【講演レポート】

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

私が副所長を務めている AI セーフティ・インスティテュート（AISI：エイシー）は、規制当局ではなく規制当局である内閣府等政府の検討を技術的に支援する組織という位置で、情報のハブとして世界各国の動向などを情報収集しています。

AI に関して、すでに規制の詳細な部分まで検討が進んでいるのではないか、というご質問をいただくことが多くありますが、実際にはまだ世界中で AI 規制をどうすべきかを検討している最中であり、詳細が決まっているものではありません。それを踏まえて、現在の世界の状況や日本がどのように対応しているかという点をお伝えしたいと思います。

AISI とは

AI はイノベーションとセーフティが車の両輪で、非常に大きな力を持つ AI を安心して開発したり活用するためにはセーフティが重要となります。

AISI は、内閣府を中心に 10 府省、5 政府関連機関が連携し、官民の取組を支援する裏方的な組織として今年 2 月に設立されました。事務局は、独立行政法人情報処理推進機構（IPA）に設置され、4 月から本格的に稼働し始めたところです。世界各国の AISI では、包括的な AI 法が成立している EU の AI オフィスで 60 名程度、他国も 30～40 名程度、日本も現在 25～26 名なので、各国ともまだ組織の立ち上げ中という段階で、具体的な検討等はこれからとなります。

AISI の役割としては、

- AI セーフティに関する調査、評価手法の検討や基準作成等の支援
  専門調査を行う部門が、各国の AI セーフティに関する主に技術的な検討の状況等を収集し、評価のあり方を考える。
- 日本の AI セーフティのハブとして、産学の最新情報を集約し、関係団体間の連携を促進
  散在する解説や事例、業界ガイドライン等に関する情報の集約と提供、現在 40 以上ある AI 関連団体間の連携を図る。現在は団体等の情報を収集している段階。
- 他国の AI セーフティ関連機関との連携
  現在最も活発に行われている活動。各国政府機関やビッグテック等と緊密に連携し、AI セーフティを保つためのルール形成のあり方について意見交換を行っている。

が挙げられます。また、スコープは、現時点では明確化せず、海外の動向や企業等のニーズを把握しながら柔軟にスコープを設定し取組を進めていくこととしています。

AI のリスクとは

AI を規制する必要があるのは、そこにリスクがあるためなので、まずはリスクを把握しそのうえでそれぞれのリスクに対してどのような対処が望ましいか（法律で厳格に規制すべきか、自主規制に任せるべきか等）を検討する必要があります。

AI に関しては、世界に先駆け日本が 2019 年に公表した「人間中心の AI 社会原則」で 7 つの原則を定めています。今年 4 月に公表された「AI 事業者ガイドライン」では、この原則をもとに AI 開発者、AI 提供者、AI 利用者それぞれのフローの中で想定されるリスクを整理し、ライフサイクルを通じた AI リスク検討の必要性を説明しています。

また、7 つの「人間中心の AI 社会原則」を細分化し、AI で想定されるリスクを 10 の原則に基づき共通の指針、主なリスクを整理しています。AI 事業者ガイドラインは、AI を安全安心に利用するためのガイドなので、AI リスク一覧ということではなく、何を考える必要があるか、考える視点が示されています。

英国の AI セーフティの原則では、AI はフィジカル、ソーシャル、経済、心理的なリスクがあるされています。フィジカルとは、車載 AI が暴走して事故を起こし物理的に人に被害を及ぼすもの、ソーシャルは偽情報等により社会に混乱を招くもの、エコノミカルは株価への影響等財政的損失を負わせるもの、サイコロジカルなリスクとは AI が生成した情報や画像を見てショックを受ける、雇用への影響を与えるというものです。

この 4 つの観点から、それぞれの原則に対して現時点で考えられる主なリスクをまとめています。

表 1：AI に関して想定されるリスク

	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 • ステークホルダー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
4)	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

プライバシー保護		
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
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

出典：「AI 事業者ガイドライン」（経済産業省、総務省）

「リスク」というと、特に日本ではネガティブなイメージにとらえられがちですが、ISO 31000 ではネガティブにもポジティブ（「顧客が急増して対応しきれない」等）にも捉えられるもの「物事の不確実性に影響があるもの」として定義されています。ISO9001 でも同様に、不確実性に対する影響と定義しています。

また、リスクのレベル感に関しては、例えば欧州 AI 法では、受け入れがたいリスク（生命の危機を生じさせる等）、ハイリスク（社会がマヒする等）、限定的なリスク（業界や一部地域等）、ミニマムリスク（スパムメール等）と分類されています。また、リスクへの対応として、回避するのか受容するのか、あるいは転嫁、低減するかと言った点を考える必要があり、そのうえで規制も必要だという声も上がってきます。

規制の考え方

AI への取組に関しては、現在ほとんどの国がリスクベースアプローチをとっています。従来のコンプライアンス型は、ルールを作成しそれが遵守されているかどうかをチェックするもので、広くリスクに

見合った低減措置は可能ですが、効果に関しては懐疑的です。一方、リスクベースアプローチでは低リスク分野に必要以上のリソースを割かず、高リスクな領域や想定されるリスクに対してアプローチする考え方です。

もう 1 つの検討課題は、ソフト・ローによるアプローチをとるか、ハード・ローによる対応を取るかという点が挙げられます。ソフト・ローというのは、ガイドラインや基本的に強制力を持たないものですが、技術変化に対して迅速かつ柔軟に制定変更が可能です。一方、法律等のハード・ローは強制力を伴う一方で、制定までに非常に時間がかかります。これに関しては、各国で考え方が異なっています。最近では、アジャイルガバナンスという言葉も聞かれます。基本的な部分を法律等で対応し、細かな運用等はソフト・ローで対応するといったことも検討されています。いずれにおいても、組織やプロセスを対象とするのか、プロダクト・サービスを対象とするのか、どういうタイミングで確認するかという点などを整理していく必要があります。

「「AI 制度に関する考え方」について」

規制の考え方という点では、今年 5 月に内閣府 AI 戦略チームが「「AI 制度に関する考え方」について」という資料を取りまとめ公表しています。これは、「規制が必要だ」「ソフト・ローで対応すべき」等様々なご意見がある中で、日本としてどうあるべきかを考え始めるために作成されたものです。ここにまとめられた各国の考え方を見ても、リスクベースアプローチという点では一致していても各国で観点や対応が異なっています。共通しているのは、規格やガイドラインといったソフト・ローと法律等ハード・ローの組み合わせにより、効率的で技術の変化に迅速に対応することが必要だという認識です。日本でも、AI 戦略会議でこの資料が議論され、AI 制度研究会が設置され、制度の要否から検討されることとなるなど、まだまだ各国ともやっと検討段階に入ったという状況です。

多様なリスクへの各国の対応

現在各国で検討されている内容がリスク対応にフォーカスしているように思われるかもしれませんが、リスク対応とイノベーションは対立概念ではなく、イノベーションするために安心安全に使うためのルール作りという位置づけです。

EU では、広範なハード・ロー（AI 法）をソフト・ローで補完しようとしています。偏見等を重大リスクと位置づけたり、製品事故が想定されるものへのリスク評価義務付け、汎用 AI モデルへの透明性要件の遵守義務を課し、違反に対しては課徴金による制裁を行う方向で 2 年後の施行に向けて詳細を詰めています。また、AI 関連だけでなくデータに関する法律等も精力的に整備し、着実にデータ社会/AI 社会に対して布石を打っています。

一方、米国は AI 開発大手がボランタリー・コミットメントを出したり、大統領令として規制整備の指示やスケジュールを発出、G7 では 2023 年に日本が議長国となって広島 AI プロセスを主導し、高度 AI システムに関する国際指針や AI 開発者に対する国際行動規範を策定しました。

日本は、どちらかと言えばソフト・ロー対応として AI 事業者ガイドライン等も早い段階で出し、アジャイルガバナンスという形で見直していこうとしています。

多くの方からご質問をいただく世界的な相互運用性に関する検討も、当然考えていかなければならない部分ではありますが、今はまだそこまでの議論はされていません。相互認証のような形を取るのか、

どこかの国の認証を取得すれば OK とするのか、あるいは＋αの確認を求めるようにするのか等は今後の検討となりますが、各国でルールがバラバラで良いとはベンダーも行政側も思わないので、将来的には整合が取れる形に進んでいくのではないかと思われます。

AI 制度研究会による検討

7 月 19 日に開催された AI 戦略会議において、今後、AI の制度の必要性や対象範囲を検討する「AI 制度研究会」の設置が決定しました。この研究会において、これから議論が本格化すると思われますので注視していく必要があると思います。

評価やテスティングについて

リスクの取り扱いに関しては、法律による規制以外にも評価やテスティングという手法もありますが、何をどのように評価するかという課題があります。これまでご紹介した「AI 事業者ガイドライン」は米国 NIST の「AI リスクマネジメントフレームワーク(AI RMF)」と対をなすものとして用語の対応関係を整理し、5 月に日米クロスウォーク 1 の成果として公表しています。

現在は、評価の観点に力を入れており、各国のガイドライン等を参照しつつ作業を進めています。また、8 月には米国 NIST AI RMF と AI 事業者ガイドラインを項目レベルで相互参照できるものを日米クロスウォーク 2 の結果として公表する予定としています。

最後に

AISI は、内閣府が進める AI 戦略の下、一枚岩のチームとなって関係府省庁と連携して今後本格的な活動に入っていきます。また、AISI では汎用的な内容を検討していきますが、業界ごとの自主ガイドライン等、民間主導の取組も行われていますので、そういったところとも連携し、他業界でも参考になる共通的な内容を各分野のガイドに組み込んでいくといったことも必要になってくると思います。

AI の変化が速く、世界の取組も日々変わっている中で、各国とも AISI の本格活動に向けて体制を構築しているため、今はまだみなさんから寄せられる具体的な詳細に関する質問に答えられる段階にありませんが、社会的受容性にも配慮しながら課題の優先度やニーズを見極めながら整備に向けた活動を行っています。

今後、テーマ別の委員会等も活動開始を予定しており、AISI では現在、人材を募集中です。ご関心のある方はぜひ AISI の Web サイトをご確認ください。

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

1990 年 4 月 NTT データ通信株式会社 入社（現 株式会社 NTT データ）
2008 年 7 月経済産業省 CIO 補佐官
2012 年 8 月内閣官房 政府 CIO 上席補佐官
2021 年 9 月デジタル庁 データ戦略統括
2023 年 7 月 IPA デジタル基盤センター センター長 現職兼務
2024 年 2 月 AI セーフティ・インスティテュート事務局長 兼務
2024 年 4 月 AI セーフティ・インスティテュート副所長 兼務

本内容は、2024 年 7 月 22 日に開催された JIPDEC セミナー「AI 規制について　欧米の動向と日本の状況」講演内容を取りまとめたものです。