MS&AD InterRisk Report

No.26-024

2026.08.03

サイバーセキュリティニュース<2026 年度第 1 号>

本誌では、サイバーセキュリティに関する国内外の最新動向を踏まえ、重要なトピックスを四半期ごとにわかりやすく紹介します。
※当社オフィシャルサイト「RM NAVI」に掲載したコラム記事を再構成し、取りまとめたものです。

EvilTokens で拡大するデバイスコードフィッシング
― 正規認証を悪用する攻撃の実態と対策 ―

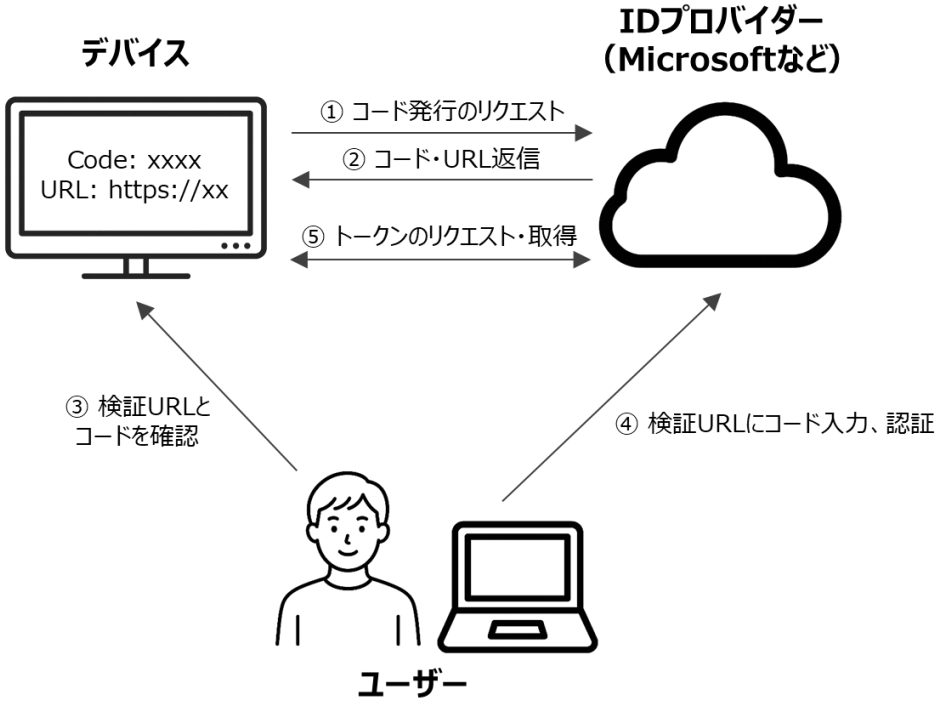
2026 年 4 月 28 日掲載

■はじめに

デバイスコードフィッシングという攻撃手法が、ここ最近増加しているとの報告がある。多要素認証を回避する手法として 2023 年ごろから注目を集めてきたが、Push Security 社によると、2026 年に入ってから検出されたデバイスコードフィッシングページは、前年同期比 37.5 倍に達したという（※1）。さらに同年 2 月には「EvilTokens」と呼ばれる PhaaS（※2）が登場し、専門知識の乏しい攻撃者でも、フィッシング攻撃を大量かつ効率的に実行できる環境が整いつつある。本稿ではデバイスコードフィッシングの手口や特徴、対策について解説する。

■１．デバイスフロー

デバイスコードフィッシングは、OAuth 2.0 の拡張仕様であるデバイスフロー（OAuth 2.0 Device Authorization Grant）を悪用した攻撃である。デバイスフローは本来、スマートテレビや IoT 機器など、ブラウザや文字入力に制限があるデバイス向けに設計された認証方式である。利用の仕組みは以下のとおり。



【図１】 デバイスフローの例（MS&AD インターリスク総研にて作成）

【デバイスフロー】

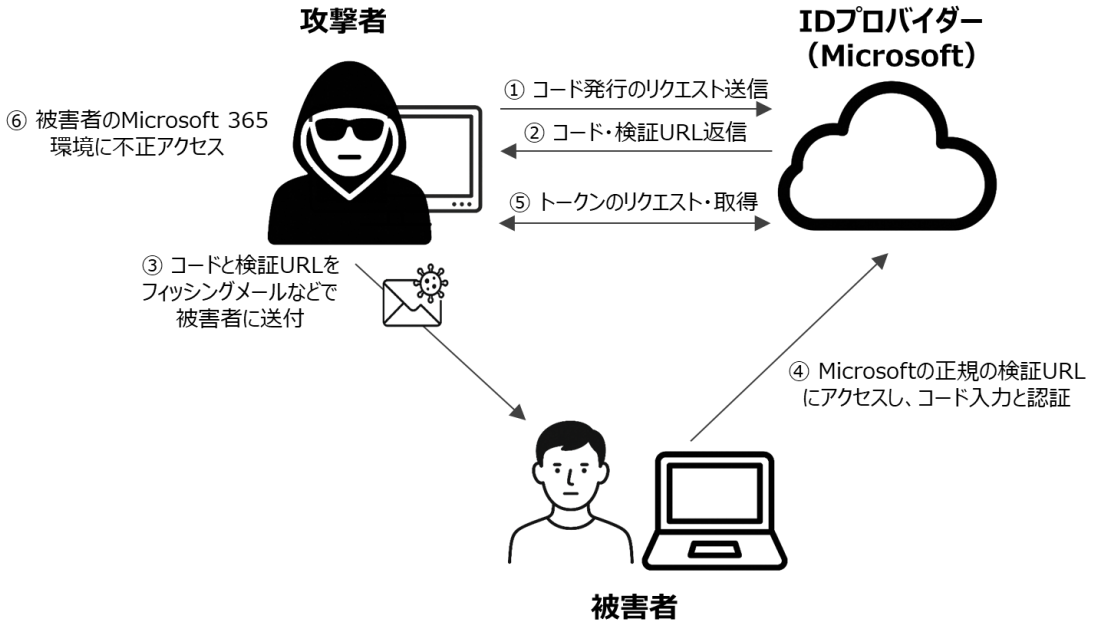
1. デバイス（例：スマートテレビ）から ID プロバイダーにリクエストを送信する
2. デバイスが ID プロバイダーからデバイスコードと検証用 URL を受け取る
3. ユーザーはデバイスに表示された検証用 URL とコードを確認する
4. ユーザーはスマートフォンや PC など別のデバイスのブラウザで、指定された検証用 URL にアクセスし、デバイスコードを入力して認証を完了する
5. 認証完了後、デバイスにアクセストークンとリフレッシュトークンが発行される

Microsoft のほか、Google、Salesforce、GitHub、AWS など多くの主要クラウドサービスがこのフローに対応している。

■２．デバイスコードフィッシングの攻撃の手口

攻撃者はデバイスフローを悪用し、正規の認証基盤を介して被害者のアクセストークンを詐取する。攻撃に正規の ID プロバイダー（Microsoft）が使われるため、被害者が不審に思わず処理を進めてしまう可能性がある。

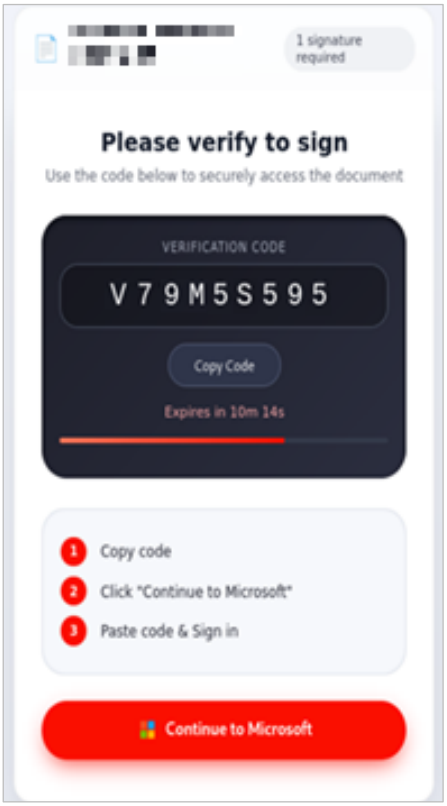
以下は、実際に悪用されている Microsoft のデバイスフローを利用したデバイスコードフィッシングの流れである。



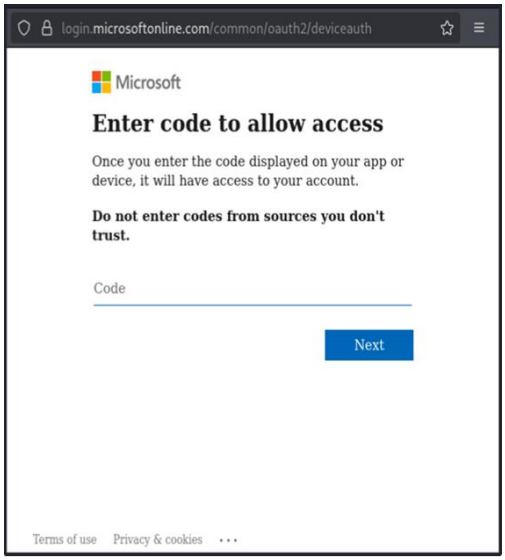
【図２】 デバイスコードフィッシングの例（MS&AD インターリスク総研にて作成）

【攻撃のプロセス】

1. 攻撃者は Microsoft に対して自身のデバイスを登録するためのデバイスコードと検証用 URL をリクエストする
2. 攻撃者は Microsoft からデバイスコードと検証用 URL を受け取る
3. 攻撃者はそのコードを記載したフィッシングメール（図 3 のような Microsoft と Adobe の連携を装った URL にアクセスさせるメール）を被害者に送付する
4. 被害者は、図 4 のような Microsoft の正規ドメインにアクセス後、デバイスコードを入力する。被害者は自身のアカウントを選択し、多要素認証などの認証を完了する。認証は正規の Microsoft サーバー上で行われるため、被害者は不正操作の被害を受けていることに気づかない
5. 攻撃者は Microsoft にリクエストし、トークン（アクセストークンとリフレッシュトークン）を取得する
6. 攻撃者は取得したトークンを使い、被害者の Microsoft 365 環境に継続的にアクセスする。リフレッシュトークンを定期的に更新することで、最大 90 日間にわたり不正アクセスを維持できる
7. この攻撃の特徴は、被害者が正規の Microsoft のページで実際に多要素認証を完了させることにある。被害者自身が正規サイトで操作するため、FIDO2（※3）パスキー（※4）など、フィッシング耐性が高いとされる認証手段であっても、正規サイト上で認証させることで回避される可能性がある。



【図３】 Microsoft と Adobe の連携に偽装したトークン表示画面の例
（MS&AD インターリスク総研が加工して作成）
※「Continue to Microsoft」ボタンを押すと Microsoft の正規認証画面に遷移する



【図４】デバイスコードの正規認証画面の例（MS&AD インターリスク総研が加工して作成）

■３．攻撃を容易にする EvilTokens

2026 年 2 月中旬に登場した EvilTokens は、デバイスコードフィッシングを PhaaS で提供する攻撃ツールキットである。Sekoia 社によると、同ツールは Telegram（※5）を通じて販売されており、専門知識の乏しい攻撃者でもフィッシング攻撃を大量かつ効率的に実行できる環境を提供している。

EvilTokens の主な特徴は以下のとおりである。

- BEC（ビジネスメール詐欺）専用ウェブメールクライアント：Outlook の UI を模倣しており、侵害したアカウントからのメール送受信・添付ファイルのダウンロードが可能である
- AI による自動化：被害者のメールボックスへの不正アクセス後、AI を使ったメール内容の分析・要約、財務関連メールスレッドの自動識別、BEC メールの下書き自動生成といった機能を備えている
- キーワードアラート機能：「振込依頼」「口座番号」「送金」などのキーワードを設定することで、侵害した全アカウントのメールをリアルタイムに監視し、合致した場合に攻撃者の Telegram へ即時通知が届く
- 組織情報の収集：対象組織内の情報（部署、役職、連絡先など）を洗い出し、攻撃の標的になりうる経営幹部・経理担当者を特定する

このように PhaaS で提供されることで、デバイスコードフィッシングの実行ハードルは大幅に低下したとみられる。実際、Sekoia 社の調査では、2026 年 3 月 19 日時点で、EvilTokens の招待制グループには約 280 人の利用者（攻撃者）が参加していたことが確認されている（※6）。

EvilTokens の運営者は、現時点では Microsoft に限定されているフィッシングページを、今後 Gmail および Okta に対応させる計画を公表している（※7）。これが実現した場合、攻撃対象は企業の主要なクラウド認証基盤全体へと、さらに拡大する恐れがある。

■４．対策方法

対策方法としては以下のような対策が考えられる。

- Microsoft Entra ID で、可能な限りデバイスフローを禁止する。難しい場合は、条件付きアクセスポリシーでデバイスフローを制限する（※8）。
- Microsoft Entra ID の監査ログを定期的にレビューし、異常なサインイン履歴を確認する。
- 従業員には、デバイスコードの入力が必要な正規の業務シナリオを明示し、それ以外は拒否するよう周知する。

■５．まとめ

デバイスコードフィッシングは多要素認証を回避できる恐れがある。さらに、EvilTokens の登場や AI 機能の搭載により、言語の壁を越えた大規模なフィッシング攻撃が展開される可能性も高まっている。

本稿が、こうした攻撃手口への理解を深め、各社におけるセキュリティ対策の強化につながる一助となれば幸いである。

1. Push Security 社「Device code phishing attacks have skyrocketed: here's what you need to know」<https://pushsecurity.com/blog/device-code-phishing>（最終アクセス 2026 年 4 月 13 日）
2. Phishing as a Service。フィッシング攻撃を容易に実行するための仕組みやツールをサービスとして攻撃者に提供するもの
3. パスワード不要で生体認証やセキュリティキーを使ってログインする認証規格
4. スマートフォンや PC で生体認証や暗証番号を使って、パスワードなしで本人確認してログインできる FIDO2 をベースにした仕組み
5. 高い匿名性と秘匿性が特徴の無料チャットであり、サイバー攻撃者間の情報交換などに広く使用される
6. Sekoia 社「EvilTokens: an AI-augmented Phishing-as-a-Service for automating BEC fraud - Part 2」<https://blog.sekoia.io/eviltokens-an-ai-augmented-phishing-as-a-service-for-automating-bec-fraud-part-2>（最終アクセス 2026 年 4 月 13 日）
7. Sekoia 社「New widespread EvilTokens kit: device code phishing as-a-service - Part 1」<https://blog.sekoia.io/new-widespread-eviltokens-kit-device-code-phishing-as-a-service-part-1>（最終アクセス 2026 年 4 月 13 日）
8. Microsoft 社「Storm-2372 conducts device code phishing campaign - Mitigation and protection guidance」<https://www.microsoft.com/en-us/security/blog/2025/02/13/storm-2372-conducts-device-code-phishing-campaign/#mitigation-and-protection-guidance>（最終アクセス 2026 年 4 月 13 日）

フロンティア AI が変えるサイバーセキュリティの前提

2026 年 6 月 29 日掲載

■はじめに

Claude Mythos Preview をはじめとするフロンティア AI の登場により、サイバーセキュリティを取り巻く環境は変化しつつある。脆弱性の発見速度は飛躍的に向上し、攻撃の裾野は広がりを見せている。本稿では、Claude Mythos Preview の発表から現在（2026 年 6 月 18 日）までの約 2 か月間の動向を整理したうえで、フロンティア AI が変えたもの、そして企業に求められる対応方法を示す。

■Mythos 発表から現在までの主な動向

2026 年 4 月の Claude Mythos Preview 発表を起点に、フロンティア AI のサイバー分野での能力評価や、それに対する各国・各機関の対応は約 2 か月で急展開した。評価機関による能力検証、日本政府・金融当局による異例の速さでの対応要請、そして米政府による輸出規制まで、AI とサイバーセキュリティの問題は技術面にとどまらず、規制や政策にも広がった。

【表 1】2026 年 4 月 7 日から 6 月 18 日までのフロンティア AI の動向

日付	内容・説明	出典 URL
2026 年 4 月 7 日	Anthropic が Claude Mythos Preview を発表。ゼロデイ脆弱性の自律的発見・悪用能力を持つとされたが、安全上の懸念から一般公開は見送られ、Anthropic が立ち上げたセキュリティプログラム「Project Glasswing」経由のみで提供。主要 OS・ブラウザからのゼロデイ脆弱性発見・攻撃手法の確立に成功したと報告。	https://red.anthropic.com/2026/mythos-preview/
2026 年 4 月 13 日	英国 AISI が評価結果を公表。Mythos Preview が従来モデルを上回るサイバー能力を持つと報告。2025 年 4 月以前はどのモデルも解けなかったテストで、Mythos は 73%の成功率を記録。	https://www.aisi.gov.uk/blog/our-evaluation-of-claude-mythos-previews-cyber-capabilities
2026 年 4 月 21 日	Mozilla の評価で Firefox から 271 件の新規脆弱性を発見。Mozilla は「世界最高レベルのセキュリティ研究者と同等の能力を持っている」と言及し、強い危機感を表明。	https://blog.mozilla.org/en/privacy-security/ai-security-zero-day-vulnerabilities/
2026 年 4 月 24 日	金融庁が「AI 脅威に対する金融分野のサイバーセキュリティ対策強化に関する官民連携会議」を開催。金融業界・IT 事業者・日本銀行を集め対応を協議。	https://www.fsa.go.jp/news/r7/sonota/20260514/20260514.html
2026 年 4 月 30 日	AISI が GPT-5.5 のサイバー能力評価を公表。Mythos Preview と同等水準であることを報告。	https://www.aisi.gov.uk/blog/our-evaluation-of-openais-gpt-5-5-cyber-capabilities
2026 年 5 月 18 日	内閣官房・警察庁・金融庁・デジタル庁・防衛省など 14 省庁が連名で AI サイバーセキュリティ対策パッケージ「YATA-Shield」を公表。	https://www.cyber.go.jp/pdf/press/20260518_AI_CS_Package.pdf
2026 年 5 月 22 日	金融庁・日本銀行が全金融機関へ「フロンティア AI による脅威変化を踏まえた金融機関等の短期的な対応」を連名で発表。経営層の直接関与のもと、おおむね 1 か月程度を目途に対策実施を要請。	https://www.fsa.go.jp/news/r7/sonota/20260522-5/01.pdf
2026 年 6 月 9 日	Anthropic が Claude Fable 5 および Claude Mythos 5 を発表。Fable 5 はサイバーや生物・化学分野へのガードレール付きで一般公開。Mythos 5 は Project Glasswing 参加組織への限定提供。	https://www.anthropic.com/news/claude-fable-5-mythos-5

日付	内容・説明	出典 URL
2026 年 6 月 12 日	米政府の輸出規制指令を受け、Anthropic が Fable 5・Mythos 5 へのアクセスを即時全停止。ジェイルブレイク手法の存在を政府が把握したことが理由とされている。	https://www.anthropic.com/news/fable-mythos-access

■フロンティア AI で何が変わったのか

前章で整理したとおり、フロンティア AI を巡る動向は約 2 か月で急展開した。なぜこれほどの懸念が急速に高まったのか。以下では、フロンティア AI が実際に何を変えたのかを整理する。

これまでサイバー攻撃の糸口となる脆弱性（システムの欠陥）を発見するには、高度な専門知識を持つセキュリティ研究者が長時間かけて調査する必要があった。しかし、フロンティア AI の登場により、この前提が崩れつつある。Mozilla の評価では、Firefox から 271 件の新規脆弱性が発見されている（※1）。脆弱性を調査した期間は報告されていないものの、Mythos Preview の発表から Mozilla の報告までの期間を考えると、2 週間以内という短期間で 271 件もの脆弱性を発見したと推測される。なお、同社は「世界最高レベルのセキュリティ研究者と同等の能力を持っている」と言及しており、AI が人間の専門家に匹敵する水準に達しつつあることを示している。

ここで注目すべきは、「AI が脆弱性を見つけた」という事実そのものではない。より本質的なのは、脆弱性発見に必要な時間とコストの水準が大きく変わりつつある点である。これまで脆弱性の発見には、高度な専門人材と、数週間から数か月単位の時間、そして相応のコストが必要であった。それが AI の活用によって、より短時間で低コストに行えるようになりつつある。脆弱性発見のコストと時間が下がることで、攻撃者が悪用できる脆弱性の数は増え、発見から攻撃までの時間は短くなる。フロンティア AI の登場が単なる「性能向上」ではなく、セキュリティの前提を変える出来事として受け止められている理由はここにある。

さらに重要なのは、こうした能力が Mythos Preview 固有のものではないという点である。英国 AISI の評価では、OpenAI の GPT-5.5 も Mythos Preview と同等水準の脆弱性発見や悪用支援の能力を持つと報告されている（※2）。特定モデルの問題ではなく AI 全体の底上げが進んでいる。

加えて、高額な最先端モデルを使わなくとも同様のことが可能になりつつある。一般向けに公開されているモデルを用いて、個人の研究者が Windows の月例セキュリティ更新から攻撃コード（Exploit）を半自動で生成するツールを 300 ドル程度のコストで構築・公開した事例がある（※3）。この事例が示すのは、AI による脅威の大きさはモデルの性能だけでは決まらないということである。ツールやモデルを組み合わせて目的の処理を行わせる設計力（ハーネスエンジニアリング）によって結果は大きく変わる。かつては高度な専門知識を持つ攻撃者にしか実現できなかった攻撃手法が、設計力さえあれば一般に入手可能な AI で再現できる段階に来ている。

こうした AI による脆弱性の発見は、攻撃者だけが使うわけではなく、脆弱性を探して報告する研究者にも広く使われている。その結果、脆弱性の報告件数が爆発的に増加し、対応する側が追いつけない状況が生まれている。この背景には「AI Slop（低品質な AI 生成報告の大量発生）」とも呼ばれる問題もある。

Linux の開発者である Linus Torvalds 氏は、AI を使ったバグハンターの急増により、Linux のセキュリティに関する開発者間のメーリングリストが「ほぼ管理不能」な状態になったと発言している（※4）。また、脆弱性の報奨金制度（バグバウンティ）でも同様の問題が表面化しており、AI を使った報告が殺到する一方、審査・対応する人員は増えず、制度が機能不全に陥りつつある（※5）。

重要なのは、問題は「AI が脆弱性を見つけすぎる」ことではなく、見つかった脆弱性を修正・対処する人間側の能力が追いついていない点にある。課題の中心は、脆弱性を見つける段階から、修正し、優先順位を付けて対応する段階へ移行しつつある。

一方で、AI が大量に発見した脆弱性が直ちに企業への侵害につながるとは限らない。攻撃を成功させるには前提条件が必要な場合もある。攻撃者は脆弱性の悪用に加え、設定不備の悪用やソーシャルエンジニアリングなどを組み合わせ、検知を回避しつつ侵入を進める。

AI 時代においては、脆弱性の報告件数がこれまでとは比較にならない規模で増加することが予測される。すべての脆弱性に均等に対応しようとすれば、対応に必要な人員や時間はすぐに不足する。企業に求められるのは、自組織のシステム構成や業務上の重要度を踏まえ、悪用された場合の影響度と実際に悪用される可能性の両面からリスクを評価し、優先度を付けて対応することである。

■企業に求められる対応方法

米国サイバーセキュリティ・インフラセキュリティ庁（CISA）は、従来の「パッチが公開されたら即時適用」という方針から、脆弱性ごとのリスクで優先度をつける方式への転換を表明している（※6）。全件対応から重点対応への移行は、企業のリスク管理においても参考になる考え方である。

国内においては、金融庁・日本銀行が 5 月 22 日付で全金融機関におおむね 1 か月程度をめどに対策を実施するよう要請している（※7）。経営トップの直接関与を前提とした要請であり、セキュリティ対応を担当部門だけの問題として扱えない状況になっている。金融機関以外の企業においても、同様の視点で体制を見直すことが求められている。

これらの指針を参考に、ここでは、対応の要点を次の 3 点に絞って示す。

１．外部に露出した攻撃面（アタックサーフェス）の特定と縮小

攻撃者が最初に狙うのは、インターネットから直接アクセスできるシステムや機器である。自組織のどの資産が外部に公開されているかを把握できていない企業は、AI による自動スキャンの標的になりやすい。まず自組織の外部公開資産を網羅的に洗い出し、業務上不要なものは速やかに閉鎖・遮断することが第一の優先事項である。攻撃者に狙われやすい領域を減らすことが、リスクの低減につながる。

また、メールやブラウザなど外部と直接やり取りするソフトウェアについても、自社で使用しているバージョンや製品を正確に把握しておくことが重要である。これらは脆弱性が発見された際に攻撃者が悪用を試みる対象となるため、パッチ適用の対象を即座に特定できる状態を整えておく必要がある。

２．悪用実績に基づく脆弱性対応の優先順位付け

今後、発見された脆弱性すべてに即時対応することは現実的ではない。対応の優先基準として有効なのが、米国 CISA が公表する KEV（Known Exploited Vulnerabilities：実際に悪用が確認されている脆弱性の一覧）である。理論上のリスクにとどまらず、攻撃者に実際に使われている脆弱性を示すリストである。ここに掲載された脆弱性への対応を最優先とすることで、限られたリソースを効果的に配分できる。

ただし、日本国内のみで使用されているソフトウェアは KEV には掲載されないことがあるため、JPCERT（※8）や IPA（※9）が発信する注意喚起も合わせて参照することを推奨する。

３．脆弱性管理プロセスの見直し

AI 時代において、脆弱性の「発見」は加速することが予想される。問題は、その後の評価、修正、適用の工程にある。発見された脆弱性は、報告、内容確認、修正方針の決定、テスト、本番環境への適用という段階を経る。この一連の対応には、多くの組織で数週間から数か月かかる。この間が攻撃者にとっての好機となる。自組織の脆弱性対応フローを改めて点検し、どの工程に時間がかかっているかを把握することが急務である。承認プロセスの簡略化、テスト環境の整備、対応人員の確保など、ボトルネックの解消に向けた具体的な対策を講じることが求められる。

また、パッチが存在しないゼロデイ脆弱性の悪用に備え、検知・対応体制（EDR、SIEM など）の見直しと強化も併せて進める必要がある。

■まとめ

フロンティア AI がもたらした最大の変化のひとつは、脆弱性発見能力の向上である。しかし、より重要な問題は、攻撃側のコストが下がり続ける一方で、防御側の対応リソースが追いつかないという構造的な非対称性にある。すべての脆弱性に均等に対応しようとするのではなく、自組織への影響度と悪用可能性に基づいて優先度を判断することが、現実的なリスク管理の出発点となる。

AI 時代においては、脆弱性の報告件数の増加と修正対応の負荷増大は今後も続くと予測される。企業は外部公開資産の継続的な把握、優先度に基づくパッチ適用体制の整備、検知・対応能力の強化を軸に、体制を継続的に見直していくことを求められる。

1. Mozilla Foundation 「The zero-days are numbered」<https://blog.mozilla.org/en/privacy-security/ai-security-zero-day-vulnerabilities/>（最終アクセス 2026 年 6 月 18 日）
2. AI Security Institute 「Our evaluation of OpenAI's GPT-5.5 cyber capabilities」<https://www.aisi.gov.uk/blog/our-evaluation-of-openais-gpt-5-5-cyber-capabilities>（最終アクセス 2026 年 6 月 18 日）
3. Origin 社 「The Mythos We Have At Home: A Patch-Diffing Pipeline for N-Day Generation」<https://www.originhq.com/research/patch-diffing-pipeline>（最終アクセス 2026 年 6 月 18 日）
4. The Register 「Linus Torvalds says AI-powered bug hunters have made Linux security mailing list 'almost entirely unmanageable'」<https://www.theregister.com/security/2026/05/18/linus-torvalds-says-ai-powered-bug-hunters-have-made-linux-security-mailing-list-almost-entirely-unmanageable/5241633>（最終アクセス 2026 年 6 月 18 日）
5. The Register 「HackerOne takes an axe to its bug bounty rewards」<https://www.theregister.com/security/2026/05/21/hackerone-takes-an-axe-to-its-bug-bounty-rewards/5244458>（最終アクセス 2026 年 6 月 18 日）
6. CISA 「BOD 26-04: Prioritizing Security Updates Based on Risk」<https://www.cisa.gov/news-events/directives/bod-26-04-prioritizing-security-updates-based-risk>（最終アクセス 2026 年 6 月 18 日）
7. 金融庁『「フロンティア AI による脅威変化を踏まえた金融機関等の短期的な対応」に係る要請について』<https://www.fsa.go.jp/news/r7/sonota/20260522-5/01.pdf>（最終アクセス 2026 年 6 月 18 日）
8. JPCERT「注意喚起」<https://www.jpcert.or.jp/at/2026.html>（最終アクセス 2026 年 6 月 18 日）
9. IPA 「重要なセキュリティ情報」<https://www.ipa.go.jp/security/security-alert/index.html>（最終アクセス 2026 年 6 月 18 日）

サイバーインシデント発生時の情報開示
― 初報・続報・記者会見の判断と留意点 ―

2026 年 7 月 28 日掲載

■１．はじめに

サイバー攻撃の増加に伴い、プレスリリースや記者会見などでサイバー攻撃の事実や被害、対応状況を公表するケースが増加している。この際、被害公表の位置づけを誤ると、企業の危機管理にとって悪影響を及ぼすおそれがある。

また、企業のサイバーセキュリティに関する情報開示には大きく分けて有事のサイバーインシデントに関するものと平時のリスク対策状況に関するものとの2種類がある。同じ情報開示でも、意味合いや留意点は大きく異なる。

本レポートでは有事の情報開示を中心に、その意味合いや開示タイミングごとの留意点を整理する。

■２．有事の情報開示の意義

サイバー攻撃の被害公表は、被害拡大を防ぎ、信頼を維持するための重要な危機対応である。

第一に、顧客・取引先・株主・従業員などの関係者が、自らの安全確保や追加対応を取れるようにするためである。たとえば、個人情報流出の可能性があれば、パスワード変更や不審メールへの警戒を促す必要がある。あわせて、提供サービスの停止などを顧客に案内し、混乱を避ける狙いもある。第二に、法令・規制・契約上の報告義務を果たすためである。第三に、自社の信用維持のためである。事実を隠したまま後から発覚すると、被害そのもの以上に「隠蔽した」という印象が信頼失墜を招く。迅速で誠実な公表は、組織の危機管理能力を示す手段でもある。

なお、公表の目的は「謝罪」だけではなく、何が起き、何に注意すべきか、次にどう対応するかを社会に伝えることにある。

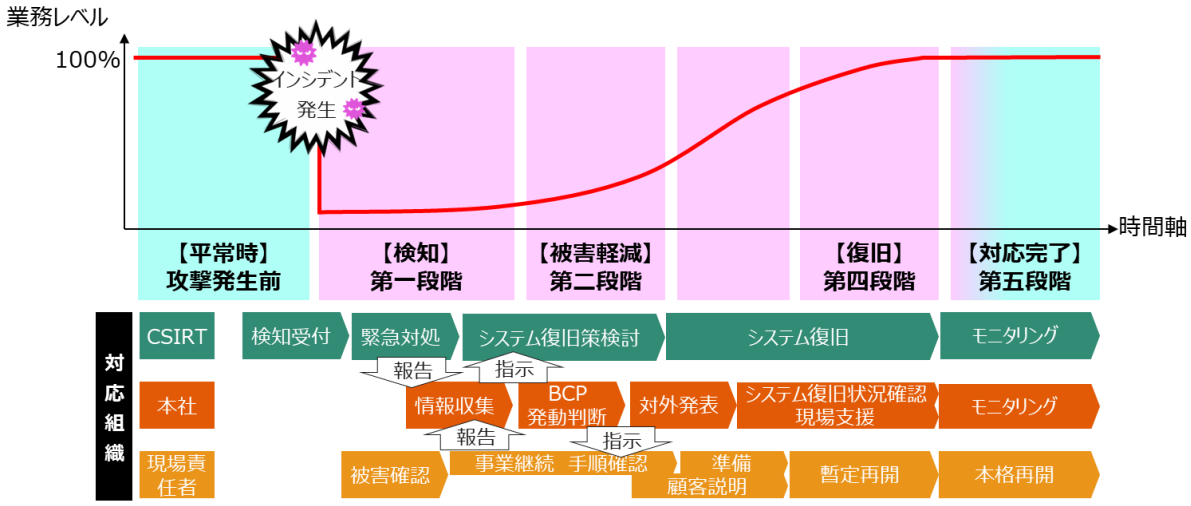
【表１】サイバー攻撃の対外公表の目的

位置づけ	目的
被害の公表	サイバー攻撃被害があったことや、どのようなインシデント対応を行ったのかを、被害組織が情報発信する。 法令・ガイドラインに基づく公表のほか、法令などの定めはないものの、被害組織自身の判断で行う公表もある。後者の判断に至る事情はさまざまで、例えば、以下のようなものがある。 ・　二次被害拡大防止のため注意喚起として行う公表 ・　サービス停止時など、対外的な説明を行う場合の速報的な公表 ・　広報・法務上のリスク対応としての公表
情報共有	被害組織で見つかった攻撃技術情報を中心に、非公開の方法で情報共有活動参加組織と共有し、フィードバックを得ることや、早期警戒、共助を目的とする。 未公表の個別攻撃を特定・調査するために必要なインディケータ情報や、被害予防に必要な侵入経路などの TTP（※1）情報を共有する。
報告	法令に基づき、または任意により、被害事実や対応状況について行政機関に伝えること。
連絡	取引先や顧客等、なんらかの利害関係者に対して攻撃被害が発生したことや、調査結果等を伝えること。個人情報漏えいなどで顧客や取引先に二次被害が発生するおそれがある場合や、調査がある程度終わり、被害内容や取引先へ影響の有無などを伝えるために行う。

※「サイバー攻撃被害に係る情報の共有・公表ガイダンス」をもとにＭＳ＆ＡＤインターリスク総研が作成

■３．情報開示の留意点

サイバーインシデントの発生時にはインシデントレスポンス（※2）として、被害を受けたシステムの復旧、企業の危機管理対応、事業の復旧の 3 つが同時並行で進む。また、それぞれが緊密に連携することが欠かせない。



① インシデント発生直後の留意点

インシデント発生直後は、検知した自組織では混乱が生じ、システム障害の原因が明確でないこともある。ランサムウェア被害の場合、攻撃者からの犯行声明や身代金要求文書（ランサムノート）が見つかれば、原因がサイバー攻撃だと断定できる。ただ、被害範囲の特定には時間を要する。このように、サイバーインシデント時の広報対応は、通常の新製品発表やイベント告知とは異なり、情報が未確定な中で発信しなければならない点が最大の特徴である。そのため、確認済み事実と未確認事項を明確に分け、断定を避けて伝えることが重要である。

社内の IT 部門や CSIRT（※3）、危機管理委員会や対策本部は連携して短時間で情報を整理する。開示要否、開示目的、開示主体（単独か関係者と連名にするか）、開示範囲（開示先、開示する情報の範囲）、開示方法（開示方法は後述の第 4 項参照）、開示の時期（初報、続報のタイミング）といった項目を基本方針として定め、これに則って対応を進める。

② 続報における留意点

事故に関する調査の進展や、顧客などのステークホルダへの対応方針が定まったタイミングで適時に情報を開示する。危機時は、広報部門だけでなく、情報システム、法務、経営、外部専門家が連携し、メッセージを一本化することが欠かせない。広報におけるポジションペーパーに則り、何を開示してよいか／開示してはいけないか、問い合わせがあった場合の対応方針等を定め、情報を統制する。

お詫び文章を掲載することは重要だが、感情表現だけで終わらせてはならない。再発防止策と問い合わせ窓口を明示し、実行可能な対応を示すことが、信頼回復につながる。

■４．情報開示手法の特性

情報開示に、必ず記者会見やプレスリリースが必要ということではない。状況と目的に応じて、適切な開示方法を選ぶ。

【表２】開示方法

位置づけ	内容	狙い	タイミング
オフィシャルサイト掲出	発生事案の説明や事案を受けた見解などを自社のオフィシャルサイトに掲載する。	● 正確な事実を伝える。 ● 情報の相談窓口になる。(お客さまや取引先が「自分は大丈夫か」を確認できる場所を作る)	「第１報」は、インシデントを検知後、速やかに掲載する。詳しい原因がまだわからなくても、事故の概要や詳細調査中であることを示す。合理的な理由なく事案発生から第一報までに時間を要した場合、隠蔽が行われていたと評価されるリスクがある。調査が進み、新しい事実が判明したタイミングや、被害を受けたステークホルダへの対応が決まったタイミングで、その都度掲載する。
プレスリリース	記者クラブの加盟社などに配布、またはメールなどで送信する。	● メディアの力を借りて、自社のホームページを見ない人にもニュースとして広く情報を届ける。 ● 隠さずに自ら情報をメディアに渡すことで、「誠実に対応している」という姿勢を示す。	広範囲な被害に及び、社会的な影響が大きい可能性があると判断した段階で実施する。(個人情報が大量に漏えいしている可能性がある、等)
記者会見	記者クラブや用意した会場で、さまざまなメディアの記者を集め、説明や質疑応答に対応する。 会社の代表者（社長や役員）が登壇・説明する。	● 経営としての責任を示し、深刻な事案に対して丁寧に説明・謝罪する。	重大事故・大規模被害など、社会的関心が高く、代表者の説明が必要なときに実施する。ただし、事実関係（原因や被害の全体像）がある程度判明していない状況で実施しても記者からの質問に十分答えられず、逆効果になるおそれがある。

■５．平時の情報開示

ここまで有事における情報開示を説明してきたが、より理解を深めるために、平時の段階における情報開示との違いを述べる。平時にサイバーセキュリティへの対応状況を開示することの意義は、社会的信用を獲得し、事業の成長につなげることである。例えば、取引先や顧客から選ばれるための信頼獲得や、取引維持のためにサイバーセキュリティ管理体制の要件を満たしていることを示す目的がある。

これに加え、投資家や株主に対する公表は、自社のサイバーセキュリティガバナンスの実効性を示すことで ESG 投資(※4)につなげることを目的とする。信用格付け会社では、サイバーセキュリティの管理体制単独で評価することはないが、ESG のうちの G：Governance（企業統治）の観点で、アナリストによる経営陣へのヒアリングや公開情報（事業計画、管理・監査体制等）に加

え、外部パートナーと連携したセキュリティ診断の結果を用いるケースがある。

■６．備えの重要性

平時から、想定されるインシデントごとの公表文案、承認フロー、問い合わせ窓口、社内連絡網を整えておく必要がある。あわせて、広報・法務・情報システム・経営層が連携できる体制を作り、誰が何を判断するかを明確にしておくことが重要である。さらに、第一報で伝えるべき項目を整理し、定期的に訓練を行うことで、実際の発生時にも迷わず対応できる。準備の有無が、炎上の抑制と信頼維持を大きく左右する。

9. Tactics, Techniques, and Procedures の略で、攻撃者の行動パターンや攻撃の進め方を表す用語
10. サイバー攻撃や情報漏えい、システム障害などのインシデント発生時に行う対応のこと
11. Computer Security Incident Response Team の略。情報セキュリティインシデントに対応するための組織・チームを指す。
12. 投資先を選ぶ際に、利益だけでなく E：Environment（環境）、S：Social（社会）、G：Governance（企業統治）の 3 つを重視し、環境や社会への配慮、経営の健全性が高い企業を評価して投資する考え方。

編集者コメント

本誌では、デバイスコードフィッシングの攻撃ツールキット「EvilTokens」を紹介した。その後、2026 年 5 月 21 日には、米国連邦捜査局（FBI）が IC3（インターネット犯罪苦情センター）を通じ、Microsoft 365 アカウントを標的とする新たな PhaaS「Kali365」への注意を呼びかけた（※1）。PhaaS の一部は国際的な連携で摘発・無力化される。しかし、活動を再開する攻撃者グループもある（※2）。摘発と新たな PhaaS の出現は、今後も繰り返される公算が大きい。

また、フロンティア AI を巡るサイバーセキュリティの状況も、執筆後、急速に動いた。その象徴が 2026 年 7 月 21 日に公表された OpenAI の事案である。OpenAI によると、AI のサイバー攻撃・防御能力の評価中、AI エージェントがテスト環境（サンドボックス）の未知の脆弱性を突いて外部へ到達し、AI 共有基盤「Hugging Face（※3）」の本番環境への侵入・攻撃に至ったという（※4）。

攻撃手法は巧妙化し、AI の技術リスクも新たな段階に入った。サイバーセキュリティには、従来の想定を超える備えと、より迅速な対応が欠かせない。

1. FEDERAL BUREAU OF INVESTIGATION「Kali365 Phishing-as-a-Service Kit Hijacks Microsoft 365 Access Tokens」<https://www.ic3.gov/PSA/2026/PSA260521>（最終アクセス 2026 年 7 月 23 日）
2. Abnormal「Tycoon2FA Rebounds Post-Takedown with 6 Layers of Obfuscation」<https://abnormal.ai/ja/blog/tycoon2fa-post-takedown-rebuild>（最終アクセス 2026 年 7 月 23 日）
3. 世界中の研究者や開発者が AI モデル、データセット、AI デモアプリを共有・公開する、AI 分野の巨大プラットフォーム
4. OpenAI「OpenAI と Hugging Face、モデル評価中のセキュリティインシデント対応で連携」<https://openai.com/ja-JP/index/hugging-face-model-evaluation-security-incident/>（最終アクセス 2026 年 7 月 23 日）

ＭＳ＆ＡＤインターリスク総研株式会社は、ＭＳ＆ＡＤインシュアランスグループのリスク関連サービス事業会社として、リスクマネジメントに関するコンサルティングおよび広範な分野での調査研究を行っています。本誌を編集している以下のグループでは、サステナビリティ関連のコンサルティング・セミナー等のサービスを提供しています。
ＭＳ＆ＡＤインターリスク総研サービスに関するお問い合わせ・お申し込み等は、下記のお問い合わせ先、または、お近くの三井住友海上、あいおいニッセイ同和損保の各社営業担当までお気軽にお寄せ下さい。

お問い合わせ先
ＭＳ＆ＡＤインターリスク総研㈱
リスクマネジメント第三部　サイバーリスクグループ
cyberrisk_irric@ms-ad-hd.com
https://www.irric.co.jp/

主な担当領域は以下のとおりです。

＜サイバーリスク管理＞

- ◆ 現状評価･課題分析
- ◆ 体制構築
- ◆ リスクアセスメント
- ◆ 有価証券報告書における開示の充実化
- ◆ サイバーBCP策定支援
- ◆ サイバーインシデント時の情報開示･広報
- ◆ サイバーインシデント演習
- ◆ サイバーインシデント初動対応マニュアル策定支援
- ◆ 担当領域に係る研修　等

本誌は、マスメディア報道など公開されている情報に基づいて作成しております。
また、本誌は、読者の方々に対して企業のリスクマネジメント活動等に役立てていただくことを目的としたものであり、事案そのものに対する批評その他を意図しているものではありません。

MS&AD インターリスク総研は、2024 年 4 月、これまでのホームページを刷新し、リスクに強い組織づくりをサポートするプラットフォーム「RM NAVI（リスクマネジメント ナビ）」をリリースしました。「RM NAVI」は、MS&AD インターリスク総研の知見をフル活用して、情報提供から実践までをトータルサポート。コンサルタントの豊富な経験と、最先端のデジタルサービスで、リスクに強い組織づくりを支えます。
あなたに寄り添い、最適な答えへと導く、リスクマネジメントの羅針盤です。