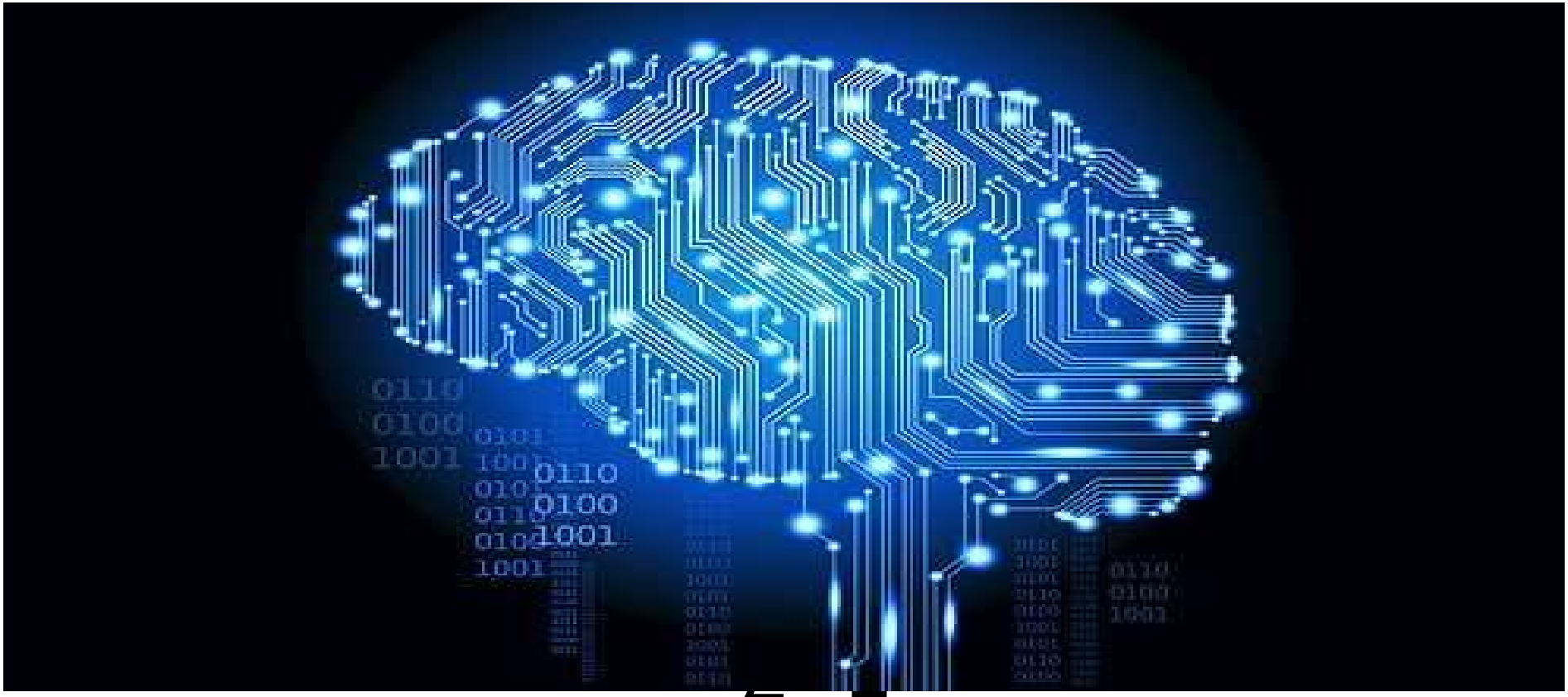
AIガバナンスの枠組みの構築に向けて

Ver 2.0



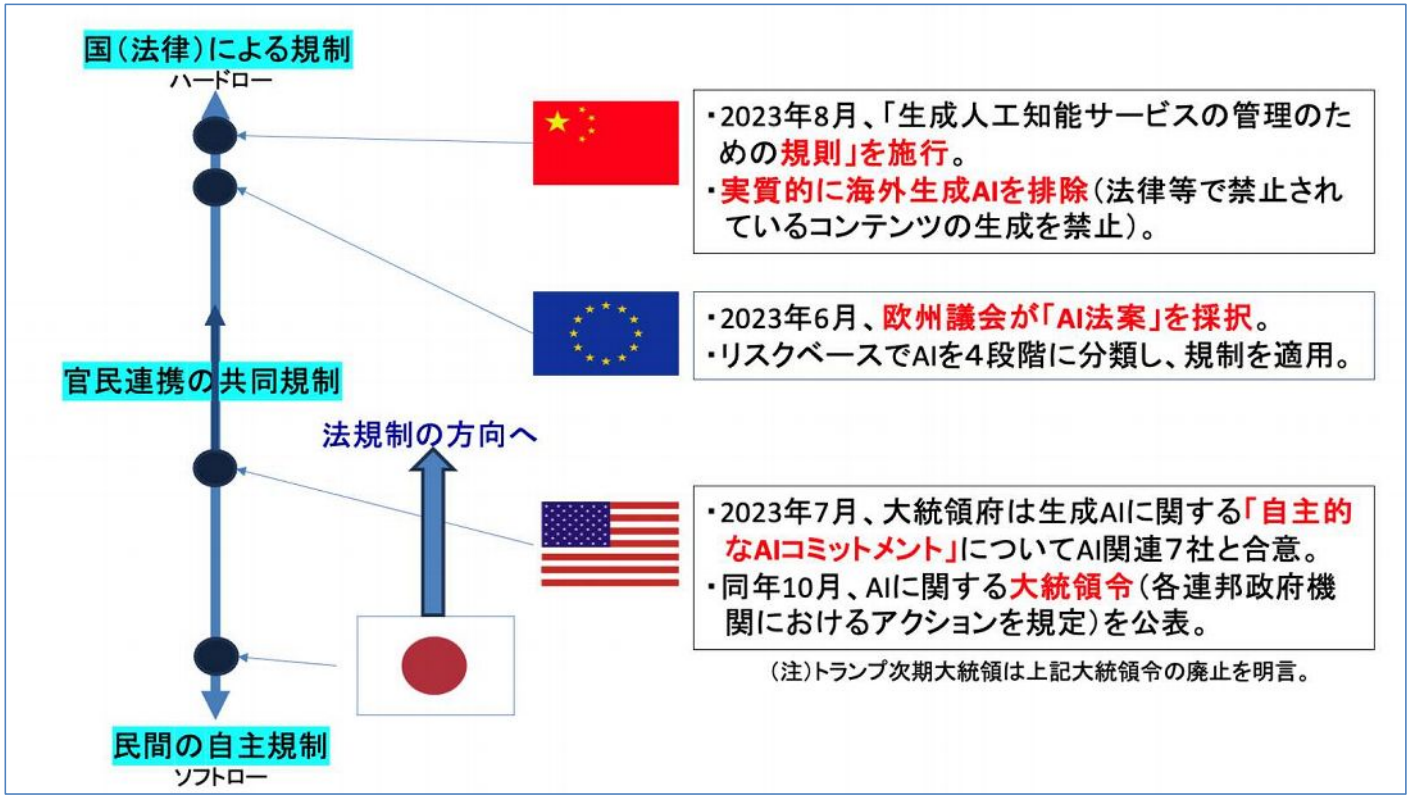
2024年12月

デジタル政策フォーラム

Digital Policy Forum Japan

本文書の目的

✔欧州や中国などで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）



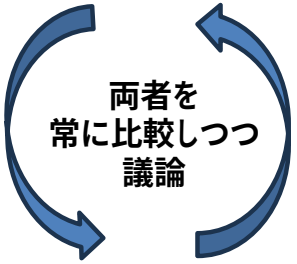
✔DPFJは2024年7月に公表した本文書(ver1.0)において論点整理。

（その後、有識者へのインタビュー、企業関係者からのヒアリング）

→ver2.0において検討の方向性（試案）を提示。

基本的考え方

1　リスクの最小化



人間による制御可能性が失われるリスクやAIが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、過度な規制の導入はイノベーション促進の観点から不適当）

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主権（data sovereignty）を技術的に担保しつつ、利便性の高いサービスを享受可能

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

(1)**リスク**管理のあり方
(2)**規制**のあり方と実効性の確保
(3)**外的リスクに対する脆弱性対策**
(4)**生成物**の取扱い

(5)AIの**積極的活用**

(6)健全な**エコシステム**の構築
(7)**産業振興とグローバル連携**
(8)国際的**コンセンサス**の醸成
(9)**倫理的問題**への対処

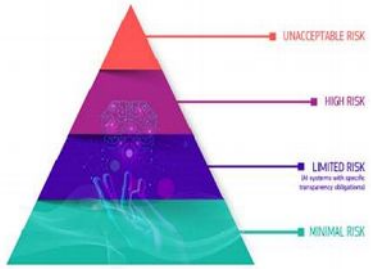
リスクの最小化(1/2)

(1)リスク管理のあり方

AIの段階別リスク管理の困難性

■**段階別リスク管理**は、リスクの管理手法、リスク判断の主体、第三者への説明責任等が未確立。
(AIのシステムログを基にリスクのスコアリング化等を検討)
■AIの抱える**リスク源が多様**(MIT調査では700項目超)で全容把握が困難。
■リスク管理そのものは重要。産学官連携による**AIリスクのレポジトリー**の作成・分析を積極的に推進。

欧州AI法におけるリスクベースアプローチ



許容できないリスク 禁止
(人の生命や基本的人権に対する直接的な脅威)
ハイリスク 規制 →事前の適合性評価・DBへの登録・EU適合宣言書の作成
(人の健康や安全、基本的人権、社会的・経済的利益に影響を与える可能性)
限定リスク 透明性の義務 →AIとのやりとりであることを利用者に知らせることが必要。
(深刻なリスクはないが、透明性確保のための要件を満たす必要)
最小リスク 規制なし →EUで利用されている大半のAIはこのカテゴリー

(出典) EU "Regulatory Framework Proposal on Artificial Intelligence" https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai

主体別のリスク管理

AIを実装 / サービス提供

AI開発者	AIサービス提供事業者	エンドユーザー
基本原則にとどめる(人権・民主主義・法の支配と十分な整合性を確保すること等) ※欧州評議会「AI条約」(2024年9月)を参照。	不当な差別的取扱いの禁止	誤認による被害防止

リスク管理の手法

AI開発者	AIサービス提供事業者	エンドユーザー
✔リスクの自己評価 ✔第三者評価(監査制度) →社会的に不可欠な重要サービスと密接に関連しているもの	✔各業法における利用者保護に係る既存の枠組みの中で検討・運用	✔AIに関するリテラシー教育の推進

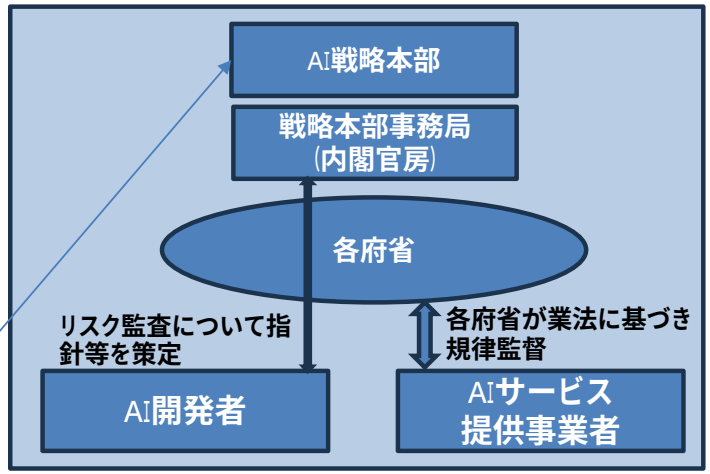
AI開発者と(AIを実装した)サービス提供事業者の責任分界点の明確化

リスクの最小化(2/2)

(2)規制のあり方と実効性の確保

AI基本法の制定

AI基本法と国の役割



(注) サイバーセキュリティ基本法を参照。

(3)外的リスクに対する脆弱性対策

AIに係るサイバー攻撃対策

■**脆弱性調査**(red teaming)**の運用ガイドライン**の策定(官民連携)
■**AIに対するサイバー攻撃・AIによるサイバー攻撃**に関する対処検討(オープン性の確保とAI悪用の可能性につき同時並行的に検討)

データ空間の健全性の確保

■**学習データの取扱ルール(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**
■AI生成物であることを示すラベリング(**電子透かし**)**の導入**
→OP(オリジネータープロファイル)の導入等についても併せて検討

利便性の向上

(5)AIの積極的活用

課題解決のためのAI活用の推進

■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防止**のための一定のセーフガード措置
■**環境対策、防災・減災、文化などの分野**でAIを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リスクに関する周知啓発活動の推進（**AIリテラシーの向上**）

行政サービスにおけるAI活用の推進

■行政サービスにおける**AI活用の制度的枠組みの整備**
（基本指針の策定、リスクアセスメントの実施等）
■**ベストプラクティスの共有**促進

AI活用と労働市場

■デジタル技術は既存領域の壁を打ち破り新しい市場領域を生み出すことで新たな雇用を生み出すもの。
■**AIを労働生産性の向上及び新たな市場領域を創出するツールとして活用**（政策支援を実施）

健全な市場の育成

(6)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の防止の仕組み検討
■域外適用の規定の妥当性の検証

(7)産業振興とグローバル連携

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
■技術仕様としてのオープン性と実効面でのオープン性の区別
（例：技術仕様はオープンだが学習データは非公開など）

産業としてのAI総合戦略の推進

■関連する先端性の高い技術開発、半導体の製造・流通、言語モデルの開発、データ流通のための環境整備、知財・著作権などの権利処理仕組み等、**経済安全保障の視点を含む俯瞰的なAI総合戦略の策定・推進**

(8)国際的コンセンサスの醸成

■**国内ルールと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の促進
■AIの軍事利用に関する規範形成
（例：REAIM Summit 軍事領域における責任あるAIに関する会議）

（**出典**）Yuval Abraham "Lavender': The AI machine directing Israel's bombing spree in Gaza" (April 3, 2024) +972 Magazine

(9)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
（例：「AIに自意識を持たせること」や「自己複製・改変能力を持たせること」の是非）

今後の作業計画

✔**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等**を開催。

✔**2025年夏を目処にver3.0への更新**を実施（予定）。

✔**他のフォーラム等との連携**を積極的に推進。