NTT DATA
Trusted Global Innovator

AIシステムの開発・運用・利活用に、安心と信頼をもたらす

「AIガバナンス」

AIガバナンス、AI規制、AI倫理、AIリスク、AI指針

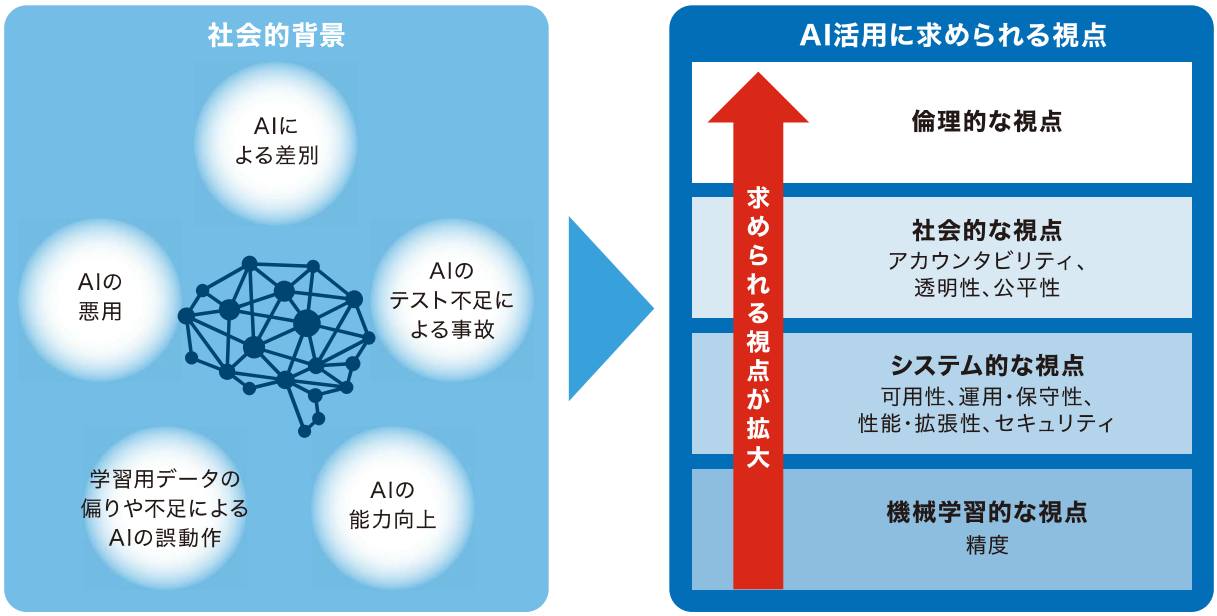
CONTENTS

CHAPTER.1
はじめに

ディープラーニングの登場をきっかけにAIの能力が向上したことで、AIはサイバー空間から実世界で活用されるようになり、その活用範囲は自動運転や融資審査、画像診断など人命や経営に関わる高い信頼性が求められる領域に拡大してきました。特に2022年の後半からデータを生成するAIが急激に進化しており、画像生成AIや、文書生成AIは広く使われるようになりました。これらのAIは、人間の生成物と比しても劣らない出力をすることから、その利便性に注目が集まっています。

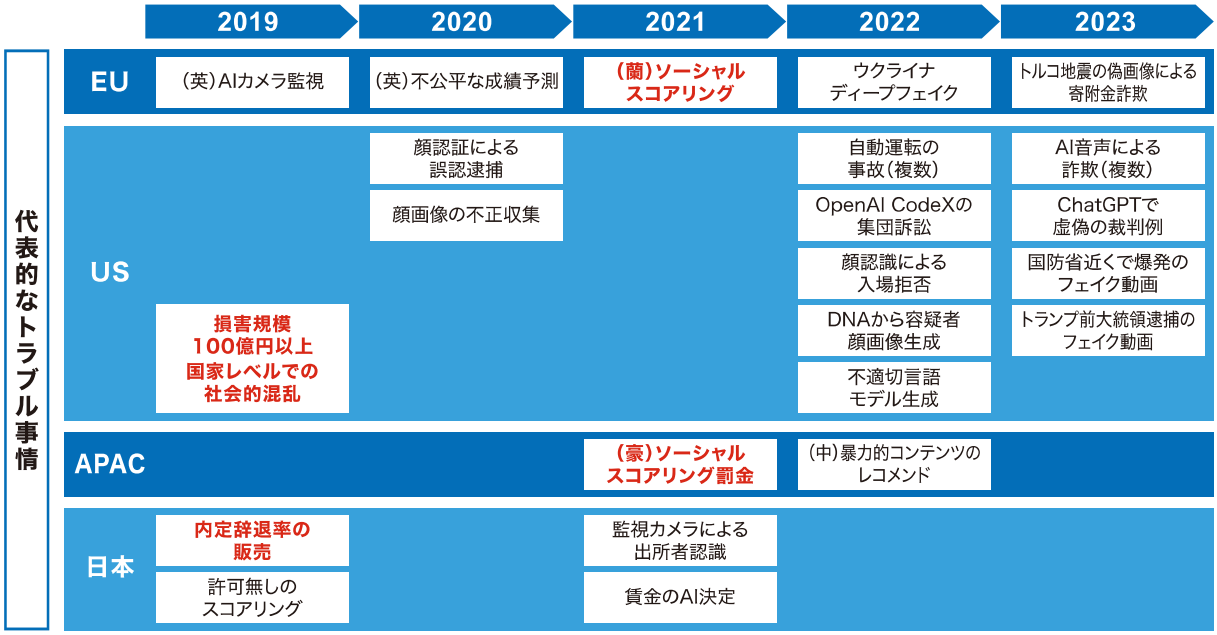
一方でAIの悪用やテスト不足による事故、学習用データの偏りや不足による誤動作といった問題が散見されるようになりました。生成AIでも、誤った出力がされる問題やフェイクニュースのような悪用できるコンテンツが容易に生成できる問題に加えて、入力からの情報漏洩の問題や学習データや生成物による著作権侵害の問題などが指摘されています。このため、AI活用に求められる視点は、精度や可用性やセキュリティなどシステム的な視点だけでなく、AIのアウトプットに対するアカウンタビリティや透明性といった社会的な視点、さらには、本当にそのタスクにAIを適用して良いのかといった倫理的な視点に拡大しています（図1）。

図1　AI活用に求められる視点の変化



実際のところ、過去5年間のAIによる問題事例は世界的に増加傾向にあり、2022年後半からは生成AIに関する問題が急増しています。例えば公的機関による審査にAIを適用してマイノリティーに差別的な判定をした問題や、顔認識による誤認逮捕や入場拒否の問題、生成AIでつくられた偽情報拡散の問題などが世界中で起きています（図2）。

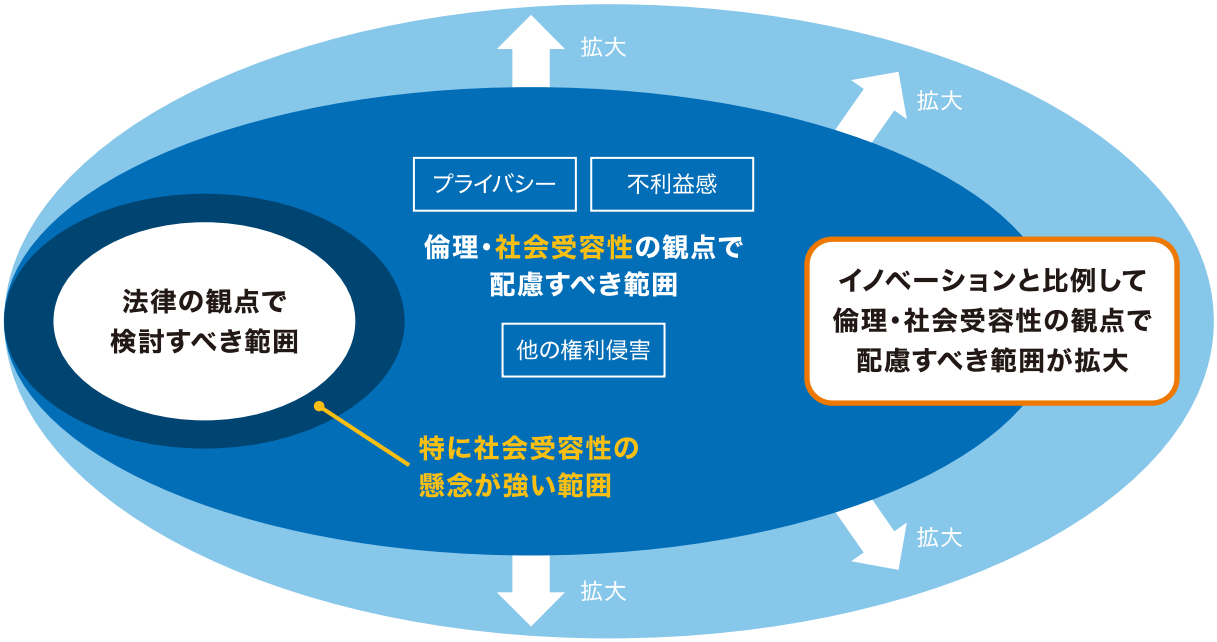
図2 増加するAIによる問題事例



※インシデントDB:世の中のAI関連トラブル情報をまとめたDB(https://incidentdatabase.ai/)、および、ニュース記事から作成
運営主体はPartnership on AI(Facebook、Amazon、Google、IBM、Microsoftにより設立された米国の非営利団体、現在約100社がパートナー)

こうした問題事例に共通するのは、「倫理・社会受容性」の問題であり、単に法律を遵守していれば良いわけではなくなってきていることを意味しています。AIの入出力に対する何らかの権利侵害の懸念は、結果として社会受容性を低下させ、企業のレピュテーション低下を招きビジネスの継続性に影響を生じるリスクがあります。また、「倫理・社会受容性」の観点は、例えば、個人情報保護法とプライバシーの関係が典型ですが、イノベーションと比例して配慮すべき範囲が拡大していくため、変化に対する追従が必要です(図3)。

図3 倫理・社会受容性の問題

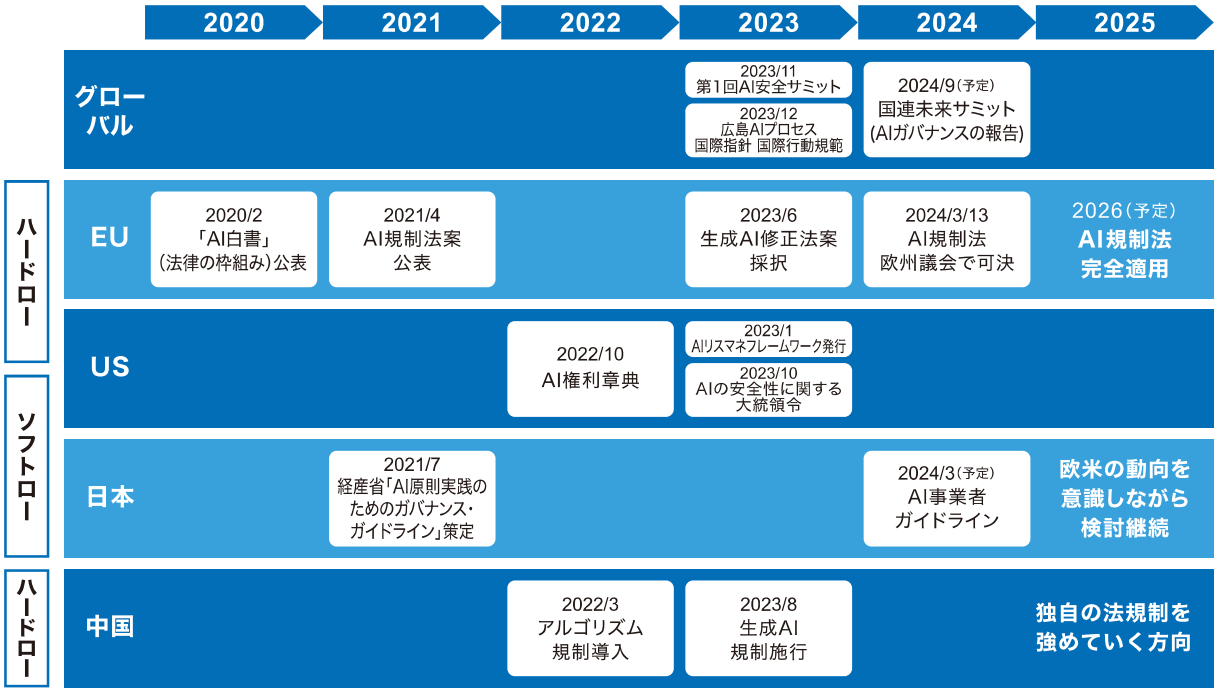


CHAPTER.2
AI規制に関する国内外の動向

このようにAIの進化と共にAIに起因する問題が多発している状況を踏まえ、世界各国で法規制やガイドラインなどAIのガバナンスを強化する動きが進んでいます（図4）。

2023年のG7広島サミットでは広島AIプロセスとして生成AIについて議論する場が設置され、12月に首脳合意がされました。ここでは各国の動向とG7をはじめとする多国間の取り組みについて紹介します。

図4　AI規制の国内外動向



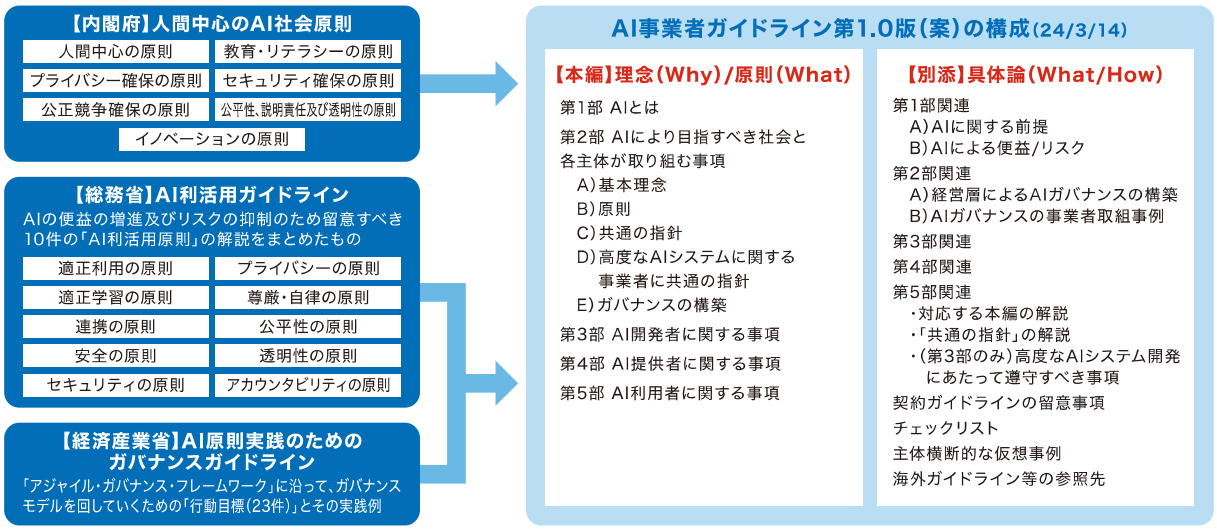
1. 日本

2022年までは総務省や経済産業省からAIの利活用における原則やガイドライン、および、各省庁からAIの出力の誤りが人命に直結する医療や自動運転、財産への影響が大きい金融分野などで個別のガイドラインが公開され、これらを皆で守るソフトローの路線でした。しかし、生成AIの登場により、AIに関する国内外の状況が大きく変化したことで、国際的な議論動向や国内与党からの提言も踏まえ、政府にAI戦略会議が設置され、戦略の見直しが行われました。

具体的には、AI戦略会議の議論を踏まえ、内閣府、総務省、経済産業省が連携して事業者向けの新たな統合ガイドラインである「AI事業者ガイドライン」の策定が進んでいます。これは、総務省の「AI開発ガイドライン」と「AI利活用ガイドライン」、経済産業省の「AI原則実践のためのガバナンスガイドライン」の3つをベースにして、生成AIなど新しい要素も取り込んだ上で、事業者向けの1つのガイドラインにまとめるものです（図5）。順調に検討が進めば2024年3月中にガイドラインの初版が公開される見込みです。

また、2023年12月に自民党から出された「AIの安全性確保と活用促進に関する緊急提言」では、AI事業者ガイドラインの遵守の法制化が提唱されています。

図5 **AI事業者ガイドライン（案）の構成**



2. アメリカ

2022年まではプライバシーや金融、雇用など分野ごとにAIの利用規制や、これらの分野への自動意思決定システムに公平であることの透明性の義務を課す「アルゴリズム説明責任法案」など、分野ごとの法規制が進められてきました。また、AIそのものには、AIを用いたシステムを設計、使用する際に考慮すべき5つの原則を定めた「AI権利章典」の検討や、AIによる市民の自由意志や差別、組織の評判やセキュリティ、グローバルな金融システムやサプライチェーンへの悪影響を想定した「AIリスクマネジメントフレームワーク」といったハードローとソフトローの両面からの規制強化が行われてきました。

2023年8月には、生成AIの急速な発展を受けて、バイデン大統領がOpenAIやマイクロソフト、グーグルなど7社のトップと会談を行い、開発企業がサービス発売前の段階で、外部専門家による検証やリスク評価を通じ、AIの安全性や信頼性を保証することで合意しました。

その後、2023年10月末にAIの安全性に関する大統領令が発行され、これに基づいて法規制が行われる見込みとなり、今後、どういった規制強化が行われるのか注目されています。

3. EU

EUのAI規制法は2021年4月に法案が公開され、2023年12月9日に大筋合意、2024年3月13日に欧州議会で可決されました。今後は加盟国で構成する理事会による承認を経て成立し、2026年にも完全適用されるといわれています。

EUのAI規制法は、「リスクベース・アプローチ」により倫理の観点からリスクを4つにカテゴライズし、対策を義務づけています（図6）。

「受容できないAI」は「人々の安全や権利に対して明らかに驚異のあるAIシステム」で、人の潜在意識への働きかけや、公的機関による社会的スコア利用が該当し、これらは原則、禁止とされています。

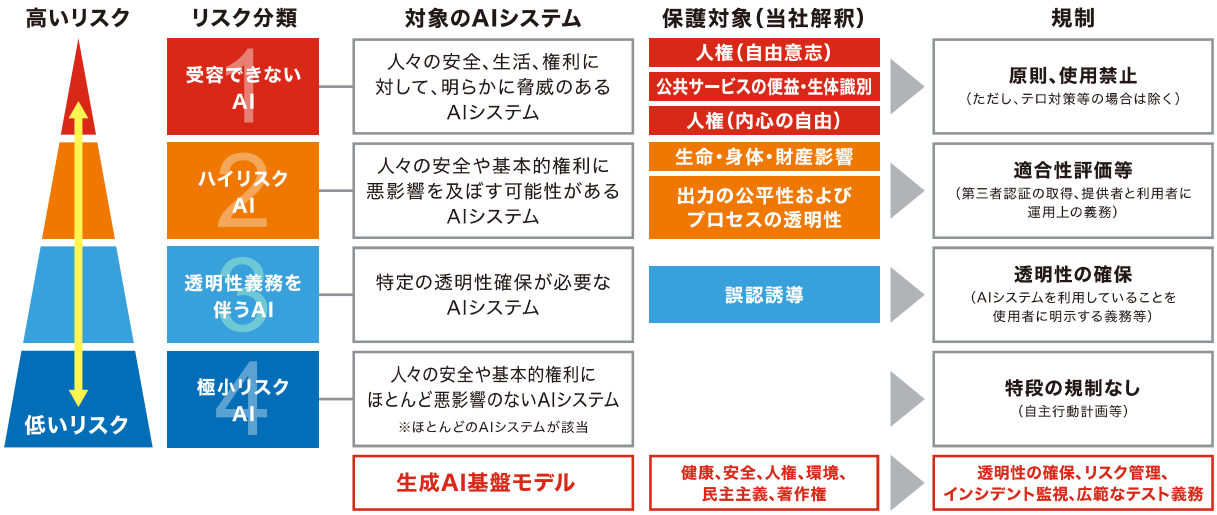
「ハイリスクAI」は「人々の安全や権利に悪影響を及ぼす可能性があるAIシステム」で、医療や自動運転、人材採用などが該当し、これらには適合性評価として第三者認証の取得や提供者と利用者に運用上の義務を課すようになっています。

「透明性義務を伴うAI」としては、チャットボットや感情推定といったAIが該当し、これらは人の誤認誘導とならないように、AIシステムを利用することの明示といった透明性の確保の義務が課されます。

なお、生成AIは、様々なユースケースが想定されることから上記のような分類ではなく、「Foundation Model」としてこの枠組みとは別に、モデルの提供者に対して運用上の義務と透明性義務の規制、さらに学習データの開示を義務づけるなど、生成AI特有の規制が課されるようになっています。

違反した場合は最大で3,500万ユーロ（約55億円）あるいは 全世界売上高の7%の大きい方の金額という巨額の賠償金が課されます。

図6 EUのAI規制法の概要



総務省AIネットワーク社会推進会議（第19回）配布資料、経済産業省第1回 AI原則の実践の在り方に関する検討会配布資料
市川類先生講演資料「生成AIの社会的リスクと世界のAI規制・ガバナンス政策動向」より作成

4. G7

2023年のG7広島サミットではグローバルAIガバナンスが主要トピックとして取り上げられ、責任あるイノベーションと実装の推進が宣言されました。その内容は大きく2つあります。

1つはAIガバナンスの相互運用性の確保です。米欧日それぞれでAIに対する規制の考え方が異なっていることを踏まえ、たとえ制度が違ってもAIガバナンスとして相互に運用できるようにするために、国際機関を通じた国際技術標準の開発および採用の推進を図ることです。

もう1つが生成AIに関する議論のための「広島AIプロセス」の創設です。これは年内に結論が得られるように定期的に議論が進められ、12月に広島AIプロセス包括的政策枠組みが承認されました。4つの文書と今後の作業計画が取りまとめられ、2024年の議長国イタリアに引き継がれました。主に、12か条からなる「国際指針」と、12か条の指針に対してAI開発企業が取るべき対策事例を例示した「行動規範」の2つの柱で構成されています。具体的には、AI関連企業が製品を市場に投入する前に外部の専門家のチェックを受けることや、人がつくったものと区別するためAIの生成物には原則として「電子透かし」の導入などが適用されるべきとしています。

5. その他

中国は2022年にレコメンド等のアルゴリズムで差別を禁止する規制法、2023年8月に生成AI規制を施行するなど独自で法規制を強化しています。

OECDが2019年5月に公開したAI原則は、世界中で作成されているAI原則やガイドラインの根本になるものといわれています。2022年の生成AIの登場を踏まえて、AIの定義と原則の見直しが進められています。

2023年11月にイギリスでAI安全サミットの第1回が開催され、G7に加えて中国も参加し合計28か国で生成AIや汎用AIをフロンティアAIとして、その安全性に取り組むことを合意しました。2024年も韓国とフランスで開催の予定で、今後の議論が注目されます。

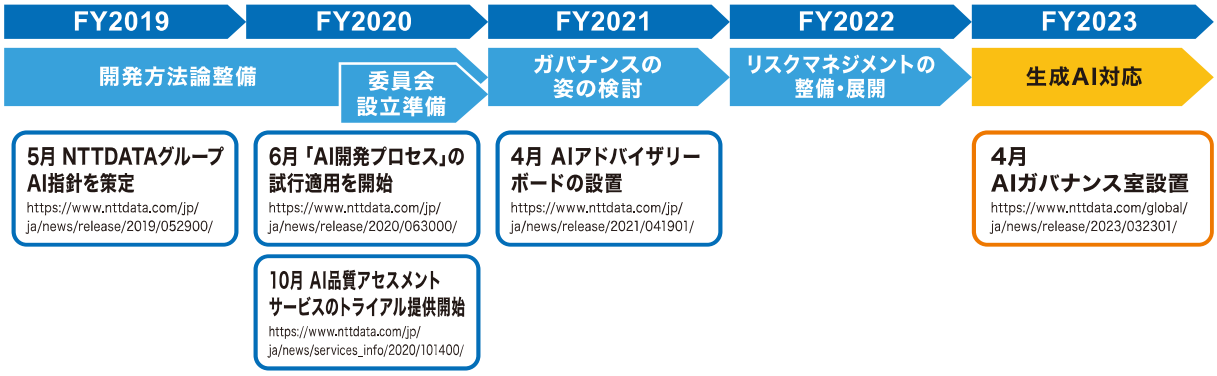
国連でも「AIに関する諮問機関」が2023年10月に設置され、世界中から委員が集められてAIガバナンスに関する議論を行い、2024年9月に開催予定の国連未来サミットで報告予定といわれています。

このように、企業におけるAIの利用に対して、様々な問題が拡大している一方で、違反した場合には巨額の制裁金が科されるような強い法規制の波が押し寄せてきています。そのため、企業にはAI利活用におけるリスクをどのように検知し対処するか、自組織固有のガバナンスの整備が求められているといえます。

CHAPTER.3
安心・安全で信頼性のあるAIの社会実装に必要な「AIガバナンス」

以上を背景に、NTT DATAでも、AIを統制するためのガバナンス活動を推進し、品質・倫理の両面から公平かつ健全なAI活用による価値創造と持続的な社会の発展に向けた活動を実施してきました（図7）。

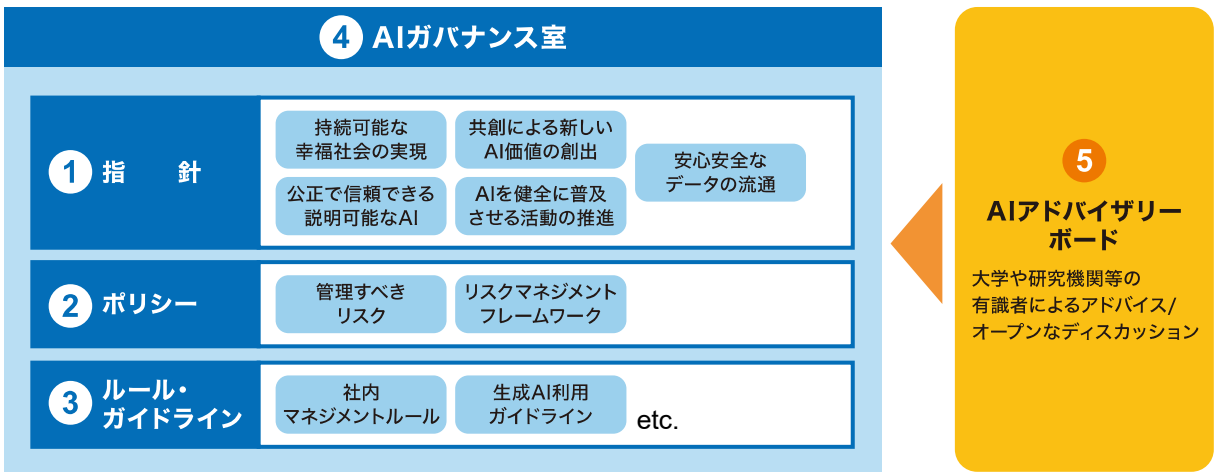
図7 NTT DATAにおけるAIガバナンスの整備



NTT DATAにおけるAIガバナンスとは、「AI・データ活用における法制度や社会規範（倫理）の遵守および、社会の理解と受容性を高める活動を推進し、AI・データ活用技術がもたらすリスクの正しい理解のもとに、公正で信頼できるAI・データ活用の仕組みを実現することで、社会とお客さまがベネフィットを最大限に享受することを可能にし、サステナブルな社会を実現できるようにすること」と定義されています。

実効的なAIガバナンスを確立するために、「AI指針」や「AIリスクマネジメントポリシー」、ポリシーを実効的なものにするための社内ルールやガイドラインの整備に加えて「AIアドバイザリーボード」や「AIガバナンス室」の創設といった取り組みを拡大・継続してきました（図8）。生成AIに対しても、外部動向を踏まえたあるべき生成AIへのガバナンスを具体化していく活動を推進しています。ここからはそれぞれの概要を紹介します。

図8 AIガバナンスの取り組み



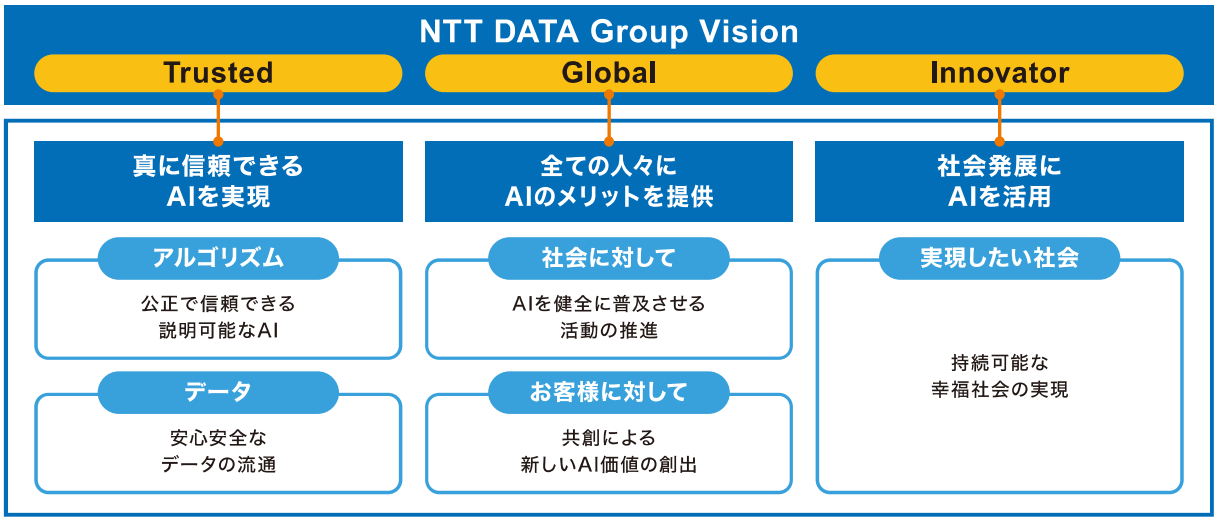
1. AI指針の策定

AIを単なる効率性確保の手段として利用するのではなく、個人・ビジネス・社会がAIのメリットを享受できる「人間とAIが共生する社会」を目指すために、5か条から成るNTT DATAグループのAI指針を2019年5月に公開しています。

https://www.nttdata.com/jp/ja/news/release/2019/052900/

AI指針は、当社がAIに関する開発、運用、利活用、判断をする上で参考とするガイドであり、国内外のAI原則やSDGsの理念、NTT DATAのGroup Visionである「Trusted Global Innovator」を踏まえて策定しました（図9）。

図9 NTT DATAグループのAI指針



2. AIリスクマネジメントポリシー

グループ全体として管理すべきAIリスクと、そのマネジメントフレームワークを定めた「AIリスクマネジメントポリシー」を制定しました。生成AIの急速な発展によりAIガバナンスの整備が急務となっていることを踏まえ、現在、グループ全体で本ポリシーに基づいたルール・プロセスの整備を進めています。

3. ルール・ガイドライン

ポリシーに基づいたリスクマネジメントを実現するために、AI全般について具体的な実効性を持ったマネジメントルールや、生成AIに特化して開発・提供・利用、3つの立場の観点から留意事項と対処方針をまとめた社内向けのガイドラインを整備し、お客さま向けサービス・システムの開発や提供にも活用しています。

4. AIガバナンス室

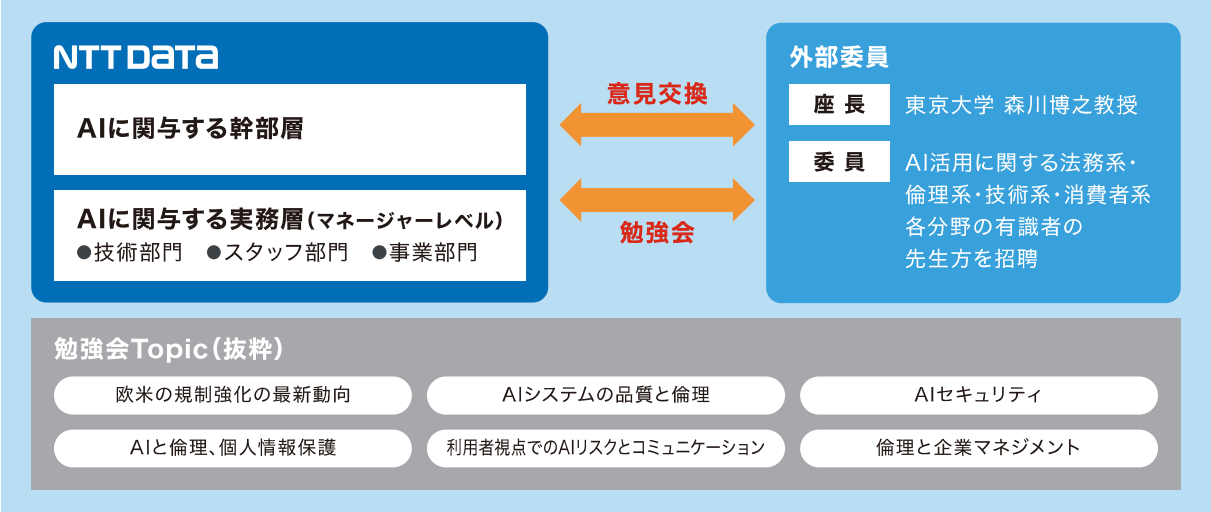
AIガバナンス室は、AIの不適切な利用によって生じるリスクを実効的にマネジメントし、AIの適正活用を推進するための組織です。NTT DATAが提供するAIシステムおよび当社内システムを対象に、AIによる事業リスクを検知するルール・体制を整備し、プロジェクトと一体となって対処することで、社会とお客さまがAI活用の恩恵を最大限に享受し、サステナブルな社会を実現できる環境を整備します。

https://www.nttdata.com/global/ja/news/release/2023/032301/

5. AIアドバイザリーボード

AIアドバイザリーボードは、社会デザイン/ソフトウェア工学/法務・倫理/レジリエンス・SDGsなど様々な分野を専門とする大学・研究機関の有識者で構成されるグローバルな組織です。AI利活用に関する最新の技術動向、法令・規制、市民社会の認識について、有識者とNTT DATAのメンバーが議論する場です（図10）。これにより得られた知見は社内向けのガイドラインなどAIガバナンスの確立に活かされています。

図10 AIアドバイザリーボード



NTT DATAでは、AIを使ったシステム開発・運用を対象に、AIガバナンスに対する社員の意識を高め、倫理や法務面も含めて問題化を事前に回避できるよう、今後もAI技術の進展に併せてポリシーを改定し、必要なルールやガイドラインを策定することで、お客さまに安心・安全なAIサービスの提供を継続する環境を整備していきます。