報道資料

令和６年１月 24 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

AI 使用に関する国際ガイダンスへの共同署名について

１．概要

本日、内閣府科学技術・イノベーション推進事務局と内閣サイバーセキュリティセンター(NISC)は、ともに、豪州通信電子局(ASD)豪州サイバーセキュリティセンター(ACSC)が、カナダ、ニュージーランド、英国、米国の関係当局とともに作成したAI 使用に関する国際ガイダンス”Engaging with Artificial Intelligence(AI)”（以下「本件文書」という。）の「共同署名」（本ガイドライン作成への協力機関として組織名を列記：日本のほか、10 か国が参加）に加わり、本件文書を公表しました。

（注：昨年 11 月に我が国が共同署名した「セキュアAI システム開発ガイドライン」は、AI システムの「開発」に関するガイダンスを示す目的で作成されたのに対し、本件文書は AI システムの「使用」に関するガイダンスを示す目的で作成されたもの。）

なお、本件文書の参考文献として、昨年 12 月の G7 デジタル・技術会合で合意され、同月に G7 首脳が承認した「広島 AI プロセス包括的政策枠組み」が記載されています。

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、本件文書は関係のサイバーセキュリティ当局の有する周知の知見を整理し、AI システムの利用者に注意喚起する内容であり、我が国の利用者にも資することなども踏まえ、共同署名に加わりました。

今後も、引き続き、産業界とも継続的に対話を重ねつつ、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）本文：AI システムに対する脅威を 6 つ列挙の上、これに対する 12 の緩和策を注意喚起するもの。

（ア）導入

この文書は AI システムを安全に「使用」するガイダンスを提供。なお、安全な AI システムの「開発」については、（英米が主導した）文書”Guidelines for Secure AI System Development”を参照願いたい。

（イ）AI とは

機械学習、自然言語プロセス、生成 AI を説明し、AI が意図的又は過失で被害をもたらすリスクがあり、リスク管理の必要性を強調。

（ウ）AI に関する課題

脅威として、①データポイズニング、②インプット改ざん攻撃（プロンプトインジェクション・敵対的サンプル）、③生成 AI ハルシネーション、④プライバシー・知的財産に関する懸念、⑤モデル窃取攻撃・学習データ漏えい、⑥匿名化データの再特定を列挙。

（エ）緩和策

AI システムに関し、それぞれ、①サイバーセキュリティ枠組みの実施、②プライバシー・データ保護義務への影響評価、③多要素認証の実装、④特権アクセスの管理、⑤バックアップ、⑥試行、⑦サプライチェーンを含むセキュアバイデザイン確保、⑧限界や制限の理解、⑨関係スタッフの能力・資格、⑩検査・ヘルスチェック、⑪ログ監視、⑫インシデント対応を列挙。

（２）参考文献：各国のサイバーセキュリティ枠組み、AI セキュリティに関する文書を列挙。（例えば、広島 AI プロセス包括的政策枠組みや、米国 NIST の Cybersecurity Framework, AI Risk Management Framework 等。）

３．関連リンク

豪州 ACSC ホームページ "Engaging with Artificial Intelligence(AI)" 【国際ガイダンス仮訳】

【本報道発表に関する問い合わせ先】
内閣官房内閣サイバーセキュリティセンター
基本戦略総括グループ
国際戦略グループ