サイバーセキュリティレポート

2026.05

NTT セキュリティ・ジャパン株式会社
プロフェッショナルサービス部 OSINT モニタリングチーム

目次

【1 ページサマリー】

当レポートでは 2026 年 5 月中に生じた様々な情報セキュリティに関する事件、事象、またそれらを取り巻く環境の変化の中から特に重要と考えられるトピック 3 点を選び、まとめたものである。各トピックの要旨は以下のとおりである。

第 1 章『強力な AI は誰のものかーOpenMythos が示す AI 設計の民主化』

- 22 歳の個人開発者が第三者の学術論文を基に、Anthropic 社の非公開 AI モデル「Claude Mythos」の再構築を試みた。これをオープンソースプロジェクト「OpenMythos」として公開し、研究者等から注目を集めている。
- OpenMythos と Anthropic 社には公式な関係はなく、その内容が正確に Claude Mythos の構造を反映している保証はない。一方で、Forbes 誌は、公開情報から AI モデルの内部設計が再構築され得ること自体に意義があると報じている。
- OpenMythos の登場は、「誰が強力な AI を持てるか」という前提的な問いが変わる可能性を示した。今後、強力な AI の利用が広がる社会の中で自社・自組織の防御態勢をアップデートする舵取りが求められる。

第 2 章『台湾高速鉄道への近接攻撃』

- 2026 年 4 月、台湾の大学生が市販のソフトウェア無線(SDR)機器を用いて台湾高速鉄道の通信システムを侵害し、偽の緊急警報信号を送信することで、列車の運行を停止させる事案が発生した。
- 一般に入手可能な機器によるソフトウェア無線を用い、認証情報や通信設定の更新が長期間行われない等の運用管理の不備に付け込むことで、高度な攻撃技術が無くても正規の制御信号になりすました信号を送信することができた。
- 同様の無線通信は広く利用されており、鉄道分野に限らず重要インフラのシステムにおいて、無線からのシステム侵害は想定されるべきものと捉える必要がある。

第 3 章『AI がもたらすバグバウンティの構造転換』

- 米 Google をはじめとする企業は、AI による脆弱性発見プロセスの変化を受け、バグバウンティの見直しを進めている。
- AI は調査の効率化や検証範囲の拡大を可能にした一方、発見された多数の脆弱性と、それらを検証・修正するスピードのバランスに変化をもたらし、新たな課題を生じさせている。
- 今後こうしたバランスの変化はセキュリティ全体に波及する可能性があり、AIと人間がそれぞれの強みを活かして補完し合うことで、実効性の高いセキュリティ運用の確立が期待される。

1. 強力な AI は誰のものかーOpenMythos が示す AI 設計の民主化

1.1. 概要

22 歳の個人開発者が第三者の学術論文を基に、Anthropic 社の非公開 AI モデル「Claude Mythos Preview」(以下、Claude Mythos)の再構築を試み、これをオープンソースプロジェクト「OpenMythos」として GitHub(開発者がプログラムのソースコードを管理・共有するためのプラットフォーム)上に公開した[1]。同プラットフォームのコミュニティでは開発者や研究者等から大きな関心・支持を集めている[2]。

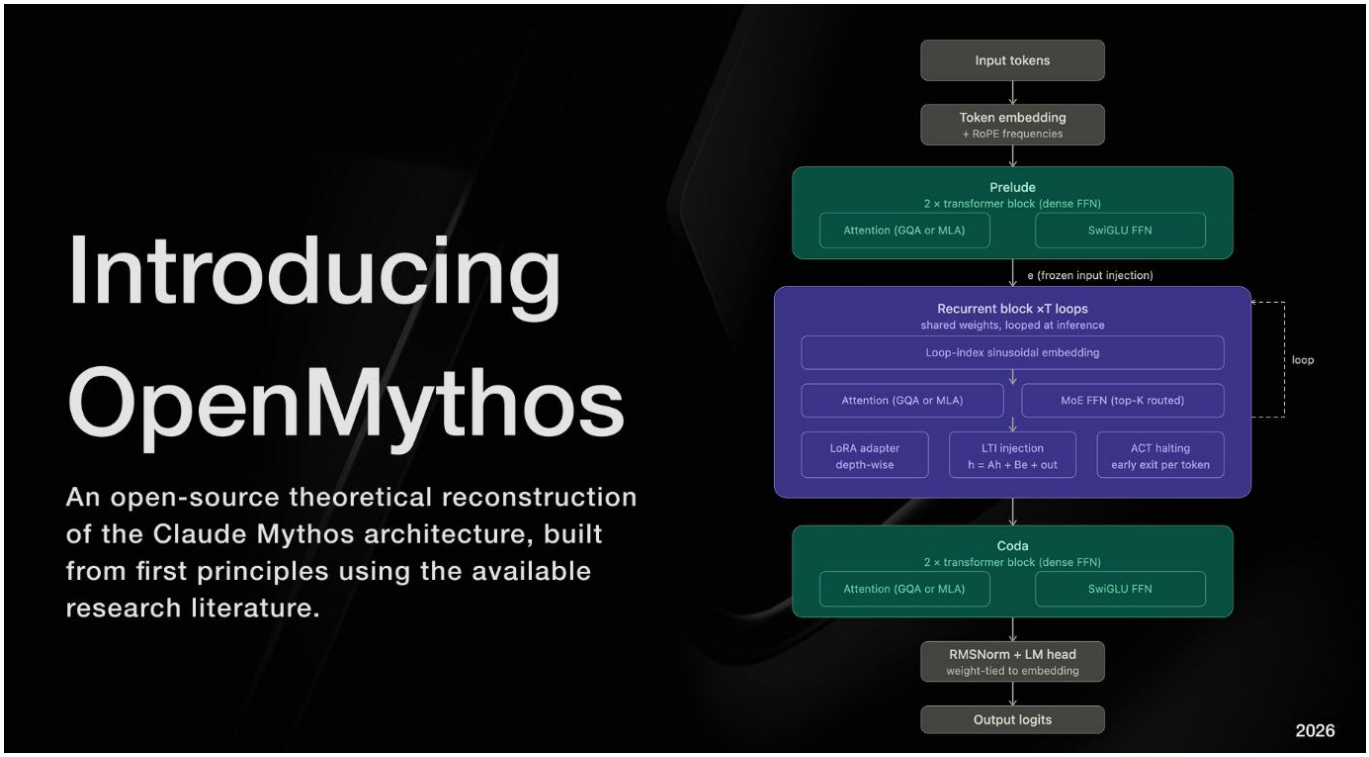

図 1 OpenMythos の説明画像(開発者 Kye Gomez 氏の X 投稿より)[3]

[1] 出典: MarkTechPost『Meet OpenMythos: An Open-Source PyTorch Reconstruction of Claude Mythos Where 770M Parameters Match a 1.3B Transformer』
https://www.marktechpost.com/2026/04/19/meet-openmythos-an-open-source-pytorch-reconstruction-of-claude-mythos-where-770m-parameters-match-a-1-3b-transformer/

[2] 出典: GitHub『kyegomez/OpenMythos』
https://github.com/kyegomez/OpenMythos

[3] 出典: X『@KyeGomezB』
https://x.com/KyeGomezB/status/2045659150340723107?s=20

1.2. OpenMythos とは何か

【OpenMythos について】

名称は似ているが、OpenMythos と Claude Mythos の両者に公式な関係はない[1, 2]。Claude Mythos の公表前に、Anthropic 社が運用する CMS(コンテンツ管理システム)の設定ミスから、同社の(Claude Mythos 関連を含む)未公開データが誰でも閲覧可能な状態となったことがあったが[4]、OpenMythos の開発はこの流出を利用したものではない[1]。また、Anthropic 社は Claude Mythos のモデル構造に関して技術論文等の資料を公開しておらず[1]、OpenMythos 開発者の Kye Gomez(カイ・ゴメス)氏は、複数の既存の(Anthropic 社が関係していない)論文で紹介された研究の内容を組み合わせることで、Claude Mythos の設計を推測した[5]。ソースコードは GitHub 上で公開されており、誰でも無償で入手・利用・改変できる[1]。なお、Gomez 氏が採用した、設計思想に関する論文は、従来の約半分のパラメータ数(AI 処理の規模を示す指標)で、より大規模な AI モデルと同等の性能を発揮できることを述べており[6]、将来的により安価で低スペックのハードウェアにおいても高度な AI を動かせる可能性を示している。

現時点で OpenMythos のベンチマーク評価(モデル性能をデータセットと指標で比較・数値化する評価手法)は確認されていない。また、同プロジェクトは、AI が実際に動作するために必要な、「重み」と呼ばれる学習済みデータを提供していないが[7]、その代わり、推奨するデータセットを公開しており、ユーザーはこれを使用して OpenMythos に学習させることができる[8]。

【開発者のプロフィール】

開発者の Gomez 氏は米フロリダ州マイアミ出身の 22 歳。10 歳でプログラミングを始め、13 歳の時には AI モデルを初めて作成したが、この時の目的は母親の Gmail アカウントをハッキングし、ゲームのオンラインストアで使用できるコード番号を入手することであった[9]。現在同氏は、AI 分野を中心に活動する研究者であり、数年前に創業した AI 企業「Swarms」の CEO でもある[10]。

【OpenMythos の評価】

過去に Gomez 氏が公開した成果物では、出典を明示することなく他者の研究成果を改変しリリースしているとして、機械学習のコミュニティで批判の声が上がったこともあった[5]。今回の OpenMythos の公開にあたっては、Gomez 氏は参照した論文を明示的に引用しており、決して根拠のない独自性を主張しているわけではない。ただし、本プロジェクトの内容が正確に Claude Mythos の構造を反映している保証はなく、本人が言うようにあくまでも「公開されている研究と推論のみ」[2] に基づく

---

[4] 出典: Awesome Agents『Anthropic's Mythos Model Exposed by CMS Misconfiguration』
https://awesomeagents.ai/news/anthropic-mythos-capybara-leak/

[5] 出典: Awesome Agents『OpenMythos Recasts Claude Mythos as Looped MoE Transformer』
https://awesomeagents.ai/news/openmythos-recurrent-depth-transformer/

[6] 出典: Sandy Research『Parcae: Doing More with Fewer Parameters using Stable Looped Models』
https://sandyresearch.github.io/parcae/

[7] 出典: Forbes『Did A 22-Year-Old Dropout Reverse-Engineer The World's Scariest AI?』
https://www.forbes.com/sites/craigsmith/2026/05/02/a-22-year-old-dropout-just-reverse-engineered-the-worlds-scariest-ai/

[8] 出典: GitHub『kyegomez/OpenMythos - Recommended Training Datasets』
https://github.com/kyegomez/OpenMythos/blob/main/docs/datasets.md

[9] 出典: Refresh Miami『18-year-old Miamian Kye Gomez is developing AI to make life less boring』
https://refreshmiami.com/news/18-year-old-miamian-kye-gomez-is-developing-ai-to-make-life-less-boring/

[10] 出典: Kye Gomez『Kye Gomez | AI Researcher and Founder of Swarms.ai』
https://kyegomez.com/

ものである点には留意が必要である[5]。

米経済誌 Forbes は、OpenMythos が Anthropic 社の設計と一致するかどうかよりも、公開情報から AI モデルの内部設計が再構築され得ること自体に意義があると報じており、当プロジェクトの公開によって具体的な検証材料を研究者らに提供した点を評価している[7]。実際、このような評価は GitHub 上での反応にも表れており、OpenMythos は 1 万以上の Star(GitHub ユーザーが関心や支持を示す指標)を集めていることから、研究者や開発者等から大きな注目を集めていることがうかがえる[2]。

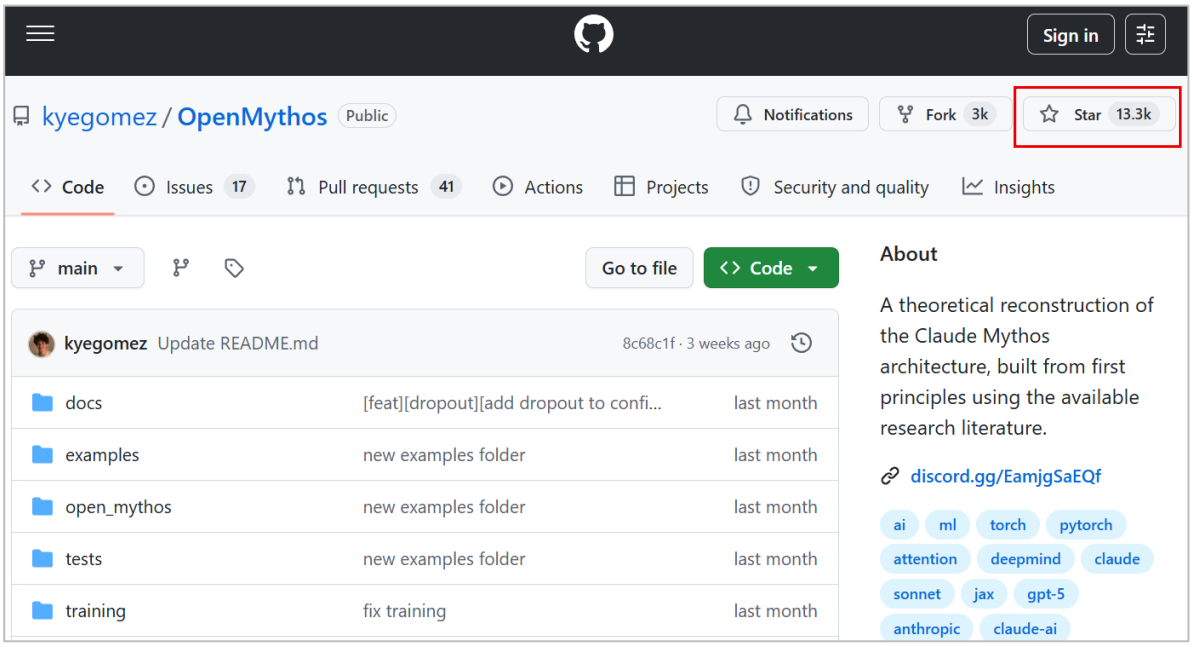

図 2 GitHub のスター数(赤枠 2026 年 5 月 21 日時点)[2]

【Anthropic 社の見解】

現時点で、Anthropic 社は OpenMythos に関するコメントを出していない。一方で同社は、オープンソースや中国の開発者が Claude Mythos と同等の能力に到達するまでの期間について、おおむね 6〜12 か月程度との見方を示している[11]。

1.3. AI 設計思想の民主化

最先端 AI の開発をめぐる変化を考える上で注目すべきは、現在、「誰が・どのくらいの速さで・どこまで」その設計思想を再構築できるようになっているかということである。

OpenMythos は、従来は大規模 IT 企業等の限られた環境において非公開として扱われてきた最先端の AI 設計手法を、オープンソースコミュニティの手によって短期間で実験・検証可能な形で提示した。これにより、AI モデルの設計に関する知見を実際に試行可能な形で共有する動きが生まれており、研究開発の裾野が拡大しつつあるという点で大きな意義がある[7]。

[11] 出典: SecurityWeek『Chinese Cybersecurity Firm's AI Hacking Claims Draw Comparisons to Claude Mythos』 https://www.securityweek.com/chinese-cybersecurity-firms-ai-hacking-claims-draw-comparisons-to-claude-mythos/

1.4. まとめ

OpenMythos の登場は、「誰が強力な AI を持てるか」という前提的な問いが変わる可能性を示した。従来は主に大規模 IT 企業等に限られていた最先端 AI の設計が、政府の規制や企業の意図に影響されないオープンソースのコミュニティにおいて自由に議論し検証できる可能性が広がることは自然な流れとなりつつある。今後、強力な AI の利用が広がる社会においては自社・自組織の防御態勢をアップデートする舵取りが求められる。

2. 台湾高速鉄道への近接攻撃

2.1. 概要

台湾で、23 歳の男が市販のソフトウェア無線(SDR)機器を使用し、台湾高速鉄道(THSRC)の無線通信システムをハッキングする事件が発生した。攻撃者は偽の「緊急警報」信号を送信し、4 本の列車を停止させ、計 48 分間の遅延を引き起こした[12]。本事件は、市販機器による近接攻撃でも重要インフラを物理的に停止させ得ることを示し、鉄道の運用技術(OT)における深刻な脆弱性を露呈させた[13]。

2.2. 攻撃者について

攻撃者は 23 歳の台湾人大学生であり、無線通信の愛好家と報じられている[14]。

攻撃について本人は「うっかりして(無線機の)操作を誤った」と主張しているが[15]、警察の捜査では自宅から 11 台のプロ用ソフトウェア無線機、PC 等が押収されている。これらの状況から、意図的に信号解読が行われた可能性が高いとみられる。

図 3 押収された機器(いずれも容易に入手可能な市販品)[16]

ソフトウェア無線は一般に市販されており、PC に接続してソフトウェアを書き換えることで、異なる無線通信方式や周波数等に柔軟に対応することができる特徴を持つ。柔軟性が高い反面、周囲の無線通信の傍受、信号の解析、さらには特定の通信方式を模倣した信号の生成・送信が可能となる。

---

[12] 出典：Taipei Times『Student's hack prompts THSRC review』
https://www.taipeitimes.com/News/taiwan/archives/2026/05/05/2003856781

[13] 出典：Dark Reading『Taiwan Bullet Train Hack Highlights Cybersecurity Gaps in Rail Systems』
https://www.darkreading.com/ics-ot-security/taiwan-incident-highlights-cybersecurity-gaps

[14] 出典：知新聞『高鐵無線電遭破解有共犯！男大生學弟提供關鍵參數 烏日站外觸發緊急按鈕露餡』
https://www.knews.com.tw/news/85322068F027005083050EC2DDFC3DE4

[15] 出典：SETN 三立新聞網『竊聽高鐵無線電！23 歲男大生「誤觸緊急按鈕」害 3 車急停 驚人背景曝光』
https://www.setn.com/News.aspx?NewsID=1831451

[16] 出典：民視新聞網『堪稱「國家級駭客」！ 男大生入侵高鐵、北捷迫列車停駛』
https://www.ftvnews.com.tw/news/detail/2026430C13M1

本事件においても、攻撃者はこのような特性を活用し、台湾高速鉄道の無線通信プロトコルやパラメータを解析・再現することで、正規の通信を装った信号を意図的に送信したと考えられる。

さらに、捜査の結果、攻撃者は台湾高速鉄道に限らず、消防局や空港 MRT (地下鉄駅)の通信周波数についても受信・解析ができる状態にあったことが確認されており、公共性の高い無線インフラに広くアクセス可能であった実態が明らかとなった[12]。

2.3. 無線インフラの運用不備

本事件の主な要因は、無線通信における長年の運用管理の不備にある。特に、外部から直接アクセス可能な無線通信領域において、認証情報や設定が適切に管理されていなかったことが、近接攻撃による侵入を許し、攻撃の成立に直結した。

【運用の不備】

無線通信のエリア内で電波を送信することで、通信システムに対して信号を送信することが可能となるが、それだけでシステムを侵害することは困難である。しかし、台湾高速鉄道が利用しているシステムでは、通信に用いられるパラメータや認証情報が約 19 年にわたり変更されていなかった可能性が指摘されている[17]。このような状況の下では、通信の観測や解析によって得られる情報が蓄積され、それらをもとに通信方式や関連する設定を把握しやすい状態となっていたと考えられる。

さらに攻撃者は、認証に用いられる通信パラメータを友人(共犯者)から入手していたことで[18]、これらの情報を組み合わせて認証に必要な条件を再現し、正規の通信になりすますことができたと考えられる。

これらは主として運用管理上の問題に起因するものであり、本来であれば認証情報や通信パラメータの定期的な更新および適切な管理により防止可能であった。

【事件の影響】

5 月 4 日、日本の国会にあたる「立法院」の交通委員会での審議の中で、議員が「高速鉄道の無線通信システムは既に台湾の交通機関の中で最高水準にある。もし高速鉄道ですらハッキングされるなら、(在来線である)台湾鉄路はさらに脆弱ではないのか」と述べ、無線システム設備の更新時期や、設備の保管に関する手順書等の見直しを検討する必要性を訴えた。これに対し、台湾交通部は、台湾高速鉄道と台湾鉄路を対象とした点検を行っており、地下鉄についても各事業者に対応を要請する方針であることを説明した[19]。

2.4. まとめ

無線通信は、その管理状態によっては近接からの攻撃を受け得る構造を有している。本事例は、このような特性のもとで認証情報や通信設定の管理が不十分な場合、セキュリティ機能が十分に発揮されず、外部から送信された信号を受理してしまう要因となり得ることを示している。

また、同様の無線通信システムは日本を含む鉄道分野で広く利用されているが、こうした特性は鉄道に限らず、無線通信を利用する他の領域にも共通して見られるものであり、本事例は分野横断的なリスクが存在することを示唆している。

---

[17] 出典：ETtoday 新聞雲『高鐵無線電 7 道驗證仍被盜接　立委：系統 19 年未換』
https://www.ettoday.net/news/20260430/3158010.htm

[18] 出典：中央通訊社『高鐵無線電遭盜接案　檢方逮 20 歲共犯、8 萬元交保』
https://www.cna.com.tw/news/asoc/202604300366.aspx

[19] 出典：自由時報『高鐵無線電遭「盜接」導致停駛延誤　交通部：1 個月內檢討』
https://news.ltn.com.tw/news/Taipei/breakingnews/5424768

3. AI がもたらすバグバウンティの構造転換

3.1. 概要

米 Google はバグバウンティ(Bug Bounty [脆弱性報奨金制度])の見直しを進めている。この背景には、AI の普及により脆弱性発見のあり方が大きく変化したことがある。生成 AI の活用により、調査の自動化と高度化が進み、従来は人手に依存していた検証範囲が大幅に広がるとともに、発見までの時間も大きく短縮された。このような状況を踏まえ、バグバウンティを導入している Google 等の企業は、AI による網羅的な調査を前提とした上で、新たな時代に適応した制度設計への転換を進めている。

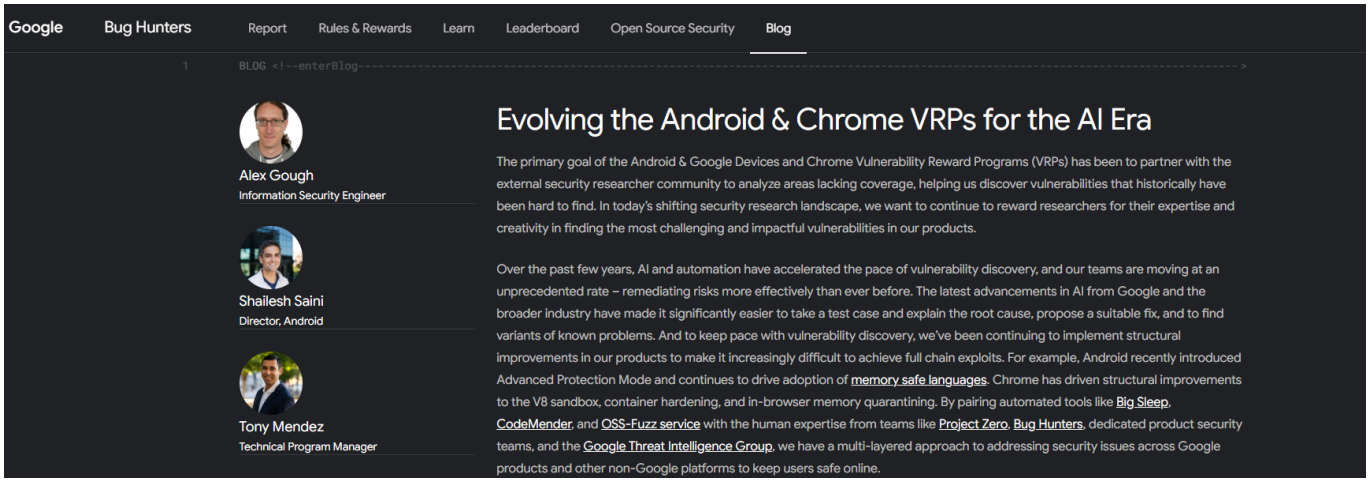

図 4 Google のバグバウンティ見直しに関する発表[20]

3.2. バグバウンティとは

バグバウンティとは、自組織の製品やサービスの脆弱性を外部のセキュリティ研究者やホワイトハッカーに発見・報告してもらい、その対価として報奨金を支払う制度である。この制度の目的は、攻撃者によって悪用される前に脆弱性を特定し、迅速に修正することでセキュリティリスクを低減することである。世界中の多様な専門知識を活用できるため、従来の診断方法では見落とされがちな脆弱性の発見が期待できるほか、成果報酬型であることから費用対効果が高いという利点がある[21]。この制度は、主に高度な専門知識を持つ研究者が主体となって脆弱性を発見・検証することを前提として成立してきた[22]。

3.3. AI がもたらした影響

AI は、脆弱性調査の効率を飛躍的に高めるとともに、これまで一部の熟練したセキュリティ研究者に依存していた分析プロセスを大きく拡張させた。現在では、AI を用いた高度な自動化技術が脆弱性の検出に広く応用されつつある。例えば、膨大

[20] 出典: Google『Evolving the Android & Chrome VRPs for the AI Era』
https://bughunters.google.com/blog/evolving-the-android-chrome-vrps-for-the-ai-era

[21] 出典: 三井物産セキュアディレクション株式会社『バグバウンティ』
https://www.mbsd.jp/glossary/bug_bounty

[22] 出典: Technology Org『Bug Bounty Schemes Buckle Under Flood of AI-Generated Junk Reports』
https://www.technology.org/2026/05/18/ai-slop-bug-bounty-programs-suspended/

なコードを短時間でスキャンしたり、従来は見落とされがちだった異常や構成ミスを検出したりすることも可能になるとされており、これは、作業担当者への依存度を下げ、品質を標準化することにも大きく貢献している[23]。AI の補助を通じて、より多くのリサーチャーが脆弱性の特定から検証、報告までを効率的に行えるようになり、バグバウンティ参加者の裾野が広がっている。

しかし、この変化は新たな構造的課題も生んでいる。AI による「発見・報告」の高速化が、人間の「検証・修正」の処理能力を圧倒しつつあるためである。次々と届く報告の検証に膨大なリソースが奪われ、本当に対応すべき重要なバグを見極めるだけで疲弊する「トリアージ疲労」や、有効と判定された件数に対して開発者の修正能力が追いつかず、脆弱性が未解決のまま放置される「修正プロセスのボトルネック」が深刻化している[24, 25]。AI の進化は、脆弱性を大量に発見する段階から、それらをいかに迅速かつ効率的に修正・管理するかという次の段階への移行を促しているのである。こうした脆弱性調査における発見と修正のバランスの変化は、その成果を受け止めるバグバウンティの再設計へとつながっている。

3.4. 各社の対応

バグバウンティをめぐる大きな変化に適応するよう、同制度を導入している組織には評価基準の再設計や運用の最適化が求められている。

【Google】[26]

Android: Google は、脆弱性の深刻度に応じて段階的に報奨金を設定している。今回の見直しにより、特に Google Pixel のセキュリティチップにおける脆弱性のうち、侵害後も攻撃者によるアクセスや制御が継続する「永続化(Persistence)」を伴うものについては、報奨金の上限額を 100 万ドルから 150 万ドルへと引き上げた[27]。このような高額報奨金の背景には、AI による広範な調査を前提としつつも、自動化された AI では発見が難しく、かつ人間による文脈理解や分析が必要な脆弱性を重点的に評価する狙いがある。

Chrome: 詳細な説明よりも、再現可能な証拠を備えた簡潔な報告を最優先する方針へと変更した。また、AI によって一部の攻撃手法の実証が容易になったことを受け、それらを対象としていたボーナス制度は廃止された。

---

[23] 出典:セキュアイノベーション『【2025 年最新】AI による脆弱性診断はどこまで進化しているのか？』
https://www.secure-iv.co.jp/blog/19975#blog_midashi2_2

[24] 出典: Dark Reading『AI-Led Remediation Crisis Prompts HackerOne to Pause Bug Bounties』
https://www.darkreading.com/application-security/ai-led-remediation-crisis-prompts-hackerone-pause-bug-bounties

[25] 出典: HackerOne『The Vulnerability Apocalypse Is a Remediation Crisis』
https://www.hackerone.com/blog/continuous-threat-exposure-management-remediation-crisis

[26] 出典: Google『Evolving the Android & Chrome VRPs for the AI Era』
https://bughunters.google.com/blog/evolving-the-android-chrome-vrps-for-the-ai-era

[27] 出典: SecurityWeek『Google Adjusts Bug Bounties: Chrome Payouts Drop as Android Rewards Rise Amid AI Surge』
https://www.securityweek.com/google-adjusts-bug-bounties-chrome-payouts-drop-as-android-rewards-rise-amid-ai-surge/

Description	Maximum Reward
Titan M2 with Persistence	Up to $1,500,000
Titan M2 without Persistence	Up to $750,000
Secure Element Data Exfiltration	Up to $375,000
Software-Based Lockscreen Bypass	Up to $150,000

図 5 Android 向け脆弱性報酬プログラムの最大報奨金一覧[28]

【HackerOne】

バグバウンティの代表的なプラットフォームとして知られる HackerOne は、主要なオープンソースソフトウェア(OSS)を対象とした脆弱性報奨金プログラム「Internet Bug Bounty (IBB)」を運営していたが、新規受付を停止した。背景には、AI の活用などにより脆弱性の発見スピードが大幅に向上した一方で、OSS 側の修正が追いつかなくなり、両者のバランスが崩れたことがある。本来、IBB は脆弱性の発見を確実な修正につなげ、継続的なセキュリティ向上を実現することを目的としていたが、現在は制度の見直しが進められている[29]。

【GitHub】

開発者がプログラムのソースコードを管理・共有するためのプラットフォームである GitHub は、セキュリティ上の重大な影響には至らない報告であってもコードやドキュメントの修正につながるものについては、報奨金ではなくノベルティグッズで表彰する方針を発表した。重要度の高い脆弱性に報奨金のリソースを集中させる狙いがある[30]。

3.5. まとめ

これまでバグバウンティにおいては、脆弱性の報告件数が重視される傾向にあったが、AI の普及により、その前提は大きく変化しつつある。確かに、AI によって脆弱性の発見は飛躍的に増加したが、その一方で、修正・管理が追いつかず、重要度や対応優先順位を見極めることが新たなボトルネックとなっている。こうした変化は、バグバウンティ分野にとどまらず、セキュリティ分野全体にも波及していく可能性があり、様々な混乱が生じることが想定される。

また、当面、AI がセキュリティ運用における検出や分析を担い、人間が評価と優先順位付けを行うという役割分担が様々な面で広がると考えられる。一方で AI の進化に伴い、評価や判断の領域においても AI が人間の能力を上回る場面が増えていくかもしれない。このような状況においては、両者それぞれの強みを活かした柔軟な役割分担を構築し、実効性の高いセキュリティ運用を確立していくことが今後の重要な課題となる。

以上

---

[28] 出典: Google『Android and Google Devices Security Reward Program Rules』
https://bughunters.google.com/about/rules/android-friends/android-and-google-devices-security-reward-program-rules

[29] 出典: HackerOne『Inernet Bug Bounty - Program guidelines』(受付ページ)
https://hackerone.com/ibb?type=team

[30] 出典: GitHub『Raising the bar: Quality, shared responsibility, and the future of GitHub's bug bounty program』
https://github.blog/security/raising-the-bar-quality-shared-responsibility-and-the-future-of-githubs-bug-bounty-program

免責事項

本記事の内容は、正確であることに最善を尽くしておりますが、内容を保証するものではなく、本記事の利用に起因して発生したいかなる損害、損失についても補償しませんのでご留意ください。記事内に誤植や内容の誤り、その他ご指摘等、お問い合わせ事項がある場合は、お手数ですが下記までご連絡ください。

[お問い合わせ先]
NTT セキュリティ・ジャパン株式会社
プロフェッショナルサービス部 OSINT モニタリングチーム
メールアドレス: nsj-ps-osintmonitoring@security.ntt