March, 2026

THY
Taiwan International
Law Offices

NEWSLETTER

米国における人工知能リスクマネジメントフレームワークの紹介

近年、人工知能（Artificial Intelligence, AI）は急速に発展し、ビッグデータ分析とアルゴリズムによるディープラーニング（深層学習）を通じて、人間の知能を模倣することができ、情報の統合、精緻な分析、継続的な学習及び自律的な行動といった機能を発揮するに至っている。AI が示す自動化と高効率という特性は、各学問分野の発展に飛躍なブレークスルーをもたらし、その応用範囲は統計学、ソフトウェア工学から言語学、神経科学、さらには哲学や人文科学の領域にまで広がり、とどまるところを知らない。しかし、AI は強力なビッグデータ分析能力及び表現能力を持つがゆえに、その応用過程においては個人情報の収集と秘密保持、倫理的問題、差別的言動といったリスクを伴う可能性がある。そのため、各国は AI の発展を推進すると同時に、潜在的な危害をいかに回避するかについても検討を進めている[1]。

本稿では、米国が 2023 年に発表した人工知能リスクマネジメントフレームワーク（Artificial Intelligence Risk Management Framework、以下「AI RMF 1.0」又は「本フレームワーク」）について紹介する。

一、AI RMF 1.0 の概要

米国が 2020 年に発表した国家人工知能イニシアチブ法（National Artificial Intelligence Initiative Act）に対応するため、AI をめぐる一連の対応政策が打ち出された。AI に伴う様々なリスクとその固有の複雑さを考慮すると、既存のリスクマネジメントシステムだけでは AI が直面する懸念に十分に対応することは困難であった。このため、米国国立標準技術研究所（National Institute of Standards and Technology、NIST）は、民間組織及び政府機関との多岐にわたる議論を経て、AI RMF 1.0 を発表した。これは法規範ではなく、法的拘束力も持たない。その目的は、安全でトラストワージな（信頼できる）AI システムやサービスの構築を目指し、AI が関連しうるリスクに対して対応可能なリスクマネジメントシステムを確立し、AI の開発者、設計者及び利用者に参照・依拠される指針を提供することにある。また、科学技術の急速

[1] 弊所 2024 年 8 月のニュースレターでは、AI の開発に対応するため台湾及び EU が制定した法規範を紹介した。詳細は弊所のウェブサイトを参照。

な変化に対応するため、NIST は本フレームワークの内容を随時見直し、調整していくとしている[2]。

二、AI RMF 1.0 の運用

AI の強大な力は、世界に与える影響も増大させており、ポジティブインパクトを最大化し、ネガティブインパクトを最小化するために、より信頼性の高い AI システムを構築して、リスクマネジメントシステムを整備することが不可欠である。この目的の達成にむけて、本フレームワークは二つのパートで構成されている。パート 1 では基礎情報としてリスクマネジメントの基本概念と優れた枠組みを説明し、パート 2 ではコアとプロファイルに関するリスクマネジメントの実施指針を解説している。以下にその概要を紹介する[3]。

（一）パート 1：基礎情報（Foundational Information）

本パートは、「**リスクの枠組み**（Framing Risk）」、「**オーディエンス**（Audience）」、「**人工知能のリスクとトラストワージネス**（AI Risks and Trustworthiness）」、「**AI RMF の有効性**（Effectiveness of the AI RMF）」というトピックで構成されている。

本パートの趣旨は、完備された AI マネジメントシステムを構築するには、まず AI が直面するリスク及びオーディエンスを認識する必要があるという点にある。AI がもたらすネガティブインパクトは、個人のプライバシー権から、マイノリティへの差別、ひいては社会構造全体への影響にまで及び、関連規範への違反につながる可能性もある。次に、潜在的リスクを明確にした後に、より効果的なリソース配分、組織運営と管理効率の向上を図るため、リスク評価と優先順位付けを行い、同時に AI のリスクを他のサイバーセキュリティや個人情報保護などのリスクと統合する必要がある。「**AI のリスクとトラストワージネス**」では、信頼できる AI システムが備えるべき特性として、例えば、安全性、アカウンタビリティ、有害なバイアスをマネジメントした公平性などを概説している。また、NIST が本フレームワークに寄せる高度な適応性と継続的な進化への期待に応えるため、利用者が本フレームワークによって AI リスクを管理する能力が実際に効果的に向上したか否かを定期的に評価することを奨励している。

---

[2] 2025 年 5 月現在までには改訂はない。詳細は下記サイトをご参照。（最終閲覧日：2025 年 5 月 21 日）https://www.nist.gov/itl/ai-risk-management-framework/ai-rmf-development

[3] National Institute of Standards and Technology, Artificial Intelligence Risk Management Framework (AI RMF 1.0), NIST AI 100-1 (2023), https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-1（最終閲覧日：2025/05/21）

（二）パート 2：コアとプロファイル（Core and Profiles）

本フレームワークはパート 2 において、この理念を実現するため、リスクマネジメントシステムの中核に、GOVERN（統治）、MAP（マッピング）、MEASURE（測定）、及び MANAGE（管理）の四つの機能を包含する必要があるとさらに言及している。これらの四つの機能は相互に連携し、リスクマネジメントプロセスの全段階をカバーする[4]。

- GOVERN（統治）：
  リスクマネジメントシステムに関する方針を策定し、組織を構築し、プロセスを確立することで、技術と組織の価値観を整合させることを指す。
- MAP（マッピング）：
  AI の潜在的リスクを評価し、分野横断的な AI アクターと協働することを指す。
- MEASURE（測定）：
  システムのトラストワージネス（信頼性）とリスクを評価し、意思決定の的確性を確保することを指す。
- MANAGE（管理）：
  既存のリスクへの対応計画を策定することにより、リスク事象を効果的に対処し、フィードバック機能を通じてシステムの改善と最適化を行うことを指す。

三、結論

人工知能の発展と向上は、国家による推進及び企業による効果的な活用にかかっている。台湾も世界で確固たる足場を築くため、AI の発展を推進する多くの政策を打ち出している[5]。発展と同時に、その背後にある懸念や関連する法規範も重視されるべきであり、優れたリスクマネジメントシステムが構築されて初めて、AI は信頼に足りるツールとなる。将来、台湾の政府や産業界が AI を運用する際には、リスクマネジメントとシステムの適法性にも注意を払うべきであろう。

[4] 本フレームワークに加え、NIST は補足として「AI RMF プレイブック（AI RMF PLAYBOOK）」も発行している。https://airc.nist.gov/docs/AI_RMF_Playbook.pdf を参照（最終閲覧日：2025/05/21）。

[5] 例えば、台湾は 2018 年に「台湾 AI 行動計画」を打ち出し、2023 年には「人工知能基本法（草案）を公告、2025 年には「産業イノベーション条例」を改正し、各産業における AI 製品及び AI サービスの導入を奨励し、産業構造の最適化を図っている。