

AI の脅威に対する米国の取り組み

- 1) 米国 AI 安全研究所の新設：商務省傘下の国立標準技術研究所（NIST）に、米国 AI 安全研究所（US AISI）を設立する。同研究所は、AI のリスク特定・軽減に向けたガイドンスやツール、ベンチマーク、ベストプラクティスの作成を通じて、AI により生じるリスクを管理するための全体的な枠組み作り・運用を所掌する。この所掌に沿って、まずは規制当局が今後 AI に関する規則の策定・施行を検討する際に必要となる技術的なガイダンスの開発に取り組んでいく。
- 2) 米国政府による AI 利用に関する政策ガイダンスの策定：国民の権利と安全に影響を与える AI の使用に対しセーフガードを設けるべく、米国政府の AI 利用に関するガイダンスを策定する。このガイダンスは、「AI 権利章典のための青写真」や、NIST の「AI リスク管理フレームワーク」など、これまでに発表されたガイダンス（[2022 年 10 月 5 日記事](#)、[2023 年 2 月 1 日記事参照](#)）に基づくもので、保健、教育、雇用、連邦政府の支援、法執行、移民、交通、重要インフラ等幅広い分野をカバーする。政府における責任ある AI イノベーションを推進、透明性と説明責任の向上、連邦政府職員の保護、AI の機密利用によるリスク管理に向けた具体的なステップの概要を示すものとなる予定で、この中には AI の影響評価の実施・リスクの特定・監視・軽減措置や、AI オペレーターの訓練、AI 活用に当たっての公告及び協議の手順、AI による被害を申告するための手法の提供などが盛り込まれている。本ガイダンスについては、行政管理予算局（OMB）から[草案](#)が発表されており、2023 年 12 月 5 日まで[パブリックコメント](#)が受け付けられている。
- 3) AI と自律性の責任ある軍事利用に関する政治宣言：同宣言は 2023 年 2 月に発表されたもので、世界各国が AI 機能（軍や防衛施設の自律的機能やシステムを可能にするもの

を含む) の利益を責任ある合法的な方法で活用できるようにするために、軍事 AI 機能の責任ある開発、展開、使用に関する一連の規範を定めたもの。11 月 1 日までに日本を含む 30 カ国が参加したと発表した。

- 4) 公益の利益のために AI を推進する資金提供者イニシアチブ：AI に関連する 10 の慈善団体は、世界の一般市民の最善の利益を促進することを目的とした AI イニシアチブに総額 2 億ドル以上の投資を発表。また、同団体は 5 つの柱（1. AI が民主主義と権利を保護することを証明する、2. 公益の利益のために AI のイノベーションを推進する、3. AI 主導の変化の中でも労働者が活躍できるようにする、4. AI の透明性と説明責任を向上させること、5. AI に関する国際的なルールと規範を支持する）に沿った活動を推進するために、新たな慈善寄付を調整するための資金提供者ネットワークを形成する。
- 5) AI による不正電話の検知とブロック：ホワイトハウスは、詐欺に使われる可能性のあるロボコールやロボテキストを検出してブロックできる AI モデルを構築する目的で、バーチャルハッカソンを開催し、AI による防御で消費者をロボコールから守る。
- 6) コンテンツ認証に関する国際規範：米国は各国に対し、各国政府が作成するデジタル・コンテンツと AI により生成・操作されたコンテンツを識別・追跡できるようにし、欺瞞的で有害なメディアに対する回復力を確保するべく、デジタル署名をはじめとする国際基準の開発・実装を支援するよう求めている。
- 7) 政府の AI 開発、調達、利用に責任ある権利尊重の慣行を取り入れる誓約：国務省を通じて、世界 38 カ国からなるフリーダム・オンライン連合（FOC）と協力し、政府の開発、調達、AI の利用に責任ある権利尊重の慣行を取り入れる誓約書を作成する。

以上