GMO Flatt Security | Takumi by GMO



今からでも間に合う！

Claude Mythosで変わる脆弱性対策の基礎



Speaker

GMO Flatt Security
ソフトウェアエンジニア
梅内 翼

6.25（木）
12:00~12:40@online

登壇者・会社紹介

GMO Flatt Security

©2026 GMO Flatt Security Inc. All Rights Reserved.

自己紹介



梅内 翼（うめうち つばさ）

GMO Flatt Security株式会社 ソフトウェアエンジニア

- 2023年5月 GMO Flatt Security株式会社に入社
- Web アプリケーションや Firebase 等のセキュリティ診断の経験を経て、ソフトウェアエンジニアとしてセキュリティ診断AIエージェント「Takumi byGMO」の開発に従事
- 漫画・ドラマ『トリリオンゲーム』の技術監修
- 著書『詳解セキュリティコンテスト CTFで学ぶ脆弱性攻略の技術』(マイナビ出版)

GMO Flatt Security

© 2026 GMO Flatt Security Inc. All Rights Reserved.

開発・提供

GMO Flatt Security

バグハント × AI開発
世界トップクラスの
エンジニア集団

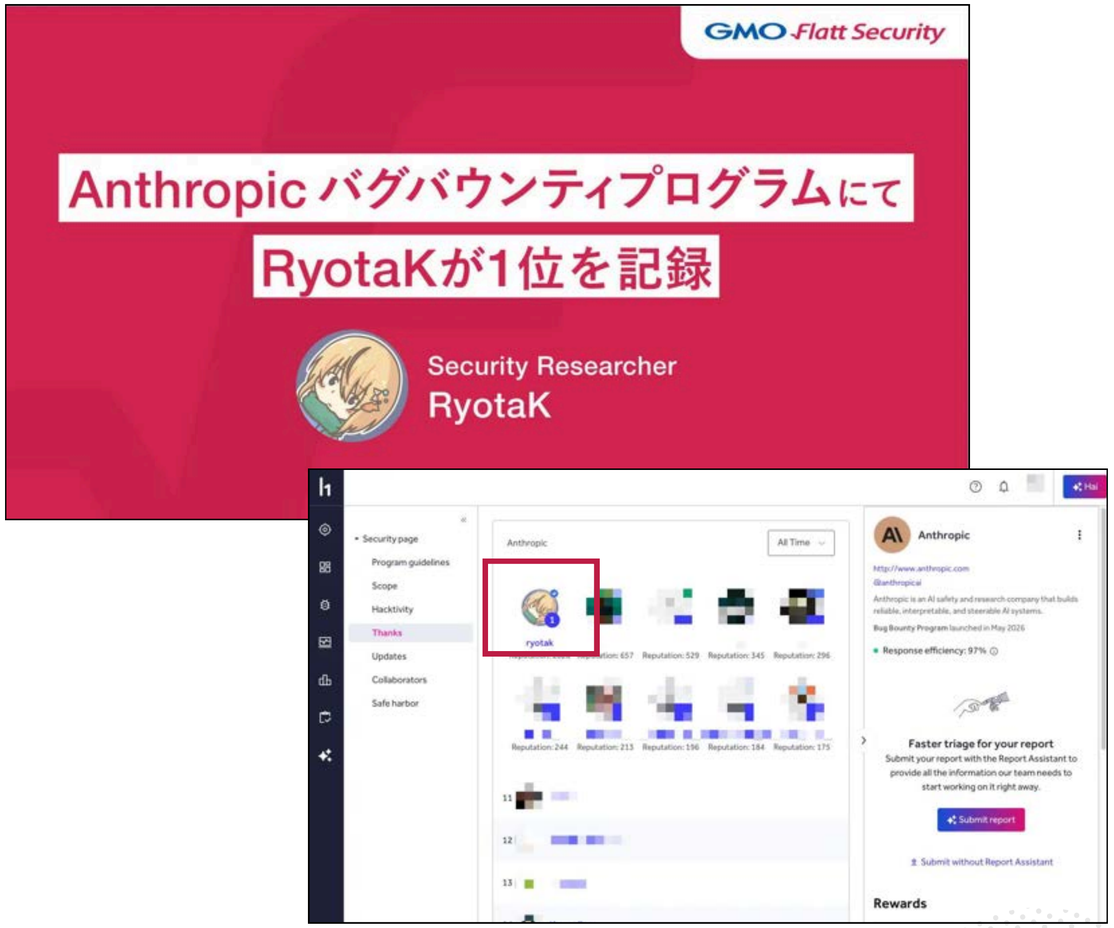
防衛省サイバーコンテスト(2023-2025) **3連覇**

Ubuntuへの脆弱性報告報奨金(単一の脆弱性) **$30,000**

その他脆弱性報告先実績
Google, Cloudflare, GitHub,
Salesforce, Microsoft, Meta

Canonical
at home
SBI DigiTrust
NTT DATA 株式会社NTTデータユニバーシティ
en
GLORY
CyberAgent
Cygames
SmartHR
Timee
TIER IV
TOKYO GAS
TREND MICRO
NIKKEI
PKSHA COMMUNICATION
pixiv
freee
Findy
FINATEXT HOLDINGS
Preferred Networks
MIZUHO みずほリサーチ&テクノロジーズ
MITSUI & CO. DIGITAL ASSET MANAGEMENT
MITSUBISHI ELECTRIC Changes for the Better 三菱電機インフォメーションシステムズ株式会社
LayerX

Anthropicに世界で最も脆弱性報告を行っている会社



「Anthropic バグバウンティプログラム」の順位表（2026年5月13日時点）

ClaudeをはじめとしたAnthropic製品に対して、世界で最も脆弱性を報告した研究者として評価

※ バグ報告プラットフォーム「HackerOne」にてAnthropic バグバウンティプログラム（Anthropicが自社製品への脆弱性報告を受け付け、報告者に対して金銭的な報酬を提供する取り組み）にて、弊社セキュリティ研究者が1位を記録

プレスリリース : https://prtimes.jp/main/html/rd/p/000000069.000027502.html

GMO Flatt Security

© 2026 GMO Flatt Security Inc. All Rights Reserved.

本ウェビナーのねらい

本日お話しする内容

- AI モデルと脆弱性診断の進化についての**技術的な解説**
- ポスト Mythos 時代を見据えた**脆弱性対策の考え方**とソリューション

本日お話ししない内容

- Claude Mythos 等のフロンティア AI モデルについての**政策的な動向**
- Claude Mythos を除く**各フロンティア AI モデルの具体動向**

今後も Claude Mythos やフロンティア AI モデルに関するウェビナーを開催予定です。
アンケートにて、今後のご要望をぜひお寄せください！

GMO Flatt Security

© 2026 GMO Flatt Security Inc. All Rights Reserved.

目次

1 脆弱性診断のこれまでとこれから

2 Claude Mythos の基礎: 登場から現在まで

3 今後、脆弱性対応において起こること

4 ポスト Mythos 時代を見据えた脆弱性対策

GMO Flatt Security

©2026 GMO Flatt Security Inc. All Rights Reserved.

Chapter 1

脆弱性診断のこれまでとこれから

GMO Flatt Security

©2026 GMO Flatt Security Inc. All Rights Reserved.

従来の脆弱性診断の流れ

一定期間内に、人間のセキュリティエンジニアがツール等を用いて
システムに潜む悪用可能なバグ・不備を網羅的に洗い出す取り組み



一般的な規模の Web アプリケーションでは短くても2週間〜1か月程度要する
見積もりや契約等の期間も考慮すると、**完了までに数ヶ月要するケースが大半**

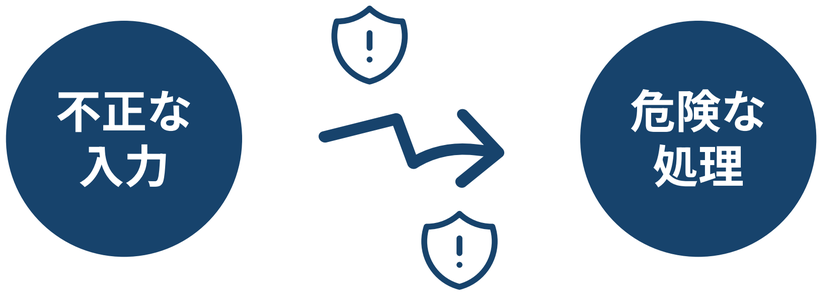
[1] OWASP Application Security Verification Standard (ASVS) (https://owasp.org/www-project-application-security-verification-standard/)

GMO Flatt Security

©2026 GMO Flatt Security Inc. All Rights Reserved.

従来の脆弱性スキャナ: SAST（Static Application Security Testing)

1. ソースコードの静的解析によって脆弱性を発見
2. 脆弱性のパターンを示す検査ルールに合致する箇所を脆弱性として特定



不正な入力が
無害化機構の影響を受けずに
危険な処理に伝播し得るか

```yaml
rules:
- id: jwt-go-none-algorithm
  [...]
  patterns:
  - pattern-either:
    - pattern-inside: |
        import "github.com/golang-jwt/jwt"
        ...
    - pattern-inside: |
        import "github.com/dgrijalva/jwt-go"
        ...
  - pattern-either:
    - pattern: |
        jwt.SigningMethodNone
    - pattern: jwt.UnsafeAllowNoneSignatureType
```

検査ルールに基づき、ソースコードを解析 [1]

[1] Go における JWT 処理の検査ルール (https://github.com/semgrep/semgrep-rules/blob/develop/go/jwt-go/security/jwt-none-alg.yaml) を一部改変

GMO Flatt Security

©2026 GMO Flatt Security Inc. All Rights Reserved.

従来の脆弱性スキャナ: DAST (Dynamic Application Security Testing)

1. アプリケーションをクロールし、**攻撃対象**になり得るエンドポイントを発見する
2. 発見したエンドポイントに対して**入力可能なパラメータを特定**する
3. 特定したパラメータに攻撃用のペイロードを埋め込んだ**リクエストを送信し、挙動を確認**する

診断対象のアプリケーション



①攻撃対象を特定

本来の HTTP リクエスト

```
POST /api/items HTTP/1.1
Host: example.com
Authorization: eyJ...

{
 "item": {
  "name": "Apple",
  "price": 100
 }
}
```

② 入力可能なパラメータを特定

攻撃用のペイロードリスト

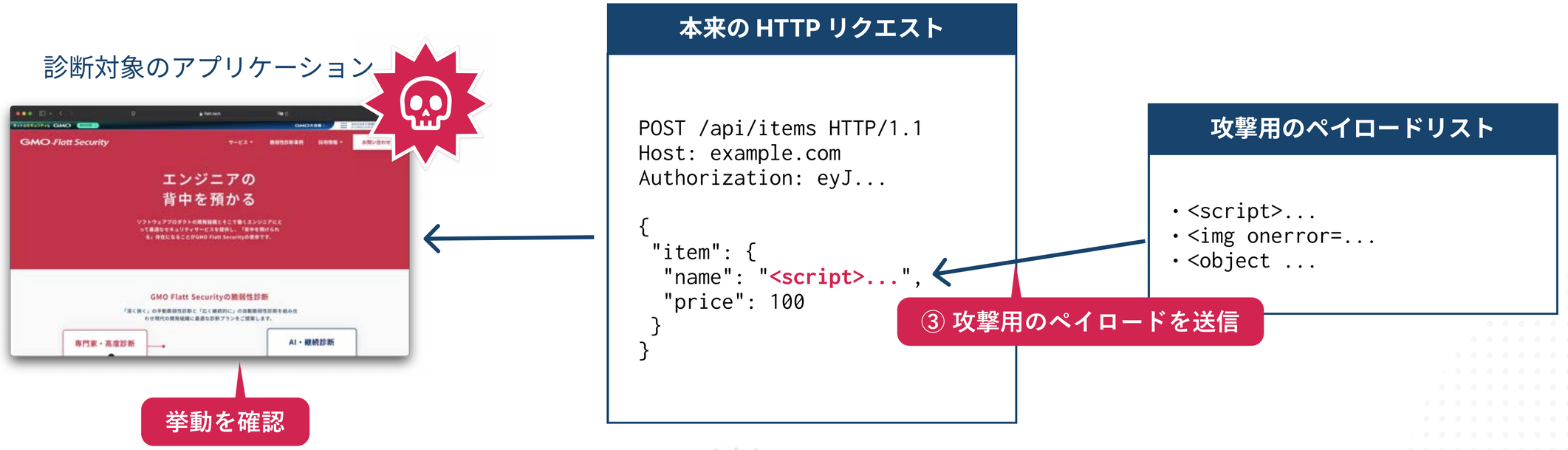
- <script>...
- <img onerror=...
- <object ...

GMO Flatt Security

©2026 GMO Flatt Security Inc. All Rights Reserved.

従来の脆弱性スキャナ: DAST (Dynamic Application Security Testing)

1. アプリケーションをクロールし、**攻撃対象**になり得るエンドポイントを発見する
2. 発見したエンドポイントに対して**入力可能なパラメータを特定**する
3. 特定したパラメータに攻撃用のペイロードを埋め込んだ**リクエストを送信し、挙動を確認**する



GMO Flatt Security

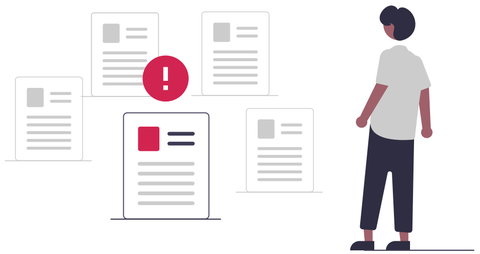
©2026 GMO Flatt Security Inc. All Rights Reserved.

従来の脆弱性スキャナの課題・限界

AI エージェントを活用しない既存の脆弱性スキャナには、診断能力に限界がある

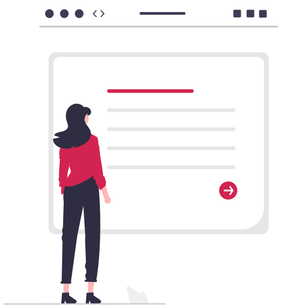
✕ SASTの課題

- 真に正確な検査ルールを
  定義・管理し続けることが困難
- 認識していない脆弱性に対する
  検査ルールを定義することができない



✕ DASTの課題

- 非典型的な脆弱性の検出が困難
- 対象のアプリケーションの挙動に
  基づいた適応的な診断ができない



GMO Flatt Security

©2026 GMO Flatt Security Inc. All Rights Reserved.

従来のセキュリティスキャナにおける課題 (SAST)

- **真に正確な検査ルールを定義・管理し続けることは困難**
  - ユーザーからの入力によって左右されるパラメータが、それを無害化する機構を通過せずに危険な処理まで到達していることを表すルールとして**正確に定義することは専門家であっても簡単ではない**
  - 一度ルールを定義しても、リアーキテクチャやリファクタリングによる影響を受ける

- **認識していない脆弱性に対する検査ルールを定義することができない**
  - 「問題のあるコードを発見した場合に、そのようなパターンを検出するためのルールを定義する」という**事後的なアプローチが基本**

GMO Flatt Security

©2026 GMO Flatt Security Inc. All Rights Reserved.

従来のセキュリティスキャナにおける課題 (DAST)

- **非典型的な脆弱性の検出が困難**
  - 「1人1回までしか使えないクーポンの制限を、短時間に複数回のリクエストを送信することで回避できる」のような**サービス固有のロジックによる脆弱性は、機械的な検出が困難**

- **対象のアプリケーションの挙動に基づいた適応的な診断ができない**
  - 人間の診断員はしばしば「このパラメータにこの値を送ったらこうなるから、このような値を送ったら脆弱性になるのでは」といった適応的なアプローチを採用する
  - 多くの DAST では単にペイロードを送信し、その際の挙動を観察するのみで、**仮説に応じたスキャン戦略の適応的な修正は基本的に行わない**

©2026 GMO Flatt Security Inc. All Rights Reserved.

なぜ AI エージェントを脆弱性の発見に利用するのか

AI エージェントを活用した脆弱性スキャナは、従来スキャナの課題を全て解決可能

- 真に正確な検査ルールを定義・管理し続けることは困難

→ **脆弱性の知識を与えられた AI エージェントは検査ルール・パターンを独自に構築する**

- 認識していない脆弱性に対する検査ルールを定義することができない

→ **脆弱性の知識を与えられた AI エージェントは未知の脆弱性も推論して発見できる**

- 非典型的な脆弱性の検出が困難

→ **AI エージェントは思考を通して何が脆弱性になり得るかを柔軟に判断する**

- 対象のアプリケーションの挙動に基づいた適応的な診断ができない

→ **AI エージェントはスキャン戦略をアプリケーションの挙動に応じて適応的に変化させる**

GMO Flatt Security

©2026 GMO Flatt Security Inc. All Rights Reserved.

AI を活用した脆弱性スキャナは、高速・高頻度で診断可能に

AI エージェントは人間の精度とツールの速さを両立し得るので、
脆弱性診断やセキュリティ業務に活用しない手はない



人間（ベンダー依頼）

診断期間：×数ヶ月

能力：◎適応的な診断が可能

高精度だが、時間がかかる

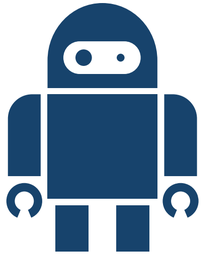


非 AI ツール

診断期間：◎1日

精度：×典型観点のみ

迅速だが、精度が低い



AI エージェント

診断期間：◎1日

能力：◎適応的な診断が可能

迅速かつ、高精度

GMO Flatt Security

©2026 GMO Flatt Security Inc. All Rights Reserved.

Chapter 2

Claude Mythos の基礎: 登場から現在まで

GMO Flatt Security

©2026 GMO Flatt Security Inc. All Rights Reserved.

Claude Mythos 以前のモデルの能力の進化

サイバーセキュリティ領域に限らず、遂行可能なゴールの抽象度が徐々に向上
抽象的な目標を与えるだけで様々なタスクが達成可能に

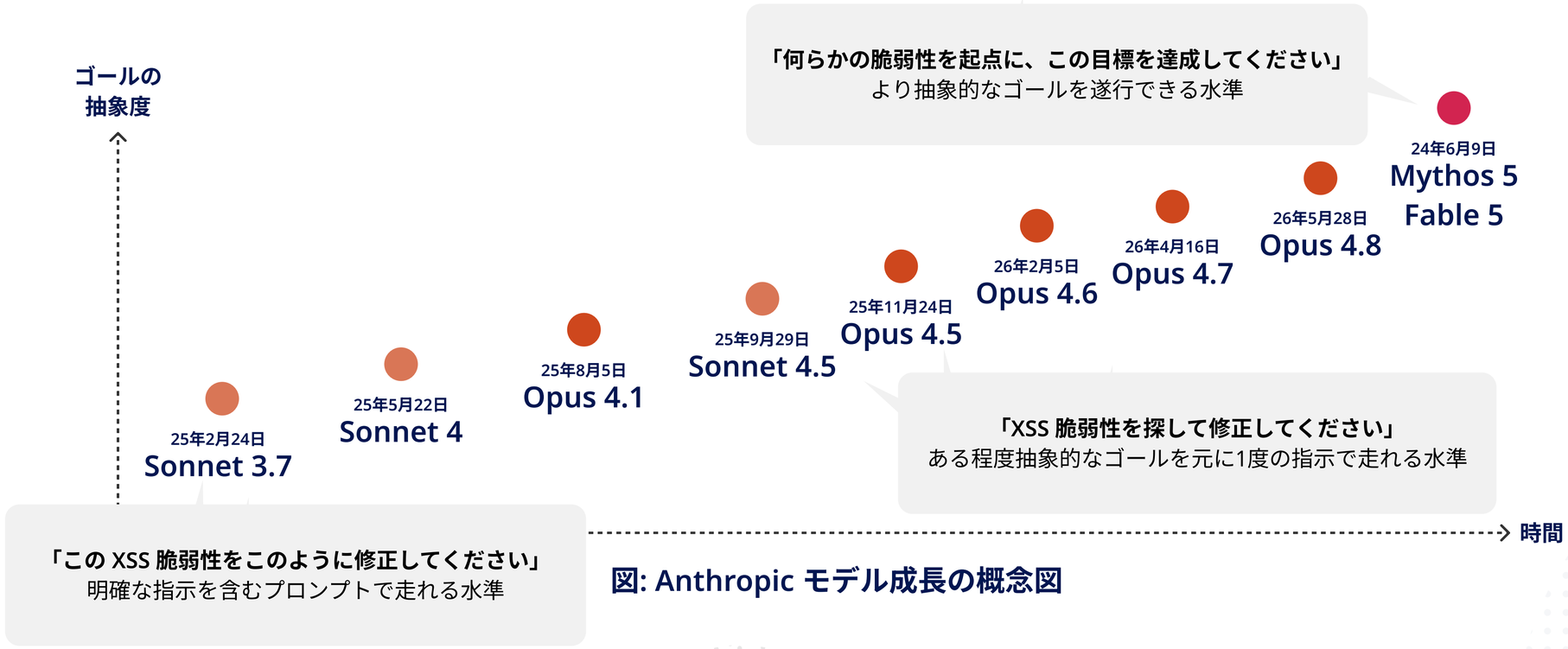


図: Anthropic モデル成長の概念図

GMO Flatt Security

©2026 GMO Flatt Security Inc. All Rights Reserved.

Claude Mythos 以前のモデルの脆弱性の発見能力

脆弱性の発見能力は、少なくとも1年前時点には実用化レベルに至っていた



オープンソースソフトウェアの
ゼロデイ脆弱性発見に実用化
（2025/3, 弊社）



脆弱性報告プラットフォームにて
人間のリサーチャーを抑え、AIが1位を獲得
(2025/6, XBOW)

GMO Flatt Security

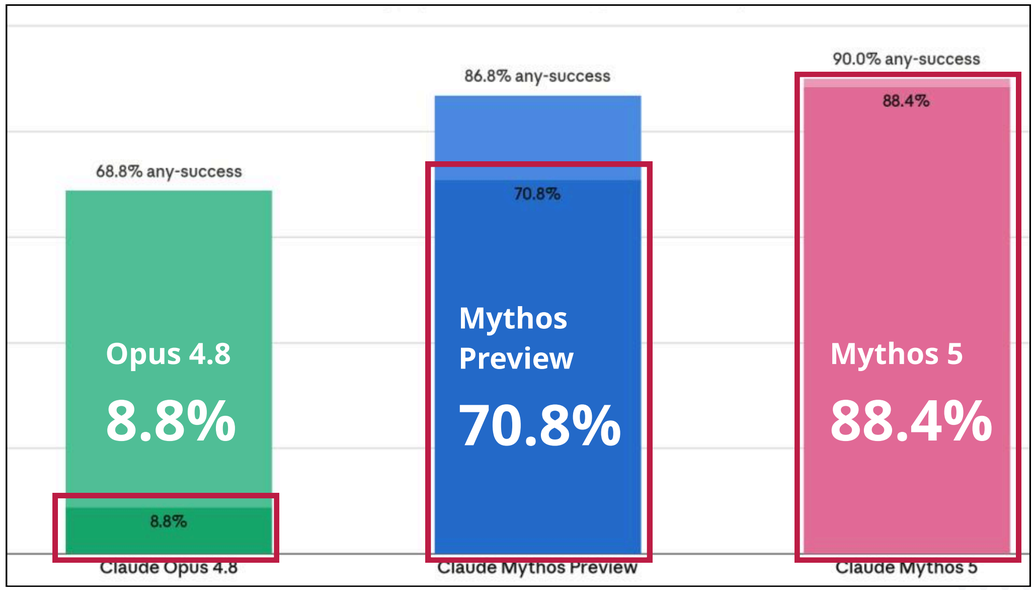
©2026 GMO Flatt Security Inc. All Rights Reserved.

Claude Mythos で進化した、脆弱性の悪用能力

2026年4月7日に発表されたMythos Preview および後継の Mythos 5では
脆弱性の発見能力に加えて悪用能力が大幅に強化 [1]

Anthropic は Firefox (ブラウザ) の
脆弱性の悪用試験を実施

- Opus 4.8 は **8.8%**と部分的な悪用コード生成に留まる
- Mythos 5 は **88.4%**とほぼ完璧に悪用コード生成に成功している



[1] より引用テキストを一部追記

[1] System Card: Claude Fable 5 & Claude Mythos 5 (https://www-cdn.anthropic.com/d00db56fa754a1b115b6dd7cb2e3c342ee809620.pdf)

GMO Flatt Security

©2026 GMO Flatt Security Inc. All Rights Reserved.

Project Glasswing の始動

脆弱性の発見・悪用可能性の高さを鑑み、
同時期に Anthropic が Project Glasswing [1] の始動を公表

- **Anthropic が立ち上げたサイバーセキュリティ強化のためのコンソーシアム**
  - Claude Mythos を Google, AWS, Apple 等のソフトウェア産業における主要企業・団体に限って配布
  - 強力なモデルが攻撃者の手に渡る前に、重要なソフトウェアやインフラ等の脆弱性を自律的に発見・修正することが目的
- **例：Mozilla Firefox における Claude Mythos 活用** [2]
  - Firefox では Mythos を活用して、Firefox 150 のリリースにおいて271件の脆弱性の修正を一括で完了したことを発表
- **2026年6月上旬には主要なインフラシステムを提供するベンダーへのアクセス提供が拡大**

[1] Project Glasswing (https://www.anthropic.com/glasswing/)
[2] Behind the Scenes Hardening Firefox with Claude Mythos Preview (https://hacks.mozilla.org/2026/05/behind-the-scenes-hardening-firefox/)

GMO Flatt Security

©2026 GMO Flatt Security Inc. All Rights Reserved.

Claude Fable 5 の公開と輸出制限

2026年6月9日に独自のセーフガードを組み込んだ Claude Fable 5 をリリース
しかし、6月12日に米国によって安全保障上の懸念から輸出制限が命令

- **Claude Fable 5の概要**
  - Project Glasswing パートナー限定モデルである Claude Mythos 5 に対して
    セーフガードが組み込まれたモデル (Mythos-class モデル)
  - サイバーセキュリティや生物学など悪用可能性のリスクが高い指示に対して、
    回答の拒否や Claude Opus 4.8 へのフォールバックを行う (セーフガード)

- **非常に厳格なセーフガードにより、サイバーセキュリティ用途でほぼ利用不可**
  - Fable 5ではいわゆる善意の防御目的で依頼された指示もほぼすべて拒否される状況に
  - 弊社内で実施した実験においても、ほぼすべての脆弱性発見タスクをモデルが遮断

- **セーフガード回避の実証報告により、米国外への提供が停止される事態に**
  - 2026年6月12日より全ユーザーの利用を停止しており、6月25日現在も利用ができない状態

GMO Flatt Security

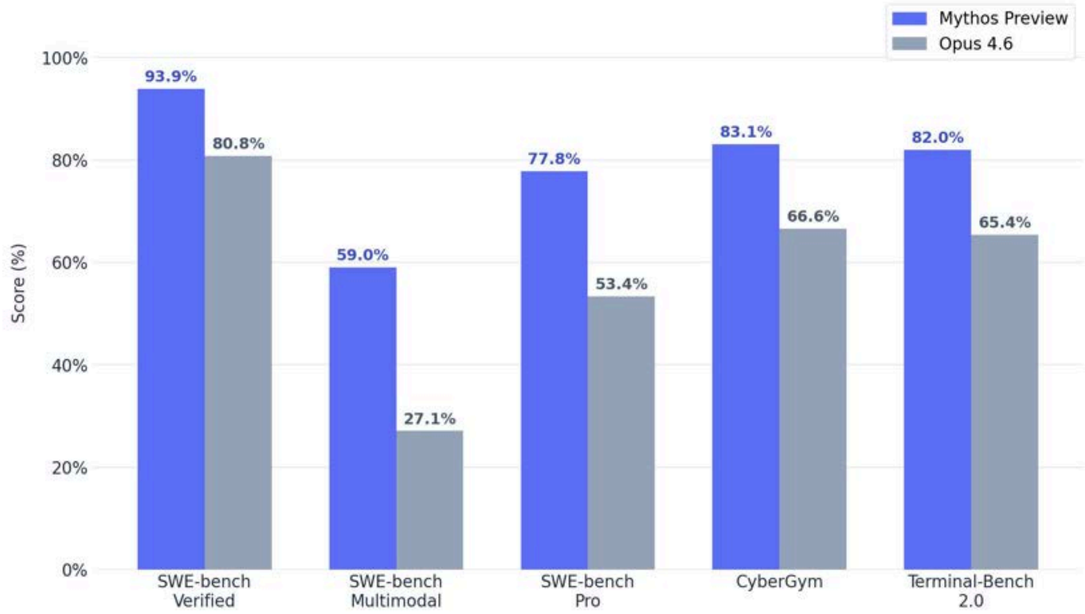
©2026 GMO Flatt Security Inc. All Rights Reserved.

Claude Mythos の性能

脆弱性悪用コードの実装など、一部の能力は大きく向上していると報告されているものの
ベンチマークスコア上ではあくまで漸進的な進歩に留まっている

例えば下記の能力は、Opus 4.6 との間に極端な差が存在するわけではない

- ソフトウェア課題の自律的解決 [1]
- 脆弱性分析タスク [2]
- ターミナル環境における習熟度 [3]



※ Project Glasswing (https://www.anthropic.com/glasswing/) の内容をもとに作成

[1] SWE-bench Verified, SWE-bench Multimodal, SWE-bench Pro, SWE-bench (https://www.swebench.com/)
[2] CyberGym (https://www.cybergym.io/)
[3] Terminal-Bench 2.0, Terminal-Bench (https://www.tbench.ai/)

GMO Flatt Security

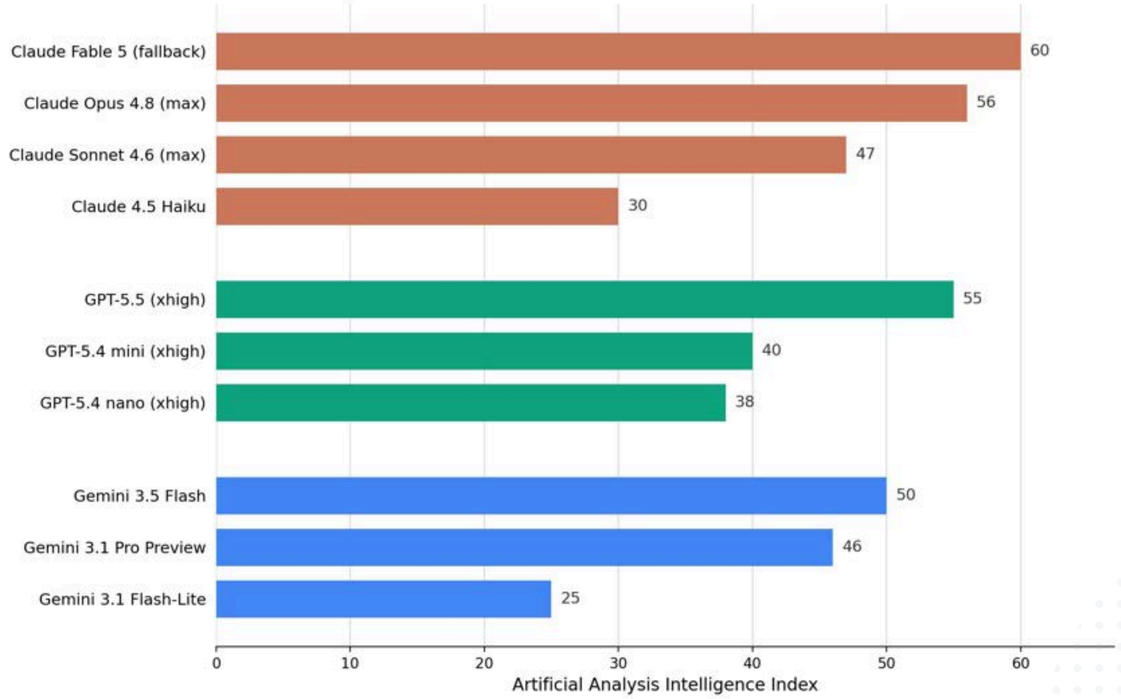
©2026 GMO Flatt Security Inc. All Rights Reserved.

主要プロバイダごとのベンチマークの比較

性能トップは Claude Fable 5 だが、競合プロバイダの主要モデルと性能は拮抗している
喧伝されているような脅威をもたらしうるモデルは Mythos / Fable に限らない

主要モデルの性能は同様に高く、Anthoropic 社のモデルのみが特出している訳ではない

- **Claude Fable 5** : 60
- **GPT-5.5** : 55
- **Gemini 3.5 Flash** :50



※各スコアは9つの指標をベースに計算された値（GDPval-AA v2, τ³-Banking, Terminal-Bench v2.1, SciCode, AA-LCR, AA-Omniscience, Humanity's Last Exam (HLE), GPQA Diamond, CritPt）
2026/06/23 時点の https://artificialanalysis.ai/leaderboards/models の結果をベースに作成

GMO Flatt Security

©2026 GMO Flatt Security Inc. All Rights Reserved.

Claude Mythos 利用者からの懐疑的な評価

Claude Mythos でなくとも高い精度での脆弱性の発見自体は可能であり、
既存の主要モデルでも十分に対策が可能という評価も

- **cURL 開発者による「マーケティングスタント」であるという批判** [1]
  - Mythos によるソースコード解析の結果、cURL に対して報告された脆弱性は低深刻度のものが 1 件だけで、その他はすべて誤検知か単なるバグだった
  - Mythos のみが特段脆弱性の発見において優れているモデルであるわけでないと主張
- **AISLE Security による Claude Mythos の評価** [2]
  - Mythos が発見した脆弱性を他の小型のモデルでテストした際に、脆弱性の発見能力についてはほぼ同等の分析結果に
  - 様々なモデルで評価したところ、各モデルのサイバーセキュリティに関する能力はモデルのアクティブパラメータ数やトークンあたりの価格にきれいに比例しなかった

[1] Anthropic's bug-hunting Mythos was greatest marketing stunt ever, says cURL creator (https://www.theregister.com/security/2026/05/11/anthropics-bug-hunting-mythos-was-greatest-marketing-stunt-ever-says-curl-creator/5238111) より (※ 開発者本人が Mythos にアクセスしてスキャンを実行したわけではないことに注意)
[2] AI Cybersecurity After Mythos: The Jagged Frontier (https://aisle.com/blog/ai-cybersecurity-after-mythos-the-jagged-frontier)

GMO Flatt Security

©2026 GMO Flatt Security Inc. All Rights Reserved.

先端モデル活用にはハーネスやパイプラインの整備が重要

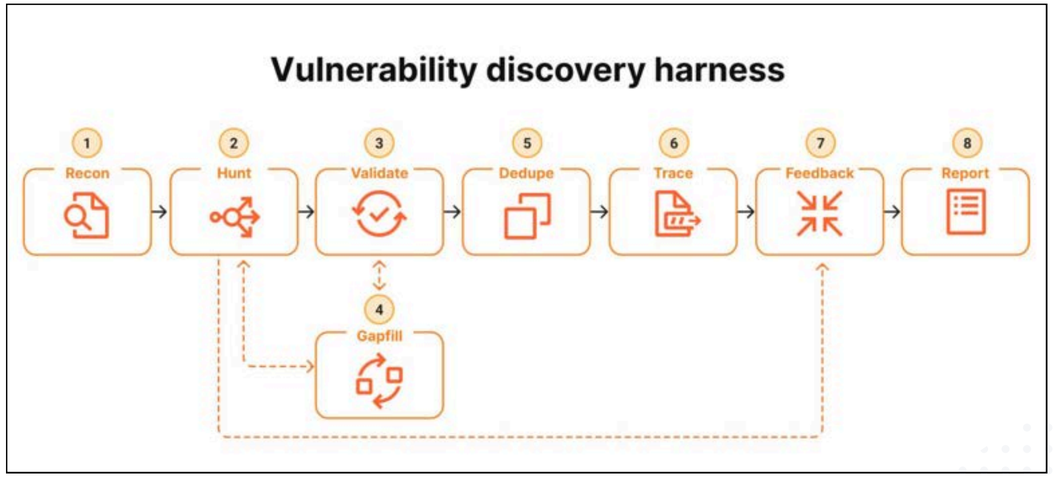
Firefox の事例 [1] やProject Glasswing に参加するCloudflare のレポート [2] では、
モデル単体の利用では不十分で、ハーネスやパイプラインの重要であると言及

ハーネス (脆弱性発見を支える足回り)

- 仮説検証用のテストケースを自動生成
- 既存のファジング基盤の上で脆弱性の発見を並列化

パイプライン (脆弱性発見後の対応フロー)

- 見つかった脆弱性をトリアージし、修正パッチのリリースまでを自動化



Cloudflare では、AI エージェントの精度や効率向上のため
上記の工程を行うハーネス・パイプラインを構築している [2]

[1] Behind the Scenes Hardening Firefox with Claude Mythos Preview (https://hacks.mozilla.org/2026/05/behind-the-scenes-hardening-firefox/)
[2] Project Glasswing: what Mythos showed us (https://blog.cloudflare.com/cyber-frontier-models/)

GMO Flatt Security

©2026 GMO Flatt Security Inc. All Rights Reserved.

セキュリティAI エージェントにおけるハーネスとは何か

ハーネスとは、モデルの動作や外部システムとの接続を制御・最適化するもの
高性能なモデルがあっても、それ単体で安全かつ効率的に脆弱性を発見できるわけではない

コンテキスト制御

コンテキストウィンドウに合わせてプロンプトや知識をモデルの会話履歴に投入したり、上限到達前にコンテキストの圧縮を行ったりするための仕組み

安全な動作環境

発見したバグを検証するためのエクスプロイトコードをビルド・実行する際に利用される、使い捨ての安全なサンドボックス環境

ツールの付与

Web ブラウザやターミナル、デバッガ、ファザーなど、AI エージェントが脆弱性を発見したり、PoC が実際に動くかどうかを検証したりするためのツール

オーケストレーション

タスクをエージェントが処理しやすい形式に分解したり、ロールや実行順序を制御したりすることで、複数の AI エージェントを協調的に動作させる仕組み

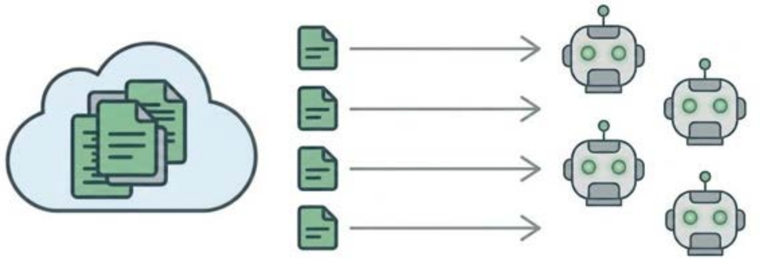
GMO Flatt Security

©2026 GMO Flatt Security Inc. All Rights Reserved.

セキュリティ AI エージェントとハーネスの具体例 (1/2)

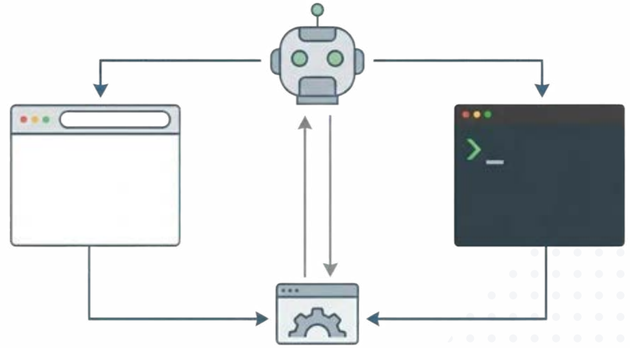
機能レベルでの分割

- 大規模なコードベースやシステムを特定の単位で分割することで
  特定の AI エージェントが検査の対象とする範囲を制限する
- 調査する価値のあるコンポーネント・機能のみを検査対象にする



適切なツールの付与

- AI エージェントが脆弱性を発見・検証したり、
  特定のコードを実行したりするのに必要十分な
  ターミナル環境やブラウザアプリケーション等を付与する



GMO Flatt Security

©2026 GMO Flatt Security Inc. All Rights Reserved.

セキュリティ AI エージェントとハーネスの具体例 (2/2)

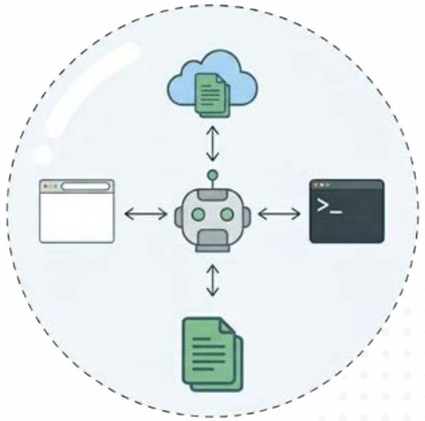
ナレッジベースの提供

- AI エージェントが脆弱性の発見や検証に必要な
  脆弱性そのものの知識やパターンに関するナレッジベース、
  システムの仕様等に関するナレッジベースへのアクセスを付与する



サンドボックス環境

- AI エージェントが自由に (危険な) 実験用のコードや
  エクスプロイトコードを安全に実行するための環境を付与する
- 自身が生成したエクスプロイトコードや修正用パッチの
  正当性を検証するために必要不可欠



GMO Flatt Security

©2026 GMO Flatt Security Inc. All Rights Reserved.

Chapter 3

今後、脆弱性対応において起こること

GMO Flatt Security

©2026 GMO Flatt Security Inc. All Rights Reserved.

この先、脆弱性対応はどうなるのか？

Forrester のリサーチ [1] によると主に3つのフェーズが訪れると予測されている

直近	脆弱性発見のコスト低下に伴う脆弱性情報の氾濫と価値低下
～ 1年半	脆弱性管理システムの破綻と 脆弱性発見を軸としたスキームからの脱却
～ 5年	AI を中心としたセキュリティガバナンスの確立 セキュリティエンジニアに求められる知識・スキルの変化

[1] Project Glasswing: The 10 Consequences Nobody's Writing About Yet (https://www.forrester.com/blogs/project-glasswing-the-10-consequences-nobodys-writing-about-yet/) を参考に作成

GMO Flatt Security

©2026 GMO Flatt Security Inc. All Rights Reserved.

この先、脆弱性対応はどうなるのか？

Forrester のリサーチ [1] によると主に3つのフェーズが訪れると予測されている

直近	**脆弱性発見のコスト低下に伴う脆弱性情報の氾濫と価値低下**
～ 1年半	脆弱性管理システムの破綻と 脆弱性発見を軸としたスキームからの脱却
～ 5年	AI を中心としたセキュリティガバナンスの確立 セキュリティエンジニアに求められる知識・スキルの変化

[1] Project Glasswing: The 10 Consequences Nobody’s Writing About Yet (https://www.forrester.com/blogs/project-glasswing-the-10-consequences-nobodys-writing-about-yet/) を参考に作成

GMO Flatt Security

©2026 GMO Flatt Security Inc. All Rights Reserved.

脆弱性発見のコスト低下に伴う脆弱性情報の氾濫と価値低下

- **高いサイバーセキュリティの能力を持った最先端モデルの一般化やハーネスの拡充によって、「脆弱性を発見」するためのコスト・時間が低下し、脆弱性の発見が大衆化**
  - 昨年に米国 DARPA (国防高等研究計画局) が開催した、AI による脆弱性発見＆修正大会 (‘25)での脆弱性発見・修正単価は、既に平均 $152 と安価化 [1]
  - 「セキュリティベンダに定期的に脆弱性診断を依頼する」といったスピード感のもとで行われる脆弱性の発見・トリアージが手遅れになり得る

- **OSS やバグバウンティプラットフォームにおいて脆弱性対応の限界を迎える**
  - 代表的なプラットフォームでは、2021 年以降有効なレポート数が 20% 増加 [2]
  - 加えて、AI による出力は「脆弱性に見えるが実際には脆弱性ではないもの 」も多く含まれ得るため、正確な検証が不可欠
  - 結果、最先端のモデルにより脆弱性の発見はスケール・高速化していくが、それをトリアージ・修正する人間の能力はスケールせず限界を迎える

[1] aicyberchallenge.com (https://aicyberchallenge.com/)
[2] HackerOne: Hacker-Powered Security Report 9th Edition (https://www.hackerone.com/resources/hacker-powered-security)

GMO Flatt Security

©2026 GMO Flatt Security Inc. All Rights Reserved.

この先、脆弱性対応はどうなるのか？

Forrester のリサーチ [1] によると主に3つのフェーズが訪れると予測されている

直近	脆弱性発見のコスト低下に伴う脆弱性情報の氾濫と価値低下
～ 1年半	**脆弱性管理システムの破綻** **脆弱性発見を軸としたスキームからの脱却**
～ 5年	AI を中心としたセキュリティガバナンスの確立 セキュリティエンジニアに求められる知識・スキルの変化

[1] Project Glasswing: The 10 Consequences Nobody's Writing About Yet (https://www.forrester.com/blogs/project-glasswing-the-10-consequences-nobodys-writing-about-yet/) を参考に作成

GMO Flatt Security

©2026 GMO Flatt Security Inc. All Rights Reserved.

脆弱性管理システムの破綻と現在のスキームからの脱却 (1/2)

脆弱性管理システム (CVE 等) をベースとしたスキームが大量の脆弱性報告によって破綻

- AI が大量の 0-day 脆弱性を発見することで、既存の脆弱性管理システムが持つキャパシティを超過
  (脆弱性が発見・報告されてから正式に発表されるまでの期間が長くなる)
- 実際、2026年4月 NIST から CVE の増加に対応するためのポリシーの変更も発表されている [1]

＋

SCA の破綻

既知の脆弱性の登録が遅れたり、あるいは、登録すらされなくなるため、
バージョン情報等に基づいた脆弱性の対応が困難になる

開示ルールの破綻

脆弱性発見から公開までの修正期間に猶予を与えるルールが存在するが、
AI による脆弱性発見の高速化によって、**修正が追いつかなくなる** [2]

パッチ SLA の破綻

パッチ公開までの期間に対する SLA は、攻撃者がエクスプロイトを開発する
までの期間をベースに設定されていたが、これも**高速化されるため破綻する** [2]

[1] NIST Updates NVD Operations to Address Record CVE Growth (https://www.nist.gov/news-events/news/2026/04/nist-updates-nvd-operations-address-record-cve-growth)
[2] AI Vulnerability Discovery: The New Defender Economics (https://bloo.io/resources/articles/ai-vulnerability-discovery)

GMO Flatt Security

©2026 GMO Flatt Security Inc. All Rights Reserved.

脆弱性管理システムの破綻と現在のスキームからの脱却 (2/2)

「脆弱性の発見」を主軸とした現在のスキームからの脱却

1. リリース前や定期的に脆弱性診断を実施して、見つかったものを修正する
2. 攻撃者に先立って脆弱性を発見して、攻撃される前に修正する

脆弱性を誰でも安価かつ高速に発見できる世界においては、
単に脆弱性を発見することの価値が薄れていく

- 自組織やプロダクトにまつわるビジネス上のコンテキストを考慮した戦略が不可欠
- 発見された脆弱性の候補をどのように優先付け・検証した上で、
  高速かつ正確にパッチ実装・リリースまでのフローを実現できるかが焦点に

GMO Flatt Security

©2026 GMO Flatt Security Inc. All Rights Reserved.

この先、脆弱性対応はどうなるのか？

Forrester のリサーチ [1] によると主に3つのフェーズが訪れると予測されている

直近	脆弱性発見のコスト低下に伴う脆弱性情報の氾濫と価値低下
～ 1年半	脆弱性管理システムの破綻と 脆弱性発見を軸としたスキームからの脱却
～ 5年	**AI を中心としたセキュリティガバナンスの確立** **セキュリティエンジニアに求められる知識・スキルの変化**

[1] Project Glasswing: The 10 Consequences Nobody’s Writing About Yet (https://www.forrester.com/blogs/project-glasswing-the-10-consequences-nobodys-writing-about-yet/) を参考に作成

GMO Flatt Security

©2026 GMO Flatt Security Inc. All Rights Reserved.

攻撃者の視点から考える脆弱性対応の未来

攻撃者はAI を用いて脆弱性発見・悪用をスケールさせていく
防御側も AI を活用したスケーラブルな脆弱性対応のための仕組みの構築が不可欠

- **攻撃者の目的はこれからも変わらない**
  - AI の高度化によって攻撃に関するコストは低下するが、それが攻撃者の目的を変えるわけではない
  - 脆弱性を突くことで利益が得られる構造が成立する領域 (ランサムウェア、ソフトウェアサプライチェーン) において、攻撃者の数やその能力がスケールする恐れがある
- **「脆弱性発見の大衆化」がすぐそこまで来ている**
  - 強力なモデルの出現により、脆弱性発見・悪用のコスト・難易度も低下している
  - 強力な道具を手に入れた攻撃者に対抗する手段として、**防御側も AI を活用した自律的なトリアージプロセスを確立することが必須**
- **攻撃者は自身に利益がある限り AI に投資する**
  - 攻撃者による高度なモデルへの投資は、それ以上の**利益を生む合理的な投資**になりつつある
  - 防御側はセキュリティへのコスト投入を「事業継続を担保するための投資」として捉え直す必要がある

GMO Flatt Security

©2026 GMO Flatt Security Inc. All Rights Reserved.

AIを中心とした時代の脆弱性対応における主要な論点

1. Vibe Coding の隆盛により脆弱性が埋め込まれる機会も増加

- Claude Code 等によるコーディングエージェントの隆盛に伴い、誰もがコードを書ける時代に、
- AI が生成するコードは、当然常にセキュアであるわけではない [1]

2. 最も深刻な課題は、脆弱性の発見速度と人間によるトリアージ速度のギャップ

- AI が瞬時に 0-day を発見する一方で、人間による修正・検証・リリースのサイクルがボトルネック化
- AI の発見速度に追従するような**自律的なトリアージの仕組みの構築が不可欠**
- 引き続き、修正パッチをリリースするための運用体制の構築も重要に

3. AI が発見する脆弱性はすべてが重要であるわけではない

- AI は客観的に脆弱性を見つけ出すが、**組織やプロダクト特有のコンテキストを理解しているわけではない**
- 組織・プロダクトごとに脅威モデリングやビジネスに対するインパクト分析を実施し、
  適切なリスクスコアリングに基づいたトリアージ体制を整える必要がある

[1] AI Coding Security Vulnerability Statistics 2026: Alarming Data (https://sqmagazine.co.uk/ai-coding-security-vulnerability-statistics/)

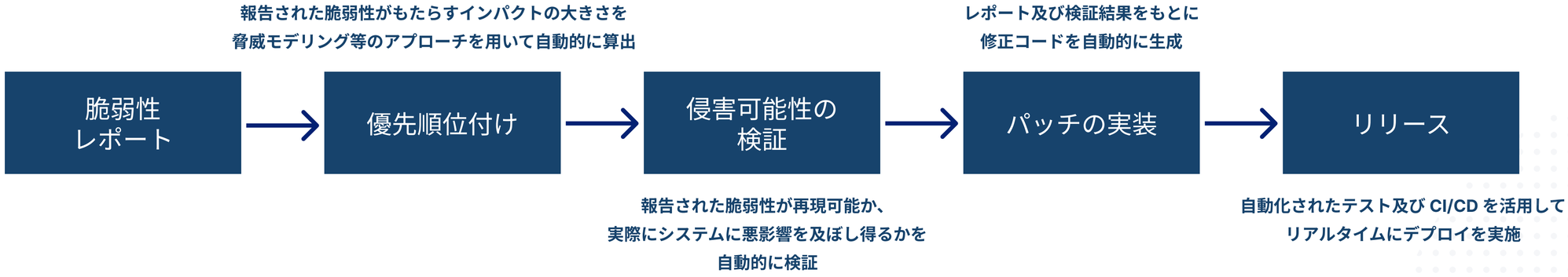
GMO Flatt Security

©2026 GMO Flatt Security Inc. All Rights Reserved.

論点2: 脆弱性の発見速度と人間によるトリアージ速度のギャップ

見つかった脆弱性を人間が確認して、優先順位付けして、修正するという
従来の基本的な脆弱性対応が手遅れになり得る

- **優れたモデルは高速かつ大量に脆弱性を発見し、それを悪用するコードを実装し得る**
  - Firefox の事例では、脆弱性発見から1時間以内にエクスプロイトコードが実装された例も存在 [1]
  - 脆弱性が公開されてから最初の攻撃が確認されるまでの平均時間は8時間に短縮 [2]
  - 一方で、ベンダーによるパッチ適用ウィンドウは往々にして30~90日程度要する



[1] Measuring LLMs' impact on N-day exploits (https://www.anthropic.com/research/n-days)
[2] Zero Day Clock (https://zerodayclock.com/)

GMO Flatt Security

©2026 GMO Flatt Security Inc. All Rights Reserved.

論点3: 自組織・プロダクトに即した脆弱性のトリアージ

AI は高速で脆弱性を発見するが、そのすべてが対応する必要のあるものとは限らない

- **AI はあくまで客観的にシステムの脆弱性を発見する**
  - 典型的 (技術的)な脆弱性や、仕様に反するリスキーな挙動を脆弱性として検出する
  - 特定の組織・プロダクトに応じたチューニングが行われていない限り、
    ある脆弱性がもたらすビジネス上のインパクトを正確に理解しているわけではない
  - 資産管理や脅威モデリングを通して、被攻撃時にインパクトをもたらし得る資産や、
    具体的な侵害のシナリオについて想定しておくことが不可欠

- **最も重要な資産を中心とした脆弱性検出・対応アプローチへの転換**
  - すべての機能をまんべんなく検査し、検出された脆弱性をひとつずつ対応するのではなく、
    侵害されるとインパクトの大きい資産・パスを特定しておき、その可能性を検討する
  - いわゆる従来型の網羅的な診断から、目的指向のペネトレーションテストへの転換が必要

GMO Flatt Security

©2026 GMO Flatt Security Inc. All Rights Reserved.

Chapter 4

ポスト Mythos 時代を見据えた脆弱性対策

GMO Flatt Security

©2026 GMO Flatt Security Inc. All Rights Reserved.

ポスト Mythos 時代を見据えた技術的なスタンスの転換

1. AI を活用した高速・高精度な脆弱性対応の実現

- 脆弱性を発見・優先付けし、検証し、必要なパッチを実装・適用するという
  脆弱性管理のライフサイクルを高速・高精度で実現させていくことが必要
- AI をこのパイプライン全体でうまく活用し、スケールするような仕組みを構築する必要がある

2. スポット型の検査から継続的な検証モデルへの移行

- 継続的に最新のコード・動作環境をスキャンし続ける継続的な検証モデルに移行する
- 診断時点でセキュアであるだけでなく、継続的にセキュアであることを検査している体制自体が必要

3. 予防モデルから封じ込めモデルへの移行

- 「脆弱性を事前に発見する」だけの対応モデルは近い将来に限界を迎える
- 攻撃が成立してもシステム全体が壊滅的な状態に陥らないように、
  システムのマイクロセグメンテーション等に基づく Blast-radius (被害拡大範囲) の抑制が重要

GMO Flatt Security

©2026 GMO Flatt Security Inc. All Rights Reserved.

ポスト Mythos 時代を見据えた組織的・戦略的なスタンスの転換

1. セキュリティ対策への投資に対する見直し

- 前述の通り、攻撃者は AI に対する課金額を上回る利益が得られるのであれば、進んで投資する
- 改めてセキュリティ対策を「コスト」から「事業継続のための投資」と捉え直す必要がある [1]

2. セキュリティ人材の要件に対する見直し

- 従来セキュリティエンジニアの主たるスキルであった「脆弱性を見つけること」の価値が失われる
- 自組織やプロダクトにとってビジネス上の大きなインパクトをもたらす脆弱性を判別したり、それに基づいてパッチ実装やリリース戦略を計画的に行うための意思決定力が求められる

3. 組織的な対応力・レジリエンスの醸成

- 自社のシステム・プロダクトに関与するすべてのメンバーが、AI によって高度化・高速化されるサイバー攻撃への対応を円滑に実施できるように訓練しておく必要がある

[1] 国家サイバー統括室 - Project YATA-Shield (https://www.cyber.go.jp/pdf/press/20260518_AI_CS_gaiyou.pdf)

GMO Flatt Security

©2026 GMO Flatt Security Inc. All Rights Reserved.

セキュリティAI エージェント「Takumi byGMO」による脆弱性対応



脆弱性の検出と修正を高精度・高速に行う セキュリティ特化のAIエージェント

Takumiは、”今”開発組織が備えるべきリスクの解消をご支援。AIによる**高度なセキュリティ診断・ペネトレーションテスト・脆弱性の自動修正**等をご提供。

月額7万円(税抜)

atama+ Azoom Kanmu ゲームエイト Game8 KOMEHYO HOLDINGS SORACOM _KNOWLEDGE WORK NIKKEI

PartnerProp HACOBU freee Findy FINATEXT HOLDINGS prime Number PLAID MITSUI & CO. DIGITAL ASSET MANAGEMENT

GMO Flatt Security

©2026 GMO Flatt Security Inc. All Rights Reserved.

Takumi を活用した高速・高精度な脆弱性対応 (1/2)

ソースコード・実アプリケーションの双方に対応した高速かつ高精度な脆弱性検出を実現

AI 脆弱性診断機能

- AI エージェントが網羅的に疑似攻撃を行い、システムの脆弱性を洗い出す
- 最適なモデルを独自のハーネスのもとで実行



AI ペネトレーションテスト機能

- 複数の脆弱性を組み合わせた悪用可能性を検証
- 高性能なモデルの性能を活かし、スコープとゴールの設定のみで、複雑なシナリオも調査可能



※ 2026/06/25 現在、段階的に提供開始中

GMO Flatt Security

©2026 GMO Flatt Security Inc. All Rights Reserved.

Takumi を活用した高速・高精度な脆弱性対応 (2/2)

発見された脆弱性の再現可能性を迅速に検証可能
システムのコンテキストを考慮した対応方針を自動で判断

AI 脆弱性検証・修正機能

- 発見された脆弱性が実際に再現可能か自動で検証
- リポジトリと連携して自動で修正提案を作成



修正パッチのAIトリアージ機能

- Dependabot が作成した修正パッチの提案を精査し対応の必要があるかをリアルタイムに判断



GMO Flatt Security

©2026 GMO Flatt Security Inc. All Rights Reserved.

専門のエンジニアによるセキュア開発コンサルティング

開発プロセスに加え、運用プロセスにおける脆弱性の封じ込めを考慮した
セキュアなソフトウェア開発・運用を実現

- **脆弱性によるインパクトや影響範囲を抑えるためには、開発初期段階での検討が必須**
  - 脆弱性による影響を抑え込むためには、システムのアーキテクチャレベルでの検討が不可欠
  - 開発初期から専門のエンジニアによるレビューを実施することで、
    脆弱性を埋め込まないセキュアな開発と運用時における効果的な脆弱性の検知・封じ込めを両立

設計	• **システムの仕様やアーキテクチャをレビューし、事前に問題点を抽出** • **脅威モデリングを実施し、重点的に守るべきポイントを事前に洗い出し**
開発	• 設計段階で決定した内容に基づくセキュアな開発の実現
リリース・運用	• Takumi を活用した高速・高精度な脆弱性の検知と対応 • **運用後に明らかになった問題を解決するための中長期的なコンサルティング支援**

GMO Flatt Security

©2026 GMO Flatt Security Inc. All Rights Reserved.

本日のウェビナーのまとめ

GMO Flatt Security

©2026 GMO Flatt Security Inc. All Rights Reserved.

まとめ: Claude Mythos で変わる、脆弱性対策の基礎

- **AI エージェントによる診断は既存の脆弱性診断が抱えていた課題を解決する**
  - AI エージェントによる脆弱性診断は、従来はツールにより実現されていた迅速な診断の実施と、人間により実現されていた高い精度での検出・レポーティングという要件を同時に満たす

- **Claude Mythos は脆弱性の発見・悪用能力が高いが、既存主要モデルの能力でも十分対策に足る**
  - Mythos 以前のモデルも、高い脆弱性の発見能力を有しており、ベンチマーク上も極端な差はない
  - 加えて、脆弱性対策には高度な**モデル単体では不十分で、ハーネスの整備が重要である**

- **攻撃者の AI 活用に対して、守り手側も Takumi byGMO のような AI を活用する必要がある**
  - 脆弱性発見のためのコスト低下に伴い、CVE 等の既存の脆弱性管理システムが破綻しつつある
  - ベンダーへの定期的な診断依頼によってプロダクトのセキュリティを担保していく仕組みが手遅れになり得る
  - 継続的かつ高速・高精度な診断だけでなく、**攻撃者側の AI 投資同様に守り手側も AI を活用した対策が必須**

GMO Flatt Security

©2026 GMO Flatt Security Inc. All Rights Reserved.