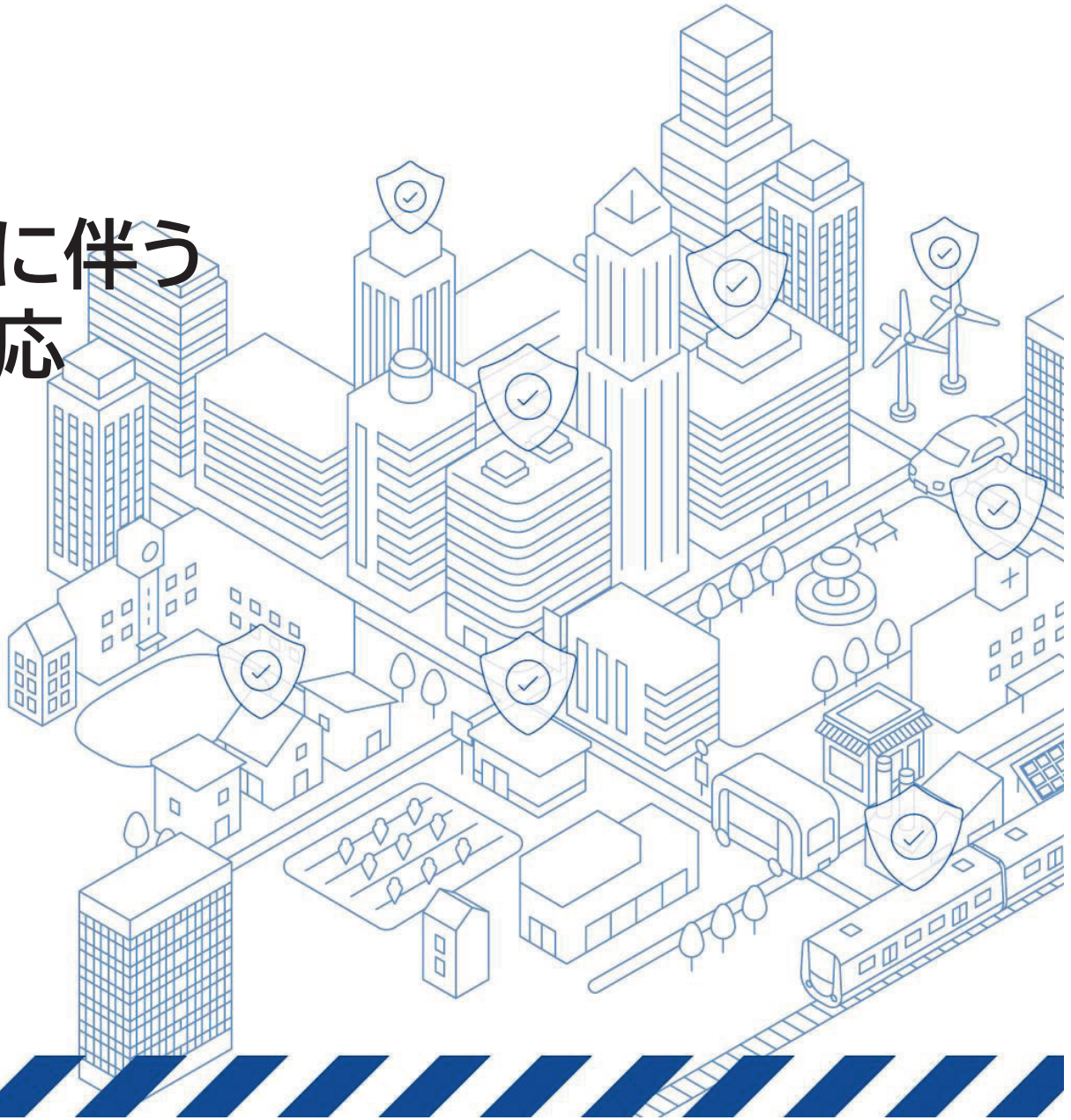
資料１

AI技術の進展・普及に伴う サイバー脅威への対応



2026年3月19日
内閣官房
国家サイバー統括室

AI技術の進展・普及に伴うサイバー脅威への対応

- **AI技術の急速な進展・普及**により、サイバーセキュリティにおける新たな脅威に直面。
  - 脆弱性探索やマルウェア生成、侵入・横展開等の一連の攻撃行為にAIが活用されることで、**攻撃のスピード・規模が劇的に増加**し、**攻撃者優位をより助長する懸念**。
  - また、AI利活用の更なる進展に伴い、**AI自体のセキュリティ確保もより一層重要に**。
- こうした**AI技術の進展・普及に伴うサイバー脅威に対応する**ため、下記**3つの観点**から、**施策の具体化**を進めていくことが必要。

①AIを活用したサイバーセキュリティ確保（AI for Security）

AIを利用してサイバー防御を強化する

総合的に推進

③AIを悪用したサイバー攻撃への対応

AIを悪用して行われるサイバー攻撃に対応する

②AIに係る安全性確保（Security for AI）

サイバー攻撃からAI自体を防御する

AIを利用したサイバー関連主要トピック

●Microsoft 「デジタル防衛レポート」（2025.10）

国家アクターによるAI利用したサイバー攻撃が増加している。足下では、特定のナラティブ拡散や、学習データ汚染、他者へのなりすまし等影響工作へのAI利用が急増。

●アンソロピック 「初めて報告されたAI主導型サイバー諜報活動」（2025.11）

Claudeを活用して複数の組織を標的にサイバー攻撃を実施。脅威アクターはAIを使ってサイバー攻撃の80~90%を実行し、人間の介入は重要な意思決定ポイントのみ。経験やリソースが少なくても、効率的に大規模攻撃が可能に。

●英国AISI 「フロンティアAIトレンドレポート」（2025.12）

フロンティアAIモデルの性能評価。サイバー領域では、一部まだ改善の余地はあるものの、10年以上の経験を要する専門家レベルのタスクを完了できる初のAIモデルが登場。AIが人間の補助なしで完了できる範囲は、約8カ月ごとに倍増。

●ダボス会議 「AI時代のサイバー防御」等（2026.1）

AIによる攻撃の量・スピード・巧妙さが加速しており、攻撃者の優位性が拡大。
また、技術自立の必要性や、地政学的観点でのサイバー防御力強化について議論。

●OpenAI 「AIの悪用を阻止する」レポート（2026.2）

AIを悪用した国家主導のサイバー攻撃の「手口」を分析。中国系のサイバー攻撃グループが、国内外の敵対者に対する影響工作の進捗状況の定期報告書の作成にAIを利用。また、生成AIを使った高市総理の加工画像が、極右とのつながりを示唆するコメントとともに、SNS上で拡散。

（出典）各種公表情報を元にNCO作成

国際動向：AI政策

米国

- **Winning the Race: AMERICA'S AI ACTION PLAN［2025年７月発表］**
  - ①AIイノベーションの加速、②米国AIインフラの整備、③国際的なAI外交・安全保障の主導の3本柱で構成される国家戦略。

①	**AIイノベーションの加速**	官僚的手続き・過度な規制の撤廃、オープンソース・オープンウェイトAIの促進、世界水準の科学データセット構築など
②	**米国AIインフラの整備**	データセンター・半導体工場・エネルギーインフラの迅速な許認可、AIに対応した電力網の整備、半導体製造の国内回帰など
③	**国際的なAI外交・安全保障の主導**	フルスタックでの米国AI技術の同盟国・パートナー国への輸出、AI計算資源の輸出管理強化など

EU

- **EU AI Act［2025年8月施行］**
  - 人間中心の信頼できるAIの導入促進と、AIシステムの有害な影響に対して健康・安全・民主主義・法の支配・環境保護等の基本的権利の保護・確保、イノベーション支援を目的とする。
  - リスクベースアプローチを採用。４つのリスクレベル（Unacceptable、High Risk、Limited Risk、Minimal Risk）を設け、各々のリスクに応じた要件・規制を設定するとともに汎用AIモデルに関する規律を規定。提供者だけでなく導入者にかかる要件も存在。
- **Europe's AI leadership with an ambitious AI Continent Action Plan［2025年４月発表］**
  - 欧州がAI分野での世界的リーダーとなることを目的として、①大規模AIデータ・インフラの構築、②大規模・高品質データへのアクセス拡大、③EU戦略部門でのアルゴリズム開発とAI採用の促進、④AIスキル・人材強化、⑤規制の簡素化の5つの柱から構成される。

（参考）国内動向：AIに係る安全性確保

- **AI事業者ガイドライン**［総務省・経産省：2025年３月］
  - 人間中心、安全性、公平性、プライバシー保護、セキュリティ確保、透明性、アカウンタビリティ等の観点から、AI開発者・AI提供者・AI利用者別の取組事項等を整理。
- **AIセーフティに関する評価観点ガイド**［AISI：2025年３月］
  - AIセーフティの評価の観点を、有害情報の出力制御、偽誤情報の出力・誘導停止、公平性と包摂性、ハイリスク利用・目的外利用への対処、プライバシー保護、セキュリティ確保、説明可能性、ロバスト性、データ品質、検証可能性とし、評価は基本的にAI開発・提供管理者が実施すると整理。
- **AIシステムに対する既知の攻撃と影響**［AISI：2025年3月］
  - AIシステムに対する特有のセキュリティ攻撃を俯瞰するため、学術論文等で発表された攻撃とその影響を、攻撃の類型別に整理。AIシステム特有のセキュリティ対策が必要不可欠とした。
- **AIのセキュリティ確保のための技術的対策に係るガイドライン**
  ［総務省：2025年度末策定予定］
  - AIに対するプロンプトインジェクション攻撃、DoS攻撃等の攻撃の類型別に対策の概観等を整理
- **行政の進化と革新のための生成AI活用の調達・利活用に係るガイドライン**
  ［2025年5月：デジタル社会推進会議幹事会決定］
  - 政府における生成AIの調達・利活用に係る対応事項等を整理し、調達・契約チェックシート等を提供

本日の議論の論点

- AI技術の急速な進展・普及により、サイバーセキュリティ確保における**新たな脅威**に直面。
  - 脆弱性探索やマルウェア生成、侵入・横展開等の一連の攻撃行為にAIが活用されることで、**攻撃のスピード・規模が劇的に増加し、攻撃者優位をより助長する懸念**。
  - また、AI利活用の更なる進展に伴い、**AI自体のセキュリティ確保もより一層重要に**。
- こうした中で、欧米諸国では、AI利活用推進の観点も踏まえつつ、**AIセキュリティに関する官民連携の強化（米）**や**AIに対する規制強化（欧州）**等の方針が示されているところ。
- 我が国においても、引き続き、AI基本計画やサイバーセキュリティ戦略等を踏まえ、AI技術の進展・普及に伴うサイバー脅威に的確に対応するため、**①AIを活用したサイバーセキュリティの確保（AI for Security）**、**②AIに係る安全性確保（Security for AI）**、**③AIを悪用したサイバー攻撃への対応**の３つの観点から取組を強力に推進していくことが必要。
- 本日は、それぞれの観点（上記①～③）における**課題認識**や**取り組むべき施策の方向性**について御意見頂きたい。

（論点例）

- 政府機関におけるAIを活用したサイバー対処能力強化について
- AIセキュリティに関する官民連携の強化について　　等