CSAJ Congress 2024

AI-working group

AI セキュリティ

はじめに

- **本講演は、CSA日本支部における AI-working group の活動を紹介するものです。**
  - ✓ **注）CSA（Cloud Security Alliance）とは、米国本部を中心に、世界60支部をもつ非営利団体です。クラウドセキュリティ分野におけるベストプラクティスを政府機関や企業団体に提供しています。**
  - ✓ **注）CSA日本支部は、2010年世界で2番目のCSA公認支部として設立されました。**

- **目次**
  - **① AI-working groupの紹介**
  - **② 情報セキュリティとAIセキュリティの違い**
  - **③ ハードローとソフトローの違い**
  - **④ AI組織の責任（RACIモデル）**
  - **⑤ AIレジリエンス（レジリエンススコア）**
  - **⑥ AIセキュリティとプライバシー（ガイダンス、ベストプラクティス、ガードレール）**

① AI -WGの紹介
（目標と成果）

- CSA日本支部のAI-WGは、現在17名が活動しているworking-groupです。
  - 米国CSA本部のAI-WGと連携し、調査研究や啓蒙活動を展開しています。

  - AI-WGの目標は、
    - AIセキュリティに関わる、ガイドライン、ベストプラクティス、ガードレールを提供すること。
      - ✓ ガイドライン　（法律や内規を守るための指針）
      - ✓ ベストプラクティス　（評価された手法、成功事例も含む）
      - ✓ ガードレール　（問題のある操作をしないように、警報や防止をする仕組）

  - これまでの成果物は、
    - AI Organizational Responsibilities - Core Security Responsibilities
    - AI Resilience - A Revolutionary Benchmarking Model for AI Safety
    - AI Model Risk Management Framework
    - AI Cybersecurity - Google Cloud
    - AI LLM Model Threats Taxonomy, etc.

      ✓ 注）成果物は、CSA会員専用ページから何時でも入手できます。

① AI -WGの紹介
（ 今、取扱っている課題1 ）

- 今、 AI-WGで取扱っている主な課題は、

  - AI組織の責任
    （RACIモデル、報告指標、報告メカニズムなど）
    ✓ 注）メカニズムとは、機械のように指標が伝わって目的を達成する仕組の事。

  - AIレジリエンス
    （産業別規制と課題、ベンチマークモデル、レジリエンススコアなど）

  - AIデータプライバシーとセキュリティ
    （バイデン大統領令、 NISTガイドライン 、ベストプラクティスとガードレールなど）

② 情報セキュリティとAIセキュリティの違い（要素の違い1）

- 情報セキュリティの要素は、「CIA」で、**脅威となる攻撃から防御**すること。

○ ＜情報セキュリティの基本3要素＞

1. Confidentiality（機密性）	許可してない者にアクセスさせない事
2. Integrity（完全性）	破壊や消去及び改ざんされないようにする事
3. Availability（可用性）	許可された者が常に利用できるようにする事

○ ＜情報セキュリティの追加4要素＞

4. Authenticity（真正性）	利用者や情報が本物であることを明確にする事
5. Reliability（信頼性）	意図した通りの動作・結果を得られる事
6. Accountability（責任追跡性）	利用者やシステムの行為を記録し追跡する事
7. Non-repudiation（否認防止）	作成者・利用者の行為を否認できない事

② 情報セキュリティとAIセキュリティの違い（要素の違い2）

- AIセキュリティの要素は、「TAA」で、**誤用・悪用から保護**すること。

○ ＜AIセキュリティの基本3要素＞

1. Transparency（透明性）　AIアルゴリズム、学習データ、機能などの透明性を確保する事
2. Accountability（説明責任）　AI品質やハルシネーションに対する説明責任が果たせる事
3. Auditability（可監査性）　AIシステムが監査可能な方法で日々運用され、監査やレビューが効果的に出来る事

○ ＜AIセキュリティの追加4要素＞

4. Data Operations（データ運用性）　データの取得と学習について説明責任を果たす事
5. Model Operations（モデル運用性）　モデルの調達と作成について評価責任を果たす事
6. Model Deployment and Serving（モデル展開）　自動スケーリング・レート制限・監視の実装責任を果たす事
7. Operations and Platform（運用プラットフォーム）　プラットフォームのCI/CDに保証責任を果たす事

② 情報セキュリティとAIセキュリティの違い（規制の違い１）

- 情報セキュリティの規制は、**主に「ハードロー」と「規格認証」**。

- ＜情報セキュリティの規制＞

1. ハードロー（しかし、法令範囲に限界、海外では他国の通信も収集）
   - サイバーセキュリティ基本法　2024年改正；NISCの強化と改組、サイバー空間、海洋宇宙空間、電磁波領域への取組
   - 不正アクセス禁止法　2024年改正；ID不正取得の罪を新設など
   - 電子署名法　2024年改正；電子文章と電子署名は実体文書と同じ
   - マイナンバー法　2024年改正；かざし利用の推進など

2. ソフトロー（しかし、マークを取得しても漏洩は減少していない）
   - ISO27001国際規格　国際標準化機構が定めた、情報セキュリティの国際規格。
   - ISMS認証（申請部署が対象）　ISO27001と同等の基準に適合している事を認証機関が審査して与える認証。
   - Pマーク認証（書類作成が大）　JIS15001の基準と個人情報保護法に適合している事をJIPDECの認定機関が審査して与える認証。

③ 情報セキュリティとAIセキュリティの違い（規制の違い2）

- AIセキュリティは、**主に「ソフトロー」と「ベストプラクティス」**。

- ＜AIセキュリティの規制＞

1. ハードロー（**法規制は国内外とも未整備**）
   - 日本では、法規制は未だ存在せず、ガイドラインが定められたのみ。
   - 米国では、連邦法は未だ存在せず、カリフォルニア州法案は2024年に拒否権発生。
   - EUでは、EU-AI法で、EU市場の安全性を確保する目的で、AIシステムの定義を定め、最大で年間売上高の7%または3500万ユーロの罰金を、2025年から段階的に適用。

2. ソフトロー（**各国のガイドラインは始まったばかり**）
   - 日本では、内閣サイバーセキュリティ戦略本部とデジタル庁が政策、NISCが監視と分析、IPAの中に2024年AISIを創設しAI安全性の評価。
   - 米国では、 2023年バイデン大統領令のAI安全性や連邦政府のAI利用など8つの措置を定め、NISTがガイドライン策定、CSAと専門家がベストプラクティスとガードレールを助言、2023年AISI（米国AI安全研究所）を創設。
   - EUでは、EU-AI法に基づくガイドラインを2026年から適用予定(リスクベースアプローチ、信頼できるAI推進、汎用AIモデルの規制、罰則規定、イノベーション促進、国際協力の推進）

③ 情報セキュリティとAIセキュリティの違い（AI向け攻撃手法）

- 代表的なLLM向け攻撃手法

保護すべき資産		概要	各資産に対する攻撃	例
LLMシステム		LLMシステムや、出力結果を処理するサービス	アプリが動作するプラットフォームの脆弱性を悪用する	・構成要素の脆弱性の悪用
LLMシステムを構成する要素	訓練データ	モデル開発のデータ（訓練データ、テストデータ）	訓練データやテストデータを改竄する	・データポイゾニング
	モデル	入力データに対し、出力結果を導き出す仕組み	訓練済みモデルの改変、訓練プログラムの改変、改変したプログラムの配信	・モデルポイゾニング ・モデル抽出/窃取
	クエリ	LLMシステムに出力を生成させる命令文（入力プロンプト、システムプロンプト）	特定の反応を引き出すために、LLMシステムに細工したクエリの送信を連発する	・直接プロンプトインジェクション ・プロンプトリーキング
	ソースコード	モデル開発のプラットフォームやソースコード	ライブラリのオープンソースコードに細工する	・バックドアポイゾニング
	リソース	アプリ実行時にLLMが取り込む文章、webページなど	アプリ実行時にLLMが取り込むリソースに細工する	・間接プロンプトインジェクション

③ 情報セキュリティとAIセキュリティの違い（AI向けセキュリティ製品の例）

- 代表的なAI向けセキュリティ製品

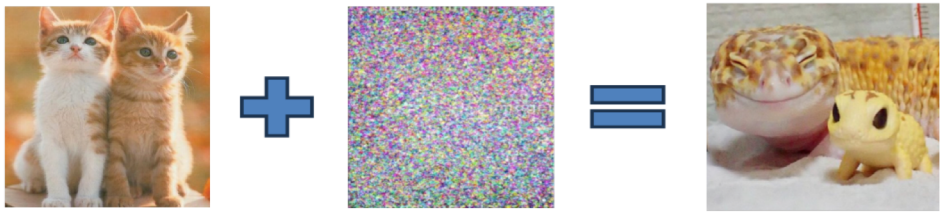
  - McAfee MVISION Cloud Guard for AI:
    McAfee社が提供するAIシステム向けのセキュリティプラットフォームです。AIモデルの保護、データの保護、AIシステムの脆弱性診断、AIシステムの不正検知などの機能を提供。

  - IBM Cloud Security for AI:
    IBM社が提供するAIシステム向けのセキュリティサービスです。AIモデルの保護、データの保護、AIシステムの脆弱性診断、AIシステムの不正検知などの機能を提供。

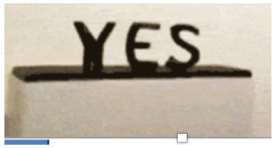
  - Google Cloud Armor:
    Google Cloud Platform上で動作するAIシステム向けのセキュリティサービスです。AIシステムへのDDoS攻撃や不正アクセスから保護。

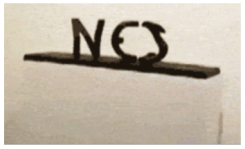
参考
（今までも、社会的弱点を悪用する例）

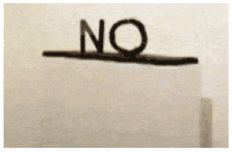
- サブリミナル効果；写真にノイズを混入し、AIに学習させる。
  ＞AIから見ると「いもり」が写り、誤った出力を回答してしまう。
  ✓注）以下のノイズ画像はサンプルであり、本物ではありません

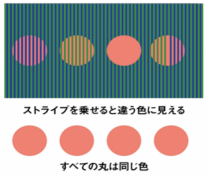


- 錯視画効果；物体は３次元、網膜には２次元映像、脳が認識するのは錯視画像
  ＞「左側で見るとYES」、「中央でNCS」、「右側で見るとNO」、ムンカー錯視









- ザイオンス効果；TVコマーシャルに画像を挿入し、視聴者に何度も見せる。
  ＞脳から見ると「意識せず興味や好意」を抱いてしまう。
  ✓注）瞬き0.2秒、2秒間もあれば強く記憶し、新たな信者の獲得になった例。

③ 情報セキュリティとAIセキュリティの違い（AI向けセキュリティ構築１）

- データセキュリティ（**主に、暗号化と認証**）
  - 暗号化：機微データを保存時、移動時、通信時および使用時に暗号化する手法。
  - 認証プロトコル：多要素認証やゼロトラストセキュリティモデルなどを採用し、機微情報やAI機能へのアクセスを正規ユーザーに厳格許可し、不正アクセスのリスクを軽減する。
  - 定期的な監査：定期的なセキュリティ監査と脆弱性評価を実施し、潜在的なセキュリティリスクを特定し軽減する手法。

- 敵対的攻撃に対するセキュリティ：
  - ART（Adversarial Robustness Training）：意図的に敵対的な例を使って生成AIモデルを訓練する手法。
    - ✓ 注）ARTは、計算コストが膨大で、学習分布外の敵対的攻撃に対し、実用的な限界を示す研究者もいる。
  - セキュリティテスト：敵対的攻撃シミュレーションを定期的に実施し、 脆弱性を特定し、モデルの防御力を向上させる手法。

③ 情報セキュリティとAIセキュリティの違い（AI向けセキュリティ構築2）

- プライバシー保護（**主に、プライバシーバイデザインとPET**）
  - プライバシー・バイ・デザイン：
    データの最小化、同意、セキュリティなどを生成AIシステムの設計と実装に直接組み込む手法。

  - PET（Privacy-enhancing technologies）；
    機微性の高いユーザデータを保護するプライバシー強化技術を採用する手法。
    - ✓ 差分プライバシー：計算されたノイズをデータセットに加え情報を匿名化する技術。個人のプライバシーを保護しながら分析を行う手法。
    - ✓ 統合された学習：複数のデバイスやサーバーに分散されたデータで生成AIモデルを訓練する技術。機微データを一箇所に集める必要性を回避する手法。
    - ✓ 準同型暗号：暗号化されたデータを復号せずに計算する技術。機微情報をセキュアに分析する手法。

③ 情報セキュリティとAIセキュリティの違い（AI向けセキュリティ構築3）

- 人間による監督（**主に、人間の介入と報告メカニズム**）
  - ヒューマン・イン・ザ・ループ：重要な意思決定プロセスにおいて、人間の関与を維持するために人間の介入を許可する手法。
  - フェイルセーフ・メカニズム：必ず侵入されることを前提とし、異常時に安全側に導く設計手法。特にエスカレーションパスを加える事で、ユーザ報告にも対処し易くする。

- 人間に説明責任を果たさせる
  - オーナーシップと責任：開発、デプロイ、監視の役割と責任を定め、個人に責任を持たせる手法。
  - 監査証跡：モデルの開発、トレーニング、および使用に関するログを収集し維持する手法。
  - 報告メカニズム：社内外の**利害関係者**がAIシステムに関する懸念事項や潜在的な課題を報告できるオープンなチャネルを構築する手法。
  - インシデントレスポンス：インシデントレスポンス計画と報告メカニズムを確立し実装する。
  - 倫理審査委員会：影響評価と企業価値観との整合性をとる、倫理委員会や審査委員会を設置する手法。
  - バイアスと公平性の監査：データセット、アルゴリズム、および結果における潜在的なバイアスを特定し、緩和するための監査を定期的に実施する手法。

③ ハードローとソフトローの違い（拘束力から観た違い）

- ハードローとは（**法的拘束力＝法的義務と罰則の規則**）
  - 憲法、法律、政令、省令、条例、法令で定めた内規
  - 法的拘束力は、明確に持ち、違反すると法的制裁を受ける
  - 規範は、国や地方自治体の立法機関によって制定される
  - 執行と強制力は、裁判所等が手続や証拠を判断し、その執行が強制力を持つ
  - 適用範囲は、一般的に広範囲に適用され、変更には通常、法改正のプロセスが必要

- ソフトロー（**自主的拘束力＝努力義務と社会的制裁の規範**）
  注）但し労働協約など法令で定めた内規には法的義務が発生する）
  - 政府のガイドライン、国際的な規範、企業や産業界の自主規制
  - 法的拘束力は、持たず、不履行になると社会的信用の失墜がある
  - 規範は、国際標準化組織等によって定められ、第3者機関が監査を行う
  - 執行と強制力は、証拠がなくても履行されるが、その執行に対し強制力が無い
  - 適用範囲は、特定の組織や業界内で適用され、状況に応じて変更や適応が可能

③ ハードローとソフトローの違い（側面から観た違い）

- ハードロー（**適用の側面は国家/加盟国の管轄権内**）
  - 法規制の適用;「AIセキュリティは国内外の法規制も考慮すべき」
    - ✓ 域外移転;EU越境移転規制、中国データ越境安全評価弁法。域内でデータ漏洩すると、域外企業で有っても規制対象となり得る。
    - ✓ 著作権;日本では音声が著作権に明確に含まれて無いが、海外では対象となるのが多い。
    - ✓ 管轄権;日本は属地主義で、立法・司法・執行権が自国領域のみで有効。日本企業が海外に在るAIデータセンタをクラウドサービスとして利用していると、他国の司法や執行権の対象になり得る。

- ソフトロー（**適用の側面は当事者間の合意管轄権内**）
  - ✓ 脅威の違い;「情報は技術的脅威、AIは社会的脅威（生命・身体・財産への危害）」
  - ✓ 要素の違い;「情報はCIA要素、AIはTAA要素（透明性・説明責任・監査可能性）」
  - ✓ 防御の違い;「情報は物理的境界防御、AIは時間的非境界防御（MLOps開発・ディプロイ・運用・廃棄/削除）」

④ AI組織の責任（RACIモデル）

- RACIモデルについて
  - AIセキュリティのフレームワークとしてRACIモデルを採用し、各担当者の役割と責任を定める。
    - R (Responsible)：実行責任者＞データ管理部/開発部などタスク遂行の直接責任者
    - A (Accountable)：説明責任者＞最高データ責任者/開発マネージャなどタスク結果の最終責任者
    - C (Consulted)：相談先＞法務部/DevOps部などタスクに関する相談先
    - I (Informed)：報告先＞IT部/データサイエンティスト部などタスク情報の報告先
      - ✓ 注）主なタスクは、（真正性管理、匿名化、データ最小化、アクセス制御、保管と送信）とし、これらのタスク毎にRACIモデルを設定している。
  - RACIモデルを採用したマトリックスは、以下の手順で作成する。
    1. RACIの役割と責任を定める
    2. プロジェクトの主要なタスクを洗い出す
    3. 各タスクに対して、RACIマトリックス図を作成する
    4. RACIマトリックス上にて、RACIの役割者と責任者を決める
    5. RACIマトリックスを、プロジェクトで共有し、ステークホルダーの合意を得る

④ AI組織の責任
（ 一般的なRACIマトリックスの作成例）

- RACIマトリックスの作成例
  - ✓ 注）AI向けRACIマトリックスは作業中の為、一般的なRACIマトリックス例を以下に示す）
  - ○ RACIマトリックスは、タスク毎に役割者を分ける。
    - R実行責任者とA説明責任者は、プロジェクトに1名づつとするが、タスクを兼務する事が多い。
    - C相談先の担当者とI報告先の担当者は、プロジェクトに複数名いて、問題/課題毎に担当者が異なる事が多い。

実行責任者Rと報告責任者Aを兼任する場合もある

相談先Cと報告先Iを兼任する場合もある

	佐藤	鈴木	高橋	田中	伊藤	渡辺	山本
企画書作成	R/A		I	C/I	R		
要件定義書作成	A	R	I	C/I	R		
機械設計書作成	A	R	I	C/I	R		
開発計画と実施	I	A		C		R	R
テスト計画と実施	I	A		C		R	R
運用設計	I	A		C		R	R
マニュアル作成	I	A		C		R	R
ユーザー対応	I	A		C		R	R

複数の実行責任者が割り当てられる場合もある

※役割を設定するときは報告責任者のAから決める

④ AI組織の責任
（AI責任共有モデルの縦軸）

- AI責任共有モデルの縦軸について、
  - AIアプリケーションの安全なオペレーションには、複数の利害関係者の協力が必要。
  - これら複数の関係者を、3つに分け、責任共有モデルの縦軸とした。

    1）AIプラットフォーム層（主にAIプラットフォームプロバイダー）
    - モデルのセキュリティ、モデルのチューニング、モデルの説明責任
    - モデルの設計と実装、モデルの学習とガバナンスの責任

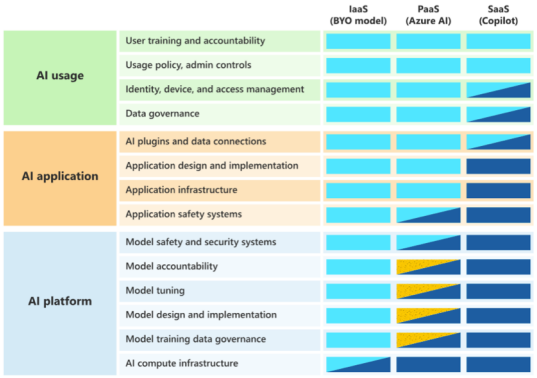
    2）AIアプリケーション層（主にAIアプリケーションの所有者と開発者）
    - AI プラグインとデータ接続、アプリケーションの設計と実装の責任
    - アプリケーション基盤、AI 安全システムの責任

    3）AIユーセイジ層（主にAIユーザ；使用者/利用者）
    - ユーザートレーニングと説明責任
    - 許容される使用に関するポリシーと管理者による管理責任
    - アイデンティティアクセス管理とデバイス制御の責任、データガバナンスの責任

④ AI組織の責任
（AI責任共有モデルの横軸）

- AI責任共有モデルの横軸について。
  - AI責任共有モデルは、横軸としてサービスモデル毎に定めた。
    - ✓ サービスモデルは、3分類（SaaS、PaaS、IaaS）とした。
    - ✓ 責任共有モデルとは、「管理主体が責任を持つ」と言う原則。
      - 原則の基、AIプラットフォームプロバイダー、AIアプリケーション所有者とAI開発者、AI-customer/AI-usage の間で、役割と責任を明確にし、職務分離を保証する。

        - 注）「AI-usage」は、2つの意味があると解釈し、注意して議論した。
          - ✓ AI usage＝使用者＞AI開発で、直接的にAIを使用する者（AIエンジニア、データサイエンティストなど）
          - ✓ AI usage＝利用者＞AIサービスを通して、間接的にAIを利用する者（経営者/従業員、消費者など）
          - ✓ 左記は、マイクロソフトの事例



⑤ AIレジリエンス
（AIレジリエンスの定義）

- AI レジリエンスについて、
  以下の3能力から構成されると定めた。
  - ✓ 抵抗する能力(レジスタンス)
  - ✓ 立ち直る能力(レジリエンス)
  - ✓ 元の状態に戻る能力(プラスティシティ;可塑性)

- 3能力で構成されたAI レジリエンスとは、
  - 「レジスタンス、レジリエンス、プラスティシティ」の3柱から構成され、以下に明示する。
    1. AI レジスタンスは、侵入、操作、誤用、悪用に直面した際、「必要最低限のパフォーマンスを維持し、抵抗するシステム能力」を言う。
    2. AI レジリエンスは、インシデン発生後、「要求される時間と容量および能力に関し、必要な最低限の性能に立ち直る能力」を言う。
    3. AI プラスティシティは、障害発生後、「“作るか壊すか” 迅速な対処を可能にするゲージの機能を示し、元の状態に戻る能力」を言う。

⑤ AIレジリエンス
（AIレジリエンスのベンチマーク1）

- AIベンチマークについて、
  - 既存の性能ベンチマークは、使用目的に対する適合性を、従来サービスの指標と、新サービスの指標で表し、比較分析するものです。

  - ベンチマークの課題は、
    - ✓ 一部のAIシステムは、既に人間のベースライン性能を上回っており、評価指標に問題点と限界が飽和状態に近づいており、比較できないものもある。

    - ✓ モデルの健全性、精度、ドリフト、バイアス、生成AIの品質を評価するサービスもあるが、自社サービスである事や、ディプロイ後の性能劣化を監視するものが多く評価不十分である。
      - 簡略化されたテストや多面評価不足
      - データセット依存と過学習のリスク
      - 異なるAIシステムを組み合わせたシステム間の評価ができない
      - 日本語専用のベンチマークの遅れ

⑤ AIレジリエンス
（AIレジリエンスのベンチマーク2）

- AIレジリエンス・スコアについて
  - AIレジリエンス・スコアの設計概念は、
    - ✓ 知能を比較するのではなく、知能の違いを理解することに重点を置いた「インテリジェンスの認識」と言う概念である。
      - G. C. v. d. B.-V. R. A. M. B. e. a. J. E. (Hans) . Korteling, “Human versus Artificial Intelligence,” *Front. Artif. Intell.,* vol. 4, 25 03 2021.
    - ✓ 未だ広く使われている概念ではないが、知的システムが人間の性能を凌駕するにつれて、そのAIベンチマークは極めて重要になる。

  - AIレジリエンス・スコアの数値化は、
    - ✓ レジスタンスとレジリエンスおよびプラスティシティ という 3 つの柱を考慮して、AIの回復力を反映する 0 から 10 までのレジリエンススコアを提案しています。
    - ✓ スコアは、例えば、16:5-8-3 のように、3 つの柱の合計と 3 つの柱のそれぞれを個別に表す事とする。
    - ✓ 3つの柱のスコアの分布は、異なるAIシステムの多様性を反映することができる。これにより、異なるAIシステムを組み合わせる場合、リスクとその軽減に関して、より情報に基づいた判断が可能になる。

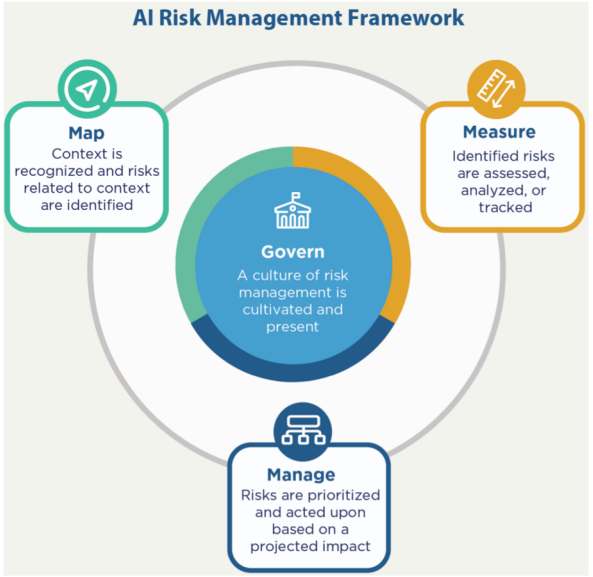
⑥ AIセキュリティとライバシー（バイデン大統領令 AI-8要請）

- AIセキュリティとプライバシーについて

  - 2023年のバイデン大統領令 AI-８要請；
    1. 安全性とセキュリティーの新基準 ＞NISTがテスト基準を設定（AI100シリーズ）
    2. 米国民のプライバシー保護 ＞データプライバシー法案を要請
    3. 公平性と公民権の推進 ＞差別を防ぐガイダンスの提供を要請
    4. 消費者、患者、学生の権利保護 ＞医療や教育分野で、活用ツールの導入支援を要請
    5. 労働者の支援 ＞危害軽減と利益最大化をする ベストプラクティスの開発を要請
    6. イノベーションと競争の促進 ＞研究リソースの運用と研究促進を要請
    7. 外国における米国のリーダーシップの促進 ＞政府機関でAI専門家の採用加速
    8. 政府によるAIの責任ある効果的な利用の保証 ＞AI利用に明確なガイダンスを発行

⑥ AIセキュリティとライバシー（NISTのガイドライン）

- 大統領令に基づく、NISTのガイドライン
  - NIST AI 100;2023
    - ✓ NIST新基準; AI 100-1 (AI-RMF),100-2 (AI-AML),100-3 (AI-LofT),100-4 (AI-CC),100-5 (AI-std) を発表
    - ✓ AI 100-1のAI-RMF(Artificial Intelligence Risk Management Framework) は、 AIリスク管理フレームワークとして、組織がAIシステムに関連するリスクをより適切に管理できるようにすることを目的としている。
    - ✓ AI製品、サービス、システムの設計、開発、使用、評価に信頼性を組み込む能力を向上させることができる。
      - 第1部は、AIリスクの枠組み
      - 第2部は、 AIリスクマネジメントのフレームワーク
        a. GOVERNとは、AI組織全体のガバナンス体制を確立するための機能
        b. MAPとは、AIの潜在的なリスクを特定するために使用
        c. MEASUREとは、AIリスクを包括的に測定・評価
        d. MANAGEとは、具体的な対応策を実装し、継続的なモニタリング
  - その他、
    - ✓ NIST SSDF 安全ソフト開発、
    - ✓ NIST SP 800-53 Rev.5 組織のセキュ＆プライバシー管理
    - ✓ ISO42001:2023AIMS(Artificial intelligence Management system)



⑥ AIセキュリティとライバシー（CSAのベストプラクティスとガードレール）

- NISTのガイドラインに基づく、ベストプラクティスとガードレールについて、
  - 例）データへのアクセス制御の場合
  - ベストプラクティス
    - ✓ RACIモデルは、（Responsible）セキュリティチーム、（Accountable）最高情報セキュリティ責任者、（Consulted）データガバナンス機関/データ管理者/IT チーム、（Informed）運用チームとする。
    - ✓ 評価基準は、年間0.5%未満の不正データアクセスインシデントの達成とする。
    - ✓ 実装戦略は、レイヤード・セキュリティ・モデルの実装とする。このモデルは、多要素認証（MFA）、役割ベースのアクセス制御（RBAC）、最小特権の原則（PoLP）などの先進技術も統合する必要がある。
    - ✓ 継続的な監視と報告：アクセスログを監視し、監査を実施する。
    - ✓ アクセス制御のマッピング：アクセス許可を定期的に監視・管理する。
  - ガードレール
    - ✓ ISO/IEC 42001、ISO/IEC 27001、ISO/IEC 27701、NIST 800-53、および OWASP Top 10 A07:2021-Identification and Authentication Failures を実施し、プライバシー＆データ保護とユーザ操作ミスなどを防止する。

FYI:
最新のCSA調査 “NHI: Non-Human Identity Security”

- CSA米国本部が「人間以外のアイデンティティ・セキュリティ」の現状調査を実施
  - 調査の目的
    - ✓ 調査の主な目的は、NHIの重要な側面について理解を深めることである。
      - 人間以外のアイデンティティに関する認識と懸念
      - 人間以外のアイデンティティに関するセキュリティ、プライバシー、APIキー管理とポリシー
      - サードパーティ・ベンダーとの接続に関する課題
  - 800人以上の業界専門家から得られた調査結果の例
    - ✓ 薬事規制（ヘルスケア／製薬／医療機器）の認識
      - 厳しい内規があるがグローバルではない。機械学習と生成AIの違いを明確にすることが要請されている。特にヘルスケアにおける生成AIは、さまざまなステークホルダーと相互作用し、セキュリティ、プライバシー、誤用や乱用を防止する対策が未だ十分でない。
    - ✓ 薬事当事者/専門家の意見
      - 薬事セキュリティ確保への不安は高く、自己肯定感（自信を持って行う感情）が低い
      - 権限とAPIキーの管理に課題がある
      - 断片的なアプローチがセキュリティ・インシデントに繋がっている、など

FYI:
最新のCSA調査 “NHI: Non-Human Identity Security”

ヒューマン・アイデンティティ
（個人/家族の氏名、生年月日など）

マシン・アイデンティティ
（Robot/IoT/OSの品名、ディプロイ年月日など）

nonヒューマン・アイデンティティ
（AI/自動運転の学習データ？LLMモデル？）

ご清聴ありがとうございました

これを機会に
CSA会員およびworking-groupへの参加をお待ちしてます
https://www.cloudsecurityalliance.jp/site/

プレゼン資料や成果物を作成する上で、

AI-working group メンバーの協力に感謝します