JIPDEC

JIPDECセミナー　講演資料
「AI規制について　欧米の動向と日本の状況」

独立行政法人情報処理推進機構
デジタル基盤センター長　AIセーフティ・インスティテュート 副所長・事務局長
平本　健二氏

本資料は、2024年7月22日（月）開催、JIPDECセミナーで配布した資料です。
セミナーお申込み者様限定での配布となりますので、WEB、SNS等への掲載、転載はご遠慮ください。

2023.07.22

AIと規制について

ー　欧米の動向と日本の状況　ー

(AISIは規制担当部局ではないので、本資料・講演は、情報ハブとしての一般的な情報提供)

2024-07-22

AIセーフティ・インスティテュート（AISI）

AISI Japan AI Safety Institute

AISI（AIセーフティ・インスティテュート）とは

AISIは、内閣府を中心に10府省、5政府関連機関が連携する**官民の取組を支援する機関**である。（2024年2月設立。独立行政法人情報処理推進機構（IPA）に事務局）

♦ 役割

- 政府への支援として、AIセーフティに関する調査、評価手法の検討や基準の作成等の支援を行う
- 日本におけるAIセーフティのハブとして、産学における関連取組の最新情報を集約し、関係企業・団体間の連携を促進する。
- さらに、 他国のAIセーフティ関係機関と連携する。

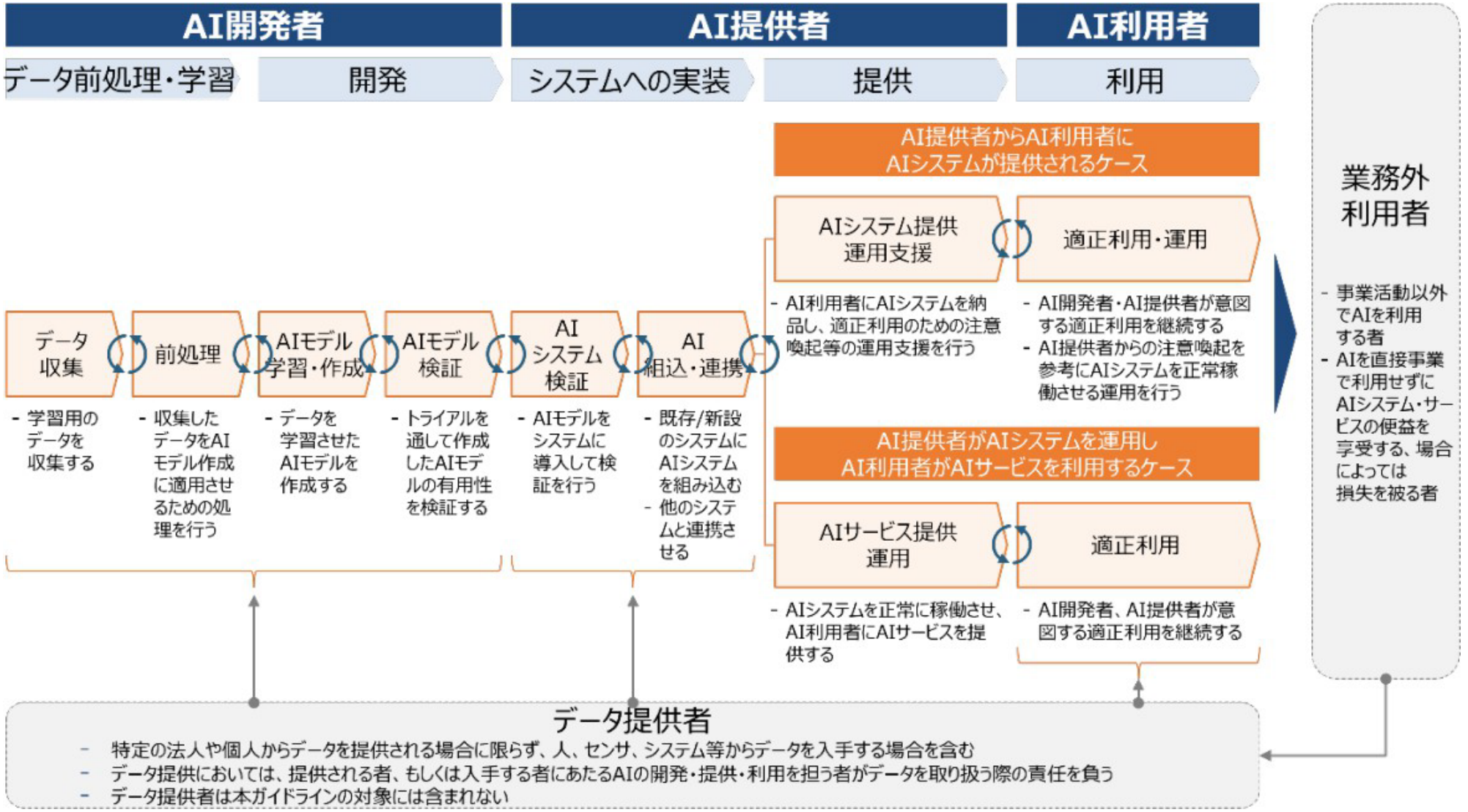
♦ スコープ

- AIによる以下の事象や検討事項の中で、諸外国や国内の動向も見ながら柔軟にスコープを設定し取組を進めていく。
  - 社会への影響、ガバナンス、AIシステム、コンテンツ、データ

AIのリスクとは

- 規制を考える前提

AI開発と活用の流れ


AI開発者
AI提供者
AI利用者
データ前処理・学習
開発
システムへの実装
提供
利用
AI提供者からAI利用者に
AIシステムが提供されるケース
AIシステム提供
運用支援
適正利用・運用
- AI利用者にAIシステムを納品し、適正利用のための注意喚起等の運用支援を行う
- AI開発者・AI提供者が意図する適正利用を継続する
- AI提供者からの注意喚起を参考にAIシステムを正常稼働させる運用を行う
データ収集
前処理
AIモデル学習・作成
AIモデル検証
AIシステム検証
AI組込・連携
- 学習用のデータを収集する
- 収集したデータをAIモデル作成に適用させるための処理を行う
- データを学習させたAIモデルを作成する
- トライアルを通して作成したAIモデルの有用性を検証する
- AIモデルをシステムに導入して検証を行う
- 既存/新設のシステムにAIシステムを組み込む
- 他のシステムと連携させる
AI提供者がAIシステムを運用し
AI利用者がAIサービスを利用するケース
AIサービス提供
運用
適正利用
- AIシステムを正常に稼働させ、AI利用者にAIサービスを提供する
- AI開発者、AI提供者が意図する適正利用を継続する
業務外
利用者
- 事業活動以外でAIを利用する者
- AIを直接事業で利用せずにAIシステム・サービスの便益を享受する、場合によっては損失を被る者
データ提供者
- 特定の法人や個人からデータを提供される場合に限らず、人、センサ、システム等からデータを入手する場合を含む
- データ提供においては、提供される者、もしくは入手する者にあたるAIの開発・提供・利用を担う者がデータを取り扱う際の責任を負う
- データ提供者は本ガイドラインの対象には含まれない

AIに関して想定されるリスク

AISI Japan AI Safety Institute

- AIには、Physical, Social, Economical, Psychologicalなリスクがある。

	共通の指針	主なリスク
1) 人間中心	① 人間の尊厳及び個人の自律 ② AIによる意思決定・感情の操作等への留意 ③ 偽情報等への対策 ④ 多様性・包摂性の確保 ⑤ 利用者支援 ⑥ 持続可能性の確保	・人間の尊厳及び個人の自律を損なうリスク （プロファイリング時の配慮の必要性等） ・AIにより意思決定・感情の操作をされてしまうリスク ・偽情報などのリスク ・多様性や包摂性が確保されないリスク ・地球環境への影響のリスク
2) 安全性	① 人間の生命・身体・財産、精神及び環境への配慮 ② 適正利用 ③ 適正学習	・動作が止まる、低下するリスク ・意図しない動作のリスク ・ステークフォルダがリスクを知らないリスク ・目的外に利用してしまうリスク ・学習データに十分な品質がないリスク ・学習データのコンプライアンスリスク
3) 公平性	① AIモデルの各構成技術に含まれるバイアスへの配慮 ② 人間の判断の介在	・バイアスによる公平性を損なうリスク ・潜在的なバイアスが発生するリスク ・人間の介在が不足するリスク ・バイアスの評価プロセスが不十分なリスク

AIに関して想定されるリスク

AISI Japan AI Safety Institute

	共通の指針	主なリスク
4) プライバシー保護	① AIシステム・サービス全般におけるプライバシーの保護	・プライバシーを侵害するリスク
5) セキュリティ確保	① AIシステム・サービスに影響するセキュリティ対策 ② 最新動向への留意	・不正操作のリスク ・AIシステム自体へのセキュリティ侵害へのリスク ・不正データが使われるリスク
6) 透明性	① 検証可能性の確保 ② 関連するステークホルダーへの情報提供 ③ 合理的かつ誠実な対応 ④ 関連するステークホルダーへの説明可能性・解釈可能性の向上	・検証ができないリスク ・ステークホルダーに十分な情報提供がされないリスク ・合理的でない情報提供を求められるリスク
7) アカウンタビリティ	① トレーサビリティの向上 ②「共通の指針」の対応状況の説明 ③ 責任者の明示 ④ 関係者間の責任の分配 ⑤ ステークホルダーへの具体的な対応 ⑥ 文書化	・トレーサビリティ情報が入手できないリスク ・共通の指針への対応状況が報告されないリスク ・責任が明確にならないリスク ・ステークホルダーと適切なコミュニケーションが取れないリスク ・各種情報をドキュメンテーションできていないリスク

AIに関して想定されるリスク

	共通の指針	主なリスク
8) 教育・リテラシー	① AIリテラシーの確保 ② 教育・リスキリング ③ ステークホルダーへのフォローアップ	・AI利用者が判断能力を持たないリスク ・AIにより雇用が奪われるリスク ・ステークホルダーが技術などの進化に追随できないリスク
9) 公正競争確保		・AIに関して公正な競争が阻害されるリスク
10) イノベーション	① オープンイノベーション等の推進 ② 相互接続性・相互運用性への留意 ③ 適切な情報提供	・AIのイノベーションが阻害されるリスク ・相互運用性が確保されないリスク ・AIに関する情報が十分に伝達されないリスク

RISKの定義とレベル

- RISKの定義
  - ISO 31000
    - effect of uncertainty on objectives
  - ISO 9001
    - effect of uncertainty
- RISKの定義
  - 様々な機関が定義している
    - 右の図は欧州AI法の定義

ISO 31000におけるリスクへの対応

AISI Japan AI Safety Institute

1. Avoiding the risk by deciding not to start or continue with the activity that gives rise to the risk
2. Accepting or increasing the risk in order to pursue an opportunity
3. Removing the risk source
4. Changing the likelihood
5. Changing the consequences
6. Sharing the risk with another party or parties (including contracts and risk financing)
7. Retaining the risk by informed decision

規制の考え方

リスクベース・アプローチ

- 従来型のコンプライアンス型のアプローチではなく、リスクに見合った低減措置を講ずること
- 低リスク分野に必要以上にリソースを割くの避け、高リスクの分野に対する

ソフト・ローとハード・ロー

ソフト・ロー

- ガイドラインなど、強制力を伴わない
- 制定や変更が迅速にできる

ハード・ロー

- 法律など強制力を伴う
- 制定や変更に時間がかかる

資料２－２

「AI制度に関する考え方」について（概要）

令和6年５月

AI戦略チーム

概要

暫定

- AIはイノベーション。一方で、様々なリスクがあり、イノベーション促進のためにも、適切なガードレールが必要。
  （リスクの例）
  製品・サービスの安全性に関するリスク（誤作動など）、人権侵害（プライバシーや公平性など）、安全保障・犯罪増加等のリスク、財産権侵害のリスクなど
- 各国とも、リスクの大きさに応じた対策（リスクベースアプローチ）。
  - EUは主として人権等の観点からAI全体に関して規律。国際規格も併用。
  - 米国はビックテックによるボランタリーコミットメントを基本としつつ、大統領令で、既存法令を活用した大規模汎用モデル開発者からの報告を求めるなど。
  - 日本はAI全体に対してAI事業者ガイドラインで迅速に対応。
- 各国はソフトロー（規格・ガイドライン）とハードロー（法律・基準）の組合せを指向。日本においても制度の要否も含め検討は必要。

様々なリスクを想定し、各主体の役割等を検討していく必要がある。一つの考え方として、以下のような整理の仕方がある。

影響大・高リスクのAI開発者

国民の安全・安心の観点から**ソフトローを補完する法制度の要否の検討**。
変化の速さや多様性を踏まえ、規制の運用は官民連携型の第三者機関が担う「**共同規制型**」、「ゴールベース」も重要。

影響大・高リスクのAI提供者・利用者

基本的には**業法・規制法**があり、その下で対応。業法等のない**重要インフラ等では、技術変化や利用状況に応じて機動的な対応が望まれるが、議論の積上げが必要**。
法令とAI事業者ガイドラインの併用も。

政府

政府による**AIの適切な調達・利用**（他分野への波及効果）。
違法行為へのAI利用など**リスク情報を調査**し、**悪用される蓋然性の高いAIに対する改善・排除措置を検討**。

プロバイダー

偽・誤情報に関して、**オンラインプラットフォーマー等による不適切なコンテンツの削除**など、情報流通全体の枠組みの中で対応。

多様なリスクへの各国の対応

- AIには、製品・サービスの安全性に関するリスク（誤作動など）、人権侵害（プライバシーや公平性など）に関するリスク、安全保障・犯罪増加などに関するリスク、財産権侵害のリスクなど、様々なリスクがある。
- リスクへの対応（ガードレール）とイノベーションは対立概念ではなく、イノベーション促進のためにも適切なガードレールが必要。
- リスクの程度に応じて対策を講じるリスクベースアプローチが適切。EUは人権、米国は安全保障などの観点から、ソフトロー（規格・ガイドライン）とハードロー（法律・基準）の組合せを指向。日本はガイドラインで迅速に対応。

EU 広範なハードローをソフトローで補完

欧州理事会・欧州委員会・欧州議会は**AI法案**に大筋合意し、欧州議会は最終案を承認（2024年3月）。

主として人権侵害、差別・偏見リスクを重大リスクと捉え、センシティブな情報を扱うAIは禁止、製品事故等の危険性がある高リスクなAIにはリスク評価や基準遵守義務、誤使用等のリスクのあるAIには表示義務等。

汎用AIモデルには、透明性要件の遵守義務。影響力の大きいモデルには、より多くの義務。

義務違反には高額の課徴金など罰則。

法制定後2年後に施行（例外あり）。

国際規格、欧州規格等も活用する可能性。

米国 ソフトローをベースにしつつ、目的に応じてハードロー

AI開発大手が**ボランタリー・コミットメント**（2023年7月）。

大統領令を発出（2023年10月）し、イノベーション促進、リスク対応を各省庁に指示。先進的なAIシステムを開発する大手企業による自主的規律遵守を基本としつつも、既存の法令（国防生産法等）を活用し、主として安全保障の観点から、大規模汎用モデル等の開発企業に報告を求めるなどとしている。

G7

2023年、日本はG7議長国として広島AIプロセスを主導し、**高度AIシステムに関する国際指針**、**AI開発者に対する国際行動規範**を策定。

2024年はイタリアが議長国。G7以外へのアウトリーチなど広島AIプロセスをさらに前進。

日本 ソフトローによる対応

2016年の**G7香川・高松情報通信大臣会合**を契機に、G7・G20やOECD等の議論をリードし、貢献。

AIの変化の速さ・複雑さを踏まえ、イノベーションを阻害しない観点から、ソフトローによって目的達成に導くゴールベースの考え方。広島AIプロセスの成果も引用し、**AI事業者ガイドライン**を策定。

- 改善・修正を繰り返すアジャイル・ガバナンスが有効。一方で、リスクの高いAIに関しては一定の規制を導入すべきとの指摘も。
- 幅広い関係者の意見を聴取し、国民の安全・安心を守る観点からAI制度について検討が重要。
- 規制を導入した場合でも、民間の専門的能力、**AIセーフティ・インスティテュート（AISI）**の活用、国際的な連携が必要。

基本的な考え方

■ **リスクや技術進歩に応じた柔軟な制度**

ソフトローを最大限活用しつつ、リスクの高いAIに対して必要な法的規制を検討。国際的な議論等に即応するため、制度の柔軟性が重要。

■ **国際整合性の確保**

G7広島AIプロセス国際指針・国際行動規範を踏まえ、スタートアップや外国企業が安心して事業活動を展開できるようにすべき。

■ **AI事業の主体及びリスクの高低に応じた考え方**

様々なリスクが懸念され、その対応に関しては国内外において様々な議論がある。

一つの考え方として、事業主体を開発者、提供者・利用者に分けた上で、事前対応（許認可、第三者認証、自己宣言・開示等）と事後対応（安全性や脆弱性のリスクがあるAIに対する対応措置）を整理すると以下のとおり。技術の変化、国際動向等も踏まえ、さらなる検討が必要。

（参考）AI関係者を巡る制度検討のイメージ

	影響大・高リスク	影響小・低リスク
AI開発者	① **確実なリスク対応** 米国では大規模なモデルに報告義務 EUハイリスクなAIに様々な義務	② **リスク対応** ルールを遵守していることの開示等
AI提供者・利用者	③ **個別業規制等による基準遵守等** リスクの高い装置・機械類等の安全基準等	④ **リスク対応** AIガバナンスポリシーの策定・公表等
	⑤ **政府による適切なAIの調達・利用** **リスクに関する知見の集積、情報共有**	
プロバイダー	**不適切なコンテンツへの対応** オンラインプラットフォーマーによる対応（EUのデジタルサービス法） テック企業による欺瞞的AI選挙コンテンツの削除等	

①影響大・高リスクのAI開発者

EU・米国はハードローも検討。日本においてもセキュリティ、インシデント対応など安全・安心の観点から、**ソフトローを補完する法制度の要否を検討。**変化の速さや多様性を踏まえ、規制の大枠は決めつつ、運用は官民連携型の第三者機関が担う「**共同規制型**」、「ゴールベース」も検討。

②影響小・低リスクのAI開発者

ソフトローでの対応。目的等に応じた安全確保等やプロダクト認証制度の検討も。

③影響大・高リスクのAI提供者・利用者

業法・規制法の下で対応。業法・規制法がない**重要インフラ等では、技術変化や利用状況に応じた機動的な対応や議論の積上げが必要。法令とAI事業者ガイドラインの併用**も。

④影響小・低リスクのAI提供者・利用者

ソフトローでの対応。ガバナンスを第三者が認証する制度も。

⑤政府によるAIの適切な調達・利用、AI利用に関するリスク情の調査等

政府による**適切な調達・利用**が必要。他への波及効果を有する。
違法行為に利用される事例など**リスク情報を調査・収集**し、**悪用の蓋然性の高いAIに対する事後的な改善・排除措置を検討。**

具体的な考え方

■ 制度に関しては、様々な論点について検討が必要。

影響大・高リスクのAI開発者に対する考え方

前頁①

目的

EUは人権、米国は安全保障など視点の違いはあるが、細かな行為義務の規定ではなく、体制整備や情報開示に力点。

日本においても国民の安全・安心を目的とし、犯罪的行為への悪用の抑止、経済安全保障の観点も含めて検討。

公正競争が確保されているか、引き続き注視。

対象範囲

欧米はモデルの使途、規模などにより規制対象を規定しており、総合的な検討が必要。国外事業者も対象とすることが必要。

制度の考え方

広島AIプロセス国際指針、国際行動規範をベースに検討。技術革新の速さ・不確実性に柔軟に対応するため、ハードローとソフトローの組合せによる「共同規制」も。

1) 広島AIプロセスでは、最先端の基盤モデル及び生成AIシステムを含む高度なAIシステムが議論の対象。米国大統領令では、大規模かつ広範に適用可能なAI（数百億パラメーター、10^{26}FLOPs超）を国防生産法に基づく報告義務対象。EUのAI法案は、汎用AIモデルに透明化義務等を課し、システミックリスクを伴う影響力の大きい汎用AIモデル（10^{25}FLOPs超）にはさらにリスクの評価と軽減、敵対的テストの実施等を課す。

悪用される蓋然性の高いAIに対する考え方

前頁⑤

AIが悪用されて被害が生じた時/生じる可能性が高い時、政府/関係機関は実態を調査・公表し、国民や事業者に対して注意喚起。さらに、必要に応じて改善命令・実名公表等を行うべき。

（参考）ソフトウェアのサイバーセキュリティ確保に関しては、IPA（情報処理促進機構）が調査を行い、必要に応じて事業者または利用者が講ずべき措置を公表する仕組みが情報処理促進法に基づき運用されている。

影響大・高リスクのAI提供者・利用者に対する考え方

前頁③

①人の生命、身体等に直接影響を及ぼすおそれのある分野

安全性の観点から法規制が存在。

医療機器は厚労大臣による承認が必要。AI活用の有無によって、審査に関する基本的考え方は変わらないが、AIを活用した医療機器の評価法や評価指標がある。

自動車は保安基準適合性を国が確認。自動運転車の自動運行装置の適合性も審査。

重要インフラは業法が存在。設備計画や点検等にAIが利用されているが、人による判断の支援であり、制度改正は検討されていない。プラントに関しては、信頼性の高いAIの実装促進の観点から、経産省が「プラント保安分野AI信頼性評価ガイドライン」を公表。

②権利侵害や差別的対応のおそれのある分野

労働基準法や男女雇用機会均等法では、国籍・信条等による差別や男女差別を禁止。現状のAIは人による判断の支援であり、制度改正は検討されていない。

クレジット分野では、包括信用購入あっせん業者は、支払可能見込額の算定においてAI等の技術的手法を利用できるが、「不当な差別、偏見その他の著しい不利益」のおそれがある方法は禁止。

③その他分野

生成AI利用が想定されるコールセンター分野は業界のセキュリティガイドラインで対応。

教育分野は文科省が、「初等中等教育段階における生成AIの利用に関する暫定的なガイドライン」を公表、「大学・高専における生成AIの教学面の取扱いについて」を周知。

④制度の考え方

内閣府は内閣サイバーセキュリティセンター等と連携しつつ、関係省庁等の協力を得て、重要インフラ等を中心とした分野におけるAI利用の実態やリスク管理の状況等について調査するとともに、その結果をAI戦略会議等に報告。

偽・誤情報、知的財産権のリスクに対する考え方

- 偽・誤情報対策に関して、**デジタル空間における情報流通の健全性確保の在り方に関する検討会（総務省）**において検討。
- 生成AIと知的財産に関して、**AI時代の知的財産権検討会（知的財産戦略推進事務局）**において検討。
- 生成AIと著作権に関して、**AIと著作権に関する考え方について（文化審議会著作権分科会法制度小委員会）**を2024年3月とりまとめ。

AIを利用した偽・誤情報等の生成・拡散

技術的対策は、
①AI生成物に電子透かし等を付加、
②コンテンツに出所や来歴等の情報を付与、
③オンラインプラットフォーマーがAI生成物を判別しラベリング
等により、受信者がAI生成物や信頼性ある情報を識別可能とする等が考えられる。

EUは、デジタルサービス法や官民協調による行動規範等により、不適切な情報への対処をオンラインプラットフォーマー等に要請。

日本においても、ネット上の違法・有害情報が問題化。総務省における検討の結果、大規模プラットフォーム事業者に対して①対応の迅速化、②運用状況の透明化を義務づける**情報流通プラットフォーム対処法**[1] が2024年５月に成立。

[1] 特定電気通信役務提供者の損害賠償責任の制限及び発信者情報の開示に関する法律の一部を改正する法律案による改正後の特定電気通信による情報の流通によって発生する権利侵害等への対処に関する法律（平成13年法律第137号）

総務省は2023年11月より、**デジタル空間における情報流通の健全性確保の在り方に関する検討会**を開催し、総合的な対策を検討。2024年夏頃を目途にとりまとめ予定。

なお、なりすまし広告等に起因した被害への対策の検討に当たっては、AIにより情報の改ざん・偽情報の生成が精緻化・巧妙化すること踏まえることが必要。

AIと知的財産権等の関係

AI時代の知的財産権検討会では、4月中間とりまとめ案について議論。

- 知的財産法では直接の保護対象として明記していない労力や作風、声、肖像等も含め、知的財産法のルールのみでは解決できない点も複合的に関わることを踏まえ、AIガバナンスとの連動が必要。
- AIの進歩と知的財産権の保護が両立するエコシステムの確立に向けて、幅広い関係者が法・技術・契約を適切に組み合わせ、アジャイルに取り組む必要。

との観点から知的財産法に係る法的考え方を整理。技術による対応策、契約による対価還元策について検討するとともに、法、技術、契約の各手段は、相互補完的に役割を果たす関係があることを確認。

AI技術の進歩の促進と知的財産権の適切な保護が両立するエコシステムの実現に向けて、AI開発者、AI提供者、AI利用者等の関係主体に期待される取組事例についてとりまとめ。

AIと著作権の関係

生成AIと著作権に関する考え方を整理し、周知するため、令和６年３月に「AIと著作権に関する考え方について」を取りまとめ。

- 開発・学習段階に関して、著作権法第30条の４の適用範囲を明確化。

  第30条の４（抜粋） 著作物は（中略）当該著作物に表現された思想又は感情を自ら享受し又は他人に享受させることを目的としない場合には（中略）利用することができる。ただし（中略）著作権者の利益を不当に害することとなる場合は、この限りでない。

- 生成・利用段階に関して、著作権侵害にあたりうる場合等について、現行の著作権法における考え方を整理・明確化。
- AI生成物が著作物として認められる場合について考え方を整理。

各関係者が法的リスクを自ら把握し、権利の実現を図ることで、著作権者等の権利保護、AIの適正な開発及び利用の環境を実現。

今後は、考え方の周知・啓発、議論の継続、関係者間の相互理解を促進。

「ＡＩ制度研究会」の設置について

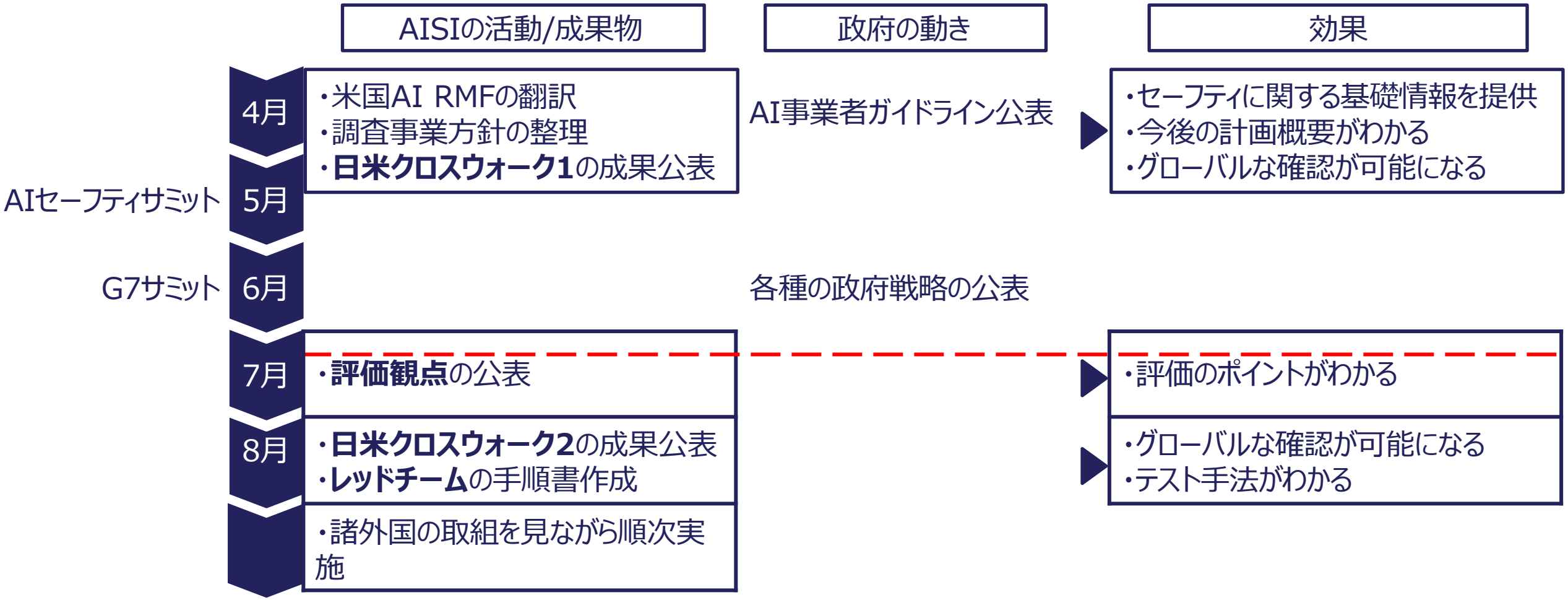
令和６年７月１９日ＡＩ戦略会議決定

- 「統合イノベーション戦略 2024」(2024 年 6 月 4 日閣議決定)に基づき、イノベーション政策強化推進のための有識者会議「ＡＩ戦略」（ＡＩ戦略会議）の下、「ＡＩ制度に関する考え方」等を踏まえ、ＡＩ制度の在り方に ついて検討することを目的として、「ＡＩ制度研究会」（以下「研究会」という。）を設置する。

評価やテスティングについて

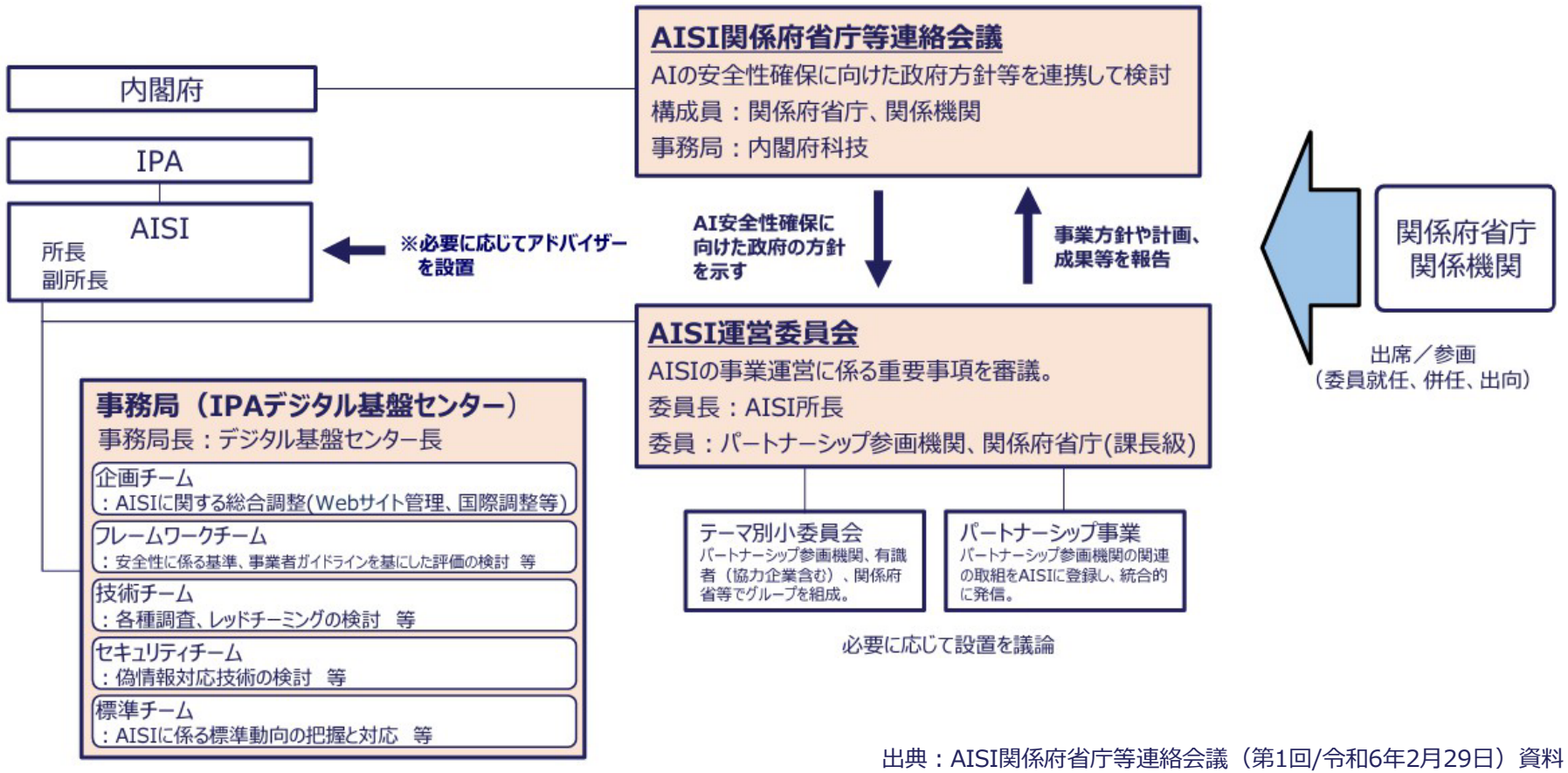
AISIの活動と成果予定物

AISI Japan AI Safety Institute



AISIの推進体制

- 内閣府を事務局とする「AISI関係府省庁等連絡会議」を設置し、重要事項を審議（年間２～３回の開催を予定）。AISIの中に、AISI所長を委員長とする「AISI運営委員会」を設置（月１回の開催を予定）。
  - 運営委員会の下に、必要に応じて、「テーマ別小委員会」や「パートナーシップ事業」（研究機関等の関連の取組みをAISI事業として発信）を設置。



【関係府省庁】
・内閣府（科学技術・イノベーション推進事務局）
・国家安全保障局
・内閣サイバーセキュリティセンター
・警察庁
・デジタル庁
・総務省
・外務省
・文科省
・経産省
・防衛省

【関係機関】
・情報処理推進機構（IPA）
・情報通信研究機構
・理化学研究所
・国立情報学研究所
・産業技術総合研究所

出典：AISI関係府省庁等連絡会議（第1回/令和6年2月29日）資料２

AISI

Japan AI Safety Institute

世界最先端で働きたい人材大募集中

本資料は、2024年7月22日（月）開催、JIPDECセミナーで配布した資料です。セミナーお申込み者様限定での配布となりますので、WEB、SNS等への掲載、転載はご遠慮ください。