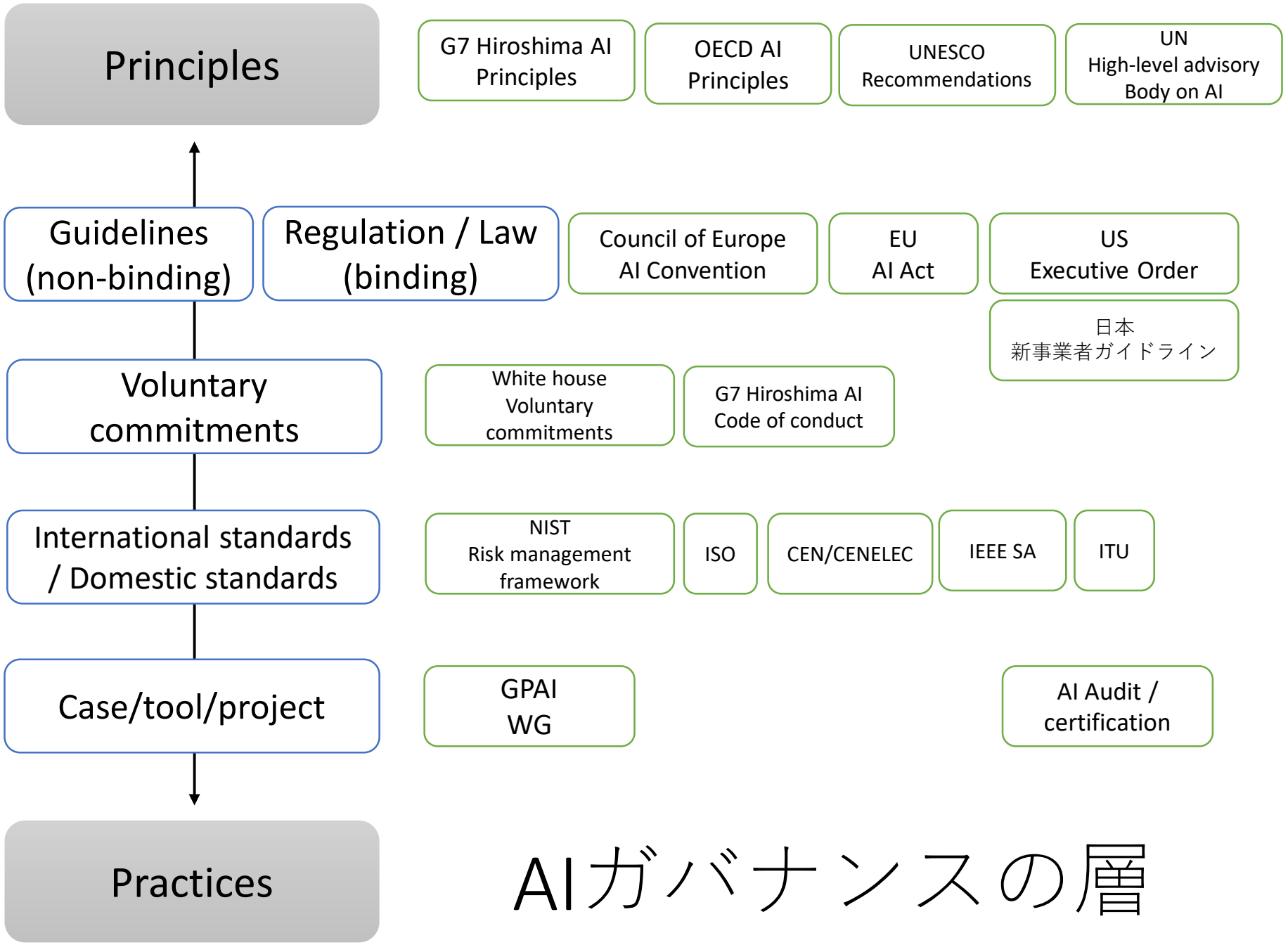
資料３－１

AIガバナンスの Principles and Practices

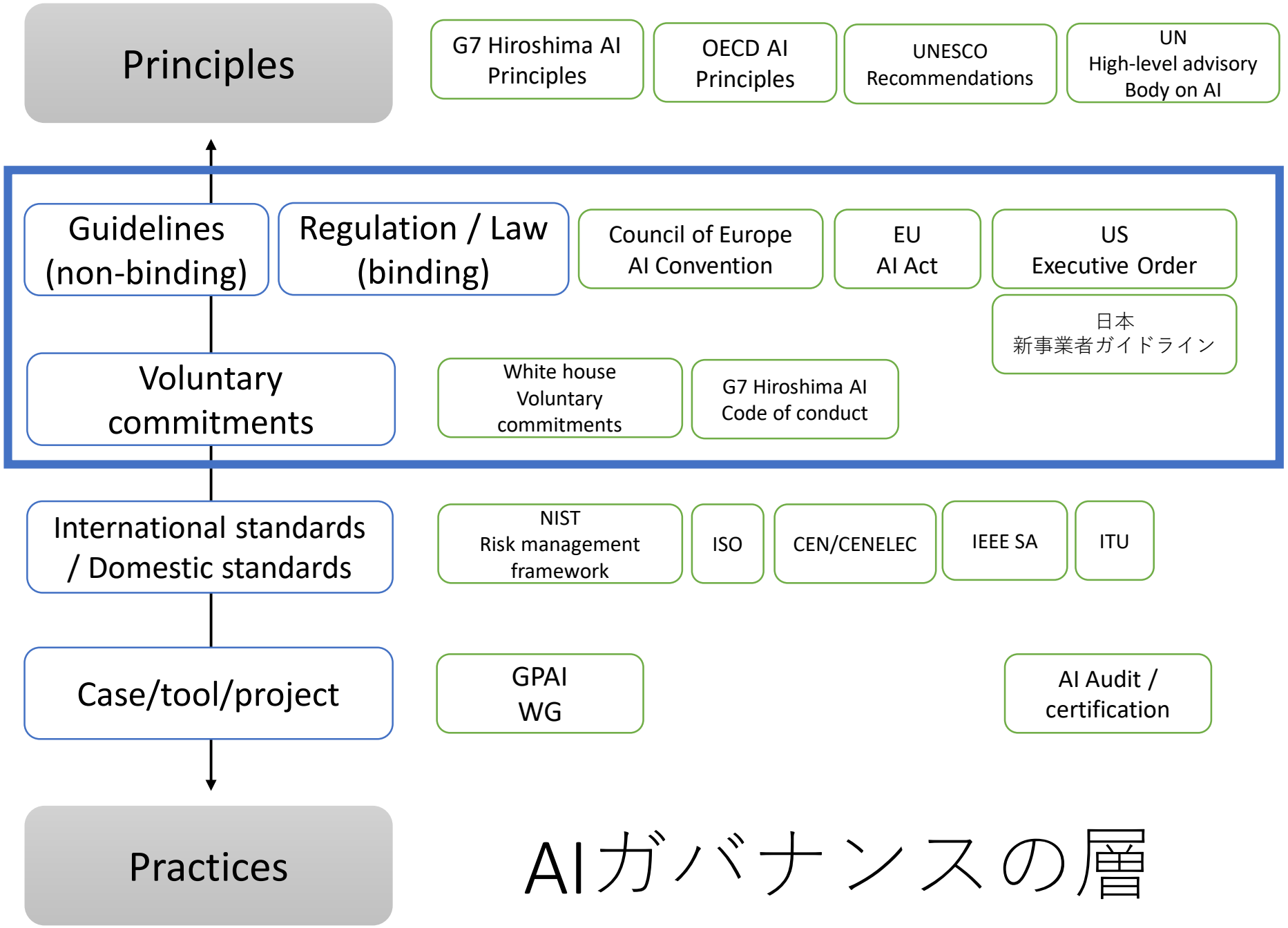
東京大学
江間有沙

デジタル空間における情報流通の健全性

- Principle
  - **どういう状態**が「健全」かの議論と合意
- Practice
  - 「健全性」を**誰**が**どのように**担保しているかを**どのように確認するか**の方法論の議論


Principles
G7 Hiroshima AI Principles
OECD AI Principles
UNESCO Recommendations
UN High-level advisory Body on AI
Guidelines (non-binding)
Regulation / Law (binding)
Council of Europe AI Convention
EU AI Act
US Executive Order
日本 新事業者ガイドライン
Voluntary commitments
White house Voluntary commitments
G7 Hiroshima AI Code of conduct
International standards / Domestic standards
NIST Risk management framework
ISO
CEN/CENELEC
IEEE SA
ITU
Case/tool/project
GPAI WG
AI Audit / certification
Practices

AIガバナンスの層


Principles
G7 Hiroshima AI Principles
OECD AI Principles
UNESCO Recommendations
UN High-level advisory Body on AI
Guidelines (non-binding)
Regulation / Law (binding)
Council of Europe AI Convention
EU AI Act
US Executive Order
日本 新事業者ガイドライン
Voluntary commitments
White house Voluntary commitments
G7 Hiroshima AI Code of conduct
International standards / Domestic standards
NIST Risk management framework
ISO
CEN/CENELEC
IEEE SA
ITU
Case/tool/project
GPAI WG
AI Audit / certification
Practices

AIガバナンスの層

What

- **Fairness**
- Accountability
- Transparency
- Trust
- Explainability
- Sustainability
- Safety & Security
- Privacy
- etc

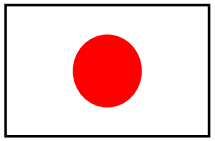
How

- Multi-Stakeholder
- Diversity & Inclusion
- International / Regional

Who

- Academic communities
- Industry Associations
- NPOs/NGOs
- Nation / Region
- International organizations

M: 56.9 %
F : 42.1 %



M: 86.2 %
F : 13.8 %

Male-female ratio of AI principle committee members

Principles

Practices

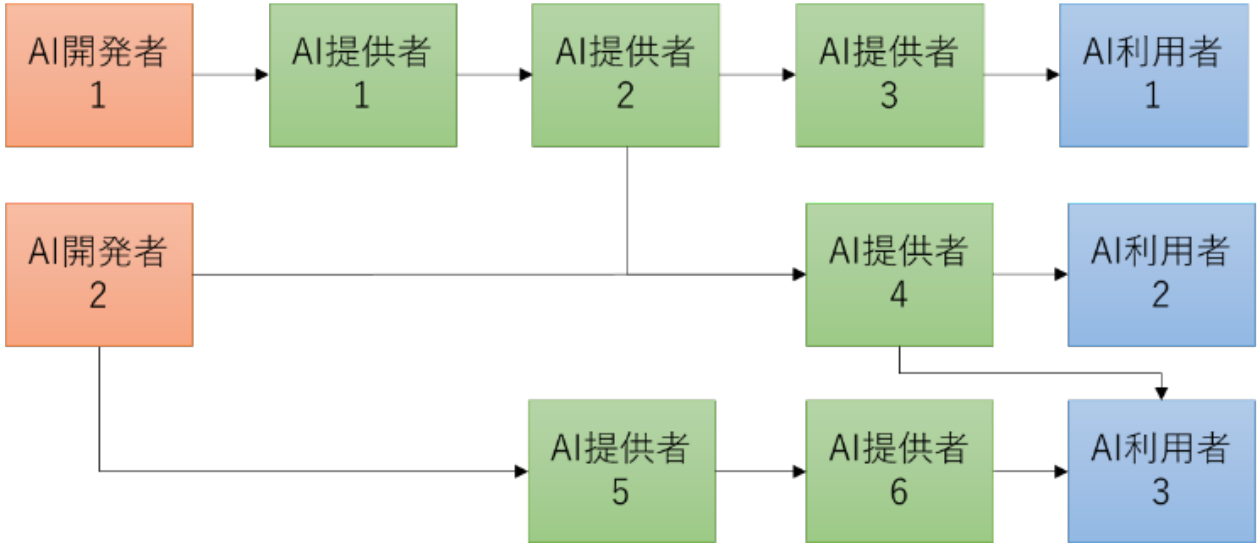
Focus on areas
- Medicine/ health
- Mobility
- Law Enforcment
- Finance

Focus on regions
- Global South

組織や国をまたぐAI開発・提供

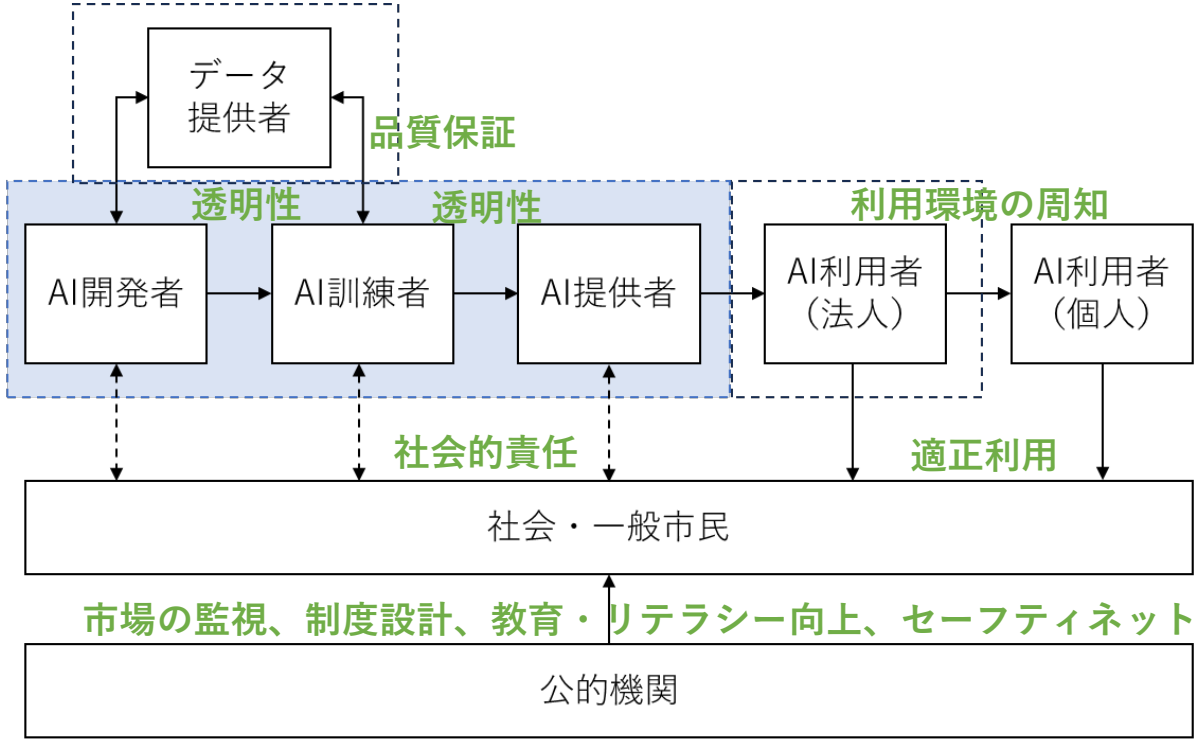
(1) AI開発から提供まで同一組織が担う例

(2) AI開発から提供までが異なる組織で担われている複雑な例



東京大学未来ビジョン研究センター政策提言「AIの責任ある展開に向けて：広島AIプロセスへの政策提言」https://ifi.u-tokyo.ac.jp/news/16642/

AIシステム関係者と責任



AI開発者・AI学習者・AI提供者は同一組織の場合もあれば、別々の組織のあ場合もる。データ提供者とAI開発者が同一組織の場合もある。自社製品を利用する場合は、AI開発者/AI提供者と利用者（法人）が同一組織となる場合もありうる。

責任の向きを示す。例えばAI開発者はAI学訓練とデータ提供者に対して責任を負う（表3）。

企業の社会的責任や、社会からのフィードバック等の間接的な責任の向きを示す。

■事業者間（青枠内）

- **事業者間**で契約や約束を取り交わし、リスク対応や責任を取る
- **公的機関**は契約の公平性等を監視する

■事業者ー消費者間

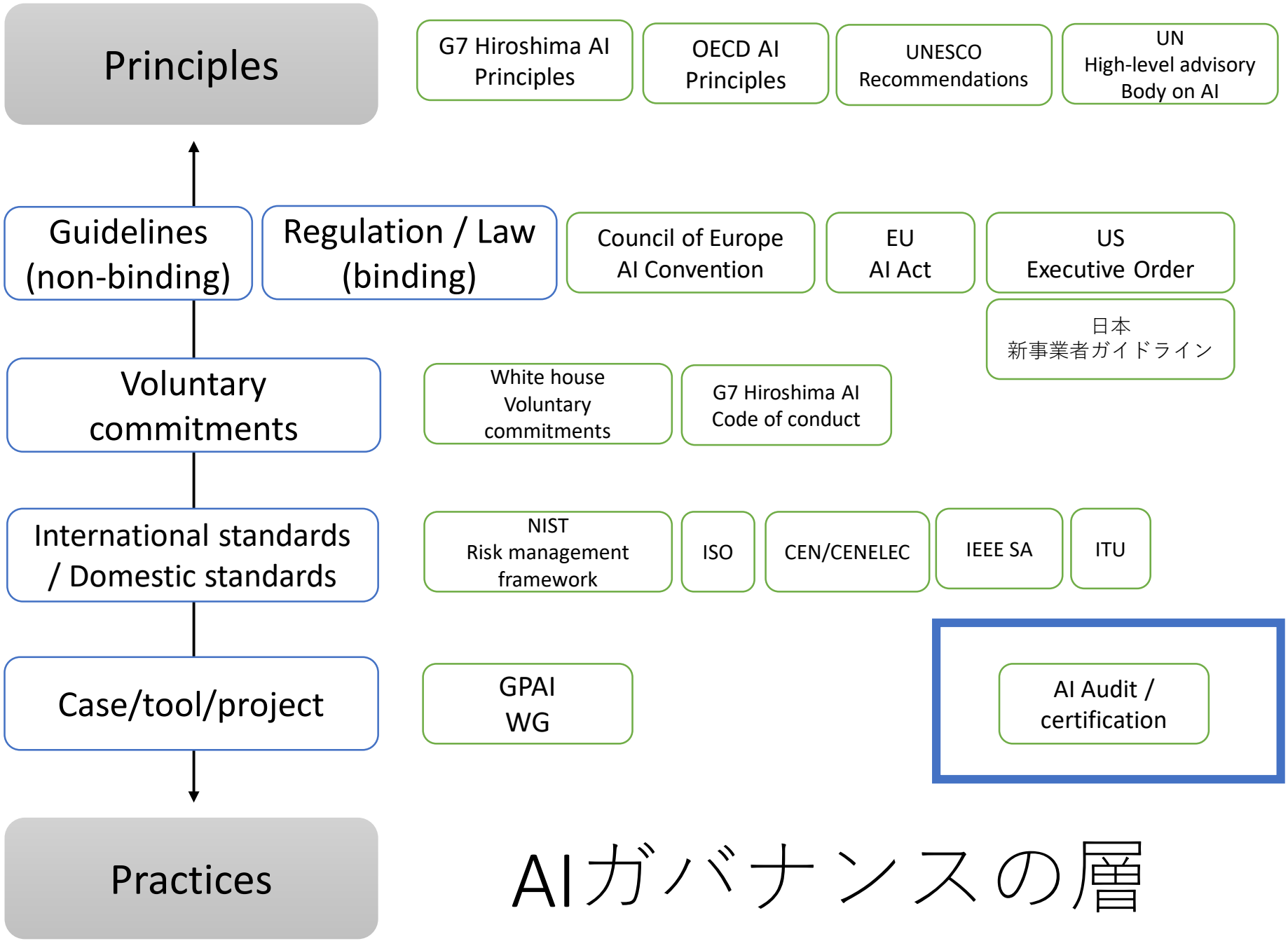
- **AI提供者**が事前・事後対応を行う
- **利用者**も適正利用する
- **公的機関**も事件・事故時の原因究明や被害救済の仕組等を形成する

東京大学未来ビジョン研究センター政策提言「AIの責任ある展開に向けて：広島AIプロセスへの政策提言」https://ifi.u-tokyo.ac.jp/news/16642/

様々な規律の在り方

		国家によるエンフォース（強制的な執行）の有無	
		国家がエンフォースする	国家はエンフォースしない
規律を形成する主体	国家	法令・一部ガイドライン	ガイドライン・行政指導
	産業界・個別企業	法令を背景とする標準化（強制規格）	業界ガイドライン・社内ポリシー・業界標準
	それ以外（市民団体・学術団体等）	慣習法	市場・投資・モラル・規範・学会基準・慣習・評判

東京大学未来ビジョン研究センター政策提言「AIの責任ある展開に向けて：広島AIプロセスへの政策提言」https://ifi.u-tokyo.ac.jp/news/16642/


Principles
G7 Hiroshima AI Principles
OECD AI Principles
UNESCO Recommendations
UN High-level advisory Body on AI
Guidelines (non-binding)
Regulation / Law (binding)
Council of Europe AI Convention
EU AI Act
US Executive Order
日本 新事業者ガイドライン
Voluntary commitments
White house Voluntary commitments
G7 Hiroshima AI Code of conduct
International standards / Domestic standards
NIST Risk management framework
ISO
CEN/CENELEC
IEEE SA
ITU
Case/tool/project
GPAI WG
AI Audit / certification
Practices

AIガバナンスの層

2. AI監査をめぐる論点 – 監査の立証命題②

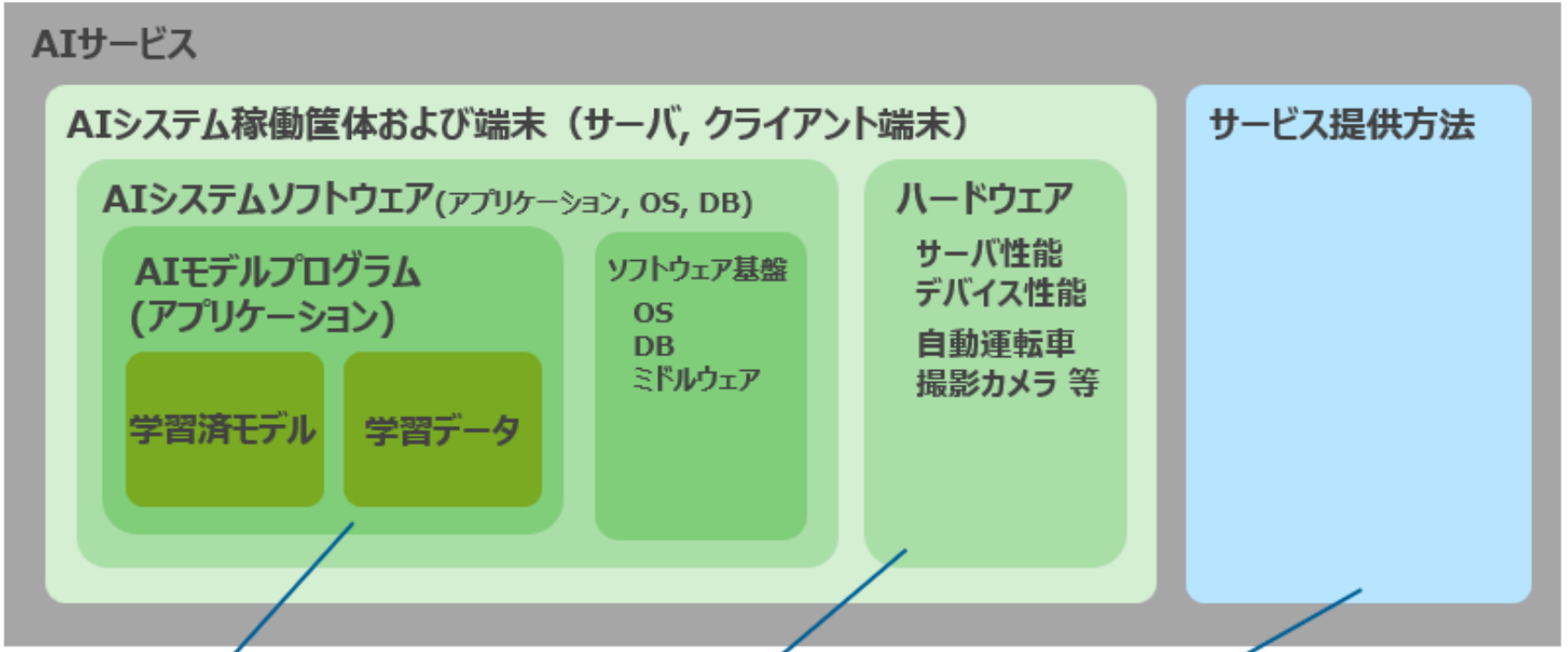
AI監査に特徴的な立証命題（一例）

立証命題	解説
公平性	• AIシステムの出力結果に不適切なバイアスがかかっていないか　等 • 公平の定義についても予め共通認識を持つ必要がある
透明性	• AIシステムの出力結果について再現ができるか、学習データや採用されている特徴量（パラメータ）について説明が可能か　等
安全性	• AIシステムが利用者に危害を加える可能性はないか、不具合が発生した場合に適切に停止状態に移行するか　等 • AIシステムが組み込まれているハードウェアについても考慮する必要がある
セキュリティ	• 学習データに対する攻撃を予防・発見できるか、意図的に不適切な出力を誘導するような本番入力データを予防できるか　等
プライバシー	• 個人が共有を望まない属性データを拒否できるか、誤った個人評価を適時適切に訂正できるか　等

- それぞれの立証命題ごとに実施する監査手続やクライテリアは大きく異なる。そのため、対象となる立証命題について共通認識を持った上で議論、監査を実施する必要がある
- 既存の基準/規準のみでは全ての立証命題のカバーは困難

東京大学未来ビジョン研究センター政策提言「AIガバナンスに資するAI監査の実践に向けて」https://ifi.u-tokyo.ac.jp/news/16814/

2. AI監査をめぐる論点 – AI監査の対象②



東京大学未来ビジョン研究センター政策提言「AIガバナンスに資するAI監査の実践に向けて」https://ifi.u-tokyo.ac.jp/news/16814/

AI監査をめぐる論点 – AI監査の対象③

AIサービス、システムを提供する組織 – 経営層、業務ユーザ部門、開発・運用部門 等

体制、方針

各種規程類
ルール

日々の業務
ルールの遵守

全社対象の広範な統制活動
AI利活用方針、ポリシー
教育体制　等

個別業務での統制活動
リリース前のテスト実施、承認
学習データの内容妥当性確認　等

各種統制活動の整備、運用

COSO, COBIT等(*)のフレームワーク

(*)COSO：トレッドウェイ委員会組織委員会（Committee of Sponsoring Organizations of the Treadway Commission）が提唱する内部統制フレームワーク
COBIT： ISACA, ITGIが提唱する事業体全体を対象とした事業体の情報と技術のガバナンスとマネジメントのためのフレームワーク

東京大学未来ビジョン研究センター政策提言「AIガバナンスに資するAI監査の実践に向けて」https://ifi.u-tokyo.ac.jp/news/16814/

2. AI監査をめぐる論点 – AI監査のタイミング

AIライフサイクル

AI監査のタイミングをめぐる論点

継続学習による精度の更新
- 監査を実施した時点と監査結果を利用する時点でAI出力結果の精度が異なるケースが存在

監査対象や監査実施主体によるタイミングの違い
- 対象（AIサービス/システム自体と内部統制）、実施主体（内部監査/外部監査）により それぞれタイミングが異なる
- AIサービスそのものの妥当性、AIシステム開発要否の妥当性についても重要な論点であるため、 リリース前や企画段階での監査もより重視して検討される可能性がある

東京大学未来ビジョン研究センター政策提言「AIガバナンスに資するAI監査の実践に向けて」https://ifi.u-tokyo.ac.jp/news/16814/

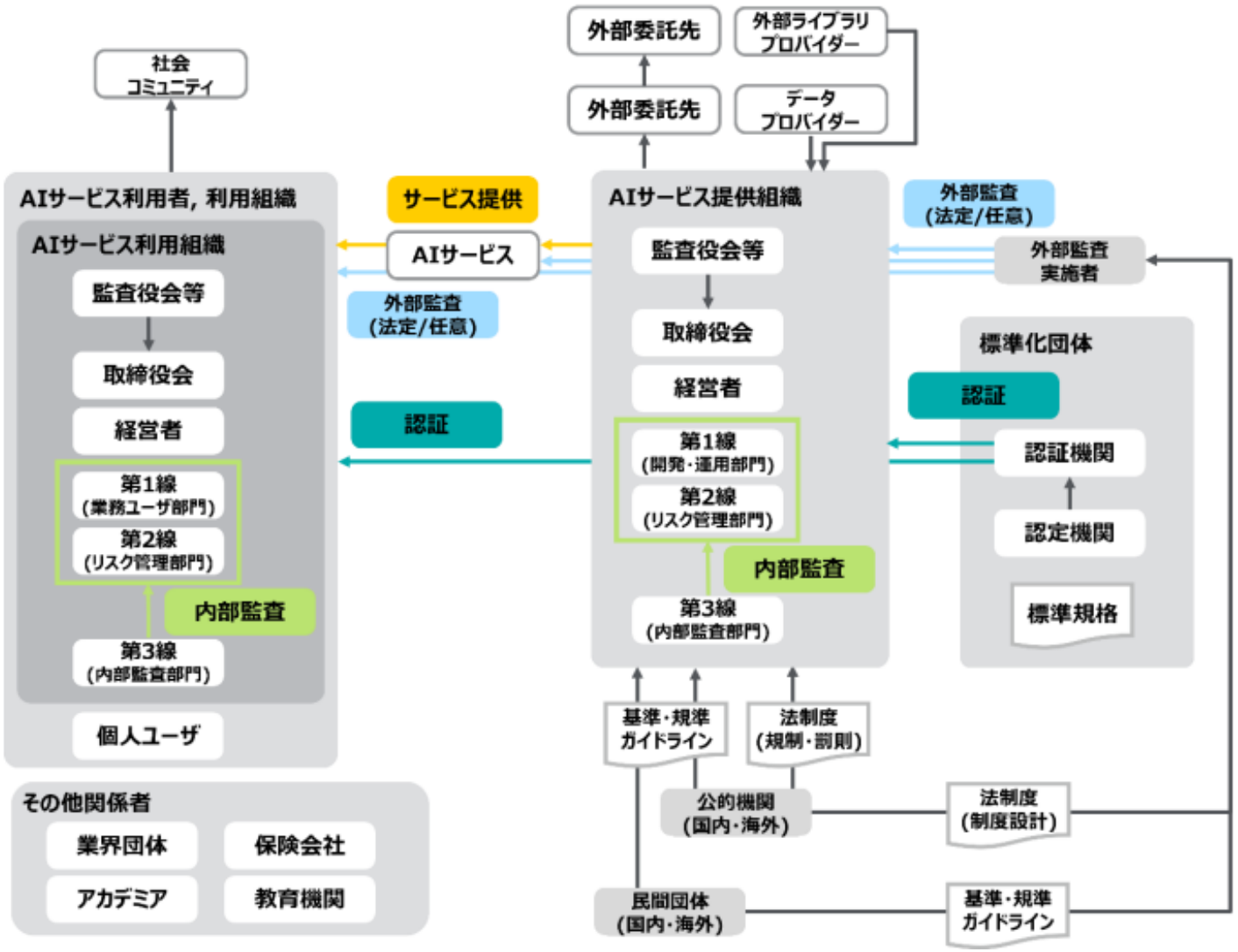
2. AI監査をめぐる論点 – AI監査の実施者要件

	専門性要件	独立性要件
内部監査 外部監査	• 従来システムの監査に必要な監査論の理解、業界知識およびIT領域の知識、経験 • AI固有の技術的知見や法規制等の知識。新たな資格制度の必要性も要検討 • 個人で全てをカバーするのは現実的には困難	• 従来の監査と同様、被監査企業や部門と利害関係のない独立性を確保する必要がある • 独立性要件は外部監査、内部監査それぞれにおいて定義

	組織要件	監査人の法的責任
外部監査	• 監査品質や実施者の独立性等の観点で一定の基準を満たす組織 • 実施組織についての認定やモニタリングの必要性 • 組織要件を満たさない監査実施機関が実施した結果は実態が正しく反映されていない可能性有	• 正当な注意を払いつつも監査結果が誤っていたケースにおける法的な責任範囲 • 責任が重すぎると監査人不足の可能性有 • 監査人を守る免責要件や保険制度の必要性

東京大学未来ビジョン研究センター政策提言「AIガバナンスに資するAI監査の実践に向けて」https://ifi.u-tokyo.ac.jp/news/16814/

AI監査の関係者と関係組織



東京大学未来ビジョン研究センター政策提言「AIガバナンスに資するAI監査の実践に向けて」https://ifi.u-tokyo.ac.jp/news/16814/

3. AI監査は何故難しいのか

- AI監査制度設計の未整備
- 被監査対象の範囲に起因する複雑性
- AI監査の需要と供給のアンバランス
- AI監査の実施基準設定の困難性
- AI技術の複雑性

東京大学未来ビジョン研究センター政策提言「AIガバナンスに資するAI監査の実践に向けて」https://ifi.u-tokyo.ac.jp/news/16814/

補足
生成AIと監査をめぐる論点と課題

- AI監査の立証命題
  - 認識AIや予測AI：公平性や透明性等が重点
  - 生成AIや対話AI：著作権や偽情報・誤情報、個人の尊厳、感情操作、人と機械の関係性
    - これらを適切に評価する規準開発も未熟

- AI監査の対象
  - 対話型生成AIというインタフェース
    - 利用にあたっての情報の扱いを適切に説明しているか等も監査対象になりうる
    - 入力された支持を制御するロジック
  - 生成物に対する適切な表示の有無

デジタル空間における情報流通の健全性

- Principle
  - **どういう状態**が「健全」かの議論と合意
  - 様々な層におけるAIガバナンスの議論と国内外の議論の協調をどのように進めていくか

- Practice
  - 「健全性」を**誰**が**どのように**担保しているかを**どのように確認するか**の方法論の議論

- 動きの速い技術はPrinciples to practicesではなく、Principles and Practicesになっている