

報告書 2022

～ 「安心・安全で信頼性のある AI の社会実装」の更なる推進 ～

2022 年 7 月 25 日

AI ネットワーク社会推進会議

目次

はじめに	1
第1章 AIネットワーク化をめぐる最近の動向	4
1. 国内の動向	
2. 海外の動向	
3. 国際的な議論の動向	
4. 国際標準化に関する動向	
5. 国際シンポジウム「AIネットワーク社会フォーラム 2022」	
第2章 「安心・安全で信頼性のあるAIの社会実装」の推進の取組 (取組事例の取りまとめ)	27
1. ヒアリングの概要	
2. ヒアリングにおける発表・意見交換のポイント	
第3章 AI開発ガイドライン及びAI利活用ガイドラインに関するレビュー	47
1. AI開発ガイドライン及びAI利活用ガイドラインに関するレビュー	
2. AI開発ガイドライン及びAI利活用ガイドラインの見直しに関する論点の整理	
第4章 今後の取組	60
1. AI倫理・ガバナンス	
2. 取組事例の周知・共有	
3. 人材育成	
結びに代えて	65

- <別紙1> 国際シンポジウム「AIネットワーク社会フォーラム 2022」の概要
- <別紙2> AI開発ガイドライン及びAI利活用ガイドラインに関するレビュー（詳細）
- <別紙3> AIネットワーク社会推進会議及びAIガバナンス検討会構成員一覧

<別冊>

「安心・安全で信頼性のあるAIの社会実装」に関するグッドプラクティス集（取組事例集）

はじめに

私達の生活において、人工知能（Artificial Intelligence：A I）が使われることが多くなってきている。また、企業活動においても、A Iを使って、モノを生産したり、サービスを提供したりすることが増えてきており、A Iの社会実装が進んできている。

総務省では、2016年10月から「A Iネットワーク社会推進会議」（以下「本推進会議」という。）を開催し、A Iネットワーク化に関する社会的・経済的・倫理的・法的課題について検討を行ってきた。本推進会議の検討の成果として、2017年7月に「国際的な議論のためのA I開発ガイドライン案」（以下「A I開発ガイドライン」ということがある。）、2018年7月に「A I利活用原則案」、2019年8月に「A I利活用ガイドライン」を取りまとめ、公表した。その後、本推進会議（A Iガバナンス検討会を含む場合がある。以下同じ。）において、「安心・安全で信頼性のあるA Iの社会実装」¹をテーマに掲げ、A Iの開発者やサービスプロバイダー、ビジネス利用者、消費者的利用者などのステークホルダー²からヒアリングを行い、その取組等を取りまとめ、2020年7月に「報告書2020」、2021年8月に「報告書2021」を公表した³。

また、2016年4月に日本で開催されたG7⁴香川・高松情報通信大臣会合を契機として、G7やOECD（経済協力開発機構）⁵などにおいて国際的な議論が活発に行われ、2019年5月にOECDにおいて「A Iに関する理事会勧告」が採択されるとともに、同年6月にG20⁶において「G20 A I原則」が採択された。その後、OECDにおいて、A Iに関する取組の情報共有や社会実装に向けた検討が進められているほか、GPAI（Global Partnership on A I）⁷やUNESCO（国際連合教育科学文化機関、以下「ユネスコ」ということがある。）⁸などにおいても、A Iの倫理・ガバナンス等に関する国際的な議論が行われている。

さらに、国内では、「第5期科学技術基本計画」（平成28年1月22日閣議決定）⁹において、我が国の

¹ ここで、「安心」とは、個人の主観的な判断に依存するものであるが、その前提として信頼の醸成が必要とされ、社会的に合意されるレベルの安全を確保しつつ信頼が築かれる状態のことをいう。「安全」とは、生命・身体・財産に損害がないと客観的に判断される状態のことをいう（文部科学省「安全・安心な社会の構築に資する科学技術政策に関する懇談会」報告書 第2章における記述等を参考に記載。）。なお、本推進会議で取りまとめた「国際的な議論のためのA I開発ガイドライン案」及び「A I利活用ガイドライン」の「安全の原則」において、『利用者及び第三者の生命・身体・財産に危害を及ぼすことがないように配慮する。』とされている。

² ステークホルダーの区分については、「A I利活用ガイドライン」参照。

³ ヒアリングによる取組等の取りまとめのほか、「報告書2020」においては、A Iネットワーク化の進展に伴い形成されるエコシステムの展望、「報告書2021」においては、新型コロナウイルス感染症とA I利活用に関する検討を整理している。

⁴ G7サミット（主要国会議）は、フランス、米国、英国、ドイツ、日本、イタリア、カナダの7箇国及び欧州連合（EU）の首脳が参加して毎年開催される国際会議のこと（国名は議長国順）。

⁵ 最近の動向については、第1章3.（1）参照。

⁶ G20サミット（金融・世界経済に関する首脳会合）は、G7各国に、アルゼンチン、オーストラリア、ブラジル、中国、インド、インドネシア、メキシコ、韓国、ロシア、サウジアラビア、南アフリカ、トルコ及び欧州連合（EU）首脳が参加して毎年開催される国際会議のこと（国名はアルファベット順）。

⁷ 人間中心の考え方に立ち、「責任あるA I」の開発・利用を実現するために、価値観を共有する政府、国際機関、産業界、有識者等からなる官民国際連携組織で、2020年6月に創設された。最近の動向については、第1章3.（2）参照。

⁸ 最近の動向については、第1章3.（5）①参照。

⁹ 次に掲げるURLのウェブサイトに掲載。

<<https://www8.cao.go.jp/cstp/kihonkeikaku/5honbun.pdf>>

目指すべき未来社会の姿として、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により、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」¹⁰が提唱され、「第6期科学技術・イノベーション基本計画」（令和3年3月26日閣議決定）¹¹において、『Society 5.0の実現に向け、サイバー空間とフィジカル空間を融合し、新たな価値を創出することが可能となるよう、質の高い多種多様なデータによるデジタルツインをサイバー空間に構築し、それを基にAIを積極的に用いながらフィジカル空間を変化させ、その結果をサイバー空間へ再現するという、常に変化し続けるダイナミックな好循環を生み出す社会へと変革することを目指す。』とされている¹²。

このような状況を踏まえて、本推進会議において、「報告書2021」の公表後も、「安心・安全で信頼性のあるAIの社会実装」を推進するために、引き続き、AIの開発者やサービスプロバイダー、ビジネス利用者などのステークホルダーからヒアリングを行い、AIの社会実装に関する取組事例を収集、蓄積してきた。また、AI開発ガイドラインやAI利活用ガイドラインを取りまとめた後数年が経過し、この間に国内外において多くのAI倫理・ガバナンスに関する原則、指針、ガイドライン等が策定されていることや「報告書2021」において両ガイドラインの射程を超えて、「安心・安全で信頼性のあるAIの社会実装」に関する優れた取組であると評価することができるものが示されたことなどを踏まえて、AI開発ガイドライン及びAI利活用ガイドラインのレビューを行ってきた。さらに、サイバー空間とフィジカル空間の融合やデジタルツイン等に関する検討を深めるため、2回にわたってヒアリング¹³を行ったほか、将来的なAIの開発・利活用を見据えて、量子コンピュータとAIに関する検討を深めるためのヒアリング¹⁴を行った。加えて、EU（欧州委員会）が2021年4月に公表した「人工知能に関する調和の取れたルールを定める規則の提案」に関する研究を深めるため、3回にわたってヒアリング¹⁵を行った。

本報告書は、AIネットワーク化をめぐる国内外の動向及び国際的な議論の動向等を概観するとともに、AIの社会実装に関して先進的あるいは意欲的な取組を行っている事業者、有識者及び関係団体からのヒアリングにより収集した取組事例を紹介している。さらに、AI開発ガイドライン及びAI利活用ガイドラインのレビューを行い、その結果等を踏まえて、見直しに関する論点を整理しつつ、今後の取組等を整理している。

本報告書において紹介している取組事例は、次表のとおりである。

¹⁰ サイバー空間（仮想空間）とフィジカル空間（現実空間）を融合させたシステムのことを「Cyber-Physical System：CPS」という。

¹¹ 次に掲げるURLのウェブサイトに掲載。

<<https://www8.cao.go.jp/cstp/kihonkeikaku/6honbun.pdf>>

¹² この点に関し、AIネットワーク化の進展を通じて目指すべき社会像として本推進会議が提唱した「智連社会」を形成するに当たってのつとるべき基本理念として『物理空間とサイバー空間の調和』が掲げられており、『AIネットワークを活用して物理空間とサイバー空間を連結し、両者の調和を図ることにより、ヒト・モノ・コト相互間の空間を超えた協調の実現を可能とすること。』とされている。本推進会議の前身の会議体である「AIネットワーク化検討会議」が取りまとめた「中間報告書」（2016年4月）15-16頁及び「報告書2016」（2016年6月）注（43）並びに本推進会議が取りまとめた「報告書2017」（2017年7月）7-8頁及び「報告書2018」（2018年7月）8-9頁参照。

¹³ 本推進会議 AIガバナンス検討会 合同会議（2022年2月8日及び同年4月27日）。

¹⁴ 本推進会議 AIガバナンス検討会 合同会議（2021年10月27日）。藤井啓祐氏（大阪大学基礎工学研究科教授）による発表資料については、次に掲げるURLのウェブサイトに掲載。

<https://www.soumu.go.jp/main_content/000775388.pdf>

¹⁵ 本推進会議 AIガバナンス検討会 合同会議（2021年10月27日、2022年2月8日及び同年4月27日）。また、ヒアリングの内容等については、第1章2.（参考）参照。

○ ソフトバンク株式会社	: ソフトバンク AI 活用事例ご紹介
○ パナソニック株式会社	: パナソニックにおける AI 開発の考え方と活用事例
○ シャープ株式会社、株式会社 AIoT クラウド	: シャープグループの AI への取り組み
○ 中山敬一（九州大学生体防御医学研究所教授）	: 人工知能（AI）の医学生物学への応用
○ 植田琢也（東北大学医療 AI 人材育成拠点プログラム）	: Global×Local な医療課題解決を目指した最先端 AI 研究開発人材育成教育拠点 – Clinical AI –
○ 株式会社メルカリ	: メルカリグループにおける AI に関する取り組みのご説明
○ 一般社団法人日本データマネジメント・コンソーシアム	: 実務者が考える AI・データ活用における倫理フレームワーク
○ 三井住友海上火災保険株式会社	: 三井住友海上のデジタルイノベーション推進取組
○ ダイキン工業株式会社	: ダイキン工業における AI 人材育成の取り組みについて
○ 西日本旅客鉄道株式会社	: JR 西日本におけるデータアナリティクスの取組
○ KDDI 株式会社	: KDDI における AI 事例と AI ガバナンスに向けた取組み
○ 有限責任あずさ監査法人	: AI の適切性検証への取組み
○ 株式会社エヌ・ティ・ティ・データ	: AI に係るサプライチェーンリスクと NTT データの対応
○ 富士通株式会社	: AI サプライチェーンに対する富士通の取組
○ 一般社団法人 AI データ活用コンソーシアム	: データ活用と課題と取組み

（注）事業者等の名称はヒアリング当時のもの。

本報告書において、上記の表に掲げられている事業者、有識者及び関係団体の協力により、AI の社会実装に関する多数の具体的な取組事例を取りまとめることができた。これらの取組事例については、AI の開発・利活用に取り組んでいる者又は取り組むことを検討している者などにとって、非常に有用・有益な情報になるものと考えられる。

協力が得られた事業者、有識者及び関係団体に感謝するとともに、本報告書が、多くの関係者に共有され、AI の開発・利活用の取組の推進に貢献し、「安心・安全で信頼性のある AI の社会実装」が更に推進されることを願いたい。

第1章 AIネットワーク化をめぐる最近の動向

本章においては、本推進会議における検討に際して参考としたAIネットワーク化をめぐる国内外の動向及び国際的な議論の動向等について、「報告書2021」公表以後のものを中心に概観する¹⁶。

1. 国内の動向

(1) AI戦略2022

政府は、2021年10月から2022年2月まで開催した新AI戦略検討会議等における検討を踏まえ、同年4月22日に開催した第11回統合イノベーション戦略推進会議において「AI戦略2022」¹⁷を決定した。前年の「AI戦略2021」からの主な改訂事項は、次のとおりである。

- ・ 社会実装の充実に向けて新たな目標及び取組を設定
- ・ パンデミックや大規模災害等の差し迫った危機への対処のための取組を具体化

【社会実装の推進のための目標】

- AIの信頼性の向上
- AI利活用を支えるデータの充実
- 人材確保等の追加的な環境整備
- 政府におけるAI利活用の推進
- 日本が強みを有する分野とAIの融合

【差し迫った危機への対処のための目標】

- 「国家強靱化のためのAI」の確立
- 「地球強靱化のためのAI」でのリーダーシップの確立
- 「強靱かつ責任あるAI」でのリーダーシップの確立

(2) デジタル社会の実現に向けた重点計画

政府は、2022年6月7日に、「デジタル社会の実現に向けた重点計画」¹⁸を閣議決定した。AIの普及に伴い、広範な分野においてAI利活用に向けた取組が記載されているが、例えば「AIの社会実装に向けた取組の加速」において、次のように記載されている。

- ・ 今後の更なるAIの実用化に向けて、「AI戦略2022」に基づき、ディープラーニングを重要分野として位置付け、企業による実装を念頭において取り組む。¹⁹
- ・ AIの社会実装に際しては、AIを有効かつ安全に利用できる社会（AI-Readyな社会）への変革を推進する必要があるため、「人間中心のAI社会原則」においてまとめられた、AI-Readyな社会における社会的枠組みに関する7つのAI社会原則²⁰を踏まえて取り組む。政府機関におけるAI利活用の推進においても、上記AI社会原則を基に、汎用的なAIの採用促進や推進体

¹⁶ 「報告書2021」公表前のAIネットワーク化をめぐる動向については、「報告書2021」第1章参照。なお、本章における記載については、網羅的なものではなく、総務省の調査等により把握することができた範囲において本推進会議における検討に際して参考としたものを記載している。

¹⁷ 次に掲げるURLのウェブサイトに掲載。

<https://www8.cao.go.jp/cstp/ai/aistrategy2022_honbun.pdf>

¹⁸ 次に掲げるURLのウェブサイトに掲載。

<<https://www.digital.go.jp/policies/priority-policy-program/>>

¹⁹ この点に関し、「責任あるAI」の実現に向けた「説明可能なAI」等の技術開発、AI利活用を支えるデータの充実等に向けたデータの取扱いルールについての再点検などの環境整備、高度AI人材の育成に向けた取組の強化など具体的な取組が記載されている。

²⁰ 「①人間中心の原則」、「②教育・リテラシーの原則」、「③プライバシー確保の原則」、「④セキュリティ確保の原則」、「⑤公正競争確保の原則」、「⑥公平性、説明責任及び透明性の原則」、「⑦イノベーションの原則」の7つの原則を指す。

制の強化等に取り組む。

(3) AIを用いたクラウドサービスに関するガイドブック

総務省は、2022年2月15日に、学識経験者、弁護士及び事業者等の有識者からなる「AIクラウドサービス検討会」における検討等を総合的に踏まえて、AIを用いたクラウドサービスの開発及び提供の際に留意すべき事項を取りまとめ、「AIを用いたクラウドサービスに関するガイドブック」²¹を公表した。

同ガイドブックにおいて、AIを用いたクラウドサービス開発の具体的な流れが体系的に整理され、各プロセスにおける開発者の留意事項がまとめられている。また、事業者が利用者に対してAIを用いたクラウドサービスを提供する際に推奨される自主的な取組について、具体例を示しつつ提言されている。

(4) AIを用いたクラウドサービスの安全・信頼性に係る情報開示指針(ASP・SaaS編)

総務省は、2022年2月15日に、従来から策定・公表している「クラウドサービスの安全・信頼性に係る情報開示指針」²²に「AIを用いたクラウドサービスの安全・信頼性に係る情報開示指針(ASP・SaaS編)」²³を追加した。

同指針は、AIを用いたクラウドサービスの提供の増加等を踏まえ、ASP・SaaS等のクラウドサービス事業者がAIを用いたクラウドサービスを提供しようとする際のリスクへの対応方針として、新たに「クラウドサービスの安全・信頼性に係る情報開示指針」に追加されたものである。

(5) AI原則実践のためのガバナンス・ガイドライン

経済産業省は、2022年1月28日に、「AI原則実践のためのガバナンス・ガイドライン Ver. 1.1」²⁴を公表した。

同ガイドラインにおいて、AIの社会実装の促進に必要なAI原則の実践を支援するため、AI事業者が実施すべき行動目標が提示されるとともに、それぞれの行動目標に対応する仮想的な実践例やAIガバナンス・ゴールとの乖離を評価するための実務的な対応例が例示されている。

(6) 企業活動にAI倫理を導入していく上での注意点と提言

一般社団法人AIビジネス推進コンソーシアムは、2021年8月31日に、企業活動においてAI倫理を導入するためのリスクの整理や注意点、企業がAI倫理の問題に対応する際の課題を取りま

²¹ 次に掲げる URL のウェブサイト上所掲。

<https://www.soumu.go.jp/main_content/000792669.pdf>

²² 総務省は、利用者によるクラウドサービスの比較・評価・選択等を容易にすることを目的として、これまでに「クラウドサービスの安全・信頼性に係る情報開示指針」と総称する7つの情報開示指針を策定し、公表している。

²³ 次に掲げる URL のウェブサイト上所掲。

<https://www.soumu.go.jp/main_content/000792672.pdf>

²⁴ 次に掲げる URL のウェブサイト上所掲。

<https://www.meti.go.jp/shingikai/mono_info_service/ai_shakai_jisso/pdf/20220128_1.pdf>

なお、同ガイドラインは、2021年7月9日に公表された「AI原則実践のためのガバナンス・ガイドライン Ver. 1.0」に関し、同年9月15日まで行われた意見募集に寄せられた意見や「AI原則の実践の在り方に関する検討会」における議論を踏まえて、取りまとめられたものである。

とめ、「企業活動にA I 倫理を導入していく上での注意点と提言～リスクベース・アプローチを踏まえた検討～」²⁵と題する報告書を公表した。

同報告書において、A I 倫理をどのように企業活動に取り入れていくべきか、次のとおり提言されている。

- ・ 提言1…自社のA I 倫理に対するポリシーを設定すべきである。
- ・ 提言2…A I により生み出す価値とリスクの両面を見える化すべきである。
- ・ 提言3…外部のステークホルダとの共通認識を作るべきである。

(7) A I プロダクト品質保証ガイドライン

A I プロダクト品質保証コンソーシアム (QA4A I コンソーシアム)は、2021年9月15日に、「A I プロダクト品質保証ガイドライン 2021.09版」²⁶を公表した。

同ガイドラインは、「A I プロダクト品質保証ガイドライン 2020.08版」を改訂して、「各組織においてA I 技術への過度の期待を予防し、適切な活用や適時のリリースを行うための、A I プロダクト²⁷の品質保証に対する共通的な指針を与える」ものとして策定されたものであり、今般の主な改訂事項は、次に掲げる4つである。

- ・ 3章 技術カタログの改訂 : 公平性の具体的な定義の追記等
- ・ 5章 生成系システムの改訂 : テスト設計事例の追記等
- ・ 7章 産業用プロセスの改訂 : A I プロダクトの利用における法令上、倫理上の問題の有無と
いう観点からの留意事項に対する説明の修正等
- ・ 8章 自動運転の一部修正 : 自動運転関連標準の記述の修正等

2. 海外の動向

(1) 欧州連合 (E U)

① A I 分野の国際的な取組に関するハイレベル会合

欧州委員会と欧州理事会は、2021年9月14日及び15日に、同委員会によるA I に関する政策パッケージの発表(同年4月)に続くものとして、A I 分野の国際的な取組に関するハイレベル会合をスロベニアで開催した。

同会合において、A I 政策を具体的な活動へ移す方策として、人間中心アプローチの国際支援イニシアチブである InTouchAI.eu の発足や globalpolicy. A I について言及がなされた。また、欧州のA I 規制を国際的な標準としていきたい旨の方向性が示された。

② 文化・クリエイティブ分野におけるA I の機会と課題に関する研究

著作権と新技術に関する研究 : 著作権データ管理とA I

欧州委員会は、2022年3月16日に、デジタル化の10年間における文化・クリエイティブ分野

²⁵ 次に掲げる URL のウェブサイト上所掲。

<https://aibpc.org/cp-bin/wordpress/wp-content/uploads/2021/08/AIBPC_企業活動に AI 倫理を導入していく上での注意点と提言_リスクベース・アプローチを踏まえた検討_-20210831.pdf>

²⁶ 次に掲げる URL のウェブサイト上所掲。

<<https://www.qa4ai.jp/download/>>

なお、2019年5月に初版が公表され、今回が3回目の改訂版の公表となる。

²⁷ 同ガイドラインにおいて、「A I 技術を用いた製品やサービス」と規定されている。

の活動を推進するため、「文化・クリエイティブ分野におけるA Iの機会と課題に関する研究」及び「著作権と新技術に関する研究:著作権データ管理とA I」と題する2つの報告書を公表した²⁸。

前者において、欧州の文化的コンテンツ（建築、出版、映画、音楽、ニュース）の流通におけるA I利活用の可能性とその妨げとなり得る課題として、言語の多様性、A Iへの過依存、雇用構造の変化等を挙げるとともに、課題への対策として、システムの相互運用性やA Iスキルに係る教育等が新しく資金注入すべき領域である旨が提言されている。

後者において、文化・クリエイティブ分野における権利メタデータの管理、ライセンス供与、報酬と分配における技術の役割を確認しつつ、著作権データ管理の課題を示して、権利メタデータの重要性を訴えるとともに、分野間の権利メタデータの相互運用性を高めるための手段が提示されている。

③ 欧州議会における決議採択

欧州議会は、2021年10月6日に、A Iのアルゴリズムにはバイアスが存在しており、特に法執行機関による捜査や国境検問所などの場面でA Iが差別に利用されることを防ぐには人間による監督と強力な法的権限による制限が必要であるとの主張により、法執行機関が顔認証技術や犯罪を予見する技術を使用することを禁止するよう求める決議を採択した²⁹（賛成377、反対248、棄権62）。

また、2022年5月3日に、デジタル時代の人工知能に関する特別委員会（Special Committee on Artificial Intelligence in a Digital Age:A I D A）の最終勧告を採択した³⁰（賛成495、反対34、棄権102）。

同勧告において、A Iを技術として規制すべきではなく、規制介入の種類、強度、タイミングはA Iシステムの特定の用途に関連するリスクのレベルに合わせて検討すべきとした上で、透明性、説明可能性、公正性などの基礎的な欧州の価値とともにA Iとロボティクスが人間中心で人間を補完するものであるという原則に基づいて、A Iの巨大な可能性をどのように探究するかという公的な議論を推進すべき旨が提言されている。また、責任あるA Iの利活用のための共通基準の世界的な合意に向けて、同じ考え方を持つ民主主義の国家が協力し、国際的な議論を形成していくべきであるとされている。

²⁸ “Study on Opportunities and Challenges of Artificial Intelligence (AI) Technologies for the Cultural and Creative Sectors”及び“Study on copyright and new technologies”

次に掲げる URL のウェブサイト上所掲。

<<https://digital-strategy.ec.europa.eu/en/library/study-opportunities-and-challenges-artificial-intelligence-ai-technologies-cultural-and-creative>>及び

<<https://digital-strategy.ec.europa.eu/en/library/study-copyright-and-new-technologies>>

²⁹ 本件に関する欧州議会の報道発表は、次に掲げる URL のウェブサイト上所掲。

<<https://www.europarl.europa.eu/news/en/press-room/20210930IPR13925/use-of-artificial-intelligence-by-the-police-meps-oppose-mass-surveillance>>

³⁰ 本件に関する欧州議会の報道発表は、次に掲げる URL のウェブサイト上所掲。

<<https://www.europarl.europa.eu/news/pt/press-room/20220429IPR28228/artificial-intelligence-meps-want-the-eu-to-be-a-global-standard-setter>>。

また、採択文書については、次に掲げる URL のウェブサイト上所掲。

<https://www.europarl.europa.eu/doceo/document/TA-9-2022-0140_EN.html>

なお、A I D Aは、2021年11月9日に、「デジタル時代の人工知能」草案を公表し、2022年3月22日に、同草案を最終勧告として採択した。

④ 機械学習アルゴリズムの保護

欧州連合の専門機関である欧州ネットワーク・情報セキュリティ機関は、2021年12月14日に、機械学習アルゴリズムの分類と機械学習システムを標的とする脅威を分析した「機械学習アルゴリズムの保護」と題する報告書を公表した³¹。

同報告書において、機械学習アルゴリズムの分類学を提示した上で、機械学習に関するセキュリティ管理について分析されている。また、機械学習の6つの脅威を特定し、37のセキュリティ管理項目が推奨されているが、機械学習特有の管理項目が半数以下（16項目）であるため、機械学習に対するセキュリティ管理は、従来のサイバーセキュリティ管理と共に実施すべきとされている。

⑤ 法執行分野における顔認証技術の使用に関するガイドライン

欧州データ保護会議は、2022年5月12日に、EU及び加盟各国の法制定者や法執行機関を対象とする「法執行分野における顔認証技術の使用に関するガイドライン」を採択し³²、同年5月16日から6月27日まで意見募集を実施した。

同ガイドラインにおいて、次に掲げる特定のケースの顔認証技術の使用について、禁止する必要があるとされている。

- ・ 公共の場における個人の遠隔生体認証
- ・ 民族、ジェンダー、政治的・性的嗜好や他の差別的事由に基づき個人を分類するAI顔認証システム
- ・ 自然人の感情を推察するための顔認証技術又は類似の技術
- ・ 法執行の文脈での個人データの処理であり、大規模で無差別的な方法に基づく個人データの集約（オンラインでアクセス可能な写真や顔写真のスクレイピング）によるデータベースに依拠するもの

(2) フランス

フランス政府は、2021年11月8日に、2018年3月に公表したAI戦略の第2弾として、AI人材の育成・誘致等を中心とした「国家AI戦略」を公表した³³。

同戦略は、研究開発に重点を置いた第1弾とは異なり、人材育成や技術の商用化・展開にも重点が置かれていることが特徴であり、公的投資15億ユーロ、民間投資5億ユーロなど総額22億ユーロ（組込型AIや信頼できるAIの展開等：約12.2億ユーロ、人材育成：約7.8億ユーロ、革新的技術の研究：約1.3億ユーロ等）にも及ぶ計画となっている。また、「国家の能力を向上させる」、

³¹ “Securing Machine Learning Algorithms”

次に掲げる URL のウェブサイト上所掲。

<<https://www.enisa.europa.eu/publications/securing-machine-learning-algorithms>>

³² “Guidelines 05/2022 on the use of facial recognition technology in the area of law enforcement”

次に掲げる URL のウェブサイト上所掲。

<https://edpb.europa.eu/our-work-tools/documents/public-consultations/2022/guidelines-052022-use-facial-recognition_en>

³³ “STRATÉGIE NATIONALE POUR L’INTELLIGENCE ARTIFICIELLE – 2e phase”

次に掲げる URL のウェブサイト上所掲。

<<https://www.enseignementsup-recherche.gouv.fr/sites/default/files/2021-11/dossier-de-presse---strategie-nationale-pour-l-intelligence-artificielle-2e-phase-14920.pdf>>

「組込型A I及び信頼できるA Iの分野でフランスをリーダーにする」、「経済へのA Iの展開を加速させる」という3つの大きな目標が掲げられている。

(3) ドイツ

① ネットワーク分野における人工知能

連邦ネットワーク庁は、2021年12月15日に、通信、電力、ガス、郵便、鉄道等のネットワークサービス部門におけるA I利活用に関する報告書「ネットワーク分野における人工知能」を公表した³⁴。

同報告書において、A Iが絶大なポテンシャルを有しており、ネットワークサービス部門へのA I導入を積極的に図るべきとの見知が示されるとともに、企業間のギャップの解消、A I専門労働者の不足、アルゴリズムの複雑性、法的枠組みの未整備などが課題である旨が指摘されている。また、A I法制度について、EUが提案する法制枠組みに準ずるとして、ドイツ国内のビジネス環境に適した法整備を行うため、今後も関連事業者との協議を継続する意向が示されている。

② 監査可能なA Iシステムに向けて

連邦情報セキュリティ局は、2022年5月23日に、A Iシステムの監査に関するホワイトペーパー「監査可能なA Iシステムに向けて」を発行した³⁵。

同ホワイトペーパーにおいて、様々なユースケースにおけるA Iシステムの監査可能性を包括的にインベントリ化し、監査可能性を異なるアプリケーション間や時間軸で比較することができるようにするため、A Iのライフサイクル（計画フェーズ、データ取得フェーズ、訓練フェーズ、運用フェーズ等）や評価項目（セキュリティ、安全性、説明可能性等）に応じて、スコアを割り当てる「認証準備マトリクス」（Certification Readiness Matrix：CRM）が提案されている。

(4) 英国

① 国家A I戦略

英国政府は、2021年9月22日に、初めての10箇年計画となる「国家A I戦略」を公表した³⁶。同戦略の主な内容は、次のとおりである。

- ・ 国家A I研究イノベーションプログラムの立ち上げによる研究者間の調整と協力の改善によるビジネス・公共部門の市場投入能力の強化
- ・ A IオフィスとUKリサーチ&イノベーションの共同プログラムの立ち上げ
- ・ A I技術の大規模展開を推進するために必要なコンピューティングパワーの可能性と能力に

³⁴ “Künstliche Intelligenz in den Netzsektoren”

次に掲げる URL のウェブサイト上所掲。

<<https://www.bundesnetzagentur.de/SharedDocs/Downloads/DE/Sachgebiete/Digitales/KI/KI-Bericht-Marktdialog.pdf>>

³⁵ “Towards Auditable AI Systems”

次に掲げる URL のウェブサイト上所掲。

<https://www.bsi.bund.de/SharedDocs/Downloads/EN/BSI/KI/Towards_Auditable_AI_Systems_2022.pdf?jssessionid=64A6880D58497D957001FBEDB6D08E29.internet472?__blob=publicationFile&v=4>

³⁶ “National AI Strategy”

次に掲げる URL のウェブサイト上所掲。

<<https://www.gov.uk/government/publications/national-ai-strategy>>

関する共同レビューの発表

- ・ 知的財産庁を通じたA Iの著作権と特許に関する協議の開始
- ・ A Iスタンダードハブの試行による国際ルール設定に関する英国の関与の調整と公共部門におけるA I倫理と安全性に関するガイダンスの更新

② アルゴリズムの透明性基準

内閣府中央デジタル・データ・オフィスは、2021年11月29日に、国家A I戦略及び国家データ戦略に掲げられている公共部門におけるアルゴリズムが支援する意思決定の透明性を高めるための適切かつ効果的な方法を模索するというコミットメントを踏まえ、「アルゴリズムの透明性基準」を取りまとめ、公表した³⁷。

同基準は、公共部門の組織に向けて策定した37項目からなる「アルゴリズムの透明性データ基準」及びその利用説明書である「アルゴリズムの透明性テンプレートとガイダンス」で構成され、その有効性を検証するため、公的機関において試験運用（パイロット）が実施された。

③ 効果的なA I保証エコシステムに向けたロードマップ

データ倫理・イノベーションセンター（Centre for Data Ethics and Innovation：CDEI）は、2021年12月8日に、5年以内に世界をリードするA I保証エコシステムを構築することを目標とし、それに向けて必要な手順を示したロードマップを公表した³⁸。

同ロードマップにおいて、A I技術が社会に受容されて利益を生み出すために、A Iの効果、正確性、安全性といった要素に対する信頼を担保する必要があるとして、A I保証サービスのプロバイダーは、他の領域（財務報告、品質管理等）における保証と同様に、次に掲げる2つの事項を満たさなければならないとされている。

- ・ 信頼性を証明するエビデンスを確実に評価すること
- ・ ユーザー（導入企業等）に適切なレベルでエビデンスを示すこと

次のステップはロードマップの遂行であり、その1つとして、同センターは、2022年1月12日に、デジタル・文化・メディア・スポーツ省及び人工知能局と共に、A Iガバナンスを改善し投資や雇用を促進するため、また、グローバルなA I技術基準の発展に貢献するため、ツール開発やコミュニティの場として「A Iスタンダードハブ」の試験運用を発表した。

④ A Iバロメーター（第2版）

CDEIは、2021年12月17日に、A I及びデータの利用に関連する機会、リスク、ガバナンスの課題について分析を行った「A Iバロメーター（第2版）」を公表した³⁹。

³⁷ “Algorithmic Transparency Standard”

次に掲げる URL のウェブサイト在所掲。

<<https://www.gov.uk/government/collections/algorithmic-transparency-standard>>

³⁸ “The road map to an effective AI assurance ecosystem”

次に掲げる URL のウェブサイト在所掲。

<<https://www.gov.uk/government/publications/the-roadmap-to-an-effective-ai-assurance-ecosystem>>

³⁹ “AI Barometer 2021”

次に掲げる URL のウェブサイト在所掲。

<<https://www.gov.uk/government/publications/ai-barometer-2021>>

同報告書は、約1,000社の企業を対象に実施した調査をもとに、英国全土でAI及びデータがどのように活用されているのかを明らかにしている。例えば、産業分野によって、データ駆動型テクノロジーの普及に大きな差があることが判明したほか、データへのアクセスと共有に関する障壁が明らかになった。

(5) 米国

① 全国AI諮問委員会

商務省は、2021年9月8日に、AIに関する問題について大統領や他の連邦機関に助言するための全国AI諮問委員会(National Artificial Intelligence Advisory Committee:NAIAC)を創設したことを発表した⁴⁰。

NAIACは、国家AIイニシアチブ法に基づいて創設されたもので、学术界、産業界、非営利団体、市民団体、国立研究所などのAIに関する幅広い分野の専門家で構成され、米国のAI分野における競争力の現状、イニシアチブの実施状況、AIを取り巻く科学の状況、AIと労働力に関する問題、国際連携、法的課題等について提言を行うこととされており、2022年5月4日に初回会合が開催された。

同会合において、同法に定められている「法執行とAI小委員会(NAIAC-LE)」が正式に設立された。同小委員会は、法執行におけるAI利活用に関する課題を検討し、バイアス、データセキュリティ、AIの安全や法執行への採用可能性、AI利活用がプライバシーの権利と一致することを確実にする法的基準を含むトピックについて、大統領に助言することとなっている。

② 顔認証を含むAI導入に関する意見募集

国土安全保障省(Department of Home and Security:DHS)は、2021年11月5日から12月6日まで、顔認証を含むAIの導入について、一般市民の認識に関する意見募集を実施した⁴¹。

DHSは、既に税関、国境警備、交通機関の治安維持及び調査等の重要業務において、AIを利用した技術を活用又は試行しているが、このような技術の活用について、一般市民の間で十分に議論されているものではないことから意見募集を行ったものである。DHSは、技術の導入を成功させるためには、一般市民がどのように新しい技術の利点やリスクを認識しているかを理解しておく必要があると考えているとコメントしている。

③ 責任あるAIガイドライン

国防総省国防イノベーションユニットは、2021年11月15日に、「責任あるAIガイドライン」を公表した⁴²。

同ガイドラインは、民間企業や政府職員を含む国防総省のAIシステム開発の関係者に対し、そ

⁴⁰ 商務省の報道発表は、次に掲げるURLのウェブサイトに掲載。

<<https://www.commerce.gov/news/press-releases/2021/09/department-commerce-establishes-national-artificial-intelligence>>

⁴¹ 意見募集の詳細及び寄せられた意見については、次に掲げるURLのウェブサイトに掲載。

<<https://www.regulations.gov/docket/DHS-2021-0015-0001>>

⁴² "Responsible AI Guidelines"

次に掲げるURLのウェブサイトに掲載。

<<https://www.diu.mil/responsible-ai-guidelines>>

それぞれのAIプログラムが、国防総省のAI倫理原則に準拠しているかどうかを判断するための包括的なフレームワークを提供するものである。

また、同ガイドライン内のレポートにおいて、AIシステムを開発・調達する技術に同ガイドラインを適用する方法が説明されており、プランニング、開発、導入に至るまでの様々な段階でとるべき手順がステップ・バイ・ステップで提示されている。

④ AI研究者ポータル

国家AIイニシアチブ局は、2021年12月20日に、AIイニシアチブ・アプリケーションをトレーニングするデータセットやテストベッド環境へ容易にアクセスすることを実現するために、「AI.gov」⁴³に、AI研究者を対象とする「AI Researchers Portal」を開設した⁴⁴。

ポータルを介してアクセスすることができるツールは、AI関連の連邦助成金の情報、データセット、全米科学財団等が運用するコンピューティングリソース、研究プログラムディレクトリ、40箇所のテストベッドなどであり、開設を発表するツイート⁴⁵において、「米国のAI研究コミュニティと連邦政府が提供する多くのリソースをつなぐ中心」になるとされている。

⑤ AIにおけるバイアスに関するレポート（改訂版）

国立標準技術研究所（National Institute of Standards and Technology：NIST）は、2022年3月15日に、AIにおけるバイアスに関するレポートの改訂版を公表した⁴⁶。

同レポートは、2021年夏に公表されたドラフト版に対して寄せられた意見を反映したもので、AIシステムが単独で動作するものでない以上、全体のコンテキストに着目すべきであり、バイアスが生み出す有害な影響を特定・管理する能力を向上させるための一歩として、その原因を調べる範囲について、AIソフトウェアの学習に使用される機械学習プロセスやデータに加え、技術の開発方法に大きな影響を及ぼす幅広い社会的要因まで拡大することを提言している。

⑥ AIリスク管理フレームワーク（初期ドラフト）

NISTは、2022年3月17日に、信頼でき、責任あるAIの開発・利活用を促進することを目的とするAIリスク管理フレームワークの初期ドラフトを公表した⁴⁷。

同フレームワークは、2021年12月に公表されたコンセプトペーパー等に寄せられた意見をもとに、AIシステムに関連する事業者や社会のリスクに対する理解を深め、リスク管理を支援するために作成されたものである。AIシステムに関するステークホルダ、ライフサイクルなど各種概念を定義し、AIに係るリスクの特性を分類した上で、リスク管理に必要な4機能（マッピング、測

⁴³ 「報告書2021」第1章2.（6）④参照。

⁴⁴ ポータルのURLは、次のとおり。

<<https://www.ai.gov/ai-researchers-portal/>>

⁴⁵ ツイートのURLは、次のとおり。

<<https://twitter.com/WHOSTP/status/1471877506504568834>>

⁴⁶ “Towards a Standard for Identifying and Managing Bias in Artificial Intelligence”

次に掲げるURLのウェブサイトに所掲。

<<https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.1270.pdf>>

⁴⁷ “AI Risk Management Framework: Initial Draft”

次に掲げるURLのウェブサイトに所掲。

<<https://www.nist.gov/system/files/documents/2022/03/17/AI-RMF-1stdraft.pdf>>

定、管理、ガバナンス)が整理されている。

NISTは、AI関連コミュニティからコメント・意見を受けた上で、同フレームワークの策定作業を進めるとして、2022年4月29日まで意見募集を実施した。今後、第2次ドラフトを同年夏から秋頃に公表し、冬頃に正式版(Ver1.0)を公表することが予定されている。

⑦ AIが生み出す雇用差別に対処するガイダンス

司法省及び雇用機会均等委員会は、2022年5月12日に、雇用者がAIやその他のソフトウェアツールを利活用して雇用に関する決定を行う際の障害者差別に関するガイダンス及び技術支援文書を公表した⁴⁸。

同ガイダンス等において、雇用主がAIツールを利活用して雇用に関する決定を行う場合、障害がある者に対して米国障害者法上違法な差別をもたらす可能性があるという警告がなされるとともに、雇用主はツールを選択する際に当該ツールが様々な障害にどのような影響を及ぼすかを考慮する義務を負うこと、アルゴリズムによる意思決定ツールを利活用する際に合理的配慮を提供するためのプロセスを整備すべきことなどが規定されている。

⑧ イノベーションによるオンライン上の違法・有害情報との戦い

連邦取引委員会は、2022年6月16日に、オンライン上の違法・有害情報に対処することを目的としたAIツールの使用促進は慎重に進めるべきとする報告書「イノベーションによるオンライン上の違法・有害情報との戦い」を公表した⁴⁹。

同報告書は、2021年歳出予算法により、オンライン上の違法・有害情報の発見、削除その他の適切な対応をするために、AIをどのように活用することができるかを調査し、議会に報告するよう求められていることを受けたものである。

同報告書において、AIには、内在的な設計上の欠陥と不正確さ、開発者の偏見を反映し誤った結果を導く可能性、技術開発のために膨大なデータを必要とするため商業的な監視を動機付ける可能性等があることが指摘され、「危害を減らすという重要な目的においても、このようなツールの使用の義務付けや過度の依存については細心の注意が必要」と結論付けられている。

(6) 中国

① 信頼できる人工知能に関する白書、人工知能白書2022

通信院は、2021年7月9日に、「信頼できる人工知能に関する白書」を公表した⁵⁰。

⁴⁸ “Algorithms, Artificial Intelligence, and Disability Discrimination in Hiring”及び“The Americans with Disabilities Act and the Use of Software, Algorithms, and Artificial Intelligence to Assess Job Applicants and Employees”

次に掲げるURLのウェブサイトに所掲。

<<https://beta.ada.gov/ai-guidance/>>及び<<https://www.eeoc.gov/laws/guidance/americans-disabilities-act-and-use-software-algorithms-and-artificial-intelligence>>

⁴⁹ “Combatting online harms through innovation”

次に掲げるURLのウェブサイトに所掲。

<[https://www.ftc.gov/system/files/ftc_gov/pdf/Combatting Online Harms Through Innovation: Federal Trade Commission Report to Congress.pdf](https://www.ftc.gov/system/files/ftc_gov/pdf/Combatting%20Online%20Harms%20Through%20Innovation%20-%20Federal%20Trade%20Commission%20Report%20to%20Congress.pdf)>

⁵⁰ 「可信人工知能白皮書」

次に掲げるURLのウェブサイトに所掲。

<<http://www.caict.ac.cn/kxyj/qwfb/bps/202107/P020210709319866413974.pdf>>

同白書において、信頼できるAIのパノラマフレームワークを系統的に示し、信頼できるAIの特徴となる要素を述べるとともに、信頼できるAIとAIの科学技術倫理・管理の関係が分析されている。また、信頼できるAIの技術、産業と業界の実践などの面に焦点を当て、制御可能で信頼できる、透明で解釈可能な、プライバシーを保護できる、責任を明確にできる、多元的な包容力を持つ信頼できるAIを実現する道筋を分析し、更に信頼できるAIの未来の発展について提案されている。

通信院は、2022年4月12日に、「人工知能白書2022」を公表した⁵¹。

同白書において、世界におけるAIの政策、技術、実装、ガバナンスに関する最新動向が包括的にレビューされ、課題の整理が行われている。また、AIの近時の発展を踏まえて、AIの持続可能かつ健全な発展が目指されていることが示されている。

② 人工知能標準化白書（2021年版）

国家人工知能標準化総体チーム及び全国情報標準化委員会人工知能分委員会が指導し、中国電子技術標準化研究院等が編成を行った「人工知能標準化白書（2021年版）」を2021年7月9日に公表した⁵²。

同白書において、AIの現状及び発展の動向、中国国内外のAIの標準化の状況並びに標準的な体系の構築の状況及び留意点などが整理されている。

③ 新世代の人工知能倫理規範

科学技術部国家新世代人工知能ガバナンス専門委員会は、2021年9月25日に、「新世代の人工知能倫理規範」を公表した⁵³。

同規範において、AIが常に人間の制御下にあることを確保することを目標としており、6つの基本的な倫理要件並びに人工知能管理、研究開発、供給及び使用の活動における18の具体的な倫理要件が示されている。

【基本的な倫理要件】

- ・ 人類の幸福の増進
- ・ 公平と公正の推進
- ・ プライバシーとセキュリティの保護
- ・ 制御可能性と信頼性の確保
- ・ 責任者の明確化
- ・ 倫理意識の向上

【具体的な倫理要件】

<人工知能管理規範>

- ・ 人工知能の持続可能な発展を推進すること

<供給規範>

- ・ 品質管理を強化すること

⁵¹ 「人工知能白書（2022年）」

次に掲げるURLのウェブサイトに所掲。

<<http://www.caict.ac.cn/kxyj/qwfb/bps/202204/P020220412613255124271.pdf>>

⁵² 「人工知能標準化白書（2021版）」

次に掲げるURLのウェブサイトに所掲。

<<http://www.cqsoft.org/ueditor/php/upload/file/20210917/人工知能標準化白書（2021版）.pdf>>

⁵³ 「新世代人工知能倫理規範」

次に掲げるURLのウェブサイトに所掲。

<http://www.most.gov.cn/kjbgz/202109/t20210926_177063.html>

- | | |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <ul style="list-style-type: none"> ・ リスク防止を強化すること <p><研究開発規範></p> <ul style="list-style-type: none"> ・ アルゴリズムなどについて、安全性・透明性を強化すること ・ データの収集・アルゴリズムの開発において、偏見差別を避けること | <ul style="list-style-type: none"> ・ ユーザーの権益を保障すること <p><使用規範></p> <ul style="list-style-type: none"> ・ 誤用・濫用を避けること ・ 不正使用を禁止すること |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
- 等

(参考) 人工知能に関する調和の取れたルールを定める規則の提案

欧州委員会は、2021年4月21日に、「人工知能に関する調和の取れたルールを定める規則の提案」(以下「AI規制法案」という。)を公表した⁵⁴。同法案は、信頼できるAIのための法的枠組みを提案することにより、信頼のエコシステムを形成することを目的とするもので、採択に向けて、欧州議会及び欧州理事会における審議や調整が行われている。

同法案については、EU域内のみならず、日本を含めて国際的に大きな影響を及ぼすものと考えられることから、本推進会議においても、その動向等を注視しているところである。同法案に対する国内外の受け止め、反応等について、本推進会議において関係者から聴取したところ、その概要については、次のとおりであった⁵⁵。

<国内>

○ 一般社団法人日本経済団体連合会の意見

- ・ 基本的な考え方
 - 経団連が目指す「信頼できる高品質AIエコシステムの構築」と方向性は一致している。
 - しかしながら、禁止AIやハイリスクAIの定義等に曖昧さや解釈の余地が残されており、イノベーションや国家安全保障に影響を及ぼすおそれがある。
 - 施行前に、定義の明確化や説明の追加、ガイドライン等の提供を行うべき。
- ・ 具体的な意見
 - 法執行目的で使用されるリアルタイム遠隔生体認証システムは、ユースケースごとの危険性を考慮し、公的にアクセスできる空間の範囲を具体例等で明確にすべき。
 - 禁止AIやハイリスクAIの対象範囲をより限定的にし、規制の対象範囲や根拠、リスクの測定・評価方法を明確にすべき。
 - AIのリスクマネジメントに関するリスクトレードオフも考慮したガイドライン等を示すべき。
 - 一定のリスク精査を果たすことを条件に、ハイリスクAIのプロバイダーが免責を得られるようにすべき。

⁵⁴ 「報告書2021」第1章2.(1)③参照。

AI規制法案の本文及び概要(いずれも日本語(仮訳))については、本報告書の参考資料として、総務省のウェブサイトにおいて公表している。次に掲げるURLのウェブサイトに所掲。

<<https://www.soumu.go.jp/iicp/research/results/ai-network.html>>

⁵⁵ 3回にわたり発表及び意見交換を行ったものである。前掲注(15)参照。

なお、国内については、(一社)日本経済団体連合会のほか、日本電気株式会社や富士通株式会社などがAI規制法案に関する意見募集に対して意見を提出している。

また、欧州及び米国の意見については、総務省「諸外国における人工知能(AI)への規制の動向に関する調査研究」に基づくものである。

- 違反の種類や内容、得られた便益の大きさ、違反の悪意の有無などに応じて、適切なペナルティを定めるべき。

<欧州>

○ 欧州各国における研究者、弁護士事務所、企業等の意見

- ・ AI規制法案の全体に関する見解
 - 次のような見解が示され、AI規制法案を法制化すること自体への異論は見られなかった。
 - AIに関するEUのリーダーシップを確保し、市場を守る統一ルールを整備、イノベーションのサポートを行うものであり、人権や社会の保護、人間中心のアプローチのため、倫理の枠組みだけではなく、規範的な枠組みを採用するものと認識している。
 - 人権の尊重・保護（人間中心のアプローチ）とともに、欧州の競争力の強化が重要であるとして作られたものであると認識している。
 - EUにおける信頼できてセキュアなAIの発展とEU市民の価値・基本権・安全を守ろうとする内容であると認識している。
- ・ 「AIシステム」の定義に関する見解
 - より定義を明確化すべきである。
 - 「統計的アプローチ」も含まれてしまう点は問題である。
 - ソフトウェアのほとんどが含まれてしまいかねない。
 - ・ 許容できないリスクのあるAIに関する見解
 - 定義が曖昧で、禁止の範囲が狭い。
 - ソーシャルスコアリングや公的機関による監視のためのAIの明確な禁止は合理的である。
 - 予測可能性の確保のため、規範的に狭く定められたものと考えられるが、それでも事業者の負担になり得る。
 - ・ ハイリスクAIに関する見解
 - ハイリスクAIの範囲は限定されていると考えるが、技術的要件の明確化が必要である。
 - ハイリスクAIの範囲が広い。AIのカテゴリーとユースケース、解釈の手引きを明示すべき。
 - ハイリスクAIの定義・義務としてAI規制法案が示す内容が今後の議論によって、大きく変わることは予想されないが、明確な基準が望まれる。
 - ・ 中小企業へのサポートに関する見解
 - 現在のAI規制法案のままではイノベーションに対する萎縮効果が生じかねない。
 - 規制のサンドボックスを好意的に捉える向きもある。
 - 中小企業へのサポートの規定は曖昧過ぎるため、修正が必要である。

<米国>

○ 米国における弁護士事務所の意見

- ・ 米国における規制の在り方について
 - 米国が、EUほど範囲が広く、また、ビジネスにとって対応すべき課題の多いAI規制を作るとは思われない。しかし、新たな制定法がなくても、連邦の機関は、現行の（典型的には分野別の）規制権限と一貫するようにAIを規制するよう努力を払っている。
 - TTC（米EU貿易技術評議会）は、差別・バイアス、詐欺、透明性について、あるいは、

米国とEUが概ね合意するその他の分野において、米国のAI規制とEUのAI規制を、少なくともある程度平仄の取れたものとする可能性がある。また、TTCは、一定の国であって、反民主主義的な方法で技術を用いるもの又は米国若しくは欧州及びアジア太平洋地域における同盟国の安全保障に課題を生じさせるものとの関係で、そのような国への依存を限定し又はそのような国の影響を弱めるための政策にも最大の力を発揮し得る。

- 米国は、AIを統制するに当たって、ハードローとソフトローをミックスしたものをを用いる。一般的には、米国はできる限りソフトローを用いる。立法者や規制当局は、経験によってソフトローではAIによる害から一般公衆を守るに不十分であると分かった場合に、初めてハードローに目を向ける。最近、そのようなハードローを適用し又は制定することに向けたいくつかの動きが見られる。
- 新法を制定するよりも現行法を適用することの方がかなり容易であるため、米国の規制当局は、AIに現行法を適用することに非常に着目している。それと同時に、連邦レベル及び州レベルにおいて、制定法の立法又はその提案が見られる。
- 政府はAIのリスクに焦点を合わせている。企業も同様にすべきである。米国において、AIは規制されないという考え方は間違いである。(注：一般的な日本人が想像するよりも、ハードローが既に採用され又は今後採用されるであろう範囲は広いという見方について、当該法律事務所から異論は述べられなかった。)
- AIに関する日米の協働について
 - 米国と日本は、倫理的で安全で信頼できるAIの論点に関し、多国間の組織(OECD、G7、G20、GPAIなど)の場において又は二国間の対話により、協働を継続するであろう。
 - 米国は、AIの技術標準の策定や規制遵守を求めることを目的として、TTCの下でのEUとの協調と同様の協調を日本ともしたいであろう。

3. 国際的な議論の動向

(1) OECD

① 閣僚理事会

2021年10月5日及び6日に、閣僚理事会が開催された。

閣僚声明⁵⁶において、「我々は、『AI(人工知能)に関するOECD原則』の実施並びに信頼性のある自由なデータ流通(DFFT)及び個人データへの信頼性のあるガバメント・アクセスに関する高次原則のあり得べき策定の促進を通じた取組を含め、包摂的なデジタル経済を前進させるために協働することにコミットする。」とAI原則の実装を含むデジタル経済の促進について言及されている。

また、同年10月4日に、サイドイベントとしてAI原則の実装に関するセッションが開催された⁵⁷。

⁵⁶ 「2021年OECD閣僚理事会第二部 閣僚声明(仮訳)」

次に掲げるURLのウェブサイト上所掲。

<<https://www.mofa.go.jp/mofaj/files/100243592.pdf>>

⁵⁷ “Putting the OECD AI Principles into Practice: progress and future perspectives”

次に掲げるURLのウェブサイト上所掲。

<<https://oecd.ai/en/mcm>>

② デジタル経済政策委員会

2021年12月2日に、デジタル経済政策委員会（Committee on Digital Economy Policy: CDEP）のAIセッションが開催された。日本の取組について、総務省から、事業者等へのヒアリングを踏まえて具体的な取組事例・グッドプラクティスを取りまとめた本推進会議の「報告書 2021」を紹介するとともに、経済産業省から、参考事例を交えた「AI原則実践のためのガバナンス・ガイドライン」⁵⁸を紹介した。

また、非公式専門家会合（OECD Network of Experts on AI: ONE AI）⁵⁹の3つのワーキンググループ（WG）のうち、「AI政策に関するWG（WG on AI policies）」をCDEPの作業部会（AIガバナンス作業部会（Working Party on Artificial Intelligence Governance: WP AI GO））へ移行すること⁶⁰及びONE AIの1年間の運営延長について合意された。

③ AIガバナンス作業部会

2022年5月24日及び25日に、AIガバナンス作業部会（WP AI GO）の初回会合が開催された。

同作業部会は、AIに関する理事会勧告をはじめとするOECDのAIに関するスタンダードの履行を支援し、信頼できるAIの責任あるスチュワードシップを促進するために各国のAI政策に関する情報交換やドキュメントの作成、オンラインプラットフォーム（OECD AI オブザーバトリー）の更なる機能開発等について議論する場として設置されたものである。

同会合において、2022年度の議長及び副議長（日本を含む。）が承認された。また、各国のAIガバナンスに関する政策が紹介・共有されるとともに、ONE AIの各WGの活動・進捗状況の報告等とそれらに関する議論がなされた。さらに、OECD AI オブザーバトリーについて、AI関連の文献数や組織間の連携状況等に関する指標により、各国のAI政策を定量的に分析・可視化するユーザーインタフェースなどの機能拡張案が紹介された。

日本は、総務省から、AI原則の遵守に関して、ソフトロースタンスであること及び監査による客観的評価の重要性について言及するとともに、言語技術に関して、「グローバルコミュニケーション計画 2025」⁶¹やMICSUS（高齢者介護支援マルチモーダル音声対話システム）⁶²等を紹介した。加えて、AIと効率的なエネルギーシステムとの関係性及び通信インフラ政策に関するパネルディスカッションにおいて、株式会社東芝から、分散電源や蓄電池など散在するエネルギー源をIoT機器により遠隔制御し1つの発電所のように機能させるVPP（Virtual Power Plant（仮想発電所））⁶³とその応用技術であるエネルギー需要や太陽光発電量の予測AI技術等を紹介した。

⁵⁸ 当時は、改訂前の「AI原則実践のためのガバナンス・ガイドライン ver. 1.0」を紹介。

次に掲げる URL のウェブサイトに掲載。

<https://www.meti.go.jp/shingikai/mono_info_service/ai_shakai_jisso/2021070902_report.html>

改訂後の「AI原則実践のためのガバナンス・ガイドライン ver. 1.1」については、本章1.（5）参照。

⁵⁹ 本推進会議構成員の中から、須藤議長及び杉山構成員が参加している。

⁶⁰ 「③ AIガバナンス作業部会」参照。

⁶¹ 総務省 「グローバルコミュニケーション計画 2025」

次に掲げる URL のウェブサイトに掲載。

<https://www.soumu.go.jp/menu_news/s-news/01tsushin03_02000298.html>

⁶² 情報通信研究機構 高齢者介護支援マルチモーダル音声対話システム（ミクス）

次に掲げる URL のウェブサイトに掲載。

<<https://keihanna-fair.jp/exhibition/ai/899>>

⁶³ （株）東芝 VPP（バーチャルパワープラント）

(2) GPAI (Global Partnership on AI)

① 第2回プレナリー会合

2021年11月11日及び12日に、GPAIサミット(第2回プレナリー会合)が開催され、閣僚級理事会において、次期議長国(2022年秋から2023年秋)に日本が就任することが承認された。また、ベルギー、チェコ、デンマーク、アイルランド、イスラエル、スウェーデンの新規加盟、2022年からユネスコのオブザーバ参加が承認された。

参加国からは、OECDのAI原則に基づく責任あるAIの推進の重要性、マルチステークホルダの参加推進の重要性、民主主義の価値を共有する途上国へのアプローチによる多様性の強化、アジア・太平洋地域初の次期議長国としての日本への期待等が示された。

② GPAI シンポジウム

総務省は、2022年2月9日に、経済産業省との共催によりGPAIシンポジウムを開催した。

AI倫理の国際動向、理論から実践、国内外のステークホルダへの期待などのテーマについて、GPAIに参加する有識者を含む産学のAI専門家による意見交換等を実施した⁶⁴。400名超の聴衆の参加があり、その多くから、GPAIの活動内容やAI原則の実践に関する理解が深まったという評価があった。

③ GPAI 仕事の未来：日本調査からの報告と提案

WG3「Future of Work (仕事の未来)」⁶⁵日本チームは、2022年3月15日に、「GPAI仕事の未来：日本調査からの報告と提案」と題するレポートを公表した⁶⁶。

同レポートは、AIが仕事に与える影響及び実態把握を問題意識として、既に海外において先行調査されているものをベースに日本の状況に合わせた調査項目等を検討し、企業や地方公共団体にインタビュー調査(11件)を行った結果を取りまとめたものである。

調査結果から、人材不足やサービス品質の向上などAI利活用の目的が類型化され、ガバナンス体制の構築や人材育成などの取組が行われていることが明らかになった。他方で、AI利活用の課題として、透明性や公平性などの技術的な課題のみならず、AIと人間の役割分担の再定義、AIと人間の信頼関係の構築、AIへの過度な依存など利用する人間側の課題等が挙げられている。また、調査から得られた知見をもとに、次に掲げる2つの事項が提案されている。

- ・ 「仕事の未来」調査に関する国・地域別の報告書の必要性
- ・ 学生を中心として調査する方法論の推進と効果検証

(3) G20

2021年10月30日及び31日に、G20ローマ・サミットが開催された。

次に掲げるURLのウェブサイトに掲載。

<<https://www.global.toshiba/jp/products-solutions/renewable-energy/products-technical-services/vpp.html>>

⁶⁴ 本推進会議構成員の中から、須藤議長及び実務構成員が参加し、AIガバナンス検討会構成員の中から、江間構成員が参加した。

⁶⁵ AIガバナンス検討会構成員の中から、江間構成員が参加している。

⁶⁶ 次に掲げるURLのウェブサイトに掲載。
<<https://ifi.u-tokyo.ac.jp/news/12582/>>

首脳宣言⁶⁷において、「我々は、信頼できる人間中心の人工知能（AI）の責任ある利用と開発から生じる利益を十分に認識しつつ、競争とイノベーションを促進するための中小零細企業と新興企業に特有のニーズ、多様性と包摂性、さらにはAIの研究、開発及び応用を促進するための国際協力の重要性を考慮しつつ、『G20 AI原則』の実行を推進する。我々は、『中小零細企業や新興企業のAI導入を促進するG20の政策事例』を歓迎する。」とAI原則の実行の推進について言及されている。

（４）欧州評議会 AIに関するアドホック会合（AdHoc Committee on AI：CAHAI）

2021年11月30日から12月2日にかけて、第6回総会が開催され、AIに関する法的枠組みについての文書（「人権、民主主義及び法の支配に関する欧州評議会基準に基づく人工知能についての法的枠組みにおける考え得る要素」）が採択された⁶⁸。

同文書において、AIシステムの使用により生じ得るリスクの予防・緩和のために、3つの価値（人権、民主主義、法の支配）に基づいた法的拘束力を有する分野横断的な法的文書が必要である旨が述べられ、各分野の詳細な指針については、法的拘束力を有する／有しない追加的な法定文書が将来的に必要となる可能性が指摘されている。

また、分野横断的な法的枠組みを補足する文書として、公共行政分野に法的拘束力を有する／有しない追加的文書を導入することが提案されている。特に、透明性、公平性、責任、アカウントビリティ、説明可能性及び補償といった観点から、AIシステムの使用、デザイン、調達、開発及び実装が監督され、3つの価値が遵守されるようメカニズムを構築すべき旨が推奨されている。

なお、CAHAIの活動は、AIに関する委員会（Committee on AI：CAI）に引き継がれ、2022年4月4日にハイレベル会合、同日から6日にかけて第1回総会が開催された⁶⁹。総会において、CAHAIの成果に関する意見交換や法制度の策定に向けて前進するための方策の検討が行われた。CAIは、2023年11月までに法制度について起草することとなっている。

（５）国際連合

① ユネスコ（UNESCO：国際連合教育科学文化機関）

2021年11月9日から24日にかけて、第41回総会が開催され、AI倫理勧告が採択された⁷⁰。

⁶⁷ 「G20 ローマ首脳宣言（仮訳）」

次に掲げる URL のウェブサイトに掲載。

<<https://www.mofa.go.jp/mofaj/files/100259275.pdf>>

⁶⁸ “Possible elements of a legal framework on artificial intelligence, based on the Council of Europe’s standards on human rights, democracy and the rule of law”

次に掲げる URL のウェブサイトに掲載。

<https://search.coe.int/cm/Pages/result_details.aspx?ObjectID=0900001680a4e8a5>

⁶⁹ “Inaugural Meeting of the Committee on Artificial Intelligence (CAI)”

次に掲げる URL のウェブサイトに掲載。

<<https://www.coe.int/en/web/artificial-intelligence/-/inaugural-meeting-of-the-committee-on-artificial-intelligence-cai->>

⁷⁰ “UNESCO member states adopt the first ever global agreement on the Ethics of Artificial Intelligence”

次に掲げる URL のウェブサイトに掲載。

<<https://en.unesco.org/news/unesco-member-states-adopt-first-ever-global-agreement-ethics-artificial-intelligence>>

なお、決議案に対して、ロシア、中国等から、追加的な技術的修正意見や議論が提起されたものの、多くの国が勧告案を歓迎したこともあり、AI倫理勧告及び決議の内容については、実質的な変更なく採択された。

同勧告において、“国家がA Iに関する立法、政策等を策定する際の国際法に合致した価値観、原理及び行動の普遍的な枠組みを提供すること”や“A Iシステムのライフサイクルのすべての段階に倫理が組み込まれるよう個人や企業等の行動を誘導すること”等を目的として、次に掲げる価値及び原則（A Iシステムのライフサイクルにおけるすべての関係者によって尊重されるべき事項）並びに政策措置の領域（勧告に基づき加盟国が措置すべき分野等）が定められるとともに、A I倫理に関する監視・評価等の必要性が示されている。

<価値>

- ① 人間の尊厳
- ② 人権及び基本的自由の尊重
- ③ 豊かな環境と生態系
- ④ 多様性と包摂性の確保
- ⑤ 平和と共存

<原則>

- ① 比例性と無害性
- ② 安心・安全
- ③ 公正・無差別
- ④ 持続可能性
- ⑤ プライバシーとデータ保護
- ⑥ 人間による監督と決断
- ⑦ 透明性と説明可能性
- ⑧ 責任とアカウンタビリティ
- ⑨ Awareness とリテラシー
- ⑩ マルチステークホルダによる適応的ガバナンス

<政策措置の領域>

- ① 倫理的影響評価
- ② 倫理的ガバナンスと管理
- ③ データ政策
- ④ 開発と国際協力
- ⑤ 環境と生態系（エコシステム）
- ⑥ ジェンダー
- ⑦ 文化
- ⑧ 教育と研究
- ⑨ コミュニケーションと情報
- ⑩ 経済と労働
- ⑪ 健康と社会的福利

② 高等弁務官事務所

2021年9月15日に、「デジタル時代におけるプライバシー権」と題する報告書を公表した⁷¹。公表に際して、バチエレ国連人権高等弁務官は、国際人権法に準拠しないA Iアプリケーションについて、各国は明示的に禁止すべきとの考えを示した⁷²。

同報告書において、A Iの活用によるプライバシー等の権利への影響を分析し、4つの主要分野（法執行機関、国家安全保障、刑事司法、国境管理）における影響の例が提供されている。また、有害な結果を防止・最小化し、A Iが提供する利益の享受を促進するための保障措置の設計と実施に関する勧告が提供されている。

(6) APEC

2021年12月7日から10日にかけて、ペルー主催によるA Iに関するワークショップが開催された。総務省は、「開発のためのA I（リスクと機会）」と題するセッションにおいて、本推進会議の取組等（A I開発ガイドライン、A I利活用ガイドライン、「報告書2020」や「報告書2021」における事業者の取組事例等）について紹介した。このほか、マルチステークホルダの観点による議論

⁷¹ “The right to privacy in the digital age: report (2021)”

次に掲げる URL のウェブサイト に所掲。

<<https://www.ohchr.org/en/calls-for-input/calls-input/2021/right-privacy-digital-age-report-2021>>

⁷² “Urgent action needed over artificial intelligence risks to human rights”

次に掲げる URL のウェブサイト に所掲。

<<https://news.un.org/en/story/2021/09/1099972>>

の重要性やAI分野における政府調達の方向性について意見交換が行われた。

(7) 二国間における政策対話等

① インターネットエコノミーに関する日米政策協力対話

2021年11月11日及び12日に、インターネットエコノミーに関する日米政策協力対話が行われ、AIセッションにおいて、AI技術への公共の信頼を深め、信頼性のあるAIの責任あるガバナンスを促進するための官民連携等に関する議論がなされた。

総務省から、本推進会議の「報告書2021」の内容（「AI倫理・ガバナンス」、「AI開発・利活用」、「人材育成」に関するベストプラクティス、グッドプラクティスを取りまとめた旨）を紹介するとともに、EUのAI規制法案等に関する議論を注視し、日本はソフトローのアプローチを堅持する方向であり、日米で連携して対応したい旨を発言した。

日米両国は、OECDのAIに関する勧告への支持を再確認し、継続的にOECDやGPAIを通じて信頼におけるAIの責任ある開発と展開について協働することを確認した。また、米国から、日本がGPAIの次期議長として今後の役割を担うことを歓迎する旨が示された。

② 日EU ICT戦略ワークショップ

2021年11月17日に、日EU ICT戦略ワークショップが行われ、AIセッションにおいて、EUから、AI規制法案の状況説明があり、これに対して、日本から、技術変革に柔軟に対応するにはソフトローが有効であるとの考えを説明するとともに、GPAIやユネスコなどの国際的な議論への参画状況、これまでのAI利活用ガイドライン等の策定を通じた貢献、AI原則実践のためのガバナンス・ガイドライン、AI原則の実装に関する産業界の状況を紹介した。日EU双方から、マルチの場での議論における日EU協力の重要性が指摘され、意見交換と議論を継続していくこととされた。

また、2022年4月8日に、同ワークショップが行われ、AIセッションにおいて、多国間会合の場における日EU協力の重要性が改めて指摘され、意見交換と議論を継続していくことで合意した。日本から、G7やG20などのAI原則に係る国際場裡における議論の経緯や今後の取組（特に、2022年秋以降に日本が議長を務めることとなっているGPAIに関する取組、人間中心のAIを実現するための取組）について説明した。EUから、AI規制法案やAI原則に向けた取組の状況、政策レベルでの協力の重要性、AIに係る理解・共有の重要性などについて説明があった。

③ 日EU ICT戦略政策対話

2022年2月3日に、日EU ICT戦略政策対話が行われ、AIセッションにおいて、日EU双方から、AI原則に関する見解の説明があり、国際連携等の取組やこれらに関する類似点・相違点等について議論がなされた。

総務省から、OECD、GPAI、ユネスコ等の国際的な情勢を説明するとともに、本推進会議の取組状況や「報告書2021」を紹介し、日EUで目指す方向性は一致しており、アプローチの違いについては継続的に密に連携して取り組んでいく旨を伝えた。EUから、価値観は日EUで共通しているが、ソフトローでは不十分で信頼を得るには法的な枠組みが必要である旨のコメントがあった。

④ 日独 ICT 政策対話

2022年3月23日及び24日に、日独 ICT 政策対話が行われ、23日の政府間会合において、グローバルデジタルガバナンスのセッションの一環として、AI ネットワーク社会やデータ利活用の推進に向けた取組に関する議論がなされた。

総務省から、本推進会議の取組状況や「報告書 2021」に記載されている事業者の取組事例等を紹介した。ドイツから、EU のデジタルサービス法案やデータ法案に関するドイツ国内の対応等について説明があった。

4. 国際標準化に関する動向

AI システムは、従来のソフトウェア技術やハードウェア技術に係る国際標準が適用可能な部分も多い一方で、ブラックボックス問題やバイアスのような倫理的課題、学習データに起因する問題など AI 特有の課題を有している。また、AI に係る技術開発の進展が速いため、品質保証技術のように研究段階のものも多い分野を含め、開発・運用と並行する形で標準化の取組が進められている。

主な国際的な標準化団体の動向として、IEEE（米国電気電子学会）は、2016年に、AI 倫理に係るプロジェクト（P7000 シリーズ）を開始している。また、ISO（国際標準化機構）／IEC（国際電気標準会議）は、2017年に、JTC1（第一合同技術委員会）にSC42（AI に関する分科会）を創設し、SC42内の各ワーキンググループにおいて様々なプロジェクトを開始している。

これらの標準化団体においては、AI 倫理の技術的側面、AI の概念や用語の定義、AI システムやライフサイクルの定義、ユースケース、データの定義、品質管理など原則から実装に向けて、多岐にわたってプロジェクトやワーキンググループ等が活動中である。さらに、最近では、政府調達要件の明確化やマネジメント、テスト等に関する標準化活動が活発化している。「報告書 2021」公表以降の主な取組は、次のとおりである⁷³。

(1) IEEE（米国電気電子学会）

IEEE の標準化委員会は、2021年9月23日に、AI の政府調達に関するプロジェクト（P3119）⁷⁴とそのワーキンググループを承認した。次に掲げる事項を目的として、AI の政府調達に係る用語と一連のプロセスを定義する活動を行っている。

- ・ AI 及びADS（Automated Decision Systems）の調達における社会・技術的及び責任あるイノベーションの課題に対処するための統一された定義とプロセス要件を確立すること
- ・ 支援機関が、公益のために責任を持ってAI システムを調達するためのプロセスに適応することを支援すること
- ・ プロセスモデルにおいて、倫理的に整合した価値観と強力な市民参加を促進すること

⁷³ 本節における国際標準化に関する動向については、本推進会議議長によるヒアリング（第5回ヒアリング（2022年3月24日））において、発表及び意見交換等を行ったものを中心に記載している。本推進会議議長によるヒアリングについては、次章参照。

また、これまでのIEEE やISO／IEC など国際標準化に関する動向については、「報告書 2018」第1章2.（1）、「報告書 2019」第1章2.（1）及び「報告書 2021」第1章3.（8）参照。

⁷⁴ “Standard for the Procurement of Artificial Intelligence and Automated Decision System”
次に掲げる URL のウェブサイトに掲載。
<<https://standards.ieee.org/ieee/3119/10729/>>

(2) ISO (国際標準化機構) / IEC (国際電気標準会議)

ISO/IEC JTC1/SC42 は、AIシステムのマネジメントに関するプロジェクト「ISO/IEC 42001」において、組織がAIのマネジメントシステムを確立、実施、維持するための要求条件やガイダンスを検討中である。また、AIシステムのライフサイクルに対応した検証や妥当性の確認を行うためのテスト技術や評価指標に係るプロジェクト「ISO/IEC TS 29119-11」において、標準仕様書の検討を行っている。

(3) 国際標準化に関連する各国等の動向

① 欧州

- ・ 欧州委員会 共同研究センターは、2021年7月に、「AIウォッチ：AI標準の風景」を公表した⁷⁵。AI規制法案に関連する標準をピックアップし、同法案が求める要件との整合性を調査するとともに、ギャップを特定し、欧州標準化ロードマップの策定に活かすこととしている。
- ・ CEN (欧州標準化委員会) 及びCENELEC (欧州電気標準化委員会) は、ISO/IEC JTC1/SC42において策定されている標準を分析し、AI規制法案の実施に当たって活用することができる標準及びJTC21 (AIに係るCEN-CENELEC合同技術委員会) で策定すべき標準等をJTC21として特定した。今後、欧州委員会や他の欧州標準化機関との調整を行った上で、具体的な標準化活動が開始される予定である。

② 米国

- ・ NISTは、2019年8月に、「米国のAIでのリーダーシップ：技術標準と関連ツール開発のための連邦の関与計画」を公表し⁷⁶、標準化の重点分野として、次に掲げる9分野を明示した。
 - Concepts and terminology (コンセプトと用語)
 - Data and knowledge (データとナレッジ)
 - Human interactions (人間との相互作用)
 - Metrics (メトリクス)
 - Networking (ネットワーキング)
 - Performance testing and reporting methodology (性能評価)
 - Safety (安全性)
 - Risk management (リスク管理)
 - Trustworthiness (信頼性)
- ・ NISTは、2021年1月から2年間でAIリスク管理フレームワーク⁷⁷を開発し、定期的に更新することとしており、また、必要に応じて、国際的な基準 (ISO/IEC標準等) と整合させることとしている。
- ・ 科学技術政策局は、2021年10月に、AI権利章典 (Bill of Rights for an Automated Society) を策定する必要性を訴え、同年11月に、公聴会やワークショップを開催した

⁷⁵ “AI Watch, AI standardisation landscape state of play and link to the EC proposal for an AI regulatory framework”
次に掲げる URL のウェブサイト に所掲。

<<https://op.europa.eu/en/publication-detail/-/publication/36c46b8e-e518-11eb-a1a5-01aa75ed71a1/language-en>>

⁷⁶ “Plan Outlines Priorities for Federal Agency Engagement in AI Standards Development”

次に掲げる URL のウェブサイト に所掲。

<<https://www.nist.gov/news-events/news/2019/08/plan-outlines-priorities-federal-agency-engagement-ai-standards-development>>

⁷⁷ 本章2.(5)⑥参照。

5. 国際シンポジウム「AIネットワーク社会フォーラム 2022」

総務省は、2022年3月1日に、AIの普及・利活用の鍵となるデータの利活用・流通に関する議論を行うとともに、国際的な動向も踏まえつつAIに係る規制やガバナンスの在り方等に関する議論を行うことにより、AIに関連する様々な社会的課題の解決に資することを目的として、国際シンポジウム「AIネットワーク社会フォーラム 2022」を開催した。本推進会議及びAIガバナンス検討会構成員のほか、国内外の幅広い分野から有識者や経営者等が参加し、意見交換等を行った。フォーラムの概要は、別紙1のとおりである⁷⁸。

同シンポジウムにおいて、データ活用について、個人情報の保護とデータ所有者の尊重とのバランスを取った上で、データ共有や流通の促進が重要であり、それに向けた環境整備が必要であるとの意見があった。

また、AI倫理・ガバナンスについては、規制等の海外動向を踏まえた日本における規制や事業者におけるAIガバナンスの在り方等に関する議論が行われ、次のような意見があった⁷⁹。

- ・ インターネットの民間利用開始以来の破壊的な技術トレンドの中にあつて、AIは極めて重要な要素であり、AI原則及びAIガバナンスの取組が大きな注目を集めている。
- ・ 日本政府は、AIの適切な開発と社会実装のため、有識者会議の検討等を踏まえて、2019年3月に「人間中心のAI社会原則」⁸⁰を取りまとめた。これらの原則はAIを利活用して新たな社会を構築する上で核となるものである。
- ・ 総務省のAIネットワーク社会推進会議において、AI社会のあるべき姿を考え、AI利活用の仕方やデータ流通などについて多面的な検討を行ってきた。
- ・ 2023年にG7が日本で開催される予定であり、AIの在り方について積極的に提案する準備が進んでいる。様々な動きがある中、国際社会と協調しながら、新たな社会を作っていくことが希求される。
- ・ 規制はAI倫理を守る多くの理由のうちの1つに過ぎず、例えば、企業価値や評判、社会的正義やビジネスチャンスなどの面からもAI倫理を考えることが重要となる。規制を遵守することは重要であるが、AI倫理というのは単なる規制遵守を超えるものであり、法律はAIを正しい方向に導く一部でしかない。
- ・ AIガバナンスは、公平性や透明性、信頼性などの原則を定め、それを実践する段階へと進んでいる。AIガバナンスに関連する課題解決の取組は、「競争領域」ではなく「協調領域」であり、様々な関係者を巻き込んで議論を重ね、エコシステムを構築することが重要である。
- ・ AI利活用の目的は、人間の知性を補うことであり、置き換えることではない。データやそこから導き出される知見は、作成者に帰属し、新たな技術も、透明かつ説明可能でなければならない。
- ・ AIによる社会価値の創出には、機能や性能と同時に、社会から信頼を獲得し、受け入れられることが重要である。社会に信頼されるAIソリューション提供の仕組みは市場拡大にとって有効であるが、AIガバナンスがイノベーション創出の阻害要因や参入障壁とならないための配慮が不可欠である。

⁷⁸ 英語版については、次に掲げるURLのウェブサイトに所掲。

<https://www.soumu.go.jp/main_content/000811496.pdf>

⁷⁹ 講演やパネルディスカッションの概要については、別紙1と同じ内容のものが既に総務省ウェブサイトに掲載されているが、本章における記載については、それから一部加筆・修正をしているところがある。

⁸⁰ 前掲注(20)参照。

- AIの規制に対して、「ハードロー」や「ソフトロー」に関する議論が行われているが、二分法的な議論ではなく、これから顕在化してくる多様で具体的なリスクに対して、合理的な規制は何か、知恵を出し合うことが重要である。
- EUのAI規制法案について、転ばぬ先の杖がよいという価値観は理解できるが、イノベーションの遅れによるリスクも危惧される。規制内容の精査が必要である。
- AIは発展途上の技術であり、技術もリスクも社会の受容性も時間とともに変化していく前提で考えることが重要ではないか。
- EUのハイレベルな考え方は重要であるが、民間の立場からは、ユースケースベースの重要性を主張し、協調しつつ、現実的に遵守可能な形を作っていくことが大事である。
- プライバシーや人間の尊厳に対する考え方、働き方など、国や地域で違いがある中で、そのような違いをどのように扱い、調整するかは非常に難しい課題である。
- AIの規制については、動かしながら状況に応じて変えていくべきであり、一番厳しい規制に合わせる必要性も出てくるのではないか。また、相互運用性の確保が重要である。
- EUにおいても公共分野でAI利活用が進んでいるケースもあり、AIは、使う側のメリット等の理解を得ながら進めることが重要である。
- EUとの協調路線を模索することが重要であると考えられるが、そのためにも人材育成や議論が不可欠である。
- エンジンとしてAIを利活用するだけでは暴走するので、人間中心の社会においては、「ELSI（エルシー）」（倫理的・法的・社会的課題）の検討にハンドルとブレーキの役目を担わせることが重要である。
- 各国の法制、産業構造、行政システム等に違いがあることをEU等に対して明確に示した上で、国際協調を進めていく必要がある。

今後、このような意見を踏まえ、AIの社会実装の推進に向けた取組を継続していくことが重要である。

第2章 「安心・安全で信頼性のあるA Iの社会実装」の推進の取組（取組事例の取りまとめ）

本推進会議において、これまでA I開発ガイドライン（2017年）やA I利活用ガイドライン（2019年8月）を策定してきたが、A I利活用ガイドラインを取りまとめた後は、「安心・安全で信頼性のあるA Iの社会実装」を進めるために、A Iの開発者やサービスプロバイダー、ビジネス利用者、消費者的利用者などのステークホルダー⁸¹からヒアリングを行い、「報告書2020」（2020年7月）及び「報告書2021」（2021年8月）において、その取組等を取りまとめて公表した。

本章においては、「報告書2021」の公表後、2021年11月から2022年3月まで行った本推進会議議長によるヒアリング（以下「本ヒアリング」という。）を踏まえて、事業者、有識者、関係団体等（以下「事業者等」という。）の取組を紹介するとともに、「安心・安全で信頼性のあるA Iの社会実装」の推進に向けた留意点（グッドプラクティスのポイント）を整理した⁸²。

なお、各事業者等の取組については、別冊（グッドプラクティス集）に取りまとめているので、参照されたい（各事業者等の発表資料については、総務省ウェブサイトに掲載⁸³）。

1. ヒアリングの概要

本推進会議は、「報告書2021」を取りまとめた後、本ヒアリングにおいて、A Iの社会実装に関して先進的あるいは意欲的な取組を行っている事業者等からの発表をもとに意見交換を行った。本ヒアリングにおける主な論点については、今後の「安心・安全で信頼性のあるA Iの社会実装」の推進の観点から、「報告書2020」におけるヒアリングの論点を踏襲し⁸⁴、具体的には、

- 開発者や利用者（A Iサービスプロバイダー、ビジネス利用者）が、どのような取組を行うことにより、「安心・安全で信頼性のあるA Iの社会実装」が進むか、あるいは、社会における受容性が向上するか。
- それらの取組を進めるために、事業者等において、どのような課題があり、課題解決のために何をすべきか。
- 社会における受容性の向上を図り、「安心・安全で信頼性のあるA Iの社会実装」を進めるために、どのような環境整備を図っていくことが必要か。

といった内容を中心に、意見交換を行い議論を深めた⁸⁵。

⁸¹ 前掲注（2）参照。

⁸² 「報告書2020」において、今後も継続的にヒアリングを行い、その内容を「共有知」として整理することが必要である旨が指摘されている（「報告書2020」結びに代えて参照。）。

⁸³ 公表の承諾が得られたものについては、次に掲げるURLのウェブサイトに所掲。

<https://www.soumu.go.jp/main_sosiki/kenkyu/ai_network/02iicp01_04000232.html>

⁸⁴ 「報告書2021」に掲載されている取組事例を取りまとめるために行ったヒアリングにおいても、同様の論点を踏襲していた。論点の詳細については、「報告書2020」はじめに参照。

⁸⁵ 更に具体的に、人材育成やA Iサプライチェーンといった観点からヒアリングを行ったケースもある（特に第5回ヒアリング（2022年3月24日）においては、A Iサプライチェーンの観点から意見交換を行い、議論を深めた。）。

本ヒアリングの実施状況は、次表のとおりである。

	発表者	タイトル
2021年		
第1回 ヒアリング (11月24日)	ソフトバンク株式会社 【ソフトバンク】	ソフトバンクAI活用事例ご紹介
	パナソニック株式会社 【パナソニック】	パナソニックにおけるAI開発の考え方と活用事例
	シャープ株式会社、株式会社AIoTクラウド 【シャープ】	シャープグループのAIへの取り組み
第2回 ヒアリング (12月22日)	中山敬一(九州大学生体防御医学研究所教授) 【中山教授(九州大学)】	人工知能(AI)の医学生物学への応用
	植田琢也(東北大学医療AI人材育成拠点プログラム) 【植田教授(東北大学)】	Global×Localな医療課題解決を目指した最先端AI研究開発 人材育成教育拠点-Clinical AI-
	株式会社メルカリ 【メルカリ】	メルカリグループにおけるAIに関する取り組みのご説明
2022年		
第3回 ヒアリング (1月25日)	一般社団法人日本データマネジメント・コンソーシアム 【JDMC】	実務者が考えるAI・データ活用における倫理フレームワーク
	三井住友海上火災保険株式会社 【三井住友海上】	三井住友海上のデジタルイノベーション推進取組
	ダイキン工業株式会社 【ダイキン工業】	ダイキン工業におけるAI人材育成の取り組みについて
第4回 ヒアリング (2月15日)	西日本旅客鉄道株式会社 【JR西日本】	JR西日本におけるデータアナリティクスの取組
	KDDI株式会社 【KDDI】	KDDIにおけるAI事例とAIガバナンスに向けた取組み
	有限責任あずさ監査法人 【あずさ監査法人】	AIの適切性検証への取組み
第5回 ヒアリング (3月24日)	株式会社エヌ・ティ・ティ・データ 【NTTデータ】	AIに係るサプライチェーンリスクとNTTデータの対応
	富士通株式会社 【富士通】	AIサプライチェーンに対する富士通の取組
	一般社団法人AIデータ活用コンソーシアム 【AIDC】	データ活用と課題と取組み

(注) 事業者等の名称はヒアリング当時のもの。括弧(【】)内は、本文における略称。

本章における記述については、別冊（グッド プラクティス集）第2節（「2. 各事業者等における取組」）の各事業者等の取組の概要並びにヒアリングにおける発表及び意見交換の概要の記述をもとに⁸⁶、本章第2節（「2. 本ヒアリングにおける発表・意見交換のポイント」）において、「A I 倫理・ガバナンスに関する取組」、「A I 開発・利活用に関する取組」、「人材育成に関する取組」及び「A I サプライチェーンに関する取組」の観点から、A I 開発ガイドラインやA I 利活用ガイドラインにおける原則の項目ごと、A I 利活用の分野ごとなど横断的・横串を通した形で整理した。なお、これらの項目や分野等をまたがる取組については、内容がより近いと考えられる項目や分野等に記述することとしている（場合によっては、再掲の形で重複して記述することもある。）。

また、「安心・安全で信頼性のあるA I の社会実装」の推進の観点から、これからA I の開発や利活用を始めようとする者、A I の開発や利活用を行っているが何らかの課題に直面している者、A I の開発や利活用を更に積極的に進めようとする者などにとって参考となるよう、課題認識やその解決方策など共有を図ることが特に有益・有用と考えられるものを中心に整理した。

なお、記述の表現振りについては、実際のヒアリングにおける発表及び意見交換を踏まえたものになっていることに留意されたい。

本章の記述は、本ヒアリングにおける発表及び意見交換等をもとに整理したものであるが、各事業者等においては、記述されている取組のみならず、様々な取組を行っている。記述されていない取組についても、今後、必要に応じて、ヒアリング等を行い、「安心・安全で信頼性のあるA I の社会実装」の推進に役立てていく考えである。

2. ヒアリングにおける発表・意見交換のポイント

（1）A I 倫理・ガバナンスに関する取組

各事業者等のA I 倫理・ガバナンスに関する取組について、「指針・ガイドライン・原則」、「組織・体制」、「セキュリティ」、「プライバシー」、「公平性」、「透明性・アカウントビリティ」、「適正利用」、「品質保証・開発レビュー」及び「外部との連携・協働」の観点から整理した。

<グッドプラクティスのポイント>

- A I の開発者やサービスプロバイダーにおいて、A I 開発ガイドライン、A I 利活用ガイドライン等を参考に、A I 倫理・ガバナンスを目的とした原則を策定・制定する取組が見られる。

A I の開発者・利用者等は、本ヒアリングにおいて示された事例を参照しつつ、A I 利活用の様態やA I の性質等に応じて、指針・ガイドライン・原則等を策定・制定し、適切に実践・運用することが望ましい⁸⁷。

- A I の開発者やサービスプロバイダーにおいて、A I 倫理・ガバナンスを実践・運用するために、倫理委員会を設置したり、社長直下に専門部署を新設したりする取組が見られる。

⁸⁶ 別冊（グッド プラクティス集）について、別冊だけで1つの事例集として完結するように作成したため、別冊第1節は、本章第1節と同じような内容の記述となっている。

⁸⁷ 「報告書 2021」においても、様々な具体的な取組事例が紹介されていることから、本報告書と合わせて参照すると、指針・ガイドライン・原則等の策定・制定の参考になるものと考えられる。また、指針・ガイドライン・原則等の策定・制定のみならず、「組織・体制」、「セキュリティ」や「プライバシー」など他の項目についても、「報告書 2021」において、具体的な事例が紹介されていることから、本報告書と合わせて参照されたい。「報告書 2021」第3章2.（1）、4. 及び別紙3参照。

AIの開発者・利用者等は、本ヒアリングにおいて示された事例を参照しつつ、AI倫理・ガバナンスを適切に実践・運用するための組織・体制を構築することが望ましい。

- セキュリティの確保、プライバシーの保護、公平性の確保やバイアスの排除を重視した取組が見られる。特に、画像認識において、エッジ側で画像処理を行って認識結果だけをクラウドにアップロードしたり、骨格を推定することにより、AI側に個人情報が残らない形で人間を識別する取組は注目すべきものである。

AIの開発者・利用者等は、本ヒアリングにおいて示された事例を参照しつつ、AI利活用の様態やAIの性質等に応じて、セキュリティの確保、プライバシーの保護、公平性の確保やバイアスの排除を実現するための適切な取組を行うことが望ましい。

- AIの判断根拠を可視化したり、過去の状況をトレースするなど透明性の確保・アカウントビリティを重視した取組が見られる。特に、ウェブサイトにて個人情報の利用目的として技術開発に関することをきちんと明示して、エンドユーザーに対してアカウントビリティを果たす取組は注目すべきものである。

AIの開発者・利用者等は、本ヒアリングにおいて示された事例を参照しつつ、AI利活用の様態やAIの性質等に応じて、透明性の確保、アカウントビリティを果たすための適切な取組を行うことが望ましい。

- 多くの事業者等が、AIの判断をそのまま使う、すべてAIに委ねるという運用ではなく、AIを人間が使うツールとして位置付けて、AIの判断を最終的に人間が確認する、人間をサポートするためにAIを使うという、いわゆる「human in the loop (人間参加型)」⁸⁸といわれる運用を行っている。これは、プライバシー、公平性、透明性・アカウントビリティなどとも密接に関連するとともに、人間とAIとの関係を考える上で、非常に重要な観点である。

AIの開発者・利用者等は、本ヒアリングにおいて示された事例を参照しつつ、AI利活用の様態やAIの性質等に応じて、「human in the loop」を実現するための仕組みを構築し、適切に運用することが望ましい。

- 自社において品質保証や評価を実施するためのプロセスやルールを定める取組に加えて、第三者によって、評価を行う仕組み（チェックリスト、フレームワーク）が提供されたり、モニタリングを行うサービスが提供される取組は注目すべきものである。このような取組は、セキュリティの確保、プライバシーの保護、公平性の確保やバイアスの排除などを実現するための方策の1つとして有用・有効であるものと考えられる。

AIの開発者・利用者等は、本ヒアリングにおいて示された事例を参照しつつ、品質保証や評価などを適切に実施する取組を進めることが望ましい。なお、自社のみでは、このような仕組みを構築することが困難である場合には、外部が提供する仕組み・サービスを積極的に活用することが期待される。

- 政府、関係団体、海外を含む外部の専門家・有識者等と連携・協働を行っている取組が見られる。このような外部のステークホルダとの連携・協働により、事業者等自身の取組が深化するとともに、AIの社会実装の推進に貢献しているものと考えられる。

AIの開発者・利用者等は、本ヒアリングにおいて示された事例を参照しつつ、積極的に外部と

⁸⁸ここでは、「人工知能などによって自動化・自律化が進んだ機械やシステムにおいて、一部の判断や制御に人間を介在させる」という意味で用いている。

の連携・協働を図り、マルチステークホルダによるAIの社会実装を進めていくことが望ましい。

ア 指針・ガイドライン・原則⁸⁹

【パナソニック】

- 安心・安全を顧客に届けるための最低限の約束として、AIでやってはいけないことのガイドラインを独自に設定し、AI倫理原則を策定している。

【KDDI】

- AI利活用とガバナンスを両立させる1つの要素として、信頼できるAIが重要であり、このために、「原則の策定・公開」、「研究開発」、「社内運用の整備」の3つの取組を行っている。原則の策定・公表については、2021年8月に、総務省が公表しているAI開発ガイドライン、AI利活用ガイドラインを参考にしながら、適切なAIの開発・利活用のために重要と考えられる9つの原則を策定・公表した。

イ 組織・体制

【パナソニック】

- 社内にAI倫理委員会を設置しており、AI倫理の徹底とアップデートする仕組みを早期に立ち上げるべく取組を進めているところである。

【富士通】

- AI倫理の推進に当たって、自社の規律を適切に行うとともに、社内外に浸透させるための継続的な取組が重要であると考えており、2022年2月に、「AI倫理ガバナンス室」を社長直下に新設した。

ウ セキュリティ

【NTTデータ】

- セキュリティについて、外部から調達したモデルやプログラムがどのようなものをベースに作られたのか分からず、バックドア的なものが含まれているリスクがある。これに対しては、ソースコードを確認する、提供元の信頼性をもとに判断するといった対応にならざるを得ず、法務審査・レビューをきちんとやるというのが企業側として現実的な方法であるものと考えている。

また、バックドアによってデータが抜き取られるというリスクもある。これに対しては、支配的なコンポーネントやプロダクトに関して、公的なお墨付きのようなものがあると利用する側としては安心できるのではないかと考えている。

エ プライバシー

【パナソニック】

- 自律アップデート型エッジデバイス&サービスについて、エッジ側だけで画像処理を行い、認識

⁸⁹ 本ヒアリングの対象の事業者等のうち、富士通、NTTデータについては、既にヒアリングを実施し、「報告書2020」に記載済みである。「報告書2020」第3章参照。

結果のメタデータだけをクラウドにアップロードすることによって、プライバシーの問題と伝送量逼迫の問題を回避することができる。

【JR西日本】

- 駅の監視カメラの画像を使って、車椅子に乗っていたり、白杖を持っていたり、あるいは、盲導犬を連れているなど介助を必要とする利用者を素早く検知して、介助・手伝いができないかというところから開発を始めている。なお、AIで判断しているのは、骨格推定のところであり、顔認証や行動追跡等により個人を特定するものではなく、個人情報はAI側に残らないようになっている。
- 防犯カメラの活用については、センシティブであり、非常に気を遣うところである。ウェブサイトに、個人情報の利用目的を掲載しており、その中で、技術開発に関することも記載されている。AIの開発に役立てるとともに、開発したAIに関しては、個人の行動ログを収集・分析して特徴を捉えることや顔認証はやっておらず、個人情報を含まない形でAIを活用している。

【NTTデータ】

- プライバシー、知的財産について、目的外利用のリスクがあり、様々なデータをAPIを通じて提供しているが、それが再利用されてしまうリスクがある。これに対しては、利用規約できちんと確認することが重要である。

オ 公平性

【ソフトバンク】

- データバイアスを完全に除去することは難しく、幅広いデータを大量に集めて精度を高めていくという地道な努力を続けることが重要である。また、人種や性別などによる差異が出ないように技術的に工夫しているが、AIによるバイアスが問題になるところでは、人間が対応して、カバーするようにしている。

【NTTデータ】

- データバイアスに関するリスクがあり、正解ラベルにバイアスがあると、AIが公平性に欠ける推論をしてしまうことがあり得る。これに対しては、男性か女性かといった人間の普遍属性に関わるパラメーターが、どの程度出力に影響を与えているのか確認することが重要である。

カ 透明性・アカウントビリティ

【パナソニック】

- AIはブラックボックスであるため、学習データとそれにより得られた結果の因果関係を顧客の納得を得ながら開発を進める必要があり、品質保証のプロセスを顧客と共同で構築する必要がある。
- AIの判断根拠を高精度に可視化して誤検知対策を導入している。これまで勘と経験で推定していた誤検知の要因が瞬時に特定することができるようになり、対策期間を半分に短縮することができた。

【メルカリ】

- AIの判断や推定結果が説明可能かという点に対して、これを防ぐために、過去の状況をトレー

スできるように、実験計画を作成して情報を保存することとしている。

実験計画には、背景や内容、評価指標や評価方法などを記載することになっている。実験内容は社内でオープンになっており、関連部署や他のエンジニア等からコメントをもらえるようになっていたため、リスクの顕在化を防げることに一役買っている。

- 与信審査モデルというセンシティブなサービスを提供しており、モデルの性能評価や特徴量の効果等を厳密に精査している。与信審査の結果に対して、説明可能な状態を作ろうと様々な取組を行っている。

【JR西日本】

- 防犯カメラの活用については、センシティブであり、非常に気を遣うところである。ウェブサイトに、個人情報の利用目的を掲載しており、その中で、技術開発に関することも記載されている。

【KDDI】

- 説明可能AIについて、今のところ、ユーザーにとって事後的に推論根拠を説明することを中心に研究を行っている。推論根拠として知識グラフを活用するということを特徴としながら研究を進めているのが現状である。

キ 適正利用

【ソフトバンク】

- データバイアスを完全に除去することは難しく、幅広いデータを大量に集めて精度を高めていくという地道な努力を続けることが重要である。また、人種や性別などによる差異が出ないように技術的に工夫しているが、AIによるバイアスが問題になるところでは、人間が対応して、カバーするようにしている。
- 利用用途において、例えば、採用の可否の判定に直結するといった影響を与えることがないような使い方をすることが第一の対策である。人間がエントリーシートを読む工数を削減することが目的であり、AIに可否の判定を決めることを求めているわけではない。
- 新卒採用選考におけるAI利活用について、AIが不合格と判断したエントリーシートについては、必ず人間が確認するようにしており、AIの判断だけで不合格になることはない。また、面接についても、不合格の判定が出たら人事担当者が実際に確認するようにしており、面接の可否判断がAIだけで行われることはない。
- AI利活用として、大量に処理する、同じような処理をするなど人間が行うと手間、時間、コスト、労力を必要とする業務については、AIを適用しやすい。他方、人間による判断が重要である、あるいは、判断そのものが難しい業務については、AIの適用が難しい面がある。また、そのような業務にAIを適用する場合には、人間の目で最終確認を行うプロセスを経るようにしている。
- 問題が発生することが許容できない業務やタスクについては、途中で人間が入ってカバーすることにしており、その分の工数がかかることになるが、このような取組によって100%と言えないAIの精度をカバーして実用化を進めることとしている。

【パナソニック】

- AIのビジネス実装を進めるに当たっては、AIをツールとみなして積極的に活用する、必要な

技術を見極めて現場での困りごとを解決するというスタンスである。データとドメイン知識を有する技術者（大工）が、道具であるAIを使いこなし、困りごとを解決するということで、DAICC（Data&AI for Co-Creation：ダイク）というコンセプトで推進している。

【メルカリ】

- 自社ルールに違反する取引や不正な取引について検知できるようにする必要があり、出品や入金の特ランザクションから、リスクがありそうな行動を特定することができるモデルを作った。AIがリスクのある行動をしている利用者を検知し人間による目視確認、追加確認措置を実施し、不正の有無を確認するようにしている。
- 模倣品の検知について、画像や説明文から識別するアプローチをとっているが、誤検知や抽出漏れの可能性があるため、AIが検知したものを人間が最終的に目視でチェックをして判断するという対策を行っている。
- ヒューマンセントードという考えもあり、誤検知の場合には、顧客にペナルティを与えてしまうことになるため、人間による一定のチェックは必要である。他方、人手、コストの問題があるので、今後の検討課題である。

【三井住友海上】

- 個人データが保険商品や保険の金額に影響するかという点について、例えば、自動車事故においては、その結果だけではなく運転の挙動と事故との因果関係の検証が必要であるため、運転データからは、保険会社としての保険料の反映に行き着かないのが実態である。日本では基本的な保険料を算出するための料率に関する法律に基づいてベースの料率が算出されている。世の中の事故そのものをいかに減らしていくかということの方が重要であると考えている。

【JR西日本】

- 人間とAIとのバランスについて、安全上、判断のミスが許されない保安装置などにAIを導入することは、まだまだハードルが高いと考えている。基本的には、これまで人間が検査していたことに関し、人間では見えないところがあったり、ルールの解釈にブレがあったりする中で、AIを活用すると7割程度の回答が出てくるので、人間がやるよりも高頻度でデータを取って判断の回数を増やすことによって業務を置き換えていくという考え方でやっている。
- 最終的な判断は人間に任せるというところで、人間を支援するツールとしてAIを活用している。

【KDDI】

- AIかん水施肥システムについて、AIで自動化できる作業があると、人間は品質向上に係るような作業に注力することができる。人間がやる部分とAIに任せ部分とうまく分けながら、収量増と高品質化を達成するという取組を行っている。

ク 品質保証・開発レビュー

【JDMC】

- 倫理フレームワークは、情報を取得して実際にデータを活用するといったビジネスプレイヤーを想定したフレームワークとなっている。なお、AIの利活用などによって、取得したデータそのも

ののほか、派生データが増えるとともに、それらが多段階的に活用されたり、外部に提供されたりしていることに注目して問題点を整理している。

- 実務的に有用であることが重要であることから、実際に使えるものということを考えて、日本において問題が発生している事例を集めて組み上げた形となっている。倫理という概念から演繹的に具体化していくよりも、帰納的に多くの事例を集めることから始めて、事例を抽象化して、チェックポイントや倫理的な観点にまとめていき、それを実際のビジネスの事例に当てはめて、使いやすいかどうかということを繰り返し検証して、フレームワークを作成した。
- 倫理フレームワークの全体の構成として、5つの倫理項目に整理した。一番下に土台となる「情報保護」があり、その上で、「安全・安心」を確保し、「公平性」、「説明責任」のあるデータ活用を実現していき、これらによって全体として「正当性」が示されるという考え方で整理している。
- 倫理フレームワークは、非常に網羅性があるものと認識しており、実際のビジネスに照らして使ってみると、非常に細かい点についても記入することになり、このようなフレームワークを活用することによって、想定していなかった問題・課題を洗い出すことができるのではないかと考えている。
- 実際の活用プロセスとしては、チェックポイントによって洗い出したリスクを全体的に俯瞰し、それぞれのリスク要因を把握した上で、最終的にリスクをどのように低減するのかを策定するというプロセスを検討している。

【KDDI】

- 社内運用の整備について、指針を具体的にどのように運用するのか、どのように守っていくのかというところで、従前からプライバシーに関しては、PIA (Privacy Impact Assessment : プライバシー影響評価) を行っており、プライバシーのほかにも、公平性・非差別、アカウントビリティ、透明性、人間の判断の介在などガバナンスが必要なポイントについては、PIAの中で、同時にチェックするところから始めていこうと考えている。その後、チェックできていない領域をどのようにするか議論を行っていこうと考えている。

【あずさ監査法人】

- 従来は、モデルを1回作った後は、多少のバージョンアップなどはあるものの、それほど大きく検証する必要はなかったが、AIの場合は、継続的な学習を通してモデルが変容していくので、変容の仕方について、オーバーフィッティング（過学習）の問題にも留意して検証しなければならない。
- 継続的な学習がAIの特徴であるため、従来の監査のように1年に1回ということではなく、AIモデルの検証は、それをモニタリングする監査システムを同時に実装して継続的にモニタリングしていく枠組みが必要ではないかと考えている。
- AI検証のフレームワークについて、AIに求められる要素として、例えば、「正確性」、「説明可能性」、「差別の排除」、「迅速かつ堅牢」というものに照らして、AIに何をやらせるのか、そのためどのような準備をするのかといったインプットの部分、また、その結果としてAIが生み出すもの、それが期待どおりの結果となっているのか確認するといったアウトプットの部分について、問題がないか確認する枠組みとなっている。
- 管理フレームワークとして、このような「ソリューション&データマネジメント」だけではなく、

その前提となる「プロジェクトマネジメント」、事業者全体としての「エンタープライズマネジメント」、技術要素としての「テクノロジーマネジメント」という様々なレイヤーで多層的、多重的に検証することが重要である。

- 検証の観点について、「公平性」、「説明可能性／解釈可能性」、「AIモデルの精度」、「セキュリティ／データ保護」、「追跡可能性／監査可能性」、「事業継続性」、「データ品質」、「ガバナンス」という8つの観点で整理しているが、様々な要素を複合的に確認することが重要である。
- 特に、どのような考え方でAIを作ろうとしているのか、どのように実現しようとしているのかということについて「ガバナンス」を効かせるという8番目の観点は大きなポイントではないかと考えており、実際に開発されたもの、運用されているものと齟齬がないか確認することが検証の中で大きなウェイトを占めるのではないかと考えている。
- 検証のプロセスについて、「戦略・ガバナンス」から始まり、「要件定義」、「データ準備・前処理」、「モデル構築」、「モデルの検証」、「モデルの実装と運用／監視」というところで、多重的に検証していくことになるが、すべてやらなければいけないということだけでなく、リスク感度に応じて重点的なところを一定の頻度で実施していくことになるのではないかと考えている。

【NTTデータ】

- 品質について、外部調達したものの精度が求めているタスクに合っているかというリスクがある。これに対して、現状としては、タスクに応じて、P o Cにおいて精度を検証するしかないが、将来的には、AIの品質基準が定められて一定の目安ができることが望ましいと考えている。

【A IDC】

- AI・機械学習の品質の偏りに関して、データに起因するものが多いが、データのアノテーション（ラベル作業）をアウトソースしている場合、悪意を持った作業者がいることも否定できず、問題が発生した場合に、責任の所在をどのように考えるべきか、また、データ取引における契約において、このようなリスクに対して、どのような形としておくべきなのかといった検討を行ってきた。

ケ 外部との連携・協働

【パナソニック】

- AIは極めて進化のスピードが速い分野であるため、いち早く実用化の道筋が見えやすい技術を中心に必要な最先端の技術を取り込むことが重要であり、スタンフォード大学やバークレー大学といった世界トップレベルの大学との連携を深めている。

【三井住友海上】

- 組織体制について、保険会社が単独ではなく、社外の顧客、取引先など様々なステークホルダーと一緒にイノベーションを進めていくことが必要であるとして、ビジネスイノベーション部というアライアンスを行う部門を新設した。

【A IDC】

- 知的財産、AIに関する研究、データ収集、データ基盤の構築等に関する活動を行っており、様々な大学、研究機関、行政機関等と連携して取組を進めている。

(2) AI開発・利活用に関する取組

各事業者等のAI開発・利活用に関する取組のうち、新型コロナウイルス感染症（以下「新型コロナ」という。）の感染拡大が続いていることや課題先進国として国際的に情報発信することが重要であるといった観点から、特に注目すべき分野として、「新型コロナ対策」、「医療・ヘルスケア」及び「高齢者・障害者」に関する取組について整理した。

<グッドプラクティスのポイント>

- 混雑状況の可視化や非対面・非接触の配送など新型コロナ対策として、AIを利活用した取組が見られる。

AIの開発者・利用者等は、本ヒアリングにおいて示された事例を参照しつつ、新型コロナの感染拡大の状況を踏まえ、引き続き、AIを利活用した新型コロナ対策の取組を推進することが望ましい⁹⁰。

- 医療・ヘルスケア分野において、がんの予後の正確な推定や創薬の取組、デザイン思考を取り入れた医療とAI・データサイエンスをつなぐ人材育成の取組⁹¹は注目すべきものである。このような取組が進むことによって、医療側、患者側など多方面にメリットがもたらされることが期待される。

AIの開発者・利用者等は、本ヒアリングにおいて示された事例を参照しつつ、医療・ヘルスケア分野におけるAIの開発・利活用等の取組を推進することが望ましい。

- 聴覚障害者とのコミュニケーション、高齢者施設における見守りや歩行支援、鉄道の駅での要介助者の早期発見など、多くの事業者等が、AIを利活用して高齢者・障害者を支援する取組を行っていることは注目すべきものである。

これからの日本において、高齢者・障害者を支援する取組の重要性・必要性は、ますます高まっていくことが想定され、AIの開発者・利用者等は、本ヒアリングにおいて示された事例を参照しつつ、高齢者・障害者を支援するためのAI利活用の取組を推進することが望ましい。

ア 新型コロナウイルス感染症対策としてのAI開発・利活用

【ソフトバンク】

- 人の混雑状況を示す満空カウントにおいては、プライバシーを保護するため、人がいることは分かるが、誰がいるかは分からないようにアイコンを表示して、混雑状況を可視化している。この技術は新型コロナ対策として、オフィスや会議室の混雑状況の可視化にも利用することができ、社外にも提供する予定である。

⁹⁰ 「報告書 2021」においても、様々な具体的な取組事例が紹介されていることから、本報告書と合わせて参照すると、新型コロナ対策としてのAI開発・利活用の参考になるものと考えられる。また、新型コロナ対策としてのAI開発・利活用のみならず、「医療・ヘルスケアに関するAI開発・利活用」及び「高齢者・障害者に関するAI開発・利活用」についても、「報告書 2021」において、具体的な事例が紹介されていることから、本報告書と合わせて参照されたい。「報告書 2021」第3章2. (2)、4. 及び別紙3参照。

さらに、AI利活用により障害者を支援する取組については、次に掲げるURLのウェブサイトにも事例が紹介されているので、参照されたい。

<<https://www.minnanosyougai.com/article1/ai> で障がい者の生活は大きく変わる？>

⁹¹ この点について、次項（「(3) 人材育成に関する取組」）において述べている『技術的な専門性に加えてAIやIoTのことも分かる人材の育成を目指す取組』と共通するものであり、注目すべきものと考えられる。後掲注(95)参照。

【パナソニック】

- 小型低速ロボットを活用した住宅街向け配送サービスの実証実験を行っている。配送員不足の深刻化に加えて、非対面・非接触など新たな生活様式への対応が求められている中、これまで培ってきたロボット技術のノウハウを活かして、新たな配送サービスの実現に向けた取組を加速していく。

イ 医療・ヘルスケアに関するAI開発・利活用

【中山教授（九州大学）】

- AIを活用して、正確ながんの予後の判定システムを作った。約2万種類ある人間の遺伝子の発現量を調べたところ、286種類の遺伝子について、発現量の高低により予後に差があることが分かった。研究を進めると、このうち184種類の遺伝子については、人種や環境等に関係なく予後に影響を与えることが分かった。
- さらに、184種類のうち23種類の遺伝子があれば、有意なデータが得られることが分かったため、23種類に絞り込むとともに、それぞれ重み付けを行い最適なパラメーターを決めて、m P S (Molecular Prognostic Score : 分子予後スコア) を作った。m P S を6つの群 (A群～F群) に分けて予後を追跡すると、それぞれの群がきれいに予後と合致した。
- がんの予後を正確に推定することができると、①患者の目線で、自分があと何年生きることができるか正確に分かる、②医師の目線で、治療方針の検討に役立つとともに、無駄な治療や過剰な治療をなくすことができる (例えば、A群、B群の患者に対して、強い抗がん剤を投与する必要がなくなる。)、③そのほか、例えば、生命保険において、きちんとした保険料率を算出して、患者に適した保険を作ることができるようになる。
- 薬を創るためには、「タンパク質と化合物が結合するか」、「どの程度の力で結合するか」を調べる必要があるが、人間のタンパク質は約2万種類、低分子化合物は1060種類あり、調べるのは非常に困難である。
- 1次元構造 (ひも状) が折りたたまれて3次元構造 (まり状) になっていることに着目して、1次元構造においても、対応関係があるのではないかと考え、タンパク質と化合物を1次元の数値ベクトルに置き換えて、結合力と関連性を判定するソフト「LIGHTHOUSE」を作った。
- LIGHTHOUSEは、3次元のドッキングシミュレーションとほぼ同じ正確性で判定することができる。また、3次元の構造情報がなくても判定することができるので、非常に広範囲、かつ、大量にスクリーニングすることができる (LIGHTHOUSE : 6,000種類/分 (パソコン利用時)、ドッキングシミュレーション : 2～3種類/分 (スパコン利用時))。
- LIGHTHOUSEの活用により、新たな抗がん剤の候補となり得る化合物や新型コロナウイルスに有効である薬を発見することができた。

【植田教授（東北大学）】

- 病院の中にデータサイエンティストを入れて、医療の価値などを協働しながら創っていこうと、病院の中の1つの階を整備して、実証スペース「OPEN BED Lab」を作った。ここでは、医療現場の視点を取り入れたデザイン思考に基づくAIの研究開発を行うことを重視している。
- デザイン思考というのは、人間を中心に据えたイノベーションを創出するアプローチであり、医療においても、技術先行ではなく、デザイン思考を取り入れて、医療の現場で医師が大切にす

いる医療的な価値を最初に考えて、それとマッチするA Iとの融合点を見つけることが重要である。

この融合点を見つける人材を育成することが重要であり、大学学部教育(医学部)、病院初期研修、大学院教育の3段階で一貫した教育プログラムの中で、医療人でありながらデータサイエンスが分かる人材、両方の融合域を見つけることができる人材を育てようとしている。

- デザイン思考を学ぶプログラムを取り入れており、例えば、病院の待ち時間を減らすにはどのようなセッティングがよいのかといった一般的なデザイン思考の過程を学ぶプログラムのほか、バイオデザインそのもの、A Iを開発するためにどのようなことに留意する必要があるのかといった過程を学ぶプログラムとなっている。
- 東北大学、北海道大学、岡山大学が連携して、GLOCAL (Global×Local) なClinical A Iというプロジェクトを進めている。医師不足や高齢化の進展など医療課題が多いのが都市部ではなく地方であるため、地方の大学が連携して、このような課題に対応するために、ソリューションとしてA Iがどのようにマッチするのかということを考えられる人材を育てようとしている。なお、大学や病院ごとに特色があり、東北大学は病院が中心、岡山大学は薬学部系が中心になっている。
- 医療側は医療的な価値を知っており、データサイエンス側は洗練されたアルゴリズムを持っているが、両者をつなぐ人材が少ない。両者をつなぐ人材、間に入ってハブとなる人材を育てるプログラムとなっており、医療現場の感覚、医者への価値観を持ちながら、その中でデザイン思考を活かしてA Iの研究を行う人材を育てたいと考えている。

ウ 高齢者・障害者に関するA I開発・利活用

【ソフトバンク】

- 聴覚障害者と健聴者が当たり前でコミュニケーションすることができる社会にすることを目指して、手話と音声でリアルタイムに会話することができるシステムの構築を行っている。このシステムでは、端末のビデオ通話において、(聴覚障害者から健聴者へは) A Iが身体動作を追跡して手話の特徴を抽出してテキストに変換し、また、(健聴者から聴覚障害者へは) 音声を自動でテキスト化することによってコミュニケーションを行うことができるようになっている。

【パナソニック】

- 高齢者施設向け介護分野において、見守り安心サービスとして、非接触でルームセンサーが入居者の体動、心拍、呼吸等を検知し、生活リズムをロギングして分析することによって、生活の改善、睡眠の改善等を実現している。また、高齢者向けの歩行支援ロボットは、個別の状況に応じて、最適な負荷を調整しながら歩行トレーニングを支援している。

【JR西日本】

- 駅の監視カメラの画像を使って、車椅子に乗っていたり、白杖を持っていたり、あるいは、盲導犬を連れているなど介助を必要とする利用者を素早く検知して、介助・手伝いができないかというところから開発を始めている。

【KDD I】

- A Iかん水施肥システムについて、宮城県東松島市において、グループ会社のKDD I エボルバが既存のアプリケーション・機器を導入してやっており、農産物の栽培を通じて復興支援と障害者

の雇用創出に取り組んでいる。水やりや肥料を与えるタイミングや量は、生産者の経験とノウハウが重要であるが、AIで補うことによって、農業経験によらず安定した農作物の栽培環境を実現するべく障害者とともに取り組んでいる。

【AIDC】

- 脳性麻痺、視覚障害者が話す言葉について、健常者は聞き取るのが難しいが、支援者は理解することができる。何か共通点、相関があるのではないかと考えられ、AIによって、このような課題を解決することを1つの目指す姿としている。

(3) 人材育成に関する取組

AIに関連する人材の不足が指摘され、人材育成・確保が課題となっている中⁹²、各事業者の人材育成⁹³に関する取組⁹⁴について整理した。

<グッドプラクティスのポイント>

- 多くの事業者が、積極的にAI人材、デジタル人材の育成に力を入れて取り組んでおり、着実に成果を上げている取組も見られる。また、大学との連携、段階別のカリキュラムの設定、きめ細やかなレベル分けを行った上での育成、技術的な専門性に加えてAIやIoTのことも分かる人材の育成を目指す取組⁹⁵などが行われており、人材育成の在り方や手法を検討する上で、参考になるものと考えられる。

特に、社外向け（高校生、小学生向け）にAIリテラシー教育を提供している取組は注目すべきものであり、このような取組が進むことによって、AIリテラシーの底上げにつながっていくことが期待される。

AI人材について、日本の技術力、国際競争力を高めるためには、人材育成は急務の課題であり、AIの開発者・利用者等は、本ヒアリングにおいて示された事例を参照しつつ、AI人材の育成・確保の取組を強力に推進することが期待される⁹⁶。

【ソフトバンク】

- 幅広いAI人材の育成に取り組んでおり、データ分析やAI利活用に関する様々なオンライントレーニングを提供したり、統計検定やアルゴリズム実技検定といった実践的な資格試験に対する資

⁹² 「AI戦略2022」においても、『AIについて先進的な米国などの他国に比べて、我が国ではAIの利活用を支える関連人材が不十分となっている。』『関連の人材の育成・確保は、緊急的課題であるとともに、初等中等教育、高等教育、リカレント教育、生涯教育を含めた長期的課題であり、AI戦略2019策定時から取り組んできている課題である。』とされている。

⁹³ 事業者によっては、『人材』を『人財』と表記しているところがあり、本報告書においても、各事業者の表記にならうこととする。

⁹⁴ 事業者等の社内における人材育成のほか、社会における受容性の向上も重要な観点であることから、利用者のリテラシーの向上等に関する取組も対象としている。

⁹⁵ この点について、前項（「医療・ヘルスケアに関するAI開発・利活用」）において述べている『デザイン思考を取り入れた医療とAI・データサイエンスをつなぐ人材育成の取組』と共通するものであり、注目すべきものと考えられる。前掲注（91）参照。

⁹⁶ 「報告書2021」においても、様々な具体的な取組事例が紹介されていることから、本報告書と合わせて参照すると、人材育成の参考になるものと考えられる。「報告書2021」第3章2.（3）、4.及び別紙3参照。

格取得支援制度を活用した報奨金を設定したりするなど様々なA I教育の機会を提供している。

- 社内におけるA I人材の確保について、統計などのコンピュータサイエンスの幅広い知識が必要であるため、主にそのような教育を行っている大学等から採用するとともに、中途採用によって確保している。また、入社後の実践的な業務の中で知識やスキルを身に付けていくようなO J T的な教育も広く行っている。
- 2022年5月に、A Iを使いこなし事業に活かすA I活用人材の育成を目標とした高等学校向けの教育プログラムの提供を開始した。A I活用リテラシーコースとA I活用実践コースの2つのコースで構成されている。学生は、学習用のデータを使ってA Iモデルを作成し、実際にウェブサービスや人型ロボット Pepper にA Iを組み込むといったことなどにより、学校や地域社会における身近な課題解決を目指して取り組むことになっている。

【パナソニック】

- 2016年度からA I人材を強化する取組を進めてきており、2021年にはA I技術を使いこなせるレベルの人材を独力で1,000人超を育成することができた。これにより、グループの各事業会社において、A I利活用を検討することができる体制が整ってきた。
- 大学と企業の両方に所属することを可能とする制度（クロスアポイントメント制度）を利用することによって、尖った人材も育ってきており、そのような者が著名な学会において連続で採択されたり、国際的なコンペにおいて優秀な成績を収めるなどA I技術を牽引している。
- 人材育成についても、組織に定着させるという考え方に基づいて、人材の質を高める方向に注力している。A Iを使いこなす人材の育成から、今後は、A Iエンジニアリング、品質保証の考え方に優れたA Iアーキテクト人材の育成が重点的な取組と考え、育成プログラムを強化している。

【シャープ】

- 2020年2月から、プログラミング教育に加えて、A Iリテラシー教育の取組を開始している。A Iリテラシー教育については、今後A Iが活用されていく社会で生きていくために、A I技術を開発することができる人材を増やしていきたいという課題の中で、ロボホンを活用して、身近に使われているA I技術を学び、早い時期からA I技術への興味・関心を喚起して、A Iについて考えるきっかけを持ってもらいたいというのが狙いである。
- 例えば、ロボホンとパンダを画像認識によって見分けるプログラムを作るという課題では、プログラムを作るだけではなく、実際にプログラムを使ってA Iに学習させる、判定させるというところまでを体験してもらうようにしており、単に課題を作るだけではなく、どのような仕組みになっているのかという体験を通じて学んでいくことができるカリキュラムを提供している。
- 実証授業において、子ども達は、実際にプログラムを作り、学習させて、判定させるという過程を通じて、試行錯誤を繰り返しながら、どのような方法が最も効率的に学習させることができるのかといったことを学ぶ、体験するという形になっている。

非常にクリエイティブで自分自身が考えて授業に取り組む様子が見られるとともに、実証授業後のコメントにおいて、授業前に比べて授業後の方がA Iに関する興味・関心が高まったという結果が得られた。

【三井住友海上】

- 社内にデジタル人財を育成する制度を設け、主にデータを分析する層、主にデジタルビジネスを行う層に区分して育成に取り組んでいる。

データ分析層については、3段階のカリキュラム（①提携先大学の情報連携学部における基礎的な知識の習得、②職場に戻ってからの実践的なデータ分析、③業務として分析プロジェクトの実施）により、専門人財を育成している。

デジタルビジネス層については、別の提携先大学の講座において、最新のデジタル技術などを用いた様々な新しいビジネスモデルの知見を習得したりするとともに、職場の上司の理解も必要であろうということで、上司も一緒になって実務として伴走していくといった研修プログラムを設けている。

【ダイキン工業】

- 2025年に向けた中期経営計画において、ダイキン情報技術大学という育成機関を中心にデジタル人材を1,500名育成し、その中でビジネスのイノベーションとプロセスのイノベーションを進めていこうというビジョンを掲げている。

- もともと技術の開発や活用を行うシステム人材が非常に少ない状況であったが、社内での育成では間に合わないということで、2017年12月に、ダイキン情報技術大学をスタートした。データサイエンティストに必要とされるテーマ実行力、分析力、データエンジニア力の3つの軸について、独り立ちできるレベルを目指そうと取り組んでいる。

- 最終的に目指す人材として、入社時から有している技術的な専門性に加えて、このような育成を通じてAIやIoTのことも分かるII（パイ）型の人材を育成していきたいと考えている。

- 新入社員向け、既存社員向け、基幹職層向け、幹部層向け、役員向けと階層別の教育・育成を行っている。

例えば、新入社員向けについては、1年目は座学のほかに課題解決型の演習でクラウドの構築といった教育を行っており、習熟度の確認として、基本情報処理技術者や統計検定の試験などを受けているが、合格率は全国平均を上回っている。2年目はPBL（Project Based Learning：問題解決型学習）を行い、全社向けの報告会なども行っている。この取組において、集合知の力が非常に大きいこと、2年間でキャリアプランをじっくり考える時間が持てることがメリットである。他方で、基礎教育と実践との間に壁があるため、それをどのように乗り越えるかということが課題となっている。

- これまでの4年間の取組で着実に前進していると考えており、この取組を知った上で入社を希望する学生が非常に増えており、また、離職率も平均と比べると非常に低く、仲間意識が生まれている。

【JR西日本】

- CBM（Condition Based Maintenance）という技術分野とマーケティングの分野に分かれているが、案件をプロデュースして課題設定を行うビジネス担当と分析担当であるデータサイエンティストがバディになって案件に取り組むところが特徴的である。

- 社内における人事、処遇については検討しているところである。鉄道業界は、データサイエンスの分野は未開の地が多く、今は当社の取組が「業界初」というプレミアが付くため、そのようなと

ころにやり甲斐を感じているのではないかと思う。ただし、それだけでは将来的には立ち行かなくなるので、給与や権限、チーム編成、仕事のやり方・スピード感などは変えていく必要があると考えている。

- データサイエンティストの知見を有する社員のキャリアパスは2つのパターンがあり、1つは高度なデータ分析を担う組織として、鉄道本体に技術をフィードバックしていくパターンである。他方で、現場の社員も将来的には一般教養的にデータ分析を行って活用できるようにしていく必要があるため、デジタルのチームに一定期間ジョインしてもらい、その間にデータ分析の素養を身に付けて、その後、事業部門に戻って現場の業務を改善していくというのがもう1つのパターンである。

【KDDI】

- ルールや原則を作っても、社員に認識してもらい浸透を図っていくことは難しいが、AIに携わるメンバーがコミュニティを作っているため、そこでAIの利活用事例についてライトニングトークで話し合ったり、倫理指針などガバナンスの重要性について浸透活動を行っている。また、全社員を対象としたeラーニングを行うとともに、社内広報を活用して浸透を図っているところである。
- AI人財の育成というより、DX人財の育成を行っており、その中の1つとして、個人情報保護やAI倫理の講座を設置するようにしている。プライバシーやAI倫理を守るという一定の意識付けを行った上で、事業で活躍するという形でバランスを取りながら取組を進めている。

(4) AIサプライチェーンに関する取組

AIが開発され、利活用されるまでには、データの取得や加工、AIモデルの作成、システムへの組込など様々なプロセスを経ることとなるが、このようなAIサプライチェーン⁹⁷におけるリスクや課題、その対応策等について整理した。

<グッドプラクティスのポイント>

- AIサプライチェーンにおいて、データの取得・加工、アルゴリズムやモデルの生成など様々な過程においてリスクがあることが明らかとなったが、AIの開発者やサービスプロバイダーにおいて、データ提供元のチェックや契約による業務保証といった取組が見られる。特に、データ流通の基盤の構築する取組は注目すべきものである。このような基盤が整備されることによって、データのソース・来歴等を可視化することができるため、信頼あるデータ流通の実現につながるものと期待される。

AIの開発者・利用者等は、本ヒアリングにおいて示された事例を参照しつつ、AIサプライチェーン上のリスクを洗い出すとともに、リスクを顕在化させない、顕在化した場合でも被害を最小限に抑えるといった取組を行うことが望ましい。

【NTTデータ】

- AIの構成要素であるデータやソフトウェアは無形資産であり、物品のように在庫されたり、輸送されるものではないという前提を踏まえると、「ほしい時に、ほしい役務・データ・ソフトウェア

⁹⁷ AIサプライチェーンについて明確な定義があるわけではないが、本報告書においては、データの取得や加工、AIモデルの作成、システムへの組込などの一連のプロセスをまとめてAIサプライチェーンと称することとする。

製品が、ほしい条件で利用することができるか」がAIサプライチェーンにおけるリスクの観点である。

- まず、公表された論文や革新的なアルゴリズムなどをベースにOSSのような形で学習・推論プログラムを作るとともに、現場や実社会のデータ（元データ）を加工して学習させるためのデータ（加工済みデータ）を作る。次に、学習・推論プログラムと加工済みデータをもとに学習済みモデルを作り、それを実際のソフトウェア、システムに組み込む。最終的に、そのソフトウェア、システムを運用してアウトプット出力を得るとというのが一般的なAIサプライチェーンであるものと考えられる。
- このようなプロセスの中で、知的財産・ライセンスに関するリスク、データのプライバシーに関するリスク、バイアスがあるデータを使っていないかといったデータの偏りなど倫理的なリスク、期待した効用が得られるかといった品質に関するリスク、データの漏洩や意図しない出力結果が出される（推論誘導）といったセキュリティに関するリスク、このようなことがAIサプライチェーンにおけるリスクの要素として考えられる。
- 知的財産・ライセンスについて、提供ポリシーや公開ポリシーが変更されることによって、使うことができなくなったり、また、途中でプライシングの考え方が変更されることによって、それまで無料で使うことができていたにもかかわらず、使用料が必要になるといったリスクがある。
- プライバシー、知的財産について、目的外利用のリスクがあり、様々なデータをAPIを通じて提供しているが、それが再利用されてしまうリスクがある。これに対しては、利用規約できちんと確認することが重要である。（再掲）
- 倫理について、データの取得が法的、倫理的に適切に行われているかというリスクがある。これに対しては、扱うデータのセンシティブ度に応じて、個別に調達先や調達仕様が問題ないか確認することが重要である。また、提供元のデータガバナンスの考え方や方法に関して、第三者的なお墨付きが得られるような仕組みがあると、安心してデータを利用することができるのではないかと考えている。
- さらに、データバイアスに関するリスクがあり、正解ラベルにバイアスがあると、AIが公平性に欠ける推論をしてしまうことがあり得る。これに対しては、男性か女性かといった人間の普遍属性に関わるパラメーターが、どの程度出力に影響を与えているのか確認することが重要である。（再掲）
- 品質について、外部調達したものの精度が求めているタスクに合っているかというリスクがある。これに対して、現状としては、タスクに応じて、POCにおいて精度を検証するしかないが、将来的には、AIの品質基準が定められて一定の目安ができることが望ましいと考えている。（再掲）
- セキュリティについて、外部から調達したモデルやプログラムがどのようなものをベースに作られたのか分からず、バックドア的なものが含まれているリスクがある。これに対しては、ソースコードを確認する、提供元の信頼性をもとに判断するといった対応にならざるを得ず、法務審査・レビューをきちんとやるというのが企業側として現実的な方法であるものと考えている。
また、バックドアによってデータが抜き取られるというリスクもある。これに対しては、支配的なコンポーネントやプロダクトに関して、公的なお墨付きのようなものがあると利用する側としては安心できるのではないかと考えている。（再掲）

【富士通】

- AI サプライチェーンの実態については、主要な取引の典型的なものに限定して議論できるのではないかと考えている。アルゴリズムの流通としては、アルゴリズムを当社が他社に提供するケースと当社が他社から提供を受けるケースである。データの流通としては、当社のユーザーや共同研究のパートナーがデータを用意するケースが大半を占めている。なお、当社がデータを購入して他社に転売するケースもあるが、それほど多いものではない。
- AI サプライチェーンにおける法令上の要請については、従来のサプライチェーンとほぼ同じであり、AI であるとはいえ、取引パターンが限定的であることを踏まえると、現在のところ、特段の法的対応の必要性は認識していない。また、契約上の処置については、データの主体が個人であれば、同意取得などの手続は既存の枠組みで行われるが、技術的な限界や倫理的な課題に関しては、ユーザーと共同で対応するというのが実態である。なお、当社がデータ収集者からデータを入手する場合には、契約上の表明保証や監査を行うことによって担保している。
- アルゴリズムを当社が他社に提供する事例（測定データの自動解析）について、データの観点では、ユーザー企業がデータを収集、アノテーションまでを行い、共同研究の契約のもとでデータを当社に共有してもらい、当社がAI モデルを開発するという流れになる。システム開発の観点では、ユーザー企業のデータを用いて当社がAI モデルを開発し、さらに当該モデルを組み込んだAI ソフトウェアを開発してユーザー企業にライセンス提供するという形となる。ユーザー企業は提供されたAI ソフトウェアを組み込んだソフトウェアを開発し、エンドユーザーに販売するという流れになる。
- 当社が他社のアルゴリズムの提供を受けて別のユーザーにサービスを提供する事例（翻訳サービス）について、データの観点では、製品開発段階でパートナー企業が自身の保有するデータを利用し（当社はアクセスできない）、エンドユーザーがファイルを当社のクラウドにアップロードして、翻訳されたデータをダウンロードする流れとなっており、アップロードされたデータや翻訳結果は削除する運用となっている。システム開発の観点では、当社はパートナー企業のサービスを組み込んで自社サービスを提供する形となっているため、品質については、パートナー企業が担保することになる。
- アルゴリズムを当社が構成して他者（地方公共団体）に提供する事例（保育所の入所マッチング）について、データの観点では、実証実験の中で、アルゴリズムの検証、有効性の検証のためにデータを共有しているが、機械学習型ではなく数理最適化型のアルゴリズムを使っているため、住民のデータが組み込まれないモデルとなっている。システム開発の観点では、実証実験の有効性を確認したアルゴリズムを研究部門から事業部門に移管し、AI 機能が組み込まれたソフトウェアを当社がエンドユーザーに販売して利用するという流れになる。
- AI サプライチェーン固有の課題に対しては、技術面での対応が重要であると考えている。精度劣化等に対しては、監視や是正といった運用者の役割が重要であり、また、この部分をサポートする技術が必要であり、例えば、運用時の性能劣化を担保する品質保証に関わる技術開発を行っている。このような技術がサプライチェーンの中で適用されていくことによって、AI ならではの課題を改善する1つのステップになるのではないかと考えている。

【AIDC】

- データ流通を考えた場合、データの偏り度合いや時間の経過により、データの価値が変化するも

のが多い。例えば、自然言語処理において、方言や高齢者層のデータを日常的に必要としているが、このようなデータの入手は難しいため、データ流通のマーケットにおいては、価値のあるデータとなる。

- データについては、情報システムで一般的に統制を効かせているが、AI・機械学習ソリューションを適用していくことになると、しっかり統制を図り、説明責任を果たせることが重要になってくるのではないかと考えている。
- このようなことに応えることができるデータ流通の取引基盤が求められており、データ取引の契約や基盤の構築の取組を進めてきた。2022年3月に、データ取引のための基盤（AIDC Data Cloud）を構築し、サービスを開始した。契約や課金モデルについて、様々なケースが想定されるので、動的に契約テンプレートを作成するといったところも連携して仕組みを実現している。ラベルデータが何をもとにしたデータなのか、開発データの来歴はどのようになっているのかということもデータ基盤で可視化できるようになっている。

第3章 AI開発ガイドライン及びAI利活用ガイドラインに関するレビュー

前章で述べたとおり、本推進会議において、AI開発ガイドラインやAI利活用ガイドライン（以下「両ガイドライン」ということがある。）を策定し、OECDやG7、G20などの国際的な議論に貢献してきた。その後、国内外の動向や国際的な議論の動向をフォローしつつ、「安心・安全で信頼性のあるAIの社会実装」を進めるために、AIの開発者やサービスプロバイダー、ビジネス利用者、消費者的利用者などのステークホルダ⁹⁸からヒアリングを行い、「報告書2020」（2020年7月）及び「報告書2021」（2021年8月）において、その取組等を取りまとめた。

AI開発ガイドラインを策定してから5年、AI利活用ガイドラインを策定してから3年が経過しており、この間、国内外において多くのAI倫理・ガバナンスに関する原則、指針、ガイドライン等（以下「指針等」ということがある。）が策定されている。また、AI利活用ガイドラインの策定後、本推進会議及び本推進会議議長によるヒアリング（以下「ヒアリング等」という。）⁹⁹において、両ガイドラインの射程を超えて、「安心・安全で信頼性のあるAIの社会実装」に関する優れた取組であると評価することができるものが示された¹⁰⁰。

本章においては、このような状況を踏まえて、AI開発ガイドライン及びAI利活用ガイドラインをレビューするとともに、見直しに関する論点を整理する¹⁰¹。

1. AI開発ガイドライン及びAI利活用ガイドラインに関するレビュー

（1）レビューの概要

前述のとおり、両ガイドラインを策定してから数年が経過しており、この間、海外において、また、国内の事業者等において、多くのAI倫理・ガバナンスに関する指針等が策定されていることから、これらの指針等との比較により、両ガイドラインの射程や内容等について検証することが重要である。

また、ヒアリング等を通じて、両ガイドラインの射程を超えて、「安心・安全で信頼性のあるAIの社会実装」に関する優れた取組が行われていることが明らかになったことから、ヒアリング等において示された取組との対応を踏まえて、両ガイドラインの射程や内容等について検証することが重要である。

さらに、国内外においてAI利活用・社会実装が進められているが、その中で、リスクが顕在化したり、インシデントが発生したりした事例があることから、このような事例を踏まえて、両ガイ

⁹⁸ 前掲注（2）参照。

⁹⁹ ここでは「報告書2021」の取りまとめに当たって行ったヒアリング等のことであり、本報告書の取りまとめに当たって行ったヒアリングとは異なることに留意が必要である。前者のヒアリング等の概要については、「報告書2021」第3章1. 参照。また、後者のヒアリングの概要については、本報告書前章1. 参照。

¹⁰⁰ 主に「報告書2021」第3章2.（1）参照。このほか、AIガバナンスに関する更に具体的な取組について、事業者からヒアリングを行い、事例を取りまとめて「AIガバナンスに関する取組事例」（2021年9月）を公表した。これらの取組の中にも、「安心・安全で信頼性のあるAIの社会実装」に関する優れた取組であると評価することができるものが示されている。

次に掲げるURLのウェブサイトに掲載。

<https://www.soumu.go.jp/main_content/000770820.pdf>（日本語版）

<https://www.soumu.go.jp/main_content/000785877.pdf>（英語版）

¹⁰¹ 「報告書2021」第3章3.（1）②において、「AI開発ガイドラインやAI利活用ガイドラインをレビューし、位置付けや射程、原則などに関し、必要に応じて、見直し等を検討することが重要である」ことが指摘されている。

ドラインの射程や内容等について検証することが重要である。

以上のことから、次に掲げる4つの観点から、両ガイドラインをレビューすることとする¹⁰²。

- ・ 海外における指針等との比較
- ・ ヒアリング等において示された取組事例との対応
- ・ 国内における指針等との比較
- ・ A I 利活用・社会実装に関する事例との対応

(2) 海外における指針等との比較からの検証

海外において策定された指針等を踏まえて、A I 開発ガイドライン及びA I 利活用ガイドラインの射程や内容等について検証する¹⁰³。

検証に当たっては、米国、カナダ、英国、イタリア、オランダ、スウェーデン、デンマーク、ドイツ、ノルウェー、フィンランド、フランス、インド、韓国、シンガポール、中国、オーストラリア及びEUの17箇国・地域並びに2つの国際機関（ユネスコ及び世界保健機関（WHO））を対象とした¹⁰⁴。また、これらの国・地域等における政府機関、業界団体等及びグローバル企業が公表している67個の指針等を調査した¹⁰⁵。

各指針等の調査を通じて、次に掲げる22の「尊重すべき価値」を確認することができた。このうち「1. 人間中心」から「16. アカウンタビリティ」については、A I ガイドライン比較表（「報告書2019」別紙2）における項目と同じものであり、「17. 堅牢性」から「22. その他（コスト、効果測定）」が新たに確認することができたものである¹⁰⁶。

- | | | |
|-------------------|-------------------|---------------|
| 1. 人間中心 | 9. 適正な学習（学習データの質） | 17. 堅牢性 |
| 2. 人間の尊厳 | 10. A I 間の連携 | 18. 責任 |
| 3. 多様性、包摂 | 11. 安全性 | 19. 追跡可能性 |
| 4. 持続可能な社会 | 12. セキュリティ | 20. モニタリング、監査 |
| 5. 国際協力 | 13. プライバシー | 21. ガバナンス |
| 6. 適正な利用 | 14. 公平性 | 22. その他 |
| 7. 教育、リテラシー | 15. 透明性、説明可能性 | （コスト、効果測定） |
| 8. 人間の判断の介在、制御可能性 | 16. アカウンタビリティ | |

¹⁰² 欧州委員会がA I 規制法案を公表するなど国内外の動向や国際的な議論の動向を踏まえた対応が重要であるものと考えられることから、このような点も考慮してレビューすることとする。

¹⁰³ 総務省「A I に係る原則・ガイドライン等の策定状況、規定事項等に関する調査研究」による。網羅的なものではなく、当該調査に基づいて整理したものである。

¹⁰⁴ 対象国の選定に当たっては、A I 準備指数、A I 出版物数、A I 特許出願数及びA I 研究者数を考慮することとし、これらのいずれかが10位以内の国を選定した。また、海外の動向や国際的な議論の動向を踏まえて、EUや2つの国際機関を加えた。

¹⁰⁵ 主としてA I 利活用ガイドラインを公表した2019年8月以降に公表されたものを対象に調査した。内訳は、政府機関（EUの欧州委員会及びA I ハイレベル専門家グループを含む。）・国際機関：40個、業界団体等：10個、グローバル企業：17個であり、指針等のリストについては、別紙2 16-23頁参照。

¹⁰⁶ 例えば、「堅牢性」と似た（同じような）概念として、A I 開発ガイドラインの「安全の原則」に関する解説において、A I システムの信頼性や頑健性に留意する旨が記載されている。また、「追跡可能性」と似た（同じような）概念として、A I 利活用ガイドラインの「透明性の原則」に関する解説において、入出力等のログを記録・保存することが期待される旨が記載されている。このように、既存のものと新しく確認されたものとの間で明確な区別が困難なものもあるが、指針等において記載されている文言や文脈等を踏まえて、整理した。

これらの「尊重すべき価値」が、どの程度の数の指針等に記載されているのかについて、表1のように整理した（67個の指針等のうち、各国の政府機関¹⁰⁷及び国際機関が公表している40個の指針等を対象として整理した。）。

(表1)

	指針等の数	人間中心	人間の尊厳	多様性 包摂	持続可能な社会	国際協力	適正な利用	リテラシー 教育	人間の判断の介入 制御可能性	(学習データの質) 適正な学習	AI間の連携	安全性	セキュリティ	プライバシー	公平性	説明可能性 透明性	アカウンタビリティ	堅牢性	責任	追跡可能性	モニタリング 監査	ガバナンス	その他	
米国	6	1	2			1	1		1	1			3	3	1	4	5	4	1	2	2	3		
カナダ	1							1								1								1
英国	6		2		1			1		2		1	1	2	5	5	5	1	2			2		1
イタリア	1									1				1		1								
オランダ	2		1				1		1	2				1	2	2	1		1	1				
スウェーデン	2		1	1			1	1					1	1	1	2								
デンマーク	2		2	2			1		1			1	1	1	2	2	1	1				1		
ドイツ	3	1	2	2	3				1	2		2	2	2	2	3	2	3	1	1	1			
ルウェー	2		2		1		2		1			1		1	2	1	1	1		1	1			
フィンランド	2		2	1			1		1			1		1	1	2	1		2	2	1			
フランス	1								1						1	1	1					1		
インド	1		1	1								1	1	1	1	1	1		1					
韓国	1		1	1		1						1		1	1	1	1	1						
シンガポール	2		1												2	2	1							
中国	3		3	2	2	2	2	2	3	1		3	2	2	3	2	2	2		2	3	3		
オーストラリア	1	1	1									1	1	1	1	1	1		1		1			
EU	2		2	1	2	1	2		1			2	1	2	2	1	2	2	2	1	1			
国際機関	2		2	2	2	1	2	1	1		1	2	1	2	1	2	1		1	1	1			
合計	40	3	25	13	11	6	13	6	12	9	1	19	14	20	31	35	25	12	13	11	16	3	2	

(注) 記載があるものはピンク色に着色。また、各国等の指針等の数に対し、半数を超える指針等に記載のある場合、赤字で表示。

また、表2のように、北米、欧州、アジア、大洋州の4つの地域別に整理したところ¹⁰⁸、「透明性、説明可能性」、「公平性」、「アカウンタビリティ」、「人間の尊厳」といったものは、いずれの地域においても重視されていることが明らかになった。

(表2)

	指針等の数	人間中心	人間の尊厳	多様性 包摂	持続可能な社会	国際協力	適正な利用	リテラシー 教育	人間の判断の介入 制御可能性	(学習データの質) 適正な学習	AI間の連携	安全性	セキュリティ	プライバシー	公平性	説明可能性 透明性	アカウンタビリティ	堅牢性	責任	追跡可能性	モニタリング 監査	ガバナンス	その他
北米	7	1	2			1	1	1	1	1		3	3	1	4	6	4	1	2	2	3		1
欧州	23	1	14	7	7	1	8	2	7	7		8	6	12	18	20	14	8	8	6	8		1
アジア	7		6	4	2	3	2	2	3	1		5	3	4	7	6	5	3	1	2	3	3	
大洋州	1	1	1									1	1	1	1	1	1		1		1		
合計	38	3	23	11	9	5	11	5	11	9	0	17	13	18	30	33	24	12	12	10	15	3	2

(注) 記載があるものはピンク色に着色。また、各国等の指針等の数に対し、半数を超える指針等に記載のある場合、赤字で表示。



北米	欧州	アジア
① 透明性、説明可能性 ② 公平性 ③ アカウンタビリティ ④ 安全性 ④ セキュリティ ④ モニタリング、監査	① 透明性、説明可能性 ② 公平性 ③ 人間の尊厳 ③ アカウンタビリティ ⑤ プライバシー	① 公平性 ② 人間の尊厳 ② 透明性、説明可能性 ④ 安全性 ④ アカウンタビリティ

¹⁰⁷ 前掲注(105)参照。

¹⁰⁸ 表1において対象としている40個の指針等から国際機関のもの(2個)を除いた38個の指針等を対象に整理した。

さらに、それぞれの「尊重すべき価値」について、両ガイドラインにおいて関係があるものと考えられる記載と調査した各種の指針等における記載について、次に掲げる例のように整理した。22項目すべての整理については、別紙2 5-15頁のとおりである。

このように整理したところ、両ガイドラインと海外における指針等において、概ね同じような概念で用いられているものが多いが、他方で、既存の「尊重すべき価値」として共通するものであっても、詳細については、両ガイドラインと海外における指針等に記載されている内容等に差異¹⁰⁹が見られるものがあることが明らかになった。

尊重すべき価値	ガイドラインにおいて関係があるものと考えられる記載（主なもの）	主な記載内容の例
1.人間中心	【AI開発ガイドライン】 ○ 基本理念 ・ 人間の尊厳と個人の自律が尊重される人間中心の社会を実現すること 【AI活用ガイドライン】 ○ 基本理念 ・ 人間の尊厳と個人の自律が尊重される人間中心の社会を実現すること	・ 国家安全保障や信頼できるパートナーシップを強化できるように、人間を中心としたAIの開発・利用に留意する。 ・ システムはそれを利用する人、あるいは、その決定によって影響を受ける人を中心に考えなければならない。 ・ AIシステムは人間を中心とした価値観を重視し、人権、多様性、個人の自律性を尊重すべきである。
2.人間の尊厳	【AI開発ガイドライン】 ○ 倫理の原則 ・ 人間の尊厳と個人の自律を尊重する 【AI活用ガイドライン】 ○ 尊厳・自律の原則 ・ 人間の尊厳と個人の自律を尊重する	・ AIシステムは、人間の尊厳、権利、自由を守る国際条約と矛盾しないように設計されなければならない。エシックス・バイ・デザインにより合理的に予測できる誤用を避けるための措置を講ずる責任がある。 ・ 国籍、性別、国や民族の出身、人種、宗教、言語にかかわらず、個人の平等な尊厳と自律性を尊重する必要がある。 ・ 人類共通の価値観に従い、人権を尊重するとともに、人間にとっての利益を追求し、国や地域の倫理や道徳を守ることに留意する。

以上のとおり、海外における指針等との比較から検証した結果、

- ・ 海外における指針等において、「堅牢性」、「責任」、「追跡可能性」、「モニタリング、監査」といった新しい「尊重すべき価値」が盛り込まれているものがあること
 - ・ 「透明性、説明可能性」、「公平性」、「アカウンタビリティ」、「人間の尊厳」といったものは、いずれの地域においても重視されていること
 - ・ 既存の「尊重すべき価値」として共通するものであっても、詳細については、両ガイドラインと海外における指針等に記載されている内容等に差異¹¹⁰が見られるものがあること
- などを確認することができた。

(3) 国内における指針等との比較からの検証

国内の事業者や団体等において策定された指針等を踏まえて、A I 開発ガイドライン及びA I 活用ガイドラインの射程や内容等について検証する¹¹¹。

検証に当たっては、国内の事業者や団体等が公表している 22 個の指針等を調査した¹¹²。

¹⁰⁹ 例えば、「16. アカウンタビリティ」について、両ガイドラインにおいては、主としていわゆる説明責任といった意味で用いられているが、海外の指針等においては、説明責任にとどまらず、事故が発生した場合に事後的な対応をしたり、救済したりすべきというように、より広い意味で用いられているものがある。

具体的な例として、世界保健機関（WHO）が公表している「ETHICS AND GOVERNANCE OF ARTIFICIAL INTELLIGENCE FOR HEALTH」において、A I 技術によって問題が発生した際には、アカウンタビリティを果たす必要がある旨が記載されている。次に掲げる URL のウェブサイトに掲載。

<<https://apps.who.int/iris/rest/bitstreams/1352854/retrieve>>

¹¹⁰ 前掲注（109）参照。

¹¹¹ 前掲注（103）参照。

¹¹² 指針等のリストについては、別紙2 36-37 頁参照。

調査した指針等を公表している事業者及び団体等は、次のとおりである。

- ・ 株式会社ABEJA
 - ・ 株式会社 エヌ・ティ・ティ・データ
 - ・ 沖電気工業株式会社
 - ・ クウジツ株式会社
 - ・ KDDI株式会社
 - ・ コニカミルタ株式会社
 - ・ 株式会社J.Score
 - ・ 株式会社スタジオム
 - ・ ソニーグループ株式会社
 - ・ 日本電気株式会社
 - ・ 日本ユニシス株式会社
 - ・ 株式会社野村総合研究所
 - ・ 株式会社日立製作所
 - ・ 富士通株式会社
 - ・ 富士フイルム株式会社
 - ・ 株式会社三菱総合研究所
 - ・ 三菱電機株式会社
 - ・ 株式会社リクルート（リクルートワークス研究所）
 - ・ 一般社団法人人工知能学会
 - ・ 一般社団法人ピープルアナリティクス&HRテクノロジー協会
 - ・ 一般社団法人日本経済団体連合会
 - ・ AIクラウドサービス検討会※
- ※ 当該検討会での検討等を総合的に踏まえて、総務省において公表。
(注) 社名は、公表当時のもの。

各指針等の調査を通じて、上記（2）と同じように、22の「尊重すべき価値」を確認することができ、このうち、「17. 堅牢性」から「22. コスト」¹¹³が新たに確認することができたものである。

これらの「尊重すべき価値」が、どの程度の数の指針等に記載されているのかについて、表3のように整理した。なお、事業者や団体等が属する業種・分野やAI利活用の目的・用途等により、どの「尊重すべき価値」を重視することが期待されるのかは異なることに留意が必要である。

(表3)

1. 人間中心	5	12. セキュリティ	19
2. 人間の尊厳	17	13. プライバシー	21
3. 多様性、包摂	5	14. 公平性	21
4. 持続可能な社会	9	15. 透明性、説明可能性	20
5. 国際協力	1	16. アカウンタビリティ	20
6. 適正な利用	17	17. 堅牢性	3
7. 教育、リテラシー	14	18. 責任	3
8. 人間の判断の介在、制御可能性	6	19. 追跡可能性	4
9. 適正な学習（学習データの質）	7	20. モニタリング、監査	3
10. AI間の連携	4	21. ガバナンス	2
11. 安全性	14	22. コスト	1

(注) 半数を超える指針等に記載のある場合、赤字で表示。

また、それぞれの「尊重すべき価値」について、両ガイドラインにおいて関係があるものと考えられる記載と調査した各種の指針等における記載について、次に掲げる例のように整理した。22項目すべての整理については、別紙2 26-35頁のとおりである。

このように整理したところ、両ガイドラインと国内における指針等において、概ね同じような概念で用いられているものが多いことが明らかになった。

¹¹³ 海外における指針等においては、「効果測定」があったため「22. その他（コスト、効果測定）」としているが、国内における指針等においては、「効果測定」は確認されなかったため「22. コスト」としている。

尊重すべき価値	ガイドラインにおいて関係があるものと考えられる箇所（主なもの）	主な記載内容の例
3.多様性 包摂	<p>【AI開発ガイドライン】</p> <ul style="list-style-type: none"> ○基本理念 <ul style="list-style-type: none"> ・人間がAIネットワークと共生することにより、その恵沢がすべての人によってあまねく享受され、人間の尊厳と個人の自律が尊重される人間中心の社会を実現すること ○利用者支援の原則 <ul style="list-style-type: none"> ➢ ユニバーサルデザインなど社会的弱者の利用を容易にするための取組に努める ○関係するステークホルダーに期待される役割 <ul style="list-style-type: none"> ・本ガイドラインの運用や見直しにおいて各国政府、国際機関、開発者、市民社会を含む利用者など多様なステークホルダー間の対話の促進に向けた環境整備に努めることが期待される ・AIをめぐる議論の多様性を確保しつつ、AIの便益の増進及びリスクの抑制について、認識の共有を図り、相互に協力するよう努めることが期待される <p>【AI利活用ガイドライン】</p> <ul style="list-style-type: none"> ○基本理念 <ul style="list-style-type: none"> ・人間がAIネットワークと共生することにより、その恵沢がすべての人によってあまねく享受され、人間の尊厳と個人の自律が尊重される人間中心の社会を実現すること ・AIの利活用において利用者の多様性を尊重し、多様な背景と価値観、考え方を持つ人々を包摂すること 	<ul style="list-style-type: none"> ・AIの活用において、不当な差別を起こさないよう、お客様およびステークホルダーの人権および多様性の尊重に努めます。 ・安全かつ堅牢で使いやすい製品・サービスの実現に取り組む中で、より多くの人がAIによる恩恵を享受できるよう、AIの利活用を積極的に進めていきます。特に、これまで培ってきた先進・独自の技術との融合や社外との共創の推進により、新たな価値をスピーディーに生み出していくことで、ヘルスケアをはじめとした幅広い領域で社会課題の解決をリードすることを目指します。 ・多様な人々の多様なライフスタイルの実現をサポートするAI技術は人類の未来のため、多様な人々の多様な夢を実現させるためにあるとの理念のもと、積極的な開発と展開を行う。
4.持続可能な社会	<p>【AI開発ガイドライン】 (記載なし)</p> <p>【AI利活用ガイドライン】</p> <ul style="list-style-type: none"> ○基本理念 <ul style="list-style-type: none"> ・個人、地域社会、各国、国際社会が抱える様々な課題の解決を図り、持続可能な社会を実現すること 	<ul style="list-style-type: none"> ・AIの力を、地球規模の問題解決への貢献や平和で持続可能な社会の発展のために活用するよう努めます。 ・社会の多様な課題を解決し、快適で強靱な持続可能社会の実現や、世界中の人々のQoL向上のためにAIの開発と利活用を計画します。 ・AI等の利活用によりさまざまな社会課題の解決を図り、持続可能な未来社会の実現を目指します。

以上のとおり、国内における指針等との比較から検証した結果、

- ・ 国内における指針等において、「堅牢性」、「責任」、「追跡可能性」、「モニタリング、監査」といった新しい「尊重すべき価値」が盛り込まれているものがあること
 - ・ 両ガイドラインと国内における指針等において、概ね同じような概念で用いられているものが多いこと
 - ・ 「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」、「セキュリティ」といったものは、ほぼすべての事業者・団体等において重視されていること
- などを確認することができた。

(4) ヒアリング等において示された取組事例との対応からの検証

主として「報告書 2021」及び「A I ガバナンスに関する取組事例」（2021 年 9 月）¹¹⁴に示されている国内の事業者の取組等を踏まえて¹¹⁵、A I 開発ガイドライン及びA I 利活用ガイドラインの射程や内容等について検証する。

直接的に両ガイドラインの目的、基本理念、原則等と関係が深いものと考えられる「多様性」、「セキュリティ」、「プライバシー」、「公平性」、「透明性・アカウンタビリティ」、「適正利用」及び「品質管理」に関し、両ガイドラインにおいて関係があるものと考えられる記載と事業者の取組事例等について、次に掲げる例のように整理した。7項目すべての整理については、別紙2 38-41 頁のとおりである。

このように整理したところ、多様性を確保すること、新しい技術の活用も含めて様々な取組により、セキュリティや公平性の確保、プライバシーの保護、説明性・アカウンタビリティを実現しよ

¹¹⁴ 前掲注 (100) 参照。

¹¹⁵ 一部「報告書 2020」に示されている取組等が含まれている。

うとすること、いわゆる「human in the loop (人間参加型)」¹¹⁶といわれる運用を行い、人間を介在させて適正な利用を行っていることが明らかになった。また、品質管理について、ライフサイクルを通じた評価プロセスの整備など優れた取組が行われていることが明らかになった。

項目・事項	ガイドラインにおいて関係があるものと考えられる記載（主なもの）	具体的な取組事例等
多様性	<p>【AI開発ガイドライン】</p> <ul style="list-style-type: none"> ○ 倫理の原則 <ul style="list-style-type: none"> ・ 人間の尊厳と個人の自律を尊重する <p>【AI利活用ガイドライン】</p> <ul style="list-style-type: none"> ○ 基本理念 <ul style="list-style-type: none"> ・ AIの利活用において利用者の多様性を尊重し、多様な背景と価値観、考え方を持つ人々を包摂すること ○ 尊厳・自律の原則 <ul style="list-style-type: none"> ・ 人間の尊厳と個人の自律を尊重する ○ 公平性の原則 <ul style="list-style-type: none"> ・ 個人及び集団が不当に差別されないよう配慮する 	<ul style="list-style-type: none"> ・ 人材の多様性を重視しており、インプットする人間が非常にダイバース（多面的）であることに重点を置いている。このダイバースには、男女も含むものであるが、異なる文化を背景に持っている学者を集めているところに特徴がある。 ・ 研究所には、コンピュータサイエンスの研究者だけでなく、心理学やフィールドワーク、認知科学に取り組んでいる社会科学系の研究者がいる。心理学や認知科学の視点を入れることが重要である。 ・ 安心・安全で信頼性のあるAIの研究開発、社会実装には、様々なバックグラウンドを持った者が必要であり、デザインや社会心理学をやっていた研究者なども含めてAI倫理専門チームを構成している。
セキュリティ	<p>【AI開発ガイドライン】</p> <ul style="list-style-type: none"> ○ セキュリティの原則 <ul style="list-style-type: none"> ・ AIシステムのセキュリティに留意する <p>【AI利活用ガイドライン】</p> <ul style="list-style-type: none"> ○ セキュリティの原則 <ul style="list-style-type: none"> ・ AIシステム又はAIサービスのセキュリティに留意する ○ 適正学習の原則 <ul style="list-style-type: none"> ➢ AIが不正確又は不適切なデータを学習することにより、AIのセキュリティに脆弱性が生じるリスクが存在することに留意することが期待される 	<ul style="list-style-type: none"> ・ AIセキュリティについては、AIの普及により、AIに対する攻撃（AIをだます、AIの情報を盗むなど）、AIを使った脅威（AIをだますなど）が出てきている。 ・ AIとセキュリティの関係を考えて場合、(a)Attack using AI（AIを利用した攻撃）、(b)Attack by AI（AI自身による攻撃）、(c)Attack to AI（AIへの攻撃）、(d)Measure using AI（AIを利用したセキュリティ対策）の4つの視点が存在する。(c) AIに対する攻撃だけでなく、AIに関連するそれ以外の論点も考慮しながら議論を進めることが重要と考えられる。

以上のとおり、ヒアリング等において示された取組事例との対応から検証した結果、

- ・ 多くの事業者等が、両ガイドラインの各原則等に対応する取組を行っていること
- ・ これらの取組に加えて、両ガイドラインの射程を超えた「安心・安全で信頼性のあるAIの社会実装」に関する優れた取組を行っていること

などを確認することができた。

（5）AI利活用・社会実装に関する事例との対応からの検証

国内外において、AI利活用・社会実装が進められているが、リスクが顕在化したり、インシデントが発生するなど課題が指摘されている事例が生じている。これらの事例を踏まえて、AI開発ガイドライン及びAI利活用ガイドラインの射程や内容等について検証する¹¹⁷。

検証に当たっては、日本、米国、英国、ドイツ、フランス、中国及びEUの7箇国・地域並びに3つの国際機関（ユネスコ、WHO及び世界経済フォーラム（WEF））を対象とし、これらの国・地域の政府機関等が公表している14個のドキュメント¹¹⁸を調査した。

また、AI利活用・社会実装の事例については、「生体認証」、「自動運転」、「画像識別」、「医療」、「犯罪予測」、「再犯予測」、「採用」、「入学選抜」、「成績・人事評価」、「広告」、「新型コロナウイルス感染症対応」に分類して整理した。

これらの分類ごとに、課題が指摘されている事例について、次に掲げる例のように整理した。すべての分類の整理については、別紙2 42-43頁のとおりである。

¹¹⁶ 前掲注（88）参照。

¹¹⁷ 前掲注（103）参照。

¹¹⁸ 政府機関等が発行する、いわゆる白書や報告書、レポート等のことを「ドキュメント」としている。ドキュメントのリストについては、別紙2 45頁参照。

AI利活用・社会実装に関する事例から抽出された課題等（概要）	
生体認証	<ul style="list-style-type: none"> ・ 米国において、黒人男性が白人警察官に首を押さえつけられて死亡した事件を巡り、大手IT事業者（数社）は警察に対して顔認証技術を提供しないことを決定した。また、人権に基づく国内法が整備されるまで顔認識技術を用いたサービスを提供しないことを表明した社もある。顔認識を含む生体認証技術に関する法制度等について検討する必要がある。 ・ マサチューセッツ工科大学の科学者が、大手IT企業（数社）の顔認識システムに人種やジェンダーのバイアスがあることを指摘しており、公平性を確保する必要がある。 ・ 中国のフェイスブック事業者のプライバシーポリシーにおいて、ユーザーの顔データが無償で提供され、その肖像権が永続的に当該事業者に帰属すること等が記載されており、プライバシーを確保する必要がある。 ・ 米国のAI事業者が3Dマスクを活用して、中国の大手IT事業者の顔認証システムをパスしており、セキュリティを確保する必要がある。 ・ プリントアウトした紙を帽子に張り付けることにより、Face ID（顔認証）システムを誤認させることができるため、堅牢性を確保する必要がある。
自動運転	<ul style="list-style-type: none"> ・ 自動運転車による死亡事故が発生しており、安全性を確保するとともに、AIシステムの開発者、利用者間で責任の所在を明らかにする必要がある。 ・ 道路標識にステッカーを張ることにより、標識を誤認識させることができるため、堅牢性を確保する必要がある。

また、課題が指摘されている事例に関し、上記（2）及び（3）で確認した「尊重すべき価値」のうち、どのような価値に関連するのかについて、表4のように整理した。

（表4）

	人間中心	人間の尊厳	多様性 包摂	持続可能な社会	国際協力	適正な利用	リテラシー 教育	人間の判断の介入 制御可能性	（学習データの質） 適正な学習	AI間の連携	安全性	セキュリティ	プライバシー	公平性	説明可能性	透明性	アカウンタビリティ	堅牢性	責任	追跡可能性	モニタリング 監査	ガバナンス	その他	
生体認証	★	★	★			★			★			★	★	★										
自動運転	★	★				★				★	★	★			★				★					
画像識別												★						★						
医療	★	★	★			★			★				★	★	★	★								
犯罪予測	★	★	★			★			★					★	★	★								
再犯予測	★	★	★			★			★				★	★	★	★								
採用	★	★	★			★			★				★	★	★	★								
入学選抜	★	★	★			★			★					★	★	★								
成績・人事評価	★	★	★			★			★				★	★	★	★								
広告	★	★	★			★						★	★	★										
新型コロナウイルス感染症対応	★	★	★			★			★			★	★	★										

（注）記載があるものはピンク色に着色。

以上のとおり、AI利活用・社会実装に関する事例との対応からの検証した結果、

- ・ リスクが顕在化したり、インシデントが発生した多くの事例において、どのような価値を重視すべきかという対象については、概ね両ガイドラインに掲げられている価値（原則）でカバーされており、それらの価値をどのように実現するのかといった実効性の問題であること
- ・ 現時点においては、両ガイドラインの射程を大きく超えてリスクが顕在化したり、インシデントが発生するおそれは大きくないものと考えられること

などを確認することができた。

なお、現行のガイドラインにおいて掲げられている価値（原則）の実効性を確保すること及び将来的なAI利活用・社会実装の動向を見据えつつ、リスクを抑制する（顕在化させない）方策等を検討することは重要である。

2. AI開発ガイドライン及びAI利活用ガイドラインの見直しに関する論点の整理

(1) AI開発ガイドライン及びAI利活用ガイドラインの見直しに関する論点の整理に向けた検討

上記1. のAI開発ガイドライン及びAI利活用ガイドラインに関するレビューにおいて、国内外における多くの指針等に、両ガイドラインに記載されていない「尊重すべき価値」が盛り込まれているものがあること、また、事業者の取組等の中には、両ガイドラインの射程を超えた「安心・安全で信頼性のあるAIの社会実装」に関する優れたものがあることなどを確認することができた。

さらに、本推進会議において、両ガイドラインの見直しに関して、例えば、次に掲げるような意見が示された¹¹⁹。

- ・ 尊重すべき価値として新しく確認された「堅牢性」、「責任」、「追跡可能性」、「モニタリング、監査」等は、運用フェーズにおけるガイドラインの観点であり、これらの概念を打ち出していくことは重要である。
- ・ リスクを最小限に抑え込むというリスクコントロールについて、どのようにアーキテクチャを作っていくのか、また、どのようにガイドラインに書き込むのかということが重要になってくる。
- ・ AIが相互依存関係を強めていくので、実態をつかみながらガイドラインを作っていく必要がある。
- ・ 現時点では、OECD、G20、ユネスコなど国際的に共有される枠組みも形成され、議論のフェーズも進展してきているため、AI開発ガイドラインについては、歴史的な役割を一定程度果たした文書として取り扱った上で、その題名、構成、内容等を見直すに相応しい時期に至ったのではないかと考える。国内において共有するものについて、どのような形式とし、どのような内容とするのかに関しては、よく考える必要がある。
- ・ 原則の数をできる限り最小限にして、どのようにして企業内のガバナンスに導入していくかということ議論の焦点として考えていくべきである。
- ・ AIが様々な場面で活用されるようになってきていることを考えると、AI開発ガイドラインやAI利活用ガイドラインのカバー範囲をもう少し検討すべきである。(中略) AIが従来の技術と異なる点や何が問題であるのかについて、限界的な部分に関してポイントを絞り、AIガイドラインとして何を決める必要があるのか、少数精鋭化を図らないと使い勝手が悪くなるのではないかと考えている。
- ・ ガイドラインの見直しを考えるに当たっては、EU等の動向をも踏まえつつ、現在の状況に応じたものとなるよう考えるべきである。
- ・ ユネスコにおいても、文化・多様性というのがキーワードになってきている。今後、文化・多様性、SDGsを念頭に置いた目指すべき社会というものを打ち出していく必要があるのではないかと。
- ・ サプライチェーンにおいて留意されることが期待される事項として、指針を示していただくための検討は有意義であると考えます。
- ・ ガイドラインについて、最終ユーザーが理解できるかという視点から、中高生だけではなく、大人向けも含め、分かりやすい解説書が必要ではないかと。

¹¹⁹ 本推進会議 AIガバナンス検討会 合同会議（2022年2月8日及び同年4月27日）における意見。

加えて、「報告書 2021」において、AI を利活用した新型コロナ対策に関する国際比較を踏まえて、公衆衛生の向上・増進とプライバシー保護等とのバランスをどのように取ることが適切かといった観点から、両ガイドラインを見直すことも必要ではないかと考えられる旨が指摘されている¹²⁰。

これらのことを踏まえ、両ガイドラインをAI の開発者や利用者にとって、より有用・有益であるものにして、「安心・安全で信頼性のあるAI の社会実装」を進めるために、本推進会議として、両ガイドラインの見直しに関する検討を行うこととし、見直しに関する論点を整理する¹²¹。

見直しに関する検討に当たっては、ガイドラインにおける各原則の実効性をいかに確保するかという観点が重要である。分野共通・分野横断的なガバナンス、分野別・分野ごとのガバナンス、標準化によるルール形成など様々な取組を組み合わせ、全体として、ガイドラインにおける各原則の実効性を確保することにより¹²²、「安心・安全で信頼性のあるAI の社会実装」の推進につながる。このことを踏まえて、人間中心の社会を実現するために、「人間中心のAI 社会原則」¹²³を具体化するものとして、どのようなガイドラインとすることが望ましいかという点から、両ガイドラインの見直しに関する検討を行うこととする。

(2) AI 開発ガイドライン及びAI 利活用ガイドラインの見直しに関する論点の整理

両ガイドラインに関するレビュー、前述の両ガイドラインの見直しに関する本推進会議における意見、さらに、国内外の動向や国際的な議論の動向等を踏まえ、両ガイドラインの見直しに関する検討に当たって、次に掲げるように、4つの論点を整理した。

- 目的、基本理念、対象とするAI の範囲（定義）、原則、解説等について、追加や修正等をするかどうか。また、先進的な事例等を踏まえて、新しい内容を盛り込むなど構成等を見直すかどうか。

例えば、

- AI 開発ガイドラインの基本理念に、『多様性』や『持続可能』といった概念を追加するかどうか。
- 新しい原則として、『堅牢性』、『責任』、『追跡可能性』、『モニタリング、監査』といった項目を追加するかどうか。なお、これまで両ガイドラインを策定する過程において、原則数は可能な限り少なくすべきといった議論との整合をどのように考えるか。
- 既に定められている各原則の解説において、具体的な開発・利活用の場面やユースケース

¹²⁰ 「報告書 2021」第2章3. (2) 参照。また、この点に関し、「報告書 2020」において、原則の一部を緩和してでも深刻な打撃を避けるための緊急対応を行うために各原則について優先度や重要性を検討することが重要である旨が指摘されている（「報告書 2020」注 (117) 参照。）。

¹²¹ 両ガイドラインそれぞれの基本理念において、『ガイドラインを不断に見直し、必要に応じて柔軟に改定する』ことがうたわれている。

¹²² 例えば、AI 開発ガイドラインの目的において、「本ガイドラインでは、AI の開発に関し留意することが期待される事項のうち、利活用の分野に共通する事項又は分野間の連携に関し留意することが期待される事項について定めることとし、分野ごとの事情に応じて留意することが期待される事項については、本ガイドラインとは別に、各々の分野ごとに当該分野に係るガイドラインの在り方に関し、策定の要否を含め、各分野の国際機関を含む関係するステークホルダによる議論が行われることが期待される。」とされている。

¹²³ 前掲注 (20) 参照。

等を示すかどうか。

- 有事（パンデミック、災害等）を想定した見直しを行うかどうか（例えば、公衆衛生の向上・増進とプライバシー保護のバランスをどのように考えるか）。
- このほか、品質確保・マネジメント、サプライチェーン、組織・体制、人材育成、データの取扱い、行政機関のA I利活用におけるコストとのバランスなどの観点から、留意することが期待される事項を整理し、ガイドラインとして取りまとめるかどうか。 等

○ 各ガイドラインの位置付け、名称等を変更するかどうか。

例えば、

- 国際的な議論のためのもの、国内の事業者等の取組を支援するもの（指針等の策定の際の参考）というように国内外を問わずガイドラインの活用を図っていくことを明確化するかどうか。その場合、A I 開発ガイドライン（『国際的な議論のためのA I 開発ガイドライン案』）の名称を変更するかどうか。 等

○ A I の開発のハードルが低くなってきていることや利活用がより身近になってきている（利用者がA I が実装されていることを認識しない（できない）ケースが増えてきている）こと、あるいは、A I の開発と利活用の境界が曖昧になってきていることなどを踏まえて、どのような取組を進めることが望ましいか。

例えば、

- これまで想定していたよりも専門的でない開発者やA I の利活用を認識していない利用者（特に消費者的利用者）など向けに対応することが望ましいことはあるか（平易な解説を加えることなど）。
- A I 開発ガイドラインとA I 利活用ガイドラインを統合するかどうか。 等

○ 今後、ガイドラインの実効性を確保するために、ガイドラインを含めたガバナンス全体の枠組みにおいて、どのような取組を進めることが望ましいか。

例えば、

- 各原則等の実効性を確保するための方策として、事業者、業界等による取組や政府（行政機関）を含めた取組、あるいは、チェックシートの策定、認定制度などが考えられるところ、このような方策を含めたガバナンス全体の枠組みについて、どのような取組を進めることが望ましいか¹²⁴。
- EUのA I 規制法案が公表され、審議が進められていることなどを踏まえて、国際的なガ

¹²⁴ この点を考えるに当たっては、例えば、A I を用いたクラウドサービスの安全・信頼性に係る情報開示指針（ASP・SaaS編）及びASP・SaaS（A Iクラウドサービス）の安全・信頼性に係る情報開示認定制度が参考になるのではないかと考えられる。

前者の情報開示指針については、前章1.（4）参照。後者の情報開示認定制度については、（一社）日本クラウド産業協会（ASPIC（旧：ASP・SaaS・AI・IoTクラウド産業協会））は、2022年4月1日に、安心・安全なA Iクラウドサービスであることを確認できる情報を開示していることを認定するため、ASP・SaaS（A Iクラウドサービス）の安全・信頼性に係る情報開示認定制度を新設した。このような取組が参考になるのではないかと考えられる。

バランスの枠組みとの関係において、どのような対応が求められるか。 等

本推進会議において、これらの論点に関して、上記（１）で示した意見のほか、例えば、次に掲げるような意見が示された¹²⁵。

- ・ モニタリング・監査について、人間が関与することが時間的に間に合わない場合も想定される。いわゆる「human in the loop（人間参加型）」において、どのように人間が介入することが望ましいのかなどの視点を盛り込んで日本の独自性を出せるとよい。¹²⁶
- ・ AIの動作の結果が出てから対応するのか、動いている途中で停止させるなどの対応をするのかによって、モニタリング・監査の概念が異なることを考慮すべきである。また、オンラインにおける監査では、人間の介入は間に合わないので、AI技術などを用いて行うこととなるが、他方、オフラインにおける監査では、人間が時間をかけて行うことができる。この差異を明確にした方が読み手にとって分かりやすいものになると思う。
- ・ EUのAI規制法案においても、ハイリスクAIのオペレーターについて、人間の介入に関して詳しく記載されている。どの程度、どのような方法による介入が望ましいのか結論は出ていないが、方向性を示すことは意義があるのではないかと思う¹²⁷。また、ヒューマン・マシン・インターフェースの観点も盛り込むとよいのではないか。
- ・ AIのネットワークの観点は、より強調すべき時代になってきたと思う。
- ・ AIが使われる社会において、トラストの捉え方について、何らかの方向性があった方がAIを開発する側にとっても指針として役立つものになると思う。
- ・ セキュリティの観点から、ガイドラインにおいて、ブロックチェーンをどのように位置付けることが適切であるかを明確にしていくことも重要ではないかと思う。
- ・ 現状においては、自動運転のように、AIは相互に通信し合いながら、つながって動いている。セキュリティを考えると、一部問題のあるAIがあった時に、他のAIに波及しないようにすることが重要であり、システムに対するレジリエンシーということが、今後AIが使われる社会において必要になるのではないか。
- ・ 人間とAIシステム間のインターフェースについて、実用化が進むほど非常に重要になってくるため、ガイドラインに何か記載してもよい時期ではないかと思う。
- ・ 人間とAIのインターフェースについて、AIの開発や利活用の進展を踏まえて考えることが重要である。また、AIの開発と利活用をきれいに分けることができない点に関し、どのように考えていくことが望ましいのか検討することが重要である。

¹²⁵ 本推進会議 AI ガバナンス検討会 合同会議（2022年2月8日、同年4月27日及び同年6月22日）における意見。

¹²⁶ AI利活用ガイドラインの「適正利用の原則」において、人間の判断の介在に関する事項は記載されており、この点をどのようにブラッシュアップするかという検討が重要であるものと考えられる。

¹²⁷ 人間の判断の介在の要否については、「報告書2019」別紙1（附属資料）「AI利活用原則の各論点に対する詳説」において、5つの基準として考えられる観点（AIの判断に影響を受ける最終利用者等の権利・利益の性質及び最終利用者等の意向、AIの判断の信頼性の程度（人間による判断の信頼性との優劣）、人間の判断に必要な時間的猶予、判断を行う利用者に期待される能力、判断対象の要保護性（例えば、人間による個別申請への対応か、AIによる大量申請への対応か等））が例示されており、今後、これらを踏まえて、検討することが重要であるものと考えられる。

- ・ AIシステムが組み込まれた製品について、セットアップが複雑であったりするため、エンドユーザーが使いこなせない場合がある。エンドユーザーがきちんと使えるようにしたり、途中で誤った使い方をしないようにという視点からの検討が必要である。¹²⁸
- ・ AIが実用化の局面で使われることが増えてきている。ビジネスモデルのことは、既存のガイドラインには書かれていないが、ビジネスモデル化においても、AIが使えるポイントが何かあるのではないか。
- ・ ガイドラインを見直した場合、特にEUのAI規制法案に対するスタンスを示すことが重要である。ガイドライン本体に盛り込むかどうかという問題はあるが、本推進会議として、海外の動向をどのように見ているのか、どのように解釈しているのかを示すことが重要である。

今後、このような論点や本推進会議における意見等とともに、次に掲げる観点¹²⁹を踏まえて、両ガイドラインの見直しに関する検討を進めていくこととする。

- ・ 現行の両ガイドラインの在り方を前提として、目的、基本理念、原則等を追加したり、変更したりするかどうか。
- ・ 現行の両ガイドラインの在り方を前提とせず、ガイドラインの在り方自体を検討の対象としたり、ガイドライン単体ではなく各原則の実効性を確保するための方策と一体的なものとして考えるかどうか。

見直しに関する検討に当たっては、技術的中立性を確保するとともに、イノベーションを阻害せず、開発者や利用者にとって過度に負担とならないものにするといった点について、引き続き、留意することが重要である。なお、ガイドラインの在り方自体を検討する場合においても、前述の通り、両ガイドラインは、政府全体として策定されている「人間中心のAI社会原則」¹³⁰を具体化するものであることを念頭に置いて検討を行う必要があることに留意が必要である¹³¹。

また、両ガイドラインが、AIネットワーク化の健全な進展を通じて、AIの便益の増進とリスクの抑制を図るといったことなどを目的とするものとされており、さらに、AI開発ガイドラインの基本理念において『イノベティブでオープンな研究開発と公正な競争を通じ、AIネットワークの便益を増進するとともに、学問の自由や表現の自由といった民主主義社会の価値を最大限尊重しつつ、AIネットワークにより権利利益が侵害されるリスクを抑制するため、便益とリスクの適正なバランスを確保すること。』が、加えて、AI利活用ガイドラインの基本理念において『AIの利活用において利用者の多様性を尊重し、多様な背景と価値観、考え方を持つ人々を包摂すること』がうたわれていることから、ガイドラインの見直しの検討に当たっては、これらのことに留意が必要である。

¹²⁸ AI利活用ガイドラインの「適正利用の原則」において、AIサービスプロバイダーからの必要な情報の提供が期待される旨が記載されており、この点をどのようにブラッシュアップするかという検討が重要であるものと考えられる。

¹²⁹ 掲げられている2つの観点は、必ずしも相反するものではなく、ガイドラインの在り方自体を見直しつつ、現行のガイドラインに定められている目的、基本理念、原則等を追加したり、変更したりすることはあり得る。

¹³⁰ 前掲注(20)参照。

¹³¹ この点に関し、両ガイドラインを見直すとした場合、政府全体として策定されている「人間中心のAI社会原則」のもとに本推進会議(総務省)が策定するガイドラインがあるという構造を明確にした上で、特徴的なことに限定した内容にした方がよいという意見があった。

第4章 今後の取組

第1章において、国内外の動向及び国際的な議論の動向等を概観し、第2章¹³²において、本ヒアリングを踏まえて、事業者等の取組を紹介するとともに、「安心・安全で信頼性のあるAIの社会実装」の推進に向けた留意点（グッドプラクティスのポイント）を整理した。また、前章において、AI開発ガイドライン及びAI利活用ガイドラインのレビューを行い、その結果等を踏まえて、今後、両ガイドラインの見直しに関する検討を行うこととし、見直しに関する論点を整理した。

本章においては、ここまでの検討等を踏まえて、今後の取組を整理する。

1. AI倫理・ガバナンス

(1) AI開発ガイドライン及びAI利活用ガイドラインの見直しの検討

前述のとおり、前章において、AI開発ガイドライン及びAI利活用ガイドラインのレビューを行い、その結果等を踏まえて、今後、両ガイドラインの見直しに関する検討を行うこととし、見直しに関する論点を整理した。また、前章において取りまとめたとおり、本ヒアリングにおいて示された取組について、両ガイドラインの各原則に対応する優れた取組であると評価することができるものに加え、両ガイドラインの射程を超えた「安心・安全で信頼性のあるAIの社会実装」に関する優れた取組であると注目すべきものがある。これらのことなどを踏まえて、前章において整理した論点を中心に、両ガイドラインの見直しに関する検討を行う。

見直しに関する検討に当たっては、ガイドラインの位置付けや在り方、対象とする範囲等について、検討すべきではないかといった意見があったことを踏まえて、既存の枠組みにとらわれず、検討を行うことが重要である¹³³。なお、前章において述べた両ガイドラインと海外における指針等に記載されている内容等との差異も考慮しつつ、見直しに関する検討を行うこととする。

また、技術的中立性を確保するとともに、イノベーションを阻害せず、開発者や利用者にとって過度に負担とならないものにするといった点について留意が必要である。さらに、それぞれのガイドラインの基本理念にうたわれていることを踏まえて、「競争的なエコシステムの確保」や「利用者の利益の保護」といった観点にも留意が必要である¹³⁴。

加えて、開発者、利用者、専門家、関係団体など幅広いステークホルダと意見交換を行いつつ、見直しに関する検討を進めていくこととする。

なお、ガイドラインを見直すこととした場合、現在の両ガイドラインの対象者¹³⁵や法的な性格¹³⁶等を改めて明確にした上で、海外への情報発信、海外からの見え方等を考慮して、見直し後のガイドラインについても、それらの異同を明確にするるとともに、名称についても適切なものとなるよう検討することが重要であるものと考えられる。

¹³² 別冊（「安心・安全で信頼性のあるAIの社会実装」の取組事例～グッドプラクティス集～）を含む。

¹³³ この点について、実積寿也「AIガイドライン、トラスト、ガバナンスー最適なルール策定に向けてー」Nextcom Vol. 50（株式会社KDDI総合研究所、2022年）が参考となる。

¹³⁴ 前章2.（2）参照。

¹³⁵ 民間事業者だけではなく、個人や行政機関、地方公共団体等も対象者として含まれている。

¹³⁶ 両ガイドラインの基本理念において、拘束的ではない（非拘束的な）ソフトローとして指針やベストプラクティスを国際的に共有する旨がうたわれている。

(2) AI開発ガイドライン及びAI利活用ガイドラインの周知・共有

「報告書2020」において、AI原則等の策定そのものは直接的に企業の収益に結びつくものではないものの、AIの開発等に対する企業としての基本的な方針を示すものとなっており、企業のステークホルダにとっては、AIの開発等において懸念される点についての不安を払拭し、AIの開発等の取組に対する信頼を醸成することにつながる旨が指摘されている¹³⁷。他方、前章において、国内における指針等の策定状況を示しているが、一部の事業者・団体等が策定・公表しているのみであり、指針等の策定の取組の普及を図っていくことが重要である。本推進会議と総務省は、引き続き、AI開発ガイドライン及びAI利活用ガイドラインについて、周知・共有の活動を推進することとする¹³⁸。

(3) 国内外の動向・国際的な議論の動向のフォローアップ及び情報発信

第1章において概観したとおり、AI倫理・ガバナンスに関して、国内外で様々な取組が行われるとともに、国際的な議論が活発に行われている。AI倫理・ガバナンスに関する検討を行うに当たっては、国内外の動向・国際的な議論の動向を踏まえることが重要であり、引き続き、本推進会議において、これらの動向をフォローアップすることとする。

また、動向をフォローアップするだけでなく、国際的な議論の場において、日本から情報発信を行い、国際的な議論を主導し、貢献していくことも重要である。本報告書において取りまとめた各事業者等の取組事例について、OECDやGPAI等のマルチの場のほか二国間の政策対話なども含めて、国際的な議論の場において、B2B (Business to Business) の取引が多いといった日本の産業構造なども考慮しつつ、情報発信を行っていくことが重要である。

特に、本年(2022年)末にOECDにおいてデジタル関係の閣僚級会合が開催される予定であること、本年(2022年)秋から来年(2023年)秋にかけて日本がGPAIの議長国を務めることになっていること、来年(2023年)に日本でG7の会合が開催される予定であることなどを踏まえると、国際的な議論において、日本が果たす役割は大きいものと考えられる。

さらに、EUのAI規制法案において、EU域内にAIシステムを提供する又はAIシステムの成果物を提供する第三国のプロバイダー及び利用者もハイリスクAIに関する規制の対象になるとされている。したがって、同法案が採択され、適用が開始されると、日本を含めて世界中に大きな影響を及ぼすものと考えられる。本推進会議として、引き続き、国際的な議論の動向、各国の受け止めや反応等を踏まえつつ、同法案に関する研究を進めることが重要である。

(4) 原則の実効性を確保するための取組

AIに関する各国の取組や国際的な議論については、2019年5月にOECDにおいてAI原則を含む理事会勧告が採択されるとともに、同年6月にG20においてG20 AI原則が採択されたことなどを踏まえて、これらの原則をいかに遵守し、信頼性あるAIの社会実装を進めるかという点にシフトしつつある。

本ヒアリングにおいて、多くの事業者等がAI開発ガイドラインやAI利活用ガイドラインの各

¹³⁷ 「報告書2020」第3章9. 参照。

¹³⁸ 両ガイドラインについては、今後、見直しに関する検討を進めていくこととしているが、現行のものであっても、事業者等が指針等を策定するに当たっては、参考になるものと考えられる。また、ガイドラインを見直した場合、当然に見直し後のガイドラインの周知・共有を図っていく考えである。

原則に対応する取組を行っていることが示されたが¹³⁹、前章において述べたように、ガイドラインにおける各原則の実効性をいかに確保するかという観点については、事業者等自身による取組に加えて、分野共通・分野横断的なガバナンス、分野別・分野ごとのガバナンス、標準化によるルール形成など様々な取組を組み合わせて、全体として、ガイドラインにおける各原則の実効性を確保することが重要であり、これにより、AIに対する信頼性が高まるとともに、社会の受容性の向上が図られることにつながっていく。また、これらの取組については、業界全体における取組や外部機関との連携、政府（行政機関）を含めた取組なども考えられるところ、それらの取組が適切に実施されることが望ましい。これに関連して、本ヒアリングにおいて、事業者等自身による取組のほか、リスクを洗い出すフレームワークの構築や外部機関によるモニタリングの仕組みの整備などガイドラインにおける各原則の実効性を確保しようとする取組が行われていることは注目すべきである¹⁴⁰。このほか、チェックシートの策定や認定制度といった取組¹⁴¹も考えられるところ、これらのことを踏まえて、ガイドラインの見直しに関する検討と併せて¹⁴²、原則の実効性を確保するための取組に関する検討を行うことが重要である。

この検討に当たっては、それら原則の実効性を確保するための取組を推進・普及させていくかといった仕組みの構築や環境整備を図っていくことが重要であることも念頭に置きつつ、総合的に検討すべきものと考えられる。

2. 取組事例の周知・共有

(1) AI倫理・ガバナンスに関する取組

本ヒアリングを通じて、多くの事業者等がAI倫理・ガバナンスに関する原則・指針・ガイドライン等を策定したり、組織・体制を整備したりするとともに、品質保証のための仕組みの整備や外部のステークホルダ等との連携・協働等の取組を行っていることが示された。これらの取組は、セキュリティやプライバシー、公平性、透明性・アカウントビリティなどAI開発ガイドラインやAI利活用ガイドラインの各原則に対応する優れた取組であると評価することができる。

また、注目すべきものとして、リスクを洗い出すフレームワークの構築や外部機関によるモニタリングの仕組みの整備などガイドラインにおける各原則の実効性を確保しようとする取組が行われていること、あるいは、サプライチェーン上のリスクを洗い出し、そのリスクを顕在化させないよう措置を講ずるなど自社のみならず取引先等を含めて各原則を遵守しようとする取組が行われていることも示された。これらの取組は、「安心・安全で信頼性のあるAIの社会実装」を一層推進させる取組であると評価することができる。

このような取組を行っているAIの開発者・利用者等は、社会的責任を果たしているものと評価することができるのではないかと考えられ、社会からの評価につながることを期待される。さらに、このような取組が社会全体に普及・浸透することにより、AIに対する社会の受容性の向上が図られることが期待される¹⁴³。

¹³⁹ 「報告書 2021」においても、多くの事業者による各原則に対応する様々な取組が示されている（「報告書 2021」第3章2.（1）、4. 及び別紙3参照。）。

¹⁴⁰ 第2章2.（1）及び別冊（グッドプラクティス集）参照。

¹⁴¹ 前掲注（124）参照。このような業界等による取組が参考になるものと考えられる。

¹⁴² これらの点に関しては、前章における両ガイドラインの見直しに関する論点でも言及されている。

¹⁴³ このような取組の社会全体への普及・浸透を促進する方策として、AI倫理・ガバナンスに真摯に取り組んでいる事業者に対する税制上の優遇措置や政府調達における優遇措置などがあるのではないかと

本ヒアリングにおいて示された取組が、開発・利活用するAIやサービスの種類、事業者の規模や業種等により、どの事業者にとっても有用・有益であるとは限らない点には留置が必要であるが、多くの事業者等が共通的に行っている取組については、今後取組を進めようとするAIの開発者・利用者等にとって、参考となる標準的な内容であるものと考えられる。また、リスクを洗い出すフレームワークの構築や外部機関によるモニタリングの仕組みの整備、サプライチェーンに関する取組などについては、既に一定の取組を行っている開発者・利用者等にとっても、有用・有益であるものと考えられる。さらに、AIの開発者・利用者等が、AI倫理・ガバナンスに関する取組を行っていることを一般の国民や利用者へ周知することにより、AIに対する社会の受容性の向上が図られることが期待される。

これらのことから、本報告書において取りまとめた各事業者等の取組事例について、周知・共有を図っていくことが重要であり、本推進会議と総務省は、外部のステークホルダーと連携して¹⁴⁴、取組事例の周知・共有の活動を推進することとする。

取組事例の周知・共有の活動の推進に当たって、AI倫理・ガバナンスについては、マルチステークホルダーによる取組を進めることによって、AIに対する信頼を高め、社会の受容性を向上させていくことが重要であり、特に、利用者、利用者団体等への展開を図っていくとともに、利用者や専門家等の意見を取り入れることが重要であるという観点から¹⁴⁵、利用者、利用者団体及び専門家を含む幅広いステークホルダーとの意見交換を行っていくこととする。

また、「AI戦略2022」において、『人々の生活様式や働き方が変化している中、これらに対応しつつ、更に国の行政機関における業務の効率化や質の高い行政サービスの提供を行っていくためには、これまで以上に積極的なAIの活用を検討していく必要がある。国の行政機関がAIを活用する際には、特に透明性、公平性、説明可能性等の確保が重要であることを理解したうえで、AIの導入促進を図ることが必要である。このため、国の行政機関におけるAI導入の基本的考え方の整理や、AI導入ガイドラインの策定等の総合的な対策を取りまとめ、実施していく。そして、国の行政機関における業務に積極的にAIを活用していく。』とされていることを踏まえると、今後、行政機関におけるAIの利活用が進むことが想定され、関係する府省庁との意見交換を行っていくこととする。

(2) AI開発・利活用に関する取組

本ヒアリングを通じて、多くの事業者等が、新型コロナ対策、医療・ヘルスケア、高齢者・障害者支援をはじめ様々な分野において、AIの開発・利活用の取組を行っていることが示された。これは、AIが幅広く多くの分野において利活用されることにおける有用性を示していると評価でき

意見があった。ただし、これらの措置については、慎重な検討・調整を要するものであることに留意が必要である。

¹⁴⁴ 例えば、総務省は、中央大学ELSIセンターと連携し、同センターが2022年6月に行ったコミュニティ・アクティビティにおいて、本推進会議の取組等の周知・共有を行った。

¹⁴⁵ 例えば、AI利活用ガイドラインの「透明性の原則」の「説明可能性の確保」に関し、『利用者の納得感や安心感の獲得、また、そのためのAIの動作に対する証拠の提示等を目的として、AIの判断結果の説明可能性を確保することが期待される。』とされている。また、同ガイドライン附属資料「AI利活用原則の各論点に対する詳説」において、『消費者的利用者等のニーズ、意見等も踏まえつつ、説明が不足している部分を明確にし、どのような説明が必要か、開発者とも連携して解決策を模索する。』とされている。

るものである。また、これらの事例を参考に、新しい技術の開発、新しいサービスの提供や既存サービスの利便性向上が図られることなどが期待される。

これらのことから、本報告書において取りまとめた各事業者等の取組事例について、周知・共有を図っていくことが重要であり、本推進会議と総務省は、外部のステークホルダーと連携して、取組事例の周知・共有の活動を推進することとする。

また、課題先進国として、日本が世界に対して、AIの開発・利活用が社会的課題の解決に貢献することを示すことは重要である。本報告書において取りまとめた各事業者等の取組事例について、国際的な議論の場において、情報発信を行っていくことが重要である。

3. 人材育成

本ヒアリングを通じて、多くの業者等が、様々な工夫をして人材育成に関する取組を行っていることが示された。

しかしながら、AI人材については、人材不足が世界共通の課題であり、日本においても、人材育成は急務の課題である。本報告書において取りまとめた各事業者の取組事例を参考にしつつ、引き続き、AIの開発者・利用者等において、人材育成に関する取組を推進することが重要である。AI人材については、技術面に優れた人材を育成していくことが期待されるほか、AIリテラシーの向上のための取組も行われており¹⁴⁶、この点について、AIの開発や利活用をめぐる倫理的・法的・社会的課題（ELSI）の抽出及び対応に関し貢献することができる人材を育成していくことも期待される¹⁴⁷。また、エンドユーザーがAIシステムが組み込まれた製品・サービスを使いこなすことができるようにAIリテラシーの向上のための取組に加えて、エンドユーザーを支援・サポートするための人材の育成やリテラシーの向上のための取組も重要である。

人材育成について、外部の教育研究機関等との連携を図ることにより、AIの開発者・利用者等自身の取組が深化するとともに、社会全体の底上げに貢献するものと考えられる。AIの開発者・利用者等が単独ではなく、外部の教育研究機関等と連携して取組を推進することが期待される¹⁴⁸。

さらに、民間セクターだけでなく、各府省庁や地方公共団体の職員についても、人材育成に関する取組を行っていくことが重要である。

¹⁴⁶ 本報告書において取りまとめた事例においても、小学生向けのテクノロジー教育や高等学校向け教育プログラムの提供、企業と大学が連携して人材育成に取り組んでいる事例が示されており、参考になるものと考えられる。詳細は、第2章2.（3）及び別冊（グッドプラクティス集）参照。

¹⁴⁷ 「AI戦略2022」において、『AIのELSIに関する取組、例えばAIの構築に際してそもそも倫理等に配慮した設計を行うことや、AI利活用のサイクルにおける監査などを通じて、「責任あるAI」（Responsible AI）を実現する努力も期待される』とされている。

また、人文学や社会科学が主体的に役割を果たすことができる取組領域が見えてくることが重要であるとの指摘もあり、そのような人材を育成することも重要であるものと考えられる。

¹⁴⁸ ELSIについて、例えば、大阪大学は、2020年4月に「社会技術共創研究センター（ELSIセンター）」を設置し、また、中央大学は、2021年4月に「ELSIセンター」を設置するなど、産学連携等を推進する取組が行われている。

結びに代えて

本報告書は、A I 倫理・ガバナンスを中心に国内外の動向及び国際的な議論の動向等を概観し、事業者、有識者及び関係団体からのヒアリングにより収集したA Iの社会実装の推進に関する取組事例を紹介しつつ、さらに、国内外において多くのA I 倫理・ガバナンスに関する原則、指針、ガイドライン等が策定されていること等を踏まえて、A I 開発ガイドライン及びA I 利活用ガイドラインのレビューを行うとともに、見直しに関する論点を整理し、今後の取組を整理した。

第1章において述べたとおり、国内外でA Iに関する様々な取組が行われるとともに、OECDやGPAIなどにおける国際的な議論も進められており、引き続き、国内外の動向をフォローしながら、国際的な議論に貢献していくことが重要である。特に、本年（2022年）末にOECDにおいてデジタル関係の閣僚級会合が開催される予定であり、また、本年（2022年）秋から来年（2023年）秋にかけて日本がGPAIの議長国を務めることになっている。さらに、来年（2023年）に日本でG7の会合が開催される予定となっており、国際的な議論において、日本は積極的に情報を発信し、貢献していくことが重要である。

第2章において、A Iの社会実装に関して先進的あるいは意欲的な取組を行っている事業者、有識者及び関係団体からヒアリングを行い、A Iの社会実装に関する多数の具体的な取組事例を取りまとめた¹⁴⁹。「報告書2021」に引き続き、これだけ多くの具体的な取組事例を掲載しているドキュメントは、他に例を見ないのではないかと考えられ、非常に有用・有益な情報が示されているものと考えられる。協力が得られた事業者、有識者及び関係団体に対して、改めて感謝の意を表すとともに、本報告書が、A Iの開発・利活用に取り組んでいる者又は取り組むことを検討している者などに広く共有され、A Iの社会実装が更に推進されることが期待される。

第3章において、A I 開発ガイドライン及びA I 利活用ガイドラインのレビューを行い、見直しに関する論点を整理したが、ガイドラインをA Iの開発者や利用者にとって、より有用・有益であるものとしていくための検討は重要である。論点は多岐にわたっているが、マルチステークホルダによる丁寧な議論を積み重ね、既存の枠組みにとらわれず、検討を行うことが重要である。

これまで本推進会議が取りまとめたA I 開発ガイドライン及びA I 利活用ガイドラインにおいて、基本理念として『人間がA I ネットワークと共生することにより、その恵沢がすべての人によってあまねく享受され、人間の尊厳と個人の自律が尊重される人間中心の社会を実現すること』が掲げられている。

人間中心の社会を実現することは、全人類に共通する課題である。包摂的で持続可能な人間中心の社会を実現すべく、誰もがA Iを適切な条件で安心して安全に利活用することができるよう取組を進めることが重要である。また、人類共通の課題の解決に向けて、世界規模で「智」のネットワークを形成し、「智」を共創していくことが重要である。

以上を踏まえて、本推進会議において、第4章で述べた今後の取組を踏まえて、引き続き、マルチステークホルダによる「安心・安全で信頼性のあるA Iの社会実装」を更に推進するための取組を行っていくこととしたい。

¹⁴⁹ 別冊（「安心・安全で信頼性のあるA Iの社会実装」の取組事例 ～ グッド プラクティス集～）も併せて参照。

国際シンポジウム
「AIネットワーク社会フォーラム2022」
の概要

国際シンポジウム「AIネットワーク社会フォーラム2022」

国内外の有識者、専門家等が参加し、AIの普及・利活用の鍵となるデータの利活用・流通に関する議論を行うとともに、国際的な動向も踏まえつつ、AIに係る規制やガバナンスの在り方等に関する議論を行うことにより、AIに関連する様々な社会的課題の解決に資することを目的として開催。

○ 日時：2022年3月1日（火）13:00～18:00（オンライン）

13:00-13:05	開会挨拶 (5)	金子 恭之 総務大臣
13:05-13:25	特別講演 (20)	「How to value data in a world with AI」 ローラ・フェルトカンブ コロンビア大学 教授
13:25-14:05	特別対談 (40)	「AI・データ産業革命」 ローラ・フェルトカンブ コロンビア大学 教授 岩田 一政 公益社団法人日本経済研究センター 理事長
14:05-15:25	パネル ディスカッション (80)	「データ×AIが切り拓く明日の世界」 ＜モデレーター＞ 大橋 弘 東京大学公共政策大学院 院長・教授 ＜パネリスト＞ 瀧 俊雄 株式会社マネーフォワード 執行役員 サステナビリティ担当 CoPA、Fintech 研究所長 玉城 絵美 琉球大学 工学部 知能情報コース 教授、H2L, Inc. CEO 武藤 真祐 医療法人社団鉄祐会 理事長、株式会社インテグリティ・ヘルスケア 代表取締役会長
15:25-15:55	特別講演 (30)	「AI, globalisation, and the future of work」 リチャード・ボールドウィン ジュネーブ国際高等問題研究所 教授 【聞き手】 関口 和一 株式会社MM総研 代表取締役所長、元日本経済新聞社論説委員
15:55-16:00	－ 休憩 －	
16:00-16:15	特別挨拶 (15)	須藤 修 中央大学 国際情報学部 教授、中央大学 ELSI センター 所長、東京大学大学院 情報学環 特任教授
16:15-16:35	特別講演 (20)	「AI Ethics: Translating Principles into Governance and Regulation」 フランチェスカ・ロッシ IBM コーポレーション IBM フェロー、IBM AI 倫理グローバルリーダー
16:35-17:55	パネル ディスカッション (80)	「AIと共生する世界に向けたガバナンス」 ＜モデレーター＞ 須藤 修 中央大学 国際情報学部 教授、中央大学 ELSI センター 所長、東京大学大学院 情報学環 特任教授 ＜パネリスト＞ 江間 有沙 東京大学 未来ビジョン研究センター 准教授 平野 晋 中央大学 国際情報学部 学部長・教授 福田 剛志 日本アイ・ビー・エム株式会社 執行役員 東京基礎研究所 所長 望月 康則 日本電気株式会社 NEC フェロー
17:55-18:00	閉会挨拶 (5)	中西 祐介 総務副大臣

※ 参加者の所属・役職については開催当時のもの

各セッションの概要 ①

特別講演

ローラ・フェルトkamp 氏
(コロビア大学 教授)



“How to Value data in a world with AI”

- AIを活用してデータからナレッジを生産することでより多くの価値が創出される。データ量が爆発的に伸びる中、企業の価値は保有しているデータ量で決まる。米国の大企業が成長している要因は、有形資産ではなく、保有しているデータという無形資産が膨大になっているからである。企業価値を高めるには、AIによるデータ価値の最大化が大きな鍵を握っている。
- データ活用は、データ集約性と労働集約性によって成果が変わる。データを管理する労働者、古い技術スキルを持つ労働者を最適な比率で雇用した上で、将来を見据え、AIを活用してナレッジを生成する労働者の雇用を増やす労働ミックスが不可欠となる。実際に採用率をみると、ここ数年、AIワーカーの伸び率は、古い技術スキル労働者、データ管理者を上回っている。また、賃金も、AIワーカー、古い技術スキル労働者、データ管理者の順に高い。
- 現状、AIワーカーの労働者に占める割合は低く、データ活用による収益への貢献度は古い技術スキル労働者の方が高い。産業革命においても、新しい技術を使う労働者は5～13%であった。AIによるデータ活用が価値創出の源泉であることは疑う余地はなく、企業における労働分配率は変わっていくであろう。

特別挨拶

須藤 修 氏

(中央大学 国際情報学部 教授、
中央大学 ELSIセンター 所長、
東京大学大学院 情報学環 特任教授)



- インターネットの民間利用開始以来の破壊的な技術トレンドの中であって、AIは極めて重要な要素であり、AI原則及びガバナンスの取組が大きな注目を集めている。
- 日本政府においては、AIの適切な開発と社会実装のため、2019年3月に「人間中心のAI社会原則」(①人間中心の原則、②教育・リテラシーの原則、③プライバシー確保の原則、④セキュリティ確保の原則、⑤公正競争確保の原則、⑥公平性、説明責任及び透明性の原則、⑦イノベーションの原則)を取りまとめた。これらの原則はAIを利活用して新たな社会を構築する上で核となるものである。
- 自分が議長を務めている総務省のAIネットワーク社会推進会議においては、AI社会のあるべき姿を考え、AI利活用の仕方やデータ流通などについて多面的な検討を行ってきた。
- 2023年にG7が日本で開催される予定であり、AIの在り方について積極的に提案する準備が進んでいる。様々な動きがある中、国際社会と協調しながら、新たな社会を作っていくことが求められている。

各セッションの概要 ②

特別講演

リチャード・ボールドウィン 氏

(ジュネーブ国際高等問題研究所
教授)



“AI, globalization, and the future of work”

<聞き手> 関口 和一 氏 (株式会社MM総研 代表取締役所長、元日本経済新聞社論説委員)

- 2019年に著した『GLOBOTICS』において、デジタル技術がホワイトカラーと専門職に対して、仕事の自動化とグローバル化を同時にもたらすと述べた。コンピューターが新たな認知スキルを獲得したことによって、翻訳や編集、グラフィックスなど、人の介在が必要であった多くの仕事が、機械学習で実現できるようになった。
- デジタル技術は、最初はなかなか進まず、あるときから極端に成長し始め、そこでディスラプション（破壊的創造）が起きる。
- 未来では、ソフトウェア・ロボットで自動化できる仕事は、人の手を離れる。新興市場の安価な労働力に任せられる仕事は、オフショアされるであろう。しかし、倫理、創造性、好奇心、モチベーションなど、人間的で複雑なものに関するビッグデータを収集することは、リモート・インテリジェンスやAIには困難である。
- 簿記や会計、一定の法律業務など、自動化できる仕事は比較的早くロボットに置き換わる一方で、競争力のある労働者はより多くの機会を得る。
- 政府には、労働者の再適応を支援する再教育や移転支援など、積極的な労働市場政策が求められる。

特別講演

フランチェスカ・ロッシ 氏

(IBMコーポレーション IBMフェロー、
IBM AI倫理グローバルリーダー)



“AI Ethics : Translating Principles into Governance and Regulation”

- AI倫理は、認識の時代、原則の時代を経て、現在は「AI倫理3.0」と呼ばれる運用の段階にある。
- 2021年に発表された「EU AI ACT」と呼ばれる規制案は1つの形であり、AIそのものではなく、AIのアプリケーションを規制していることが特色として挙げられる。規制案においては、4つのリスクレベルが定義され、レベルごとに義務（AIシステムの提供者やユーザーが対象）とアプリケーションに関する制約が規定されており、「信頼できるAI」の7つの要件も定義されている。企業はこの規制に則って、AIソリューションを構築しなければならない。
- IBMは、AIは人間の知能を拡張するもので、透明性と説明可能性が必要であるとする原則を定め、この原則の下で信頼できるAIの姿を設定している。IBMの原則は、EUの規制案における「信頼できるAI」の7つの要件と類似したものとなっており、社内ガバナンスを効かせて、原則を運用に落とし込んでいる。
- 規制はAI倫理を守る多くの理由のうちの1つに過ぎない。規制を遵守することは重要であるが、AI倫理というのは、単なる規制遵守を超えるものである。

各セッションの概要 ③

特別対談

「AI・データ産業革命」

ローラ・フェルトカンプ 氏

(コロンビア大学 教授)



岩田 一政 氏

(日本経済研究センター 理事長)



- AI技術には「予測アルゴリズム」と「発明アルゴリズム」があるが、生産性に直接影響があって、高い成長率をもたらすのはどちらであると考えるか。また、産業革命と比較して違いはあるか。(岩田)
- 産業革命において、技術によってモノの生産方式が変わったのと同様に、AI技術は、ナレッジの生み出し方を変える。資本は、設備投資に使うものと研究開発に使うものに分けられるが、高い成長を生み出すのは研究開発に当たる「発明アルゴリズム」の方であると考え。(フェルトカンプ)
- AI革命が米国経済に与える影響をどのように見ているか。(岩田)
- 大きな影響があることは確かであるが、AI技術が汎用技術となり、様々な業界で採用されるようになるには数十年かかるであろう。また、技術だけで成長し続けられるということではない。(フェルトカンプ)
- 日本経済研究センターが行った調査において、AIとIoTを最大限活用した場合、日本経済の成長率は1%以下から5%に変わるという予測結果が得られている。長期的にみれば、技術的な進歩は収れんしていくが、この影響力は大きいと思われる。(岩田)
- 企業は、AI技術が古い技術よりはるかに生産性が高く、利益につながることは分かっている。だからこそ、AIの利活用を推進しており、人材育成にも取り組んでいる。(フェルトカンプ)
- 異なる業種や産業、あるいは国を越えたデータの自由な流通は、企業の生産性にどのような影響を与えると考えるか。(岩田)
- データ活用によって、我々の生活は、より便利で効率的になっており、生産性も向上している。他方、データ共有が進むことによって、企業行動が変容する可能性がある。プライバシーの問題を含め、データの自由な流通には注意が必要である。(フェルトカンプ)
- 日本はデータ共有の仕組みができていないため、AIの利活用が遅れている。中国では既に8割以上の企業がAIを利活用しているが、日本においては4割弱にとどまっている。(岩田)
- AIの利活用は始まったばかりであり、今後、AIの利活用による労働分配率の変化や無形資産の変化などを注視していくことが必要である。(フェルトカンプ)
- データの流れはお金の流れであるとの指摘がある。データと通貨が合わさる世界が生まれるとしたら、その結果はどのようになると考えるか。(岩田)
- お金はサービスや製品の提供に対して対価を示すものであるが、データはナレッジである。ナレッジは共有されなければならないが、私たちはナレッジを共有することで成長できる。個人情報保護と情報の所有者の尊重とのバランスを取りながら、ナレッジやデータを共有していくことが重要である。(フェルトカンプ)

各セッションの概要 ④

パネルディスカッション

「データ×AIが切り拓く明日の世界」

【モデレーター】

大橋 弘 氏

(東京大学公共政策大学院
院長・教授)



【パネリスト】

瀧 俊雄 氏

(株式会社マネーフォワード 執行役員
サステナビリティ担当CoPA、
Fintech研究所長)



玉城 絵美 氏

(琉球大学 工学部 知能情報コース
教授、H2L, Inc. CEO)



武藤 真祐 氏

(医療法人社団鉄祐会 理事長、
株式会社インテグリティ・ヘルスケア
代表取締役会長)



- データが経済的な付加価値を加速度的に高める中、医療や金融などの分野においては、産業の枠を意識した分析の掘り下げが必要である。(大橋)
- 日本の金融業界は、プログラム同士をつなぐアプリケーション・プログラミング・インターフェース(API)化が進んでおり、連携時には明確な情報移転の同意が必要である。ただし、データに係る権利の考え方や法整備は不十分である。(瀧)
- クオリティデータを用いた個別化医療を実現するには、人々が価値を感じられるようにプラットフォームを構築する必要があるが、医療提供者側の負荷を軽減することが重要である。よいAIを提供しても、医療従事者の負担が減らなければ意味がない。(武藤)
- 「人が物体に作用し、臨場感を得る感覚」のデジタル化とそのデータ出力に挑んでいる。これが実現すると、部屋の中にながら仮想空間(メタバース)上で働いたり、スポーツや観光で体験共有を得るなど、多くの付加価値が生まれる。他方、様々な課題もあり、研究者ベースでは、実用化に向けたルールづくりなどの議論が既に行われている。(玉城)
- 金融のデジタル化で重要なのは情報漏洩など事故を起こさないことである。ベンチャーであっても、フィンテック事業者にはセキュリティスタンダードが必要である。(瀧)
- オンライン診療が増えつつあるが、新たなデジタルデバイドが生まれつつある。これまで知らずに我慢していた負荷を減らすことが重要である。(武藤)
- データをどのように統合し、ユーザーなどに成果を返していくのか、ユーザーの動きや変容も調査することが必要である。自己情報をコントロールしていくことが重要である。(玉城)
- DFTT(信頼ある自由なデータ流通)の実現に向けて、日本ではどのような環境整備が必要か。(大橋)
- 個人情報の移転は法律に守られており議論しやすいが、統計情報や加工情報については議論が難しい。企業が、データステートメントを出すといった対応が必要である。(瀧)
- 医療の民主化を目標として、自己決定権を持ち、きちんと統制された枠組みを作る必要がある。予防医療においても、多分野のデータの共有・活用が必要である。(武藤)
- 環境整備の課題は、「データオーナーシップ」、「同意のユーザーエクスペリエンス」、「分野を超えたデータセキュリティと共有」の3つである。(玉城)

各セッションの概要 ⑤

パネルディスカッション

【モデレーター】

須藤 修 氏

(中央大学 国際情報学部 教授、
中央大学 ELSIセンター 所長、
東京大学大学院 情報学環
特任教授)

【パネリスト】

江間 有沙 氏

(東京大学 未来ビジョン研究センター
准教授)

平野 晋 氏

(中央大学 国際情報学部
学部長・教授)

福田 剛志 氏

(日本アイ・ビー・エム株式会社
執行役員 東京基礎研究所 所長)

望月 康則 氏

(日本電気株式会社
NECフイロー)

「AIと共生する世界に向けたガバナンス」



- AIガバナンスは公平性や透明性、信頼性などの原則を定め、それを実践する段階へと進んでいる。AIガバナンスに関連する課題解決の取組は、「競争領域」ではなく「協調領域」であり、様々な関係者を巻き込んで議論を重ね、エコシステムを構築することが重要である。(江間)
- AI利活用の目的は、人間の知性を補うことであり、置き換えることではない。新たな技術も、透明かつ説明可能でなければならない。IBMは、信頼と透明性の原則を明らかにし、AIに対する精密規制についても、技術を規制するのではなく、個別に判断することとしている。(福田)
- AIによる社会価値の創出には、機能や性能と同時に、社会から信頼を獲得し、受け入れられることが重要である。AIガバナンスがイノベーション創出の阻害要因や参入障壁とならないための配慮が不可欠である。(望月)
- AIの規制に対して、「ハードロー」や「ソフトロー」に関する議論が行われているが、二分法的な議論ではなく、これから顕在化してくる多様で具体的なリスクに対して、合理的な規制は何か、知恵を出し合うことが重要である。(平野)
- EUの規制法案について、転ばぬ先の杖がよいという価値観は理解できるが、イノベーションの遅れによるリスクも危惧される。規制内容の精査が必要である。(平野)
- AIは発展途上の技術であり、技術もリスクも社会の受容性も時間とともに変化していく前提で考えることが重要である。(福田)
- EUのハイレベルな考え方は重要であるが、民間の立場からは、ユースケースベースの重要性を主張し、協調しつつ、現実的に遵守可能な形を作っていくことが大事である。(望月)
- プライバシーや人間の尊厳に対する考え方、働き方など、国や地域で違いがある。アカデミアが持つ情報の共有や発信といった機能が役に立つと考えている。(江間)
- AIの規制については、動かしながら状況に応じて変えていくべきであり、一番厳しい規制に合わせる必要性も出てくるのではないかと。相互運用性の確保が重要である。(福田)
- AIは、使う側のメリット等の理解を得ながら進めることが重要である。(望月)
- EUとの協調路線を模索することが重要であり、人材育成や議論が不可欠である。(江間)
- 人間中心の社会においては、AIが暴走しないよう、「ELSI (エルシー) 」(倫理的・法的・社会的課題) の検討にハンドルとブレーキの役目を担わせることが重要である。(平野)
- 各国の法制・産業構造・行政システム等に違いがあることに留意が必要である。EU等に対しても、そのことを明確に示した上で、国際協調を進めていく必要がある。(須藤)

AI開発ガイドライン及びAI利活用ガイドライン に関するレビュー（詳細）

(1) レビューの概要

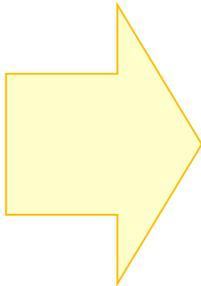
AIの倫理・ガバナンスについて、本推進会議において、AI開発ガイドライン^{※1}（2017年7月）やAI利活用ガイドライン（2019年8月）を策定し、OECDやG7、G20等における国際的な議論に貢献してきた。

その後、国内外の動向や国際的な議論の動向をフォローしつつ、有識者や開発者、利用者（AIサービスプロバイダ、ビジネス利用者、消費者的利用者）などのステークホルダからヒアリングを行い、「報告書2020」（2020年7月）及び「報告書2021」（2021年8月）において、その取組等を取りまとめた。



- 国内外において、多くのAIの倫理・ガバナンスに関する原則・指針・ガイドライン等が策定されている（特に、AI開発ガイドライン、AI利活用ガイドラインの策定後の動向に注視する必要がある。）。
- ヒアリングにおいて示された各ステークホルダの取組の中には、AI開発ガイドラインやAI利活用ガイドラインの射程を超えて、「安心・安全で信頼性のあるAIの社会実装」に関する優れた取組として評価することができるものがある。
- 国内外においてAI利活用・社会実装が進む中で、リスクが顕在化したり、インシデントが発生したりした事例がある。

次の観点から、AI開発ガイドライン及びAI利活用ガイドラインのレビューを実施。

- 
- ① 海外における原則・指針・ガイドライン等との比較
 - ② 国内における原則・指針・ガイドライン等との比較
 - ③ ヒアリング等において示された取組事例との対応
 - ④ AI開発・利活用及び社会実装に関する事例との対応

このほか、EUのAI規制法案^{※2}の動向等を踏まえた対応が重要であり、その点も考慮してレビューを実施。

※1 「国際的な議論のためのAI開発ガイドライン案」のこと(以下同じ。)

※2 「人工知能の関する調和の取れたルールを定める規則の提案」(2021年4月)のこと(以下同じ。)

(2) 海外における原則・指針・ガイドライン等との比較からの検証

○ 海外において策定された原則・指針・ガイドライン等（以下「指針等」ということがある。）を踏まえて、現行のAI開発ガイドライン及びAI利活用ガイドラインの射程、内容等について検証。

■ 調査対象国、調査対象の指針等

- ・ 16箇国^{※3}（米国、カナダ、英国、イタリア、オランダ、スウェーデン、デンマーク、ドイツ、ノルウェー、フィンランド、フランス、インド、韓国、シンガポール、中国、オーストラリア）、EU、国際機関（国際連合教育科学文化機関（UNESCO）、世界保健機関（WHO））
- ・ 政府機関、業界団体等、大手グローバル企業等が公表している67個の指針等^{※4}を調査（16～23ページ参照）
 - 政府機関^{※5}・国際機関（40個）、業界団体等（10個）、グローバル企業等（17個）

※3 調査対象国の選定に当たっては、AI準備指数、AI出版物数、AI特許出願数、AI研究者数を考慮し、いずれかが10位以内の国を選定。

また、海外の動向、国際的な議論の動向を踏まえて、EUや2つの国際機関を加えて調査を実施。

※4 主としてAI利活用ガイドラインを公表した2019年8月以降に公表されたものを対象に調査を実施。

※5 EUの欧州委員会及びAIハイレベル専門家グループを含む。

■ 尊重すべき価値

- ・ 調査を通じて、22の「尊重すべき価値」を確認^{※6}

1. 人間中心	9. 適正な学習（学習データの質）	17. 堅牢性
2. 人間の尊厳	10. AI間の連携	18. 責任
3. 多様性、包摂	11. 安全性	19. 追跡可能性
4. 持続可能な社会	12. セキュリティ	20. モニタリング、監査
5. 国際協力	13. プライバシー	21. ガバナンス
6. 適正な利用	14. 公平性	22. その他（コスト、効果測定）
7. 教育、リテラシー	15. 透明性、説明可能性	
8. 人間の判断の介在、制御可能性	16. アカウンタビリティ	

※6 1.～16.までは、AIガイドライン比較表（「報告書2019」別紙2）における項目と同じ項目。17.～22.が新たに確認した項目。

【注】 総務省「AIに係る原則・ガイドライン等の策定状況、規定事項等に関する調査研究」による。網羅的なものではなく、当該調査に基づいて整理したものである。

(2) 海外における原則・指針・ガイドライン等との比較からの検証

	指針等の数	人間中心	人間の尊厳	多様性 包摂	持続可能な社会	国際協力	適正な利用	リテラシー 教育	人間の判断の介入 制御可能性	(学習データの質) 適正な学習 AI間の連携	安全性	セキュリティ	プライバシー	公平性	説明可能性 透明性	アカウンタビリティ	堅牢性	責任	追跡可能性	モニタリング 監査	ガバナンス	その他	
米国	6	1	2			1	1		1	1		3	3	1	4	5	4	1	2	2	3		
カナダ	1								1						1								1
英国	6		2		1			1		2		1	1	2	5	5	5	1	2		2		1
イタリア	1									1			1		1								
オランダ	2		1				1		1	2			1	2	2	1		1	1				
スウェーデン	2		1	1			1	1				1	1	1	2								
デンマーク	2		2	2			1		1			1	1	1	2	2	1	1			1		
ドイツ	3	1	2	2	3				1	2		2	2	2	2	3	2	3	1	1	1		
ルウエー	2		2		1		2		1			1		1	2	1	1	1		1	1		
フィンランド	2		2	1			1		1			1		1	2	1		2	2	1			
フランス	1								1					1	1	1					1		
インド	1		1	1								1	1	1	1	1		1					
韓国	1		1	1		1						1		1	1	1	1						
シンガポール	2		1											2	2	1							
中国	3		3	2	2	2	2	2	3	1		3	2	2	3	2	2	2		2	3	3	
オーストラリア	1	1	1									1	1	1	1	1		1			1		
EU	2		2	1	2	1	2		1			2	1	2	2	1	2	2	2	1	1		
国際機関	2		2	2	2	1	2	1	1		1	2	1	2	1	2	1		1	1	1		
合計	40	3	25	13	11	6	13	6	12	9	1	19	14	20	31	35	25	12	13	11	16	3	2

(注1) 政府機関(EUの欧州委員会及びAIハイレベル専門家グループを含む。)及び国際機関を対象として整理。

(注2) 記載があるものはピンク色に着色。また、各国等の指針等の数に対し、半数を超える指針等に記載がある場合、赤字で表示。

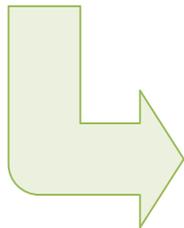
(2) 海外における原則・指針・ガイドライン等との比較からの検証

<地域ごとの比較>

	指針等の数	人間中心	人間の尊厳	多様性 包摂	持続可能な社会	国際協力	適正な利用	リテラシー 教育	人間の判断の介入制 御可能性	適正な学習 (学習データの質)	AI間の連携	安全性	セキュリティ	プライバシー	公平性	説明可能性 透明性	アカウントビリティ	堅牢性	責任	追跡可能性	モニタリング 監査	ガバナンス	その他
北米	7	1	2			1	1	1	1	1		3	3	1	4	6	4	1	2	2	3		1
欧州	23	1	14	7	7	1	8	2	7	7		8	6	12	18	20	14	8	8	6	8		1
アジア	7		6	4	2	3	2	2	3	1		5	3	4	7	6	5	3	1	2	3	3	
大洋州	1	1	1									1	1	1	1	1	1		1		1		
合計	38	3	23	11	9	5	11	5	11	9	0	17	13	18	30	33	24	12	12	10	15	3	2

(注1) 政府機関(EUの欧州委員会及びAIハイレベル専門家グループを含む。)を対象として整理。

(注2) 記載があるものはピンク色に着色。また、各国等の指針等の数に対し、半数を超える指針等に記載がある場合、赤字で表示。



北米	欧州	アジア
<ul style="list-style-type: none"> ① 透明性、説明可能性 ② 公平性 ② アカウンタビリティ ④ 安全性 ④ セキュリティ ④ モニタリング、監査 	<ul style="list-style-type: none"> ① 透明性、説明可能性 ② 公平性 ③ 人間の尊厳 ③ アカウンタビリティ ⑤ プライバシー 	<ul style="list-style-type: none"> ① 公平性 ② 人間の尊厳 ② 透明性、説明可能性 ④ 安全性 ④ アカウンタビリティ

(2) 海外における原則・指針・ガイドライン等との比較からの検証

尊重すべき価値	ガイドラインにおいて関係があるものと考えられる記載（主なもの）	主な記載内容の例
1.人間中心	<p>【AI開発ガイドライン】</p> <ul style="list-style-type: none"> ○ 基本理念 <ul style="list-style-type: none"> ・ 人間の尊厳と個人の自律が尊重される人間中心の社会を実現すること <p>【AI利活用ガイドライン】</p> <ul style="list-style-type: none"> ○ 基本理念 <ul style="list-style-type: none"> ・ 人間の尊厳と個人の自律が尊重される人間中心の社会を実現すること 	<ul style="list-style-type: none"> ・ 国家安全保障や信頼できるパートナーシップを強化できるように、人間を中心としたAIの開発・利用に留意する。【3番】 ・ システムはそれを利用する人、あるいは、その決定によって影響を受ける人を中心に考えなければならない。【41番】 ・ AIシステムは人間を中心とした価値観を重視し、人権、多様性、個人の自律性を尊重すべきである。【62番】
2.人間の尊厳	<p>【AI開発ガイドライン】</p> <ul style="list-style-type: none"> ○ 倫理の原則 <ul style="list-style-type: none"> ・ 人間の尊厳と個人の自律を尊重する <p>【AI利活用ガイドライン】</p> <ul style="list-style-type: none"> ○ 尊厳・自律の原則 <ul style="list-style-type: none"> ・ 人間の尊厳と個人の自律を尊重する 	<ul style="list-style-type: none"> ・ AIシステムは、人間の尊厳、権利、自由を守る国際条約と矛盾しないように設計されなければならない。エシックス・バイ・デザインにより合理的に予測できる誤用を避けるための措置を講ずる責任がある。【7番】 ・ 国籍、性別、国や民族の出身、人種、宗教、言語にかかわらず、個人の平等な尊厳と自律性を尊重する必要がある。【8番】 ・ 人類共通の価値観に従い、人権を尊重するとともに、人間にとっての利益を追求し、国や地域の倫理や道徳を守ることに留意する。【58番】

【注】「主な記載内容の例」に記載されている番号は、参照している原則・指針・ガイドライン等（16～23ページのリストに記載されている番号と対応）を示している。

(2) 海外における原則・指針・ガイドライン等との比較からの検証

尊重すべき価値	ガイドラインにおいて関係があるものと考えられる記載（主なもの）	主な記載内容の例
3.多様性包摂	<p>【AI開発ガイドライン】</p> <ul style="list-style-type: none"> ○ 基本理念 <ul style="list-style-type: none"> ・ 人間がAIネットワークと共生することにより、その恵沢がすべての人によってあまねく享受され、人間の尊厳と個人の自律が尊重される人間中心の社会を実現すること ○ 利用者支援の原則 <ul style="list-style-type: none"> ➢ ユニバーサルデザインなど社会的弱者の利用を容易にするための取組に努める ○ 関係するステークホルダに期待される役割 <ul style="list-style-type: none"> ・ 本ガイドラインの運用や見直しにおいて各国政府、国際機関、開発者、市民社会を含む利用者など多様なステークホルダ間の対話の促進に向けた環境整備に努めることが期待される ・ AIをめぐる議論の多様性を確保しつつ、AIの便益の増進及びリスクの抑制について、認識の共有を図り、相互に協力するよう努めることが期待される <p>【AI利活用ガイドライン】</p> <ul style="list-style-type: none"> ○ 基本理念 <ul style="list-style-type: none"> ・ 人間がAIネットワークと共生することにより、その恵沢がすべての人によってあまねく享受され、人間の尊厳と個人の自律が尊重される人間中心の社会を実現すること ・ AIの利活用において利用者の多様性を尊重し、多様な背景と価値観、考え方を持つ人々を包摂すること 	<ul style="list-style-type: none"> ・ データの処理によって特定の集団を疎外し、汚名を着せるような偏見や差別をしてはならない。不利な立場にある人や弱い立場にある人が技術開発の恩恵を受けられるように、ターゲットを絞った取組を行うべきである。新しい技術の開発と使用においては、関連する専門家グループ、ユーザーグループ、組織を巻き込むなどして、多様性を確保しなければならない。【38番】 ・ AIの開発及び活用の前段階でユーザーの多様性と代表性を反映しなければならず、性別、年齢、障害、地域、人種、宗教、国家など個人の特性による偏向と差別を最小化し、商用化されたAIはすべての人に公平に適用されなければならない。社会的弱者と脆弱な層のAI技術とサービスへのアクセスを確保し、AIが与える利益は特定の集団ではなく、すべての人に均等に分配されるように努力しなければならない。【53番】 ・ AIシステムの開発者の多様性を確保し、包括的な学習データの活用に努め、人種、性別、国籍、年齢、宗教的信条に基づいてユーザーを差別することなく、アルゴリズムのテストと検証を継続的に行うことに留意する。【60番】
4.持続可能な社会	<p>【AI開発ガイドライン】 (記載なし)</p> <p>【AI利活用ガイドライン】</p> <ul style="list-style-type: none"> ○ 基本理念 <ul style="list-style-type: none"> ・ 個人、地域社会、各国、国際社会が抱える様々な課題の解決を図り、持続可能な社会を実現すること 	<ul style="list-style-type: none"> ・ 地球環境の持続可能性を確保できるようにAIを開発・利用しなければならない。【19番】 ・ AI技術は、地球上の生物が生きるための環境の構築、持続的な人類の繁栄、将来の世代のための良好な環境の保全という人間の責任に沿ったものでなければならない。【63番】 ・ AIシステムのライフサイクルに関与するすべてのアクターは、環境・生態系の保護・回復、持続可能な開発のために設計された国際法、国内法、基準及び慣行(予防策等)に従わなければならない。【66番】

(2) 海外における原則・指針・ガイドライン等との比較からの検証

尊重すべき価値	ガイドラインにおいて関係があるものと考えられる記載（主なもの）	主な記載内容の例
5.国際協力	<p>【AI開発ガイドライン】</p> <ul style="list-style-type: none"> ○ 基本理念 <ul style="list-style-type: none"> ・ AIシステムの研究開発の在り方について、非拘束的なソフトローたる指針やそのベストプラクティスをステークホルダ間で国際的に共有すること ○ 関係するステークホルダに期待される役割 <ul style="list-style-type: none"> ・ 各国政府及び国際機関は、本ガイドラインの運用や見直しにおいて各国政府、国際機関、開発者、市民社会を含む利用者など多様なステークホルダ間の対話の促進に向けた環境整備に努めることが期待される <p>【AI利活用ガイドライン】</p> <ul style="list-style-type: none"> ○ 基本理念 <ul style="list-style-type: none"> ・ AIの利活用の在り方について、非拘束的なソフトローたる指針やベストプラクティスを国際的に共有すること 	<ul style="list-style-type: none"> ・ 各国のAIガバナンスの原則と実践を十分に尊重することを前提に、幅広いコンセンサスによる国際的なAIガバナンスの枠組み及び標準規範の形成を推進する必要がある。【57番】 ・ 我々は、「自律型」技術への平等なアクセスと、社会全体及び社会内での利益の公正な分配と機会均等に向けて、世界的に協調していく必要がある。【63番】 ・ 包括的なAIガバナンス、AIのメリットの共有、公正な技術進展、開発目標への貢献に向けて、AIシステムのライフサイクルを通じた様々なステークホルダの参加が必要である。ステークホルダには、政府、政府間組織、技術コミュニティ、市民社会、研究者、学界、メディア、教育、政策立案者、民間企業、人権機関、平等の促進に取り組む機関、差別を監視する機関、青少年と子供のための団体等が含まれる。【66番】
6.適正な利用	<p>【AI開発ガイドライン】</p> <ul style="list-style-type: none"> ○ 制御可能性の原則 <ul style="list-style-type: none"> ➢ 制御可能性を確保するため、採用する技術の特性に照らして可能な範囲において、人間や信頼できる他のAIによる監督（監視、警告など）や対処（AIシステムの停止、ネットワークからの切断、修理など）の実効性に留意することが望ましい <p>【AI利活用ガイドライン】</p> <ul style="list-style-type: none"> ○ 適正利用の原則 <ul style="list-style-type: none"> ・ 人間とAIシステムとの間及び利用者間における適切な役割分担のもと、適正な範囲及び方法でAIシステム又はAIサービスを利用するよう努める 	<ul style="list-style-type: none"> ・ AIを利用するケースが、そのAIが訓練されたユースケースと一致していることを確認しなければならない。【3番】 ・ 個人は、どのようなデータが収集され、何のために、どのような状況で使用されるか、について自分自身でコントロールできるようにするべきである。【38番】 ・ AIシステムがステークホルダの権利や利益に及ぼす影響について、ステークホルダからインフォームドコンセントを得る必要がある。予期しない事態が発生した場合、その権利や利益が侵害されないように、データとサービスを破棄できるメカニズムを確立する必要がある。【59番】 ・ 人間は、意思決定と行動を自律的システムに委ねるかどうか、いつ、どのように委ねるかを選択し、適正に利用しなければならない。【63番】

(2) 海外における原則・指針・ガイドライン等との比較からの検証

尊重すべき価値	ガイドラインにおいて関係があるもの と考えられる記載（主なもの）	主な記載内容の例
7.教育 リテラシー	<p>【AI開発ガイドライン】 (記載なし)</p> <p>【AI利活用ガイドライン】</p> <ul style="list-style-type: none"> ○ 適正利用の原則(参考) <ul style="list-style-type: none"> ➢ 消費者的利用者は、AIの判断に対し、消費者的利用者が最終判断をすることが適当とされている場合には、適切に判断ができるよう必要な能力及び知識を習得しておくことが望ましい。 ➢ アクチュエータ等を通じて稼働するAIの利活用において、一定の条件に該当することにより人間による稼働に移行することが予定されている場合には、消費者的利用者は、移行前、移行中、移行後等の各状態における責任の所在を予め認識しておくことが望ましい。また、AIサービスプロバイダから、移行条件、移行方法等についての説明を受け、必要な能力及び知識を習得しておくことが望ましい。 	<ul style="list-style-type: none"> ・ すべての市民は、AIとともに精神的、感情的、経済的に繁栄できるように教育を受ける権利を有するべきである。【20番】 ・ AI技術に対する国民の不安を解消し、安全上の注意を喚起し、既存及び将来の労働力の課題に対処するため、一般市民に対するAIに関する普及教育、関連する実務者に対する道徳的・倫理的教育、仕事が代替された人々に対するデジタル労働スキルの再教育に積極的に取り組むことに留意する。【60番】 ・ 社会のすべての構成員がAIシステムの利用について十分な情報を得た上で意思決定し、不当な影響から保護できるよう、デジタルスキルやAI倫理のトレーニング等が必要である。【66番】
8.人間の判断の介入 制御可能性	<p>【AI開発ガイドライン】</p> <ul style="list-style-type: none"> ○ 制御可能性の原則 <ul style="list-style-type: none"> ・ AIシステムの制御可能性に留意する ➢ 制御可能性を確保するため、採用する技術の特性に照らして可能な範囲において、人間や信頼できる他のAIによる監督（監視、警告など）や対処（AIシステムの停止、ネットワークからの切断、修理など）の実効性に留意することが望ましい <p>【AI利活用ガイドライン】</p> <ul style="list-style-type: none"> ○ 適正利用の原則 <ul style="list-style-type: none"> ➢ AIによりなされた判断について、必要かつ可能な場合には、その判断を用いるか否か、あるいは、どのように用いるか等に関し、人間の判断を介入させることが期待される。その場合、人間の判断の介入の要否について、基準例を踏まえ、利用する分野やその用途等に応じて検討することが期待される ○ 公平性の原則 <ul style="list-style-type: none"> ➢ AIによりなされた判断結果の公平性を保つため、AIを利活用する際の社会的文脈や人々の合理的な期待を踏まえ、その判断を用いるか否か、あるいは、どのように用いるか等に関し、人間の判断を介入させることが期待される 	<ul style="list-style-type: none"> ・ AIシステムは、意図しない危害や混乱を検知して回避する能力を持ちながら、意図した機能を果たすように設計されるべきである。また、意図しないエスカレーションやその他の動作を示す場合は、自動的に又は人間によって停止できるように設計されるべきである。【4番】 ・ ユーザーと第三者の安全を最優先しなければならず、AIシステムにはAIが動作する状況に合わせて、人間がAIを制御できるようにセーフガードが必要である。【7番】 ・ 人間が完全に自律的な意思決定を行う権利、AIが提供するサービスを受けるかどうかを選択する権利、いつでもAIとの対話から撤退する権利、いつでもAIシステムの運用を停止する権利を保証し、AIを人間の管理下に置くように留意する。【58番】

(2) 海外における原則・指針・ガイドライン等との比較からの検証

尊重すべき価値	ガイドラインにおいて関係があるものと考えられる記載（主なもの）	主な記載内容の例
9.適正な学習 （学習データの質）	<p>【AI開発ガイドライン】</p> <ul style="list-style-type: none"> ○ 倫理の原則 <ul style="list-style-type: none"> ➢ 採用する技術の特性に照らし可能な範囲で、AIシステムの学習データに含まれる偏見などに起因して不当な差別が生じないよう所要の措置を講ずるよう努めることが望ましい <p>【AI利活用ガイドライン】</p> <ul style="list-style-type: none"> ○ 適正学習の原則 <ul style="list-style-type: none"> ・ AIシステムの学習等に用いるデータの質に留意する 	<ul style="list-style-type: none"> ・ あらゆる段階でデータの責任ある利用を促進するとともにデータの完全性を確保するために、産業界には、データのパラメータと特性を理解すること、有害なバイアスの可能性を認識すること、AIシステムの展開前及び展開中にバイアスの有無をテストすることについて責任がある。【7番】 ・ 我々は使用されているすべてのデータセットの全体像を文書化し、各データセットとその出典を詳細に記述することを推奨している。【33番】 ・ 機械学習システムのトレーナーはAIシステムの学習やシステムの運用中に行われる継続学習に使用されるデータのバイアスの有無を確認する責任がある。【51番】
10.AI間の連携	<p>【AI開発ガイドライン】</p> <ul style="list-style-type: none"> ○ 連携の原則 <ul style="list-style-type: none"> ・ AIシステムの相互接続性と相互運用性に留意する <p>【AI利活用ガイドライン】</p> <ul style="list-style-type: none"> ○ 連携の原則 <ul style="list-style-type: none"> ・ AIシステム又はAIサービス相互間の連携に留意する 	<ul style="list-style-type: none"> ・ 相互運用性に関し、技術の変化、利害関係者の新たな出現、社会から疎外されたグループ等による介入を可能にするための措置を講ずる必要がある。【66番】
11.安全性	<p>【AI開発ガイドライン】</p> <ul style="list-style-type: none"> ○ 安全の原則 <ul style="list-style-type: none"> ・ AIシステムがアクチュエータ等を通じて利用者及び第三者の生命・身体・財産に危害を及ぼすことがないよう配慮する。 <p>【AI利活用ガイドライン】</p> <ul style="list-style-type: none"> ○ 安全の原則 <ul style="list-style-type: none"> ・ AIシステム又はAIサービスの利活用により、アクチュエータ等を通じて、利用者及び第三者の生命・身体・財産に危害を及ぼすことがないよう配慮する 	<ul style="list-style-type: none"> ・ AIシステムは、明示的かつ明確に定義された使用領域を持つべきであり、そのようなシステムの安全性、セキュリティ及び堅牢性は、その使用領域内の全ライフサイクルにわたってテストされ保証されなければならない。【4番】 ・ AIは安全性と技術的な堅牢性を確保して、危害を防ぎ、システムが意図したとおりに動作し、技術的に堅牢なシステム上に構築されなければならない。また、意図しない危害や予期せぬ危害のリスクを最小限に抑えなければならない。技術的な堅牢性は、システムの精度、信頼性、再現性にとっても重要である。【45番】 ・ AIの開発と活用の全ライフサイクルで潜在的なリスクを防ぎ、安全を確保するために努力しなければならない。AIを活用する過程で明らかなエラーや侵害が発生したときに、ユーザーがその動作を制御する機能を持つように努力しなければならない。【53番】

(2) 海外における原則・指針・ガイドライン等との比較からの検証

尊重すべき価値	ガイドラインにおいて関係があるものと考えられる記載（主なもの）	主な記載内容の例
12.セキュリティ	<p>【AI開発ガイドライン】</p> <ul style="list-style-type: none"> ○ セキュリティの原則 <ul style="list-style-type: none"> ・ AIシステムのセキュリティに留意する <p>【AI利活用ガイドライン】</p> <ul style="list-style-type: none"> ○ セキュリティの原則 <ul style="list-style-type: none"> ・ AIシステム又はAIサービスのセキュリティに留意する ○ 適正学習の原則 <ul style="list-style-type: none"> ➢ AIが不正確又は不適切なデータを学習することにより、AIのセキュリティに脆弱性が生じるリスクが存在することに留意することが期待される 	<ul style="list-style-type: none"> ・ AIシステムは、明示的かつ明確に定義された使用領域を持つべきであり、そのようなシステムの安全性、セキュリティ及び堅牢性は、その使用領域内の全ライフサイクルにわたってテストされ保証されなければならない。【4番】 ・ セキュリティに関して、AIシステムのライフサイクル全体にわたり、機密性、完全性、可用性を確保することが重要であり、AIシステムを開発する際には、攻撃の検知を含めたリスク評価が必要である。【39番】 ・ セキュリティの原則は、人間の物理的・情緒的な安全や環境保護に関するものであり、重要な資産の保護を含む。セキュリティの保証には、人間と機械の相互作用や攻撃・悪用に対するシステムの復元性など厳しい要件への準拠が必要である。【41番】 ・ 自律型システムは以下の安全・セキュリティの要件を満たす必要がある。【63番】 <ol style="list-style-type: none"> (1)環境とユーザーに対する安全性 (2)ハッキング等に対する信頼性と堅牢性 (3)人間と機械の相互作用に関する情緒的安全性 ・ 人間と環境及び生態系の安全とセキュリティを確保するために、AIシステムのライフサイクルを通じて、安全上のリスク及びセキュリティリスクを避けるべきである。【66番】
13.プライバシー	<p>【AI開発ガイドライン】</p> <ul style="list-style-type: none"> ○ プライバシーの原則 <ul style="list-style-type: none"> ・ AIシステムにより利用者及び第三者のプライバシーが侵害されないよう配慮する ○ 利用者支援の原則 <ul style="list-style-type: none"> ・ AIシステムが利用者を支援し、利用者を選択の機会を適切に提供することが可能となるよう配慮する <p>【AI利活用ガイドライン】</p> <ul style="list-style-type: none"> ○ プライバシーの原則 <ul style="list-style-type: none"> ・ AIシステム又はAIサービスの利活用において、他者又は自己のプライバシーが侵害されないよう配慮する 	<ul style="list-style-type: none"> ・ プライバシーとデータガバナンスの標準化に関する課題は、AIに関するデータ保護マネジメントだけでなく、全体としてどのようにデータ品質を確保するかということである。特に、機械学習のためのデータを外部プロバイダから提供される場合に課題となる。【40番】 ・ 個人情報の収集、保存、処理、使用に関する規範を策定し、それらについて個人の知る権利や選択する権利を保護する必要がある。【57番】 ・ 自国の領域内で生成された又は領域を通過するデータを規制し、プライバシー権及びその他の人権の尊重に基づき、データの効果的な規制に向けた措置を講ずる必要がある。【66番】 ・ データ保護に関する適切な法的枠組みを通じてプライバシーを保護し、有効なインフォームドコンセントを得ることが必要である。【67番】

(2) 海外における原則・指針・ガイドライン等との比較からの検証

尊重すべき価値	ガイドラインにおいて関係があるものと考えられる記載（主なもの）	主な記載内容の例
14.公平性	<p>【AI開発ガイドライン】</p> <ul style="list-style-type: none"> ○ 倫理の原則 <ul style="list-style-type: none"> ➢ 採用する技術の特性に照らし可能な範囲で、AIシステムの学習データに含まれる偏見などに起因して不当な差別が生じないよう所要の措置を講ずるよう努めることが望ましい <p>【AI利活用ガイドライン】</p> <ul style="list-style-type: none"> ○ 公平性の原則 <ul style="list-style-type: none"> ・ AIシステム又はAIサービスの判断にバイアスが含まれる可能性があることに留意し、また、AIシステム又はAIサービスの判断によって個人及び集団が不当に差別されないよう配慮する 	<ul style="list-style-type: none"> ・ 個人や集団に対して意図しない差別的影響を与える可能性を排除することが重要である。【21番】 ・ 公平さとは、適切で正確かつ一般化できるデータセットで学習・テストされたモデルであり、AIシステムは責任を持って偏見がなく実装できるように訓練されたユーザーによって運用されなければならない。【23番】 ・ AIの開発及び活用の前段階でユーザーの多様性と代表性を反映しなければならない。性別、年齢、障害、地域、人種、宗教、国家など個人の特性による偏向と差別を最小化し、商用化されたAIはすべての人に公平に適用されなければならない。社会的弱者と脆弱な層のAI技術とサービスへのアクセスを確保し、AIが与える利益は特定の集団ではなく、すべての人に均等に分配されるようにしなければならない。【53番】 ・ 意思決定を行うAIアルゴリズム、モデル及び学習用データセットは公平性を確保するべきである。【54番】
15.透明性 説明可能性	<p>【AI開発ガイドライン】</p> <ul style="list-style-type: none"> ○ 透明性の原則 <ul style="list-style-type: none"> ・ AIシステムの入出力の検証可能性及び判断結果の説明可能性に留意する <p>【AI利活用ガイドライン】</p> <ul style="list-style-type: none"> ○ 透明性の原則 <ul style="list-style-type: none"> ・ AIシステム又はAIサービスの入出力等の検証可能性及び判断結果の説明可能性に留意する ○ 適正利用の原則 <ul style="list-style-type: none"> ➢ AIサービスの公平な条件による利用を確保するとともに、必要な情報を適時に提供することが期待される 	<ul style="list-style-type: none"> ・ 各省庁はプライバシー保護、法執行、国家安全保障、その他の保護された機密情報に関して、適用される法律及び政策に従い、実行可能な範囲で議会及び一般市民を含む適切な関係者に対してAIの利用に関する関連情報を開示し、透明性を確保しなければならない。【1番】 ・ AIシステムは個々のユーザーが理解できるような説明を提供する必要がある。【2番】 ・ 透明性とは、1)どのようなデータをどのような目的で(AIに基づく意思決定を支えるために)収集するか、2)アルゴリズムを支援し意思決定を行う目的は何か、に関する開示性を指す。意思決定にAIを活用する場合は、これを意識する必要がある。透明性に関連するもう一つの重要な特徴は、AIに基づく意思決定に影響を与えたデータや意思決定の根拠を追跡できることである。【47番】 ・ 透明性・説明可能性と安全性・セキュリティ等の他の原則との間にトレードオフが存在するため、透明性・説明可能性のレベルは常に適切にすべきである。【66番】

(2) 海外における原則・指針・ガイドライン等との比較からの検証

尊重すべき価値	ガイドラインにおいて関係があるものと考えられる記載（主なもの）	主な記載内容の例
16. アカウンタビリティ	<p>【AI開発ガイドライン】</p> <ul style="list-style-type: none"> ○ アカウンタビリティの原則 <ul style="list-style-type: none"> ・ 利用者を含むステークホルダに対しアカウンタビリティを果たすよう努める ○ 利用者支援の原則 <ul style="list-style-type: none"> ・ AIシステムの学習等による出力又はプログラムの変化の可能性を踏まえ、利用者に対し適切な情報提供を行うよう努めることが望ましい <p>【AI利活用ガイドライン】</p> <ul style="list-style-type: none"> ○ アカウンタビリティの原則 <ul style="list-style-type: none"> ・ ステークホルダに対しアカウンタビリティを果たすよう努める ○ 適正利用の原則 <ul style="list-style-type: none"> ➢ AIサービスの公平な条件による利用を確保するとともに、必要な情報を適時に提供することが期待される 	<ul style="list-style-type: none"> ・ 組織内の他者、規制当局等の外部機関、AIによる意思決定の影響を受ける個人に対して、AIモデルの設計や展開時に行った行為やプロセスについて説明できる必要がある。【24番】 ・ 研究者は、システムによる決定や動作がどの程度ソースコードや学習プロセスに起因するのか、研究者が認識している限界について説明する必要がある。このような説明がない場合又は説明が遅れた場合、設計者は更なる責任を負う可能性がある。【51番】 ・ AIシステムの開発、展開及び使用に対する責任を確保するためのメカニズムを導入する必要がある。アカウンタビリティはリスク管理と密接に関連しており、第三者に説明し監査できる透明な方法でリスクを特定し軽減することが必要である。【64番】
17. 堅牢性	<p>【AI開発ガイドライン】 (記載なし)</p> <p>【AI利活用ガイドライン】 (記載なし)</p>	<ul style="list-style-type: none"> ・ 堅牢で安全なAIシステムを設計する際は、外部の脅威からシステムを防御するだけでなく、システムから発生する可能性がある悪影響から人間や環境を守ることに留意する。【41番】 ・ AIは危害を防ぎ、システムが意図したとおりに動作し、技術的に堅牢なシステム上に構築されなければならない。また、意図しない危害や予期せぬ危害のリスクを最小限に抑えなければならない。技術的な堅牢性は、システムの精度、信頼性、再現性にとっても重要である。【45番】 ・ 信頼できるAIシステムを実現するための重要な要件は、信頼性(正当に信頼できるサービスを提供する能力)と回復力(変化に直面したときの堅牢性)である。技術的な堅牢性を備えたAIシステムは、リスクに対する予防的アプローチで開発され、意図しない、予期しない危害を最小限に抑えるとともに、可能な限り防ぎながら、意図したとおりに動作することが必要である。【64番】

(2) 海外における原則・指針・ガイドライン等との比較からの検証

尊重すべき価値	ガイドラインにおいて関係があるものと考えられる記載（主なもの）	主な記載内容の例
18.責任	<p>【AI開発ガイドライン】</p> <ul style="list-style-type: none"> ○ アカウンタビリティの原則 <ul style="list-style-type: none"> ・ 利用者を含むステークホルダに対しアカウンタビリティを果たすよう努める <p>【AI利活用ガイドライン】</p> <ul style="list-style-type: none"> ○ アカウンタビリティの原則 <ul style="list-style-type: none"> ・ ステークホルダに対しアカウンタビリティを果たすよう努める 	<ul style="list-style-type: none"> ・ AIシステムは、人間の尊厳、権利、自由を守る国際条約と矛盾しないように設計されなければならない。エシックス・バイ・デザインにより合理的に予測できる誤用を避けるための措置を講ずる責任がある。【7番】 ・ 責任とは、AIに基づく意思決定が、誰の健康や安全にも脅威を与えないことを意味している。この要件は、個人の身体的・心理的健康だけでなく、データ保護やプライバシーの保護にも適用される。また、責任という要件は、意思決定が社会に見られる構造的な不平等を悪化させたり、そうでなければ個人や集団に不正や損害を与えたり、苦しみを与えたりしてはならないことを意味する。これに影響を与える方法の1つが、AIの開発者やその他の専門家の多様性を促進することである。【47番】 ・ AIシステム的设计、開発、展開に関わるすべてのステークホルダは、その行動に責任を負わなければならない。ステークホルダは、リスク及び影響評価を実施して、AIシステムがエンドユーザーに及ぼす直接的及び間接的な潜在的影響を評価し、監査プロセス(内部及び必要に応じて外部)を設定して原則の遵守を監督し、悪影響が生じた場合に苦情を解決するための仕組みを構築する必要がある。【52番】 ・ ステークホルダにはAIが適切な条件下で、適切な訓練を受けた人々によって使用されることを保証する責任がある。また、AI技術によって問題が発生した際には、アカウンタビリティを果たす必要がある。【67番】
19.追跡可能性	<p>【AI開発ガイドライン】</p> <ul style="list-style-type: none"> ○ 透明性の原則 <ul style="list-style-type: none"> ・ AIシステムの入出力の検証可能性及び判断結果の説明可能性に留意する <p>【AI利活用ガイドライン】</p> <ul style="list-style-type: none"> ○ 透明性の原則 <ul style="list-style-type: none"> ・ AIシステム又はAIサービスの入出力等の検証可能性及び判断結果の説明可能性に留意する 	<ul style="list-style-type: none"> ・ AIの設計、開発、取得、使用、特定のAIアプリケーションに関連する入力及び出力は、適切かつ実行可能な範囲において、十分に文書化され、追跡可能でなければならない。【1番】 ・ AIシステムによる意思決定は、追跡可能、説明可能、かつ透明でなければならない。これは、個人又は法人が、自分に影響を与える決定がどのようになされたかを洞察する機会を持たなければならないことを意味する。トレーサビリティは、説明可能性だけでなく、監査可能性も促進する。【45番】 ・ 透明性とは、1)どのようなデータをどのような目的で(AIに基づく意思決定を支えるために)収集するか、2)アルゴリズムを支援し意思決定を行う目的は何か、に関する開示性を指す。意思決定にAIを活用する場合は、これを意識する必要がある。透明性に関連するもう一つの重要な特徴は、AIに基づく意思決定に影響を与えたデータや意思決定の根拠を追跡できることである。【47番】

(2) 海外における原則・指針・ガイドライン等との比較からの検証

尊重すべき価値	ガイドラインにおいて関係があるものと考えられる記載（主なもの）	主な記載内容の例
<p>20. モニタリング 監査</p>	<p>【AI開発ガイドライン】</p> <ul style="list-style-type: none"> ○ 透明性の原則 <ul style="list-style-type: none"> ・ AIシステムの入出力の検証可能性及び判断結果の説明可能性に留意する ○ 制御可能性の原則 <ul style="list-style-type: none"> ・ 制御可能性を確保するため、採用する技術の特性に照らして可能な範囲において、人間や信頼できる他のAIによる監督（監視、警告など）や対処（AIシステムの停止、ネットワークからの切断、修理など）の実効性に留意することが望ましい <p>【AI利活用ガイドライン】</p> <ul style="list-style-type: none"> ○ 透明性の原則 <ul style="list-style-type: none"> ・ AIシステム又はAIサービスの入出力等の検証可能性及び判断結果の説明可能性に留意する 	<ul style="list-style-type: none"> ・ AIシステムは、完全に回答可能、監査可能であるように設計する必要があり、プロジェクト全体を通じた監視とレビューを可能にするアクティビティモニタリングを実装しなければならない。【22番】 ・ AIシステムによる意思決定は、追跡可能、説明可能、かつ透明でなければならない。これは、個人又は法人が、自分に影響を与える決定がどのようになされたかを洞察する機会を持たなければならないことを意味する。トレーサビリティは、説明可能性だけでなく、監査可能性も促進する。【45番】 ・ AIシステムとその影響についてライフサイクルを通じてアカウントビリティを果たすために、適切な監視、影響評価、デュー・ディリジェンスの仕組みが開発されなければならない。技術や制度の設計の際には、特に人権との衝突や環境・生態系への脅威に対処するために、AIシステムの監査可能性と追跡可能性を確保すべきである。【66番】
<p>21. ガバナンス</p>	<p>【AI開発ガイドライン】 (記載なし)</p> <p>【AI利活用ガイドライン】 (記載なし)</p>	<ul style="list-style-type: none"> ・ インテリジェントな技術的手段を継続的に改善し、管理メカニズムを最適化し、ガバナンスシステムを改善し、AI製品及びサービスのライフサイクル全体を通じてガバナンスを推進する必要がある。【57番】 ・ AIガバナンスの実践を通じて、AIの製品やサービスを利用する過程で明らかになった技術的な安全性や政策・規制の空白等の課題に対してタイムリーにフィードバックすることに留意する。【58番】 ・ AIの原則、政策、規制をAIの発展に合わせて、ガバナンスの改定を検討する必要がある。AIのガバナンス対策は、その適切な利用を妨げないようにするだけでなく、社会や自然にとって有益なものとなるよう、その発展状況に見合ったものでなければならない。【59番】

(2) 海外における原則・指針・ガイドライン等との比較からの検証

尊重すべき価値	ガイドラインにおいて関係があるものと考えられる記載（主なもの）	主な記載内容の例
22.その他 (コスト)	<p>【AI開発ガイドライン】 (記載なし)</p> <p>【AI利活用ガイドライン】 ○ 適正学習の原則</p> <ul style="list-style-type: none"> ・ AIによりなされる判断は、事後的に制度が損なわれたり、低下することが想定されるため、想定される権利侵害の規模、権利侵害の生じる精度、技術水準、精度を維持するためのコスト等を踏まえ、あらかじめ精度に関する基準を定めておくことが期待される 	<ul style="list-style-type: none"> ・ AIインフラの構築、運用、維持、スタッフのトレーニングや教育にどれだけの費用がかかるか、また、AI導入のための作業が潜在的な節約を上回るかどうかを検討すべきである。【22番】
22.その他 (効果測定)	<p>【AI開発ガイドライン】 (記載なし)</p> <p>【AI利活用ガイドライン】 (記載なし)</p>	<ul style="list-style-type: none"> ・ AIを使用することによる効果を理解し、測定する必要がある。【18番】

(2) 海外における原則・指針・ガイドライン等との比較からの検証

<原則・指針・ガイドライン等のリスト>

	国等	機関等	原則・指針・ガイドライン等の名称
1	米国	ホワイトハウス	Executive Order on Promoting the Use of Trustworthy Artificial Intelligence in the Federal Government URL < https://trumpwhitehouse.archives.gov/presidential-actions/executive-order-promoting-use-trustworthy-artificial-intelligence-federal-government/ >
2	米国	国立標準技術研究所	Four Principles of Explainable Artificial Intelligence URL < https://www.nist.gov/system/files/documents/2020/08/17/NIST%20Explainable%20AI%20Draft%20ONISTIR8312%20%281%29.pdf >
3	米国	国家情報長官室	Principles of Artificial Intelligence Ethics for the Intelligence Community URL < https://www.dni.gov/files/ODNI/documents/Principles_of_AI_Ethics_for_the_Intelligence_Community.pdf >
4	米国	防衛イノベーション委員会	AI Principles: Recommendations on the Ethical Use of Artificial Intelligence by the Department of Defense URL < https://media.defense.gov/2019/Oct/31/2002204458/-1/-1/0/DIB_AI_PRINCIPLES_PRIMARY_DOCUMENT.PDF >
5	米国	連邦取引委員会	Using Artificial Intelligence and Algorithms URL < https://www.ftc.gov/news-events/blogs/business-blog/2020/04/using-artificial-intelligence-algorithms >
6	米国	連邦取引委員会	Aiming for truth, fairness, and equity in your company's use of AI URL < https://www.ftc.gov/news-events/blogs/business-blog/2021/04/aiming-truth-fairness-equity-your-companys-use-ai >
7	米国	情報技術産業協議会	AI POLICY PRINCIPLES URL < https://www.itic.org/public-policy/ITIAIPolicyPrinciplesFINAL.pdf >
8	米国	ソフトウェア・情報産業協会	Ethical Principles for Artificial Intelligence and Data Analytics URL < https://history.siia.net/Portals/0/pdf/Policy/Ethical%20Principles%20for%20Artificial%20Intelligence%20and%20Data%20Analytics%20SIIA%20Issue%20Brief.pdf?ver=2017-11-06-160346-990 >

(2) 海外における原則・指針・ガイドライン等との比較からの検証

<原則・指針・ガイドライン等のリスト>

	国等	機関等	原則・指針・ガイドライン等の名称
9	米国	IBM	Everyday Ethics for Artificial Intelligence URL < https://www.ibm.com/watson/assets/duo/pdf/everydayethics.pdf >
10	米国	IBM	IBM's Principles for Trust and Transparency URL < https://www.ibm.com/blogs/policy/wp-content/uploads/2018/06/IBM_Principles_SHORT.V4.3.pdf >
11	米国	IBM	Our fundamental properties for trustworthy AI URL < https://www.ibm.com/artificial-intelligence/ai-ethics-focus-areas >
12	米国	intel	Intel's AI Privacy Policy White Paper URL < https://www.intel.com/content/dam/www/public/us/en/ai/documents/Intels-AI-Privacy-Policy-White-Paper-2018.pdf >
13	米国	Microsoft	MicrosoftのAIの基本原則 URL < https://www.microsoft.com/ja-jp/ai/responsible-ai?activetab=pivot1:primaryr6 >
14	米国	Microsoft	Responsible bots: 10 guidelines for developers of conversational AI URL < https://www.microsoft.com/en-us/research/uploads/prod/2018/11/Bot_Guidelines_Nov_2018.pdf >
15	米国	Google	Artificial Intelligence at Google: Our Principles URL < https://ai.google/principles/ >
16	米国	Meta(旧Facebook)	Facebook's five pillars of Responsible AI URL < https://ai.facebook.com/blog/facebooks-five-pillars-of-responsible-ai/ >
17	米国	BSR	Artificial Intelligence:A Rights-Based Blueprint for Business Paper 1: Why a Rights-Based Approach? URL < https://www.bsr.org/reports/BSR-Artificial-Intelligence-A-Rights-Based-Blueprint-for-Business-Paper-01.pdf >

(2) 海外における原則・指針・ガイドライン等との比較からの検証

<原則・指針・ガイドライン等のリスト>

	国等	機関等	原則・指針・ガイドライン等の名称
18	カナダ	国家財政委員会事務局	Our guiding principles URL < https://www.canada.ca/en/government/system/digital-government/digital-government-innovations/responsible-use-ai.html#toc1 >
19	カナダ	モントリオール大学	Montreal Declaration for a Responsible Development of AI URL < https://www.montrealdeclaration-responsibleai.com/the-declaration >
20	英国	貴族院AI特別委員会	Artificial Intelligence Committee AI in the UK: ready, willing and able? URL < https://publications.parliament.uk/pa/ld201719/ldselect/ldai/100/100.pdf >
21	英国	中央デジタル・データ室	Data Ethics Framework URL < https://www.gov.uk/government/publications/data-ethics-framework/data-ethics-framework-2020#overarching-principles >
22	英国	・中央デジタル・データ室 ・人口知能室	Understanding artificial intelligence ethics and safety URL < https://www.gov.uk/guidance/understanding-artificial-intelligence-ethics-and-safety >
23	英国	・中央デジタル・データ室 ・人口知能室	A guide to using artificial intelligence in the public sector URL < https://www.gov.uk/government/publications/understanding-artificial-intelligence/a-guide-to-using-artificial-intelligence-in-the-public-sector >
24	英国	情報コミッショナー事務局	Explaining decisions made with AI URL < https://ico.org.uk/for-organisations/guide-to-data-protection/key-dp-themes/explaining-decisions-made-with-artificial-intelligence/ >
25	英国	情報コミッショナー事務局	Guidance on AI and data protection URL < https://ico.org.uk/for-organisations/guide-to-data-protection/key-dp-themes/guidance-on-ai-and-data-protection/ >
26	英国	英国ファイナンス	ETHICAL PRINCIPLES FOR ADVANCED ANALYTICS AND ARTIFICIAL INTELLIGENCE IN FINANCIAL SERVICES URL < https://www.ukfinance.org.uk/system/files/AAAI-Principles-FINAL.pdf >

(2) 海外における原則・指針・ガイドライン等との比較からの検証

<原則・指針・ガイドライン等のリスト>

	国等	機関等	原則・指針・ガイドライン等の名称
27	英国	Sage Group	The Ethics of Code: Developing AI for Business with Five Core Principles URL < https://www.sage.com/~media/group/files/business-builders/business-builders-ethics-of-code.pdf?la=en >
28	英国	randstad	artificial intelligence principles URL < https://www.randstad.co.uk/s3fs-media/uk/public/2020-01/Randstad-Global-AI-principles.pdf >
29	英国	Astrazeneca	AstraZeneca's principles for ethical data and AI URL < https://www.astrazeneca.com/sustainability/ethics-and-transparency/data-and-ai-ethics.html >
30	イタリア	デジタル庁	White paper Artificial Intelligence at the service of the citizen URL < https://libro-bianco-ia.readthedocs.io/en/latest/ >
31	イタリア	Rome Call for AI ethics	Rome Call for AI ethics URL < https://www.romecall.org/wp-content/uploads/2021/02/AI-Rome-Call-x-firma_DEF_DEF_confirm_.pdf >
32	オランダ	経済・気候政策省	Strategic Action Plan for Artificial Intelligence URL < https://www.government.nl/binaries/government/documents/reports/2019/10/09/strategic-action-plan-for-artificial-intelligence/Strategic+Action+Plan+for+Artificial+Intelligence.pdf >
33	オランダ	アムステルダム市	Public AI Registers URL < https://algoritmeregister.amsterdam.nl/wp-content/uploads/White-Paper.pdf >
34	オランダ	PHILIPS	Five guiding principles for responsible use of AI in healthcare and healthy living URL < https://www.philips.com/a-w/about/news/archive/blogs/innovation-matters/2020/20200121-five-guiding-principles-for-responsible-use-of-ai-in-healthcare-and-healthy-living.html >
35	スウェーデン	デジタル政府局	Främja den offentliga förvaltningens förmåga att använda AI(行政によるAI活用の推進) URL < https://www.digg.se/4a3a73/globalassets/dokument/publicerat/publikationer/framja-den-offentliga-forvaltningens-formaga-att-anvanda-ai.pdf >

(2) 海外における原則・指針・ガイドライン等との比較からの検証

<原則・指針・ガイドライン等のリスト>

	国等	機関等	原則・指針・ガイドライン等の名称
36	スウェーデン	スウェーデン雇用者庁等	Together we are taking responsibility for a sustainable use of AI URL < https://www.almega.se/app/uploads/sites/2/2020/05/ittelekomforetagen-industry-code-for-ai-eng.pdf >
37	デンマーク	・財務省 ・産業・経済・金融省	National Strategy for Artificial Intelligence URL < https://en.digst.dk/media/19337/305755_gb_version_final-a.pdf >
38	デンマーク	データ倫理評議会	Dataetik – Sådan gør du(データ倫理 – どのように行うべきか) URL < https://dataetiskraad.dk/sites/default/files/2021-10/Dataetik%20-%20S%C3%A5dan%20g%C3%B8r%20du.pdf >
39	ドイツ	連邦政府	Artificial Intelligence Strategy of the German Federal Government URL < https://www.ki-strategie-deutschland.de/files/downloads/Fortschreibung_KI-Strategie_engl.pdf >
40	ドイツ	連邦経済・エネルギー省	GERMAN STANDARDIZATION ROADMAP ON ARTIFICIAL INTELLIGENCE URL < https://www.din.de/resource/blob/772610/e96c34dd6b12900ea75b460538805349/normungsroadmap-en-data.pdf >
41	ドイツ	データ倫理委員会	Opinion of the Data Ethics Commission URL < https://www.bmjv.de/SharedDocs/Downloads/DE/Themen/Fokusthemen/Gutachten_DEK_EN.pdf?__blob=publicationFile&v=2 >
42	ドイツ	BMW	BMW Group code of ethics for artificial intelligence URL < https://www.bmwgroup.com/content/dam/grpw/websites/bmwgroup_com/downloads/ENG_PR_CodeOfEthicsForAI_Short.pdf >
43	ドイツ	DAIMLER	Two Letters and Four Principles: How Daimler Uses Artificial Intelligence (AI) URL < https://www.daimler.com/sustainability/data/ki-guidelines.html >

(2) 海外における原則・指針・ガイドライン等との比較からの検証

<原則・指針・ガイドライン等のリスト>

	国等	機関等	原則・指針・ガイドライン等の名称
44	ドイツ	SAP	SAP's Guiding Principles for Artificial Intelligence URL < https://news.sap.com/2018/09/sap-guiding-principles-for-artificial-intelligence/ >
45	ノルウェー	地方政府・近代化省	The National Strategy for Artificial Intelligence URL < https://www.regjeringen.no/en/dokumenter/nasjonal-strategi-for-kunstig-intelligens/id2685594/?ch=7 >
46	ノルウェー	ノルウェー技術委員会	ARTIFICIAL INTELLIGENCE: OPPORTUNITIES, CHALLENGES AND A PLAN FOR NORWAY URL < https://teknologiradet.no/wp-content/uploads/sites/105/2018/11/AI-and-machine-learning-1.pdf >
47	フィンランド	経済・雇用省	Work in the age of artificial intelligence URL < https://julkaisut.valtioneuvosto.fi/bitstream/handle/10024/160980/TEMjul_21_2018_Work_in_the_age.pdf >
48	フィンランド	内閣府	Tekoäly viranomaistoiminnassa - eettiset kysymykset ja yhteiskunnallinen hyväksyttävyyys(行政における AI -倫理的問題と社会的受容性) URL < https://julkaisut.valtioneuvosto.fi/bitstream/handle/10024/161345/14-2019-Tekoaly%20viranomaistoiminnassa.pdf?sequence=1&isAllowed=y >
49	フィンランド	OP Financial Group	OP Financial Group's ethical guidelines for artificial intelligence URL < https://www.op.fi/op-financial-group/corporate-social-responsibility/commitments-and-principles >
50	フランス	情報処理と自由に関する 国家委員会	HOW CAN HUMANS KEEP THE UPPER HAND? URL < https://www.cnil.fr/sites/default/files/atoms/files/cnil_rapport_ai_gb_web.pdf >
51	フランス	ALLISTENE	Research Ethics in Machine Learning URL < http://cerna-ethics-allistene.org/digitalAssets/54/54730_cerna_2017_machine_learning.pdf >
52	インド	インド政策委員会	RESPONSIBLE AI #AIFORALL Approach Document for India Part 1 – Principles for Responsible AI URL < https://www.niti.gov.in/sites/default/files/2021-02/Responsible-AI-22022021.pdf >

(2) 海外における原則・指針・ガイドライン等との比較からの検証

<原則・指針・ガイドライン等のリスト>

	国等	機関等	原則・指針・ガイドライン等の名称
53	韓国	科学・ICT省	사람이 중심이 되는「인공지능(AI) 윤리기준」(人が中心になる「人工知能 (AI) 倫理基準」) URL < https://www.msit.go.kr/bbs/view.do?sCode=user&mPid=112&mId=113&bbsSeqNo=94&nttSeqNo=3179742 >
54	シンガポール	個人情報保護委員会	DISCUSSION PAPER ON ARTIFICIAL INTELLIGENCE (AI) AND PERSONAL DATA – FOSTERING RESPONSIBLE DEVELOPMENT AND ADOPTION OF AI URL < https://www.pdpc.gov.sg/-/media/Files/PDPC/PDF-Files/Resource-for-Organisation/AI/Discussion-Paper-on-AI-and-PD---050618.pdf >
55	シンガポール	シンガポール金融管理局	Principles to Promote Fairness, Ethics, Accountability and Transparency (FEAT) in the Use of Artificial Intelligence and Data Analytics in Singapore’s Financial Sector URL < https://www.mas.gov.sg/~media/MAS/News%20and%20Publications/Monographs%20and%20Information%20Papers/FEAT%20Principles%20Final.pdf >
56	シンガポール	シンガポールコンピュータ協会	AI ETHICS AND GOVERNANCE TOOLKIT URL < https://files-scs-prod.s3-ap-southeast-1.amazonaws.com/public/uploads/200721016+SCS+Booklet+6+Nov.pdf >
57	中国	全国新世代人工知能ガバナンス専門委員会	新一代人工智能治理原则—发展负责任的人工智能(次世代AIガバナンス原則—責任あるAIの発展) URL < http://www.most.gov.cn/kjbgz/201906/t20190617_147107.html >
58	中国	全国新世代人工知能ガバナンス専門委員会	新一代人工智能伦理规范(新時代人工知能倫理規定) URL < http://www.most.gov.cn/kjbgz/202109/t20210926_177063.html >
59	中国	北京智源人工智能研究院	Beijing AI Principles URL < https://www-pre.baai.ac.cn/news/beijing-ai-principles-en.html >
60	中国	中国人工知能産業発展連盟	人工智能行业自律公约(人工知能産業自主規制公約) URL < http://aiiaorg.cn/uploadfile/2019/0808/20190808053719487.pdf >

(2) 海外における原則・指針・ガイドライン等との比較からの検証

<原則・指針・ガイドライン等のリスト>

	国等	機関等	原則・指針・ガイドライン等の名称
61	中国	百度	AI倫理四原則(AI倫理四原則) URL < https://ai.baidu.com/ai-doc/REFERENCE/xk3dwjgfe >
62	オーストラリア	産業・科学・エネルギー資源省	Australia's Artificial Intelligence Ethics Framework URL < https://www.industry.gov.au/data-and-publications/australias-artificial-intelligence-ethics-framework/australias-ai-ethics-principles >
63	EU	欧州委員会	Artificial Intelligence, Robotics and 'Autonomous' Systems URL < https://op.europa.eu/en/publication-detail/-/publication/dfebe62e-4ce9-11e8-be1d-01aa75ed71a1/language-en/format-PDF/source-78120382 >
64	EU	AIハイレベル専門家グループ	THE ASSESSMENT LIST FOR TRUSTWORTHY ARTIFICIAL INTELLIGENCE (ALTAI) for self assessment URL < https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=68342 >
65	EU	AI4People※ ※ 企業が参画しており、 業界団体として整理	AI4People's Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations URL < https://www.eismd.eu/wp-content/uploads/2019/03/AI4People%e2%80%99s-Ethical-Framework-for-a-Good-AI-Society.pdf >
66	国際機関	国連教育科学文化機関 (UNESCO)	FIRST DRAFT OF THE RECOMMENDATION ON THE ETHICS OF ARTIFICIAL INTELLIGENCE URL < https://unesdoc.unesco.org/ark:/48223/pf0000373434?posInSet=6&queryId=ed3e3a0d-e278-46a2-898f-3e24d43f75cc >
67	国際機関	世界保健機関 (WHO)	ETHICS AND GOVERNANCE OF ARTIFICIAL INTELLIGENCE FOR HEALTH URL < https://apps.who.int/iris/rest/bitstreams/1352854/retrieve >

(3) 国内における原則・指針・ガイドライン等との比較からの検証

○ 国内の事業者、団体において策定された指針等を踏まえて、現行のAI開発ガイドライン及びAI利活用ガイドラインの射程、内容等について検証。

■ 調査対象の指針等

- ・ 国内の事業者、団体等が公表している22個の指針等を調査（36～37ページ参照）

■ 尊重すべき価値

- ・ 調査を通じて、22の「尊重すべき価値」を確認※7

- | | | |
|-------------------|-------------------|---------------|
| 1. 人間中心 | 9. 適正な学習（学習データの質） | 17. 堅牢性 |
| 2. 人間の尊厳 | 10. AI間の連携 | 18. 責任 |
| 3. 多様性、包摂 | 11. 安全性 | 19. 追跡可能性 |
| 4. 持続可能な社会 | 12. セキュリティ | 20. モニタリング、監査 |
| 5. 国際協力 | 13. プライバシー | 21. ガバナンス |
| 6. 適正な利用 | 14. 公平性 | 22. コスト |
| 7. 教育、リテラシー | 15. 透明性、説明可能性 | |
| 8. 人間の判断の介在、制御可能性 | 16. アカウンタビリティ | |

※7 1.～16.までは、AIガイドライン比較表（「報告書2019」別紙2）における項目と同じ項目。17.～22.が新たに確認した項目。

なお、22.について、海外における指針等との比較においては、「効果測定」があったため「22.その他（コスト、効果測定）」としているが、国内における指針等との比較においては、「効果測定」は確認されなかったため「22.コスト」としている。

【注】 総務省「AIに係る原則・ガイドライン等の策定状況、規定事項等に関する調査研究」による。網羅的なものではなく、当該調査に基づいて整理したものである。

(3) 国内における原則・指針・ガイドライン等との比較からの検証

- ・ 株式会社ABEJA
 - ・ 株式会社エヌ・ティ・ティ・データ
 - ・ 沖電気工業株式会社
 - ・ クウジツ株式会社
 - ・ KDDI株式会社
 - ・ コニカミノルタ株式会社
 - ・ 株式会社J.Score
 - ・ 株式会社スタジオム
 - ・ ソニーグループ株式会社
 - ・ 日本電気株式会社
 - ・ 日本ユニシス株式会社
 - ・ 株式会社野村総合研究所
 - ・ 株式会社日立製作所
 - ・ 富士通株式会社
 - ・ 富士フイルム株式会社
 - ・ 株式会社三菱総合研究所
 - ・ 三菱電機株式会社
 - ・ 株式会社リクルート（リクルートワークス研究所）
 - ・ 一般社団法人人工知能学会
 - ・ 一般社団法人日本経済団体連合会
 - ・ 一般社団法人ピープルアナリティクス&HRテクノロジー協会
 - ・ AIクラウドサービス検討会※
- ※ 当該検討会での検討等を総合的に踏まえて、総務省において公表。

(注1) 社名は、公表当時のもの。

1. 人間中心	5
2. 人間の尊厳	17
3. 多様性、包摂	5
4. 持続可能な社会	9
5. 国際協力	1
6. 適正な利用	17
7. 教育、リテラシー	14
8. 人間の判断の介在、制御可能性	6
9. 適正な学習（学習データの質）	7
10. AI間の連携	4
11. 安全性	14

12. セキュリティ	19
13. プライバシー	21
14. 公平性	21
15. 透明性、説明可能性	20
16. アカウンタビリティ	20
17. 堅牢性	3
18. 責任	3
19. 追跡可能性	4
20. モニタリング、監査	3
21. ガバナンス	2
22. コスト	1

(注2) 半数を超える指針等に記載がある場合、赤字で表示。

(注3) 事業者、団体等が属する業種・分野やAI利活用の目的・用途等により、尊重すべき価値のうち、どの項目を重視することが期待されるのかは異なることに留意が必要である。

(3) 国内における原則・指針・ガイドライン等との比較からの検証

尊重すべき価値	ガイドラインにおいて関係があるものと考えられる記載（主なもの）	主な記載内容の例
1.人間中心	<p>【AI開発ガイドライン】</p> <ul style="list-style-type: none"> ○ 基本理念 <ul style="list-style-type: none"> ・ 人間の尊厳と個人の自律が尊重される人間中心の社会を実現すること <p>【AI利活用ガイドライン】</p> <ul style="list-style-type: none"> ○ 基本理念 <ul style="list-style-type: none"> ・ 人間の尊厳と個人の自律が尊重される人間中心の社会を実現すること 	<ul style="list-style-type: none"> ・ 自由、公平、公正の理念にのっとり、人間中心の視点でAIの社会実装を行うとともに、AIが想定通りに機能するよう検証に努めます。【13番】 ・ Human Centricの理念のもと、あらゆる人の可能性を広げ、それぞれが望む形で幸福を追求したり社会に貢献したりすることを支援するツールとしてAIを位置づけます。プライバシーを守り、人が差別を受けたり人が危害を受けたりすることのないよう公平や安全を配慮して実装することにより、人が安心して利用できるAIを目指します。【14番】 ・ AI等の利活用による恩恵が多くの人に享受され、人間の尊厳と個人の自律が尊重される人間中心の社会の実現を目指します。【16番】
2.人間の尊厳	<p>【AI開発ガイドライン】</p> <ul style="list-style-type: none"> ○ 倫理の原則 <ul style="list-style-type: none"> ・ 人間の尊厳と個人の自律を尊重する <p>【AI利活用ガイドライン】</p> <ul style="list-style-type: none"> ○ 尊厳・自律の原則 <ul style="list-style-type: none"> ・ 人間の尊厳と個人の自律を尊重する 	<ul style="list-style-type: none"> ・ 人間の尊厳を尊重するためには、人間がAIに過度に依存することを避けなければなりません。人間とAIの適切な役割分担を考慮したAIサービスを開発し、提供するよう努めます。【8番】 ・ 人間の尊厳と個人の自律を尊重し、学習に用いるデータの質に留意するとともに、ステークホルダーに対し説明責任を果たすよう配慮します。【11番】 ・ 人間が常に安心して利活用できるAIの実現を目指して設計、開発を行います。AIの設計、開発および運用のライフサイクル全般にわたり、セキュリティ対策の実装、強化を行うとともに、人間の尊厳を脅かすような悪用を防ぐように努めます。また、万が一、予期せぬ事態が発生した際には、適切に対処するように努めます。【12番】

【注】「主な記載内容の例」に記載されている番号は、36～37ページに掲載されている原則・指針・ガイドライン等のリストの番号を示している。

(3) 国内における原則・指針・ガイドライン等との比較からの検証

尊重すべき価値	ガイドラインにおいて関係があるものと考えられる記載（主なもの）	主な記載内容の例
3.多様性包摂	<p>【AI開発ガイドライン】</p> <ul style="list-style-type: none"> ○基本理念 <ul style="list-style-type: none"> ・ 人間がAIネットワークと共生することにより、その恵沢がすべての人によってあまねく享受され、人間の尊厳と個人の自律が尊重される人間中心の社会を実現すること ○ 利用者支援の原則 <ul style="list-style-type: none"> ➢ ユニバーサルデザインなど社会的弱者の利用を容易にするための取組に努める ○ 関係するステークホルダに期待される役割 <ul style="list-style-type: none"> ・ 本ガイドラインの運用や見直しにおいて各国政府、国際機関、開発者、市民社会を含む利用者など多様なステークホルダ間の対話の促進に向けた環境整備に努めることが期待される ・ AIをめぐる議論の多様性を確保しつつ、AIの便益の増進及びリスクの抑制について、認識の共有を図り、相互に協力するよう努めることが期待される <p>【AI利活用ガイドライン】</p> <ul style="list-style-type: none"> ○ 基本理念 <ul style="list-style-type: none"> ・ 人間がAIネットワークと共生することにより、その恵沢がすべての人によってあまねく享受され、人間の尊厳と個人の自律が尊重される人間中心の社会を実現すること ・ AIの利活用において利用者の多様性を尊重し、多様な背景と価値観、考え方を持つ人々を包摂すること 	<ul style="list-style-type: none"> ・ AIの活用において、不当な差別を起こさないよう、お客様およびステークホルダーの人権および多様性の尊重に努めます。【4番】 ・ 安全かつ堅牢で使いやすい製品・サービスの実現に取り組む中で、より多くの人がAIによる恩恵を享受できるよう、AIの利活用を積極的に進めていきます。特に、これまで培ってきた先進・独自の技術との融合や社外との共創の推進により、新たな価値をスピーディーに生み出していくことで、ヘルスケアをはじめとした幅広い領域で社会課題の解決をリードすることを目指します。【15番】 ・ 多様な人々の多様なライフスタイルの実現をサポートするAI技術は人類の未来のため、多様な人々の多様な夢を実現させるためにあるとの理念のもと、積極的な開発と展開を行う。【20番】
4.持続可能な社会	<p>【AI開発ガイドライン】 (記載なし)</p> <p>【AI利活用ガイドライン】</p> <ul style="list-style-type: none"> ○ 基本理念 <ul style="list-style-type: none"> ・ 個人、地域社会、各国、国際社会が抱える様々な課題の解決を図り、持続可能な社会を実現すること 	<ul style="list-style-type: none"> ・ AIの力を、地球規模の問題解決への貢献や平和で持続可能な社会の発展のために活用するよう努めます。【9番】 ・ 社会の多様な課題を解決し、快適で強靱な持続可能社会の実現や、世界中の人々のQoL向上のためにAIの開発と利活用を計画します。【13番】 ・ AI等の利活用によりさまざまな社会課題の解決を図り、持続可能な未来社会の実現を目指します。【16番】

(3) 国内における原則・指針・ガイドライン等との比較からの検証

尊重すべき価値	ガイドラインにおいて関係があるものと考えられる記載（主なもの）	主な記載内容の例
5.国際協力	<p>【AI開発ガイドライン】</p> <ul style="list-style-type: none"> ○ 基本理念 <ul style="list-style-type: none"> ・ AIシステムの研究開発の在り方について、非拘束的なソフトローたる指針やそのベストプラクティスをステークホルダ間で国際的に共有すること ○ 関係するステークホルダに期待される役割 <ul style="list-style-type: none"> ・ 各国政府及び国際機関は、本ガイドラインの運用や見直しにおいて各国政府、国際機関、開発者、市民社会を含む利用者など多様なステークホルダ間の対話の促進に向けた環境整備に努めることが期待される <p>【AI利活用ガイドライン】</p> <ul style="list-style-type: none"> ○ 基本理念 <ul style="list-style-type: none"> ・ AIの利活用の在り方について、非拘束的なソフトローたる指針やベストプラクティスを国際的に共有すること 	<ul style="list-style-type: none"> ・ AIを取り巻く環境・価値観や、パートナーシップの拡大などビジネスのあり方が変化していることを認識しています。AIの利活用の推進にあたっては、国内外の顧客やパートナーなど多様なステークホルダーと対話し相互に理解を深めるとともに、グローバル企業として各国/地域の法律や行政が発する規制等を遵守し、公平かつ調和のとれたAIの利活用を追求します。【6番】
6.適正な利用	<p>【AI開発ガイドライン】</p> <ul style="list-style-type: none"> ○ 制御可能性の原則 <ul style="list-style-type: none"> ➢ 制御可能性を確保するため、採用する技術の特性に照らして可能な範囲において、人間や信頼できる他のAIによる監督（監視、警告など）や対処（AIシステムの停止、ネットワークからの切断、修理など）の実効性に留意することが望ましい <p>【AI利活用ガイドライン】</p> <ul style="list-style-type: none"> ○ 適正利用の原則 <ul style="list-style-type: none"> ・ 人間とAIシステムとの間及び利用者間における適切な役割分担のもと、適正な範囲及び方法でAIシステム又はAIサービスを利用するよう努める 	<ul style="list-style-type: none"> ・ 私たちの製品・サービスを提供する際には、人権を尊重した適正な用途で利用されるよう努めます。【10番】 ・ AIシステムを利用したビジネスを提供するにあたり、適正な範囲及び方法で人との役割分担を行うよう努めます。【11番】 ・ AIが設計された用途や動作条件から逸脱した使い方をされないよう、AIの具体的な利活用シーン(ユースケース)におけるリスクポテンシャルを踏まえた開発を行い、AIの適正な利活用に努めます。また、利用者や運用者等に対し利用方針、利用条件等を示すことでAIの適正な運用に努めます。【13番】

(3) 国内における原則・指針・ガイドライン等との比較からの検証

尊重すべき価値	ガイドラインにおいて関係があるものと考えられる記載（主なもの）	主な記載内容の例
7.教育 リテラシー	<p>【AI開発ガイドライン】 (記載なし)</p> <p>【AI利活用ガイドライン】</p> <ul style="list-style-type: none"> ○ 適正利用の原則(参考) <ul style="list-style-type: none"> ➢ 消費者的利用者は、AIの判断に対し、消費者的利用者が最終判断をすることが適当とされている場合には、適切に判断ができるよう必要な能力及び知識を習得しておくことが望ましい。 ➢ アクチュエータ等を通じて稼働するAIの利活用において、一定の条件に該当することにより人間による稼働に移行することが予定されている場合には、消費者的利用者は、移行前、移行中、移行後等の各状態における責任の所在を予め認識しておくことが望ましい。また、AIサービスプロバイダから、移行条件、移行方法等についての説明を受け、必要な能力及び知識を習得しておくことが望ましい。 	<ul style="list-style-type: none"> ・ 業界団体又は公的な研究会などに積極的に参加することや最新の学術的成果の調査などを積極的に行うことで、このような課題に対する最新の知見を獲得しています。【1番】 ・ AIの技術発展・社会に対する影響・課題に関する議論について情報収集に努めます。その上で、AIの高度化・複雑化への対応と不当利用の未然防止に向け、役職員に対するAIの知識及び倫理に関する教育の拡充に努めます。【7番】 ・ AIを活用した商品・サービスが社会に与える影響を認識した上で、より良い社会を実現するAIの発展に貢献し、AIの活用や研究開発を通じて明るい未来を形創ることができる人材の育成に積極的に取り組んでいきます。【9番】
8.人間の判断の介入 制御可能性	<p>【AI開発ガイドライン】</p> <ul style="list-style-type: none"> ○ 制御可能性の原則 <ul style="list-style-type: none"> ・ AIシステムの制御可能性に留意する ➢ 制御可能性を確保するため、採用する技術の特性に照らして可能な範囲において、人間や信頼できる他のAIによる監督（監視、警告など）や対処（AIシステムの停止、ネットワークからの切断、修理など）の実効性に留意することが望ましい <p>【AI利活用ガイドライン】</p> <ul style="list-style-type: none"> ○ 適正利用の原則 <ul style="list-style-type: none"> ➢ AIによりなされた判断について、必要かつ可能な場合には、その判断を用いるか否か、あるいは、どのように用いるか等に関し、人間の判断を介入させることが期待される。その場合、人間の判断の介入の要否について、基準例を踏まえ、利用する分野やその用途等に応じて検討することが期待される ○ 公平性の原則 <ul style="list-style-type: none"> ➢ AIによりなされた判断結果の公平性を保つため、AIを利活用する際の社会的文脈や人々の合理的な期待を踏まえ、その判断を用いるか否か、あるいは、どのように用いるか等に関し、人間の判断を介入させることが期待される 	<ul style="list-style-type: none"> ・ AI開発者は、AIシステム等の制御可能性に留意する。 AI利用者は、AIによりなされた判断について、必要かつ可能な場合には、その判断を用いるか否か、あるいは、どのように用いるか等に関し、人間の判断を介入させることが期待される。【5番】 ・ AIが導き出した提案や結果について、人が吟味して意思決定することが重要であると考えます。AIによる自動化を推進するにあたって、システム全体として、提案や結果の根拠を的確に示す仕組みを提供することを目指します。【14番】 ・ 機械学習はあくまでも道具にすぎず、その使い方を定めるのは人間です。機械学習は人類社会の繁栄に大きく貢献できる可能性を秘めているとともに、不適切な利用をすれば人類社会の利益に反する可能性もあります。機械学習は過去の事例に基づいて未来を予測しますから、偏りのある過去に基づいて予測する未来は、やはり偏りのあるものになりかねません。もし、過去と異なる「あるべき未来」を求めるとすれば、機械学習による予測や判断が公平性を欠くことがないように人間が機械学習に注意深く介入する必要があります。【19番】

(3) 国内における原則・指針・ガイドライン等との比較からの検証

尊重すべき価値	ガイドラインにおいて関係があるものと考えられる記載（主なもの）	主な記載内容の例
9.適正な学習 （学習データの質）	<p>【AI開発ガイドライン】</p> <ul style="list-style-type: none"> ○ 倫理の原則 <ul style="list-style-type: none"> ➢ 採用する技術の特性に照らし可能な範囲で、AIシステムの学習データに含まれる偏見などに起因して不当な差別が生じないよう所要の措置を講ずるよう努めることが望ましい <p>【AI利活用ガイドライン】</p> <ul style="list-style-type: none"> ○ 適正学習の原則 <ul style="list-style-type: none"> ・ AIシステムの学習等に用いるデータの質に留意する 	<ul style="list-style-type: none"> ・ 人間の尊厳と個人の自律を尊重し、学習に用いるデータの質に留意するとともに、ステークホルダーに対し説明責任を果たすよう配慮します。【11番】 ・ AI等に用いるデータの質・量に留意するとともに、AI等を適切に評価するよう努めます。また、AI等を構成するシステムやネットワークの相互連携に伴うリスク・セキュリティに留意します。【16番】 ・ 事業者が人事データに対しプロファイリング等の処理を実施する場合、元データ及び処理結果双方の正確性及び最新性が確保されるように努めなければならない。【21番】
10.AI間の連携	<p>【AI開発ガイドライン】</p> <ul style="list-style-type: none"> ○ 連携の原則 <ul style="list-style-type: none"> ・ AIシステムの相互接続性と相互運用性に留意する <p>【AI利活用ガイドライン】</p> <ul style="list-style-type: none"> ○ 連携の原則 <ul style="list-style-type: none"> ・ AIシステム又はAIサービス相互間の連携に留意する 	<ul style="list-style-type: none"> ・ 人間中心のより良い社会に向けて、人とAI、およびAI同士の協調が課題であると考えており、これらの実現についての継続的な検討を行います。【3番】 ・ AI開発者、AI利用者、並びにAIデータ提供者は、AIシステム等間の連携、及びAIシステム等がネットワーク化することによってリスクが惹起・増幅される可能性があることに留意する。【5番】 ・ AI等を構成するシステムやネットワークの相互連携に伴うリスク・セキュリティに留意します。【16番】 ・ オープンソースは、突然の修正やバージョンアップが行われることがある。互換性のない変更もある。オープンソースが変更になった場合のシステムの変更手順を決めておく必要がある。【22番】
11.安全性	<p>【AI開発ガイドライン】</p> <ul style="list-style-type: none"> ○ 安全の原則 <ul style="list-style-type: none"> ・ AIシステムがアクチュエータ等を通じて利用者及び第三者の生命・身体・財産に危害を及ぼすことがないよう配慮する。 <p>【AI利活用ガイドライン】</p> <ul style="list-style-type: none"> ○ 安全の原則 <ul style="list-style-type: none"> ・ AIシステム又はAIサービスの利活用により、アクチュエータ等を通じて、利用者及び第三者の生命・身体・財産に危害を及ぼすことがないよう配慮する 	<ul style="list-style-type: none"> ・ 私たちは、AIの利活用において、意図しないAIの作用により社会・人に悪影響を及ぼすことのないよう、安全性の確保を優先し、リスクに配慮した適用と運用に努めます。【6番】 ・ 人間が根源的に持つ権利を始めとして、人間の生命・身体・財産に危害を及ぼすことがないよう努めます。【8番】 ・ AIに依存し、AIによって人間の尊厳や能力、可能性が限定・否定される、あるいは身体や生命が危険にさらされることのないよう配慮するとともに、そうした危険性につながる意図的・非意図的なAIの悪用・誤用がなされることのないよう、基本的な人権を尊重した製品・サービスの開発・提供に取り組みます。【15番】

(3) 国内における原則・指針・ガイドライン等との比較からの検証

尊重すべき価値	ガイドラインにおいて関係があるものと考えられる記載（主なもの）	主な記載内容の例
12. セキュリティ	<p>【AI開発ガイドライン】</p> <ul style="list-style-type: none"> ○ セキュリティの原則 <ul style="list-style-type: none"> ・ AIシステムのセキュリティに留意する <p>【AI利活用ガイドライン】</p> <ul style="list-style-type: none"> ○ セキュリティの原則 <ul style="list-style-type: none"> ・ AIシステム又はAIサービスのセキュリティに留意する ○ 適正学習の原則 <ul style="list-style-type: none"> ➢ AIが不正確又は不適切なデータを学習することにより、AIのセキュリティに脆弱性が生じるリスクが存在することに留意することが期待される 	<ul style="list-style-type: none"> ・ データについては、個人情報保護やプライバシーの尊重と併せて、その適切な取得、利用、管理、セキュリティの確保を徹底し、お客様その他のステークホルダーの皆様には不当な損害が生じないよう努めます。【3番】 ・ 私たちが取扱うデータの重要性・機微性を鑑み、AIサービスにおけるセキュリティの充実を図ります。【8番】 ・ AIを搭載した製品・サービスが想定どおり動作するように品質を検証し、生命・自由を脅かすことがないよう安全性の確保に努めます。また、第三者による不正なアクセスからの保護などのセキュリティ対策にも継続的に取り組みます。【17番】
13. プライバシー	<p>【AI開発ガイドライン】</p> <ul style="list-style-type: none"> ○ プライバシーの原則 <ul style="list-style-type: none"> ・ AIシステムにより利用者及び第三者のプライバシーが侵害されないよう配慮する ○ 利用者支援の原則 <ul style="list-style-type: none"> ・ AIシステムが利用者を支援し、利用者を選択の機会を適切に提供することが可能となるよう配慮する <p>【AI利活用ガイドライン】</p> <ul style="list-style-type: none"> ○ プライバシーの原則 <ul style="list-style-type: none"> ・ AIシステム又はAIサービスの利活用において、他者又は自己のプライバシーが侵害されないよう配慮する 	<ul style="list-style-type: none"> ・ AIを活用した製品・サービスにおいて、法令および関連する社内規則に従います。お客様の意思を尊重した個人情報の保護・取り扱いを行い、プライバシーの保護に努めます。【4番】 ・ AI開発者、AI利用者、並びにAIデータ提供者は、AIシステム等の利活用において、他者又は自己のプライバシーが侵害されないよう配慮する。【5番】 ・ 人事は、個人情報などのデータを収集するに先立って、データを提供する個人に対して、データ収集の目的、収集データの種類、個人の持つ権利などに関する説明を実施し、それに対する同意を得る。人事は、この同意が、データを提供する個人の理解と納得を伴う“実質的な同意”になるよう最善の努力をする責任を負う必要がある。個人は、同意を撤回する権利を有する。【18番】

(3) 国内における原則・指針・ガイドライン等との比較からの検証

尊重すべき価値	ガイドラインにおいて関係があるものと考えられる記載（主なもの）	主な記載内容の例
14. 公平性	<p>【AI開発ガイドライン】</p> <ul style="list-style-type: none"> ○ 倫理の原則 <ul style="list-style-type: none"> ➢ 採用する技術の特性に照らし可能な範囲で、AIシステムの学習データに含まれる偏見などに起因して不当な差別が生じないよう所要の措置を講ずるよう努めることが望ましい <p>【AI利活用ガイドライン】</p> <ul style="list-style-type: none"> ○ 公平性の原則 <ul style="list-style-type: none"> ・ AIシステム又はAIサービスの判断にバイアスが含まれる可能性があることに留意し、また、AIシステム又はAIサービスの判断によって個人及び集団が不当に差別されないよう配慮する 	<ul style="list-style-type: none"> ・ AI開発者、AI利用者、並びにAIデータ提供者は、AIシステム等の判断にバイアスが含まれる可能性があることに留意し、また、AIシステム等の判断によって個人が不当に差別されないよう配慮する。【5番】 ・ AIの判断結果が、多様なステークホルダーの利益に資するとともに、人種、性別、国籍などによる差別や偏見を発生させたり助長させたりすることがないようAIの実現、運用に努めます。【13番】 ・ 人事は、AIを活用するにあたって、人々がその人種、性別、国籍、年齢、政治的信念、宗教などの多様なバックグラウンドを理由に不当な差別をされないよう、その設計思想における公平性を常に検証する。【18番】
15. 透明性 説明可能性	<p>【AI開発ガイドライン】</p> <ul style="list-style-type: none"> ○ 透明性の原則 <ul style="list-style-type: none"> ・ AIシステムの入出力の検証可能性及び判断結果の説明可能性に留意する <p>【AI利活用ガイドライン】</p> <ul style="list-style-type: none"> ○ 透明性の原則 <ul style="list-style-type: none"> ・ AIシステム又はAIサービスの入出力等の検証可能性及び判断結果の説明可能性に留意する ○ 適正利用の原則 <ul style="list-style-type: none"> ➢ AIサービスの公平な条件による利用を確保するとともに、必要な情報を適時に提供することが期待される 	<ul style="list-style-type: none"> ・ AI事業において、対象となるAI商品などの性質や利活用の場面などに応じ、想定する利用目的・方法、利活用による効果・影響、および限界について、お客様その他のステークホルダーの皆様に説明してご理解を得るように努めます。特に、個々のAI商品などにおけるAIの判断結果の透明性に配慮し、AIの判断結果がどのようなものであるかについての理解に資する情報を提供するよう努めます。【3番】 ・ 製品・サービスにおけるAIによる判断の理由が説明可能となる仕組みを、製品・サービスの企画・設計段階から導入する可能性を追求し、お客様が理解しやすい説明や情報を提供するよう努めます。【4番】 ・ AIサービスの入出力および判断結果の妥当性等を検証し、透明性のある説明ができるよう努めます。【16番】

(3) 国内における原則・指針・ガイドライン等との比較からの検証

尊重すべき価値	ガイドラインにおいて関係があるものと考えられる記載（主なもの）	主な記載内容の例
16. アカウンタビリティ	<p>【AI開発ガイドライン】</p> <ul style="list-style-type: none"> ○ アカウンタビリティの原則 <ul style="list-style-type: none"> ・ 利用者を含むステークホルダに対しアカウンタビリティを果たすよう努める ○ 利用者支援の原則 <ul style="list-style-type: none"> ・ AIシステムの学習等による出力又はプログラムの変化の可能性を踏まえ、利用者に対し適切な情報提供を行うよう努めることが望ましい <p>【AI利活用ガイドライン】</p> <ul style="list-style-type: none"> ○ アカウンタビリティの原則 <ul style="list-style-type: none"> ・ ステークホルダに対しアカウンタビリティを果たすよう努める ○ 適正利用の原則 <ul style="list-style-type: none"> ➢ AIサービスの公平な条件による利用を確保するとともに、必要な情報を適時に提供することが期待される 	<ul style="list-style-type: none"> ・ 私たちは、AIによる判断プロセスに不透明性が内包されていることを認識し、その利活用における透明性を高め、社会に対して誠実かつオープンに説明責任を果たすことに努めます。【6番】 ・ AIの利活用による効果・価値・影響について、適切な説明を行い、全てステークホルダーから理解を得られるよう努めます。【10番】 ・ AIの利用目的や利用方法、AIを搭載した製品・サービスの使用時に想定される様々な影響に関して、多様なステークホルダーとの対話を重ねて説明責任を果たすよう努めます。【17番】
17. 堅牢性	<p>【AI開発ガイドライン】 (記載なし)</p> <p>【AI利活用ガイドライン】 (記載なし)</p>	<ul style="list-style-type: none"> ・ 安全かつ堅牢で使いやすい製品・サービスの実現に取り組む中で、より多くの人々がAIによる恩恵を享受できるよう、AIの利活用を積極的に進めていきます。【15番】 ・ 公平性、アカウンタビリティ、透明性等の確保による信頼性や、プライバシー、セキュリティ、ディペンダビリティを確保した高品質AIの技術開発と、その運用基準・体制を確立する必要がある。【20番】 ・ 学習に使用していないデータに対してどの程度正確な答えが出せるかという「汎化性能」や、データにノイズが含まれていても正確な答えが出せるかという「頑健性」についても評価を行う必要がある。【22番】

(3) 国内における原則・指針・ガイドライン等との比較からの検証

尊重すべき価値	ガイドラインにおいて関係があるものと考えられる記載（主なもの）	主な記載内容の例
18.責任	<p>【AI開発ガイドライン】</p> <ul style="list-style-type: none"> ○ アカウンタビリティの原則 <ul style="list-style-type: none"> ・ 利用者を含むステークホルダに対しアカウンタビリティを果たすよう努める <p>【AI利活用ガイドライン】</p> <ul style="list-style-type: none"> ○ アカウンタビリティの原則 <ul style="list-style-type: none"> ・ ステークホルダに対しアカウンタビリティを果たすよう努める 	<ul style="list-style-type: none"> ・ 人事は、個人情報の収集とその保管、AIシステム自体と結果データのセキュリティに対して道義的責任と技術的責任を負う。【18番】 ・ 人事データを取り扱う際、グローバルに多極的に変化する情勢を的確に把握し、適法かつ適正な個人の権利利益保護と利活用のバランスを実現する見地から、ピープル・アナリティクスを専門に行う部署設立及び全社的な人事データ保護の観点に責任を持つデータプロテクションオフィサー等の役職者の選任により責任の所在を明確にするなどの組織体制を確立する。【21番】 ・ 収集したデータに不具合があり、損害が発生した場合の責任分担についても意識する必要がある。 <ol style="list-style-type: none"> ① データをデータ提供事業者から購入する場合は、購買契約における責任分担の規程に留意すること。 通常はデータ提供事業者の免責事項になっていることが多い。 ② インターネット等でオープンなデータを収集する場合は、データの不具合についての責任も自社になることに留意すること。【22番】
19.追跡可能性	<p>【AI開発ガイドライン】</p> <ul style="list-style-type: none"> ○ 透明性の原則 <ul style="list-style-type: none"> ・ AIシステムの入出力の検証可能性及び判断結果の説明可能性に留意する <p>【AI利活用ガイドライン】</p> <ul style="list-style-type: none"> ○ 透明性の原則 <ul style="list-style-type: none"> ・ AIシステム又はAIサービスの入出力等の検証可能性及び判断結果の説明可能性に留意する 	<ul style="list-style-type: none"> ・ 個人情報を適切に取得、利用、提供できる仕組み作り、データ・トレーサビリティの確保、不正アクセスの防止に努めることでデータの悪用を防ぎ、お客さまやユーザーに安心安全を届けます。【2番】 ・ AI開発者並びにAI利用者は、AIシステム等の入出力等の検証可能性及び判断結果の説明可能性に留意する。【5番】 ・ 社会の基盤としてあらゆる人に関するものになっている情報通信技術が担う社会的責任の重さを自覚して、AIについては、望ましくない、または不測の結果をもたらさないよう品質向上に努めるとともに、必要に応じて原因を究明できる仕組みを実装して、安心・安全な社会システムを目指します。【14番】

(3) 国内における原則・指針・ガイドライン等との比較からの検証

尊重すべき価値	ガイドラインにおいて関係があるものと考えられる記載（主なもの）	主な記載内容の例
20. モニタリング 監査	<p>【AI開発ガイドライン】</p> <ul style="list-style-type: none"> ○ 透明性の原則 <ul style="list-style-type: none"> ・ AIシステムの入出力の検証可能性及び判断結果の説明可能性に留意する ○ 制御可能性の原則 <ul style="list-style-type: none"> ・ 制御可能性を確保するため、採用する技術の特性に照らして可能な範囲において、人間や信頼できる他のAIによる監督（監視、警告など）や対処（AIシステムの停止、ネットワークからの切断、修理など）の実効性に留意することが望ましい <p>【AI利活用ガイドライン】</p> <ul style="list-style-type: none"> ○ 透明性の原則 <ul style="list-style-type: none"> ・ AIシステム又はAIサービスの入出力等の検証可能性及び判断結果の説明可能性に留意する 	<ul style="list-style-type: none"> ・ AIを活用したすべてのサービスの構築・提供において、公平性を担保すべく、AI活用の結果について社内外の識者によるモニタリングを継続していきます。【7番】 ・ 人事は、AIを利用するにあたって、起こり得る問題を想起し、技術的・非技術的対応策を事前、ならびに事後的に講じる能力を高め、継続的に点検を実施する【18番】 ・ AIクラウドサービスでは、運用フェーズにおいて、入力データの変化や追加学習により出力結果が変化することもあり、事業者側でも運用状況をモニタリングする必要がある。【22番】
21. ガバナンス	<p>【AI開発ガイドライン】 (記載なし)</p> <p>【AI利活用ガイドライン】 (記載なし)</p>	<ul style="list-style-type: none"> ・ 経営層自らが上記の課題に対する知見と意識を身につけるとともに、上記の課題への対応のための全社的なマネジメントを行っています。【1番】 ・ 外部の識者等で構成される委員会を設置し、その客観的な評価を取締役会と共有することで、AI倫理に関するコーポレート・ガバナンスの充実を図ります。【14番】
22. コスト	<p>【AI開発ガイドライン】 (記載なし)</p> <p>【AI利活用ガイドライン】</p> <ul style="list-style-type: none"> ○ 適正学習の原則 <ul style="list-style-type: none"> ・ AIによりなされる判断は、事後的に制度が損なわれたり、低下することが想定されるため、想定される権利侵害の規模、権利侵害の生じる精度、技術水準、精度を維持するためのコスト等を踏まえ、あらかじめ精度に関する基準を定めておくことが期待される 	<ul style="list-style-type: none"> ・ AIの開発においては、学習に使用するデータが極めて重要であり、収集方法や収集範囲等について早くから検討する必要がある。また、データの収集や加工に多くの稼働やコストを要するため、スケジュールの作成においてはデータの整備に十分な時間を見込んでおく必要がある。【22番】

(3) 国内における原則・指針・ガイドライン等との比較からの検証

<原則・指針・ガイドライン等のリスト>

	事業者、団体	ガイドラインの名称
1	株式会社ABEJA	AIポリシー URL < https://abejainc.com/ja/contact/ai_policy/ >
2	株式会社エヌ・ティ・ティ・データ	NTTデータグループ AI指針 URL < https://www.nttdata.com/jp/ja-/media/nttdatajapan/files/news/release/2019/nttdata_ai_guidelines.pdf >
3	沖電気工業株式会社	OKIグループ AI原則 URL < https://www.oki.com/jp/press/2019/09/z19033.pdf >
4	クウジツ株式会社	クウジツAI倫理ガイドライン URL < https://www.koozyt.com/aiethicsguidelines >
5	KDDI株式会社	KDDIグループ AI開発・利活用原則 URL < https://www.kddi.com/corporate/kddi/public/ai_principles/ >
6	コニカミノルタ株式会社	コニカミノルタグループ AIの利活用に関する基本方針 URL < https://www.konicaminolta.jp/about/csr/pdf/use-of-ai-basic-policy.pdf >
7	株式会社J.Score	AI活用ポリシー URL < https://www.jscore.co.jp/ai_policy/ >
8	株式会社スタジアム	AIに関する基本姿勢 URL < https://stadium.co.jp/policy/ai >
9	ソニーグループ株式会社	ソニーグループAI倫理ガイドライン URL < https://www.sony.com/ja/SonyInfo/csr_report/humanrights/AI_Engagement_within_Sony_Group_Ja.pdf >
10	日本電気株式会社	NECグループ AIと人権に関するポリシー URL < https://jpn.nec.com/press/201904/images/0201-01-01.pdf >
11	日本ユニシス株式会社 [※]	BIPROGYグループのAI倫理指針（公表当時：日本ユニシスグループのAI倫理指針） ※ 公表当時の社名。当社は、2022年4月1日に「BIPROGY株式会社」と社名を変更。 URL < https://www.biprogy.com/com/ai_ethics_principles_nihonunisys_group.pdf >

(3) 国内における原則・指針・ガイドライン等との比較からの検証

	事業者、団体	ガイドラインの名称
12	株式会社野村総合研究所	NRIグループ AI倫理ガイドライン URL < https://www.nri.com/jp/sustainability/social/policies >
13	株式会社日立製作所	社会イノベーション事業にAIを活用するためのAI倫理原則 URL < https://www.hitachi.co.jp/products/it/lumada/about/ai/ldsl/document/ai_document_jp.pdf >
14	富士通株式会社	富士通グループ AIコミットメント URL < https://www.fujitsu.com/jp/documents/about/csr/humanrights/fujitsu-group-ai-commitment-201903_ja.pdf >
15	富士フイルム株式会社	富士フイルムグループ AI基本方針 URL < https://holdings.fujifilm.com/ja/sustainability/vision/policy/ai >
16	株式会社三菱総合研究所	三菱総合研究所 AI事業推進の指針 URL < https://www.mri.co.jp/sustainability/governance/dia6ou000001oqqk-att/ai-guideline20191023.pdf >
17	三菱電機株式会社	三菱電機グループ「AI倫理ポリシー」 URL < https://www.mitsubishielectric.co.jp/corporate/sustainability/social/humanrights/aipolicy/index.html >
18	株式会社リクルート リクルートワークス研究所※	人事のためのAI原則 “暫定版” ※ 機関紙「Works」156号の特集記事の中で、Works編集部が草案を作成し、有識者等による議論を踏まえ発表されたもの。 URL < https://www.works-i.com/works/no156/ >
19	一般社団法人人工知能学会	機会学習と公平性に関する声明 URL < http://ai-elsi.org/archives/888 >
20	一般社団法人 日本経済団体連合会	AI活用原則 URL < https://www.keidanren.or.jp/policy/2019/013_honbun.pdf >
21	一般社団法人ピープルアナリティクス &HRテクノロジー協会	人事データ利活用原則 URL < https://peopleanalytics.or.jp/media/HRDataUtilizationPrinciples.pdf >
22	AIクラウドサービス検討会※	AIを用いたクラウドサービスに関するガイドブック ※当該検討会での検討等を総合的に踏まえて、総務省において公表。 URL < https://www.soumu.go.jp/main_content/000792669.pdf >

(4) ヒアリング等において示された取組事例との対応からの検証

○ 主として、「報告書2021」（2021年8月）及び「AIガバナンスに関する取組事例」（2021年9月）に示されている各事業者の取組等を踏まえて※8、AI開発ガイドライン及びAI利活用ガイドラインの射程、内容等について検証。

直接的に両ガイドラインの目的、基本理念、原則等と関係が深いものと考えられる「多様性」、「セキュリティ」、「プライバシー」、「公平性」、「透明性・アカウントビリティ」、「適正利用」及び「品質管理」に関して整理。

項目・事項	ガイドラインにおいて関係があるものと考えられる記載（主なもの）	具体的な取組事例等
多様性	<p>【AI開発ガイドライン】</p> <ul style="list-style-type: none"> ○ 倫理の原則 <ul style="list-style-type: none"> ・ 人間の尊厳と個人の自律を尊重する <p>【AI利活用ガイドライン】</p> <ul style="list-style-type: none"> ○ 基本理念 <ul style="list-style-type: none"> ・ AIの利活用において利用者の多様性を尊重し、多様な背景と価値観、考え方を持つ人々を包摂すること ○ 尊厳・自律の原則 <ul style="list-style-type: none"> ・ 人間の尊厳と個人の自律を尊重する ○ 公平性の原則 <ul style="list-style-type: none"> ・ 個人及び集団が不当に差別されないよう配慮する 	<ul style="list-style-type: none"> ・ 人材の多様性を重視しており、インプットする人間が非常にダイバース（多層的）であることに重点を置いている。このダイバースには、男女も含むものであるが、異なる文化を背景に持っている学者を集めているところに特徴がある。 ・ 研究所には、コンピューターサイエンスの研究者だけでなく、心理学やフィールドワーク、認知科学に取り組んでいる社会科学系の研究者がいる。心理学や認知科学の視点を入れることが重要である。 ・ 安心・安全で信頼性のあるAIの研究開発、社会実装には、様々なバックグラウンドを持った者が必要であり、デザインや社会心理学をやっていた研究者なども含めてAI倫理専門チームを構成している。
セキュリティ	<p>【AI開発ガイドライン】</p> <ul style="list-style-type: none"> ○ セキュリティの原則 <ul style="list-style-type: none"> ・ AIシステムのセキュリティに留意する <p>【AI利活用ガイドライン】</p> <ul style="list-style-type: none"> ○ セキュリティの原則 <ul style="list-style-type: none"> ・ AIシステム又はAIサービスのセキュリティに留意する ○ 適正学習の原則 <ul style="list-style-type: none"> ➢ AIが不正確又は不適切なデータを学習することにより、AIのセキュリティに脆弱性が生じるリスクが存在することに留意することが期待される 	<ul style="list-style-type: none"> ・ AIセキュリティについては、AIの普及により、AIに対する攻撃（AIをだます、AIの情報を盗むなど）、AIを使った脅威（AIでだますなど）が出てきている。 ・ AIとセキュリティの関係を考えて場合、(a)Attack using AI（AIを利用した攻撃）、(b)Attack by AI（AI自身による攻撃）、(c)Attack to AI（AIへの攻撃）、(d)Measure using AI（AIを利用したセキュリティ対策）の4つの視点が存在する。(c) AIに対する攻撃だけでなく、AIに関連するそれ以外の論点も考慮しながら議論を進めることが重要と考えられる。

※8 一部「報告書2020」において取りまとめた取組等が含まれている。

(4) ヒアリング等において示された取組事例との対応からの検証

項目・事項	ガイドラインにおいて関係があるものと考えられる記載（主なもの）	具体的な取組事例等
プライバシー	<p>【AI開発ガイドライン】</p> <ul style="list-style-type: none"> ○ プライバシーの原則 <ul style="list-style-type: none"> ・ AIシステムにより利用者及び第三者のプライバシーが侵害されないよう配慮する ○ 利用者支援の原則 <ul style="list-style-type: none"> ・ AIシステムが利用者を支援し、利用者へ選択の機会を適切に提供することが可能となるよう配慮する ○ アカウンタビリティの原則 <ul style="list-style-type: none"> ・ 利用者を含むステークホルダに対しアカウンタビリティを果たすよう努める <p>【AI利活用ガイドライン】</p> <ul style="list-style-type: none"> ○ プライバシーの原則 <ul style="list-style-type: none"> ・ AIシステム又はAIサービスの利活用において、他者又は自己のプライバシーが侵害されないよう配慮する ○ アカウンタビリティの原則 <ul style="list-style-type: none"> ・ ステークホルダに対しアカウンタビリティを果たすよう努める 	<ul style="list-style-type: none"> ・ 空港におけるサービスについては、オプトインの形になっており、顔情報の提供に同意が得られた場合のみ使うこととしている。 ・ ブロックチェーンを活用して、データを追跡できるようにしておき、学生自身が開示する相手先（企業）や情報の範囲を選べるようになっており、企業は情報を勝手に引き出すことはできず、個人情報保護される仕組みになっている。 ・ データの利活用に当たっては、オプトイン型で行っており、本人の同意を得た上でサービスを提供している。市民に納得してもらった上でAIを活用したサービスを使ってもらうことが重要である。
公平性	<p>【AI開発ガイドライン】</p> <ul style="list-style-type: none"> ○ 倫理の原則 <ul style="list-style-type: none"> ・ 人間の尊厳と個人の自律を尊重する <p>【AI利活用ガイドライン】</p> <ul style="list-style-type: none"> ○ 公平性の原則 <ul style="list-style-type: none"> ・ AIシステム又はAIサービスの判断にバイアスが含まれる可能性があることに留意し、また、AIシステム又はAIサービスの判断によって個人及び集団が不当に差別されないよう配慮する 	<ul style="list-style-type: none"> ・ センシティブ属性に対するバイアスの確認をどのように行うか、そもそもセンシティブ属性が何であるかを考えるところから始まるが、その判断は難しい。 ・ 多様なステークホルダへの配慮について、例えば、障害者や子どもなどステークホルダの中にも含まれるマイノリティであるような人達への配慮について考えることをチェックするようにしている。 ・ AIに関わる設計者、運用者、利用者等のステークホルダーを集めたワークショップを開催し、公平性の観点から重視すべき価値観を抽出・数値化し、それらをAIの開発に反映している。 ・ 日本、米国、欧州において価値観が異なるような多元的な公平性、性別や人種のように様々な属性が重なることにより非線形に差異が発生する交差バイアスを抽出し、AIの精度とのバランスを考慮しつつバイアスを軽減する処置を行っている。 ・ 顔認証のAIについて、人種的な差別の観点から問題となることがあるため、公共の場では利活用しないこととし、その旨を公表している。

(4) ヒアリング等において示された取組事例との対応からの検証

項目・事項	ガイドラインにおいて関係があるものと考えられる記載（主なもの）	具体的な取組事例等
透明性・ アカウンタビリティ	<p>【AI開発ガイドライン】</p> <ul style="list-style-type: none"> ○ 透明性の原則 <ul style="list-style-type: none"> ・ AIシステムの入出力の検証可能性及び判断結果の説明可能性に留意する ○ アカウンタビリティの原則 <ul style="list-style-type: none"> ・ 利用者を含むステークホルダに対しアカウンタビリティを果たすよう努める <p>【AI利活用ガイドライン】</p> <ul style="list-style-type: none"> ○ 透明性の原則 <ul style="list-style-type: none"> ・ AIシステム又はAIサービスの入出力等の検証可能性及び判断結果の説明可能性に留意する ○ アカウンタビリティの原則 <ul style="list-style-type: none"> ・ ステークホルダに対しアカウンタビリティを果たすよう努める ○ 適正利用の原則 <ul style="list-style-type: none"> ➢ 必要な情報を適時に提供することが期待される 	<ul style="list-style-type: none"> ・ 医療の例では、これまでの医療論文や知識データベースからナレッジグラフを作り、医学的な根拠、薬学的な根拠を作ることができるかどうか、過去の医療論文や知識データベースをつなぎ合わせて入力から出力まで辿り着けるか、このパスを作ることが説明可能なAIとなる。 ・ 工場のオペレーション効率を徹底的に上げたいという場合に、中身は分からなくても結果としてオペレーション効率が上がれば、このAIはよかったね、という形で使えるものと認識している。 ・ 需要予測の場合、このような条件だから、このような需要が伸びるといったことを提示するAIである。 ・ 人事や教育においては、ホワイトAIを使うべきであるというのがグローバルな倫理規定的にも十分になってきている。高度なAIを使ってしまうと、なぜ、このような人事判断が行われたのか、このような教育的な評価がなされたのかということの事後的な説明責任が果たせなくなる、アカウンタビリティがなくなってしまう。 ・ データの利活用に当たっては、オプトイン型で行っており、本人の同意を得た上でサービスを提供している。市民に納得してもらった上でAIを活用したサービスを使ってもらうことが重要である。 ・ 住宅ローン審査へのAIの適用については、公正な判断が求められるので、判断根拠の説明が可能なXAI（Explainable AI）を用いている。
適正利用 < 次頁につづく >	<p>【AI開発ガイドライン】</p> <ul style="list-style-type: none"> ○ 制御可能性の原則 <ul style="list-style-type: none"> ➢ 制御可能性を確保するため、採用する技術の特性に照らして可能な範囲において、人間や信頼できる他のAIによる監督（監視、警告など）や対処（AIシステムの停止、ネットワークからの切断、修理など）の実効性に留意することが望ましい <p>【AI利活用ガイドライン】</p> <ul style="list-style-type: none"> ○ 適正利用の原則 <ul style="list-style-type: none"> ・ 人間とAIシステムとの間及び利用者間における適切な役割分担のもと、適正な範囲及び方法でAIシステム又はAIサービスを利用するよう努める 	<ul style="list-style-type: none"> ・ 特に教育・人事領域において、完全にAIに任せるのは危険であり、あくまでもAIは補助ツールであって、人間が最終的に決めるという形が望ましい。 ・ エントリーシートの前処理など定型的なところを整理して一案を作ることや主観の入らない客観的な評価を必要とする部分はAIに任せて、それを参考にしながら人間が評価するといった役割分担が望ましい形ではないかと考えられる。 ・ 契約書のレビューについて、AIで全部できる、AIに任せておけば安心というわけではなく、最終判断は人間が行う。 ・ AI原稿自動要約について、まだシステムの実験段階であるが、最初にシステムが自動的に要約を作成し、その結果を人間がチェックして、修正を加えたものをウェブサイトに掲載したり、SNSに配信したりする運用を想定している。AIの要約の結果をそのまま配信したり、オンエアすることは考えておらず、最終的なチェックは人間がやることになっており、どのAIについても、そのようなポリシーとして運用している。

(4) ヒアリング等において示された取組事例との対応からの検証

項目・事項	ガイドラインにおいて関係があるものと考えられる記載（主なもの）	具体的な取組事例等
適正利用 〈つづき〉	<p>➢ AIによりなされた判断について、必要かつ可能な場合には、その判断を用いるか否か、あるいは、どのように用いるか等に関し、人間の判断を介在させることが期待される。その場合、人間の判断の介在の要否について、基準例を踏まえ、利用する分野やその用途等に応じて検討することが期待される</p>	<ul style="list-style-type: none"> 自動原稿作成システムについて、（中略）システムが作成した原稿案をもとに、最終的には人間が、加筆や削除などの修正の必要性を判断して原稿を作成している。 音声認識による字幕作成において、認識誤りについては、最終的に人間が修正して放送するやり方となっている。 画像診断AIは、特定の条件下においては、専門医を上回る検出率となることがある。しかし、AIに判断を任せるのではなく、AIは医師のサポート役と位置付けている。
品質管理 （ライフサイクル）	<p>【AI利活用ガイドライン】</p> <ul style="list-style-type: none"> ○ 適正利用の原則 <ul style="list-style-type: none"> ➢ 利活用の過程を通じて、AIの機能を向上させ、リスクを抑制するため、AIソフトのアップデート 及びAIの点検・修理等を提供することが期待される ○ 適正学習の原則 <ul style="list-style-type: none"> ➢ AIによりなされる判断は、事後的に精度が損われたり、低下することが想定されるため、想定される権利侵害の規模、権利侵害の生じる頻度、技術水準、精度を維持するためのコスト等を踏まえ、あらかじめ精度に関する基準を定めておくことが期待される。精度が当該基準を下回った場合には、データの質に留意して改めて学習させることが期待される 	<ul style="list-style-type: none"> AIの運用では、サービス提供後に、データの特質が変わるなど、今まで使えていたものが使えなくなるが発生する（中略）。どのように運用していくのが大きな課題である。 品質管理の上で、ライフサイクルは非常に重要である。（中略）ライフサイクルとしての品質管理をどのようにやっていくかは非常に大きな研究テーマである。 設計が行われる前にAI倫理の確認を行うことによって、AI Ethics by Designを目指すという形でいくつかのアセスメントを行っている。 リリースした後も、市場でのインシデントをモニタリングしてチェックしている。 アプリケーションを一度作って終わりではなく、何度も新しい要件を取り込みながら成長させていくサイクルができてきている。このサイクルの中で、アプリケーションが新しくなれば、データも新しく生み出されてくる。新しく生み出されるデータを使って、AIモデルを継続的に成長させていくことが必要であり（後略） AIシステムを実際に稼働させた後も、開発・運用支援基盤を提供することで、AIモデルの挙動管理を行っている。 AIの動作を監視する開発・運用支援基盤等を活用して、AIの精度劣化をモニタリングし、自動検知や自動修復することができる仕組みを導入している（又は導入に向けた開発を行っている）。 AIシステムの運用保守においては、AIの精度が必ず維持されることの保証を行うことが難しいため、AIの管理・監視・更新などの手順に関する取り決めを利用先と合意し実施している。 AIシステムが顧客のビジネスに及ぼす影響をリスクベースで整理した上で、リスクに応じた対応策を決めている。早急に対応する必要があるもの、時間をかけてモデルの見直しをすればよいものなど案件ごとにリスクの大きさなどを顧客と合意し、モニタリングの頻度や手段を検討している。

(5) AI開発・利活用及び社会実装に関する事例との対応からの検証

○ 国内外において、AI開発・利活用及び社会実装が進められている中、リスクが顕在化したり、インシデントが発生するなど課題が指摘される事例が生じている。このような事例を踏まえて、AI開発ガイドライン及びAI利活用ガイドラインの射程、内容等について検証。

■ 調査対象ドキュメント

- ・ 国内外の政府機関等が公表している14個のドキュメントを調査（45ページ参照）
 - AI開発・利活用及び社会実装の事例について、「生体認証」、「自動運転」、「画像識別」、「医療」、「犯罪予測」、「再犯予測」、「採用」、「入学選抜」、「成績・人事評価」、「広告」及び「新型コロナウイルス感染症対応」に分類し、公平性（バイアス）、プライバシー、セキュリティ等の観点から課題が指摘された事例を踏まえて整理。
 - 多くの事例において、AI開発ガイドラインやAI利活用ガイドラインに掲げられている価値（原則）でカバーされており、現時点では、両ガイドラインの射程を大きく超えてリスクが顕在化するおそれは少ないものと考えられる。なお、将来的なAI開発・利活用及び社会実装の動向を見据えつつ、リスクを抑制する（顕在化させない）方策等を検討することは重要である。

AI開発・利活用及び社会実装に関する事例から抽出された課題等（概要）

	AI開発・利活用及び社会実装に関する事例から抽出された課題等（概要）
生体認証	<ul style="list-style-type: none"> ・ 米国において、黒人男性が白人警察官に首を押さえつけられて死亡した事件を巡り、大手IT事業者（数社）は警察に対して顔認識技術を提供しないことを決定した。また、人権に基づく国内法が整備されるまで顔認識技術を用いたサービスを提供しないことを表明した社もある。顔認識を含む生体認証技術に関する法制度等について検討する必要がある。【3番、4番、5番、9番】 ・ マサチューセッツ工科大学の科学者が、大手IT企業（数社）の顔認識システムに人種やジェンダーのバイアスがあることを指摘しており、公平性を確保する必要がある。【12番】 ・ 中国のフェイスアプリ事業者のプライバシーポリシーにおいて、ユーザーの顔データが無償で提供され、その肖像権が永続的に当該事業者へ帰属すること等が記載されており、プライバシーを確保する必要がある。【11番】 ・ 米国のAI事業者が3Dマスクを活用して、中国の大手IT事業者の顔認証システムをパスしており、セキュリティを確保する必要がある。【11番】 ・ プリントアウトした紙を帽子に張り付けることにより、Face ID（顔認証）システムを誤認させることができるため、堅牢性を確保する必要がある。【7番】
自動運転	<ul style="list-style-type: none"> ・ 自動運転車による死亡事故が発生しており、安全性を確保するとともに、AIシステムの開発者、利用者間で責任の所在を明らかにする必要がある。【10番、11番】 ・ 道路標識にステッカーを張ることにより、標識を誤認識させることができるため、堅牢性を確保する必要がある。【7番】

【注】「AI開発・利活用及び社会実装に関する事例から抽出された課題等（概要）」に記載されている番号は、45ページに掲載されているドキュメントのリストの番号を示している。

【注】総務省「AIに係る原則・ガイドライン等の策定状況、規定事項等に関する調査研究」による。網羅的なものではなく、当該調査に基づいて整理したものである。

(5) AI開発・利活用及び社会実装に関する事例との対応からの検証

	AI開発・利活用及び社会実装に関する事例から抽出された課題等（概要）
画像識別	・ 画像にノイズを付与することにより、パンダをミニチュアプードル等に誤認識させることができるため、セキュリティや堅牢性を確保する必要がある。【7番】
医療	・ 患者の病状に合わせて治療プログラムを選択するAIが、白人よりも黒人の方が病気のリスクが低いと判断する傾向にあり、適切なプログラムを選択できていなかった。また、写真から医師か看護師かを判断するAIが、男性は医師、女性は看護師と判断する傾向にあり、データやアルゴリズムの公平性を確保する必要がある。【1番、3番、13番】
犯罪予測	・ 犯罪予測システムは、被害者が報告するデータを活用しており、被害者からの報告がない傾向にある場合は犯罪予測を誤る可能性がある。被害者からの報告があるか否かにかかわらず公平に扱われることを確保し、犯罪から守られるべきである。【8番】
再犯予測	・ 再犯予測システムが、白人よりも黒人の方が再犯リスクが高いと予測する傾向にあり、データやアルゴリズムの公平性を確保する必要がある。【5番、8番、11番】
採用	・ 米国の大手IT事業者が採用に活用していたAIが、女性に対して差別的な評価を行っており、データやアルゴリズムの公平性や透明性を確保する必要がある。【3番、5番、8番、10番、12番、14番】
入学選抜	・ 仏国の大学の入学先を選別するシステムが、ブラックボックス化されているため、公平な選別が行われているかどうか判断できず、透明性や公平性を確保する必要がある。【8番】
成績・人事評価	・ 英国の資格・試験統制機関であるOfqualの成績予測評価システムが、マイノリティに属する学生等に不利な評価を行うことが判明し、公平性を確保する必要がある。【5番】 ・ 米国テキサス州において、教師の業績を評価するシステムを導入したものの、そのシステムの仕組みがブラックボックス化されており、透明性や説明可能性を確保する必要がある。【11番】
広告	・ 米国の大手IT事業者が収集するユーザーの個人情報、ユーザーの同意なく英国の事業者提供され、政治広告のために活用されたとされており、プライバシーを確保する必要がある。【2番、8番、13番】 ・ 米国の大手IT事業者は、広告主が特定の人種、年齢、性別に属するユーザーを住宅、雇用等の広告配信対象から除外できるようにしており、公平性を確保する必要がある。【6番】
新型コロナウイルス感染症対応	・ 接触確認・追跡アプリ※ ⁹ を活用してデータを収集・分析する場合は、データの利用方法等を開示して、ユーザーから同意を得る必要がある。また、人種や所得水準等によらず、接触確認・追跡の効果を楽しむようにする必要がある。【3番、14番】

※9 このような機能を有するアプリについて、AIシステムを用いていない場合もあるが、収集したデータをもとに分析・判断を行う利活用の形態であり、検証のための重要なモデルケースとなるものと考えられるため、AIシステムを直接的に用いていないものも含めて対象することとする。なお、日本において(COCOA)は、追跡の機能を有していないなど、アプリにより機能や仕様などが異なっていることに留意が必要である。

(5) AI開発・利活用及び社会実装に関する事例との対応からの検証

	人間中心	人間の尊厳	多様性・包摂	持続可能な社会	国際協力	適正な利用	リテラシー 教育・	人間の判断の介入・ 制御可能性	適正な学習 (学習)データの質	AI間の連携	安全性	セキュリティ	プライバシー	公平性	説明可能性・ 透明性	アカウントビリティ	堅牢性	責任	追跡可能性	モニタリング・ 監査	ガバナンス	コスト	
生体認証	★	★	★			★			★			★	★	★									
自動運転	★	★				★				★	★	★			★			★					
画像識別												★					★						
医療	★	★	★			★			★				★	★	★	★							
犯罪予測	★	★	★			★			★					★	★	★							
再犯予測	★	★	★			★			★				★	★	★	★							
採用	★	★	★			★			★				★	★	★	★							
入学選抜	★	★	★			★			★					★	★	★							
成績・人事評価	★	★	★			★			★				★	★	★	★							
広告	★	★	★			★						★	★	★									
新型コロナウイルス感染症対応	★	★	★			★			★			★	★	★									

(注) 記載があるものはピンク色に着色。

(5) AI開発・利活用及び社会実装に関する事例との対応からの検証

	国等	機関等	ドキュメントの名称
1	日本	情報処理推進機構	・ DX白書 2021 URL < https://www.ipa.go.jp/files/000093706.pdf >
2	日本	情報処理推進機構	・ AI白書 2020 ※書籍を参照したためURLはない
3	米国	AI国家安全保障委員会	・ The Final Report URL < https://www.nscai.gov/wp-content/uploads/2021/03/Full-Report-Digital-1.pdf >
4	米国	スタンフォード大学 人間中心のAI研究所	・ 2021 AI Index Report URL < https://aiindex.stanford.edu/report/ >
5	英国	データ倫理・イノベーション センター	・ Review into bias in algorithmic decision-making URL < https://www.gov.uk/government/publications/cdei-publishes-review-into-bias-in-algorithmic-decision-making/main-report-cdei-review-into-bias-in-algorithmic-decision-making#preface >
6	英国	データ倫理・イノベーション センター	・ Online targeting: Final report and recommendations URL < https://www.gov.uk/government/publications/cdei-review-of-online-targeting/online-targeting-final-report-and-recommendations >
7	ドイツ	連邦政府情報セキュリティ局	・ Towards Auditable AI Systems URL < https://www.bsi.bund.de/SharedDocs/Downloads/EN/BSI/KI/Towards_Auditable_AI_Systems.pdf?__blob=publicationFile&v=4 >
8	フランス	情報処理と自由に関する 国家委員会	・ HOW CAN HUMANS KEEP THE UPPER HAND? URL < https://www.cnil.fr/sites/default/files/atoms/files/cnil_rapport_ai_gb_web.pdf >
9	EU	欧州委員会	・ Proposal for a REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL URL < https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:52021PC0206&from=EN >
10	EU	欧州委員会	White Paper On Artificial Intelligence URL < https://ec.europa.eu/info/sites/default/files/commission-white-paper-artificial-intelligence-feb2020_en.pdf >
11	中国	信息通信研究院	・ White Paper on Trustworthy Artificial Intelligence URL < http://www.caict.ac.cn/english/research/whitepapers/202110/P020211014399666967457.pdf >
12	国際機関	国際連合教育科学文化機関 (UNESCO)	・ ARTIFICIAL INTELLIGENCE and GENDER EQUALITY URL < https://unesdoc.unesco.org/ark:/48223/pf0000374174 >
13	国際機関	世界保健機関 (WHO)	・ ETHICS AND GOVERNANCE OF ARTIFICIAL INTELLIGENCE FOR HEALTH URL < https://apps.who.int/iris/rest/bitstreams/1352854/retrieve >
14	国際機関	世界経済フォーラム (WEF)	・ The AI Governance Journey: Development and Opportunities URL < https://www3.weforum.org/docs/WEF_The%20AI_Governance_Journey_Development_and_Opportunities_2021.pdf >

(6) レビューの結果

それぞれの観点から検証した結果、次のことを確認することができた。

海外における原則・指針・ガイドライン等との比較からの検証

- ・ 海外における指針等において、「堅牢性」、「責任」、「追跡可能性」、「モニタリング、監査」といった新しい「尊重すべき価値」が盛り込まれているものがあること
- ・ 「透明性・説明可能性」、「公平性」、「アカウントビリティ」、「人間の尊厳」といったものは、いずれの地域においても重視されていること
- ・ 既存の「尊重すべき価値」であっても、詳細については、両ガイドラインと海外における指針等に記載されている内容等に差異※10が見られるものがあること

国内における原則・指針・ガイドライン等との比較からの検証

- ・ 国内における指針等において、「堅牢性」、「責任」、「追跡可能性」、「モニタリング、監査」といった新しい「尊重すべき価値」が盛り込まれているものがあること
- ・ 両ガイドラインと国内における指針等において、概ね同じような概念で用いられているものが多いこと
- ・ 「プライバシー」、「公平性」、「透明性・説明可能性」、「アカウントビリティ」、「セキュリティ」といったものは、ほぼすべての事業者・団体等において重視されていること

ヒアリング等において示された取組事例との対応からの検証

- ・ 多くの事業者等が、両ガイドラインの各原則等に対応する取組を行っていること
- ・ これらの取組に加えて、両ガイドラインの射程を超えた「安心・安全で信頼性のあるA Iの社会実装」に関する優れた取組を行っていること

AI開発・利活用及び社会実装に関する事例との対応からの検証

- ・ リスクが顕在化したり、インシデントが発生した多くの事例において、どのような価値を重視すべきかという対象については、概ね両ガイドラインに掲げられている価値（原則）でカバーされており、それらの価値をどのように実現するのかといった実効性の問題であること
 - ・ 現時点においては、両ガイドラインの射程を大きく超えてリスクが顕在化したり、インシデントが発生するおそれは大きくないものと考えられること
- なお、現行のガイドラインにおいて掲げられている価値（原則）の実効性を確保すること及び将来的なAI開発・利活用及び社会実装の動向を見据えつつ、リスクを抑制する（顕在化させない）方策等を検討することは重要である。

※10 ガイドラインの見直しに関しては、これらの差異も考慮しつつ、検討を行うこととする。

AI開発ガイドライン及びAI利活用ガイドラインの見直しに関する論点①

- ここまでのレビューや本推進会議における意見等を踏まえて、AI開発ガイドライン及びAI利活用ガイドラインをAIの開発者や利用者にとって、より有用・有益であるものにして、「安心・安全で信頼性のあるAIの社会実装」を進めるために、ガイドラインの見直しに関する検討を行うこととし、見直しに関する論点を整理する。

ガイドラインの見直しに関する検討に当たっては、レビュー、本推進会議における意見、国内外の動向・国際的な議論の動向等を踏まえて、4つの論点を整理。

分野共通・分野横断的なガバナンス、分野別・分野ごとのガバナンス、標準化によるルール形成など様々な取組を組み合わせ、全体として、ガイドラインにおける各原則の実効性を確保することが重要であるということを踏まえて検討を行う。

⇒ 今後、整理した論点や本推進会議における意見等とともに、次に掲げる観点を踏まえて、ガイドラインの見直しに関する検討を進めることとする。

(この2つの観点は、必ずしも相反するものではなく、両立し得るものである。)

- ・ 現行の両ガイドラインの在り方を前提として、目的、基本理念、原則等を追加したり、変更したりするかどうか。
- ・ 現行の両ガイドラインの在り方を前提とせず、ガイドラインの在り方自体を検討の対象としたり、ガイドライン単体ではなく各原則の実効性を確保するための方策と一体的なものとして考えるかどうか。

ガイドラインの見直しに関する論点

- 目的、基本理念、対象とするAIの範囲（定義）、原則、解説等について、追加や修正等をするかどうか。また、先進的な事例等を踏まえて、新しい内容を盛り込むなど構成等を見直すかどうか。

例えば、

- AI開発ガイドラインの基本理念に、『多様性』や『持続可能』といった概念を追加するかどうか。
- 新しい原則として、『堅牢性』、『責任』、『追跡可能性』、『モニタリング、監査』といった項目を追加するかどうか。なお、これまで両ガイドラインを策定する過程において、原則数は可能な限り少なくすべきといった議論との整合をどのように考えるか。
- 既に定められている各原則の解説において、具体的な開発・利活用の場面やユースケース等を示すかどうか。
- 有事（パンデミック、災害等）を想定した見直しを行うかどうか（例えば、公衆衛生の向上・増進とプライバシー保護のバランスをどのように考えるか。）
- このほか、品質確保・マネジメント、サプライチェーン、組織・体制、人材育成、データの取扱い、行政機関のAI利活用におけるコストとのバランスなどの観点から、留意することが期待される事項を整理し、ガイドラインとして取りまとめるかどうか。 等

【注】 見直しに関する検討に当たっては、技術的中立性を確保し、イノベーションを阻害せず、開発者や利用者にとって過度に負担とならないものにするといった点について、引き続き、留意することが重要である。

AI開発ガイドライン及びAI利活用ガイドラインの見直しに関する論点②

ガイドラインの見直しに関する論点

- 各ガイドラインの位置付け、名称等を変更するかどうか。

例えば、

- 国際的な議論のためのもの、国内の事業者等の取組を支援するもの（指針等の策定の際の参考）というように国内外を問わずガイドラインの活用を図っていくことを明確化するかどうか。その場合、AI開発ガイドライン（『国際的な議論のためのAI開発ガイドライン案』）の名称を変更するかどうか。 等

- AIの開発のハードルが低くなってきていることや利活用がより身近になってきている（利用者がAIが実装されていることを認識しない（できない）ケースが増えてきている）こと、あるいは、AIの開発と利活用の境界が曖昧になってきていることなどを踏まえて、どのような取組を進めることが望ましいか。

例えば、

- これまで想定していたよりも専門的でない開発者やAIの利活用を認識していない利用者（特に消費者的利用者）など向けに対応することが望ましいことはあるか（平易な解説を加えることなど）。
- AI開発ガイドラインとAI利活用ガイドラインを統合するかどうか。 等

- 今後、ガイドラインの実効性を確保するために、ガイドラインを含めたガバナンス全体の枠組みにおいて、どのような取組を進めることが望ましいか。

例えば、

- 各原則等の実効性を確保するための方策として、事業者、業界等による取組や政府（行政機関）を含めた取組、あるいは、チェックシートの策定、認定制度などが考えられるところ、このような方策を含めたガバナンス全体の枠組みについて、どのような取組を進めることが望ましいか。

（参考）AIを用いたクラウドサービスの安全・信頼性に係る情報開示指針（ASP・SaaS編）

ASP・SaaS（AIクラウドサービス）の安全・信頼性に係る情報開示認定制度

- EUのAI規制法案が公表され、審議が進められてきていることなどを踏まえて、国際的なガバナンスの枠組みとの関係において、どのような対応が求められるか。 等

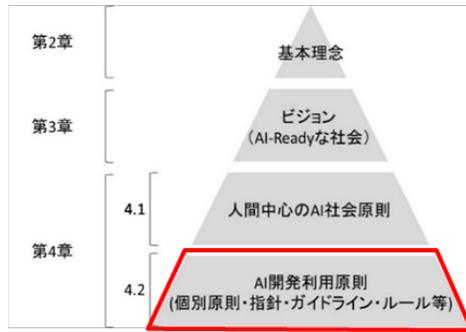
【注】 見直しに関する検討に当たっては、技術的中立性を確保し、イノベーションを阻害せず、開発者や利用者にとって過度に負担とならないものにするといった点について、引き続き、留意することが重要である。

(参考) 人間中心のAI社会原則

人間中心のAI社会原則

- 人間中心の原則
- 教育・リテラシーの原則
- プライバシー確保の原則
- セキュリティ確保の原則
- 公正競争確保の原則
- 公平性、説明責任及び透明性の原則
- イノベーションの原則

「人間中心のAI社会原則」(2019年3月 統合イノベーション戦略推進会議決定)より



人間中心のAI社会原則会議
(2018年5月～)

- …「社会(特に、国などの立法・行政機関)が留意すべき『AI社会原則』」
- …「開発者及び事業者において、基本理念及びAI社会原則を踏まえたAI開発利用原則を定め、遵守すべき」

開発者・事業者それぞれにおいて、AI開発利用原則を策定することを期待

そのための参考となるガイドラインが必要

(注) 2016年2月～9月は、「AIネットワーク化検討会議」として検討
AIネットワーク社会推進会議
(2016年10月～)

総務省の取組

AI開発ガイドライン※

➤ 開発者が留意すべき事項と解説

※「国際的な議論のためのAI開発ガイドライン案」

AI利活用ガイドライン

➤ 利用者が留意すべき事項と解説

関係省庁に共有の上、開発者・事業者提供。自主的対応を支援。

A I ネットワーク社会推進会議 構成員

(令和4年7月25日現在)

議長	須藤 修	(中央大学国際情報学部教授、中央大学E L S I センター所長、 東京大学大学院情報学環特任教授)
副議長	平野 晋	(中央大学国際情報学部教授・学部長)
構成員	Alice Xiang	(Global Head of AI Ethics Office, Sony Group Corp)
	岩本 敏男	(株式会社エヌ・ティ・ティ・データ相談役)
	遠藤 信博	(日本電気株式会社取締役会長)
	大久保 和也	(グーグル合同会社 政府渉外・公共政策部長 (デジタル利活用担当))
	大田 佳宏	(東京大学大学院数理科学研究科特任教授、 Arithmer 株式会社代表取締役社長兼CEO)
	大橋 弘	(東京大学副学長、大学院経済学研究科教授)
	大屋 雄裕	(慶應義塾大学法学部教授)
	金井 良太	(株式会社アラヤ代表取締役CEO)
(幹事)	喜連川 優	(国立情報学研究所所長/東京大学特別教授)
	木村 たま代	(主婦連合会事務局長)
	小塚 莊一郎	(学習院大学法学部法学科教授)
	近藤 則子	(老テク研究会事務局長)
	穴戸 常寿	(東京大学大学院法学政治学研究科教授)
	実積 寿也	(中央大学総合政策学部教授)
(幹事)	城山 英明	(東京大学大学院法学政治学研究科教授)
	新保 史生	(慶應義塾大学総合政策学部教授)
	杉山 将	(理化学研究所革新知能統合研究センター長、東京大学新領域創成 科学研究科教授)
(幹事)	鈴木 晶子	(京都大学名誉教授)
	鈴木 教洋	(株式会社日立製作所執行役常務CTO兼研究開発グループ長)
	高橋 恒一	(理化学研究所生命機能科学研究センターチームリーダー、 慶應義塾大学大学院政策・メディア研究科特任教授)
	谷崎 勝教	(株式会社三井住友銀行専務執行役員)
	田丸 健三郎	(日本マイクロソフト株式会社業務執行役員ナショナルテクノロジー オフィサー)
	時田 隆仁	(富士通株式会社代表取締役社長)
(幹事)	中川 裕志	(理化学研究所革新知能統合研究センターチームリーダー)
	長田 三紀	(情報通信消費者ネットワーク)
	中西 崇文	(武蔵野大学データサイエンス学部データサイエンス学科長・准教授)
(幹事)	西田 豊明	(福知山公立大学副学長、情報学部学部長・教授)
	萩田 紀博	(大阪芸術大学アートサイエンス学科長・教授、 株式会社国際電気通信基礎技術研究所特別研究所長)
(幹事)	橋元 良明	(東京女子大学現代教養学部心理・コミュニケーション学科コミュニ ケーション専攻教授)
	林 秀弥	(名古屋大学大学院法学研究科教授・同アジア共創教育研究機構教授)
	福田 雅樹	(大阪大学社会技術共創研究センター教授(兼)大学院法学研究科 教授、東京大学大学院情報学環客員教授)
(幹事)	堀 浩一	(人間文化研究機構理事)

- | | | |
|------|-------|---------------------------------------------|
| | 松尾 豊 | (東京大学大学院工学系研究科教授) |
| | 武藤 真祐 | (医療法人社団鉄祐会理事長、株式会社インテグリティ・ヘルスケア
代表取締役会長) |
| (幹事) | 村井 純 | (慶應義塾大学教授) |
| | 村上 憲郎 | (大阪公立大学大学院教授) |
| (幹事) | 森川 博之 | (東京大学大学院工学系研究科教授) |
| | 山川 宏 | (全脳アーキテクチャ・イニシアティブ代表) |
| | 山口 明夫 | (日本アイ・ビー・エム株式会社代表取締役社長) |
| | 山本 勲 | (慶應義塾大学商学部教授) |

(敬称略。議長及び副議長を除き、五十音順)

- | | | |
|----|--------|--------------|
| 顧問 | 安西 祐一郎 | (慶應義塾大学名誉教授) |
| | 西尾 章治郎 | (大阪大学総長) |
| | 濱田 純一 | (東京大学名誉教授) |

(敬称略。五十音順)

オブザーバー

デジタル庁企画官
 内閣府政策統括官(科学技術・イノベーション担当)付参事官
 個人情報保護委員会事務局参事官
 消費者庁消費者安全課長
 文部科学省研究振興局参事官(情報担当)
 経済産業省商務情報政策局情報経済課長
 経済産業省産業技術環境局研究開発課長
 情報通信研究機構 A I 研究開発本部長
 科学技術振興機構社会技術研究開発センター企画運営室長
 理化学研究所革新知能統合研究推進室長
 産業技術総合研究所情報・人間工学領域長補佐

AI ネットワーク社会推進会議
AI ガバナンス検討会 構成員

(令和4年7月25日現在)

座長	平野 晋	(中央大学国際情報学部教授・学部長)
構成員	荒堀 淳一	(富士通株式会社AI倫理ガバナンス室室長)
	江間 有沙	(東京大学未来ビジョン研究センター准教授)
	江村 克己	(日本電気株式会社シニアアドバイザー)
	大久保 和也	(グーグル合同会社 政府渉外・公共政策部長 (デジタル利活用担当))
	大屋 雄裕	(慶應義塾大学法学部教授)
	落合 孝文	(弁護士)
	影広 達彦	(株式会社日立製作所研究開発グループデジタルサービス研究統括本部 先端AIイノベーションセンタ主管研究長)
	金井 良太	(株式会社アラヤ代表取締役CEO)
	河島 茂生	(青山学院大学コミュニティ人間科学部准教授)
	喜多 淳一郎	(株式会社エヌ・ティ・ティ・データ技術革新統括本部システム技術本 部長)
	木村 たま代	(主婦連合会事務局長)
	小塚 莊一郎	(学習院大学法学部法学科教授)
	三部 裕幸	(弁護士)
	城山 英明	(東京大学大学院法学政治学研究科教授)
	高木 幸一	(KDDI株式会社技術戦略本部技術渉外部企画グループリーダー)
	高橋 恒一	(理化学研究所生命機能科学研究センターチームリーダー、 慶應義塾大学大学院政策・メディア研究科特任教授)
	武田 英明	(国立情報学研究所情報学プリンシプル研究系教授)
	田丸 健三郎	(日本マイクロソフト株式会社業務執行役員ナショナルテクノロジー オフィサー)
	中川 裕志	(理化学研究所革新知能統合研究センターチームリーダー)
	長田 三紀	(情報通信消費者ネットワーク)
成原 慧	(九州大学法学研究院准教授)	
西田 豊明	(福知山公立大学副学長、情報学部学部長・教授)	
堀 浩一	(人間文化研究機構理事)	
森本 典繁	(日本アイ・ビー・エム株式会社常務執行役員 最高技術責任者 兼 研究開発担当)	
山本 龍彦	(慶應義塾大学大学院法務研究科教授、慶應義塾大学グローバル リサーチインスティテュート (KGR I) 副所長)	
湯浅 壘道	(明治大学公共政策大学院ガバナンス研究科教授)	

(敬称略。座長を除き、五十音順。)