GMO Flatt Security

急増するAI不正アクセス　今取り組んで欲しいこと

年次×1アプリ診断から
毎日×全社AIペネトレへ

2026/11/05　GMO Flatt Security株式会社



豊田 恵二郎 @toyojuni

- GMO Flatt Security（株） 執行役員CCO / 共同創業者
- 東京大学工学部航空宇宙工学科卒
- 事業責任者としてP/Lに責任を持ちつつ、
  広報・マーケティングを管掌

01 なぜ、いま攻撃が急増しているのか

02 AIは、その会社固有の脆弱性まで突ける

03 守る側は、どう変わるべきか（今）

04 不正アクセス抜き打ちテストのご紹介

05 守る側は、どう変わるべきか（これから）

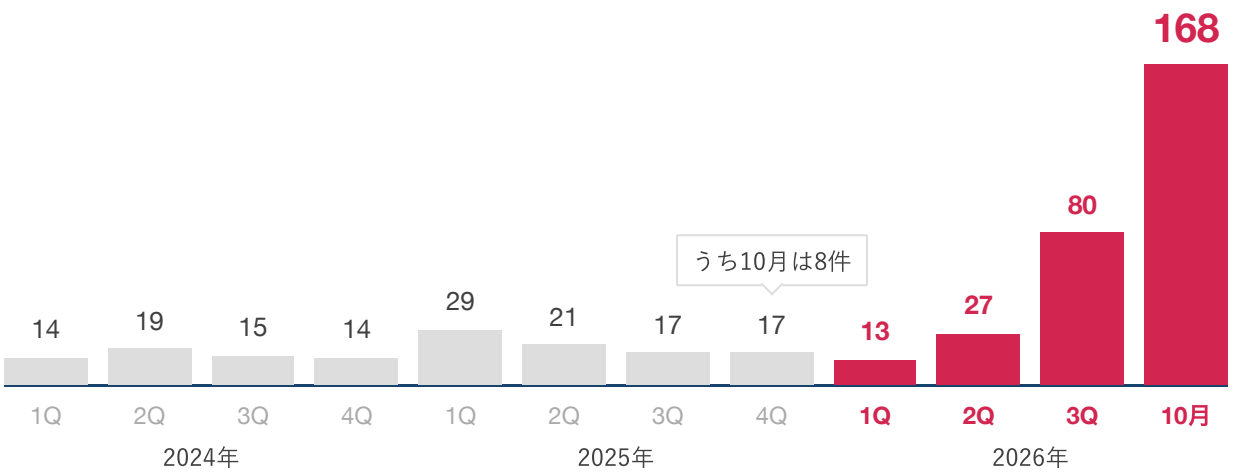
GMO Flatt Security

なぜ、いま攻撃が急増しているのか

鍵は、セーフガードのないオープンモデル

国内のWebシステムからの情報漏洩は、前年同月比21倍に急増

公表件数（四半期ごと）



2026年10月の前年同月比

21倍

2025年10月の8件に対し、2026年10月は168件。1か月で2025年1年間（84件）の2倍に達しています。

出典：マクニカ「相次ぐWEBシステムからの情報漏洩事案について」（2026年11月）の月別件数を四半期ごとに合計。2026年10月31日までの公表分で、2026年4Qは10月のみ https://security.macnica.co.jp/blog/2026/10/web-incidents2026.html

ハーネスに注目が集まりがちだが、能力の源泉はモデル

ハーネス

ARTEX など。外部システムとの通信、ツール呼び出し、
並列実行を担う枠組み



馬だけ
付け替える

モデル

能力の源泉。走る力は馬から来る



Claude



DeepSeek



GLM



Grok

出典：BleepingComputer（ARTEX、CrowdStrike の分析） — https://www.bleepingcomputer.com/news/security/hacker-used-artex-ai-and-claude-agents-to-target-south-korean-banks/

大手・クローズドはセーフガード強め。
オープンウェイト系は「無検閲」も出回り自由度高め

大手のモデル

Claude　GPT　Gemini

- 最も強いモデルは一般公開しない
- 攻撃的な用途は審査制で提供
- 既定のセーフガードで攻撃を止める

オープンウェイト

GLM　DeepSeek　Kimi

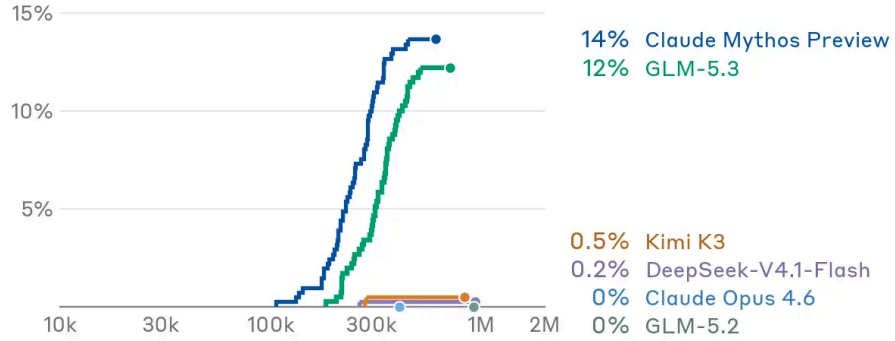
- モデル本体が丸ごと配布される
- 「無検閲」版が流通する
- 手元で動かせば、提供元の監視が届かない

出典：Anthropic「GLM-5.3 and the spread of advanced cyber capabilities」 — https://www.anthropic.com/research/glm-5-3-and-the-spread-of-advanced-cyber-capabilities

オープンウェイトのGLM-5.3が、最先端の攻撃能力に並んだ（8月）

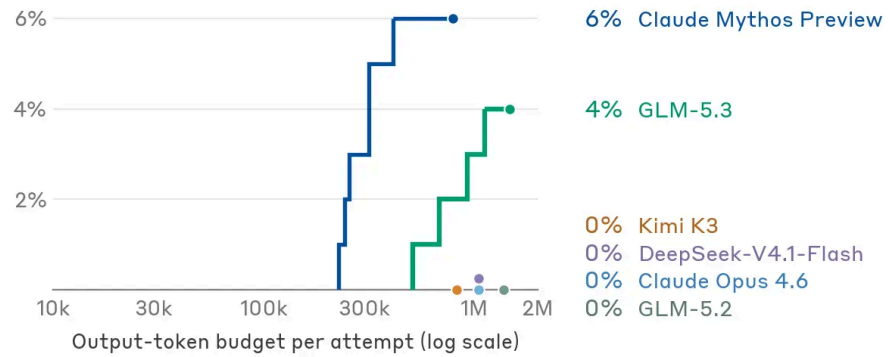
ExploitBench

Public benchmark. Share of each model's attempts that built a complete working exploit within the budget



Binary Exploitation Benchmark

Anthropic-internal benchmark. Share of 100 fuzzing entry points where the attempt reached a control-flow hijack within the budget



出典：Anthropic「GLM-5.3 and the spread of advanced cyber capabilities」（2026年9月29日） — https://www.anthropic.com/research/glm-5-3-and-the-spread-of-advanced-cyber-capabilities

オープンウェイト系モデルはすでに現場で使われている

攻撃に



韓国の金融機関への攻撃で使われた

ARTEX に DeepSeek V4.1-Flash を載せ、GLM-5.3・Grok 4.6 を併用

防御側でも

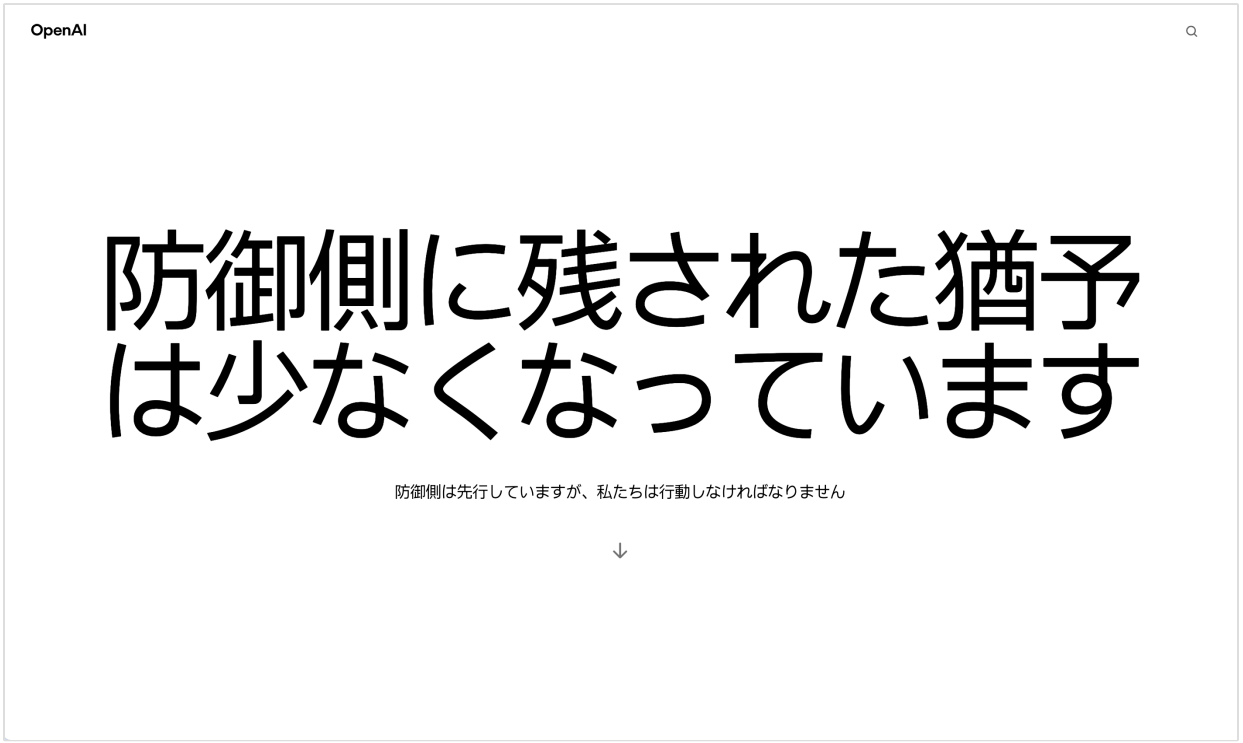


Hugging Face が侵入の解析に使った

米国モデルは安全機能で解析を拒否。自社環境の GLM-5.2 に切り替えた

出典：CrowdStrike「Unknown Threat Actor Uses AI-Driven ARTEX to Target South Korean Finance」（2026年10月7日） — https://www.crowdstrike.com/en-us/blog/unknown-threat-actor-uses-artex-to-target-south-korean-finance/
Hugging Face「Anatomy of a Frontier Lab Agent Intrusion」 — https://huggingface.co/blog/agent-intrusion-technical-timeline

守る側の「猶予」は、~~少ない~~ もう残っていない



出典：OpenAI「Defense Factory」 — https://openai.com/ja-JP/the-defense-factory/

GMO Flatt Security

いま、何が起きているのか

AIは「その会社固有の脆弱性」を攻撃できる

脆弱性には「共通の脆弱性」と「固有の脆弱性」がある

① 共通の脆弱性

依存関係の既知の脆弱性

ライブラリ・ミドルウェア等の脆弱性。使っている会社すべてが、同じ弱点によって影響を受ける。CVE 等で共有される脆弱性がこちらに該当する

例：2026年10月に公開された php2shell



② 固有の脆弱性

独自実装・設定不備による脆弱性

自社で書いたコードや、クラウド等利用ツール・プラットフォームの設定不備に起因する脆弱性。そのシステムにしか存在しない

出典：GMO Flatt Security Blog「PHPの重大な脆弱性「php2shell」の概要と対応指針」 — https://blog.flatt.tech/entry/php2shell

「共通の脆弱性」は「数打ちゃ当たる」が可能

例：WordPress本体やプラグインの脆弱性を突くコードを、
WordPressを使っていそうなWebサイトに手当たり次第に投げる



② 固有の脆弱性 も、攻撃者にとってコスパが良くなった

攻撃者のコスパ	これまで	いま
① 共通の脆弱性	○ 1つの攻撃コードを何百社にも流せる 	○ 既存の高コスパに加え、攻撃コード作成までの速度も向上
② 固有の脆弱性	△ 専門スキルが必要 多くの場合、割に合わない	○ AIが自律的に脆弱性を発見・攻撃
攻撃者の狙い	共通の脆弱性だけ	どちらの脆弱性も

その結果、攻撃は「ばら撒き」から「1社ずつ」へ

これまで



「共通の脆弱性」を狙い、同じ攻撃コードを多くの会社にばら撒く

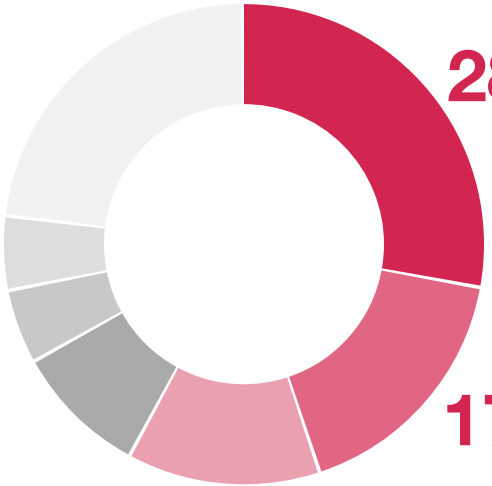
いま



- AIが1社ずつ調べ、「固有の脆弱性」を探して悪用
- もちろん「共通の脆弱性」も悪用可能
- 攻撃がスケールする。多数同時並行

Webアプリにおける「固有の脆弱性」で特に多いのは、認証・認可・その他ロジックの脆弱性

当社の診断で見つかった脆弱性の内訳（2023〜2024年）



28% 認可制御不備

17% 認証の設計・実装不備

13% ロジックの脆弱性（その他）

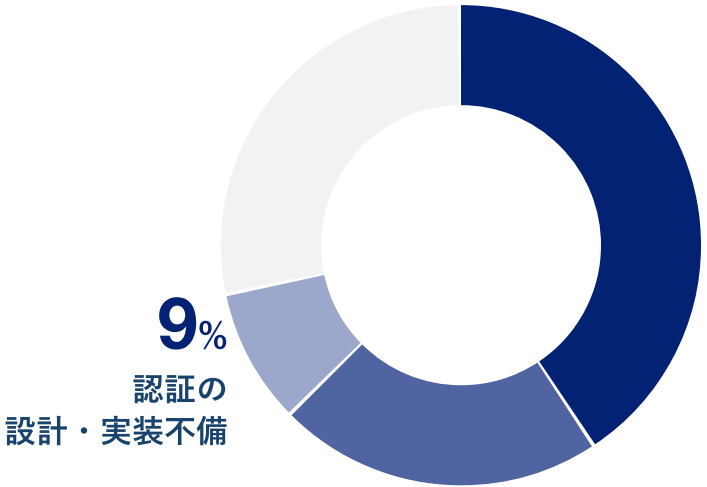
上位3つの合計

58%

出典：GMO Flatt Security Top 10 - 2025年版 — https://blog.flatt.tech/entry/flatt_top10_2025

AIは、そのロジックの脆弱性を見つけられる。
非AI DAST/SASTでは難しかった

Takumi byGMO が見つけた脆弱性の内訳（2026年5月29日〜8月4日／ブラックボックス診断の検出実績）





上位3つの合計

72%

プラットフォーム横断的な攻撃能力を有し（人間では稀）、Webアプリ侵害後の侵入もお手のもの

入口

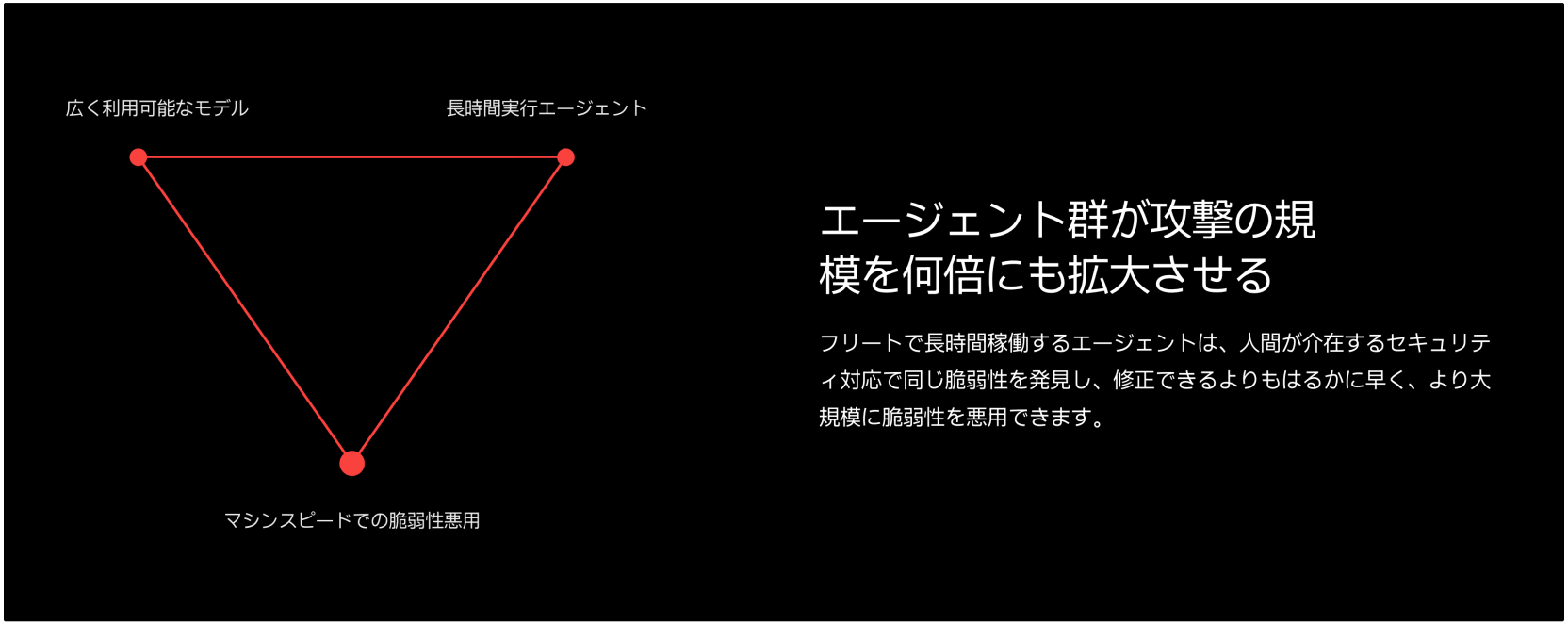
Webアプリ

その後、AI は横展開や影響の最大化を試行する

クラウド	aws	・認証情報を自動で大量発行 ・組織全体の管理権限を奪取
Kubernetes		・クラスタ管理者に昇格 ・Secretから認証情報を一括取得
CI/CD		・パイプラインで秘密情報を収集 ・悪性マニフェストを本番へ同期
IdP		・管理者ユーザーを作成 ・権限グループに追加し、SSOで侵入

出典：Wiz「Anatomy of a Real AI-Driven Cloud Attack」 — https://www.wiz.io/blog/anatomy-of-ai-driven-cloud-attacks

専門家レベル（もしくはそれ以上）の攻撃が「モデルアクセス・長時間自律実行・マシンスピード」で増幅される



エージェント群が攻撃の規模を何倍にも拡大させる

フリートで長時間稼働するエージェントは、人間が介在するセキュリティ対応で同じ脆弱性を発見し、修正できるよりもはるかに早く、より大規模に脆弱性を悪用できます。

出典：OpenAI「Defense Factory」 — https://openai.com/ja-JP/the-defense-factory/

GMO Flatt Security

守る側は、どう変わるべきか（今）

広く、同時に、止まらない攻撃に備える

結論

多くの会社で

ペンテストを診断より先に行う優先順位への転換が推奨される

脆弱性診断とペネトレーションテストの違い



脆弱性診断

すべてのマスを網羅的に確認する

	SQLi	XSS	認可	CSRF	認証
ログイン	○	○	○	×	○
会員登録	○	○	○	○	×
マイページ	○	×	○	○	○
決済	○	○	×	×	○
管理画面	×	○	×	○	○

機能 × 観点のすべてのマスを確認し、○か×をつける



ペネトレーションテスト

侵害を可能にする脆弱性やその組み合わせを試行

	SQLi	XSS	認可	CSRF	認証
ログイン					
会員登録					1
マイページ					
決済					
管理画面	3		2		

1 社内ドメインで会員登録 → 2 管理画面に入れてしまう →
3 SQLiで顧客を全件取得 → 個人情報漏洩

脆弱性診断とペネトレーションテストの違い



脆弱性診断

すべてのマスを網羅的に確認する

	SQLi	XSS	認可	CSRF	認証
ログイン	○	○	○	×	○
会員登録	○				
マイページ	○	×	○	○	○
決済					○
管理画面	×		×		

脆弱性のチェーン評価はベンダーによる

実際には予算問題で全機能を見られないことも多い

アプリケーションの外は見ない

機能 × 観点のすべてのマスを確認し、○か×をつける



ペネトレーションテスト

侵害を可能にする脆弱性やその組み合わせを試行

	SQLi	XSS	認可	CSRF	認証
ログイン					
会員登録					1
マイページ					
決済					
管理画面	2		2		

目的達成に関係しない箇所・観点は検査しない

診断よりもさらに高価

時には対象システム外の要素も悪用して目標達成につなげる

1 社内ドメインで会員登録 → 2 管理画面に入れてしまう →

3 SQLiで顧客を全件取得 → 個人情報漏洩

脆弱性診断とペネトレーションテストは「問い」が違う



脆弱性診断

問い

脆弱性の網羅的なチェック

向いている

安全基準を満たしたかの確認（リリース判定）

苦手

範囲外の資産、脆弱性の連鎖による侵害時のインパクトの見積もり



ペネトレーションテスト

問い

攻撃者が特定の侵害目標を達成できるか

向いている

致命的なリスクの有無の検証

苦手

網羅性の保証

これまで、診断→ペンテストの順番をおすすめしてきました

これまで



脆弱性診断

アプリを100%安全に



ペネトレーションテスト

余力があれば

しかし、AIによって

- **攻撃が激化した**

  致命的リスクは、一時も看過できない

- **診断・ペンテストのコストが下がった**

  AIで、安く実施できる

AIで、ペネトレーションテストが「最初の一歩」になった

- **これまで、「外から殴る」ができる人材は限られ、サービスは高価だった**
  そのため、最初の一歩にはなり得なかった
- **いまは、AIの力で安く実施できる**
  裏返せば、攻撃者も同じ力で殴ってくる
- **だから、ペネトレーションテストが最初の一歩になりうる**



完璧を目指す前に、まず「外からどこまで殴れるか」を確かめる

これまで



脆弱性診断
アプリを100%安全に

→



ペネトレーションテスト
余力があれば

いま



ペネトレーションテスト
まず致命的な脆弱性を塞ぐ

→



脆弱性診断
その後、より完璧な状態へ

まず「外から見える範囲を広く」「AIによる」「ペネトレーションテスト」が必要

GMO Flatt Security

不正アクセス抜き打ちテスト

一律98万円でAIの脅威に備える

不正アクセス抜き打ちテスト by GMO

攻撃者と同じ立場で、会社全体が不正アクセスされ得るかAIで検証

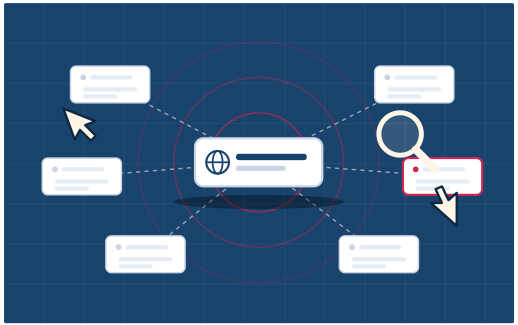
費用（税抜・一律）

98万円

※2026年中のお申し込み限定の特別料金です

目的	個人情報の持ち出し等、重大な不正アクセスができてしまわないか、**実際に侵入を試みて**確認し、あわせて**機微情報の漏洩**も調べる
対象	当社が探索し、お客様にご承認いただいた**会社全体**のインターネット公開資産
期間	**30日間**のいずれかのタイミングで、予告なく抜き打ちで実施
実施者	当社のセキュリティ専門家が、**1万件以上**の脆弱性を見つけてきたAIエージェントを用いて実施
成果物	重大な問題の**速報**と、期間終了後の**最終報告書**

攻撃者目線で、資産探索から侵入まで一気通貫。機微情報の漏洩も調査



資産探索

ドメインを手がかりに、攻撃者と同じ方法でインターネット上の公開資産を洗い出します。



ペネトレーションテスト

見つけた資産すべてを対象に、AIが攻撃者と同じ手口で実際に侵入を試みます。



機微情報の漏洩調査

GitHubなど、インターネット上の公開情報から、お客様に関係するパスワードやAPIキーなどが漏れていないかを調べます。

※機微情報の漏洩調査では、見つけた情報を使ったログインや、有効かどうかの確認は行わず、お客様へのご報告にとどめます。

AI時代の脅威にいま備える、安価・広範囲なペネトレーションテスト

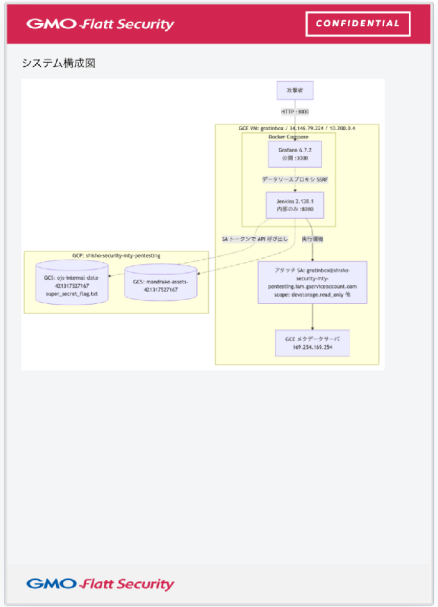
	不正アクセス抜き打ちテスト by GMO	一般的な脆弱性診断
資産探索	あり	なし
機微情報の漏洩調査	あり	なし
対象範囲	会社全体（アプリケーションからネットワークレイヤーまで、外部から到達し得るすべて）	アプリケーション単体
対象環境	本番環境を含む	原則デモ環境
費用（税抜）	98万円	100〜300万円 × アプリケーション数
検証できること	個人情報漏洩やサーバーの乗っ取りなどの、致命的なリスクの有無	深刻度の低いものから高いものまで、網羅的なリスクの洗い出し

お客様の全社資産に対し、30日間のどこかで抜き打ち侵入を試みます

1	主要ドメインの共有	お客様	必ず対象に含めたいシードドメインを共有　例：flatt.tech, shisho.dev
2	資産探索と一覧の共有	当社	DNSの受動的な調査と最小限のHTTPリクエストで探索し、 見つけた攻撃対象の一覧を共有
3	対象のご承認	お客様	他社管理など、外したい資産を確認して承認
4	抜き打ちテスト（30日間）	当社	より深い資産探索、会員登録を含む実際のアプリ操作、 脆弱性を悪用した侵入の試行

塞ぐべきクリティカルな脆弱性を、速報と最終報告書でお知らせ





期間中 速報

期間終了後 最終報告書

速報は、重大なリスクを見つけた時点で期間の終了を待たずにお送りします。

記載内容

- **発見資産の一覧**：資産探索で見つかった攻撃対象
- **判定と目的**：侵入・情報取得が達成できたか
- **キルチェーン**：初期侵入から目的達成までの手順
- **システム構成図**：攻撃経路を図で可視化
- **取得できた情報**：実際に持ち出せたデータ
- **漏れていた機微情報**：公開情報から見つけたパスワードなど

実施にあたり、2点のご了承・ご承認をいただきます



本番環境への影響のご了承

本番環境に対して、攻撃者と同じように実際に侵入を試みます。極力影響がないように実施いたしますが、本番環境に影響が出る可能性があることをご了承いただきます。また、データのバックアップ状況のご確認を推奨します。



対象範囲のご承認

当社が見つけた資産の一覧をお送りします。お客様には内容をご確認いただき、一覧内への攻撃をご承認いただきます。この際、攻撃対象から外したい対象（例：自社資産でない対象）があれば、除外いただくこともできます。

GMO Flatt Security

守る側は、どう変わるべきか（これから）

常時の検証と、ソースコードの活用

AIペンテストを、30日間/ワンショットではなく24時間365日へ

- 高まる脅威に対抗するには、抜き打ちテストのようなAIペンテストを継続的に実施する必要がある
- 期間を区切った点検ではなく、24時間365日の検証へ

不正アクセス抜き打ちテスト 年間プランとしてご提供予定

攻撃者にない優位、ソースコードを活かす

- 攻撃者にない守る側の優位は、ソースコードを活用できること
- ソースコードを使えば、効率的かつ網羅的にリスクを洗い出せる
- ペンテストの「次」として、グレーボックス形式の脆弱性診断を高頻度に実施することも重要



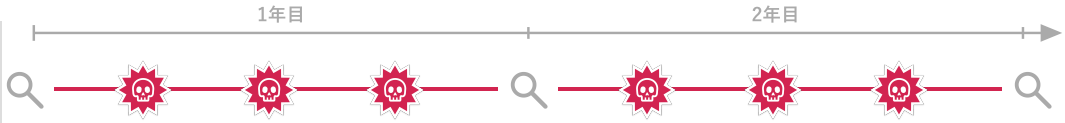
ソースコードで
網羅的にリスクを検出
（ホワイトボックス診断）

悪用可能か
実環境で判定
（ブラックボックス診断）

Takumiにより、精度・頻度共に十分な脆弱性診断を実現

手動脆弱性診断

精度：高い　回数：少ない



前回の診断から最大1年間、脆弱性が放置されるリスク

非AI DAST/SASTツール

精度：低い　回数：多い



ロジックの脆弱性や認証・認可の不備は検出できない

AI-SAST / AI-DAST

精度：高い　回数：多い



ロジックの脆弱性や認証・認可の不備まで高精度に検出
1年に何度も、繰り返し実施できる

Takumi は、AIエージェントでセキュリティ対策を高レベルに自動化。
人間レベルの脆弱性検知・修正と開発時のマルウェア対策の両輪でサポート

1 脆弱性検知・修正



脆弱性診断
AI-SAST / AI-DASTで脆弱性を網羅的に検出



脆弱性の自動修正
検出した脆弱性の修正パッチを自動生成



AI ペネトレーションテスト
AIが脆弱性を悪用し深く侵入。
リスクの最大値を検証

2 開発時のマルウェア対策



Takumi Guard
悪性パッケージを検知してブロック



Takumi Runner
CI/CDの実行トレースを可視化し、
有事に備える



Takumi Images （β版）
既知脆弱性・マルウェア最小化
コンテナイメージ

GMO Flatt Security

会社名	GMO Flatt Security株式会社
所在地	〒150-8512 東京都渋谷区桜丘町26-1 セルリアンタワー6階
設立	2017年5月23日
代表者	代表取締役CEO 井手 康貴
事業内容	サイバーセキュリティ関連サービス
顧問弁護士事務所	法律事務所ZeLo
顧問税理士事務所	RSM汐留パートナーズ税理士法人

コーポレートミッション

エンジニアの背中を預かる

ソフトウェアプロダクトの開発組織と
そこで働くエンジニアにとって最適なセキュリティサービスを提供し、
「背中を預けられる」存在になることが
GMO Flatt Securityの使命です。


GMO Flatt Security
エンジニアの背中を預かる
より多くのエンジニアがものづくりに集中できる社会を、
セキュリティ面からつくる会社