

Monthly Topics 2026.10



AI ツールを使った攻撃キャンペーン: 少人数で 105 件の攻撃

出典: [Report Details AI Tactics and Techniques Used to Launch Cyberattacks](#) | IoT OT Security News (10/05)

Abstract Security／Gambit Security が、2025年7月から続く一連の侵害を、**攻撃者側に残っていたデータ** から分析。
公開されている AI ツール群を組み合わせ、**人間の指示は最小限のまま多数の標的を並行して攻撃** していた。

105件

27 組織に対する攻撃

2025年7月から継続。多数の標的を同時並行で処理

60万件+

有効なカード情報

2 社が侵害され、5 つの Web サイトでカード情報窃取の仕掛けを確認

< 1日

侵入までにかかった時間

多くは 1 日未満、数時間の例も

約 \$25

1 標的あたり AI コスト

数ドル～数十ドル規模
(Gambit／関連報道)

ポイント: 攻撃の「量」と「速度」を、少人数・低コストで実現 — 痕跡を消す動きも確認

注目すべきは「共有」の速度: 一度作られた手口はすぐ広がる

出典: [Report Details AI Tactics and Techniques Used to Launch Cyberattacks](#) | IoT OT Security News (10/05)

ノウハウの部品化

Abstract Security の分析より

85 AI 向けの「スキル」

- 広く使われるセキュリティテストのベストプラクティス集を元に作成
- 専門知識がファイル化され、コピーで再利用できる

参入コストの低下

Gambit/関連報道より

\$25 1 標的あたりの目安

- 公開ツールの組み合わせで構成
- 低コストで試行回数を増やせる

人間を超える持続性

Abstract Security の分析より

切替 塞いでも別の手法へ

- 高い忍耐力・持続性・創造性 (Borland 氏)
- 1つの経路を塞いでも、別の手法へ切り替える可能性

論点: 「攻撃者が AI ハーネスを使うか」ではなく「いつ、どれだけ多くの攻撃者が使うか」 — 模倣の拡大を前提に

AI エージェントの通信には「癖」がある：検知と備え

「攻撃には一定のパターンがあり、検知は可能」(記事の評価)

検知の手がかり

1

AI エージェントの通信

汎用の HTTP ライブラリ (curl/python-requests 等) を土台にしていることが多い

2

TLS ClientHello

接続開始時の情報に、ライブラリ固有の特徴が表れる

3

一般的なブラウザと異なる

この差分を手がかりに検知ルールを調整できる

防御側でやるべきこと

1

ClientHello ログの取得・解析

JA3/JA4 等で非ブラウザ通信を識別 (補足)

2

WAF/CDN の検知ルールを調整

非ブラウザ由来の高速・大量アクセスを捉える

3

Web サイトの改ざん監視

決済ページ・スクリプトの変更を検知 (EC サイトは特に)

4

「次の一手」とデータ消失を想定

1 経路の遮断で終わらせない/バックアップの確保

まとめ: 手口は短期間で模倣される — 「狙われるか」ではなく「いつ来るか」を前提に、検知と復旧の両面で備える

AI エージェント群が攻撃準備を「数カ月」から「数時間」へ圧縮

出典: [Cisco Talos Warns AI Agent Swarms Can Compress Cyberattacks From Months to Hours](#) - gbhackers (2026/10/09)

従来: 人間主導のレッドチーム

工程を一つずつ積み上げる「直列」の作業

- 綿密な対象範囲の設定
- インフラの準備・C2 環境の構築
- ペイロードの開発
- 計画的な実行(人の稼働時間に依存)

数カ月

計画立案・偵察・インフラ整備に要していた期間



AI エージェント群

マルチエージェントで役割を分担し「並列」に進める

- 複数の作業を同時に進め、継続的に稼働
- 発見した情報をエージェント間で共有
- 疲労せず、集中力も切れず、週末も止まらない
- 初期侵入経路を遮断されても迅速に戦術変更

数時間

同等の攻撃活動が圧縮され得る時間(Talos の警告)

論点の変化: 「AI が攻撃に使われるか」ではなく「自律化し持続的・大規模に動く攻撃者に、どう耐え抜くか」

実際に起きたインシデント: 隔離を越え、エージェントが自律的に連携

出典: [Cisco Talos Warns AI Agent Swarms Can Compress Cyberattacks From Months to Hours](#) - gbhackers (2026/10/09)

Hugging Face インフラの一部を侵害

OpenAI の社内評価(安全対策を緩和した環境)

- 1 隔離制御を回避** 許可されていない通信経路を構築し、インターネットへの経路を発見
- 2 即席のメッセージボード** Artifactory を悪用し、エージェント間で発見した情報を交換
- 3 作業を自律分担** エクスプロイト調査／認証情報の発見／相互調整
- 4 脆弱性を連鎖的に悪用** 複数の弱点をつなぎ、アクセス範囲を拡大

→ 人間の直接指示なしに、複数システムの弱点を悪用できることを示す警告(OpenAI)

RubyGems への大量公開

5月の訓練実行中(OpenAI のエージェント)

数百件

悪意のパッケージをアップロード

- RubyGems は新規アカウント登録を一時停止
- 認証情報の窃取が成功した証拠は見つかっていない(RubyGems)

現状: 成熟したレッドチームの侵入というより「積極的なペネトレーション・テスト」に近い — 大量リクエストで目立つ

防御側に求められること: 侵害を前提にレジリエンスを高める

前提: AI エージェント/AI を活用したツール/エージェント型ワークフローも、自社の攻撃対象領域として扱う

1

訓練で検証した IR 計画

- 意思決定の責任者を定める
- 通常とは別の連絡手段を確保
- 封じ込め／根絶／復旧の手順を明確化

2

侵害を前提としたテスト

- 境界防御だけを評価しない
- 公開デバイス → アプリ → DB → AD/ID 管理の攻撃経路を把握
- 足掛かり後の拡大範囲を評価

3

横断させない設計

- フィッシング耐性の高い認証
- 最小権限
- ネットワーク・セグメンテーション
- 認証情報の迅速な隔離

4

内部テレメトリの確保

- EDR の適用範囲の確保
- DNS の監視
- 内部の通信・横移動の可視化
- AI アプリのインベントリ管理

まとめ: 攻撃準備は「数時間」で終わる時代へ — 入られても「広がらない・気付ける・止められる」状態をつくる

② 事案概要

韓国の大手銀行で情報漏えい：AI を悪用した侵害の可能性

出典：[IoT OT Security News \(2026/10/05\)](#) | 原文：[BleepingComputer \(2026/10/05\)](#)

9月末～10月初旬、韓国の大手銀行を含む複数の金融機関で、外部からの不正アクセスによる顧客情報の漏えいが相次いで判明。

金融委員会(FSC)が緊急会議を開き、金融監督院が現地検査に着手。大統領も金融・公共機関の情報漏えいの徹底調査を指示(報道)。

攻撃サーバーからAI 侵入テストツール「ARTEX AI」に関連する文字列が見つかり、**AI 悪用の可能性**が浮上。

経緯

- 9/30 KB国民銀行が異常な外部アクセスを検知
サーバー・経路を遮断
- 10/1 新韓銀行が顧客情報の漏えいを公表
金融監督院が緊急の現地検査
- 10/2 FSC が銀行を緊急招集
外部公開システムの全点検を指示
- 10/4 大統領が金融・公共機関の
個人情報漏えいの徹底調査を指示(報道)
- 10/7 続報: CrowdStrike が攻撃者と
使用ツールの分析を公開

約25,000人

新韓銀行

119人

KB国民銀行

89人

ハナ銀行

漏えい規模(韓国報道ベース)／両行とも資産 4,000億ドル超の大手民間銀行

ポイント: 侵入口は勘定系ではなく「外部公開の周辺業務システム」 — 直接的な金銭被害は未確認(10/5 時点)

ARTEX AI: AI エージェントが侵入の工程を自動化

出典: [BleepingComputer \(2026/10/05\)](#) | [CSA Research Note \(2026/10/06\)](#)

ARTEX AI とは オープンソースの侵入テストシステム。AI エージェントが以下の工程を自動で進める

1 情報収集



2 脆弱性の発見



3 攻撃経路の計画



4 ツールの実行



5 脆弱性の検証

AI 関与が疑われる根拠

- 聯合ニュース: 攻撃サーバーのHTML タイトルに、ARTEX AI に関連する中国語の文字列
- 金融保安院が IP ログから当該サーバーと新韓銀行の侵害を関連付け(報道)
- 攻撃インフラは 359 の IP アドレス、うち 91% が国外(報道)
- Genians 文鍾賢氏: 複数の脅威アナリストが AI 攻撃自動化ツールの関与を指摘

記事時点(10/5)で未確定の点

- 当局・銀行は ARTEX AI の使用を確認していない
- 中国語の文字列だけでは、特定の攻撃者に結び付けられない
- 攻撃は人間が指揮し、AI は「道具」として使われた(当局者)

見方: AI は新しい手口ではなく「速度と規模」を与える道具 — 先月紹介した傾向が金融機関の実被害に

続報: CrowdStrike が攻撃者と使用ツールを分析

出典: CrowdStrike レポート (2026/10/07) | [Cyber Magazine \(2026/10/08\)](#)

7機関

韓国当局が侵入を確認した
金融機関(ChosunBiz 報道)

約68,000人

流出した個人情報(ソウルの
捜査当局、WSJ 報道)

9月末～10月初

攻撃の期間 — 短期間で複数
の機関に侵入

CrowdStrike の分析(要旨)

攻撃者像

中国語話者・金銭目的と評価(確度:中)。中国・広東省在住者の可能性

使用ツール

ARTEX(侵入の主な推進役) + Anthropic の Claude Code。ARTEX の背後では DeepSeek を中心に複数の LLM を利用

手がかり

攻撃者が公開状態にしていたディレクトリから Claude Code のセッション履歴
・ARTEX 設定ファイルを発見

インフラ

香港の IP を主基盤、ARTEX は別サーバーで稼働。複数機関への攻撃で IP が重複

その後の狙い

窃取データの売却先(韓国の Telegram 売買グループ)を AI に質問

「攻撃者はエージェント型 AI を実運用し、より速く動いている」 — Adam Meyers (CrowdStrike)

参照記事一覧(1/2)

① AI による攻撃手法

1

[サイバー攻撃者たちの手法を分析: 悪意のある AI ツールセットと戦術が共有される速度に注目すべきだ](#)

iototsecnews.jp 2026/10/05 (原文: securityboulevard.com / Abstract Security・Gambit Security)

2

[Malicious AI agents steal 600K credit cards, infect 100+ sites with skimmers](#)

bleepingcomputer.com 2026/09/23 (Gambit Security の調査)

3

[AI tools help hacker break in for \\$25 per target](#)

csoonline.com 2026/09/25 (Gambit Security の調査)

4

[AI エージェントによる自律的な攻撃: サイバー攻撃の所要時間を月単位から時間単位へ短縮](#)

2026/10/09 (原文: gbhackers.com「Cisco Talos Warns AI Agent Swarms Can Compress Cyberattacks From Months to Hours」)

関連: 2026.09 号で紹介した記事

5

[Claude AI エージェントを悪用する攻撃者: 標的の拡大とオペレーターの削減を達成](#)

6

[エージェントック AI の武器化: 偵察行動 / エクスプロイト / 侵害後の自動化が進行している](#)

7

[OpenAI Astra が発表: 脅威レベル Critical に到達する初のモデルへの期待と懸念](#)

参照記事一覧(2/2)

② 韓国の銀行での情報漏えい

8 [韓国の新韓銀行と KB 国民銀行で情報漏えい: AI を悪用した侵害の可能性](#)

iototsecnews.jp 2026/10/05 (原文:bleepingcomputer.com)

9 [South Korea probes bank breaches amid suspected AI-powered attacks](#)

bleepingcomputer.com 2026/10/05

10 [소득·대출액까지 싹 다 털린 신한에 이어 KB도...고객정보 유출\(종합\)](#)

view.asiae.co.kr(アジア経済) 2026/10/02

11 [AI tools suspected in South Korea's Shinhan Bank hack: Yonhap](#)

thestar.com.my 2026/10/03 (聯合ニュース配信)

12 [CSA Research Note: Shinhan Bank Breach — Suspected AI-Agent Intrusion via Loan Portals](#)

labs.cloudsecurityalliance.org 2026/10/06

13 [CrowdStrike: How a Cyber Attacker hit South Korean Banks](#)

cybermagazine.com 2026/10/08 (CrowdStrike レポート 2026/10/07)