

AI関連政策の海外動向

2023年3月

事務局

目次

■ 米国

① AI権利章典

■ 英国

② AI規制政策文書

③ 効果的なAI保証エコシステムに向けたロードマップ

④ AIスタンダードハブ

■ シンガポール

⑤ モデルAIガバナンスの枠組み

⑥ A.I. Verify

■ EU

⑦-1 AI責任指令案

⑦-2 製造部責任指令の改正案

① AI権利章典（米国）

- 米国科学技術政策局（OSTP）は、「AI権利章典の青写真」を2022年10月に発表した。
- AIシステムの設計、使用、導入の際の指針となる5つの原則を提示した。
- 非実用的であったり、偏見や差別につながる有害なAIが存在する中、**国民が様々なサービスやリソースに公平にアクセスできる機会均等の権利の保障**や、**あらゆる差別・プライバシー侵害からの保護**等を目的に策定された。
- 5つの原則は下記のとおり（後述）。各原則について、**それぞれの重要性や当該原則に期待すること、原則の実践に向けた取組みの事例**が整理されている。
 - ①安全かつ有効なシステム
 - ②アルゴリズムから生じる差別からの保護
 - ③データのプライバシー
 - ④告知と説明
 - ⑤問題発生時の人間による代替、検討、対応
- 一部有識者からは、当該章典は、法的拘束力を有していないため、AIシステムを開発する企業等に対して実効性に欠けるとの指摘もある。

① AI権利章典（米国）

- 5つの原則の概要と、実践に向けた取組みの事例（一部抜粋）は下表のとおり。

原則	概要	実践に向けた取組みの事例
①安全かつ有効なシステム	• AIシステム開発の際には、多様なコミュニティや各分野の有識者と協議し、あらゆるリスクや潜在的な影響を特定、それらの軽減措置を講じること。 • 当該システムが安全かつ有効であるという評価・報告を第三者機関から得た上で、それらを可能な限り公表すること。	• DOEやDoD等、複数の政府機関でAIシステムを倫理的に使用するためのフレームワークを策定。 • NSFでは、Cyber Physical Systems programをはじめ、複数のプログラムで当該原則を遵守、かつ促進させるAIシステムの研究開発に資金提供を実施している。
②アルゴリズムから生じる差別からの保護	• アルゴリズムが生み出す人種、肌の色、民族、性別、その他あらゆる属性に基づく差別から国民を保護するため、AIシステムの設計者、開発者、ユーザーは積極的かつ継続的な対策を講じること。 • これらの保護の確証を得るため、格差試験の導入・影響評価の実施や、第三者機関による組織的な監査を実施した上で、その結果を公表すること。	• 汎用的な医療アルゴリズムに対する格差評価の研究で、黒人患者の医療アクセスに対するバイアスを特定の上、是正に向けた方法論を提案。 • 雇用機会均等委員会とDOJは、求職者や障害のある従業員に対する採用/人事関連アルゴリズムから生じる差別を可視化し、障害者関連法案を遵守する実践的な手段等を提供。

① AI権利章典（米国）

- 5つの原則の概要と、実践に向けた取組みの事例（一部抜粋）は下表のとおり。（前頁の続き）

原則	概要	実践に向けた取組みの事例
③データのプライバシー	・ ユーザーは、AIシステムにデフォルトで組み込まれた保護機能によって、データの侵害・不正使用から保護され、真に必要な場合のみデータ収集がなされること。 ・ AIシステムの開発者等は、ユーザーに対しデータ収集に関する情報を適切に開示し、ユーザーの意思を尊重すること。	・ NISTは、組織がプライバシーリスクを管理するための包括的かつ実践的なプライバシーフレームワークを提供しており、国内外の様々な分野の組織に活用されている。 ・ NY州では、公立学校での学生管理に生体認証技術を使用する際、事前に技術の使用、プライバシーや権利関係を明確化するよう法制化している。
④告知と説明	・ ユーザーは、AIシステムが使用されていることを把握し、それが自身に及ぼす影響について通知を受け、理解する必要があること。 ・ AIシステムの開発者等は、AIが果たす役割や機能、AIを用いることによる影響や、AIを用いる理由について、簡潔かつ明確に説明すること。	・ 貸金業者等は、消費者に対して信用供与を拒否する場合、当該不利益処分を決定した事由を含む説明を通知するように連邦法により義務付けられている。 ・ DARPAやNISTでは、説明可能なAIシステムの研究開発に関するプログラムを実施しており、より高度な説明可能な機械学習技術を開発中。

① AI権利章典（米国）

- 5つの原則の概要と、実践に向けた取組みの事例（一部抜粋）は下表のとおり。（前頁の続き）

原則	概要	実践に向けた取組みの事例
⑤問題発生時の人間による代替、検討、対応	・ ユーザーは必要に応じてAIシステムからオプトアウトし、代替手段として人間の担当者を選択できるようにすること。 ・ AIシステムにおける障害発生時や、ユーザーがその影響について異議申し立てを要求する場合は、その問題について検討・解決できる人間に迅速にアクセスし、救済措置等の対応を講じること。	・ 多くの企業のカスタマーサービスでは、チャットボットやAI搭載の自動応答システムと、人間によるサポート体制を併用して、複雑な要求に対応可能にしている。 ・ 少なくとも米国24州の投票是正法では、有権者照合システムの誤りで投票が無効とされた場合等に、有権者が人間の選挙管理人に対応を要求することが可能。

② AI規制政策文書（英国）

- 英国政府は、「AI規制政策文書」を2022年7月に発表した。
- AIの特性に合わせた分野横断的な原則に基づき、AI規制に向けた革新的なフレームワーク^(※)を確立することを提言している。

(※) EUのAI規則案とは異なり、原則として法的拘束力を有しないフレームワークの設計が想定されている。

- イノベーションの促進とAIEコシステムの発展を支援するために、以下のアプローチを取ることを提言している。

- ① AIが使用される**環境や状況に応じて、柔軟に規制の設計**を実施する。
- ② AIにより**現実的に発生している高リスクな事象**に焦点を絞る。
- ③ **包括的かつ分野横断的な原則を確立**した上で、これらを各セクターやドメイン内で柔軟に解釈し、規制当局と関係者間で調整する。

分野横断的な原則として以下を要求。

- | | | |
|--------------------|---------------|-------------------------|
| ・AIの使用における安全性の確保 | ・透明性と説明可能性の確保 | ・AIガバナンスにおける法人の説明責任 |
| ・技術的な安定性と設計とおりの機能性 | ・公平性への配慮の実装 | ・AIより導出された結果に対する是正措置の確保 |

- ④ 規制の継続性を重視し、既存のプロセスと連携した上で、**ガイダンスや自主的な措置など、非法定ベースでの原則**を定める。

② AI規制政策文書（英国）

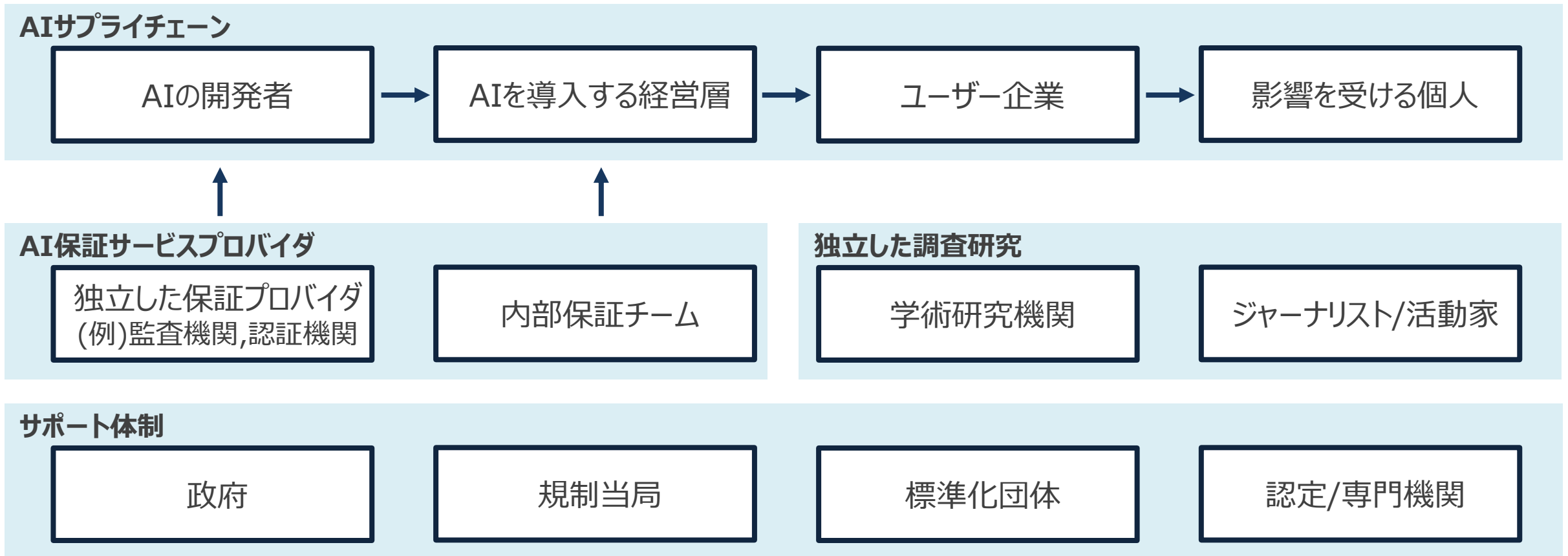
- 同国の規制フレームワークへの準拠や、国際的な標準化機関との連携・調整について表明している。
 - 革新的なフレームワークの設計・実装に向けた今後の方向性について言及している。
-
- 本政策によるアプローチは、英国にて規制策定の標準として定められている**“より良い規制フレームワーク”の規制原則に準拠**し、かつ2021年7月に発表された**デジタル規制計画とも合致**している。
 - デジタル・エコシステムの国際的な市場の重要性を視野に入れると共に、AI規則のグローバルな取組みにおける英国のプレゼンス維持・向上に向けて、欧州評議会、OECD作業部会、AIに関するグローバルパートナーシップ、ISO、IEC等の**国際的な標準化機関との緊密な連携を表明**している。
 - 今後、フレームワークの設計・実装に向けた方向性として、以下の点を提示している。
 - ① 当該フレームワークに対し、他の国際的なアプローチとの調整を図ると同時に、**優先順位の高いAIリスクに適切に対処できるか検討**する。
 - ② **規制当局の役割や権限の範囲**について検討する。
 - ③ **フレームワーク全体を適切に管理、評価できるよう設計**し、将来的な更新に備える。

③ 効果的なAI保証エコシステムに向けたロードマップ（英国）

- 英国データ倫理・イノベーションセンター（CDEI）は、「効果的なAI保証エコシステムに向けたロードマップ」を2021年12月に発表した。
 - 効果的で成熟したAI保証エコシステム実現のための6つの優先分野を提示した。
-
- 英国政府が2021年9月に発表した国家AI戦略の一環で、**AIがもたらす様々なリスクを特定かつ軽減し、信頼性の高いツール・サービスを実現するエコシステム**の構築は必要不可欠とされている。
 - 本ロードマップは、今後5年以内に**世界をリードするAI保証エコシステムの確立**を目標としており、それに向けて様々なプレイヤーの果たす役割や必要な手順を提示している。
 - 6つの優先分野は下記のとおり。それぞれの現状の課題や実現する上での要件、CDEIの果たすべき役割が整理されている。
 - ①AI保証の需要喚起 ②AI保証の市場構築 ③標準化
 - ④専門機関の設置 ⑤規制 ⑥独立した調査研究
 - ロードマップの実現に向けた取組みとして、**AI監査フレームワークの開発**や**AIの利用に関する新たなガイダンス作成**の他、**AIスタンダードハブの試験運用**についても言及されている。

③ 効果的なAI保証エコシステムに向けたロードマップ（英国）

- AI保証エコシステムの主要なプレーヤーが下図のとおり整理されている。
- AIサプライチェーンに属するプレーヤーや、実際にAI保証サービスを提供する独立した保証プロバイダーだけでなく、ジャーナリストや認定機関等の周辺プレーヤーの重要性も指摘されている。



③ 効果的なAI保証エコシステムに向けたロードマップ（英国）

- 6つの優先分野におけるロードマップの概要は下表のとおり。

ビジョン	概要	現状の課題	実現する上での要件
①AI保証の需要喚起	AIサプライチェーン全体及びAIリスク全範囲にわたり、信頼性の高い効果的な保証に対する需要を喚起する。	主に風評リスクのあるトピックなど、限定的な需要に集中している。	AIサプライチェーンの各プレーヤーが保証サービスを通じて、AIのリスクとその軽減に向けた説明責任に対する理解を深める。
②AI保証の市場構築	効果的、効率的かつ目的に沿ったサービスやツールを提供する、動的で競争力のあるAI保証市場を構築する。	主に開発者の需要に焦点を当てた、断片的な市場になっている。	政府、規制当局、民間団体が、スキルギャップを埋め、効果的なAI保証の提供に対するインセンティブを与えるようなAI保証サービスの市場を形成する。
③標準化	AI保証の標準化に向けて、適切な共通の評価基準を含めた共通言語を提供する。	様々なAI標準化の取り組みが実施されているが、保証をサポートするようなレベルに達していない。	標準化団体が、AIのリスクに対する共通の評価基準と、標準化された管理システムを提供する。

③ 効果的なAI保証エコシステムに向けたロードマップ（英国）

- 6つの優先分野におけるロードマップの概要は下表のとおり。（前頁の続き）

ビジョン	概要	現状の課題	実現する上での要件
④専門機関の設置	信頼性の高い保証サービスを提供する、責任あるAI保証の専門機関を設置する。	AI保証サービスの信頼性構築に向けて、最適な専門機関モデルに関する明確な方針が無い。	専門機関への道筋を構築するため、既存の専門機関と認定機関が協業する。
⑤規制	システムが規制要件に準拠していることを保証するガイドラインを策定することで、信頼できるAI革新を可能とする。	個々の規制当局による取り組みや、既存の保証システムのAIへの適用例もあるが、規制分野全体に広く展開する必要がある。	規制当局が、AIに関するリスクを管理・軽減するために、AIサプライチェーンに向けた保証ガイドラインの策定に積極的に取り組む。
⑥独立した調査研究	独立した研究者が、保証技術の開発とAIリスクの特定に重要な役割を果たす。	産業界と独立した研究者の間の連携が取れておらず、効果的な協業ができていない。	産業界は、独立した研究者との連携を強化し、研究者がAIリスクへの対処、かつAI保証技術のソリューション開発に貢献できるようにする。

④ AIスタンダードハブ（英国）

- 英国政府は、「AIスタンダードハブ」の試験運用の実施を2022年1月に発表した。
 - 同ハブを通じて、事業者向けの実用的なツール開発や、新たなオンラインプラットフォームを通じたAIコミュニティの整備などを実施する。
-
- 本ハブの目的は、英国政府が2021年9月に発表した国家AI戦略の一環で、**AIガバナンスの発展や、同国への関連投資・雇用の促進**の他、**AIの国際標準の整備**に寄与することで、**AIガバナンスにおける同国の存在感を高める**狙いもある。
 - 試験運用は、アラン・チューリング研究所の主導の下、英国規格協会(BSI)、国立物理研究所(NPL)と連携して実施されている。また、デジタル・文化・メディア・スポーツ省(DCMS)デジタル標準チームと人工知能局（OAI）が同ハブの支援を行っている。
 - 同ハブの実施事項は、下記の4本の柱を中心に構成されている（後述）。
 - ①ライブラリ ②コミュニティと共創 ③知見と訓練 ④調査と分析

④ AIスタンダードハブ（英国）

- 「AIスタンダードハブ」の4本の柱の内容は以下のとおり。

柱	概要
①ライブラリ	・ 国内外のAI標準化、及び関連する標準の開発に関するライブラリとして機能する。 ・ 標準化団体によって開発、公開されているAI標準を収録したデータベースにより構成される。 ・ 政府の戦略、標準化ロードマップ、AI技術に関する法規制等、多様な情報にアクセス可能となっている。
②コミュニティと共創	・ 課題の特定や優先順位付け、共通の戦略の開発、ベストプラクティスに向けたコンセンサスの構築を目的とする。 ・ 主要なテーマやトピックに関する関係グループ間の新たな連携、アイデア交換の場の調整、問題解決に向けた共創を促進するためのプラットフォームとして機能する。 ・ ワークショップやオンライン/対面での各種イベント・フォーラム等を開催する。
③知見と教育	・ eラーニング教材や対面での研修で、AI標準の開発と使用に関する手続き上の知識や、実践的な知見に関するトレーニングを実施する。 ・ 参加者がAI標準の開発に必要な知識とスキルを習得し、公開されているAI標準を解釈及び使用できるようにする。
④調査と分析	・ AIの国際的な標準化に向けて戦略的に調査と分析を実施する。 ・ 実施例として、俯瞰調査、ギャップ分析、政策提言が挙げられる。

⑤ モデルAIガバナンスの枠組み（シンガポール）

- 情報通信メディア開発庁（IMDA）とシンガポール個人情報保護委員会（PDPC）は2020年1月に第2版となるモデルAIガバナンスの枠組みを発表した。
- 同枠組みでは、実際の組織における事例を紹介することで、実務上の留意事項を含む情報を提供している。
- 本枠組みは、以下4つの観点で構成されている。
 - ①組織内部のガバナンス構造と手段 ②AIを用いた意思決定における人間の関与度合の決定
 - ③運用管理 ④ステークホルダーとの関係・コミュニケーション
- 上記4つの各項目において、Master Card、Facebook、MSD、その他複数の国内外の企業における取組みが紹介されている。
- 組織がAIガバナンスを実践する際、同枠組みとの整合性を簡易的に評価できることを目的とした、**組織のための実装及び自己評価ガイド“ISAGO”**が提供されている。
- 同枠組みと併せて、**組織におけるAIガバナンスの取組み例を多数掲載した事例集**：Volume1、及び国家プロジェクトである“AI Singapore”に参画している4社（IBM, RenalTeam, 損保HD Asia, VersaFleet）を含むVolume2が公開されている（次頁以降に概要を整理）。

⑤ モデルAIガバナンスの枠組み（シンガポール）

- モデルAIガバナンスの事例の概要(Volume1)は以下のとおり。

組織名	組織概要	取組み例
Callsign	ID認証に係るAIモデルの開発企業 (拠点：ロンドン)	AIモデルの開発に当たり、独自のAIガバナンスフレームワークの策定と、同社のデータ履歴プロセスを監査するデータ変更ガバナンス委員会を設置。
DBS Bank	シンガポールに本社を置く東南アジア 最大級の銀行・金融機関	AIを用いたマネーロンダリング防止モデルの実装に際し、内部ガバナンスフレームワークの策定、及び責任あるデータ利用委員会を設置。
香港上海銀行	65か国で事業を展開する世界最大 級の銀行・金融機関	AIを用いた融資申請の評価モデルに対するガバナンスフレームワークを開発、及びグ ローバルモデル監視委員会を設置。
MSD	世界最大級のバイオ医薬品企業	従業員のエンゲージメント管理モデルに対し、“human in the loop”のアプローチを 採用し、従業員の権利を尊重してAIガバナンスプロジェクトを実施。
Ngee Ann Polytechnic	シンガポールの高等教育機関	入学者選考を自動化するAIプラットフォームの導入に際し、効率的かつ、説明責任、 公平性を担保したフレームワークを実装。
Omada Health	サンフランシスコに拠点を置く医療機 関	AIを用いた個別化医療サービスの提供にあたり、リスクや影響を評価・管理するAI/ データガバナンス委員会を設置。
UCARE AI	医療機関の請求書見積予測のため のAIモデルを開発するスタートアップ	同枠組みに則り、内部の監査チームを設置の上、患者の個人情報の保護、データの 説明責任、正確性、透明性の確保を重点的に検証。
Visa Asia Pacific	世界最大級のデジタル決済ブランドを 展開する企業	カード保有者に打ち出す旅行キャンペーンのAIによる最適化モデルを、“human over the loop”のガバナンスアプローチでカード発行銀行に対し提供。

⑤ モデルAIガバナンスの枠組み（シンガポール）

- モデルAIガバナンスの事例の概要(Volume2)は以下のとおり。

組織名	組織概要	取組み例
Darwin	オーストラリア・北部準州の州都	スマートシティプロジェクトである“Switching on Darwin”にて、同枠組みの自己評価ガイド“ISAGO”を採用してAIガバナンスの実践に向けた自己評価を実施。
Google	巨大テック企業。2018年にAI原則を発表	シンガポールの同枠組みのサポートを主導。有名人に特化した顔認証APIツール“Celebrity Recognition API”は、自社のAI原則を遵守したもの。
Microsoft	巨大テック企業。2016年にAI原則を発表	2017年にAI倫理委員会Aetherや、責任あるAI部門の設置以来、会話チャットボットの開発を主軸にAIガバナンスの実践を世界に先駆けて主導。
TAIGER	企業・公的機関向けの業務自動化AIツールを開発するスタートアップ	顧客にAIソリューションを展開する際には、同枠組みに準拠して設計された内部AIガバナンスプロセスを実践し、製品の信頼性、透明性確保を強化。
IBM	170か国で事業を展開する多国籍IT企業	製品の品質評価を支援するAIモデルの開発段階に応じて、“human in/over the loop”のアプローチを併用し、正確性、信頼性、説明責任を確保。
RenalTeam	透析患者の入院リスクを予測するAIモデルの開発企業	“human in the loop”のアプローチを採用し、個人情報の保護とデータの機密性を確保した上でAIモデルの精度の向上を実現。
損保HD Asia	日本の法人・個人向け損害保険会社	人身事故の不正請求の検出や、保険の支払い可否の判断用AIモデルを開発し、意思決定プロセスに“human over the loop”のアプローチを採用して透明性を担保。
VersaFleet	物流向けのAI搭載ソフトウェアを開発するスタートアップ	輸送管理の自動化を実現するソフトウェアの開発プロセス全体において、内部管理チームのステークホルダーが適切に関与。

⑥ A.I. Verify (シンガポール)

- 情報通信メディア開発庁 (IMDA) とシンガポール個人情報保護委員会 (PDPC) は2022年5月にA.I. Verifyを発表した。
- A.I. Verifyは、AIガバナンスのテストフレームワークとツールキットであり、組織の適切なAI利用を検証できる。
- A.I. Verifyは、**組織の適切なAI利用を客観的かつ検証可能な方法でステークホルダーに示すことが可能**であり、**組織とステークホルダーとの間の信頼関係の構築、AIに対する利用者の信頼性向上とAIシステムのさらなる利用拡大**を目的としたものである。
- A.I. Verifyは、①テストフレームワーク (選択したAI倫理原則に関連するテスト可能な基準を指定) と、②ツールキット (テストフレームワークで説明されている技術テストを実行し、プロセスチェックを記録) で構成されている。
- A.I. Verifyの開発にあたり、**EUやOECDなどの国際的に支持されているAI倫理の原則、ガイドライン、フレームワークに準拠**するよう、①透明性、②説明可能性、③反復可能性/再現性、④安全性、⑤頑健性、⑥公平性、⑦説明責任、⑧人間の関与と監査、の8項目の倫理原則が採用されている。
- 上記8項目について、**技術テストとプロセスチェックを組み合わせることで評価が可能**であり、開発者、管理者、ビジネスパートナー向けのレポートも作成される。

⑥ A.I. Verify (シンガポール)

- A.I. Verifyは、試用版として提供されているが、既に複数の企業による利用とフィードバックが提供されており、AI検証のコミュニティが構築されつつある。
- A.I. Verifyは現段階では試用段階であり、**MVP（実用最小限の製品）として利用可能**である。
- A.I. Verifyにより検証されたAIシステムにリスクやバイアスがないこと、完全に安全であることを保証するものではないことが強調されている。
- AmazonWS、Google、Meta、Microsoft、シンガポール航空を含む**10機関が、本キットによる検証を既に実施し、結果がフィードバック**されている。
- IMDAは、A.I. Verifyに国際的なAIガバナンスの基準及び業界標準の導入をさらに進めるため、今後も業界パートナーからのフィードバックを受けながら開発を続けるとしており、**AIガバナンスの国際標準の開発に貢献するベンチマークとなること**を目指している。

⑦-1 AI責任指令案（EU）

- 欧州委員会は「AI責任指令案」を2022年9月に発表した。
 - 同案は、契約外の民事責任をAIに適用したものであり、AIシステムの提供者等に対する損害賠償責任のルールの整備を狙いとしている。
-
- 同案は、⑦-2「製造物責任指令の改正案」と共に、2021年4月に発表された**AI規則案を補完**するものとなる。
 - AIシステムから生じる、契約外の過失に起因した損害賠償責任について、**EU加盟各国の法律と矛盾しない最低限の基準をAIシステムの提供者に対して設定**することで、被害者を救済するとともに、AIシステムの利用に対する安心感の向上と、AIシステムのさらなる利用促進が目的とされている。
 - **証拠へのアクセス権と保全**：高リスクのAIシステム(AI規則案にて定義)により損害が生じた疑いがある場合、同案の適用により、被害者は裁判所に申し立てることが可能。当該システムの提供者は、特定の条件下において**関連性のある証拠の開示や保全措置を講じるよう命じられる**。
 - **因果関係の推定**：従来の製品と違い、AIにおいては、AIの出力と、AIシステムの提供者の過失との間の因果関係の立証が困難であった。しかし、同案の適用により、損害が発生した際、特定の条件下において、**AIシステムの出力(または出力できなかった事象)が、AIシステムの提供者の過失によって生じたと推定**することが可能となる。

⑦-2 製造物責任指令の改正案（EU）

- 欧州委員会は「製造物責任指令の改正案」を⑦-1:AI責任指令案と同日の2022年9月に発表した。
 - 同案は、AIシステムを利用した製品・サービスが普及する近年の情勢において、製造物の欠陥に起因する損害賠償責任の規則の全面的な改正を狙いとしている。
-
- 1985年に施行された現行の製造物責任指令は、近年のデジタル化に即しておらず、欠陥製品による被害者への救済措置に限界があった。同案の改正は、**デジタル化の進展や、循環型経済の発展に対応**するものである。
 - 同案の適用対象は、**EU域内で上市されたあらゆるソフトウェア製品、デジタルサービス、及びそれらの販売後のアップデートや機械学習に起因する欠陥**などであり、製造事業者が損害賠償責任を負う。
 - 製造物責任を負う事業者には、従来の**製造事業者**に加え、**輸入事業者、eコマースの販売事業者**も含まれる。
 - 循環型経済志向の高まりにより、製品の長寿命化を目的とした改造・改変製品の増加に対応するため、**上市済みの製品に対して改変を加えた事業者も製造物責任を負う**ことになる。
 - ⑦-1「AI責任指令案」と同様、特定の条件下で被害者による**証拠へのアクセス権や保全申し立ての手続きが容易**になり、被害者による**因果関係の立証責任も軽減**される(ただし、製造事業者は当該因果関係の推定に対して反証を行うことが可能)。