Mythos報道の落とし穴

Atsushi Teraoka
寺岡篤志
NIKKEI Inc.

Mail : atsushi.teraoka@nex.nikkei.com

Linked in:

My articles:

NOTICE!!

※大いに私見が交じります。
※所属組織を代表しての意見ではありません。
※情報は収録時点のものです。

日本経済新聞記者　寺岡篤志

■取材テーマ
サイバーセキュリティ
（コーポレート、安全保障、犯罪、政策）
調査報道
情報工作
アンチマネーロンダリング
リスク管理

経歴
07年度　某大学ラグビー部卒
08年　日経入社
14年　マウントゴックス事件でセキュリティの取材を開始
22年　九州大SECKUN　３期生
24年　英国のChevening Cybersecurity Fellowshipに参加
25年　Fulbright Scholarshipフェローとして米メリーランド大でサイバー安全保障の研究に従事
26年　セキュリティベンダへ

Mythosの問題は正しく報道されているか？

去年からのにわかサイバー祭り

※件数は６月中旬まで

- **トヨタ自動車→記事タグなし**
- **KADOKAWA→記事タグなし**
- **証券口座乗っ取り→記事タグあり、１５９件**
- **アスクル→記事タグあり、１０５件**
- **アサヒGHD→記事タグあり、６９件**
- **Mythos→記事タグあり、１０９件**

参考：

「クマ被害」→５０５件
「スペースX」→３２２件
「ナフサ高騰」→２８０件

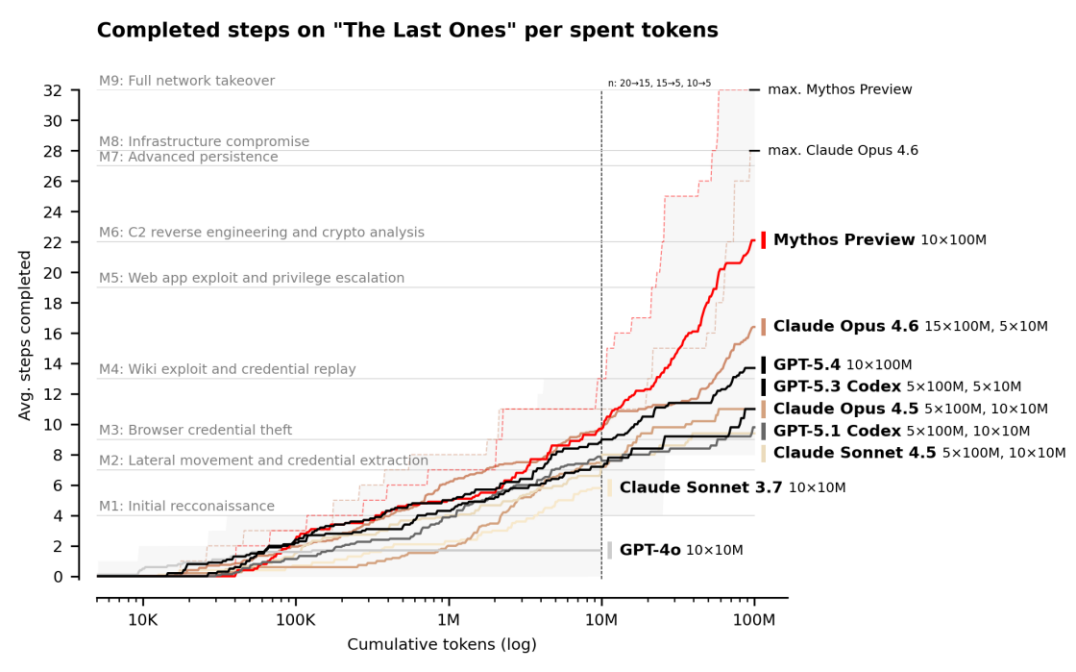
①Mythosはゲームチェーンジャーか？

①Mythosはゲームチェーンジャーか？

出典：Daniel://Steinberg://

- Curlメンテナー「大々的な宣伝は主にマーケティング目的。Still Goodだが、他のツールよりも高度なレベルにあるという証拠は見当たらない」
- 検出された「脆弱性」は５個。１つは低深刻度、３つは誤検出、１つは単なるバグ。

①Mythosはゲームチェーンジャーか？



出典：AI Security Institute

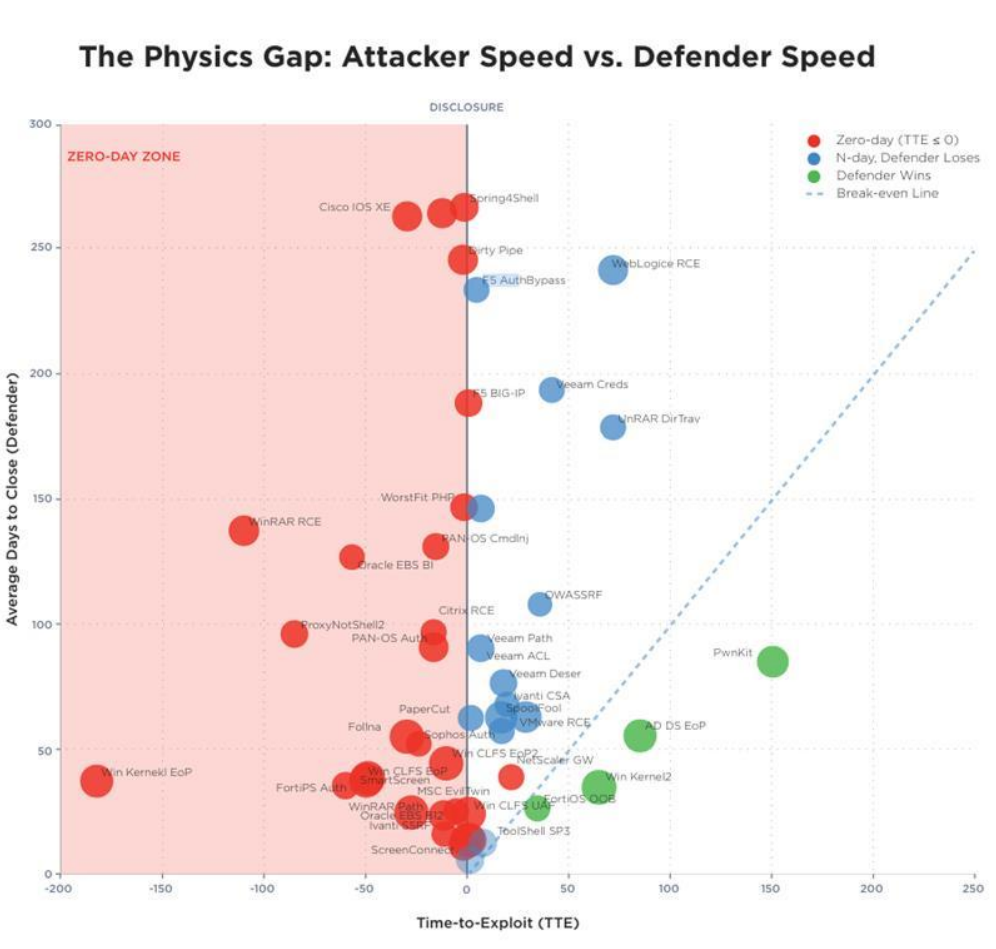
- **UK AI Security Institute「CTFで７３％をクリア」「企業ネットワークの攻撃シミュレーションを3/10で成功」**

一方で・・・

- **OTの攻撃環境をクリアできず「限界を示した」**
- **アラートをトリガーするような行動にペナルティなし。「十分に防御されたシステムを攻撃できるかどうかは断言できない」**

②今は「Mythos前夜」なのか？

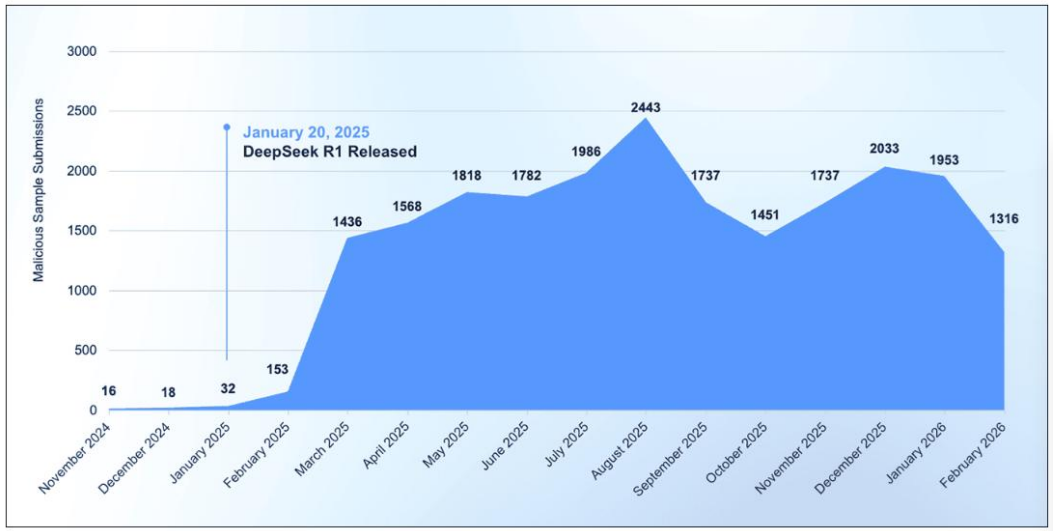
②今は「Mythos前夜」なのか？



- 2022~25年、悪用開始時期がわかっている主要なKEVの追跡調査
- 88%のKEVの修正平均時間は悪用開始に間に合っていない
- 52%はゼロデイ
- 「根本的な問題は人的ボトルネック」

出典：Qualys

②今は「Mythos前夜」なのか？



- Deep Seek R1のリリース（２０２５年１月）直後から、AI利用とみられるマルウエアが急増
- 月間検知数は最大で１３６倍に
- 開発者の推定使用言語は日本語が最多。理由は「不明」

出典：Arctic Wolf

②今は「Mythos前夜」なのか？

AN OPEN LETTER

On Transparent AI Cyber Protections

June 14, 2026

Dear Secretary Lutnick and National Cyber Director Cairncross,

We, the undersigned executives and technical leaders from across the United States and its allies, write to you to ask you to lift the export control directives on Anthropic's Fable and Mythos large language models and commit to an open, scientific and transparent process of handling AI risk assessments in the future.

First, we would like to state that we believe that:

出典：https://freefable.org/

- Mythosクラスのモデルは、脆弱性の発見と武器化に非常に優れているが、特別に優れているわけではない
- 署名者の多くは、セキュリティ監査やレッドチーム演習のために、他のモデルを日常的に利用している
- 中国のオープンウェイトモデルは米国のフロンティアモデルからわずか数か月遅れているだけ

③Mythosはインフラへの脅威なのか？

②Mythosはインフラへの脅威なのか？

どれも短期的に終わる問題じゃないよ・・・

□ 金融庁 ら脅威変化を踏まえた金融機関等の短期的な対応について

テスト環境がないから更新もできないんだよ・・・

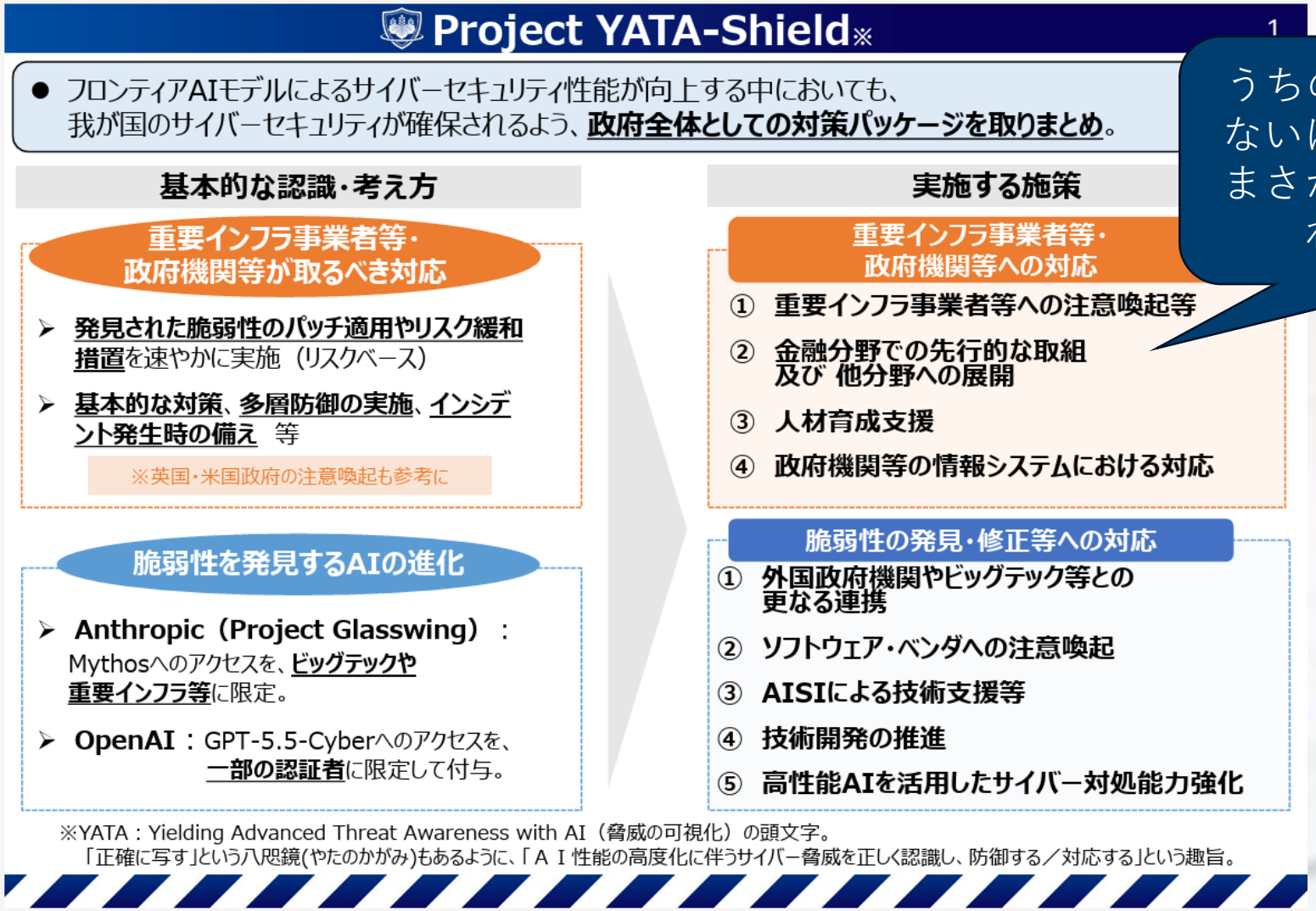
経営課題として扱う	ス／システムの特定	技術負債の解消	脆弱性対応の人的リソース追加	ベンダーとの維持保守契約の確認
パッチ適用クベー		能動的停止の検討	ISACなどとの連	

本当に国民の理解得られるの・・・？

サービス停止がむしろ増える！！

中長期的には脆弱性対応の自動化を

③Mythosはインフラへの脅威なのか？



うちのIT環境外部露出してないはずなんですが・・・まさか金融機関と同等に扱われるの・・・？

③Mythosはインフラへの脅威なのか？

□ CISA Binding Operational Directive 26-04（連邦政府機関向け）

- 公開性：インターネットに露出しているか？
- KEVステータス：KEVに掲載されているか？
- 悪用の自動化：攻撃プロセスをすべて自動化可能か？
- 技術的影響：悪用によってシステムを完全制御可能か？

高リスクのものは「３日以内」に対応すべき

（従来はKEVを１４～２１日）

脆弱性事例のうち3日以内に対応が必要なものはわずか1%で、60%以上は次回のシステムアップグレードまで延期される。

攻撃者が製品の脆弱性を悪用して主にコアネットワークを侵害している事例は確認していない。むしろ用いられるのはLiving Off the Land（LOTL）だ。LOTLへの対策としては、システム構成の強化、ネットワークのセグメンテーション、フィッシング対策に優れた多要素認証の導入などが効果的だ。

④AI vs AIはスピードの勝負なのか？

④AI vs AIはスピードの勝負なのか？

防御側 ? 攻撃側

※イラストは生成AIで作成

④AI vs AIはスピードの勝負なのか？

出典：Relia Quest

- ある大手小売企業でAI SOCのトークン超過額が50万ドルに
- 軽量モデル／フロンティアモデルの切り替え（LLMルーティング）のソリューションをPR

ただし

- 少なくとも防御側が計算コストを余分に追う
- OWASP　LLM10:2025 Unbounded Consumptionで新たな攻撃面としてアラート
- AI SOCをToken Bombで梅雨払いさせることもありうる？

④AI vs AIはスピードの勝負なのか？

出典：Schneier On Security

Cybersecurity in the Age of Instant Software

AI is rapidly changing how software is written, deployed, and used. Trends point to a future where AIs can write custom software quickly and easily: "instant software." Taken to an extreme, it might become easier for a user to have an AI write an application on demand—a spreadsheet, for example—and delete it when you're done using it than to buy one commercially. Future systems could include a mix: both traditional long-term software and ephemeral instant software that is constantly being written, deployed, modified, and deleted.

AI is changing cybersecurity as well. In particular, AI systems are getting better at finding and patching vulnerabilities in code. This has implications for both attackers and defenders, depending on the ways this and related technologies improve.

In this essay, I want to take an optimistic view of AI's progress, and to speculate what AI-dominated cybersecurity in an age of instant software might look like. There are a number of unknowns that will factor into how the arms race between attacker and defender might play out.

- 前提として、バイブコーディングと脆弱性の探索のAIを組み合わせ「ほとんど脆弱性がないソフトウェア」を実現化
- Instant Software
  - バイブコーディングによるプロプライエタリかつオンデマンドなソフトウェア
  - 使い終わったら削除
  - 入手が難しく、リバエンにハードル
  - 多様化により攻撃側の経済合理性が下がる
- Self-Healing Network
  - AIエージェントの常時スキャンおよびパッチの自動作成
  - すなわち脆弱性対応をユーザー側に移行（！）
  - 多様化により攻撃側の経済合理性が下がる
  - 防御側は一定の情報共有やインセンティブ化により経済合理性を上げられる
  - もちろん保険によるカバーなど様々なバックアップが必要不可欠

④AI vs AIはスピードの勝負なのか？

Resilient Software System

“We try to mitigate the damage and patch them and round around. We go on this merry go round. And I think the DARPA impact is going to be changing that mindset.”

- 様々なフォーマルメソッド（＝形式手法＝脆弱性の不存在を“数学的に”証明）の一連の体系
- Isabell/HOLなど様々な証明器の実装
- 約１５年に渡る研究後、昨年ようやく空軍で実証段階に
- ＡＩ×ＣＣ→ＬＬＭによるスケール化
- ＳＬＣのどの段階でも適用可能
- ある意味ではＫＥＶのアンチテーゼ
- Amazon、Google、ＭＳで一部実装

Mythos時代を正しく
怖がるための広範な情報収集を