが必要です。

まずできることは、入力する文字数を制限したり既知のコードをブロックしたりして、入力フィールドに貼り付けることができる情報を制限することです。何千行ものソースコードを手動で入力するユーザーはいません。貼り付け機能を制限することで、この種のデータ漏洩を効果的に防ぐことができます。また、ユーザーは入力しようとしている情報についてよく考えるようになります。

さらに重要なことは、組織はWebブラウザーから離れた場所で、ChatGPTやその他の生成AIプラットフォームとの対話を制限する必要があるということです。クラウド内のリモートブラウザーでアプリのコマンドを実行すれば、ユーザーとインターネットの間に追加の保護層が形成されることになり、組織は（意図的であるかどうかに関わらず）データ漏洩が発生する前に悪意のあるアクティビティを阻止する機会を得ることができます。

また、イベントログの記録やブラウザー記録の開始など、追加のセキュリティ制御をトリガーするセキュリティポリシーを適用して、解決やイベント後の分析を支援することもできます。内部関係者による侵害の調査では、その行動が意図したものであることを示す証拠が必要になることを覚えておいてください。イベントとブラウザーの記録セッションにより、ユーザーが悪意を持っていたのか、単なる不注意だったのかを可視化し、洞察を得ることができます。

© 2023 Menlo Security, All Rights Reserved.
www.menlosecurity.com

Menlo Securityについて

Menlo Securityは、Web、ドキュメント、メールからマルウェアの脅威を排除することで、サイバー攻撃から組織を保護します。これは、ナレッジワーカーの生産性を最大化するための鍵であるWebブラウザーを保護することに重点を置いています。

Menlo Securityのクラウドセキュリティプラットフォームは、単一のグローバルなクラウドベースの製品で、組織への脅威の侵入を防ぎ、データとアプリケーションへのアクセスを保護します。Menlo SecurityのElastic Isolation Core™は、ユーザー、コンテンツ、アプリケーションを分離し、セキュリティ、ポリシー、可視性を適用します。ブラウザー内部の詳細な可視化と適応型のポリシーにより、脅威を検知して対応するのではなく、脅威が発生する前に阻止できるため、組織はWeb、メール、SaaSアプリケーション、プライベートアプリケーションを狙うHEAT（Highly Evasive Adaptive Threats：検知回避型脅威）を含むすべての脅威を排除できます。

HEATcheck

Menlo Securityは、様々なHEAT攻撃に対する組織の耐性をより良く理解するために、軽量のペネトレーションアセスメントを提供しています。このアセスメントでは、攻撃者が現実に使用している様々なHEATの手法を使い、組織がそれらにどれくらい影響されやすいかを安全に推定することができます。Menlo SecurityのHEATcheckツールは、実際に悪意のあるコンテンツを配信することはありません。

ご連絡ください

皆様の組織が現在これらの主要なWeb脅威の影響を受けやすいかどうかを知り、そもそもそういった脅威が起こらないようにするにはどうしたら良いかを知りたい場合は、今すぐお問い合わせください。