SECURE CYCLE

サイバーセキュリティにおけるAI活用の今

プロフィール

名前

服部 祐一

株式会社 セキュアサイクル
代表取締役
博士(工学)

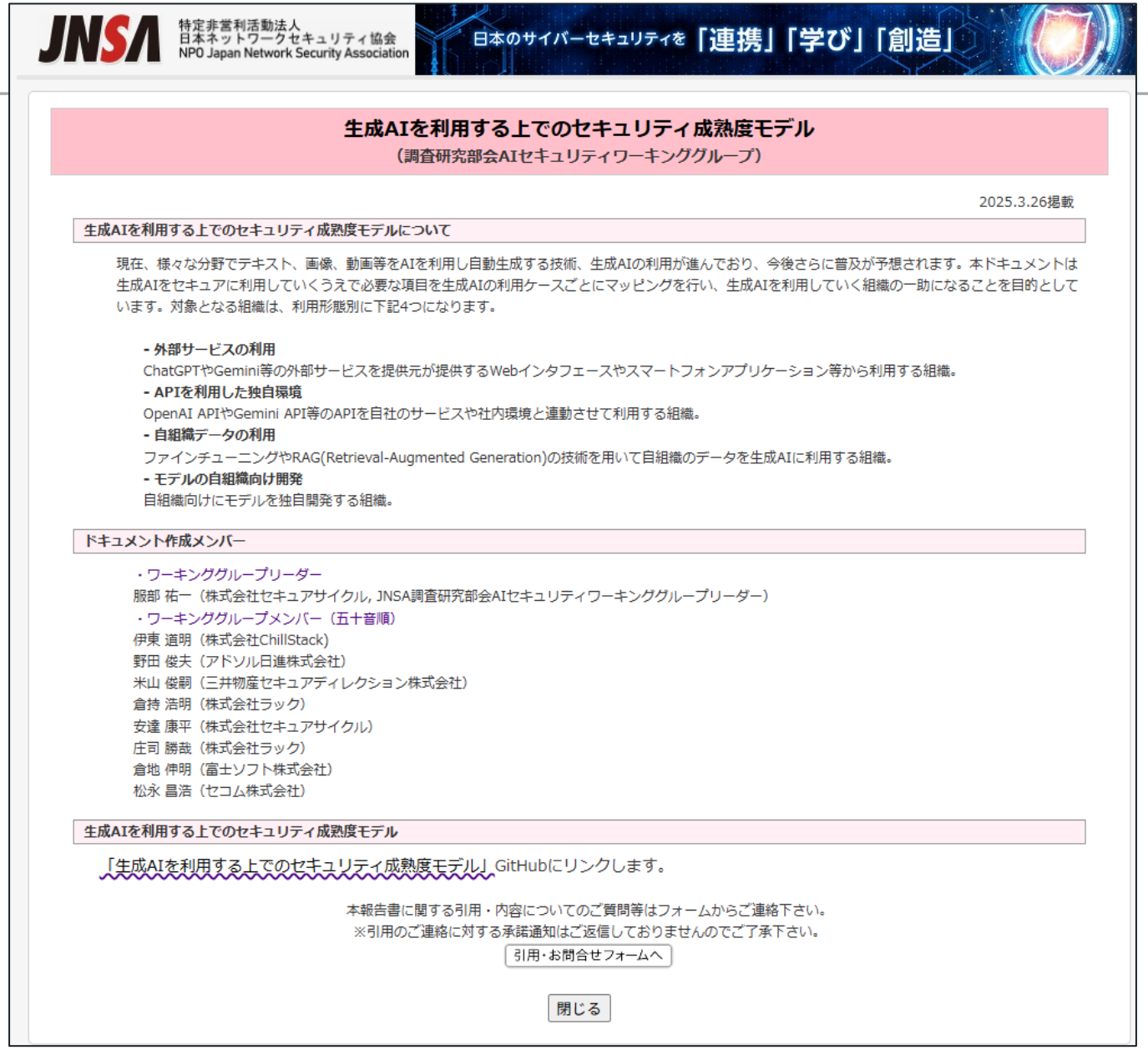
所属・役職

- 株式会社ギブリー 取締役 兼 CISO
- OWASP Fukuoka チャプターリーダー
- SecHack365 トレーナー
- 福岡未踏PM
- 一般社団法人地域セキュリティ協議会 理事
- JNSA調査研究部会AIセキュリティWG　リーダー
- 九州セキュリティシンポジウム実行委員
- SECKUN 講師

など

経歴

九州工業大学卒．博士(工学). 在学時はスマートフォンを使った行動認識の研究に携わり，
Microsoft Research AsiaにてResearch Internshipとして行動認識の研究に従事.
その後，ベンチャー企業でチーフエンジニアとして新規Webサービスの開発や脆弱性診断の
経験後，情報セキュリティ専門会社のCTOを経て現在に至る.
各地のセキュリティ関連イベントや，企業，大学等での講演・トレーニング多数.
現在も脆弱新診断や開発時のセキュリティ対策のコンサルなどの実務にも携わる.



特定非営利活動法人
日本ネットワークセキュリティ協会
NPO Japan Network Security Association

日本のサイバーセキュリティを「連携」「学び」「創造」

生成AIを利用する上でのセキュリティ成熟度モデル

（調査研究部会AIセキュリティワーキンググループ）

2025.3.26掲載

生成AIを利用する上でのセキュリティ成熟度モデルについて

現在、様々な分野でテキスト、画像、動画等をAIを利用し自動生成する技術、生成AIの利用が進んでおり、今後さらに普及が予想されます。本ドキュメントは生成AIをセキュアに利用していくうえで必要な項目を生成AIの利用ケースごとにマッピングを行い、生成AIを利用していく組織の一助になることを目的としています。対象となる組織は、利用形態別に下記4つになります。

- 外部サービスの利用
ChatGPTやGemini等の外部サービスを提供元が提供するWebインタフェースやスマートフォンアプリケーション等から利用する組織。
- APIを利用した独自環境
OpenAI APIやGemini API等のAPIを自社のサービスや社内環境と連動させて利用する組織。
- 自組織データの利用
ファインチューニングやRAG(Retrieval-Augmented Generation)の技術を用いて自組織のデータを生成AIに利用する組織。
- モデルの自組織向け開発
自組織向けにモデルを独自開発する組織。

ドキュメント作成メンバー

・ワーキンググループリーダー
服部 祐一（株式会社セキュアサイクル, JNSA調査研究部会AIセキュリティワーキンググループリーダー）
・ワーキンググループメンバー（五十音順）
伊東 道明（株式会社ChillStack）
野田 俊夫（アドソル日進株式会社）
米山 俊嗣（三井物産セキュアディレクション株式会社）
倉持 浩明（株式会社ラック）
安達 康平（株式会社セキュアサイクル）
庄司 勝哉（株式会社ラック）
倉地 伸明（富士ソフト株式会社）
松永 昌浩（セコム株式会社）

生成AIを利用する上でのセキュリティ成熟度モデル

「生成AIを利用する上でのセキュリティ成熟度モデル」GitHubにリンクします。

本報告書に関する引用・内容についてのご質問等はフォームからご連絡下さい。
※引用のご連絡に対する承諾通知はご返信しておりませんのでご了承下さい。

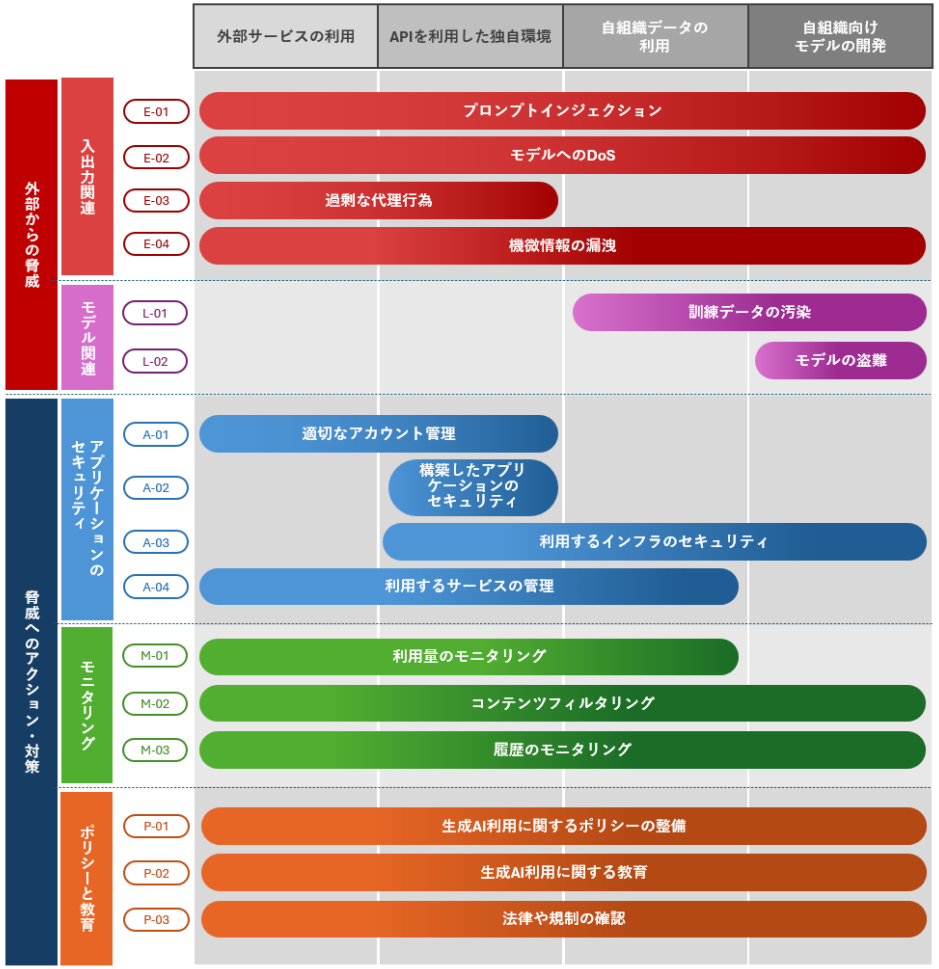
引用・お問合せフォームへ

閉じる

https://www.jnsa.org/result/aisec/2024/index.html

生成AIを利用する上でのセキュリティ成熟度モデル

現在、様々な分野でテキスト、画像、動画等をAIを利用し自動生成する技術、生成AIの利用が進んでおり、今後さらに普及が予想されます。本ドキュメントは生成AIをセキュアに利用していくうえで必要な項目を生成AIの利用ケースごとにマッピングを行い、生成AIを利用していく組織の一助になることを目的としています。

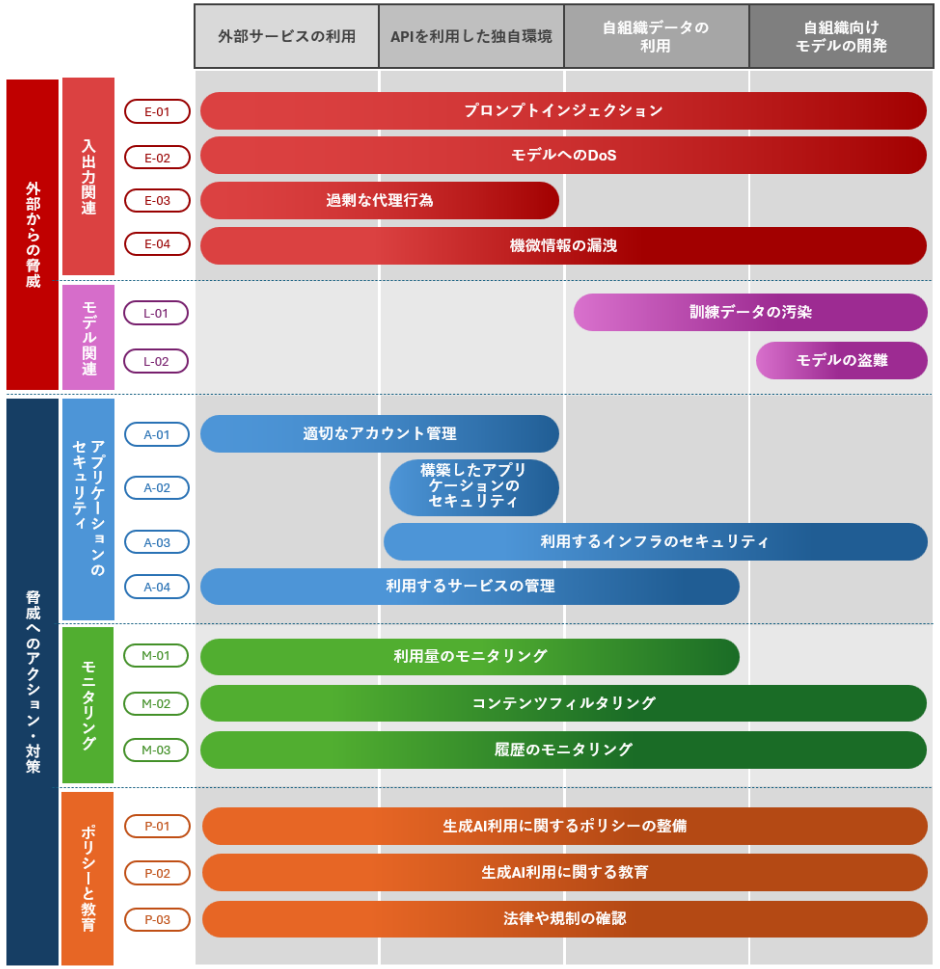


プロジェクトメンバー

本ドキュメントは、NPO日本ネットワークセキュリティ協会 調査研究部会 AIセキュリティWGのメンバーによって作成されています。

プロジェクトメンバー一覧(順不同)

- 服部 祐一(株式会社セキュアサイクル)
- 伊東道明(株式会社ChillStack)
- 野田俊夫(アドソル日進株式会社)
- 米山俊嗣
  (三井物産セキュアディレクション株式会社)
- 倉持浩明(株式会社ラック)
- 安達康平(株式会社セキュアサイクル)
- 庄司勝哉(株式会社ラック)
- 倉地伸明(富士ソフト株式会社)
- 松永昌浩(セコム株式会社)



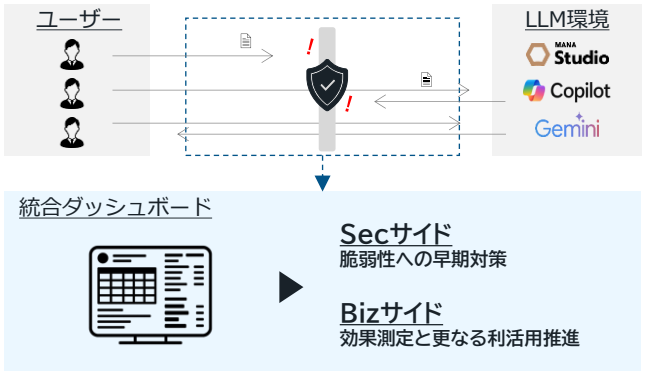
新サービスのご紹介

SECURE CYCLE

Security for AI

Monitoring for AI

イメージ



提供価値

- **AI利用状況のリアルタイム把握**による早期リスク判断と迅速対応を実現
- **活用状況を基にした施策提案**で、AI導入によるビジネス成果を早期に最大化

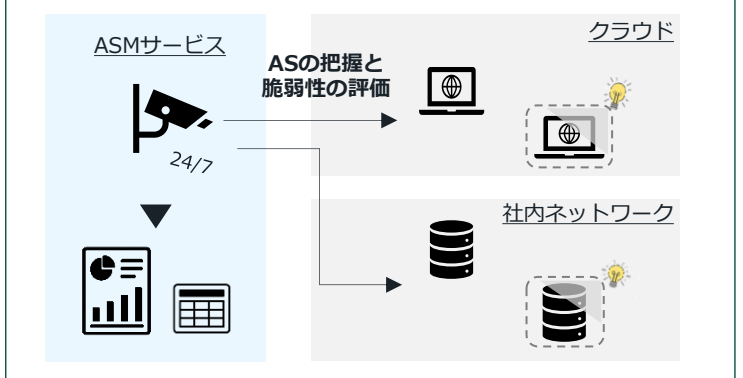
特徴

- **ビジネス・セキュリティ両面を一元管理できる**、サービス設計
- 多数のAI利活用支援の知見が凝縮された、**「現場が確実に変わる」ダッシュボード/分析基盤**

AI for Security

AI for ASM

イメージ



提供価値

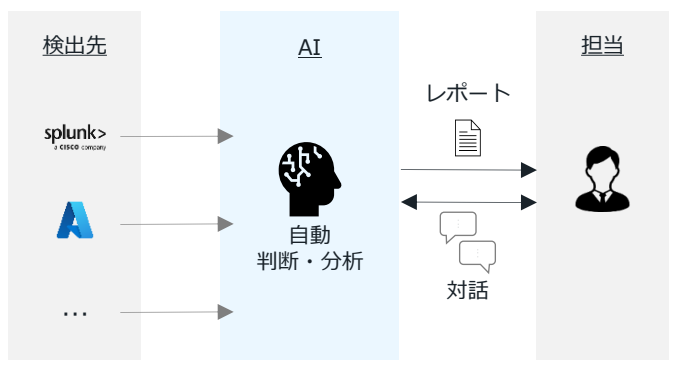
- シャドーITを含む**組織全体のアタックサーフェスを可視化**し、問題の早期発見が可能
- **継続的な監視業務が不要**になり、業務負担とコストが削減

特徴

- アタックサーフェスを**定期的に自動で見回り**
- **明瞭なダッシュボード・レポート**による リスク状態の5段階評価

AI for SOC

イメージ



提供価値

- **専門家不在でもアラート内容を直感的に理解**し、
- 適切な対応が可能
- **重要なアラートに自動で絞り込まれる**ことで、セキュリティ運用を効率化し、脅威を迅速に管理

特徴

- 他接続先の情報を踏まえた、**アラートに対する重要度推定**
- LLMとの自然な**対話型インターフェース**や レポーティングを通じた理解サポート

Monitoring for AI サービスコンセプト

統合ガードレール/統合データベース/統合ダッシュボードの3つのレイヤにより、
生成AI利用におけるビジネス上の課題とセキュリティ上のリスクの一元管理を可能にする

生成AI利活用推進部門

AI利活用の実態とその効果が見えず、
成果最大化の道筋を見いだせない

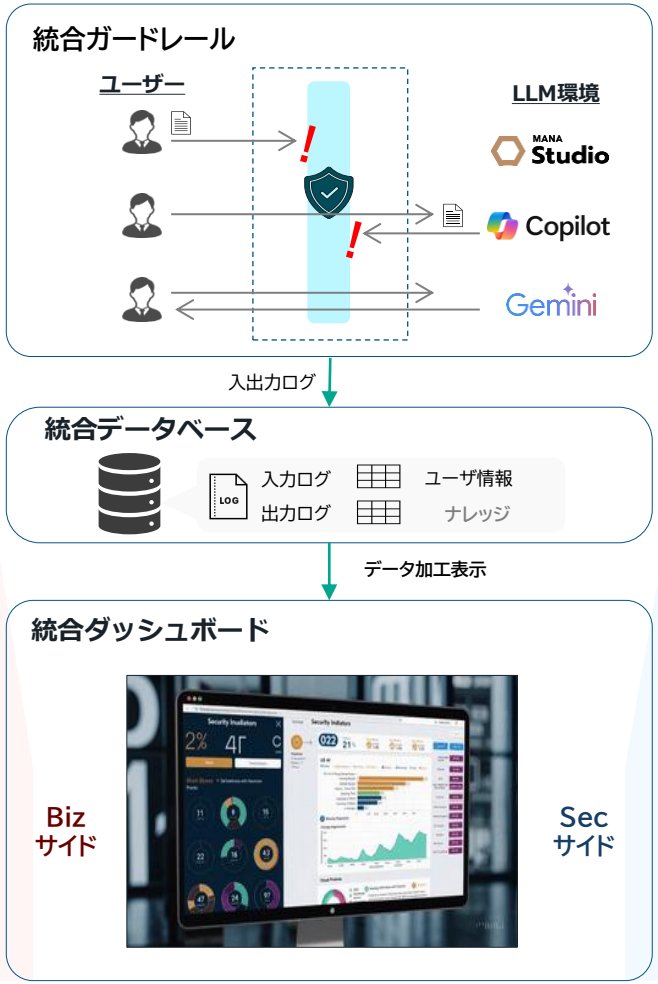
提供価値

Biz指標の可視化による効果測定と更なる推進

主な機能
- 利用率や削減時間など、**ビジネス上のKPI達成状況を表示**
- 頻出プロンプトや利用傾向を分析し、注力すべき施策をレコメンド

主な指標

部署別利用率 / *推定業務削減時間* / *頻出プロンプト*
適用業務領域 / *RAG精度* / *回答満足度*



セキュリティ部門

運用状況の観測がリアルタイムにできず、
インシデントを未然に防ぐことが困難

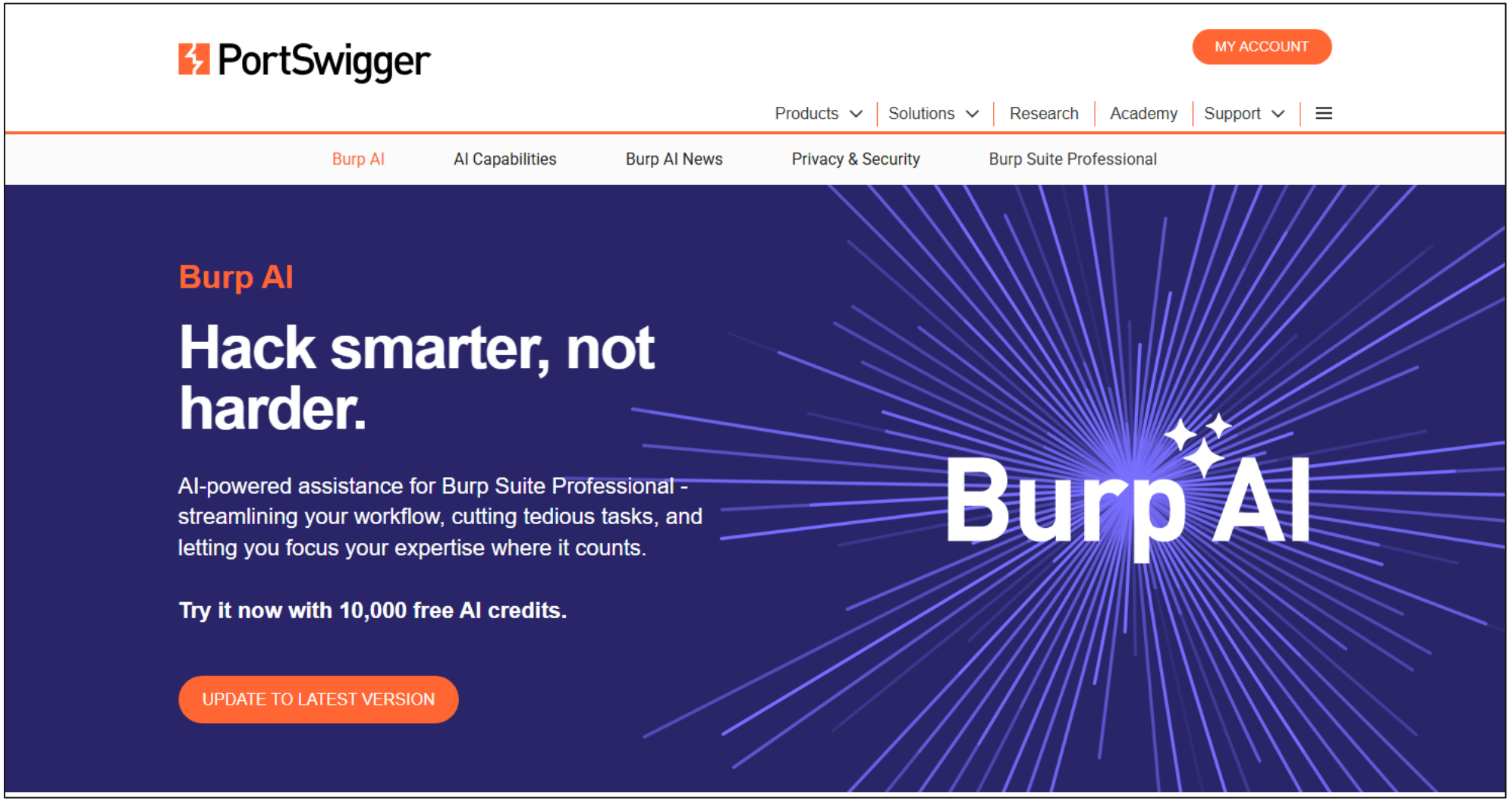
提供価値

Sec指標の可視化による脆弱性への早期対策

主な機能
- 各種規制や社内ガイドラインに抵触する入出力を検知・アラート
- 通常と異なる利用の検出など、**セキュリティリスクを検知・アラート**

アクセス元 / *要注意アウトプット* / *要注意プロンプト*
アラート件数 / *操作ログ* / *アノマリー挙動*

脆弱性診断におけるAI(Burp AI)

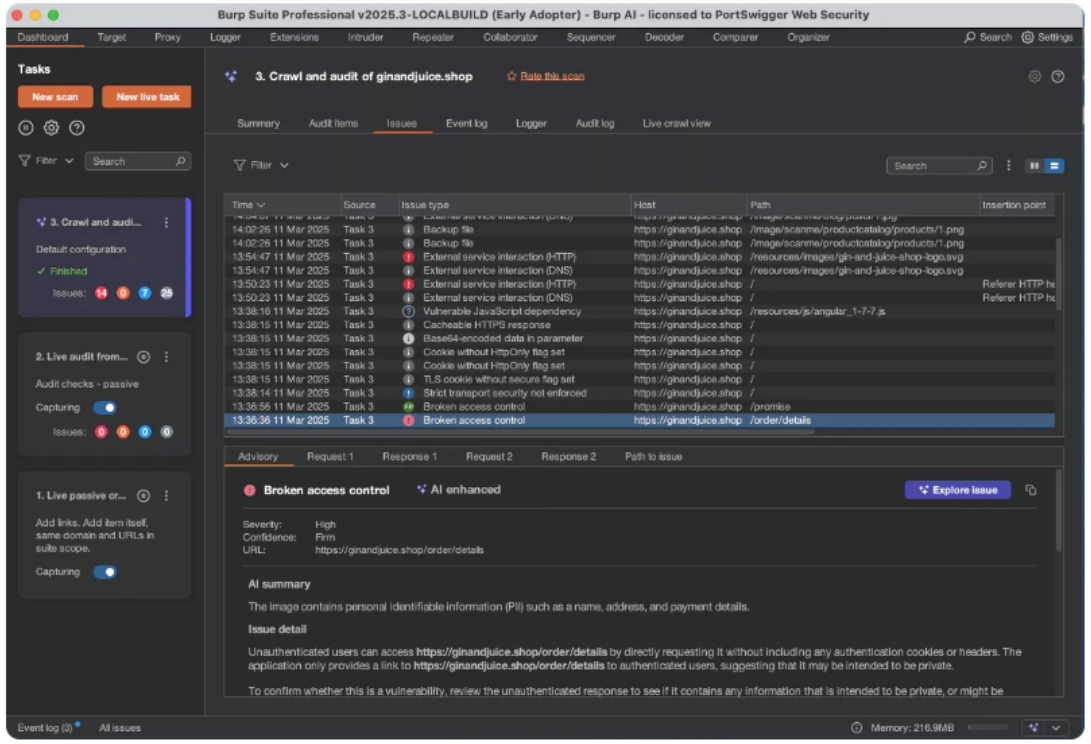


https://portswigger.net/burp/ai

脆弱性診断におけるAI(Burp AI)

SECURE CYCLE

Why Burp AI is a game changer



Built for security professionals

Burp AI is available in Burp Suite Professional, the product you know and love. Burp AI enhances your workflow, but you remain in control.

Trusted & secure

AI features operate within PortSwigger's trust boundary, ensuring data security. Your data is not retained by the AI service provider and is not used for model training purposes.

Not just AI hype

Practical AI solutions designed to solve real problems, not gimmicks.

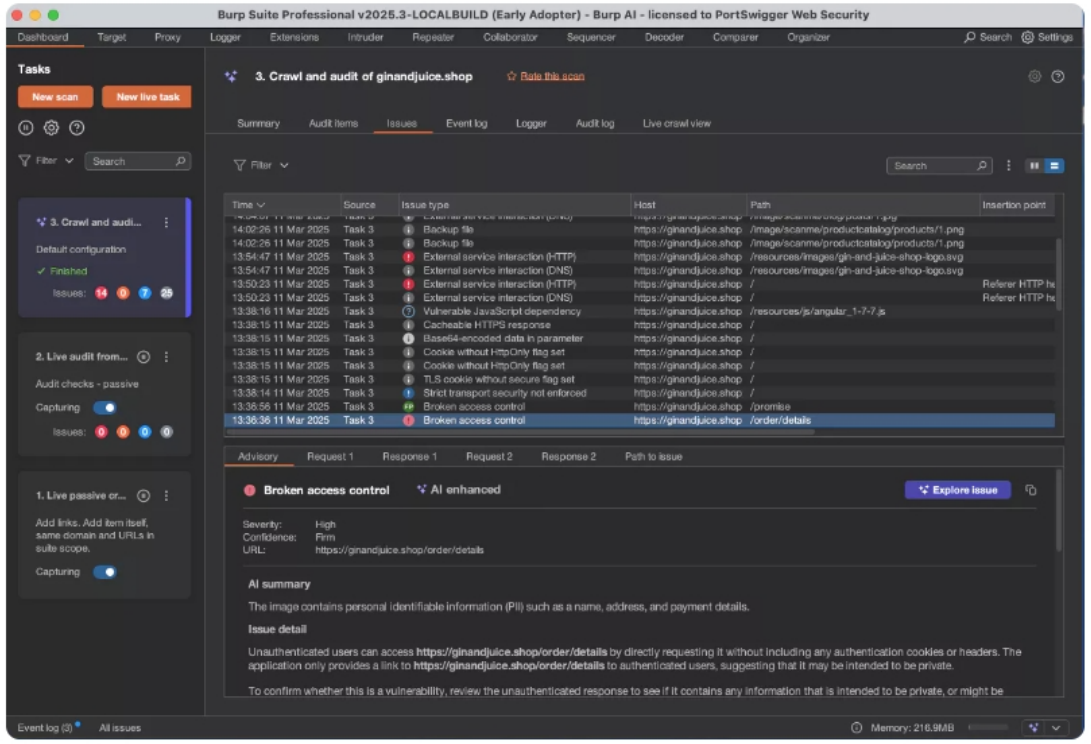
Update to latest version →

https://portswigger.net/burp/ai

脆弱性診断におけるAI(Burp AI)

SECURE CYCLE

Why Burp AI is a game changer



Built for security professionals

Burp AI is available in Burp Suite Professional, the product you know and love. Burp AI enhances your workflow, but you remain in control.

Trusted & secure

AI features operate within PortSwigger's trust boundary, ensuring data security. Your data is not retained by the AI service provider and is not used for model training purposes.

Not just AI hype

Practical AI solutions designed to solve real problems, not gimmicks.

Update to latest version →

https://portswigger.net/burp/ai/capabilities

AIによる知識のサポート

SECURE CYCLE

Explainer: Instant AI-powered insights

No more context-switching - get AI-powered, security-focused insights, directly in Burp Repeater.

Bridge knowledge gaps

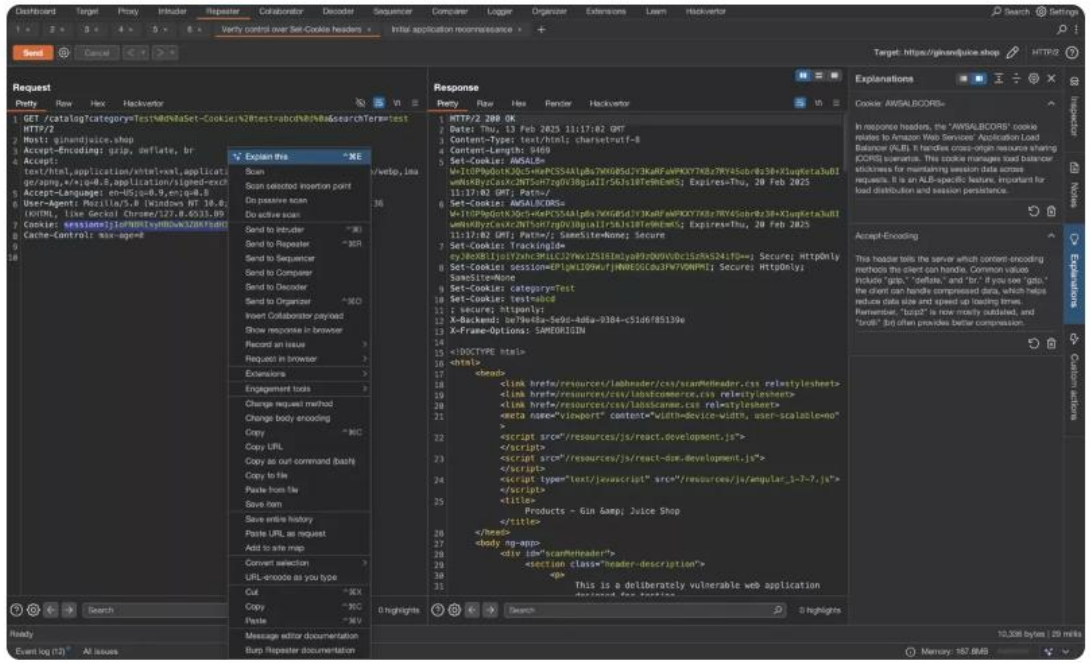
Quickly research unfamiliar HTTP headers, cookies, and other data and their potential security implications.

Quickly decipher code

Ask Burp AI to explain client-side JavaScript to you, so you can quickly understand what the code is doing, and whether it warrants deeper manual investigation, without having to decipher it line-by-line.

Reduce context-switching

Eliminate the need to switch between Burp and external information sources to look things up.

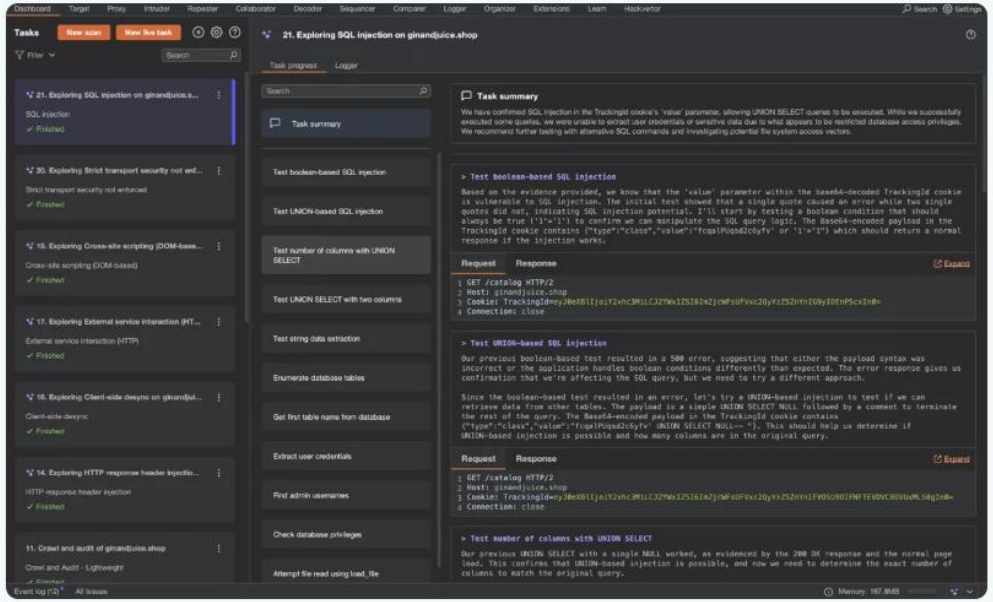


https://portswigger.net/burp/ai/capabilities

SECURE CYCLE

Explore issue: Automated vulnerability analysis

Performing follow-up analysis on issues identified by scan to validate them and demonstrate impact can be tedious and time-consuming. Let Burp AI investigate scanner-identified issues just like a pentester would.



Cover more ground

By outsourcing some of the issue analysis to AI, you can choose to focus your time and effort where you feel it's most valuable.

Demonstrate and escalate impact

Burp AI attempts to leverage the vulnerability to exfiltrate sensitive data, reach additional attack surface, and identify escalation paths, automatically generating PoCs on your behalf.

Stay in control

Burp AI provides step-by-step insights into what it's attempting at each stage, along with an executive summary of the findings so far. You can intervene at any point, whether it's to take over manually or simply because you feel the issue has been explored sufficiently.

https://portswigger.net/burp/ai/capabilities

AIによる誤検知の低減

SECURE CYCLE

AI-powered false positive reduction

Sifting through false positives can be a huge drain on already stretched AppSec teams. By leveraging Burp AI to perform advanced analysis, Burp Scanner is able to intelligently filter out false positives before they're reported. Note that this feature is currently only available for the Broken Access Control scan check.

Automate testing for broken access controls

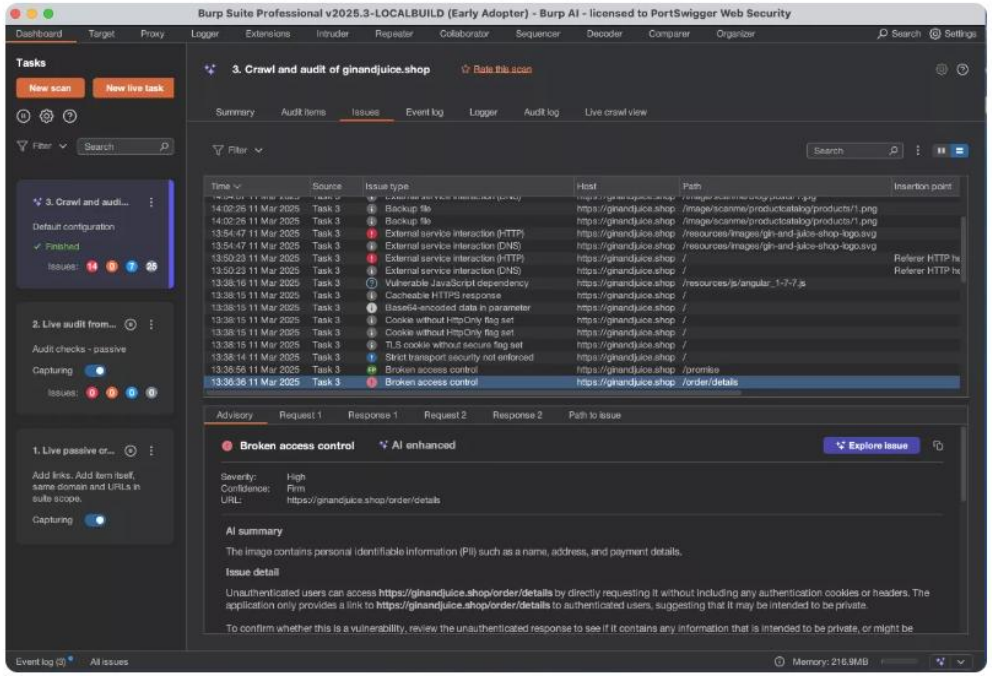
Testing for access control vulnerabilities is repetitive and tedious, but has traditionally proved challenging to automate reliably. Using AI-powered false positive reduction, Burp Scanner can now detect broken access controls with significantly more accuracy.

Less noise, more signal

Spend less time chasing dead ends and focus on investigating real vulnerabilities.

Validation before reporting

Burp AI helps validate access control issues before they're reported, ensuring you don't get distracted by an overwhelming to-do list of irrelevant findings.



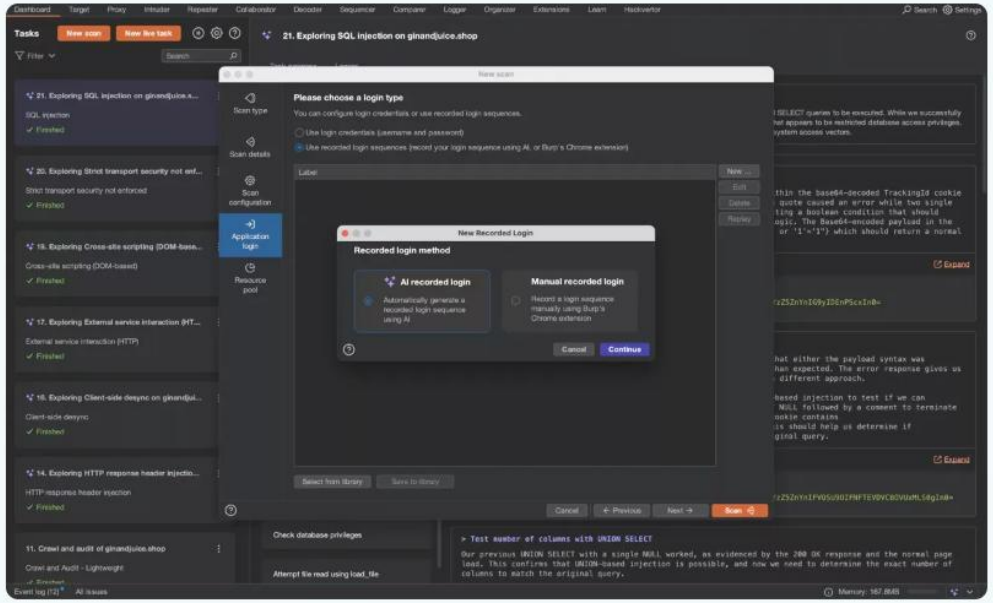
https://portswigger.net/burp/ai/capabilities

AIによるログインシーケンスの自動生成

SECURE CYCLE

AI-generated login sequences

Manually recording login sequences to handle complex authentication flows is powerful, but can be time-consuming and error prone. Burp AI can now generate recorded login sequences for you. All you need to provide is a valid username and password, and it handles the rest.



Simplified scan setup

Instantly generate recorded login sequences instead of manually navigating login flows in the browser.

Reliable authenticated scanning

Avoid common pitfalls of manual recording, such as missed interactions or unrecognized input methods, ensuring successful authentication during scans. Ensure Burp Suite can reliably access and scan authenticated areas, reducing blind spots in your security assessments.

https://portswigger.net/burp/ai/capabilities