参考資料１

米国におけるAI政策最新動向調査
―NISTの取組を中心にして―

2024年5月

独立行政法人情報処理推進機構（IPA）ニューヨーク事務所*

*本レポートはNomura Research Institute America, Inc.に委託して作成した

National Institute of Standards and Technology（NIST）の概要

NISTは科学技術分野における計測や標準、技術評価ツールの提供等を通じて、米国内の成長産業等の発展を技術面からサポートする組織

ミッション

- 経済安全保障を強化し、生活の質を高めるよう計測科学、標準、技術を改善することで、米国の技術革新及び産業競争力を強化する

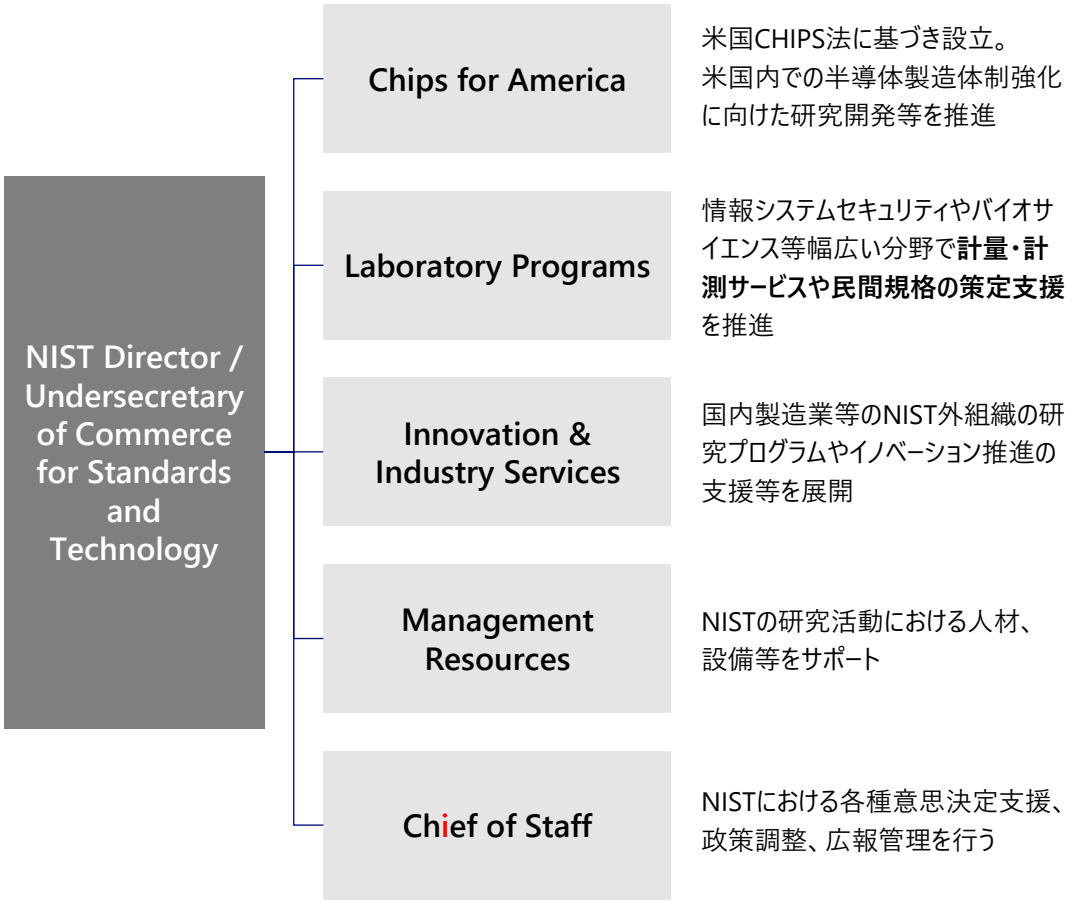
概要

- NIST（米国立標準技術研究所）は、米商務省傘下の連邦研究機関であり、告示などが法の強制力を持たない**非規制機関**
- ナノテクノロジーや量子情報科学、国土安全保障、情報技術、先進製造業といった進歩の著しい産業分野において、米国における**計測システム**の改善、**新たなテクノロジーの開発**、**標準の促進**、企業及び組織が高い品質の製品を作るために必要な**技術評価ツール**の提供などを通じて、各分野の成長・発展を技術面からサポート

コア機能

- 計測科学
- 厳格なトレーサビリティ
- 標準の開発と使用

NISTの組織構成

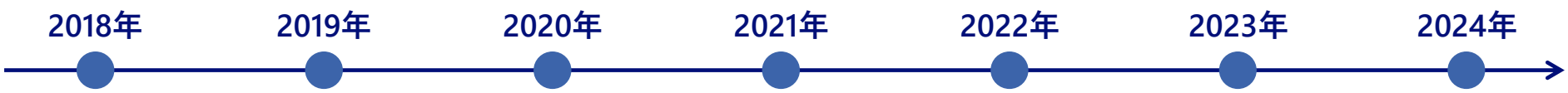


Source ）NISTウェブサイト及びJETRO/IPA New York「NISTの標準策定プロセス」（2019年 1 月）を基に作成
https://www.nist.gov/about-nist
https://www.nist.gov/director/nist-organization-structure
https://www.jetro.go.jp/ext_images/_Reports/02/2019/339d3d579a99af87/nyrp201901sp.pdf

NISTのAIに係るこれまでの主な取組

これまでの米国のAI関連政策の中で、NISTは主に、AIに関係する技術標準化やリスク管理に係る取組を推進してきた

NISTによるAI関連の主な取組



2018年

- **AI技術の基礎・応用研究および標準化プログラム**（FARSAIT）を開始。AIシステムのセキュリティ、説明可能性、透明性の測定と向上を通じてAIシステムへの信頼強化に向けた取組を推進

2019年

- 大統領令（EO13859）「人工知能における米国リーダーシップの維持」に基づき、「**技術標準および関連ツールの開発における連邦政府の関与計画**」を**策定**し、AI技術標準と関連ツールの開発に関する現況、計画、課題、機会、および連邦政府による関与の優先分野を特定

2020年

- 大統領令（EO13960）「連邦政府における信頼できるAIの利用促進」での要請事項を推進（AI使用事例の整理・報告）等

2021年

- 2020年国家人工知能イニチアチブ法に基づき、**AIリスク管理フレームワーク（AI RMF）の検討を開始**し、コンセプトペーパーを発表

2022年

- AI RMFの第一ドラフト、第二ドラフトの発表

2023年

- **AI RMF 1.0を策定**し、AI製品、サービス、システムの設計、開発、使用、評価に信頼性への配慮を組み込む能力を向上させ、自主的に使用することを促進
- **大統領令（EO14110）「AIの安心、安全で信頼できる開発と利用」**に基づく取組を開始

2024年

- **米国AIセーフティ・インスティテュート（USAISI）が発足**
- **AIセーフティ・インスティテュート・コンソーシアム（AISIC）**を設置し、AI開発者と利用者、学術界、政府・産業界の研究者、市民社会組織との連携を促進。AISICに5つのワーキンググループを設置

Source) NISTウェブサイト等に基づき作成
https://www.nist.gov/artificial-intelligence/ai-research
https://www.nist.gov/news-events/news/2019/08/plan-outlines-priorities-federal-agency-engagement-ai-standards-development
https://www.nist.gov/artificial-intelligence/EO13960
https://www.nist.gov/news-events/news/2021/12/nist-seeks-comments-concept-paper-ai-risk-management-framework
https://www.nist.gov/itl/ai-risk-management-framework
https://www.commerce.gov/news/press-releases/2023/11/direction-president-biden-department-commerce-establish-us-artificial

NISTにおけるAI関連の研究概要

また、NISTはこれまで、Fundamental AIとApplied AIに関する研究を通じて、AIの信頼性向上や特定分野におけるAIの応用に関する技術要件や検証ツール等を開発してきた

NISTによるAI関連の研究の概要

Fundamental AI Research

- 信頼できる責任あるAIのための技術要件の確立（Trustworthy and Responsible AI）と新しいAIチップのための新しい測定法、技術的アプローチの確立（Hardware for AI）を目指した研究を実施。
- 特に前者については、AIリスク測定のためにAIの設計者、開発者、評価者による適切な行動を支援するための**ツールとガイダンスの策定**を目指し、「**Bias**」、「**Explainability（説明可能性）**」、「**Security**」**の3つの観点**で研究を実施。

Applied AI Research

- NISTで実施されている各種研究テーマにAI技術を統合し、高度化していくための研究を推進。
- 研究テーマには、コンピュータビジョン、工学生物学と生物製造、画像と映像の理解、医療画像、材料科学、製造、災害回復力、エネルギー効率、自然言語処理、量子科学、ロボット工学、高度通信技術等が含まれる。

Bias

- AIシステムの評価にコンテキストを導入する手法を強化し、悪影響や有害性の理解を深めるため、例えば、以下のような取組を実施。
  - 2022年3月：「人工知能におけるバイアスの特定と管理のための標準に向けて」（NIST Special Publication 1270）では、AIバイアスを分類し、それらが及ぼし得る悪影響を整理。またバイアスを軽減するための**予備的なガイダンスを整理**。
  - 2022年11月：NIST内のNational Cybersecurity Center of Excellence (NCCoE)が主導し、クレジット審査領域でのAIサービスの利用者に利益をもたらすための推奨される**ガイダンスと実践を開発**（「Mitigation of AI/ML Bias in Context」）。

Explainability

- 説明可能なAIの中核的な考え方を探求し、最終的にはAIシステムにおける説明可能性を評価するための**ガイドを開発**。これまでに以下のようなレポートが発効されている。
  - 2021年9月：「説明可能な人工知能の4原則 (NISTIR 8312)」では、コンピュータサイエンス、工学、心理学等の多角的な観点から「説明可能なAIシステム」を整理。
  - 2021年4月：「人工知能における説明可能性と解釈可能性の心理学的基礎 (NISTIR 8367) 」では、AIシステムをより有用でアクセスしやすいものにするために解釈性と説明性を理解し区別することの重要性を指摘。

Security

- NIST内のNational Cybersecurity Center of Excellence (NCCoE)等が中心となりAIのセキュリティ対策について研究開発を実施。
  - 2023年3月にNCCoEは、「敵対的機械学習（Adversarial Machine Learning）」に関する報告書*の草案を通じて、**主要なML手法の種類と攻撃のライフサイクル段階、攻撃者の目標と目的、攻撃者の能力と学習プロセスを整理**。
  - その後、機械学習環境に対する想定される攻撃手法とその防御の有効性を**実証するためのテストベッドとして、「Dioptra」を開発**。

*2024年1月に最終版を公表（NIST AI 100-2 E2023）

Source) NISTウェブサイト等に基づき作成
https://www.nist.gov/fundamental-ai
https://www.nist.gov/applied-ai

大統領令（第14110号）「Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence」の骨子と原則

米国政府は、「人工知能の安全・安心・信頼できる開発と利用に関する大統領令」を発出し、各政府機関にAI技術に関するリスクを把握しつつAIによる利益を獲得するための施策を要求

「人工知能の安全・安心・信頼できる開発と利用に関する大統領令」の構成と8原則（8原則はSection. 2で言及されている）

大統領令(2023年10月30日）の構成（セクション）	8原則の概要（Sec.4から11それぞれに紐づいている）
Sec. 1 Purpose	
Sec. 2 Policy and Principles	
Sec. 3 Definitions	
Sec. 4 Ensuring the Safety and Security of AI Technology AI技術の安全性とセキュリティの確保	AIの安全性確保には、システム評価の標準化、リスク緩和策、生物技術やサイバーセキュリティへの対応が必要。加えて、合成コンテンツの出所を明確にすることで、適切なリスク管理を図ることも重要。
Sec. 5 Promoting Innovation and Competition イノベーションと競争の促進	米国がAIでリードするためには、責任あるイノベーション、競争、協力が必要。AI教育、研究投資、知的財産問題の解決等も通じて、AI時代のスキル獲得を支援し、世界の才能を米国に惹き付ける。
Sec. 6 Supporting Workers 労働者支援	AIの責任ある開発と利用には米国労働者の支援が不可欠。新しい職業や産業を創出するAIは、労働者の権利保護や労働環境の向上に留意し、全労働者が利益を享受できるよう集団交渉を含め支援。
Sec. 7 Advancing Equity and Civil Rights 公平性と公民権の推進	AI政策は公平性と市民権を進展させる政策と一致する必要がある。AIが差別を助長する使用を容認せず、差別を防ぐ基準で厳格な評価と規制を推進。
Sec. 8 Protecting Consumers, Patients, Passengers, and Students 消費者、患者、乗客、学生の保護	AI利活用における消費者保護は重要。AIの誤用や不正使用から生じるプライバシー侵害や差別などの害を防ぐため、連邦政府は既存の法律を施行し、適切な安全策を講じる。
Sec. 9 Protecting Privacy プライバシーの保護	AIの進展に伴い、米国人のプライバシーと自由を保護することが重要。AIは個人データの悪用リスクを高めるため、政府はデータの収集、使用、保持を法的に安全に行い、プライバシーを守る技術を活用する。
Sec. 10 Advancing Federal Government Use of AI 連邦政府によるAI活用の推進	連邦政府自身のAI利用のリスク管理と規制能力向上が重要。AIの専門家を公務に引きつけ、育成し、AIの適切な利用と統治を助けるための措置を実施。
Sec. 11 Strengthening American Leadership Abroad 海外における米国のリーダーシップの強化	連邦政府は社会、経済、技術の進歩をリードすべき。責任あるAIの展開と、国際的な協力を通じてAIのリスクを管理し、利益を最大化するための枠組みを促進する。
Sec. 12 Implementation	
Sec. 13 General Provisions	

米政府は、AIが安全であることを証明できれば、より多くの人々がAIを使うようになり、米国のAI企業に利益をもたらすと考えている。このため、世界に先駆けてAIの安全を定義し、他国・他企業がそれを採用することを望んでいる。

米AIサービス企業 CTO

Source) 大統領令第14110号及びインタビュー結果に基づき作成
https://www.federalregister.gov/documents/2023/11/01/2023-24283/safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence

大統領令（第14110号）によるNISTへの要求事項

主要セクション毎に検討項目が掲げられており、このうちNISTによる取組が要求されているのはAI安全性、プライバシー、連邦政府のAI活用、米国リーダーシップ強化

「人工知能の安全・安心・信頼できる開発と利用に関する大統領令」の8原則に紐づく主要セクションとNISTへの要求事項との関係

主要セクション	項目（青太字はNISTの役割が明記されている項目）
Sec. 4 Ensuring the Safety and Security of AI Technology AI技術の安全性とセキュリティの確保	**4.1. AIの安全性とセキュリティに関するガイドライン、基準、ベストプラクティスの開発** 4.2. 安全で信頼できるAIの確保 4.3. 重要インフラとサイバーセキュリティにおけるAIの管理 **4.4. AIとCBRN（化学、生物、放射線、核）脅威におけるリスクの軽減** **4.5. 合成コンテンツがもたらすリスクの軽減** 4.6. 広く利用可能なモデルウェイトを用いたデュアルユース基盤モデルに関する意見の募集 4.7. AI訓練のための連邦データの安全な公開と悪意のある利用の防止 4.8. 国家安全保障に関する覚書の作成
Sec. 5 Promoting Innovation and Competition イノベーションと競争の促進	5.1. 米国へのAI人材の誘致 5.2. イノベーションの促進 5.3. 競争の促進
Sec. 6 Supporting Workers 労働者支援	労働者支援
Sec. 7 Advancing Equity and Civil Rights 公平性と公民権の推進	7.1. 刑事司法制度におけるAIと公民権の強化 7.2. 政府の給付およびプログラムに関する公民権の保護 7.3. 広範な経済におけるAIと公民権の強化
Sec. 8 Protecting Consumers, Patients, Passengers, and Students 消費者、患者、乗客、学生の保護	消費者、患者、乗客、学生の保護
Sec. 9 Protecting Privacy プライバシーの保護	**プライバシーの保護**
Sec. 10 Advancing Federal Government Use of AI 連邦政府によるAI活用の推進	**10.1. AIマネジメントのガイダンス** 10.2. 政府におけるAI人材の増強
Sec. 11 Strengthening American Leadership Abroad 海外における米国のリーダーシップ強化	**海外における米国のリーダーシップ強化**

Source) 大統領令第14110号に基づき作成
https://www.federalregister.gov/documents/2023/11/01/2023-24283/safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence

大統領令（第14110号）によるNISTへの要求事項

NISTは、AIの安全性・信頼性やプライバシーリスク、連邦政府によるAI活用、AI RMFの国際標準化に係る取組に関わることが求められている(1/2)

セクション毎の要求概要とNISTが関わる取組内容

NISTが関わる項目	NISTが要求されている取組内容
Sec. 4 AI技術の安全性とセキュリティの確保	
4.1. AIの安全性とセキュリティに関するガイドライン、基準、ベストプラクティスの開発	NISTが主導 • 安全、安心、信頼できるAIシステムを開発、展開するための**ガイドラインとベストプラクティス**の策定（**4.1.(a)(i)**） a) AI RMFの付属リソースとして、**生成AI用のリソースを開発** b) 生成AIおよびデュアルユース基盤モデルの安全な開発プラクティスを組み込んだ、**セキュアソフトウェア開発フレームワークの付属リソースの開発** c) サイバーセキュリティやバイオセキュリティの分野に焦点を当て、**AI能力を評価・監査するためのガイダンスとベンチマークを作成するイニシアティブの立ち上げ** • 特にデュアルユース基盤モデルの開発者が、**AIのレッドチームテスト**を実施できるよう、以下を含む形で**適切な手順とプロセスを含むガイドライン**（国家安全保障システムの構成要素として使用されるAIを除く）**を確立**（**4.1.(a)(ii)**） a) デュアルユース基盤モデルの安全性、セキュリティ、信頼性の評価と管理に関連する**ガイドラインの調整または策定** **b) テストベッドなどのテスト環境を開発**し、その利用可能性を確保することを支援
4.4. AIとCBRN（化学、生物、放射線、核）脅威におけるリスクの軽減	NISTが主導 • 合成核酸配列プロバイダーが使用する可能性のある以下のものを開発・改良（**4.4.(b)(ii)**） a) 効果的な**合成核酸調達スクリーニングのための仕様** b) 上記スクリーニングを支援するための懸念配列データベースを管理するための、**セキュリティおよびアクセス制御を含むベストプラクティス** c) 効果的なスクリーニングのための**技術的実施ガイド** d) 適合性評価の**ベストプラクティスおよびメカニズム**
4.5. 合成コンテンツがもたらすリスクの軽減	行政予算管理局長官が主導（同長官はNIST等と協議しながら取組） • 米国政府の公式デジタルコンテンツの完全性に対する国民の信頼を強化する目的で、各省庁が作成または公表する**当該コンテンツのラベル付けおよび認証に関するガイダンス**を発行（**4.5.(c)**）

Source）大統領令第14110号に基づき作成
https://www.federalregister.gov/documents/2023/11/01/2023-24283/safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence

大統領令（第14110号）によるNISTへの要求事項

NISTは、AIの安全性・信頼性やプライバシーリスク、連邦政府によるAI活用、AI RMFの国際標準化に係る取組に関わることが求められている(2/2)

セクション毎の要求概要とNISTが関わる取組内容

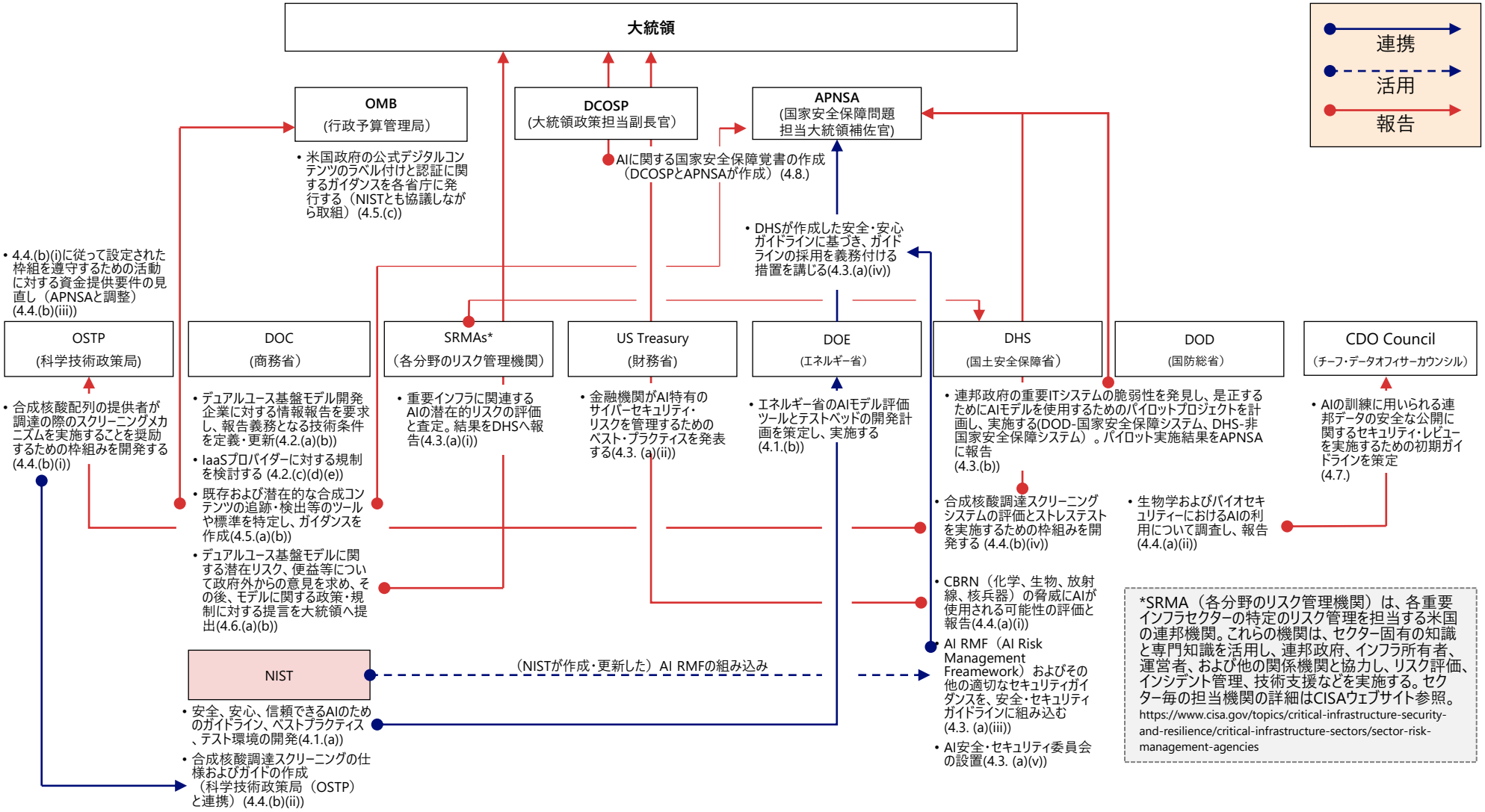
NISTが関わる項目	NISTが要求されている取組内容
Sec. 9 プライバシーの保護	
プライバシーの保護 ▶	NISTが主導 • 米国人のプライバシーを保護するために、各省庁がPETs（privacy-enhancing technologies：プライバシー強化技術）を使用することをより可能にするため、AIを含め、**各省庁が差分プライバシー保護の有効性を評価するためのガイドラインを策定**（**9.(b)**）
Sec. 10 連邦政府によるAI活用の推進	
10.1. AIマネジメントのガイダンス ▶	NISTが主導 • 人々の権利や安全に影響を与える政府のAI利用に関する**最低限のリスク管理プラクティス**（以下の内容を含む）**の実施を支援するためのガイドライン、ツール、及びプラクティスを策定**（**10.1.(d)(i)**） ○ パブリック・コンサルテーションの実施 ○ データ品質の評価 ○ 差別的効果とアルゴリズムによる差別の評価と緩和 ○ AIの使用に関する通知の提供 ○ 導入されたAIの継続的な監視・評価 ○ AIを使用して行われた不利な決定に対する人間による検討や救済措置の付与
Sec. 11 海外における米国のリーダーシップ強化	
海外における米国のリーダーシップ強化 ▶	国務長官及び国際開発庁長官が主導（NISTは、国務長官及び国際開発庁長官と連携） • AIリスクマネジメントフレームワークの原則、ガイドライン及びベストプラクティスを、社会、技術、経済、ガバナンス、人権、安全保障の状況を米国の国境を越えたコンテクストに組み込んだ「**グローバル開発におけるAIプレイブック**」を公表（**11.(c)(i)**）

Source）大統領令第14110号に基づき作成
https://www.federalregister.gov/documents/2023/11/01/2023-24283/safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence

大統領令（第14110号）によるNISTへの要求事項

「Sec.4 AI技術の安全性とセキュリティの確保」でのNIST及び各機関の関係性は以下の通り

大統領令第14110号に基づく、「Sec.4 AI技術の安全性とセキュリティの確保」におけるNISTの役割と関係機関との関係性の概観



Source) 大統領令第14110号に基づき作成
https://www.federalregister.gov/documents/2023/11/01/2023-24283/safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence

大統領令（第14110号）によるNISTへの要求事項

Sec.9, 10, 11において、NISTが関わる取組事項は以下の通り

大統領令第14110号に基づく、Sec.9, 10, 11におけるNISTの役割とNISTの取組に関係する他の機関との関係性の概観

Sec. 9 プライバシーの保護

- AIを含め、各省庁が差分プライバシー保護の有効性を評価するためのガイドラインを策定((9.(b))

その他主な関係機関の役割

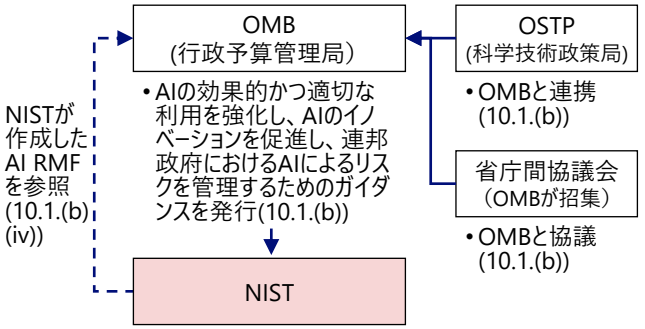
<u>行政予算管理局（OMB）：</u>

- 個人を特定できる情報を含むCAI（商業利用可能な情報）の取扱いに関する基準と手順を評価し、政府機関の活動によるプライバシーリスクを軽減する方法について政府機関にガイダンス案を提示(9.(a)(i)(ii))
- 2002年電子政府法のプライバシー規定の実施に関する政府機関向けガイダンスの改訂見込みを通知するRFI（情報提供依頼）(9.(a)(iii))
- RFIプロセスを通じて特定された必要・適切な措置の実施（新規または更新されたガイダンスの発行、RFI発行等）(9.(a)(iv))

<u>米国国立科学財団（NSF）：</u>

- DOEと協力し、PETs（privacy-enhancing technologies：プライバシー強化技術）の開発、展開、拡大を推進する研究調整ネットワーク（RCN）の創設に資金を提供(9.(c)(i))
- RCNを通じた研究活動等を通じて、政府機関が利用できる最先端のPETsソリューションの採用を奨励する研究を優先(9.(c)(ii))
- 米国・英国PETs Prize Challengeの結果を利用して、PETsの研究と採用のためのアプローチと特定された機会を案内(9.(c)(iii))

Sec. 10 連邦政府によるAI活用の推進



その他主な関係機関の役割（一部を抜粋）

<u>行政予算管理局（OMB）：</u>

- 各省庁のAI開発・利用の調整のための省庁間協議会招集(10.1.(a))
- 各省庁による業務でのAI導入やリスク管理に資する能力を追跡・評価するための方法を開発(10.1.(c))
- 各省庁のAIシステム・サービス取得契約がOMBのガイダンスと整合すること等を確保するための初期手法の開発(10.1.(d)(ii))
- 各省庁に対する毎年のAI利用状況の報告・公表指示(10.1.(e))

<u>人事管理局（OPM）：</u>

- 連邦職員による業務への生成AIの使用に関するガイダンスを作成(10.1.(f)(iii))
- OMBと連携し、雇用・職場の柔軟性、給与柔軟性・インセンティブ給与等を含む、AI人材の採用慣行改善に取り組む(10.2.(d))

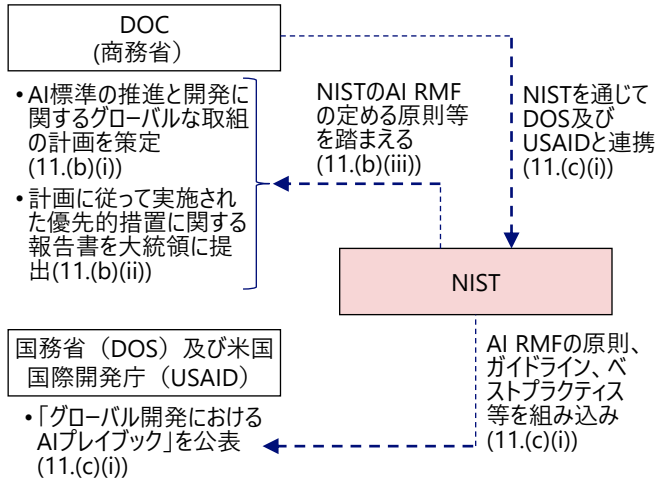
<u>共通役務庁（GSA）：</u>

- OMBと連携し、連邦政府全体での生成AIを始めとする特定タイプのAIサービス・製品の調達促進のための措置を講じる(10.1.(h))

<u>科学技術政策局（OSTP）及び行政予算管理局（OMB）：</u>

- 連邦政府におけるAI人材の増員に向け、優先ミッション分野、最も優先的に採用・育成すべき人材要件、採用経路を特定(10.2.(a))

Sec. 11 海外における米国のリーダーシップ強化



その他主な関係機関の役割

<u>国務省（DOS）：</u>

- AI関連のガイダンスや政策に対する同盟国やパートナーの理解促進と国際的な協力強化のための軍事・情報分野以外の取組主導(11.(a)(i))
- AIのリスクを管理し、利益を活用するための強力な国際的枠組みを確立するための取組の主導(11.(a)(ii))

<u>国土安全保障省（DHS）：</u>

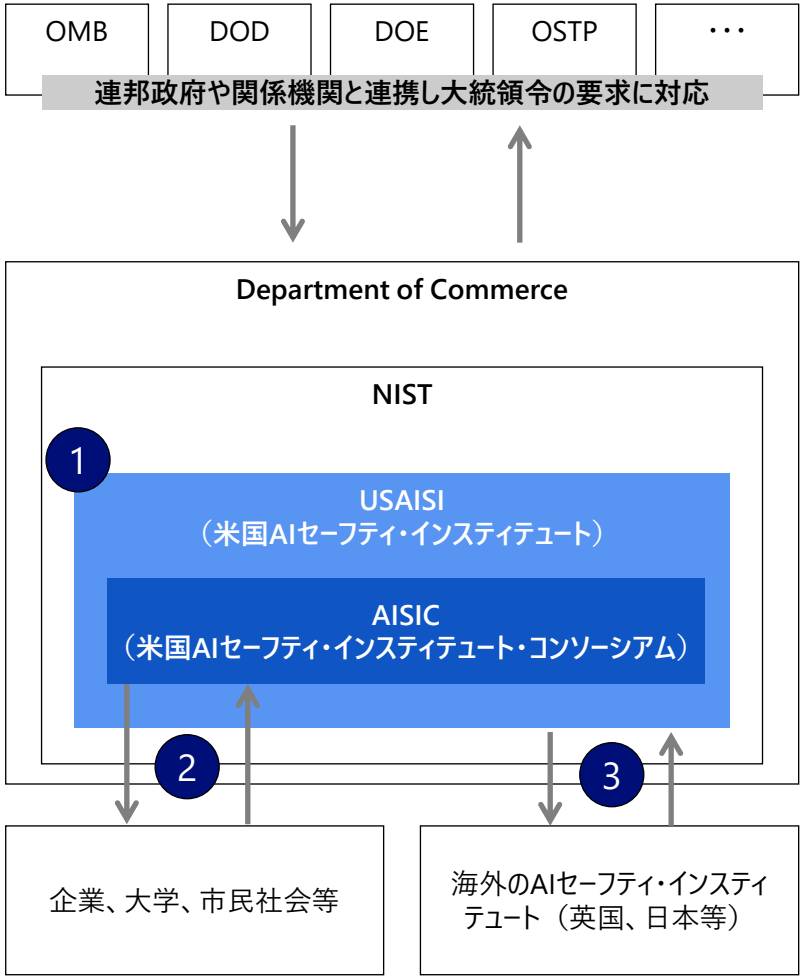
- 重要インフラシステムへのAI組み込みやAIの悪用から生じる潜在的な重要インフラの混乱を防止、対応、回復するための協力を強化する、国際的な同盟国やパートナーとの取組を主導(11.(d))
- 重要インフラの所有者及び運営者が使用するためのAIの安全・セキュリティガイドラインの採用を奨励するための多国間関与計画を策定(11.(d)(i))
- 米国の重要インフラに対する国境を越えたリスクを軽減するための優先行動に関する報告書を大統領に提出(11.(d)(ii))

Source) 大統領令第14110号に基づき作成
https://www.federalregister.gov/documents/2023/11/01/2023-24283/safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence

NIST傘下の米国AIセーフティ・インスティテュート（USAISI）と米国AIセーフティ・インスティテュート・コンソーシアム（AISIC）

商務省は、NISTに設立した米国AIセーフティ・インスティテュート(USAISI)と傘下の米国AIセーフティ・インスティテュート・コンソーシアム(AISIC)の取組を通じて大統領令の要求事項を推進

米国AIセーフティ・インスティテュート（USAISI）の位置づけと同機関の主な役割



2023年11月1日、米商務省は、特に最先端のAIモデルの評価において、AIの安全性と信頼性に関する米国政府の取組を主導する組織として、米国AIセーフティ・インスティテュート(USAISI)をNIST内に設置。

USAISI（米国AIセーフティ・インスティテュート）の主な取組内容

1. USAISIは、大統領令で商務省に割り当てられた責務を推進。
具体的には、AIモデルの安全性、セキュリティ、テストに関する標準の開発促進、AIが生成したコンテンツを認証するための標準開発、AI研究者や開発者が新たなAIのリスクを評価し、既知の影響に対処するためのテストベッド提供等を実施。

2. USAISIは、米国USAISIコンソーシアムを通じた学界、産業界、政府、市民社会のパートナーとの協力など、外部の専門知識を活用して責務を推進。
コンソーシアムには5つのワーキンググループが設置され、それぞれのワーキンググループの課題に関して、ガイドラインやベンチマーク、テスト環境の開発などに取り組む（ワーキンググループについては後述）。

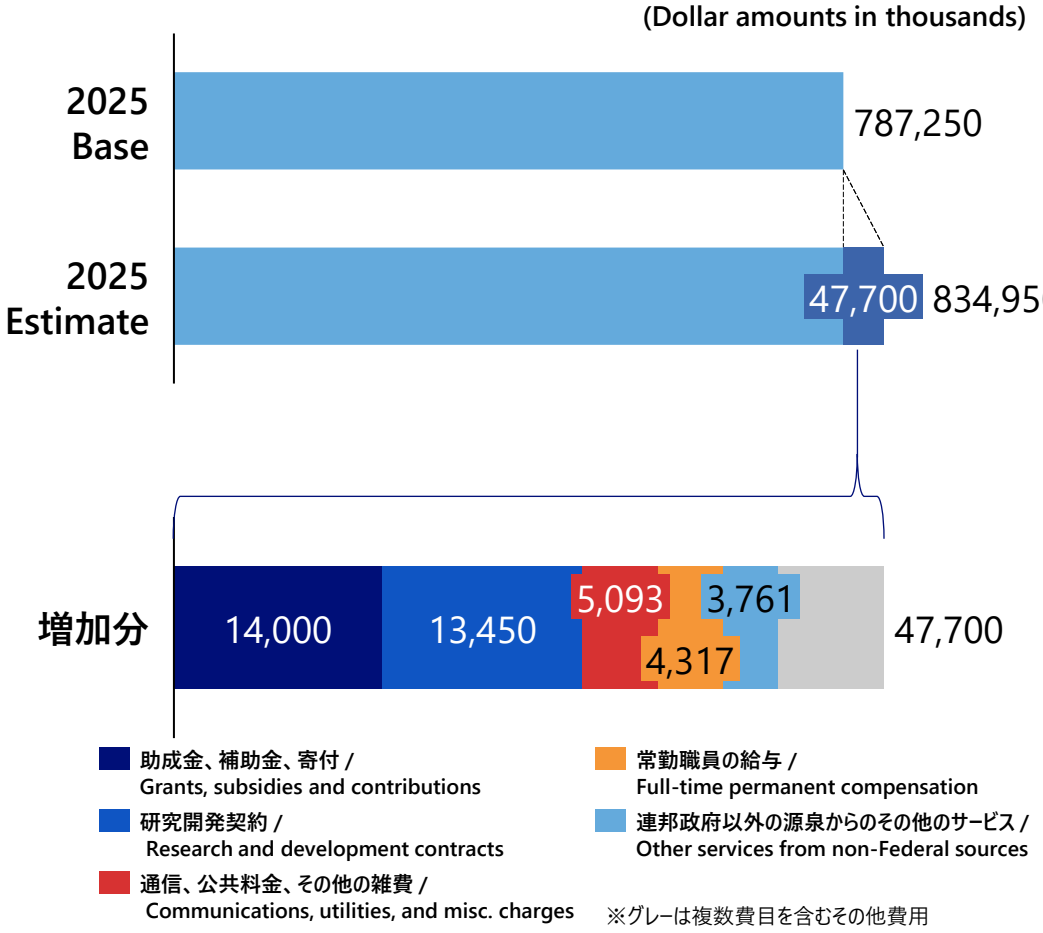
3. さらに、英国のAIセーフティ・インスティテュートを含む、同盟国やパートナー国の同様の研究所と協力し、この分野における作業の連携と調整を図る。

Source) NISTウェブサイト等に基づき作成
https://www.nist.gov/aisi/strategic-vision
https://www.nist.gov/artificial-intelligence-safety-institute/aisic-working-groups

NISTにおける2025会計年度予算（特に大統領令（第14110号）の推進関連）

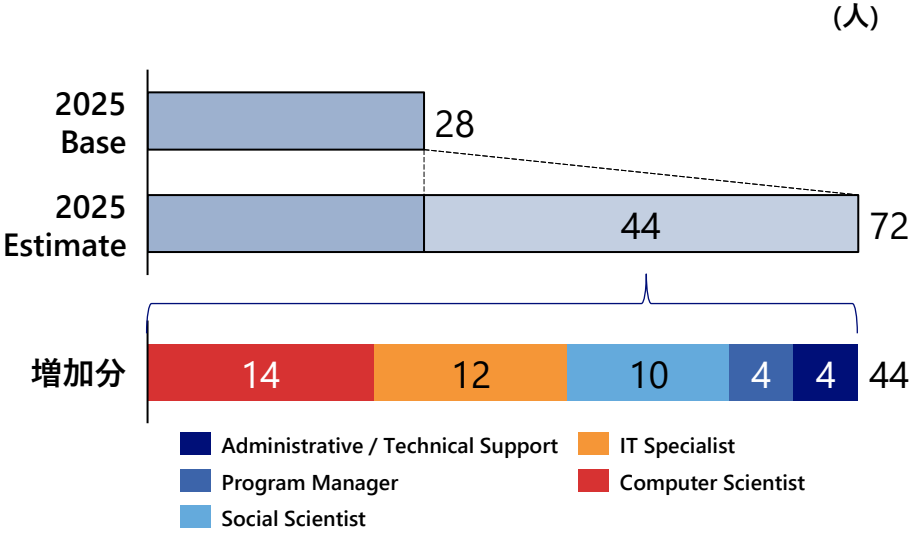
NISTは、2025会計年度における4,770万米ドルの増額要求を通じて、人的リソースを増強しつつ、USAISIのガイドライン等開発の取組や、先端AI評価能力・テスト環境等の強化を推進

NISTの2025会計年度の増額概要と主な費目内訳*



*NIST全体の予算のうち「Advancing Artificial Intelligence Research, Standards, and Testing to Meet National Needs（国家ニーズに応える人工知能の研究、標準化、テストの推進）」に充てられる予算

NISTの2025会計年度の人員増加分と内訳



2025会計年度（2024年10月1日～2025年9月30日）予算の増額について、NISTは予算書中で、大統領令（第14110号）「安全で安心、信頼できる人工知能の開発と使用」を踏まえた増額であることを説明。特に、以下の取組に重点を置いている。

1. 米国AIセーフティ・インスティテュート（USAISI）:

- AIの安全な開発と責任ある使用のための科学を構築し、産・学・市民社会と協力して、AIの危険な能力の評価・軽減のためのガイドライン、ツール、指標を開発・評価

2. AI研究、標準、実装、およびテストの推進:

- AIシステムの安全性・信頼性のより効果的な評価を開発する戦略的研究の実施、先端AIモデルのテストインフラ設立、AIシステムの透明性のための技術ガイダンス開発、AIRMFの実装支援

Source）「NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY NATIONAL TECHNICAL INFORMATION SERVICE FISCAL YEAR 2025 BUDGET SUBMISSION TO CONGRESS」を基に作成
https://www.commerce.gov/sites/default/files/2024-03/NIST-NTIS-FY2025-Congressional-Budget-Submission.pdf

NISTにおける2025会計年度予算（特に大統領令（第14110号）の推進関連）

NISTのAI関連の2025会計年度予算では、助成金・補助金・寄付、研究開発契約、人件費といった項目の予算が大幅に増額している

NISTの2025会計年度（2024年10月1日～2025年9月30日）AI関連予算要求明細

(Dollar amounts in thousands)

	Object Class	2023 Actual	2024 Annualized CR	2025 Base	2025 Estimate	Increase/Decrease from 2025 Base
11.1	Full-time permanent compensation	$282,138	$308,041	$309,838	$314,155	$4,317
11.3	Other than full-time permanent	20,999	22,101	22,633	22,633	0
11.5	Other personnel compensation	8,347	9,299	10,676	10,676	0
11.8	Special personnel services payments	0	0	0	0	0
11.9	Total personnel compensation	311,484	339,441	343,147	347,464	4,317
12.1	Civilian personnel benefits	113,223	122,960	125,248	126,624	1,376
13	Benefits for former personnel	56	56	56	56	0
21	Travel and transportation of persons	8,717	8,826	8,824	9,035	211
22	Transportation of things	460	502	471	530	59
23	Rent, communications, and utilities	0	0	0	0	0
23.1	Rental payments to GSA	159	159	1,680	1,680	0
23.2	Rental payments to others	1,978	1,978	2,022	2,022	0
23.3	Communications, utilities, and misc. charges	23,181	27,189	25,713	30,806	5,093
24	Printing and reproduction	521	574	533	601	68
25	Other contractual services	0	0	0	0	0
25.1	Advisory and assistance services	1,502	1,310	1,368	1,368	0
25.2	Other services from non-Federal sources	37,890	124,486	17,085	20,846	3,761
25.3	Other goods and services from Federal sources	58,800	64,626	61,174	62,256	1,082
25.4	Operation and maintenance of facilities	0	0	0	0	0
25.5	Research and development contracts	49,524	52,524	50,669	64,119	13,450
25.6	Medical care	0	0	0	0	0
25.7	Operation and maintenance of equipment	17,683	18,576	18,105	18,674	569
25.8	Subsistence and support of persons	0	0	0	0	0
26	Supplies and materials	32,754	33,105	33,524	34,462	938
31	Equipment	37,279	37,551	38,240	41,016	2,776
32	Lands and structures	231	231	231	231	0
33	Investments and loans	0	0	0	0	0
41	Grants, subsidies and contributions	59,158	44,158	59,158	73,158	14,000
42	Insurance claims and indemnities	0	0	0	0	0
43	Interest and dividends	2	2	2	2	0
44	Refunds	0	0	0	0	0
99.9	Total obligations	754,602	878,254	787,250	834,950	47,700

Source）「NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY NATIONAL TECHNICAL INFORMATION SERVICE FISCAL YEAR 2025 BUDGET SUBMISSION TO CONGRESS」
https://www.commerce.gov/sites/default/files/2024-03/NIST-NTIS-FY2025-Congressional-Budget-Submission.pdf

米国AISIコンソーシアム(AISIC)に設置されたワーキンググループを通じて、企業や研究機関等から専門的支援を受けながら科学的根拠と実証に裏付けされたガイドライン等を策定する

AISICの取組概要、5つのワーキンググループの概要(大統領令の関連条項)

コンソーシアムの取組概要

- コンソーシアムで、NISTは以下に取り組む。
  - AI関係者のための知識・データ共有スペースの構築
  - 研究計画実施を通じて、協調的・学際的な研究開発に従事
  - 社会や米国経済に対するAIの影響をより完全かつ効果的に理解するような、研究・評価の要件及びアプローチを優先
  - コンソーシアムメンバー間の技術及びデータの共同開発や移転を促進するアプローチの特定・推奨
  - 連邦政府の所管事項に関する連邦政府機関の意見を効率的に取り入れるメカニズムの特定
  - 将来のAIの測定に資するテストシステムとプロトタイプの査定や評価を可能にする
- 持続可能な共同研究・開発のアプローチを創出するため、コンソーシアムの作業はオープンで透明性の高いものとする。
- 信頼できる責任あるAIのための測定科学を構築・成熟させるため、関係者が協働しやすいハブを提供する。

各ワーキンググループの概要

Working Group #1: 生成AIのリスク管理 (Risk Management for Generative AI)	• 生成AI向けのAI RMFの補足リソースの開発 • 連邦政府機関向けの最低限のリスク管理ガイダンスの開発 • AI RMFの実用化
Working Group #2: 合成コンテンツ (Synthetic Content)	• コンテンツの認証・出所追跡に資する、既存の標準や潜在的科学的標準・技術開発の調査 • 合成コンテンツのラベリング・検出、不適切な画像等の生成防止、上記目的で使用されるソフトウェアのテスト、合成コンテンツの監査・管理
Working Group #3: 能力評価 (Capability Evaluations)	• AIの能力の評価と監査のためのガイダンスとベンチマークの作成（特に化学、生物、放射線、核等） • AI技術の開発を支援するためのテスト環境の開発と利用可能性の確保
Working Group #4: レッドチーム (Red-Teaming)	• 特にデュアルユース基盤モデルのAI開発者が、AIレッドチーミングテストを実施するためのガイドラインの開発
Working Group #5: 安全とセキュリティ (Safety & Security)	• デュアルユース基盤モデルの安全性とセキュリティの管理に関連するガイドラインの開発

参考情報：
- NISTは、AISICに参加する組織から専門知識やツールなどのノウハウを獲得するためNDAを締結
- 各ワーキンググループの議論の取りまとめはNISTメンバーの役割
- ワーキンググループには一組織から複数人の参加が可能。ビッグテック企業は数十名単位で参加し評価手法・ツール等の検討・開発に積極的に参画
- Slack（チャットツール）等を用いた情報/意見交換を非同期コミュニケーションで実施

Source ）NISTウェブサイト等及びインタビュー結果に基づき作成
https://www.nist.gov/aisi/artificial-intelligence-safety-institute-consortium-aisic

米国AIセーフティ・インスティテュート・コンソーシアム（AISIC）の概要

コンソーシアムには200を超える組織が参加。参加組織にとっては、AI安全評価に係る標準策定等への関与、将来動向予見、他者との関係づくりなどが参加動機となっている

コンソーシアム参加組織（一部）

業種	組織名
自動車	• Ford Motor Company
半導体	• NVIDIA • AMD
製薬	• Pfizer • Merck & Co., Inc.
金融機関	• Bank of America • Visa
エネルギー	• BP • Pacific Gas and Electric Company (PG&E)
IT	• Google • Microsoft • Open AI
通信	• AT＆T • Verizon Communications
小売	• Amazon • Walmart
エンターテインメント	• Walt Disney • Netflix
大学・教育機関	• Stanford Institute for Human-Centered AI • Massachusetts Institute of Technology (MIT)
法律	• Gibson, Dunn & Crutcher LLP • DLA Piper
非営利団体	• Future of Privacy Forum • National Fair Housing Alliance
州政府	• State of California, Department of Technology • State of Kansas, Office of Information Technology Services

※NISTは今回のメンバーを「inaugural cohort of members（創立メンバー）」としており、今後も検討トピック等に応じて参加企業を募集・拡大するとしている

コンソーシアムに参加した理由

アカデミアにある我々の参加動機は、ポジティブな方向にAIガバナンスを形成したいということ。我々はAIをより安全に活用する新しいツールがあると信じており、その合意形成に貢献したい。
その他、組織の参加理由は様々。NISTやAI分野の他の大手企業とAI分野で関係構築したいグループもあれば、標準策定に貢献したいグループもある。

米大学 准教授

小規模なハイテク企業として、策定されたガイドラインを遵守することはしばしば困難となるため、コンソーシアムへの参加を通じて、このトピックに対する我々の見解を示すことで、ガイドラインをより広く適用可能なものにしていきたい。

米AIサービス企業 CTO

我々は政策および政府問題へのアクセスを拡大し、民間企業等の組織が将来起こる変化や政策動向を予見し、それに対応できるよう支援することを目的とした組織である。コンソーシアムへの参加は、民間企業等からの政策に対するフィードバックを通じてNISTに洞察を与えることに加えて、コンソーシアムで議論される内容を理解し、AI安全評価における将来的な動向を予見することにある。

米政策インテリジェンス企業 CFO

Source) NISTウェブサイト等及びインタビュー結果に基づき作成
https://www.nist.gov/aisi/aisic-members

米国AIセーフティ・インスティテュート・コンソーシアム（AISIC）の概要

コンソーシアムに参加する組織には、以下のような技術的専門性に関する要件や取組への貢献が求められている

コンソーシアムに参加するための技術的専門性要件

1. データおよびデータの文書化 (Data and data documentation)
2. AIメトロロジー (AI Metrology)
3. AIガバナンス (AI Governance)
4. AIセーフティ (AI Safety)
5. 信頼できるAI (Trustworthy AI)
6. 責任あるAI (Responsible AI)
7. AIシステムの設計と開発 (AI system design and development)
8. AIシステムの展開 (AI system deployment)
9. AIレッドチーミング (AI Red Teaming)
10. 人間とAIのチーミングおよびインタラクション (Human-AI Teaming and Interaction)
11. テスト、評価、検証、および確認の方法論 (Test, Evaluation, Validation and Verification methodologies)
12. 社会技術的方法論 (Socio-technical methodologies)
13. AIの公平性 (AI Fairness)
14. AIの説明可能性と解釈可能性 (AI Explainability and Interpretability)
15. 労働力のスキル (Workforce skills)
16. 心理測定学 (Psychometrics)
17. 経済分析 (Economic analysis)
18. NIST AIリスク管理フレームワークを通じて安全で信頼できる人工知能(AI)システムを可能にするためのモデル、データ、および製品の支援とデモンストレーション (Models, data and/or products to support and demonstrate pathways to enable safe and trustworthy artificial intelligence (AI) systems through the NIST AI Risk Management Framework)
19. コンソーシアムプロジェクトのためのインフラストラクチャー支援 (Infrastructure support for consortium projects)
20. 施設のスペースとコンソーシアム研究者、ウェビナー、ワークショップ、カンファレンスおよびオンラインミーティングのホスティング (Facility space and hosting consortium researchers, webinars, workshops and conferences, and online meetings)

コンソーシアムで参加者に期待される貢献内容

1. AIを安全で安心、信頼できる方法で開発または導入するための業界標準の進化を促進する新しいガイドライン、ツール、方法、プロトコル、ベストプラクティスの開発
2. 潜在的な害を及ぼす可能性のあるAI能力に焦点を当てた、AI能力の特定と評価のためのガイダンスとベンチマークの開発
3. デュアルユース基盤モデルへの特別な配慮を含む、生成AIのための安全な開発プラクティスを組み込むアプローチの開発（以下を含む）：
   - モデルの安全性、セキュリティ、信頼性の評価・管理に関連するガイダンスおよびプライバシーを保護する機械学習に関連するガイダンス
   - テスト環境の利用可能性を保証するガイダンス
4. テスト環境の開発と利用可能性の確保
5. 有効なレッドチーミングとプライバシーを保護する機械学習のためのガイダンス、方法、スキル、実践の開発
6. デジタルコンテンツを認証するためのガイダンスとツールの開発
7. AI労働力スキルのためのガイダンスと基準の開発（リスクの特定と管理、テスト、評価、検証、確認（TEVV）、ドメイン特有の専門性を含む）
8. 異なる文脈で人々がAIをどのように理解し関わるかについての科学を含む、社会と技術の交点での複雑さの探求
9. AIアクター間の相互依存性を理解し管理するためのガイダンスの開発

NISTは自身がアクセスできない貴重な情報やリソースにアクセスするため、様々なバックグラウンドを持つ組織の参加を望んでいる。

米AIサービス企業 CTO

Source) NISTウェブサイト等及びインタビュー結果に基づき作成
https://www.nist.gov/aisi/aisic-members
https://www.nist.gov/aisi/artificial-intelligence-safety-institute-consortium-aisic

AISIC傘下のワーキンググループの取組概要

各ワーキンググループでは、以下のような内容を含む議論がなされている（1/2）

ワーキンググループ	取組概要	各ワーキンググループでの議論やワーキンググループ参加者の意見
Working Group #1: 生成AIのリスク管理 (Risk Management for Generative AI)	• 生成AI向けのAI RMFの補足リソースの開発 • 連邦政府機関向けの最低限のリスク管理ガイダンスの開発 • AI RMFの実用化	• WG#1は、昨年11月に実施された公開ワークショップをベースに、AI RMFの中に生成AIの内容を盛り込みアップデートすることを目的としている。その目的には、生成AIを使用する組織に対する一連の管理策の策定や、AIのサプライチェーンに対する文書化基準の確立などが含まれる。（米AIサービス企業CTO）
Working Group #2: 合成コンテンツ (Synthetic Content)	• コンテンツの認証・出所追跡に資する、既存の標準や潜在的科学的標準・技術開発の調査 • 合成コンテンツのラベリング・検出、不適切な画像等の生成防止、上記目的で使用されるソフトウェアのテスト、合成コンテンツの監査・管理	• 大統領令でも焦点を当てられている電子透かしは検知・証明ツールの一種に過ぎない。業界内では、電子透かしに集中しすぎているため十分な対応ができていないという批判もある。このようなことから、**今後ワーキンググループでは、電子透かし以外の方法についても検討範囲が広がる**のではないか。（米政策インテリジェンス企業CFO） • 合成コンテンツは選挙にも大きな影響を与える。**大統領選が近づいておりこのトピックについて政府の関心が非常に高まっている**。（米政策インテリジェンス企業CFO） • 電子透かし技術に関して業界全体で共通の標準がまだ存在しない。いくつかの企業がそれぞれ**独自の電子透かし技術を導入しているものの、一般的な標準が確立されているわけではない**。（米シンクタンクCEO）
Working Group #3: 能力評価 (Capability Evaluations)	• AIの能力の評価と監査のためのガイダンスとベンチマークの作成（特に化学、生物、放射線、核等） • AI技術の開発を支援するためのテスト環境の開発と利用可能性の確保	• WG#3では、生成AIの利用によって生じるリスクを評価するために、**リスクの高い13種類の用途に対するガイダンスを出す必要性**が議論されている。例えば、AIが危険な生物学的または化学的化合物を生成する能力や、人間のイメージや声を模倣する能力が挙げられている。今後議論が進む中で、13種類からさらに数が増える可能性もある。（米AIサービス企業CTO） • 例えば、あるAIシステムが新たな致命的なウイルスを作り出せるかどうかを測定する方法を今は誰も知らない。それを解明することは非常に重要であり、そのようなことを検討している。（米AIサービス企業CTO） • **2024年2月に改訂されたサイバーセキュリティ・フレームワーク（CSF 2.0）では、サイバーセキュリティにおいて新しいガバナンスの観点を組み込んでいる。これらの観点は、AI能力評価について議論する際にも参考にされる**だろう。（米政策インテリジェンス企業CFO） • WG#3では、能力評価のための要求事項だけでなく、その要求事項をどの程度コントロールできているのかという進捗状況を計測する方法も重要にになる。（米政策インテリジェンス企業CFO）

Source) インタビュー結果に基づき作成

AISIC傘下のワーキンググループの取組概要

各ワーキンググループでは、以下のような内容を含む議論がなされている（2/2）

ワーキンググループ	取組概要	各ワーキンググループでの議論やワーキンググループ参加者の意見
Working Group #4: レッドチーム (Red-Teaming)	• 特にデュアルユース基盤モデルのAI開発者が、AIレッドチーミングテストを実施するためのガイドラインの開発	• WG#4では、**第三者である独立監査人がレッドチームを組むことを義務付けるべきか、それとも組織内部でレッドチームを結成すべきか**という議論がある。（米大学・研究機関 准教授） • 生成AIにおけるレッドチーミングの運用に関して、どのような構造が求められるか、どのようなタスクを行うべきかについてのガイダンスが不足している状況。（米AIサービス企業CTO） • **レッドチーミングを内部で行うか外部で行うかについても、そこには潜在的な利益の衝突がある**。NISTは、そもそも**利害の衝突がどのようなものかを明らかにし、なぜ外部のレッドチーミング団体を利用すべきかについての理由を提供したい**と考えているようだ。また、**サイバーセキュリティのペネトレーションテストに類似したアプローチを適用**しようと試みているようだ。（米AIサービス企業CTO） • IBMは社内に強力なレッドチーミング組織を持っている。彼らはNISTのガイドラインを自分たちがすでにやっていることと一致させたいという思惑がある。しかし、NISTとしてはそれを避けたいと思っているだろう。このバランスをどのように調整するのかが論点ではある。（米AIサービス企業CTO） • レッドチーミングは一般的に認知されているプロセスであるものの、**AIモデルのレッドチーミングに対する明確な方法論や標準がまだ確立されていない**。（米シンクタンクCEO）
Working Group #5: 安全とセキュリティ (Safety & Security)	• デュアルユース基盤モデルの安全性とセキュリティの管理に関連するガイドラインの開発	• 現在は作業計画を検討している状況。**提案されている計画には、デュアルユース基盤モデルの安全性とセキュリティの管理に関連するガイドライン、それらの技術的な手法及び手順、モデルの安全対策とリスク軽減措置の有効性・堅牢性の検討、モデルのガバナンスに関するガイドラインを策定することが含まれている**。（米シンクタンクCEO） • NISTは、デュアルユース基盤モデルの定義を、少なくとも数十億のパラメータを含み、幅広い文脈で使用され、セキュリティ、国家経済、公衆衛生、安全への深刻なリスクをもたらすタスクで高いパフォーマンスを発揮するように簡単に修正可能であるモデルとしている。（米シンクタンクCEO） • さらにNISTは、デュアルユース基盤モデルが非専門家による化学、生物、核兵器の設計の入門障壁を下げること、攻撃型サイバーセキュリティ作戦を可能にすること、また、そのようなリスクが自動化され自律的に動作することで人間による管理・監督を必要としなくなる可能性について検討すべきと考えている。（米シンクタンクCEO）

Source) インタビュー結果に基づき作成