JNSA ワーキンググループ紹介

調査研究部会
AI セキュリティワーキンググループ

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）

AIセキュリティワーキンググループについて

AIセキュリティワーキンググループは、今後さらに活用が広がっていくAIについて、AIに対するセキュリティとセキュリティ分野へのAIの応用の両方の方向から調査研究を行っております。2024年10月に再始動しまして、主に月一回の定例会とAIに関する成果物の作成を行っております。

AIセキュリティワーキンググループの活動

主に毎月1回1時間程度でオンラインにて定例会を行っており、最近のAIセキュリティに関するドキュメントの解説や各種海外カンファレンスの参加レポート、今年度の成果物についての議論等を行っています。その他、年に数回はオフラインでの定例の実施を行っており、その際は、2-3時間程度でオンラインでは実施しずらいトレーニング等を行っています。直近のオフライン会では、AIシステムに対する脅威モデリングのトレーニングを行っています。また、オンラインでの定例会のトピックの例は下記の通りです。

- 最近のトピックに関する議論（Operator, goose, DeepSeek, MCP, AIエージェント等）
- Black Hat Asia レポート（AIシステムに対するレッドチーミング等）
- RSA カンファレンスUSA2025レポート（RAGがデータセキュリティに及ぼす影響等）
- 「Multi-Agentic system Threat Modeling Guide v1.0」の解説
- Black Hat USA レポート（脅威インテリジェンスのための生成AI活用）

生成AIを利用する上でのセキュリティ成熟度モデル

2024年度はAIの中でも近年目まぐるしく進歩している生成AIのセキュリティに焦点を当て生成AIを活用していく上でのセキュリティの調査結果をAIセキュリティワーキンググループ内の有志で作成し、「生成AIを利用する上でのセキュリティ成熟度モデル」として2025年3月に公開しました。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。また、図1にその概要図を記載しています。

・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。

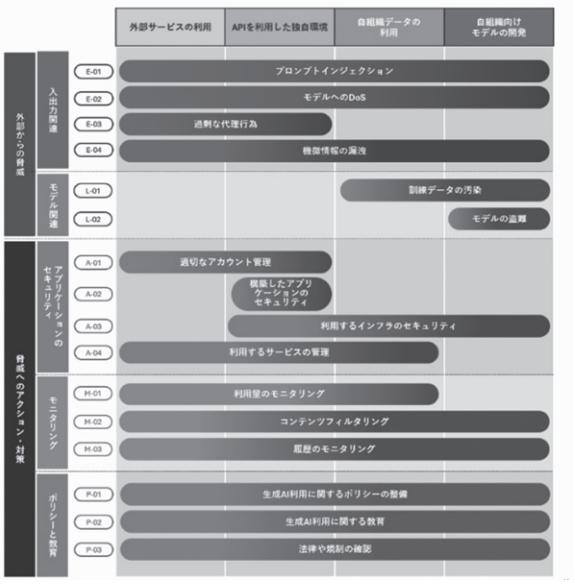

図1　生成AIを利用する上でのセキュリティ成熟度モデルの概要図

・**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。
・**自組織向けモデルの開発**
自組織向けにモデルを独自開発する組織。

今後の活動

今後の活動は、月一回の定例会を続けていくとともに、成果物としてAIを利用したシステムに対する脅威モデリング手法の評価を行う予定です。AIを利用したシステムに関する脅威モデリングは、マルチエージェントシステムの脅威モデリングフレームワークであるMAESTROや既存の脅威モデリングフレームワークであるSTRIDEをAIエージェント特有の課題に対応する形で改良した手法などがあり、これらの手法を使い同じモデルに対して評価を行うことにより、各手法のメリット・デメリットをまとめる予定です。また、作成したモデルについても公開すると共に文章自体の英語での公開も検討しています。

おわりに

AIセキュリティワーキンググループでは、今後もAIの活用について調査研究を進めていく予定ですので、AIに関するセキュリティに興味のある方は是非AIセキュリティワーキンググループにご参加ください。

※事務局注:ワーキンググループへのご参加は原則としてJNSA会員企業ご所属の方に限らせていただいております。