経済産業省 委託事業

調査報告書

令和 5 年度 デジタル取引環境整備事業

(AI ガバナンスのルールに関する調査研究及び検討会運営)

令和 6 年 3 月

ボストン コンサルティング グループ

内容

1. 検討の目的/背景

AI 関連技術は日々発展し、利用機会と可能性は拡大の一途をたどっている。その一方で、AI システムの活用を巡っては、品質向上に向けた協力推進やリスクへの対応等、様々な課題が存在する。例えば、AI システムの品質は、入力データ、AI モデル、AI システムの使い方の影響を受けるため、品質向上に関する取組について複数事業者間の認識共有や協力が欠かせない一方で、AI システムの品質維持向上のための AI ガバナンスは、各社レベルでも社会全体としても整備が十分とは言いがたい。更に、生成 AI が台頭し注目される等、取り巻く環境は劇的に変化している。

このような中、日本においても、サイバー空間とフィジカル空間を高度に融合させたシステム（CPS：サイバー・フィジカルシステム）による経済発展と社会的課題の解決を両立する人間中心の社会である「Society 5.0」の実現に向け、AI の高度な活用に対する期待が更に高まっている

また、国際社会においても EU において AI を規制する法案が合意される等、様々な議論がなされている。日本は、「群馬高崎デジタル・技術大臣会合」の閣僚宣言や、G7 広島首脳会合における首脳コミュニケにおいて、議長国として議論を主導し、「広島 AI プロセス」を推進してきており、国際的に対し、日本として発信を行っていく必要性も高い。

従前より経済産業省/総務省において、AI の開発や利活用、ガバナンスに関するガイドラインを公表してきているが、生成 AI に関するこれらの動向を踏まえ、これらのガイドラインを、関係する事業者向けによりわかりやすいものとし、発信していくが極めて重要である。

上記を踏まえ、本事業は、経済産業省/総務省及び両省の有識者等と連携し、経済産業省/総務省が公表している既存のガイドラインについて、生成 AI に係る国内外の議論等も踏まえ、開発者・提供者・利用者等様々な立場の事業者向けに、統合・見直し等を行うものである。

これにより、AI に関係する者が、国際的な動向及びステークホルダーの懸念を踏まえた AI のリスクを正しく認識し、必要となる対策をライフサイクル全体で自主的に実行できるように後押しし、イノベーションの促進及びライフサイクルにわたるリスクの緩和を両立する枠組みを関係者と連携しながら積極的に共創していくことを目指す。

2.「AI 事業者ガイドライン 検討会」及び「AI 事業者ガイドライン ワーキンググループ」の運営

検討に当たっては、「AI 事業者ガイドライン 検討会」 (以下、「検討会」と言う) 、及び、個別の論点について具体的な議論を行う「AI 事業者ガイドライン ワーキンググループ」(以下、「WG」と言う) をそれぞれ複数回開催した。

それに加え、検討会・WG 以外にも、有志の意見交換の場の設定や、メーリングリストの活用も行い、各有識者等の意見を聴取・反映しつつ議論し、ガイドラインの統合・構成の大幅な見直し・加筆修正等を実施した。その際、総務省やその受託事業者とも、週次の連絡会議等を通じ、両省の検討が整合するよう、密接に連携した。

①「AI 事業者ガイドライン 検討会」

AI 事業者ガイドラインの内容や構成について、大きな方向性を議論するため、AI に関して知見等を有する有識者を委員とする検討委員会を開催した。

(1) 構成

検討会の構成は、以下のとおり（敬称略）。

【座長】

渡部 俊也	東京大学未来ビジョン研究センター・副センター長 教授

【委員】（五十音順）

生貝 直人	一橋大学大学院法学研究科 教授
市川 類	一橋大学イノベーション研究センター 特任教授
江間 有沙	東京大学国際高等研究所東京カレッジ 准教授
岡野原 大輔	株式会社 Preferred Networks 代表取締役 最高研究責任者
北村 弘	AI リーガルリーダー / CDLE（Community of Deep Learning Evangelist）
國吉 康夫	東京大学 大学院情報理工学系研究科 教授
齊藤 友紀	法律事務所 LAB-01 弁護士
シバタ アキラ	Weights & Biases Japan 株式会社 カントリーマネージャー
中条 薫	株式会社 SoW Insight 代表取締役社長
深津 貴之	株式会社 THE GUILD 代表取締役
福岡 真之介	西村あさひ法律事務所・外国法共同事業 パートナー
福田 剛志	日本アイ・ビー・エム株式会社 東京基礎研究所 所長
舟山 聡	rinna 株式会社 Chief Legal Officer
古谷 由紀子	サステナビリティ消費者会議 代表
増田 悦子	公益社団法人全国消費生活相談員協会 理事長
松本 敬史	デロイトトーマツコンサルティング合同会社 シニアスペシャルリード
吉永 京子	慶應義塾大学大学院政策・メディア研究科 特任准教授

(2) 開催実績

検討会は、以下のとおり開催された。

開催回	日時	主な議題
第 1 回	2023 年 10 月 31 日	● AI 事業者ガイドライン案の検討状況のご共有 ● 他省庁における議論状況のご共有 ● 有識者のご意見照会結果をもとにした検討の方向性のご相談 ● 概要版、別添の構成についてのご相談
第 2 回	2023 年 12 月 15 日	● 有識者のご意見照会結果をもとにした検討の方向性のご相談 ● AI 事業者ガイドライン案の今後の活用方針方策についてのご相談
第 3 回※	2024 年 3 月 14 日	● 「AI 事業者ガイドライン案」に対するご意見及びその考え方 ● 「AI 事業者ガイドライン」第 1.0 版（案）について

※ 総務省 「AI ネットワーク社会推進会議」、「AI ガバナンス検討会」との合同開催

②「AI 事業者ガイドライン ワーキンググループ」

検討会での議論と連携して、個別の論点に係る具体的な検討等を行った。

(1) 構成員

WG の構成は、以下のとおり（敬称略）。

【委員】（五十音順）

岡田 淳　　森・濱田松本法律事務所 弁護士
小谷野 雅晴　神楽坂総合法律事務所 弁護士
齊藤 友紀　法律事務所 LAB-01 弁護士
殿村 桂司　長島・大野・常松法律事務所 弁護士
中崎 尚　　アンダーソン・毛利・友常法律事務所 外国法共同事業 弁護士
羽深 宏樹　京都大学大学院法学研究科附属法政策共同研究センター 特任教授
福岡 真之介　西村あさひ法律事務所 弁護士
古川 直裕　株式会社 ABEJA 弁護士
松本 敬史　デロイトトーマツコンサルティング合同会社 シニアスペシャルリード
丸田 颯人　長島・大野・常松法律事務所 弁護士
宮村 和谷　PwC あらた有限責任監査法人 パートナー

(2) 実施結果

WG は、以下のとおり開催された。

開催回	日時	議題
第 1 回	2023 年 9 月 23 日	● ガイドライン作成に当たって以下の事項をご相談 ➢ 背景・狙いおよび位置づけ ➢ 想定する読者と活用方法を踏まえた方向性/全体像 ➢ 原則の導出過程 ➢ ドラフト全体の流れ/記載のイメージ
第 2 回	2023 年 10 月 20 日	● 有識者のご意見照会結果をもとにした検討の方向性のご相談 ➢ ガイドラインの構成 (記載の重複部分の取り扱い等) について ➢ 原則の内容・要素・それぞれの関係性について ➢ 用語の定義について ● 概要版、別添の構成について

第 3 回	2023 年 12 月 1 日	● 有識者のご意見照会結果をもとにした検討の方向性のご相談 ● チェックリストの方向性ご相談 ● 英訳の進め方のご相談

なお、WG については、上記のとおり公式開催したもの以外に、構成員のうち有志の方々と、メーリングリストも活用しつつ、意見交換/議論を密に実施してきたところである。

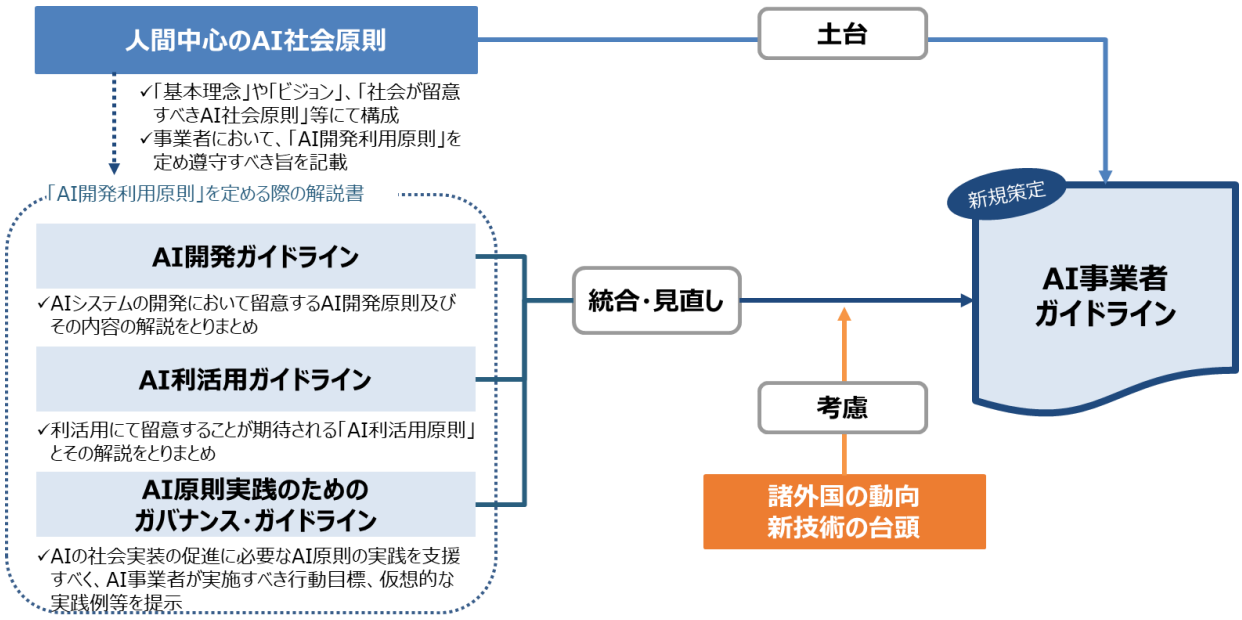
3. 調査結果

① AI に関連する主な諸原則等

本 AI 事業者ガイドラインの検討に当たっては、2019 年 3 月に策定された「人間中心の AI 社会原則」を土台としつつ、我が国における 3 つのガイドラインを統合し、諸外国の諸原則に係る議論の動向や新技術の台頭を考慮して策定した（図 1 参照）。

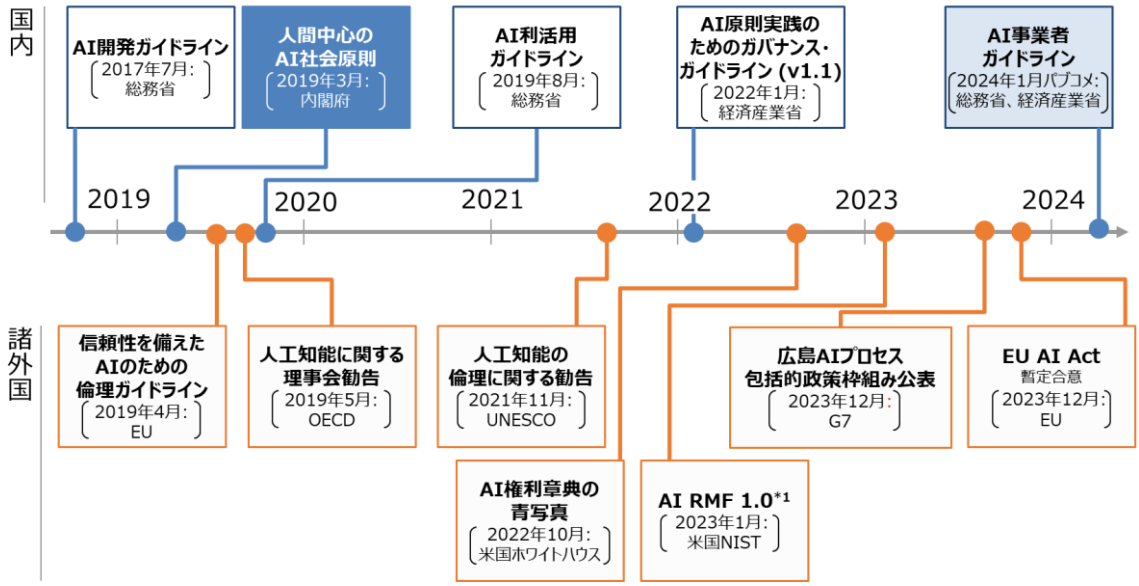
これにより、これまでのガイドラインとの整合性を担保することで、事業活動を支える AI ガバナンスの仕組みとして、連続性がある発展を遂げていくことが期待される。

< 図 1:「AI 事業者ガイドライン」の策定方針 >



諸外国においても、各種規制及びガイドラインの策定等が積極的に議論されている。本ガイドラインにおいても、諸原則や規制動向等との整合を意識しつつ検討を行った。主な諸原則は図 2 のとおりである。

< 図 2：AI に関連する主な諸原則等>



上記のうち、諸外国において主なものは下記のとおりである。

<u>「信頼性を備えた AI のための倫理ガイドライン」 (ETHICS GUIDELINES FOR TRUSTWORTHY AI)</u>

- 策定主体
  - European Commission (High Level Expert Group on AI (HLEG))
- 策定時期
  - 2019 年 4 月
- 主な構成
  - エグゼクティブサマリ
  - 導入
  - 信頼ある AI に関する枠組み (A FRAMEWORK FOR TRUSTWORTHY AI)
    - ✧ 信頼ある AI の基礎 (Foundations of Trustworthy AI)
    - ✧ 信頼ある AI の実現 (Realising Trustworthy AI)
      - 信頼ある AI の要件 (Requirements of Trustworthy AI)
      - 実現のための技術的/非技術的な手法(Technical and non-technical methods to realise Trustworthy AI)
    - ✧ 信頼ある AI の評価 (Assessing Trustworthy AI)
  - 結論
- 位置づけ/概要
  - 本ガイドラインの目的は、信頼できる AI (Trustworthy AI)を推進すること

- 信頼できる AI には、(1) 適法 (lawful) であること、適用されるすべての法律と規制を遵守すること、(2) 倫理的 (ethical) であること、 (3)堅牢 (robust)であること、 の 3 つの要素があり、この 3 つの構成要素が調和し、重複して動作することが理想的であるとしている

- URL
  - Ethics guidelines for trustworthy AI | Shaping Europe's digital future (europa.eu)

「人工知能に関する理事会勧告」(Recommendation of the Council on Artificial Intelligence)

- 策定主体
  - OECD
- 策定時期
  - 2019 年 5 月 (2023 年 11 月更新)
- 主な構成
  - 本ガイドライン検討の背景/目的
  - 用語の定義
  - 加盟国への勧告事項
    - 相互に補完/一体的に考慮されるべき各原則 (包括的成長、人間中心等)
    - 国家政策/国際協力に関する勧告事項
- 位置づけ/概要
  - 下記の 5 つの原則を相互に補完し、一体的に考慮されるべきものとして提示
    - 包括的成長、持続可能な開発、幸福(Inclusive growth, sustainable development and well-being)
    - 人間中心の価値観と公正さ(Human-centred values and fairness)
    - 透明性と説明可能性(Transparency and explainability)
    - 堅牢性、セキュリティ、安全性(Robustness, security and safety)
    - アカウンタビリティ(Accountability)
- URL
  - RECOMMENDATION OF THE COUNCIL ON ARTIFICIAL INTELLIGENCE (oecd.org)

人工知能の倫理に関する勧告 (Recommendation on the Ethics of Articifial Intelligence)

- 策定主体
  - UNESCO
- 策定時期
  - 2021 年 11 月
- 主な構成
  - 前文
  - 適用範囲
  - 目的
  - 価値と原則

- 政策の範囲
- モニタリングと評価
- 勧告の利活用、促進

- 位置づけ/概要
  - AI システムを人類、個人、社会、環境・生態系のために機能させ、危害を防止するための基礎を提供することや、平和的利用を促進することを目的とする
  - ジェンダー平等と環境・生態系の保護という包摂的な問題に強い重点を置き、具体的な政策勧告を通じて、価値と原則の明確化だけでなく、その実践的な実現にも焦点を当てた、世界的に受け入れられる規範的手段をもたらすことを目指す
  - 提示されている主な内容は下記のとおり
    - ✧ 価値 (Values)
      - 人権および基本的自由ならびに人間の尊厳の尊重、保護および促進(Respect, protection and promotion of human rights and fundamental freedoms and human dignity)
      - 環境と生態系の繁栄(Environment and ecosystem flourishing)
      - 多様性と包括性の確保(Ensuring diversity and inclusiveness)
      - 平和的、公正かつ相互連結的な社会における生活(Living in peaceful, just and interconnected societies)
    - ✧ 原則 (Principles)
      - 比例と危害を及ぼさないこと (Proportionality and Do No Harm)
      - 安全性とセキュリティ (Safety and security)
      - 公正と非差別 (Fairness and non-discrimination)
      - 持続可能性 (Sustainability)
      - プライバシーの権利とデータ保護 (Right to Privacy, and Data Protection)
      - 人間の監視と判断 (Human oversight and determination)
      - 透明性と説明可能性 (Transparency and explainability)
      - 責任とアカウンタビリティ(Responsibility and accountability)
      - 認識とリテラシー (Awareness and literacy)
      - マルチステークホルダー及び適応的ガバナンス、協働 (Multi-stakeholder and adaptive governance and collaboration)
- URL
  - Ethics of Artificial Intelligence | UNESCO

AI 権利章典の青写真 (Blueprint for an AI Bill of Rights)

- 策定主体
  - 米国ホワイトハウス (the White House Office of Science and Technology Policy)
- 策定時期
  - 2022 年 10 月
- 主な構成

- 前文
- 本フレームワークの位置づけ
- 5 つの原則
- 青写真の権利章典への適用に関する留意事項
- 「原則から実践へ」(From Principles to Practice)
  - 各原則の問題点・実践方法・実践事例をまとめた技術文書としての位置づけ

- 位置づけ/概要
  - テクノロジー・データ・自動化システムの利用は、今日、民主主義に突きつけられている大きな挑戦のひとつであり、アメリカ国民の権利を脅かしていると認識
  - これらの脅威から人々を守り、価値を強化すべくテクノロジーを活用する社会のための指針として策定
  - 5 つの原則は下記のとおり
    - 安全・効果的なシステム (Safe and effective systems)
    - アルゴリズムに由来する差別からの保護 (Algorithmic discrimination protections)
    - データプライバシー (Data privacy)
    - 通知と説明 (Notice and Explanation)
    - 人による代替、配慮、フォールバック (Human alterantives, consideration, and fallback)
- URL
  - Blueprint for an AI Bill of Rights (whitehouse.gov)

AI リスクマネジメントフレームワーク (Aritificial Intelligence Risk Management Framework : AI RMF 1.0)

- 策定主体
  - 米国 NIST
- 策定時期
  - 2023 年 1 月
- 主な構成
  - エグゼクティブサマリー
  - 基礎情報
    - リスクの枠組み
    - 想定読者
    - AI リスクと信頼性
    - AI RMF の効果
  - 具体的内容 (「AI RMF コア」と「AI RMF プロファイル」)
- 位置づけ
  - AI のリスク管理は、責任ある AI システムの開発と利用の重要な要素。責任ある AI の実践は、AI システムの設計、開発、使用に関する決定を、意図された目的と価値観に合致させるのに役立つ

- このため、社会が AI から恩恵を受けると同時に、潜在的な害悪から保護されることを目的に、実用的なものとして、AI 技術の発展に応じた AI の状況にも適応できるよう意図して策定。また、様々な程度と能力を持つ組織によって運用されることを意図
- 信頼できる AI システム (Trustworthy AI systems) は、下記の要素を含むと整理
  - 有効かつ頼ることができる (Valid and reliable)
  - 安全である (Safe)
  - セキュアであり強靭である (Secure and resilient)
  - 説明可能かつ解釈可能である (Explainable and interpretable)
  - プライバシーが保護されている (Privacy-enhanced)
  - 公正であり有害なバイアスが管理されている (Fair – with harmful bias managed)
  - 説明責任があり透明性がある (Accountable and transparent)

- URL
  - Artificial Intelligence Risk Management Framework (AI RMF 1.0) (nist.gov)

② 活用事例集

本ガイドラインを検討する過程で、多くの有識者の方々から、その活用を一層促すための工夫が必要ではないか、とご指摘をいただいた。このため、経済産業省/総務省とも連携しつつ、AI ガバナンスに積極的に取り組む事業者について、その取組についてとりまとめガイドラインに掲載して紹介することが効果的と考えられるため、本ガイドラインの別添 (付属資料) において、「コラム」として掲載することとなった。

具体的には、本ガイドラインの土台になった経済産業省「AI 原則実践のためのガバナンスガイドライン Ver. 1.1」等に依拠しつつ、ガバナンスに関する取組を進めている事業者を、両省と連携しつつ選定し、当該事業者の取組内容を確認の上、後段のとおり取りまとめている。

本ガイドラインの読者におかれては、このような事例も参照しつつ、自身に適した体制構築や取組を進めることが期待される。

以下、AI 事業者ガイドラインに掲載されている取組の内容のうち、一部を抜粋する (ガイドラインの登場順にて掲載)。本報告書に掲載されているもの以外にも事例は存在するため、ガイドライン 別添も適宜ご参照いただきたい。

(1) NEC グループの AI ガバナンスに関する取組

NEC は、2018 年に AI の利活用に関連した事業活動が人権を尊重したものとなるよう、全社戦略の策定・推進を担う組織として「デジタルトラスト推進統括部」を設置し、2019 年に「NEC グループ AI と人権に関するポリシー（以下、全社ポリシー）」を策定。ガバナンス体制として AI ガバナンス遂行責任者（CDO: Chief Digital Officer）を置き、リスク・コンプライアンス委員会や取締役会との関係性を明確化するとともに、外部有識者会議の「デジタルトラスト諮問会議」を設置し、外部とも積極的に連携する等、経営アジェンダとして AI ガバナンスに取り組んでいる（「図 1. AI ガバナンスの推進体制」参照）。

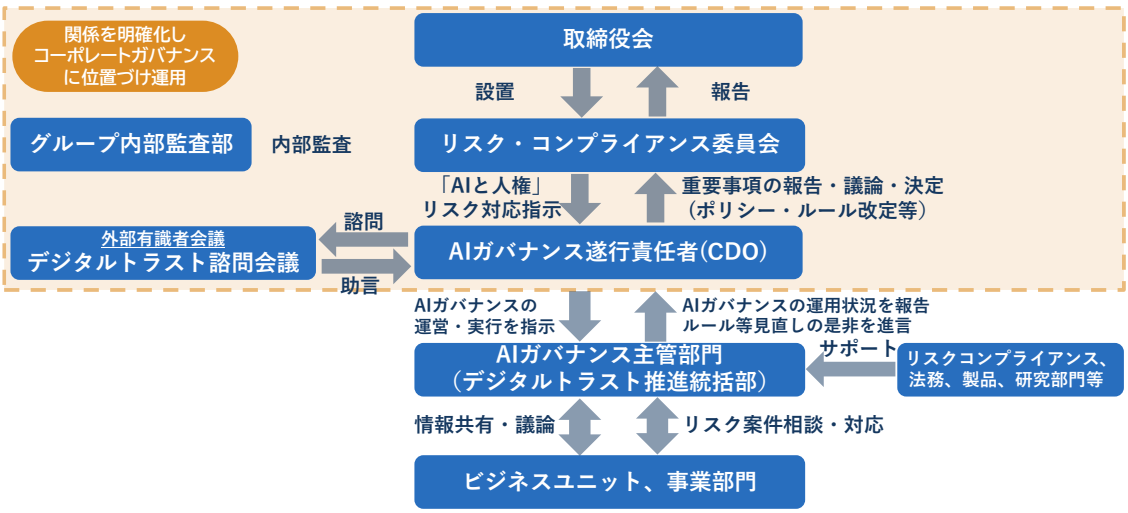


図 1. AI ガバナンスの推進体制

全社ポリシーは、デジタルトラスト推進統括部が国内外の原則や自社のビジョン・価値・事業内容等から検討し、社内の研究開発・サステナビリティ・リスク管理・マーケティング・事業部門などの関係部門や、外部の有識者・NPO・消費者など社内外の様々なステークホルダーとの対話を経て 2019 年 4 月に策定された。このポリシーは、AI 利活用においてプライバシーへの配慮や人権の尊重を最優先するために策定されたものであり、「公平性」、「プライバシー」、「透明性」、「説明する責任」、「適正利用」、「AI の発展と人材育成」、「マルチステークホルダーとの対話」の 7 つの項目から構成されている。

全社ポリシーを実践するために、デジタルトラスト推進統括部が中心となり、社内制度の整備や従業員の研修などを実施している。具体的には、ガバナンス体制や遵守すべき基本的事項が定められた全社規定、対応事項や運用フローが定められたガイドラインやマニュアル、リスクチェックシートを整備し、AI 利活用へのリスクチェックと対策を企画・提案フェーズからフェーズ毎に実施する枠組みを整えている。

また、全従業員向けの Web 研修や AI 事業関係者・経営層向けの研修も実施しており、外部有識者を講師に迎え、最新の市場動向やケーススタディも交えることで理解の促進が図られている（「図 2. AI ガバナンスの全体像」参照）。

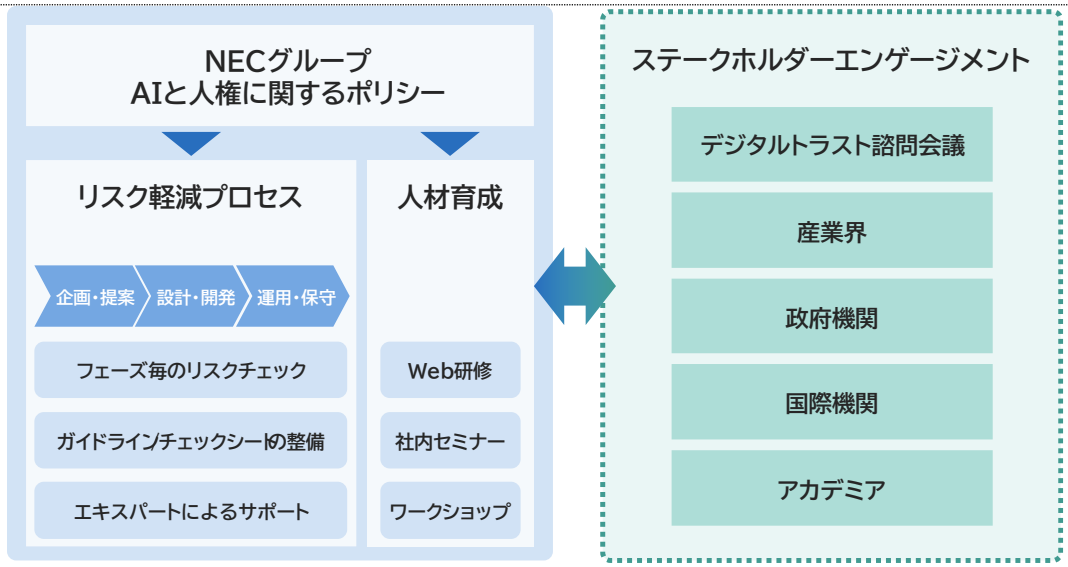


図 2. AI ガバナンスの全体像

これらの取組に際し、経済産業省「AI 原則実践のためのガバナンス・ガイドライン」（以下、旧「AI ガバナンス・ガイドライン」という）に掲載された 21 個の行動目標に対して、5 段階の「成熟度」を定義し、AI ガバナンスの現状を可視化することで、目標達成に向けた対応事項の設定や進捗管理に活用している（「図 3. 旧「AI ガバナンス・ガイドライン」の活用法」参照）。また、旧「AI ガバナンス・ガイドライン」のアジャイル・ガバナンスの考え方にもとづいて、社会環境の変化に応じた対応や社内ルール・運用の見直しを柔軟に行っている。2023 年には、生成 AI（大規模言語モデル）の社内利用に関するルール整備を行うなど、積極的な活用が進められている。

成熟度 / 行動目標	Lv.1 Performed 個々人による単発的な実施	Lv.2 Managed ポリシーに従った反復的実施	Lv.3 Defined 統一された標準的プロセスの確立	Lv.4 Measured 定量的な評価の実施	Lv.5 Optimized フィードバックに基づく継続的最適化
1-1. AIシステムから得られる正のインパクトだけではなく…	国内及び同業他社に関するAI利用のガイドライン…				
1-2. 本格的なAIの提供に先立ち、直接的なステークホルダー…	政府、市民団体等が公表している消費者アンケートや、AI利用に…				
…	…				
6-1. 行動目標1－1から1－3について、適時に再評価…	重大な「ヒヤリ・ハット」が生じた場合、特定のインシデントへ…				

図 3. 旧「AI ガバナンス・ガイドライン」の活用法

(2) 東芝グループの AI ガバナンスに関する取組

同社は、2022 年にグループ横断の AI 施策を先導する「AI-CoE プロジェクトチーム」を発足させ、グループの経営理念体系を AI 利活用の点から具体化した「AI ガバナンスステートメント」を策定。旧「AI ガバナンス・ガイドライン」を参照しつつ、このステートメントをベースとした「AI ガバナンス」を構築している。この枠組みの中で、「AI-CoE プロジェクトチーム」を中心に、プライバシー、セキュリティ、法務などの様々な分野専門家、および、事業側からの代表者を集めたワーキンググループを形成し、「AI ガバナンス」を推進する活動をしている。

具体的には、「AI 技術カタログ」の構築によるグループ保有の AI 技術資産の見える化/利活用促進や独自教育プログラムによる AI 人材育成に加え、「MLOps」(機械学習モデルのライフサイクルを管理する仕組み)や「AI 品質保証の仕組み」の整備等を通じて、AI システムの品質を保つ仕組みに取り組んでいる（「図 1. グループの AI ガバナンスの概要」参照）。

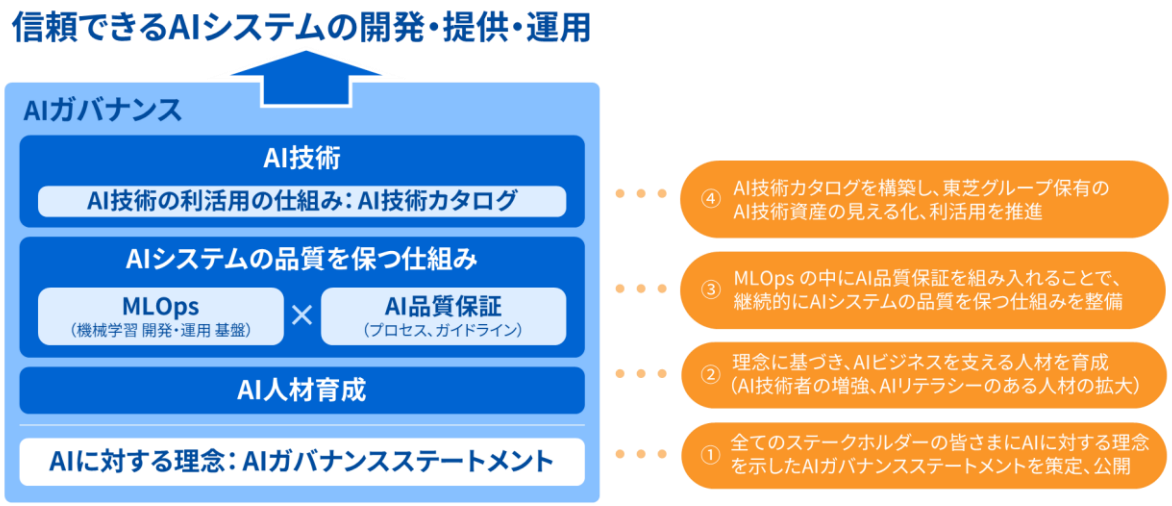


図 1. グループの AI ガバナンスの概要

この「AI ガバナンスステートメント」は、東芝グループの経営理念体系を反映し、AI に対する理念を明文化することを目的として、「人間尊重」、「安全安心の確保」、「コンプライアンスの徹底」、「AI の発展と人材の育成」、「持続可能な社会の実現」、「公平性の重視」、「透明性と説明責任の重視」の 7 つの要素から構成されている。

このステートメントをベースとし、「AI 品質保証」と「MLOps」の二軸により、AI システムの品質を保つ仕組みが構築されている。「AI 品質保証」では、「AI 品質保証ガイドライン」を策定し AI システムの開発における考え方や取り組むべき事項を整理し、当ガイドラインを踏まえた「AI 品質保証プロセス」で必要な作業や作成すべき成果物を特定し、漏れのないプロセスを整理している。また、「品質カード」を通じ、開発者目線になりがちな AI 品質保証を利用者目線で評価し、AI 品質の可視化に取り組んでいる。

「MLOps」では、ビジネス、機械学習の専門家、システム開発担当者、システム運用担当者が一体のチームとなり、運用開始後の環境変化による性能劣化などを起こさないよう、AI システムの継続的な改善に取り組んでいる。これらを連携させることで、信頼できる AI システムの開発・提供・運用を実践している。

これら AI ガバナンスの取組を始めたことにより、AI の専門家（技術者）だけでなく、東芝グループ全体で AI システムの開発・提供・運用に必要なリテラシーの向上（AI 利用に対する機会だけでなく、リスク意識の向上）がみられている。

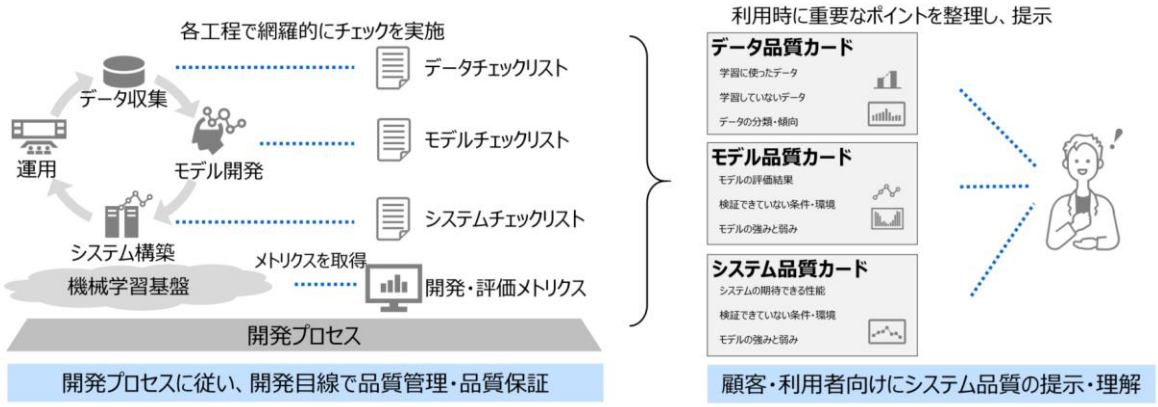


図 2. AI 品質保証ガイドライン・品質カード活用の流れ

(3) パナソニックグループの AI ガバナンスに関する取組

同社では、2019 年に、旧パナソニック(株)内に AI 倫理委員会を設置し、社内で遵守すべき「AI 倫理原則」の策定を行った。2022 年にはグループ横断で「AI 倫理原則」を運用するための組織として「パナソニックグループ AI 倫理委員会」へ改組し、同年に「パナソニックグループの AI 倫理原則」を公表した。現在は、この AI 倫理委員会が中心となり、2022 年より運用開始の「AI 倫理チェックシステム」の開発・活用や、全従業員向け AI 倫理教育などを行っている（「図 1. AI ガバナンスの体制　概要」参照）。

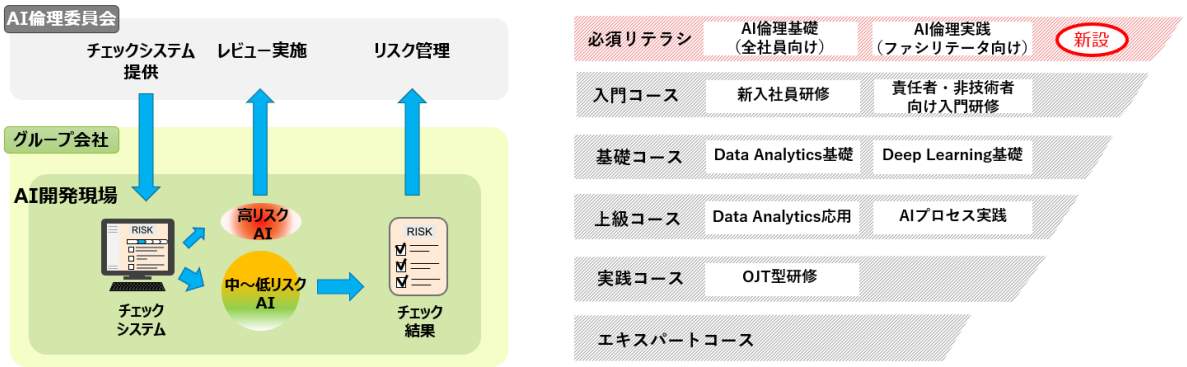


図 1. AI ガバナンスの体制　概要

「AI 倫理委員会」は、パナソニック ホールディングス(株)内に設置され、「AI 倫理原則」の公表に加え、広範な事業領域それぞれにおいて利用者・社会から信頼される活動実践に取り組んでいる。具体的には、グループの全事業会社から 1 人以上の AI 倫理担当者を選出し、法務部門や知財部門、情報システム／セキュリティ部門、品質部門とともに、グループ横断的に AI 倫理推進体制を敷いている（「図 2. AI 倫理委員会構造」参照）。パナソニックグループの多岐に渡る事業分野に対応するため、それぞれの AI 倫理担当者が事業会社グループ内の AI 倫理活動を推進し、AI 倫理委員会がそれらを支援するという形で運用されている。

図 2. AI 倫理委員会構造

AI 倫理委員会の取組の一つとして、「AI 倫理チェックシステム」が開発されている。これは、グループ内の多岐・広範囲にわたる AI 利活用において、現場負担の増大・イノベーション阻害を防ぎつつ、AI 倫理リスクチェックを効率的・効果的に行うことを目的としている。製品・サービスの特性に合わせて必要十分なチェックリストを生成できるシステムとなっており、開発中の AI が、AI 倫理原則に乖離したものになっていないか確認できる。また、各チェック項目に対して、充実した解説や対応策に関する情報・技術・ツールの提供を行い、現場が主体的に AI 倫理をチェックし改善を進められる仕組みとなっている。セルフチェック結果は集約され、「AI 倫理委員会」で内容分析が行

われ、活動に反映される。チェック項目は、経済産業省の旧「AI ガバナンス・ガイドライン」をベースに国内外のガイドラインを踏まえて初版が作成され、実運用後は現場からの意見を反映した改訂が随時行われている（「図 3．AI 倫理チェックシステム」参照）。

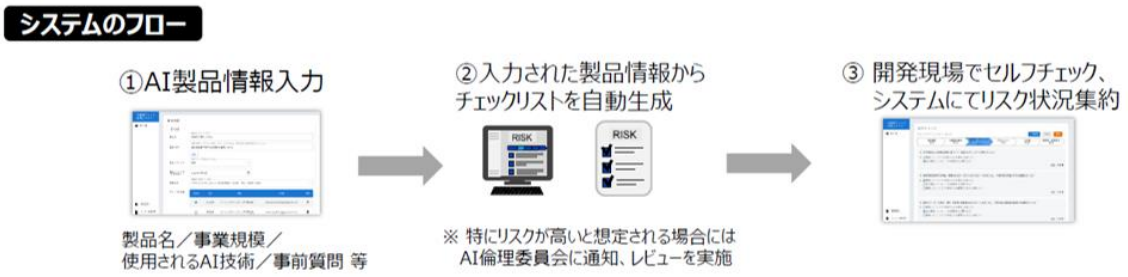


図 3. AI 倫理チェックシステム

(4) 富士通グループの AI ガバナンスに関する取組

同社は、AI の開発・提供企業として、AI に関する懸念や予期せぬ不都合の解消と適切な技術活用による持続可能な社会の創造をその責務としており、国際的な AI 倫理の議論への積極的な参加に加え、後述の「AI 倫理チェック」や海外リージョンへの AI 倫理責任者設置などの先進的な社内ガバナンスの取組を推進しつつ、AI ガバナンスの取組紹介や生成 AI 利活用ガイドラインの公開などの社外への AI 倫理普及の取組にも力を入れる。

2018 年に加盟した欧州コンソーシアム「AI4People」が提案する 5 原則を参考に、2019 年に「富士通グループ AI コミットメント」を策定、さらにその実践のために、AI の利活用方法に応じた具体的な判断基準や手順を整備した（「図 1. 富士通グループコミットメント」参照）。また、AI ガバナンスの取組に対する客観的な評価を得るべく、「富士通グループ AI 倫理外部委員会」を設置し、AI 技術のほか、生命医学、生態学、法学、SDGs、消費者問題など多様性を重視した外部の専門家をその委員として招聘している。社長をはじめ経営陣がオブザーバーとして参加する同委員会における活発な議論を提言としてとりまとめ、これを取締役会へと共有することで、AI 倫理を「企業経営上の重要課題」としてコーポレートガバナンスに組み込んでいる。

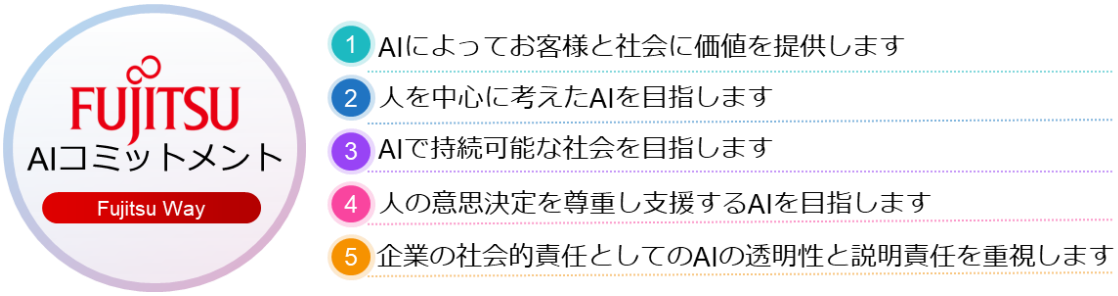


図 1. 富士通グループコミットメント

2020 年から AI 倫理に関する教育を制度化した結果、従業員の意識レベルが飛躍的に向上し、また、コンサルティングサービスとして AI ユーザー企業への倫理的観点の助言が可能になっている。

2022 年には、AI 倫理はグループ全体の経営課題との認識から、AI 倫理戦略を主導する組織として「AI 倫理ガバナンス室[1]」を会社直下（コーポレート部門）に設置した。その際、開発や営業出身者など幅広い職種の経験者をグループ各所から登用し、また、デジタルネイティブ世代が活躍できるようオープンで個人の意見が尊重される場づくりを行っている。同室では、率直な意見交換や提案が日常的に行われ、ここから生み出される各種 AI 倫理浸透策はグループ全体で推進される（「図 2. AI 倫理ガバナンスの体制」参照）。

[1] 詳細は、富士通の AI 倫理ガバナンスに関する専門サイトに掲載のホワイトペーパー『「富士通グループ AI 倫理外部委員会」による提言及び富士通の実践例』を参照のこと。
https://www.fujitsu.com/jp/about/research/technology/ai/aiethics/

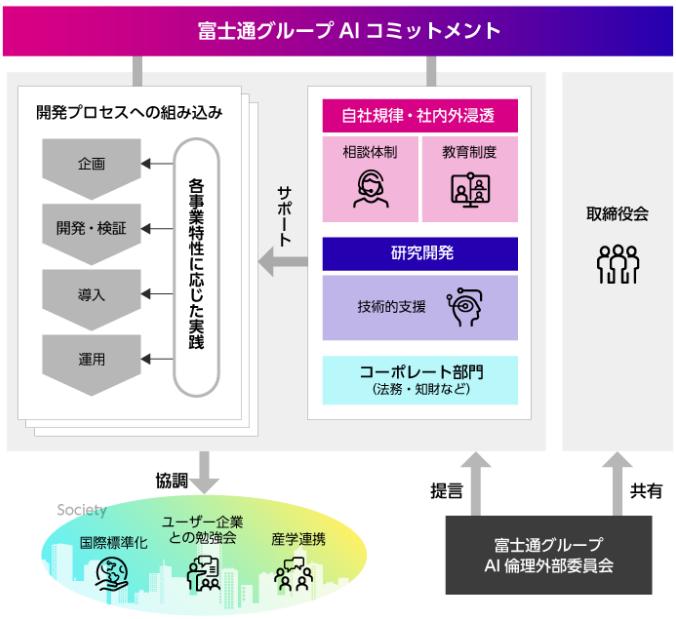


図 2. AI 倫理ガバナンスの体制

さらに、2023 年、「AI 倫理チェック」の義務化対象を日本国内全商談に拡大し、倫理的課題を抱える案件は法務・研究開発・DE＆I・事業部門などによる合議を経てその推進・改善を判断することとし、AI に関する品質保証やセキュリティの観点に留まらない統制を徹底している。海外については、本社での倫理審査に加えて、現地での AI 実装時点での倫理チェックを行うべく、各リージョンに AI 倫理責任者を配置している。

加えて、「AI 倫理影響評価」の無料公開などを通じ、安全安心で信頼できる AI の開発・提供の社内外の推進に取り組んでいる。「AI 倫理影響評価」は、内閣府、総務省、経済産業省で公表されたガイドラインに加え、OECD や EU、米国の指針等も含め、国内外各種 AI 倫理ガイドラインに準拠する上で、AI システムの開発者・運用者に関連する項目を抽出し、AI が人や社会に与える倫理的な影響を評価するために策定された。この公開に加え、ユーザー企業との勉強会、産学連携や標準化活動などにより、社会全体への AI 倫理取り組みの浸透を促進している（「図 3. AI 倫理影響評価の概要」参照）

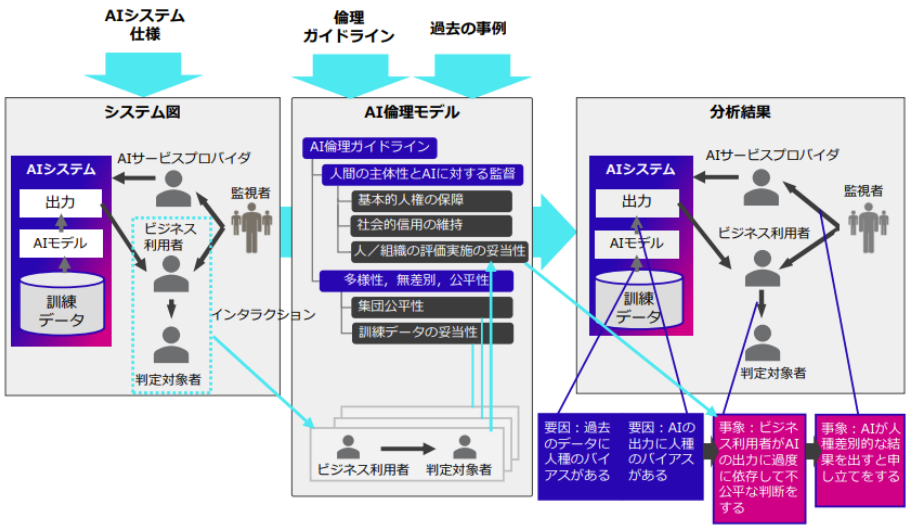


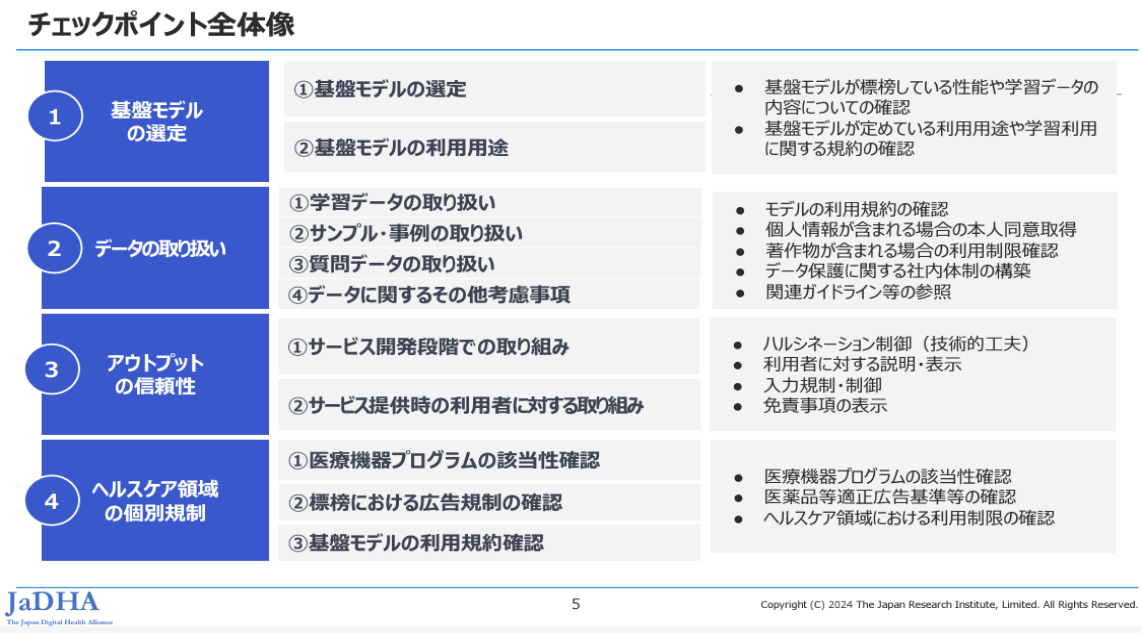
図 3. AI 倫理影響評価の概要

(5)　日本デジタルヘルス・アライアンス「ヘルスケア事業者のための生成 AI 活用ガイド」

日本デジタルヘルス・アライアンス[2]（以下、「JaDHA」という。）は、ヘルスケアサービスを提供する事業者が生成 AI による多様なサービスを創出し、AI 利用者及び業務外利用者が安心してサービス選択できる環境を構築することを目的に、2024 年 1 月 18 日に「ヘルスケア事業者のための生成 AI 活用ガイド（ヘルスケア領域において生成 AI を活用したサービスを提供する事業者が参照するための自主ガイドライン）」[3]を策定した。

本ガイドは、AI 事業者ガイドラインに全業種共通の内容が盛り込まれることを前提としつつ、ヘルスケア領域が他の領域と比較して要配慮個人情報の取扱いが多くなる点や、不確かな情報がもたらす個人への影響が極めて大きい点等を踏まえ検討を行い、AI 提供者が、ヘルスケア領域において、生成 AI を活用して安全安心に AI サービスの提供を行うためのチェックポイントをまとめている。

具体的には、生成 AI を取り巻く主体やバリューチェーンの整理を行ったうえで、AI 事業者ガイドラインで示されている「3)公平性」や「4)プライバシー保護」に関する AI 提供者として留意すべき各場面でのデータの取扱いについて、より実践的かつ具体的な注意事項を定めている。加えて、「7) アカウンタビリティ」を踏まえた AI 利用者及び業務外利用者への説明・表示等について体系的にとりまとめられているとともに、AI 提供者が活用可能なチェックリストや参考事例なども併せて公表している。



（出典：JaDHA「ヘルスケア事業者のための生成 AI 活用ガイド」）

JaDHA では、本ガイドを業界内でいち早く策定することで、新技術である生成 AI を活用したサービス推進や業界内でのイノベーション促進を期待している。加えて、スタートアップ企業や中小企業をはじめとする生成 AI を活用したサービスを検討している企業が本ガイドを参照・セルフチェックを行うことで、AI 利用者及び業務外利用者が安心して AI サービスを利用できる環境構築を目指している。

[2] 日本におけるデジタルヘルス産業の発展や課題を検討する業界団体として 2022 年 3 月に設立。現在、医薬品・医療機器メーカーからヘルステックスタートアップ企業まで、多様な属性の企業が参画している。

[3] https://jadha.jp/news/news20240118.html

なお、急速な新技術の進歩に伴い、生成 AI の技術特性や関連制度等の変化が想定されることから、本ガイドは適宜見直しを図り、ヘルスケア領域において AI を活用する事業者がタイムリーに適正な AI サービスを提供できるよう今後も業界全体の後押しを行っていく予定である。

以上