AIガバナンスの枠組みの構築に向けて

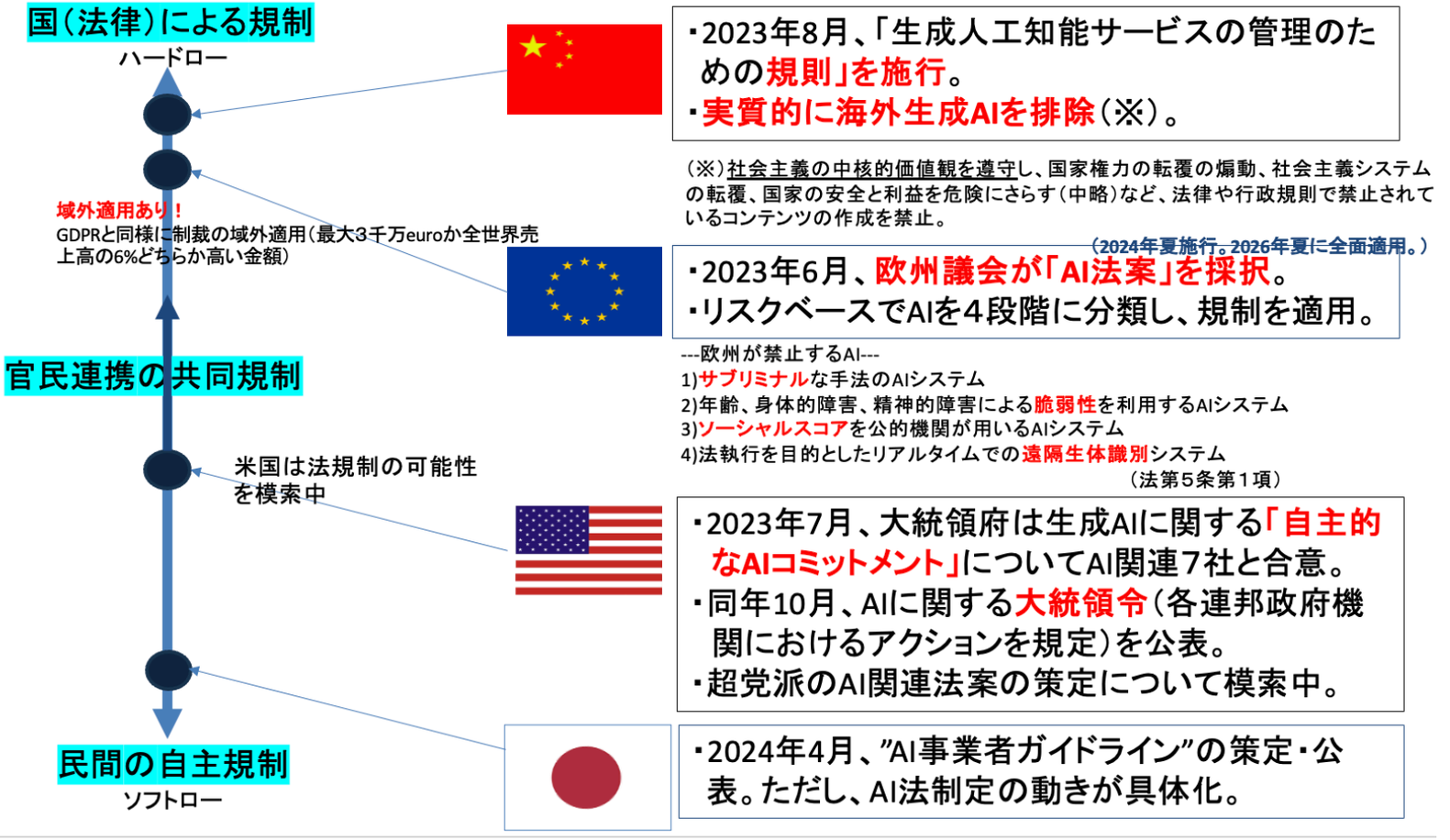
Ver 1.0

2024年7月

デジタル政策フォーラム

本文書の目的

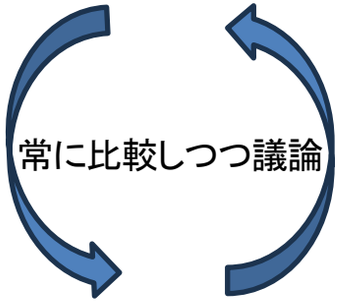
✔欧州、中国、米国などで**AIのルールづくりが進展**（理念的議論から具体的な議論へ）



✔**日本におけるAI関連法制の議論**やグローバルに広がりを見せている
AIガバナンスの議論の方向性を示す

基本的考え方

1　リスクの最小化



AIが人間を操るリスク/人間を代替することで生じるリスクの最小化（可能な限り技術的解決を目指し、必要以上の規制の導入は不適当）

(1) **リスク**管理
(2) **規制**の手法と実効性
(3) **モデル崩壊**の可能性
(4) **生成物**の取扱い

2　利便性の最大限享受

AIのパーソナル化（インテリジェンスの分散化）を通じた個人のデータ利用にかかる主権(sovereignty)を技術的に担保しつつ、利便性の高いサービスを享受

(5) **デジタル差別**の禁止
(6) AIの**積極的活用**

3　健全な市場の育成

上記を可能な限り自律的に実現する市場の創出

(7) 健全な**エコシステム**の構築
(8) **オープン性**の確保
(9) 国際的**コンセンサス**の醸成
(10) **倫理的問題**への対処

リスクの最小化

(1)リスク管理

- コントロールすべきリスクの範囲/リスクのランク分けの基準
- リスク評価の対象---AI開発者（AIを組み込むサービス提供事業者は対象外）
- 自己評価と第三者評価の組み合わせ方
- 脆弱性調査(red teaming)---技術的基準・監査制度
- AIサイバーセキュリティ対策（AIによるサイバー攻撃/AI学習データの汚染）

(2)規制の手法と実効性

- 技術革新を阻害しない/表現の自由の確保
- ハードロー（技術基準などAIリスクのセーフガード）＋共同規制（偽情報対策等）
- 規制対象---AI開発者（登録制または届出制）
- 社会的影響度を踏まえ一定規模以上を規制対象

(3)モデル崩壊の可能性

- 学習データの取扱ルール（技術基準、認証制度等）
- オープンデータ化の推進

(4)生成物の取り扱い

- 共同規制による偽情報対策
- データの完全性（非改ざん）を証明する電子透かしの導入
  →分散型の電子透かしを含め制度運用面・技術面から検討（継続的な検証と交信が必要）

（注）上記(1)から(4)の項目については、急速な技術革新が進む中、過去のAI関連ガイドライン議論の中には市場の実態からかけ離れ、必要以上に議論が為念的・抽象的なものになる傾向も散見された。このため、規制の実効性等を議論していく上で、AI開発者等による情報公開を積極的に促す仕組みを整備し、常に実態の「見える化」を進めていくことが求められる。

利便性の最大限享受

(5)デジタル差別の禁止

- AIモデルの公平性・中立性を確保するための監査制度
- AI開発者と(AIを実装した)サービス提供事業者の責任分界点の明確化

(6)AIの積極的活用

- AI活用による教育・医療の個別化(personalization)の推進
- 過度のプロファイリングを防ぐセーフガード措置
- 環境問題、防災、文化などの分野でのAI活用のための技術開発
- 個人データの取扱についてプライバシー保護の観点から検討
- AIの「集中と分散」に関する議論が必要
- AIリスクに関する周知啓発活動の推進

健全な市場の育成

(7)健全なエコシステムの構築	■ AI関連市場における巨大企業による優越的地位の濫用の防止 ■ AI起点の隣接市場での市場支配力濫用防止の仕組みの検討 ■ 域外適用の規定の妥当性の検証
(8)オープン性の確保	■ オープンソースの活用 ■ 異なるAI間の相互運用性の確保（技術標準化の促進） ■ オープン型のAI開発を促すことを前提とした研究開発支援 ■ 上記をベースとしたソリューションの開発など振興策の推進
(9)国際的コンセンサスの醸成	■ 国内ルールの国際議論との整合性を確保するための取組み ■ グローバルサウスの議論への十分な参加 ■ AIの軍事利用に関する規範形成
(10)倫理的問題への対処	■ 生命科学と同様の研究倫理規定や研究承認プロセスの確立

今後の作業計画

✔DPFJでは、AIの**技術・政策・利活用**という3つの観点から有識者のヒアリングを継続

✔**本文書の更新**を定期的に行うとともに、**オープンフォーラム**を開催

✔**2024年末を目処に最終的な文書**に取りまとめる予定

（他のフォーラム等との連携を積極的に推進）