電卓とAIの相違から紐解く

国際AI規制の本質

2024年7月

羽深宏樹

京都大学法学研究科特任教授

スマートガバナンス株式会社代表取締役CEO

弁護士（日本・NY州）

自己紹介

羽深　宏樹　Hiroki Habuka

京都大学法学研究科 法政策共同研究センター 特任教授
スマートガバナンス株式会社 代表取締役CEO
弁護士（日本・NY州）

京都大学法学研究科特任教授として、複雑なサイバー・フィジカルシステムのガバナンス手法を研究する一方、スマートガバナンス株式会社CEOとして、企業に対して先端テクノロジーのガバナンスに関するアドバイスを提供している。2022年1月まで、経済産業省ガバナンス戦略国際調整官として、デジタルプラットフォーム規制やAI・データガバナンス政策等の検討等をリードした。2020年、世界経済フォーラムGlobal Future Council on Agile Governanceによって、「公共部門を変革する世界で最も影響力のある50人」に選出。主著に、『AIガバナンス入門：リスクマネジメントから社会設計まで』（2023年，ハヤカワ新書）。東京大学法学部・同法科大学院（JD）、及びスタンフォード・ロースクール（LLM）修了。

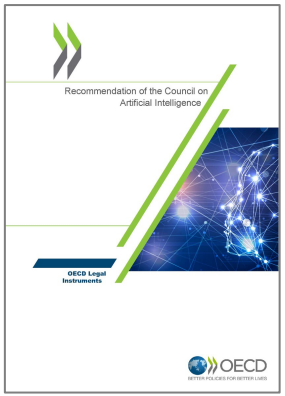
AIと人間の共生に立ちはだかる壁

SMART GOVERNANCE

- 米国の人気恋愛チャットボットサービスの圧倒的多数が、ユーザーのセンシティブな情報を第三者に販売・共有していることが発覚（2024.4）
- 香港の多国籍企業の会計担当者が、生成AIを用いて最高財務責任者（CFO）を装ったビデオ会議の相手に騙され、計2億香港ドル（約38億円）を詐取される（2024.2）
- 韓国の野菜包装工場において、ロボットが人間を野菜の箱と間違え死亡させる（2023.11）
- GM傘下の自動運転サービスCruiseが、カリフォルニア州での営業許可取消（2023.10）
- 全米摂食障害協会の提供していたチャットボットが、相談者に有害なアドバイスを提供したため、使用停止に（2023.6）

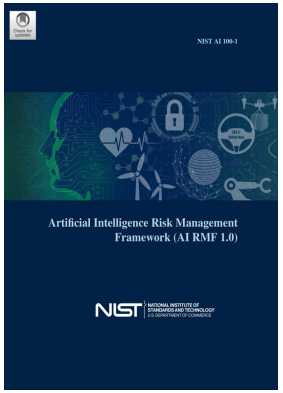
「入水」した警備ロボット（ワシントンDC、2017年）
https://www.bbc.com/news/technology-40642968

AIガバナンスはグローバルなトレンドに

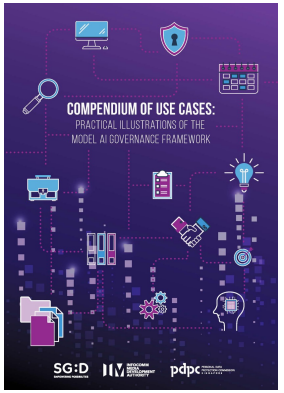


OECD Principles

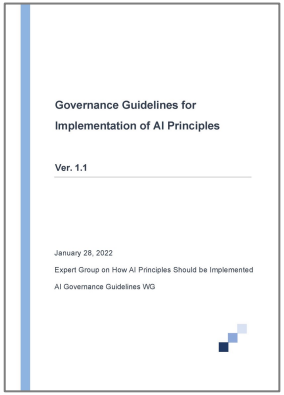
EU AI Act



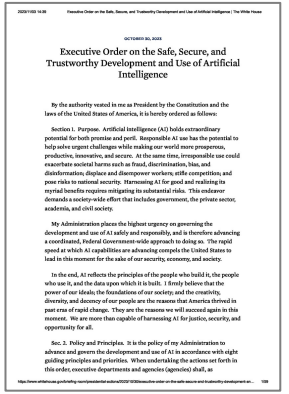
US NIST RMF



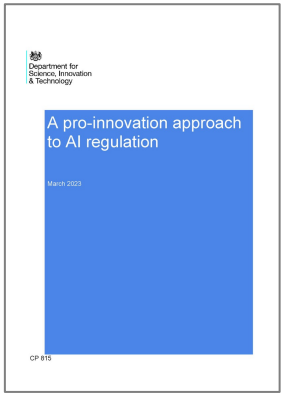
Singapore MAGF



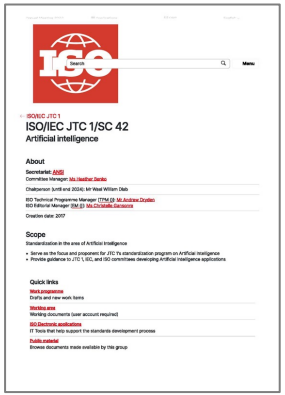
Japan AI Guidelines



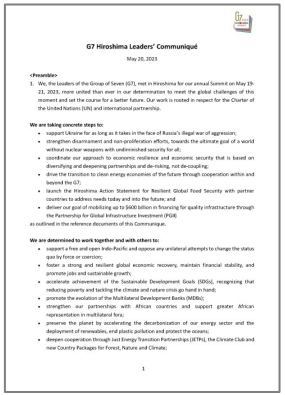
US Executive Order



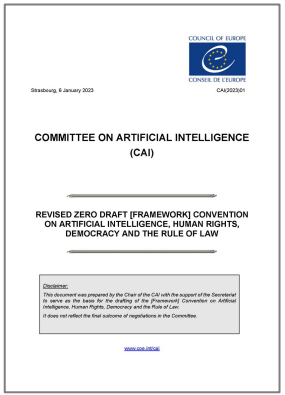
UK Approach



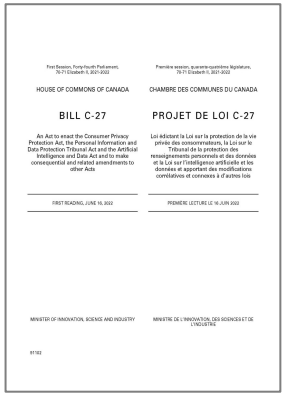
ISO/SC42



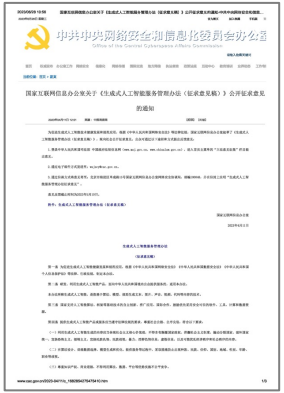
G7 Communique



Council of Europe
AI Treaty



Canada AIDA



China Gen AI Service Act

2024年4～5月のAIガバナンス関連ニュース

- 日本、AI戦略会議「AI 制度に関する考え方」について（2024.5.22）
- EU、AI法を欧州理事会で最終可決（2024.5.21）
- 日本、知的財産戦略本部「AI時代の知的財産権検討会中間とりまとめ」（2024.5.21）
- AI Safety Summit 、AI安全性に関する閣僚宣言 (2024.5.21-22)
- 欧州評議会、AI条約採択（2024.5.17）
- OECD、AI原則のアップデート（2024.5.3）
- G7、広島AIフレンズグループの立ち上げ（2024.5.2）
- 米国、NIST（国立標準技術研究所）が生成AIに関するリスクマネジメントフレームワークのドラフト公表（2024.4.29）
- 日本、自民党「AIホワイトペーパー2024」（2024.4.11）

日本だけで見ても・・・

法令

- 個人情報保護法
- 著作権法
- 各種業法（道交法、金商法、薬機法、割賦販売法、高圧ガス保安法 etc…）
- 独占禁止法　・不正競争防止法
- 民法・刑法
- 責任あるAI推進基本法（仮）

ガイドライン類

- AI事業者ガイドライン(経産省・総務省)
- 人間中心のAI社会原則（統合イノベーション戦略推進会議）
- 機械学習品質マネジメントガイドライン（産総研）
- 倫理指針（人工知能学会）

ホワイトペーパー/政策文書

- AIホワイトペーパー2024（自民党）
- AI 制度に関する考え方(AI戦略会議)
- 「AI時代の知的財産権検討会中間とりまとめ」(知的財産戦略本部)
- AI と著作権に関する考え方(文化庁)

国際関係

- EU AI Act等の域外適用
- G7広島AIプロセス
- 欧州評議会によるAI条約
- AI Safety Summit

思考のガイド

Q. AIは複雑な（深い）関数を用いてデータを統計的に分析し、与えられた命令に対して最も確率の高い答えを出す計算機です。皆さんのお手元（スマートフォン）に搭載された電卓も計算機です。なぜ電卓に対する規制やリスクマネジメントは問題にされないのに、AIに対する規制やリスクマネジメントは問題になるのでしょうか。

人間＋電卓：人間が演算の内容を決定し、機械に指示する。

機械学習：入力と出力をつなぐアルゴリズムを、機械が自律的に構築（学習）する

思考のガイド

Q. AIは複雑な（深い）関数を用いてデータを統計的に分析し、与えられた命令に対して最も確率の高い答えを出す計算機です。皆さんのお手元（スマートフォン）に搭載された電卓も計算機です。なぜ電卓に対する規制やリスクマネジメントは問題にされないのに、AIに対する規制やリスクマネジメントは問題になるのでしょうか。

人間＋電卓：人間が演算の内容を決定し、機械に指示する。

機械学習：入力と出力をつなぐアルゴリズムを、機械が自律的に構築（学習）する

＜「人間＋電卓」とAIの共通点＞

- 所与のデータに対して統計と確率（人間の感覚によるもの含む）を用いて最適解を出す
- 出力結果の利用場面は様々：ゼロリスクから生命・基本的人権へのリスクまで

→ 実際に、「AIリスク」とされるものの多くは「人間のリスク」でもある

  - 交通事故、フェイクニュース、人種/マイノリティ差別、見間違い　etc.

- したがって、今あるリスクマネジメント（内部統制）の仕組みは相当程度有効

思考のガイド（続）

＜「人間＋電卓」とAIの相違点＞

- ディープラーニング：階層が「深い」ために、演算が極めて複雑になる
  - 入力と出力の因果関係の説明が困難（ブラックボックス性）
  - アウトプットの予測が困難（予見困難性）
  - アウトプットの原因を説明することが困難（説明困難性）
- アウトプットまでに多くの主体が関与
  - データの提供者、基盤モデルの開発者、基盤モデルを活用したサービスの提供者、サービスの利用者等
  - クラウド提供者、通信サービス提供者、OS提供者、プラットフォーム事業者等
- 技術革新・普及の速さ
- 人間による信頼判断の難しさ
- クロスボーダー性

○×？

- 日本はAIについて、事業者の自主的取組を促す「ソフトロー」のアプローチを採っており、 AIの開発や利用について法的な義務を課す「ハードロー」は存在しない。

- 日本政府が公表した「AI事業者ガイドライン」はソフトローであり、これを実践する法的義務はない。

- EUのAI法は、世界で初めてAIの開発や使用を規制する法律である。

- AIを包括的（分野横断的）に規制する動きは世界中に広がりつつあり、日本政府でも、AIに対する包括的な規制の検討が始まっている。

- 法律の整備及びガイドラインや標準の開発が進むと、AIガバナンスは、創意工夫の領域から画一的なコンプライアンスの領域に移行する。

○×？

- 日本はAIについて、事業者の自主的取組を促す「ソフトロー」のアプローチを採っており、 AIの開発や利用について法的な義務を課す「ハードロー」は存在しない。

  ✕ **自動運転車に関する道交法、AI医療機器に関する薬機法等、日本にも多くのハードローが存在。**

- 日本政府が公表した「AI事業者ガイドライン」はソフトローであり、これを実践する法的義務はない。

- EUのAI法は、世界で初めてAIの開発や使用を規制する法律である。

- AIを包括的（分野横断的）に規制する動きは世界中に広がりつつあり、日本政府でも、AIに対する包括的な規制の検討が始まっている。

- 法律の整備及びガイドラインや標準の開発が進むと、AIガバナンスは、創意工夫の領域から画一的なコンプライアンスの領域に移行する。

○×？

- 日本はAIについて、事業者の自主的取組を促す「ソフトロー」のアプローチを採っており、 AIの開発や利用について法的な義務を課す「ハードロー」は存在しない。

❌ **自動運転車に関する道交法、AI医療機器に関する薬機法等、日本にも多くのハードローが存在。**

- 日本政府が公表した「AI事業者ガイドライン」はソフトローであり、これを実践する法的義務はない。

⭕ **正しい。ガイドラインはあくまで参照資料にすぎない。**

- EUのAI法は、世界で初めてAIの開発や使用を規制する法律である。

- AIを包括的（分野横断的）に規制する動きは世界中に広がりつつあり、日本政府でも、AIに対する包括的な規制の検討が始まっている。

- 法律の整備及びガイドラインや標準の開発が進むと、AIガバナンスは、創意工夫の領域から画一的なコンプライアンスの領域に移行する。

○×？

- 日本はAIについて、事業者の自主的取組を促す「ソフトロー」のアプローチを採っており、 AIの開発や利用について法的な義務を課す「ハードロー」は存在しない。

✕　自動運転車に関する道交法、AI医療機器に関する薬機法等、日本にも多くのハードローが存在。

- 日本政府が公表した「AI事業者ガイドライン」はソフトローであり、これを実践する法的義務はない。

○　正しい。ガイドラインはあくまで参照資料にすぎない。

- EUのAI法は、世界で初めてAIの開発や使用を規制する法律である。

✕　AIを「分野別」に規制する法律は各国に存在。EU AI法の特殊性は、その「包括性」にある。

- AIを包括的（分野横断的）に規制する動きは世界中に広がりつつあり、日本政府でも、AIに対する包括的な規制の検討が始まっている。

- 法律の整備及びガイドラインや標準の開発が進むと、AIガバナンスは、創意工夫の領域から画一的なコンプライアンスの領域に移行する。

○×？

- 日本はAIについて、事業者の自主的取組を促す「ソフトロー」のアプローチを採っており、 AIの開発や利用について法的な義務を課す「ハードロー」は存在しない。

× 自動運転車に関する道交法、AI医療機器に関する薬機法等、日本にも多くのハードローが存在。

- 日本政府が公表した「AI事業者ガイドライン」はソフトローであり、これを実践する法的義務はない。

○ 正しい。ガイドラインはあくまで参照資料にすぎない。

- EUのAI法は、世界で初めてAIの開発や使用を規制する法律である。

× AIを「分野別」に規制する法律は各国に存在。EU AI法の特殊性は、その「包括性」にある。

- AIを包括的（分野横断的）に規制する動きは世界中に広がりつつあり、日本政府でも、AIに対する包括的な規制の検討が始まっている。

× AIを包括的に規制するのは、現状EU AI法のみ。日本を含む諸国は分野別のアプローチが主流。

- 法律の整備及びガイドラインや標準の開発が進むと、AIガバナンスは、創意工夫の領域から画一的なコンプライアンスの領域に移行する。

○×？

- 日本はAIについて、事業者の自主的取組を促す「ソフトロー」のアプローチを採っており、 AIの開発や利用について法的な義務を課す「ハードロー」は存在しない。

  ✕ **自動運転車に関する道交法、AI医療機器に関する薬機法等、日本にも多くのハードローが存在。**

- 日本政府が公表した「AI事業者ガイドライン」はソフトローであり、これを実践する法的義務はない。

  ○ **正しい。ガイドラインはあくまで参照資料にすぎない。**

- EUのAI法は、世界で初めてAIの開発や使用を規制する法律である。

  ✕ **AIを「分野別」に規制する法律は各国に存在。EU AI法の特殊性は、その「包括性」にある。**

- AIを包括的（分野横断的）に規制する動きは世界中に広がりつつあり、日本政府でも、AIに対する包括的な規制の検討が始まっている。

  ✕ **AIを包括的に規制するのは、現状EU AI法のみ。日本を含む諸国は分野別のアプローチが主流。**

- 法律の整備及びガイドラインや標準の開発が進むと、AIガバナンスは、創意工夫の領域から画一的なコンプライアンスの領域に移行する。

  ✕ **AIの技術変化の速さや適用場面の広さを踏まえると、画一的な規範設定は不可能。事業者の創意工夫の余地がなくなることはない。**

目次

AIガバナンスの俯瞰図

AIリスク

1. 技術的リスク

(1) 誤判定
(2) バイアス
(3) 虚偽・ハルシネーション
(4) 安全性
(5) セキュリティ

2. 社会的リスク

(1) プライバシー
(2) 民主主義へのリスク
(3) 不正目的・攻撃目的利用
(4) 経済への影響（独占・仕事の代替）
(5) 財産権への影響（知財・データ）
(6) 環境負荷

3. リスクの本質

(1) 予測や説明の難しさ
(2) バリューチェーンの主体の多さ
(3) 技術革新や普及の速さ
(4) 信頼性判断の難しさ
(5) 倫理的課題の提起
(6) グローバル化
(7) 汎用AIがもたらす未知の影響

AIガバナンスの目的

・基本的人権　・民主主義
・経済成長・サステナビリティ

AI原則・AI倫理

安全性	セキュリティ	プライバシー	公平性	透明性・説明可能性	アカウンタビリティ
有効性					

AIシステムのガバナンス

AI社会のガバナンス

法規制	標準/ガイダンス	財産権	責任・制裁	救済
アジャイル・マルチステークホルダー・分散的なプロセス				

AIガバナンスの俯瞰図

AIリスク

1. 技術的リスク
(1) 誤判定
(2) バイアス
(3) 虚偽・ハルシネーション
(4) 安全性
(5) セキュリティ

2. 社会的リスク
(1) プライバシー
(2) 民主主義へのリスク
(3) 不正目的・攻撃目的利用
(4) 経済への影響（独占・仕事の代替）
(5) 財産権への影響（知財・データ）
(6) 環境負荷

3. リスクの本質
(1) 予測や説明の難しさ
(2) バリューチェーンの主体の多さ
(3) 技術革新や普及の速さ
(4) 信頼性判断の難しさ
(5) 倫理的課題の提起
(6) グローバル化
(7) 汎用AIがもたらす未知の影響

大枠で合意できている領域

AIガバナンスの目的

・基本的人権　・民主主義
・経済成長・サステナビリティ

AI原則・AI倫理

安全性	セキュリティ	プライバシー	公平性	透明性・説明可能性	アカウンタビリティ
有効性					

AI事業者ガイドライン

AIシステムのガバナンス

AI開発者　AI提供者　AI利用者

各種業法 / 責任あるAI推進法？

AI社会のガバナンス

著作権・契約
民刑法

法規制	標準/ガイダンス	財産権	責任・制裁	救済

アジャイル・マルチステークホルダー・分散的なプロセス

目次

AIに対する規制

- AIの使用を制約する**包括的な規制は存在しない**
- **スマホソフトウェア競争促進法**は、**検索結果の表示**において、自社のサービスを他社のサービ スよりも優先的に取り扱うことを原則として禁止
- **デジタルプラットフォーム透明化法**は、大規模なオンラインモール、アプリストア、デジタル広告事業者に対して、**検索ランキングを決定する主要な要素の開示**等を要求
- **金融商品取引法**は、**アルゴリズム高速取引**を行う事業者に対して、政府への登録と、リスク管理システムの整備や、取引記録の維持を要求
- プライバシーとの関係では**個人情報保護法**が、生成AIの文脈では**著作権法**が、ハードウェアと一体化した製品については**製造物責任法**が問題となる。

AIを促進するための規制

分野別

- **道路交通法**改正（2023）により、都道府県公安委員会の許可を受けた事業者によるレベル4の自動運転（特定条件下での完全自動運転）が可能に。
- **割賦販売法**改正（2020）により、認定包括信用購入あっせん業者が、データやAIを使用して与信額を決定することが可能に。
- **高圧ガス保安法**改正により、AIとドローンを活用した先進的な安全技術を持つスーパー認定事業者が、最長8年間、運転を中断せずに安全検査を行うことが可能に。
- **著作権法**改正（2017）により、AIモデルを開発するための著作権保護コンテンツの利用が直ちに著作権侵害とならないことが明確に。
- **不正競争防止法**改正（2019）により、有料データセット等に対する保護が強化。

規制全般

- **デジタル行政改革会議（旧デジタル臨調）**において、**約1万**の法令・規則・通達等におけるアナログ条項（書面義務、目視義務、定期検査、事務所常駐義務等）を撤廃し、テクノロジーによるコンプライアンスを可能とする一括法改正を実施。

大規模な基盤モデルに対する規制：「責任あるAI推進基本法（仮）」

- 自民党のAIの進化と実装に関するPTによる提案（2024.2）

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、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

12

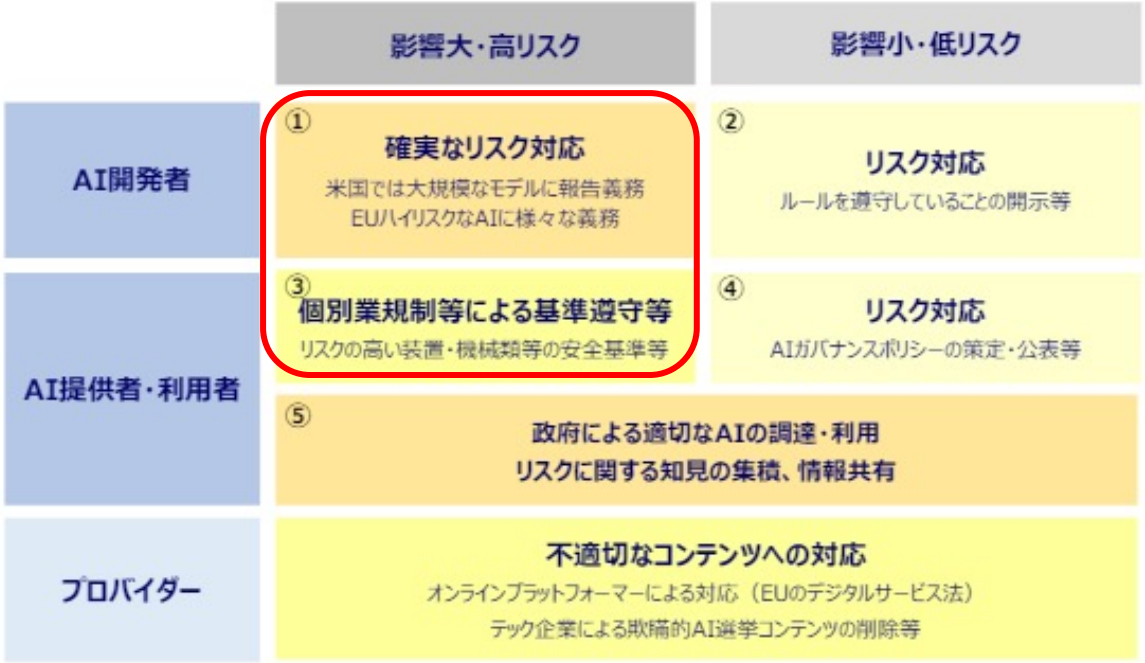
大規模な基盤モデルに対する規制：「責任あるAI推進基本法（仮）」

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	–

13

AI戦略会議「AI 制度に関する考え方」の主要ポイント

（参考）AI関係者を巡る制度検討のイメージ



https://www8.cao.go.jp/cstp/ai/ai_senryaku/9kai/shiryo2-2.pdf

（参考）文化庁「AIと著作権に関する考え方について」の主要ポイント

1．AI開発・学習段階

- 著作物に表現された思想又は感情の「享受を目的としない利用」は、原則として著作権者の許諾なく行うことが可能。ただし、「著作権者の利益を不当に害することとなる場合を除く」。（著作権法30条の4）
- 学習データである著作物の類似物を生成させること目的とした場合は、「享受目的」ありとされる可能性。
- AI学習のための著作物の複製等を防止する技術的措置が講じられている場合、これを複製する行為は「著作権者の利益を不当に害する」可能性あり。

2．生成・利用段階

- 「類似性」及び「依拠性」から判断（両方ともあれば、著作権侵害の可能性あり）
- 「依拠性」をどのような場合に認めるかについては、様々な見解がある。

> ❑ 権利者としては「AI利用者が既存の著作物にアクセス可能であったこと」や「生成物に既存の著作物との高度な類似性があること」等を立証すれば、依拠性ありと推認させることができる
>
> ❑ 生成AIの開発・学習段階で当該既存の著作物が学習されていた場合は、AI利用者が既存の著作物を認識していない場合でも、通常、依拠性があったと推認される

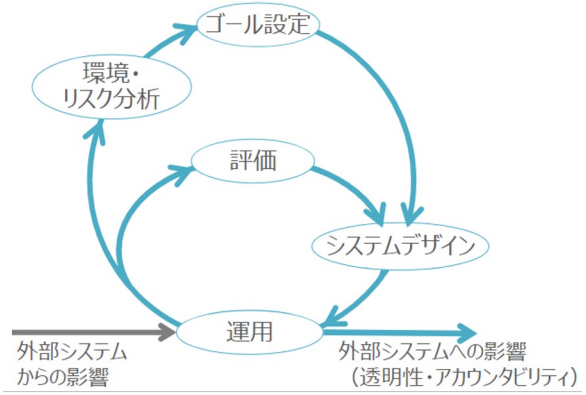
3．AI生成物は「著作物」にあたるか

- 人の「創作意図」があるか、及び人が「創作的寄与」と認められる行為を行ったかによって判断。
- 「創作的寄与」となり得るものがどの程度積み重なっているか等を総合的に考慮して判断

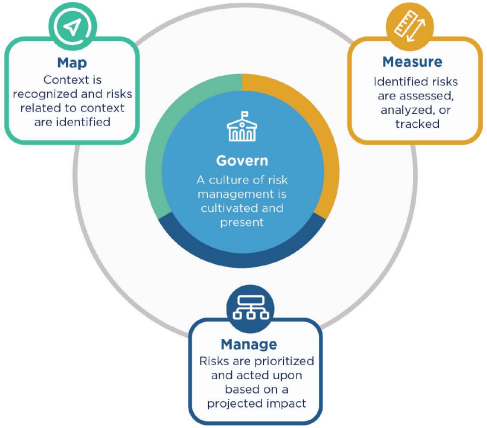
目次

AIシステムのガバナンス：基本的な考え方

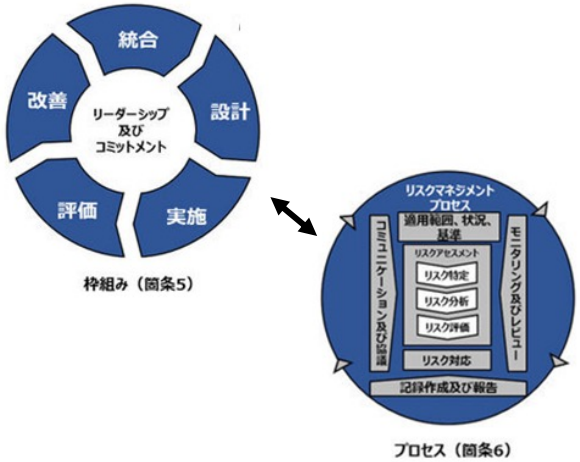
- 組織内でのAIガバナンスについて、世界中から様々なドキュメントが発行されている（その多くは拘束力を持たないソフトロー）
- ポイントとなるのは、経営層レベル＋現場レベルの「二重のループ」



経産省・総務省「AI事業者ガイドライン」



NIST AI Risk Management Framework



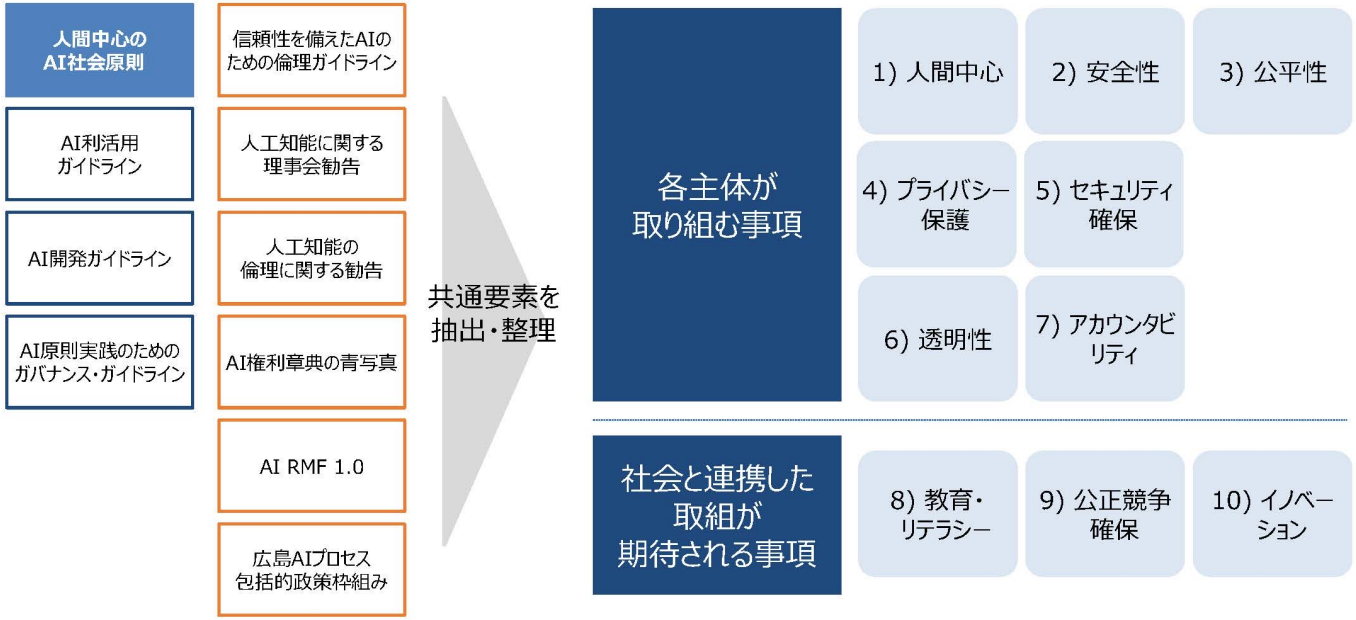
ISO 31000: 2018

AI事業者ガイドラインの構成

https://www.meti.go.jp/shingikai/mono_info_service/ai_shakai_jisso/pdf/20240119_3.pdf

各主体に共通の指針

- AIの活用による目指すべき社会の実現のために各主体が連携して取り組む内容を原則としてまとめた上で、「共通の指針」として整理する
- 「共通の指針」は、「人間中心のAI社会原則」を土台としつつ、諸外国における議論状況や、新技術の台頭に伴い生じるリスクへの対応等を反映している
- その結果、各主体が取り組む事項と、社会と連携して取り組むことが期待される事項に分類される



https://www.meti.go.jp/shingikai/mono_info_service/ai_shakai_jisso/pdf/20240119_3.pdf

ケーススタディ

製薬企業であるA社は、法令及び業界自主規制により求められる開示書類の作成に膨大なコストをかけている。今般の生成AIの飛躍的進化を前に、同社は、これらの開示資料作成に生成AIを活用することにした。あなたがA社の経営者だったとして、どのようなガバナンス（ルール・組織・技術の設計・運用）を行うべきだろうか？

Step 1 既製品の社内利用

Step 2 ファインチューニングモデルの社内利用

Step 3 ファインチューニングモデルの外部ライセンス

Step 1 既製品の社内利用

リスクの性質

- 人間によるドラフトのサポートツールとして使う限り、劇的なリスク状況の変化はない
- 但し、情報漏洩や知的財産侵害等には注意が必要

ポリシーの策定

- 生成AIのリスクについてどのような姿勢で臨むのか
- 自社のビジョンやミッションの達成に生成AIがどのように貢献するのか

組織的対応

- 記載内容の信頼性について→開示担当部門/システム開発部門
- 個人情報保護について→法務部門
- 著作権について→知財部門

[参照] 日本ディープラーニング協会『生成AIの利用ガイドライン』

教育

- 利用する従業員向けマニュアルの作成（※　細かすぎては読まれない）
- 社内試験に合格した場合にのみアクセスを許可する　等

Step 2 ファインチューニングモデルの社内利用

開発段階で生じるリスク

- 開発段階での新たな論点・・・個人情報保護法/著作権法/不正競争防止法上等
  - 学習データをどのように取得するか?データにどのような加工を行うのか?そのデータをどのような方法でどこに移転して学習させるのか? といった点の理解が不可欠に
  - 法務・コンプラ部門だけでは解決できず、事業部門・技術部門との連携が必要

新たなステークホルダーの登場

- 開発委託先：何を遵守してもらうか、どのように遵守してもらうか
  - 法令遵守/データセキュリティ/保守/開発したモデルの転用制限 etc.
- 基盤モデル開発者の規約

経営層のフィードバックサイクル例

- 各部門の連携体制の整備　　・ユーザーマニュアル等の整備
- 継続的なリスク評価体制の整備

現場でのフィードバックサイクル例

- 各部門間での円滑なコミュニケーションの実施と専門性の発揮

Step 3 ファインチューニングモデルの外部ライセンス

リスクの飛躍的増大

- 「ユーザーが何をするか分からない」というリスクを抱えることに
  - ユーザーによる法令違反のリスク
  - ユーザーによる不適切利用のリスク

考えられる対応方法の多様化・高度化

- 不適切な出力にならないような技術的な対応
- モデルの内部監査＋外部監査
- 保険

経営層のフィードバックサイクル例

- 「生成AI統括組織」の立ち上げ（技術・品質・法務・コンプラ・モニタリング・ステークホルダーコミュニケーション等の機能の集中）
- 外部のリスク環境に関する情報収集体制の確立

現場でのフィードバックサイクル例

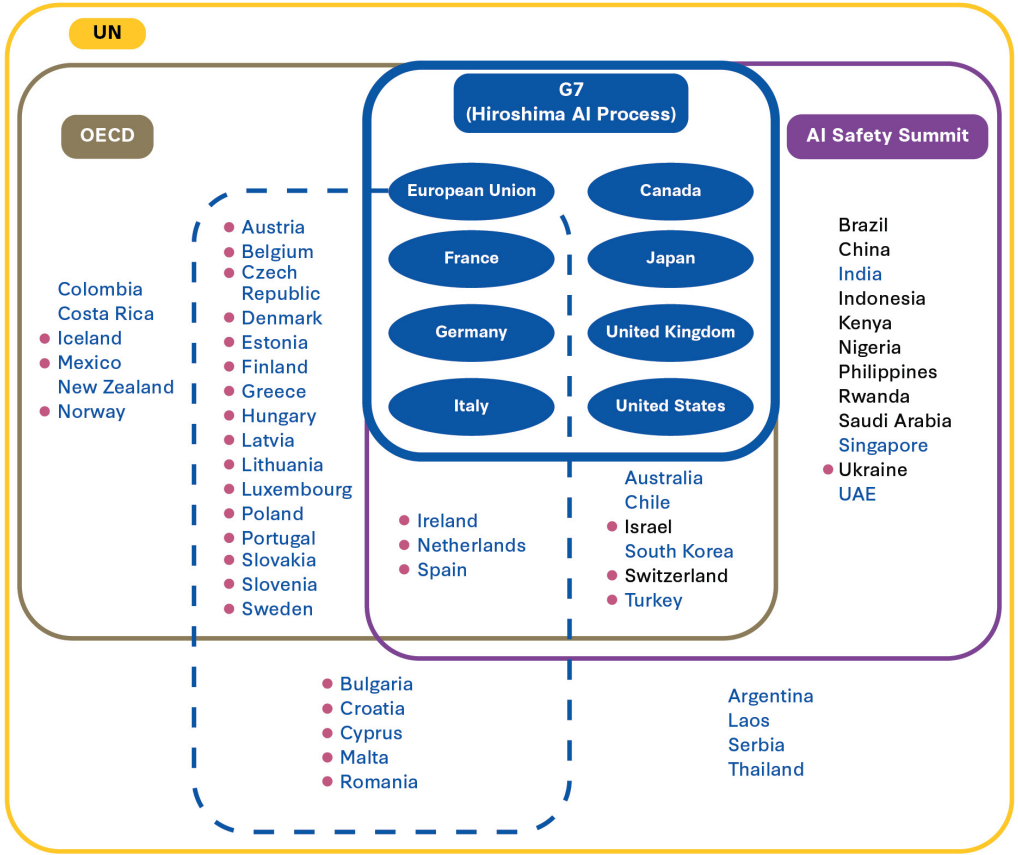
- 各リスクに対応する評価や、対応（技術的措置か、契約での責任移転か、利用料を上げてリスクを吸収するのか等）に関する継続的な評価とコミュニケーションの実施

AIシステムのガバナンスのポイント

- 初手はリスクを「特定」すること
  - AIリスクは、従来のシステムのリスクの延長上にあることも多い
  - AI原則（公平性・安全性・透明性・プライバシー等）を参照し、具体的なリスクシナリオを抽出する
- 次に、リスクを「評価」する
  - 一般には「損害×確率」と定義されるが・・・
    - 数値化が困難なリスクも多い（プライバシー、差別、民主主義 etc.）
    - リスク発現確率も、テストし続けないと分からない（レッドチーミング）
  - ステークホルダーにとってのリスクの受容度も評価が必要
- リスクへの「対応」を決定する
  - 技術的対応（ガードレイル）、組織的対応、契約による対応、保険等
- ステークホルダー（政府・民間双方）へのアカウンタビリティを尽くす
  - 開示を受ける者にとって適切な質と量の情報提供
- 上記のプロセスをサービス提供開始後も回し続ける

目次

AIガバナンスに関する国際協力体制の概観



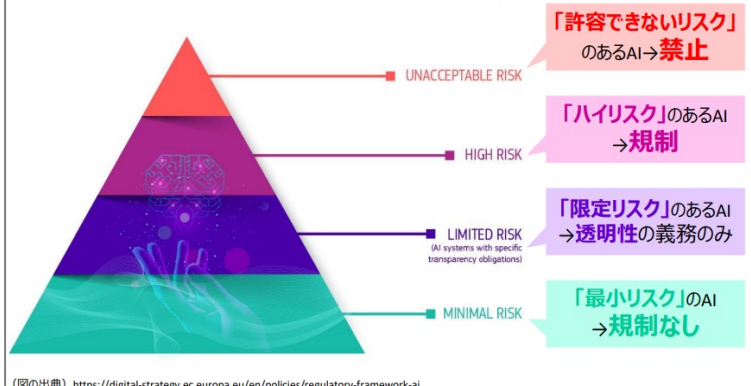
- 青字は広島AIプロセスフレンズグループを表す
- ●は「AI条約」の母体である欧州評議会の加盟国orオブザーバ国
- G7は、この表に挙げた全てのイニシアチブに参加している
- G7が民主的国家のリーダーであるのに対し、AI Safety Summitには民主的でない国も含まれる

EU AI Act

- AIに関する包括的な規制法案”AI Act”は2024年5月に欧州理事会で最終可決。
- AIを、「禁止」「ハイリスク」「中間的リスク」「低リスク」の4段階にランク付けする。
- ハイリスクAIの開発については、リスクマネジメントシステムの導入、サービス提供前の適合性評価やデータベースへの登録等が求められる。
- さらに、「汎用目的AI」への規制を追加。技術文書の保持や下流プロバイダへの情報提供といった義務に加え、システミックなリスクを持つモデルには、モデル評価の実施、リスク評価とリスク緩和措置の取り組み、適切なレベルのサイバーセキュリティ保護の確保、重大なインシデントのAIオフィスおよび各国の所管機関への報告が追加の義務として課される。

ハイリスクAIの例（付属書III）

- 生体データによる自然人の認証・分類（顔画像認証など）
- 重要インフラ（道路、水・ガス・電気の供給など）の維持・管理
- 教育・職業訓練へのアクセスの決定（入学・採用試験の採点など）
- 雇用、従業員管理等へのアクセス（採用手続において履歴書を自動仕分けするソフトウェアなど）
- 必要不可欠な私的・公的サービス・便益へのアクセス（ローン審査における信用スコアリングなど）
- 基本的権利に干渉する法執行（証拠の信用性の評価など）
- 移民、難民、出入国管理（旅行書類の真正性の検証など）
- 司法・民主的プロセスの運営（具体的事実に対する法の当てはめ等）



（図の出典）https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai

米国 (1) NIST AI Risk Management Framework

- 産業界自らによる責任あるAIに向けた取組の推進を進め、そのリスクに係る公的評価を進めつつ、政府でのAI利用に係るガイドラインの策定に取り組む。
- 2023年1月、NIST（国立標準技術研究所）が、 AIリスクマネジメントフレームワーク（AI RMF）を公表。 2024年4月には、同フレームワークを生成AIに適用した「生成AIプロファイル」を公表

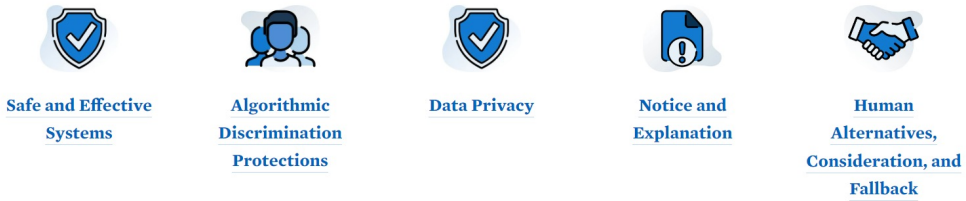
Part 1: AIのライフサイクル

Part 2: 企業が取り組むべきガバナンス

https://www.nist.gov/itl/ai-risk-management-framework

米国 (2) 事業者の自主的コミットメントと大統領令

- ホワイトハウスは、2022年に安全性、差別回避、セキュリティ、通知と説明、人間による判断へのアクセス確保を柱とする「AI権利章典のための青写真」を公表。
- 2023年7月、Amazon, Alphabet（Google）, Anthropic, Inflection, Meta, Microsoftが、AIの安全性・セキュリティ・信頼を確保するための自主的なコミットメントを公表。9月には、更に8社が加わる。
- 10月30日には、ホワイトハウスから上記のコミットメントに紐づく大統領令が発出。
  - 安全性とセキュリティに力点が置かれる。開発者等に対する報告義務の基準作成など。
  - コンテンツ認証や電子透かしに関してもガイダンスを作成する予定。
  - 立法ではなく大統領令なので、既存の制度の枠組みで対応できる内容が基本。
  - プライバシー法などについては、議会の行動を求めているが、どこまで実現できるかには疑問もあり。
- 他方、州・市レベルでは、複数の州における警察によるAIカメラ使用の禁止や、NY市の採用AIに関する規制、コロラド州におけるハイリスクAI規制などの動きがみられる。



英国

- 2023年3月29日、科学イノベーション技術省は、「AI規制：プロイノベーションアプローチ」の報告書を発表。包括規制ではなくセクターごとの規制とする方針を示す。
- 10月には、「フロンティア AI の能力とリスク」を公表し、今後生じ得るリスクを分析。
- 11月、「AI安全性サミット」を開催。29か国が「ブレッチリー宣言」に合意。
- AIガバナンスのガイドラインだけでなく、アルゴリズムの透明性のプロセスや、AIアシュアランスの枠組み、フロンティアモデルのリスク分析など、実践的なガイダンスやツールを矢継ぎ早に公表。
- 総じて、まずはリスクを見極めたうえで必要に応じた規制を行っていく方針といえ、日本としても大いに参考となる。

生成AIに関する広島AIプロセスで合意された原則一覧

リスクマネジメント	トラスト形成	倫理的・社会的配慮
1. リスクの特定、評価、低減 2. 脆弱性、インシデント、悪用パターンの特定、低減 5. ガバナンスとリスク管理ポリシーの開発、実践、開示 6. セキュリティ管理措置への投資 7. コンテンツ認証・証明 11. 個人情報及び知的財産の保護	3. 透明性とアカウンタビリティ 4.　責任ある情報共有とインシデント報告 12. 高度なAIシステムの信頼でき責任ある利用	8. 社会、安全、セキュリティ上のリスクの低減のための研究 9. 気候危機、健康・教育などのグローバルな課題の優先 10. 国際的な技術標準の開発と採用

まとめ

- AI規制の本質は、「安全性」や「公平性」といった原則それ自体にあるわけではなく、予見不可能性や説明不可能性、変化の速さ、リスクシナリオの多さ、関連主体の多さ、クロスボーダー性等にある
- これらの点に対応するためのルール形成は、製作中の溶けたガラスのような状態であり、いきなり詳細に飛びつくのではなく、はじめに全体の骨格を抑えることがポイント
- 政府や業界団体が一律のルールを作ることには様々な限界があり、現に国内外の関連する法律は極めて抽象的である
- したがって、企業自らが、組織、ルール、技術、プロセスを組み合わせてアカウンタビリティを尽くす必要
- アカウンタビリティの対象は、AIアルゴリズム、それを組み込んだシステム、そのシステムを運用する現場のルール、モニタリングメカニズム、これらを統括する組織、その組織における権限分配、そして組織カルチャーなど、多岐にわたる
- 組織としてこれらに対応するためには、トップダウンの硬直的ガバナンスではなく、倫理的カルチャーと心理的安全性に基づくアジャイルなガバナンスが必要

AIガバナンス協会

- AI実務リードする60社超が参加する、AIガバナンス団体として国内随一の規模
- 官民・国内外と連携して、AIに関する政策提言やツール提供を行う（会員随時受付中）

書籍のご紹介

「AIガバナンス入門 — リスクマネジメントから社会設計まで」

- **冨山和彦氏推薦**

  コンパクトだが、本格派である。AIの時代のガバナンスの本質かつ急所を突いた、必読の本だ。

- **読売新聞書評**

  どのような立場であれ我々はＡＩのあり方を自分ごととして考えなければならないと著者は説く。本書はわかりやすくＡＩガバナンスの全体像を示す。

本書は、世界で話題沸騰しているAIガバナンスについて、何が問題なのか、企業は、国は、そして個人はどう対応すべきなのかを網羅的に解説する。溢れる情報に踊らされず、冷静に質の高い意思決定をするための一冊。

ご清聴ありがとうございました!

- 我々は、大海原を航行中に船を作り直さなければならない船乗りのようなものだ。
  — オットー・ノイラート『アンチ・シュペングラー』(1921)

- 永久の未完成、これ完成である。
  — 宮沢賢治『農民芸術概論綱要』（1926）

- 幸せは歩いてこない、だから歩いていくんだね
  — 水前寺清子「三百六十五歩のマーチ」（1968）