報道資料

令和５年 11 月 28 日
内閣府科学技術・イノベーション推進事務局
内閣官房内閣サイバーセキュリティセンター

セキュア AI システム開発ガイドラインについて

１．概要

内閣府科学技術･イノベーション推進事務局及び内閣サイバーセキュリティセンターは、英国国家サイバーセキュリティセンター（NCSC）が米国サイバーセキュリティ・インフラストラクチャー安全保障庁（CISA）等とともに作成した「セキュア AI システム開発ガイドライン」（以下「本件文書」という。）の「共同署名」（本件文書作成への協力機関として組織名を列記：日本のほか G7 各国を含む計 18 か国が参加）に加わり、本件文書を公表しました。なお、本件文書は広島 AI プロセスを補完するものであり、参考文書として、高度な AI システムを開発する組織向けの広島プロセス国際指針及び国際行動規範が記載されています。

本件文書は、セキュアバイデザイン（IT 製品（特にソフトウェア）について、セキュリティを確保した設計を行うこと）の観点から、ソフトウェアのうち AI に焦点を当てて、AI を使用するシステムのプロバイダーによるセキュアな AI システムの構築を支援するための指針となっております。

内閣府科学技術･イノベーション推進事務局及び内閣サイバーセキュリティセンターは、本件文書が､高度な AI システムを開発する組織向けの行動規範を作成する G7 広島 AI プロセスを補完し、AI のセキュリティ部分について詳述する文書であること、関係国に対する G7 広島 AI プロセスのアウトリーチ活動にも資すること、各国の関係機関と日本の協力関係の基礎となること等の事由により、共同署名に加わりました。

今後は、技術の進歩が早い分野であることも踏まえ、本件文書の具体化に当たり、産業界とも継続的に対話を重ねつつ、引き続き、AI 及びサイバーセキュリティ分野での国際連携の強化に努めてまいります。

２．本件文書の内容

（１）導入

- サイバーセキュリティは AI システムの前提条件。
- 本件文書は､AI を使用するシステムのプロバイダーのためのガイドラインを提言するものであり、意図したとおりに機能し、必要なときに利用でき、機密データを権限のない第三者に漏らすことなく動作する AI システムを構築することを支援する。
- AI コンポーネントのプロバイダーが､バリューチェーンの先にいるユーザのセキュ

リティ結果に責任を負う（セキュアバイデザイン）。

（２）AI ガイドライン

ア　セキュアな設計

- システムに対する脅威をモデル化する。AI 特有の脅威の影響を評価し、意思決定を文書化する。
- システム設計に際し、機能性やパフォーマンスと同等に、セキュリティを考慮する。
- アーキテクチャ、設定、訓練データ、訓練アルゴリズムなどの要件に基づいて AI モデルを選択する際に、セキュリティ上の利点とのトレードオフを考慮する。

イ　セキュアな開発

- サプライチェーンのセキュリティを確保する。
- 関連アセットを保護し、SBOM 等によりデータ、モデル、プロンプトを文書化する。

ウ　セキュアな導入

- システムのライフサイクルを通じて、使用するインフラのセキュリティを確保する。
- 継続的にモデルを保護し、機密情報へのアクセス、改ざん、外部への窃取の試みを検知し妨げるため、適切なコントロール措置を実践する。
- あらかじめインシデント管理手順を定め、適切かつ効果的なテスト後に AI システム等をリリースし、ユーザに適切な使用方法等を明示する。

エ　セキュアな運用とメンテナンス

- システムの挙動を監視し、潜在的な侵入、侵害等を検知できるようにする。
- デフォルトで自動化アップデートを実装し、モデル等の変更による挙動の変化につき、ユーザが評価できるようサポートする。

３．関連リンク

- 英国 NCSC ホームページ（Guidelines for secure AI system development）【ガイドライン和訳】

【本報道発表に関する問い合わせ先】
内閣府科学技術・イノベーション推進事務局
菅田参事官、藤井企画官、鈴木参事官補佐
Tel:03-6257-1337
内閣官房内閣サイバーセキュリティセンター
山口参事官、村田参事官、金井官、髙橋官
Tel:03-3580-3188