責任あるAIの推進のための法的ガバナンスに関する素案

2024年2月

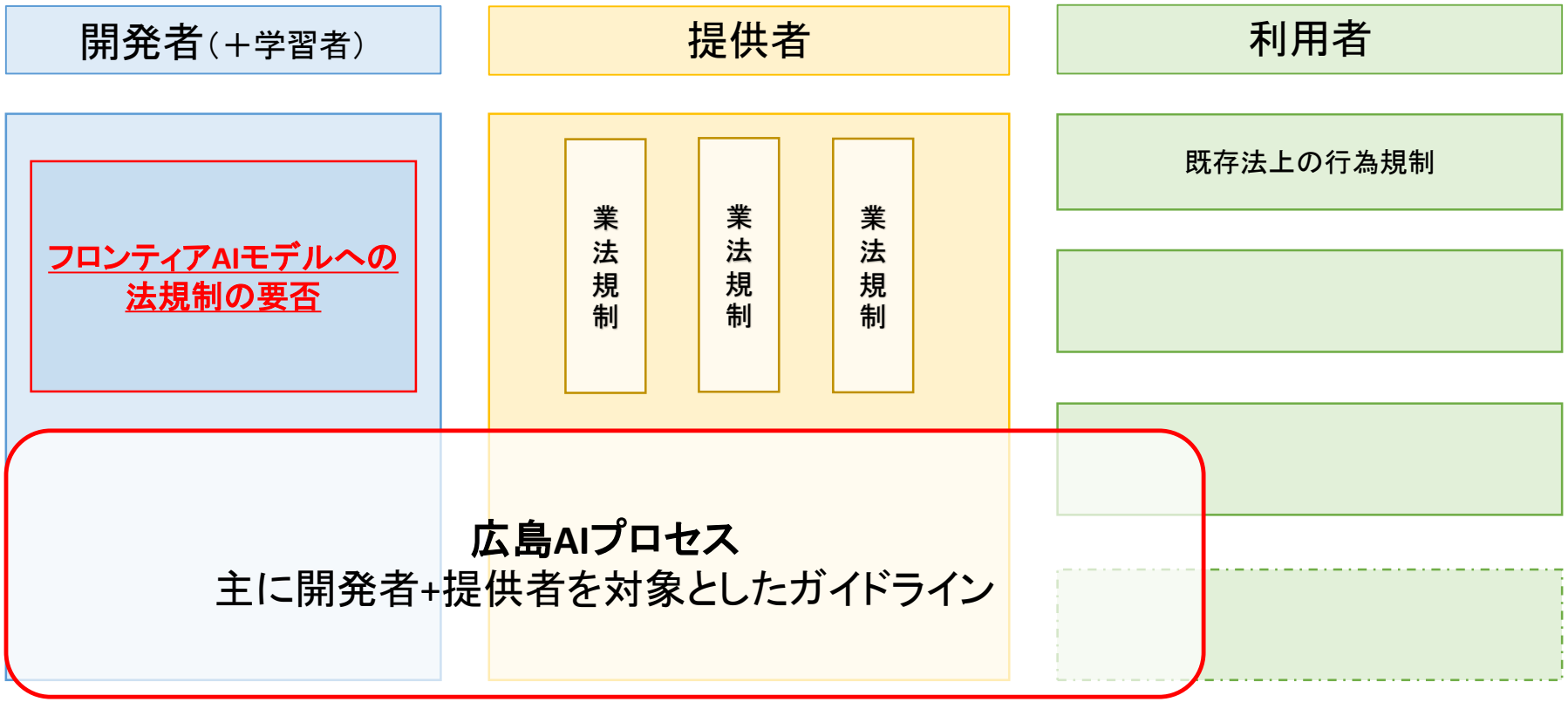
自民党AIの進化と実装に関するPT　WG有志

殿村桂司弁護士、岡田淳弁護士、生貝直人一橋大学教授、

丸田颯人弁護士、小谷野雅晴弁護士

1.各国で進むフロンティアAIモデルへの規制

ガイドライン項目の実効性担保の手法は各国の政策的判断に



欧米中ではフロンティアAIモデルに対し法規制へ

欧州

- 2023年12月に欧州連合の主要機関は人工知能の包括規制案について大筋合意。
- 基盤モデルAIについて、透明性確保や健康・安全・基本的権利・環境・民主主義・法の支配に対する合理的に予見可能なリスクの特定・軽減・緩和といった要件が追加。
- 汎用AIのうち、システミックリスクを有する汎用AIシステムについてはシステミックリスク評価及び軽減の義務が加重。
- AI規則の違反については、違反類型に応じて、最大で3500万€または前事業年度の世界全体における売上総額の7%に相当する金額の巨額の制裁金が課される。

米国

- 2023年10月30日に、「AIの安全、安心、信頼できる開発と利用に関する大統領令」が公布。
- 大統領令は連邦政府機関に対して法的拘束力を有するにとどまり、民間事業者の権利義務や罰則を直接定めているわけではない。
- しかし、2024年1月29日には大統領令にしたがい、デュアルユース基盤モデルを訓練するために米国のIaaS提供者とそのサービスの外国再販業者が取引する場合は国防生産法に基づき報告や記録の提供を求める旨の商務省規則案が公表されており、同規則違反は刑罰の対象に。

中国

- 2023年8月より「生成AIサービス管理暫定弁法」が施行。
- 生成系人工知能サービスを提供する者に対して、アルゴリズム設計、モデル生成においての差別禁止等を要求。また、世論形成力、社会動員能力を有するモデルを提供する場合については、安全評価の実施と結果の提出義務を課している。
- 政府による監督検査の権限に加え、違反した場合には制裁金等の罰則が科される旨定められている。

生成AIの急速な普及による社会の変化に対応するため、従前からAI規則を検討していた欧州だけでなく米国・中国も、AIをガイドライン等のソフトローではなく法的拘束力を持ったハードローによって規制する方向へ進展している。

一方、日本は、「AI事業者ガイドライン」等のソフトローによる規制が中心。

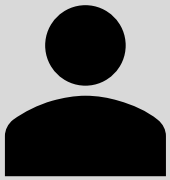
日本だけ法規制がない場合のリスクシナリオ

シナリオA

米国大手
A社

欧米のフロンティアモデルについて、平時においては広島AIプロセス等に則り、事業者の自主判断により、一定の情報提供が日本政府にもなされることを期待。

いざ何らかのトラブルが発生した場合、法規制が整備された欧米当局に対する対応が優先され、欧米当局には詳細な情報が提供されているのに、日本政府は事態把握の精度と速度で劣後することが懸念され、被害拡大のリスクも。

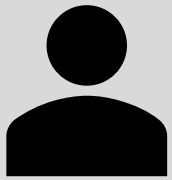


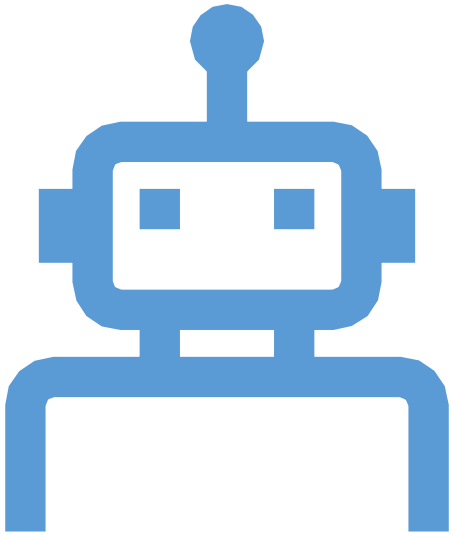
シナリオB

新興国
B社

広島AIプロセスなどの国際的枠組みに参加していない国の企業が開発したフロンティアモデルを用いたサービスが日本国内で展開。

そもそもモデルの仕様、学習データ、セキュリティ、情報管理体制などについて、日本政府として一切把握できずリスク評価すら困難。放置すれば経済安全保障の観点からも重大なリスクとなる可能性も





2.責任あるAI推進基本法（仮称）の骨子

2.1　法的枠組みの全体像

責任あるAI推進基本法（仮）フロンティアAIモデルに対する官民の共同規制

立法趣旨

- 生成AIを含むAIの利活用により基本的人権をはじめとする国民の権利利益が侵害される**リスクを最小化**しつつ、
- AIによるイノベーションを含むAIの健全な発展による**利益を最大化**するため、
- 安全、安心で信頼できる責任あるAIの設計、開発及び導入並びに人間を中心としたAIの利用を可能とするような、開かれた環境の整備を促進する。

法律の構造

①責任あるAI利活用の促進

②特定AI基盤モデル開発者の指定

③特定AI基盤モデル開発者の体制整備義務

④義務遵守状況の報告義務と監督

⑤罰則等

AIによる利益の最大化

AIによるリスクの最小化

国

責任あるAI利活用の推進のための支援

◆ 官民におけるAIの利活用を推進し、社会課題の解決を目指す

施策例：AIの技術革新を推進する官民パートナーシップの構築・強化

◆ AI人材の育成・誘致と研究開発力の強化

施策例：AIの研究開発のための助成金・補助金等交付

◆ 先進的AIの安全性に関する研究機関の機能強化

施策例：今般創設されたAIセーフティ・インスティテュート（以下「AISI」）の機能強化

AI開発者
AI提供者
AI利用者
等のステークホルダー

国

特定AI基盤モデル開発者の指定
モニタリング・監督
罰則

体制整備状況についての定期的報告

特定AI基盤モデル開発者

意見聴取

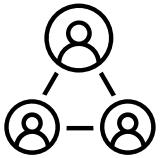
共同規制

遵守

モニタリング結果の公表

体制整備義務の具体化
規格や行動規範の制定

国民



民間等
（業界団体・標準化機関やステークホルダー）

各国のフロンティアAIモデル開発者に関する規制の比較

	米国	欧州	日本 （本素案）
規制内容	・15社による8項目の自主誓約 ・デュアルユース品については国防生産法の規律に	・許容できないリスクについては禁止、ハイリスクAIについては規制 ・開発者に対してはモデルが遵守すべき要件設定	・7項目の体制整備義務
罰則	・国防生産法に違反すれば罰則対象に	・罰則対象：最大3500万ユーロまたはグローバル売り上げの7%	・罰則あり。金額等は今後の検討
施行時期	・未定。但し、本年1月29日に規則案が公表	・法律成立後2年以内	・今後の検討

2.2　フロンティアAIモデルの指定と体制整備

フロンティアAIモデルの指定と体制整備

特定AI基盤モデル開発者の指定

国：一定の規模・目的のAI基盤モデル開発者を「特定AI基盤モデル開発者」に指定する

論点

✓「基盤モデル」の「開発者」を規制の対象とする必要性・許容性の整理
✓「規模」「目的」を何を指標にして評価・区分するか（例：パラメータ数、学習データ、汎用目的か否か）
✓指定は一方的に行うか、まず届出をさせるか。一方的に行う場合、指定のための調査権限を国に認めるか
✓届出すべきであるのに届出しない事業者に制裁するか
✓適用の地理的範囲（日本で提供されるサービスに「利用」されるモデルに限定するか。）

特定AI基盤モデル開発者の体制整備義務

国：米国の「自主誓約」を参考に、事業者に以下の項目を含む体制の整備義務を課す

□特にリスクの高い領域におけるAIについては自社・外部による安全性検証（Red team test等）を行う
□リスク情報を企業・政府間で共有する
□未公表の重み付けを守るサイバーセキュリティへの投資
□第三者による脆弱性等の検出と報告
□生成AIの利用を利用者に通知する仕組みの採用
□AIの能力、限界等の公表
□AIがもたらす社会的リスクの研究推進

民間：各事業者又は業界団体が上記の義務内容を具体化する規格や行動規範を制定・公表する

EU AI Actや米国の大統領令等及び関連ガイダンス等を参考に内容を具体化する

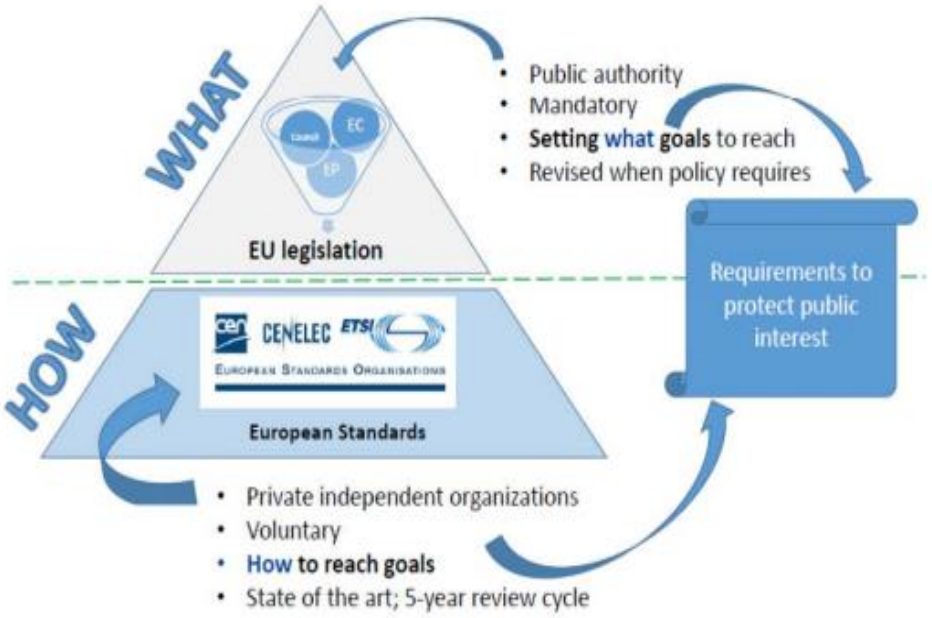
論点

✓EU AI Actの整合規格のように民間組織にAIの品質担保のための規格策定を委ねるか
✓利害関係者を含めた議論に基づく具体的な行動規範の制定の要否（例：EUデジタルサービス法では、欧州委員会が利害関係者を招請して行動規範を策定している。EUのAI規則においては行動規範をそれぞれの提供者または利用者によって策定されることも想定される旨明らかにされている。）
✓民間機関による認証制度等を設けるべきか

体制整備の項目は米国の「自主誓約」等と原則合致

米国自主誓約の主要項目	本法の体制整備義務
レッドチームテスト：Commit to internal and external red-teaming of models or systems in areas including misuse, societal risks, and national security concerns, such as bio, cyber, and other safety areas.	特にリスクの高い領域におけるAIについては自社・外部による安全性検証（Red team test等）を行う
危険リスクの共有：Work toward information sharing among companies and governments regarding trust and safety risks, dangerous or emergent capabilities, and attempts to circumvent safeguards	リスク情報を企業・政府間で共有する
サイバーセキュリティ投資：Invest in cybersecurity and insider threat safeguards to protect proprietary and unreleased model weights	未公表の重み付けを守るサイバーセキュリティへの投資
第三者検証：Incent third-party discovery and reporting of issues and vulnerabilities	第三者による脆弱性等の検出と報告の促進
ウォーターマーク：Develop and deploy mechanisms that enable users to understand if audio or visual content is AI-generated, including robust provenance, watermarking, or both, for AI-generated audio or visual content	生成AIの利用を利用者に通知する仕組みの採用
能力、仕様等の公表：Publicly report model or system capabilities, limitations, and domains of appropriate and inappropriate use, including discussion of societal risks, such as effects on fairness and bias	AIの能力、限界等の公表
社会的リスクに関する研究：Prioritize research on societal risks posed by AI systems, including on avoiding harmful bias and discrimination, and protecting privacy	AIがもたらす社会的リスクの研究推進
社会課題解決に向けた開発促進：Develop and deploy frontier AI systems to help address society's greatest challenges	―

体制整備の具体的水準については民間の知見を活用



CEN-CENELEC “Drafting Harmonized Standards in support of the Artificial Intelligence Act (AIA)” より

EUのAI規則では法律が抽象的な要求事項のみを定義し、具体的な遵守事項は民間が規格（CEマーク）として具体化することが想定されている。

本法においても特定AI基盤モデル開発者の体制整備義務を法律上は抽象的に定め、各事業者又は業界団体が上記の義務内容を具体化する規格や行動規範を制定・公表することを想定。

2.3　報告義務と監督

体制整備に関する報告義務とモニタリングレビュー

義務遵守状況の報告義務と監督

国：特定AI基盤モデル開発者に、定期的に体制整備義務の遵守状況を国または第三者機関（例：AISI）に報告する義務を課す

論点

✓国への報告にとどまらず対外的な開示まで求めるか

国及び民間：国及び国は報告内容に基づき特定AI基盤モデル開発者のモニタリングレビューを行う。国は民間等の利害関係者の意見を聴取することができる

国：国は評価の結果を公表するとともに、一定の場合には是正を特定AI基盤モデル開発者に求めることができる

国：特定AI基盤モデル開発者が義務を遵守していない場合やインシデントが発生した場合等に報告徴求や立入検査をできる

罰則等

国：報告義務・命令違反に対して課徴金・刑罰を科す

論点

✓科す場合の制裁金又は罰金の金額はどうするか

✓どのような場合に刑罰を科すか

民間：認証等の取消・一時停止等

参考：デジタルプラットフォーム取引透明化法における共同規制モデル

特定デジタルプラットフォーム提供者の役割

取引条件等の情報の開示

利用者に対する**取引条件の開示**や**変更等の事前通知**を義務付けることで、取引の透明性を向上させる。

【開示項目の例】
- ✓取引条件変更の内容及び理由の事前通知
- ✓他のサービスの利用を有償で要請する場合に、その内容及び理由
- ✓データの利用範囲
- ✓出品の拒否・停止の理由
- ✓検索順位を決する基本的な事項

【行政措置・罰則】
- ✓勧告・公表で改善を促す。
- ✓是正されない場合に限り措置命令
- ✓措置命令違反には罰則

自主的な手続・体制の整備

特定デジタルプラットフォーム提供者は、**指針**に基づいて必要な措置をとり、**公正な手続・体制の整備**を行う。

【指針の内容の例】
- ✓取引の公正さを確保するための手続・体制の整備
- ✓苦情処理・紛争解決のための体制整備
- ✓関係者と緊密に連絡を行うための体制整備（国内管理人の選任）
- ✓取引先事業者の事情等を理解するための仕組構築

【行政措置】
- ✓措置の適切・有効な実施のため特に必要な場合は、勧告・公表で改善を促す。

運営状況の報告書

報告書の内容

(1)事業の概要
(2)苦情処理の状況
(3)情報開示の状況
(4)自主的な手続・体制整備の状況
(5)自己評価結果

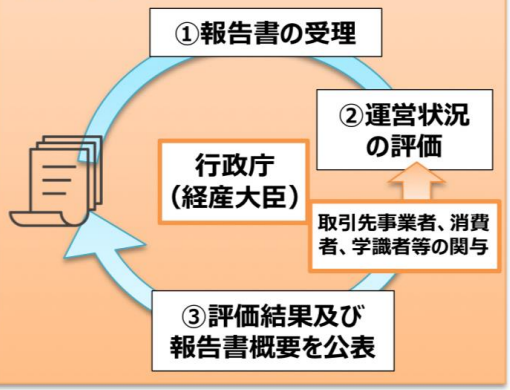
※不提出、記載事項漏れの場合には罰則

評価結果を踏まえた自主的改善

行政庁の役割

レビュー（評価）の実施

特定デジタルプラットフォーム提供者の運営状況について、**取引先事業者や消費者、学識者等も関与してレビュー**を行い、結果を公表する。



公取委への措置請求

独占禁止法違反のおそれがあると認められる事案を把握した場合、経産大臣は公取委に対し、同法に基づく対処を要請する。

出典：デジタルプラットフォーム取引透明化法の概要（経済産業省）

責任あるAI推進基本法（仮）の骨子

立法趣旨

立法趣旨：生成AIを含むAIの利活用により基本的人権をはじめとする国民の権利利益が侵害されるリスクを最小化しつつ、AIによるイノベーションを含むAIの健全な発展による利益を最大化するため、安全、安心で信頼できる責任あるAIの設計、開発及び導入並びに人間を中心としたAIの利用を可能とするような、開かれた環境の整備を促進する。

①責任あるAI利活用の促進

国：官民におけるAIの利活用を推進し、社会課題の解決を目指す

施策例：AIの技術革新を推進する官民パートナーシップの構築・強化

国：AI人材の育成・誘致と研究開発力の強化

施策例：AIの研究開発のための助成金・補助金等交付

国：先進的AIの安全性に関する研究機関の機能強化

施策例：今般創設されたAISIの機能強化

②特定AI基盤モデル開発者の指定

国：一定の規模・目的のAI基盤モデル開発者を「特定AI基盤モデル開発者」に指定する

論点

- ✓ 「基盤モデル」の「開発者」を規制の対象とする必要性・許容性の整理
- ✓ 「規模」「目的」を何を指標にして評価・区分するか（例：パラメータ数、汎用目的か否か）
- ✓ 指定は一方的に行うか、まず届出をさせるか。一方的に行う場合、指定のための調査権限を国に認めるか
- ✓ 届出すべきであるのに届出しない事業者に制裁するか
- ✓ 適用の地理的範囲（日本で提供されるサービスに「利用」されるモデルに限定するか。）

民間：届出義務を課す場合は、対象となる事業者は届出を行う

③特定AI基盤モデル開発者の体制整備義務

国：事業者に以下の項目を含む体制整備に関する義務を課す

- ☐ 特にリスクの高い領域におけるAIについては自社・外部による安全性検証（Red team test等）を行う
- ☐ リスク情報を企業・政府間で共有する
- ☐ 未公表の重み付けを守るサイバーセキュリティへの投資
- ☐ 第三者による脆弱性等の検出と報告
- ☐ 生成AIの利用を利用者に通知する仕組みの採用
- ☐ AIの能力、限界等の公表
- ☐ AIがもたらす社会的リスクの研究推進

③特定AI基盤モデル開発者の体制整備義務（続）

民間：各事業者又は業界団体が上記の義務内容を具体化する規格や行動規範を制定・公表する

論点

- ✓ EU AI Actの整合規格のように民間にAIの品質担保のための規格策定を委ねるか
- ✓ 利害関係者を含めた議論に基づく具体的な行動規範の制定の要否（例：EUデジタルサービス法では、欧州委員会が利害関係者を招請して行動規範を策定している）
- ✓ 民間機関による認証制度等を設けるべきか

④義務遵守状況の報告義務と監督

国：特定AI基盤モデル開発者に、定期的に③の義務の遵守状況を国または第三者機関（例：AISI）に報告する義務を課す

論点

- ✓ 国への報告にとどまらず対外的な開示まで求めるか

国及び民間：国は報告内容に基づき特定AI基盤モデル開発者のモニタリングレビューを行う。国は民間等の利害関係者の意見を聴取することができる

国：国は評価の結果を公表するとともに、一定の場合には是正を特定AI基盤モデル開発者に求める

国：特定AI基盤モデル開発者が義務を遵守していない場合やインシデントが発生した場合等に報告徴求や立入検査をできる

⑤罰則等

国：義務・命令違反に対して課徴金・刑罰を科す

民間：認証等の取消・一時停止等

ご清聴ありがとうございました