

報告書 2019（案）

令和元年●月●日

AI ネットワーク社会推進会議

目次

要旨	I
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	3
1. 国内の動向	3
2. 海外の動向	6
3. 国際的な議論の動向	12
4. (参考) 各ガイドライン等の比較	15
第2章 AI 利活用ガイドライン策定の考え方	17
1. 背景・経緯	17
2. AI 利活用ガイドラインの位置づけ	19
3. AI 利活用ガイドラインの概要	21
4. 今後の展開	39
第3章 今後の課題	41
1. AI の開発及び利活用の促進並びに AI ネットワーク化の健全な進展に関する事項 41	
2. AI ネットワーク化が社会・経済にもたらす影響に関する事項	46
3. AI ネットワーク化が進展する社会における人間をめぐる課題に関する事項	48
4. その他の課題	50
(参考) 報告書 2018 に掲げられている「今後の課題」	51
結びに代えて	53

【別紙1】 AI 利活用ガイドライン

【別紙1 (附属資料)】 AI 利活用原則の各論点に対する詳説

【別紙2】 AI ガイドライン比較表

要旨

AIに関する技術開発と利活用が急速に進展する中、今後、AIシステムがインターネット等を通じて他のAIシステム等と接続し連携する「AIネットワーク化」が進展していくことにより、社会的な課題の解決など人間や社会・経済に多大な便益がもたらされることが期待される一方、透明性や制御喪失などのリスクに基づく人々のAIに対する不安などが、AIの開発及び利活用の促進やAIネットワーク化の健全な進展の阻害要因となるのではないかと懸念されている。

これらのことを踏まえると、AIの便益を増進させ、リスクを抑制するとともに、人々のAIに対する不安を取り除き、信頼を醸成するため、AIの開発者、利用者やデータ提供者が留意することが期待される事項等を整理することが求められる。

総務省では、今後のAIネットワーク化の進展を見据え、2016年（平成28年）10月から「AIネットワーク社会推進会議」を開催し、本推進会議は、2017年（平成29年）7月に、AIの開発者が留意することが期待される事項を整理した「国際的な議論のためAI開発ガイドライン案」を含む『報告書2017』を取りまとめた。また、『報告書2017』の公表後、AIの利用者やデータ提供者が留意することが期待される事項に関する検討を行い、2018年（平成30年）7月に「AI利活用原則案」を含む『報告書2018』をとりまとめた。また、本推進会議は、『報告書2018』の公表後、利活用に関する指針の策定を目指し、前述の「AI利活用原則案」の内容の深掘りを行ってきた。

本報告書は、本推進会議における検討及び関連する国内外の動向を踏まえ、AI利活用ガイドラインの基本的な考え方と内容を示すと共に、今後の課題を整理している。

第1章では、『報告書2018』の公表以降の最近のAIネットワーク化に関する国内外の動向及び国際的な議論の動向を概観している。

第2章では、本推進会議が作成したAI利活用ガイドライン（別紙1）について、検討の背景・経緯を概観した上で、検討の方向性を説明するとともに、その概要を紹介している。特に、基本理念については、「人間中心のAI社会原則」との関係について確認するとともに、関連する主体や、AIの利活用において各主体がたどるフェーズについてそれぞれ整理している。その上で、AI利活用原則及びその論点についての詳説をまとめるとともに、それらの論点をどのフェーズで留意すべきかについて整理を行っている。

なお、同ガイドラインに掲げるAI利活用原則は、次のとおりである。

- ① **適正利用の原則**：利用者は、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。
- ② **適正学習の原則**：利用者及びデータ提供者は、AI システムの学習等に用いるデータの質に留意する。
- ③ **連携の原則**：AI サービスプロバイダ、ビジネス利用者及びデータ提供者は、AI システム又は AI サービス相互間の連携に留意する。また、利用者は、AI システム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。
- ④ **安全の原則**：利用者は、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、利用者等及び第三者の生命・身体・財産に危害を及ぼすことがないよう配慮する。
- ⑤ **セキュリティの原則**：利用者及びデータ提供者は、AI システム又は AI サービスのセキュリティに留意する。
- ⑥ **プライバシーの原則**：利用者及びデータ提供者は、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。
- ⑦ **尊厳・自律の原則**：利用者は、AI システム又は AI サービスの利活用において、人間の尊厳と個人の自律を尊重する。
- ⑧ **公平性の原則**：AI サービスプロバイダ、ビジネス利用者及びデータ提供者は、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が不当に差別されないよう配慮する。
- ⑨ **透明性の原則**：AI サービスプロバイダ及びビジネス利用者は、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。
- ⑩ **アカウントビリティの原則**：利用者は、ステークホルダに対しアカウントビリティを果たすよう努める。

第3章では、第2章における検討の結果等を踏まえ、今後の課題を次のように整理している。

- AI の開発及び利活用の促進並びに AI ネットワーク化の健全な進展に関する事項
 - ・ AI 開発ガイドライン案及び AI 利活用ガイドラインの周知・展開、継続的な見直し、関係するステークホルダが取り組む環境整備に関する課題 等
- AI ネットワーク化が社会・経済にもたらす影響に関する事項
 - ・ AI ネットワーク化が社会・経済にもたらす影響に関するシナリオ分析 等
- AI ネットワーク化が進展する社会における人間をめぐる課題に関する事項
 - ・ 人間と AI との関係の在り方に関する検討、ステークホルダ間関係の在り方に関する

る検討 等

本報告書が、AI ネットワーク化に関する社会的・経済的・倫理的・法的課題をめぐる国内外の検討に寄与し、AI の開発及び利活用の促進並びに AI ネットワーク化の健全な進展が図られることを願いたい。

はじめに

AI（人工知能）の研究開発及び利活用は日々進展を遂げており、関連する報道を見ない日はなく、人々の AI に対する期待は非常に高まっている。AI は、今後、個人、地域社会、各国、国際社会の抱える様々な課題¹の解決のため、その利活用が期待されており、特に AI システムがインターネット等を通じて他の AI システム等と接続し連携する「AI ネットワーク化」が進展することにより、人間及びその社会や経済に多大な便益が広範にもたらされることが期待される。我が国においても、人口減少・少子化を迎える中、新しい社会・経済を構築していく上で、AI の利活用は非常に有効であるものと考えられている。

他方、AI の判断のブラックボックス化や制御喪失などのリスクとともに、人々の AI に対する不安などが、AI の開発及び利活用の促進や AI ネットワーク化の健全な進展の阻害要因となるのではないかと懸念されている。

AI の便益を人々や社会が享受するためには、AI の開発をさらに推し進めるとともに、利活用の裾野をより一層広げていくことが必要である。そのためには、まずもって人々の AI に対する不安を取り除き、信頼を醸成する取組を進めることが必要である。AI の開発者や利用者が、それぞれ AI の開発や利活用において、AI の便益を増進させリスクを抑制する観点から期待される取組を行うとともに、開発者や利用者に加えて、政府や業界団体等を含めた関係するステークホルダが AI の開発及び利活用の促進や AI ネットワーク化の健全な進展に向けた環境整備のための取組を行うことにより、AI に対する社会的な受容性の向上がもたらされるものと考えられる。

こうした中、2016 年（平成 28 年）4 月に日本で開催された G7（先進 7 か国）香川・高松情報通信大臣会合において、ホスト国である日本は、AI 開発原則のたたき台を紹介し各国関係閣僚による議論が行われた。その結果、G7 において「AI 開発原則」及びその内容の解説からなる「AI 開発ガイドライン」の策定に向け、引き続き G7 各国が中心となり、OECD（経済協力開発機構）等国際機関の協力も得て議論していくことで合意した。その後、欧米各国においても、AI に関する社会的・経済的・倫理的・法的課題をめぐる検討が本格化し、国際的な議論が加速している。

以上のような背景と問題意識に鑑み、総務省情報通信政策研究所では、2016 年（平成 28 年）10 月から「AI ネットワーク社会推進会議」を開催し、「AI ネットワーク社会推進フォーラム」（国際シンポジウム）²における議論等を踏まえ、2017 年（平成 29 年）7 月に、AI の開発者が留意することが期待される事項を整理した「国際的な議論のため AI 開発ガイド

¹ 国際社会の抱える課題については、国連の「持続可能な開発目標」（SDGs）（http://www.un.org/ga/search/view_doc.asp?symbol=A/70/L.1）などを参照。

² 総務省が 2017 年（平成 29 年）3 月に東京大学において開催した国際シンポジウム。日米欧の産学民官のステークホルダが参加し、AI ネットワーク化のガバナンスの在り方について活発な議論が行われた。

ライン案」を含む『報告書 2017』³を取りまとめた。

その後、本推進会議では、AI の利用者やデータ提供者が留意することが期待される事項に関する検討を行うとともに、その検討に資するよう、AI ネットワーク化が進展した社会の将来像を描き AI の利活用における課題等を抽出するため、AI 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に関する検討等を行った。2018 年（平成 30 年）7 月には、AI の利活用において留意することが期待される事項を整理した「AI 利活用原則案」を含む『報告書 2018』⁴を取りまとめた。

これらの検討成果は、国際社会においても積極的に紹介され、特に、OECD が理事会勧告案を策定するために 2018 年（平成 30 年）9 月に設置した専門家会合（AI expert Group at the OECD）においては、我が国から参加した有識者より、AI 開発ガイドライン案、AI 利活用原則案のそれぞれについて、その内容のみならず、検討の背景や検討にあたり行われた議論の状況についても紹介するなど、OECD 理事会勧告の原案策定にあたり大きな貢献を行った。このため、2019 年（令和元年）5 月に公表された OECD 理事会勧告（Recommendation on Artificial Intelligence）は、我が国において検討してきた上述の各原則、ガイドラインと整合のとれたものとなっている。

このように、AI に関する原則については、我が国が国際的な議論を主導してきた結果、その概念については、概ね国際的にコンセンサスが得られつつある。他方で、今後は、この「原則」をどのように実現していくべきか、そのためには、各ステークホルダがどのような措置を講ずるべきかといった、具体的な措置について検討が進められることが期待されている。

報告書 2018 の公表後、本推進会議は、同年 10 月から 2019 年（令和元年）5 月までに開催した 10 回の会合（本推進会議の会合（親会）2 回のほか、親会の下に置かれた AI ガバナンス検討会 8 回）における検討等を踏まえて、本報告書（案）を作成した。その後、同年 6 月●日から●月●日までにかけて、本報告書（案）に関する意見募集を行った上で、AI ガバナンス検討会（第 9 回）及び本推進会議（第 13 回）の会合での議論を経て、本報告書を最終的に取りまとめた。

本報告書は、本推進会議における検討と関連する国内外の動向及び国際的な議論の動向を概観した上で、AI 利活用ガイドラインの基本的な考え方と内容を示すとともに、今後の課題を整理している。本報告書が、AI ネットワーク化に関する社会的・経済的・倫理的・法的課題をめぐる国内外の検討に寄与し、AI 利活用ガイドラインの内容が国内のみならず国際的にも共有されることで、AI に対する人々の信頼の醸成につながり、AI の開発及び利活用の促進や AI ネットワーク化の健全な進展が図られることを願いたい。

³ AI ネットワーク社会推進会議「報告書 2017 -AI ネットワーク化に関する国際的な議論の推進に向けて-」（2017 年（平成 29 年）7 月 28 日。以下「報告書 2017」という。）

⁴ AI ネットワーク社会推進会議「報告書 2018 -AI の利活用の促進及び AI ネットワーク化の健全な進展に向けて-」（2018 年（平成 30 年）7 月 17 日。以下「報告書 2018」という。）

第1章 AI ネットワーク化をめぐる最近の動向

AI に関する動きはめまぐるしく、AI に関連する報道を目にしない日がないほどである。AI に関する検討に当たっては、関係ステークホルダとのコンセンサスを図っていくことが重要であり、本推進会議においても、AI ネットワーク化に関する国内外の動向及び国際的な議論の動向を踏まえて検討を行うこととしている。

報告書 2018 公表後の1年は、AI の開発、利活用に関する原則についての議論が大きな進展を遂げた。G7 では、2018 年（平成 30 年）12 月にカナダのモントリオールにてマルチステークホルダ会合が開催され、AI の社会的、経済的、倫理的、法的課題について、G7 各国の政府機関のみならず、企業関係者、技術系・人文系の研究者、消費者団体、学生等が一堂に会し、議論を行った。また、OECD では、2016 年の G7 情報通信大臣会合における我が国からの提言を契機として、専門家会合を開催するなどして精力的な議論が行われた結果、2019 年（令和元年）5 月に AI 原則が理事会勧告としてとりまとめられるに至った。今後は、国際的にコンセンサスの得られた原則について、各ステークホルダがどのような役割を果たすことにより実現を図っていくか、すなわち、原則を実現するための具体的な方策に関する議論の本格化が見込まれる。

本章においては、上記を含め、主として報告書 2018 公表後の AI ネットワーク化をめぐる動向を概観する⁵。

1. 国内の動向

(1) AI 戦略（AI 戦略実行会議）

政府は、2018 年（平成 30 年）6 月 15 日に閣議決定された「統合イノベーション戦略」に基づき、イノベーションに関連が深い司令塔会議である総合科学技術・イノベーション会議、高度情報通信ネットワーク社会推進戦略本部、知的財産戦略本部、健康・医療戦略推進本部、宇宙開発戦略本部及び総合海洋政策本部並びに地理空間情報活用推進会議について、横断的かつ実質的な調整を図るとともに、同戦略を推進するため、内閣に統合イノベーション戦略推進会議を設置した。

また、同「統合イノベーション戦略」において、AI 技術は特に取組を強化すべき主要分野の1つとして、世界の潮流を踏まえた方向性を打ち出すこととされていることを踏まえ、同年7月27日、上記統合イノベーション戦略推進会議で決定した、「イノベーション政策強化推進のための有識者会議」の設置に関する規定に基づき、その1つとして「AI 戦略」（AI 戦略実行会議）を設置し、統合イノベーション戦略推進会議「AI 戦略 2019」（有識者提案）を報告した。

⁵ 報告書 2018 公表以前の AI ネットワーク化をめぐる動向については、報告書 2018 第1章参照。

同戦略案は、「人間尊重」、「多様性」、「持続可能」の3つの理念を掲げ、Society 5.0 を実現しSDGs に貢献していくこととされており、上記3理念を実装するための4つの戦略目標（人材、産業競争力、技術体系、国際）を設定し、それぞれの目標を達成するために、「未来への基盤作り（教育改革、研究開発）」、「産業・社会の基盤作り（社会実装、データ関連基盤、デジタル・ガバメント、中小・新興企業支援）」、「倫理（AI 社会原則：下記（2）参照）」に関する取組を特定している。

（2）人間中心のAI社会原則会議（「人間中心のAI社会原則検討会議」より移行）

内閣府は、AI をよりよい形で社会実装し共有するための基本原則となる「人間中心のAI社会原則」を策定し、同原則をG7、OECD 等の国際的な議論に供するため、AI に関する倫理や中長期的な研究開発・利活用などについて、産学民官による幅広い視野からの調査・検討を行うことを目的として、人工知能技術戦略会議の下に、「人間中心のAI社会原則検討会議」を設置し⁶、2018年（平成30年）5月8日に第1回会合が開催された。以降、同年12月13日までに8回の会合が開催され、同年12月27日、素案を提示した後、パブリックコメントを実施した。その後、政府の「AI戦略」に反映させることを目的として、AI戦略実行会議の下、「人間中心のAI社会原則会議」が新たに設置され（「人間中心のAI社会原則検討会議」を承継）、2019年（平成31年）2月28日に会合を開催、「人間中心のAI社会原則（案）」をとりまとめ、同年3月29日の統合イノベーション戦略推進会議に提案し、同日付で同案が決定された⁷。

同原則はAIのメリットを最大限に引き出し、人々の不安を解消するために、社会が留意すべき原則をまとめたものとなっており、以下の通り、基本理念、Society 5.0 実現に必要な社会変革、及び、AI社会原則から構成されている：

- 基本理念：
 - ① 人間の尊厳が尊重される社会（Dignity）
 - ② 多様な背景を持つ人々が多様な幸せを追求できる社会（Diversity & Inclusion）
 - ③ 持続性ある社会（Sustainability）
- Society 5.0 実現に必要な社会変革「AI-Ready な社会」：
 - ① 人
 - ② 社会システム
 - ③ 産業構造
 - ④ イノベーションシステム（イノベーションを支援する環境）
 - ⑤ ガバナンス
- AI社会原則：

⁶ 内閣府、総務省、文部科学省、厚生労働省、経済産業省、国土交通省が合同で開催する会議として設置された。

⁷ 次に掲げる URL のウェブサイトに掲載。
<<https://www.cas.go.jp/jp/seisaku/jinkouchinou/>>

- 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

また、上記に続いて、開発者・事業者を名宛人にした「AI 開発利用原則」についての言及がなされているが、上記の AI 社会原則を踏まえたものとするべきであること、及び、現在、多くの国、団体、企業等において議論されていることから、我々は早急にオープンな議論を通じて国際的なコンセンサスを醸成し、非規制的で非拘束的な枠組みとして国際的に共有されることが重要であると述べられている。

(3) 保健医療分野 AI 開発加速コンソーシアム

厚生労働省は、2017 年（平成 29 年）6 月、「保健医療分野における AI 活用推進懇談会」において AI 開発を進めるべき重点 6 領域を選定し、報告書として取りまとめたが、諸外国の AI 開発が急速に進む中で、AI 開発及び利活用を加速させるための課題や対応策及び本邦における今後の研究開発の方向性を探る必要があることを踏まえ、「保健医療分野 AI 開発加速コンソーシアム」を設置し、2018 年（平成 30 年）7 月 23 日に第 1 回会合を開催した⁸。

同コンソーシアムにおいて、前述の重点 6 領域を中心とした AI 開発状況の確認、及び、画像診断支援を例に AI の開発段階におけるロードブロック（課題）の特定と解消のための検討等が行われている中で、AI を用いた診断、治療等の支援を行うプログラムの利用と医師法第 17 条の規定との関係に関し、同省医政局医事課より 2018 年（平成 30 年）12 月 19 日に行われた以下の周知が取り上げられている：

AI を用いた診断・治療支援を行うプログラムを利用して診療を行う場合、

- 診断、治療等を行う主体は医師であること
- 医師はその最終的な判断の責任を負うこと
- 当該診療は医師法第 17 条の医業として行われること

(4) 日本経済団体連合会

一般社団法人日本経済団体連合会は、産業界の現状として日本企業の AI 及びその関連分野における国際的な産業競争力を向上させるべく、「AI 活用原則タスクフォース」

⁸ コンソーシアムの座長は、本推進会議の北野構成員（株式会社ソニーコンピュータサイエンス研究所代表取締役社長兼所長）が務めている。

を設置し⁹、第1回会合を2018年（平成30年）1月17日に開催した。以降、検討を深め、2019年（平成31年）2月19日に「AI活用戦略～AI-Readyな社会の実現に向けて～」を策定、公表した¹⁰。

同戦略では、デジタル革新と多様な人々の創造力の融合によって実現する未来社会のコンセプト「Society 5.0（創造社会）」を実現する中核技術を「AI」とした上で、AI研究の技術的背景の概観、産業界ならではの視点から見たAI活用を進めるにあたって留意すべき原則、企業、個人、社会制度と産業基盤、それぞれがAIを活用するための準備、AI-Ready化を進める際のガイドライン、産業分野でAI活用を進めるにあたってのフレームワークについて述べている。

また、国内外でまとめられた原則として、本推進会議で議論されている「AI開発ガイドライン案」や「AI利活用原則案」にも触れている。

2. 海外の動向

(1) 米国電気電子学会（IEEE）

米国電気電子学会（IEEE）の自律的及び知的システムの倫理（Ethics of Autonomous and Intelligent Systems(A/IS)）に関するグローバル・イニシアチブは、2019年（平成31年）3月25日に報告書「倫理的に調整された設計（Ethically Aligned Design（EAD）第1エディション）」を公表した¹¹。本エディションでは、特に「原則から実行へ（From Principles to Practice）」を重視しており、特に本報告書で掲げられる倫理課題から技術への橋渡しを目的とした標準であるIEEE P7000™シリーズの検討、用語集の作成・公表¹²、及び、関連する検討会の設立などに触れている。

EADの概念的枠組は、①人間の普遍的な価値（Universal human values）、②政策の自己決定とデータ仲介（Political Self- and Data Agency and Data Agency）、③技術的な信頼性（Technical Dependability）の3つの柱から構成されており、それらに紐付けられる一般原則として、以下の8項目が掲げられている（下線は本エディションで新たに加わ

⁹ タスクフォースの主査は、本推進会議の北野構成員（株式会社ソニーコンピュータサイエンス研究所代表取締役社長兼所長）が務めている。

¹⁰ 次に掲げるURLのウェブサイトに掲載。

<<http://www.keidanren.or.jp/policy/2019/013.html>>

¹¹ The IEEE Global Initiative on Ethics of Autonomous and Intelligent System, “Ethically Aligned Design, First Edition: A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems” (2019)

次に掲げるURLのウェブサイトに掲載。

<<https://ethicsinaction.ieee.org/>>

¹² “Ethically Aligned Design, First Edition Glossary” (2019)

次に掲げるURLのウェブサイトに掲載。

<https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/ead1e_glossary.pdf>

った原則) :

- ① 人権 (Human Rights)
- ② 幸福 (Well-being)
- ③ データ仲介 (Data Agency)
- ④ 効能 (Effectiveness)
- ⑤ 透明性 (Transparency)
- ⑥ アカウンタビリティ (Accountability)
- ⑦ 悪用への警戒 (Awareness of Misuse)
- ⑧ 技能 (Competence)

また、EAD は、一般原則に関する章を含め、全 11 の章¹³から構成されており、第 2 版¹⁴と比較し、以下の 2 つの章が追加となっている。

- 原則から実行へ (From Principles to Practice)
- 持続的発展のための自律的及び知的システム (A/IS for Sustainable Development)

(2) Future of Life Institute (FLI)

米国 Future of Life Institute (FLI) は 2019 年 (平成 31 年) 1 月 2～7 日にかけて汎用 AI (AGI) のに焦点を当て、人類にとって最善の未来を築くために AGI をどのように設計できるかを探るための会議“Beneficial AGI 2019”を開催した¹⁵。

各分野のステークホルダが集い、技術的な安全性や、戦略・ガバナンス等に関してワークショップ、パネルを実施した。

(3) データ保護プライバシー・コミッショナー国際会議 (ICDPPC)

データ保護プライバシー・コミッショナー国際会議 (ICDPPC) は 2018 年 (平成 30 年) 10 月 23 日、第 40 回の同会議において「AI における倫理及びデータ保護についての宣言」を採択した¹⁶。

同宣言において、AI の開発における人権保護のための本質的価値として以下の 6 つ

¹³ ①原則から実行へ、②一般原則、③A/IS における伝統的倫理観、④幸福、⑤感情コンピューティング、⑥パーソナルデータと個々のアクセス制御、⑦倫理的な研究及び設計を導く方法、⑧持続的発展のための A/IS、⑨A/IS への価値の埋め込み、⑩政策、⑪法律の 11 章から構成されている。

¹⁴ 第 2 版及び、IEEE P7000™ シリーズの内容については「報告書 2018」17 ページ参照。

¹⁵ 次に掲げる URL のウェブサイトに掲載。

<<https://futureoflife.org/beneficial-agi-2019/?cn-reloaded=1>>

¹⁶ “DECLARATION ON ETHICS AND DATA PROTECTION IN ARTIFICIAL INTELLIGENCE”

次に掲げる URL のウェブサイトに掲載。

<https://www.privacyconference2018.org/system/files/2018-10/20180922_ICDPPC-40th_AI-Declaration_ADOPTED.pdf>

の指導原則を支持している:

- ① 基本的人権の尊重及び公正の原則に従ったデザイン、開発及び使用
- ② 継続的な注意及び警戒と、説明責任
- ③ AI システムの透明性及び明瞭性の改善
- ④ プライバシー・バイ・デフォルト及びプライバシー・バイ・デザインによる責任をもったデザイン及び開発
- ⑤ すべての個人の権限強化の推進、及び、個人の権利行使の促進と、広範な参画の機会の創設
- ⑥ 違法なバイアス又は差別の軽減及び緩和

ICDPPC では、同宣言に沿った指針を定めるため、AI ワーキンググループが創設され、2019 年 10 月予定の第 41 回 ICDPPC に向けて、より具体的な指針を策定する予定である。

(4) Public Voice

米・電子プライバシー情報センター (EPIC) により設立された団体 Public Voice は、2018 年 (平成 30 年) 10 月 23 日、「AI に対するユニバーサルガイドライン」を公表した¹⁷。

同ガイドラインでは、AI の設計や利活用の改善を目的として以下 12 原則を提案している:

- ① 透明性の権利
- ② 自己決定権利
- ③ (AI を使っていることの) 周知義務
- ④ 公平性義務
- ⑤ 評価とアカウントビリティ義務 (評価してリリースし、リリース後は責任を負うべき)
- ⑥ 正確性、信頼性、妥当性義務
- ⑦ データ品質義務
- ⑧ 公共安全性義務
- ⑨ サイバーセキュリティ義務
- ⑩ 秘密裏に行われるプロファイリングの禁止
- ⑪ (政府等による) 単一のスコアリングの禁止
- ⑫ システム停止できる義務

また、AI システムの主な責任は、同システムに資金を供給し、開発し、そして展開

¹⁷ “Universal Guidelines for Artificial Intelligence”

次に掲げる URL のウェブサイトにも掲。

<<https://thepublicvoice.org/ai-universal-guidelines/>>

する機関にあるべきと言及している。

(5) 欧州委員会

欧州委員会により 2018 年（平成 30 年）6 月に選定された 52 名のメンバーによる AI ハイレベル専門家グループ（HLEG）が、同年 12 月 18 日、「信頼できる AI（Trustworthy AI）のための倫理ガイドライン案」を公表した。その後、パブリックコメントを経て、2019 年（平成 31 年）4 月 8 日に同ガイドラインを公表した¹⁸。

同ガイドラインでは、信頼できる AI を満足する 3 要素として、合法的であること、倫理的であること、及び、頑健であることを掲げており、その上で、基本的人権に基づき尊重すべき 4 つの倫理原則、及びそれらを実現するための 7 つの要求を以下の通り掲げ、さらに、それらの実現（運用）を評価するために考えられるチェックリスト¹⁹を列挙している。

4 つの倫理原則：

- ① 人間の自律性の尊重（respect for human autonomy）
- ② 危害の防止（prevention of harm）
- ③ 公平性（fairness）
- ④ 説明可能性（explicability）

7 つの要求条件：

- ① 人間の営みと監視（human agency and oversight）
- ② 技術的な頑健性と安全性（technical robustness and safety）
- ③ プライバシーとデータガバナンス（privacy and data governance）
- ④ 透明性（transparency）
- ⑤ 多様性、無差別、公平性（diversity, non-discrimination and fairness）
- ⑥ 環境及び社会の幸福（environmental and societal well-being）
- ⑦ アカウンタビリティ（accountability）

なお、掲げられた内容は非拘束的なものとして AI を開発・利用する全ての関連するステークホルダを対象としているため、企業・団体等の各ステークホルダは、本ガイドラインの内容を咀嚼し、自らの憲章・行動規範等に適用することにより、信頼できる AI への関与を表明することができる。また、今後は上記チェックリストに対し、オンラインでのフィードバック、及び公的・私的セクターの代表者による深いレビューを行い、2020 年を目処にとりまとめる予定である。

¹⁸ High Level Experts Group on Artificial Intelligence, “Ethics Guideline for Trustworthy AI”
次に掲げる URL のウェブサイトに掲載。

<<https://ec.europa.eu/futurium/en/ai-alliance-consultation/guidelines>>

¹⁹ 掲げられたチェックリストは、信頼できる AI を実現するための要件を特定していくために作られたものであり、網羅性を担保するものではない。

(6) AI4People

欧州 AI4People²⁰は、2018年（平成30年）11月26日に報告書「良いAI社会（good AI society）のための倫理的な枠組み：機会、リスク、原則及び推奨」を公表した²¹。

同報告書では良いAI社会に向けた5つの原則と、同原則を尊重しつつ、AI活用の機会を促す一方で、リスクを回避または最小限に抑えることを可能とするための20の推奨項目を提案している。

原則としては、生命倫理でも用いられる **Beneficence**（幸福、便益）、**Non-maleficence**（危害を加えないこと）、**Autonomy**（自己決定）、**Justice**（正義）に **explicability**（説明可能性）を加えた5つである。

他方、推奨項目としては、評価方法（AIシステムによる誤り等を是正するための民事裁判所などの司法機関の能力（機能）の評価）、仕組み等の構築（社会的に重要な決定を行うAIシステムの処理プロセスを説明・理解できるようにするための仕組みの構築等）、財政面での促進（EUレベルで、社会的に望ましく環境にやさしいAI技術の利用と開発の促進）、及び支援（倫理的義務を伴う医師や法律家などの職業における自己行動規範の支援）が挙げられている。

(7) 独国

独連邦政府は連邦教育研究省、連邦経済エネルギー省及び連邦労働社会省より提案された「AI戦略」を2018年（平成30年）11月に閣議決定し公表した²²。同戦略の骨子は同年7月に提案されていたもので、以後、協議・パブリックコメントを経て閣議決定に至ったものである。

同戦略はAIの研究開発や利活用について世界を先導するレベルに持ち上げることを意図したもので、具体的な行動領域として、イノベーションを牽引するためのドイツ・欧州の研究強化、国内外での関係、社会との対話の実施と政治の活動枠の発展など全12項目を提示している。

特に「国内外での関係」では、国内、EUでの連携を踏まえた上で、G7、G20等において、各国のガイドラインとも連携を図っていく点に言及している。また、「社会との対話の実施と政治の活動枠の発展」では、AIが揺籃期にある中で、策定するAI戦略は業界各者への継続的なフィードバックが必要である点に言及している。

²⁰ AI4People は“good AI society”の構築を目指し2017年11月に設立されたフォーラム。

²¹ “AI4People—An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations”

次に掲げる URL のウェブサイト上所掲。

<<http://www.eismd.eu/ai4people/>>

²² 次に掲げる URL のウェブサイト上所掲。

<<https://www.bmwi.de/Redaktion/EN/Pressemitteilungen/2018/20181116-federal-government-adopts-artificial-intelligence-strategy.html>>

(8) 米国

ホワイトハウスが新設した人工知能（AI）特別委員会が、2016年に策定された「国家人工知能研究開発戦略プラン」の改定を行うため、意見募集を2018年（平成30年）9月26日より10月26日〆切で実施した。

上述の通り、現在の戦略プランは2016年に策定されたものだが、特別委員会は、AI技術がこの2年間で飛躍的に進化したことを受け、米国のAIに関する政策も修正すべきではないかという考えから戦略プランの改定の検討に着手した。意見募集の内容は戦略プラン改定の是非および改定の方向性についてとなっている。

また、米トランプ大統領は2019年（平成31年）2月11日、人工知能における米国のリーダーシップ維持のための大統領令（AIイニシアチブ）に署名した²³。大統領は、声明で、「AIにおける米国のリーダーシップを維持することは、米国の経済及び安全保障にとって極めて重要」としている。

同大統領命令は、連邦政府におけるAIの研究開発予算優先化、AIの研究開発のためのデータ・コンピューティングリソース提供、信頼性・安全性・相互運用性等を備えたAIシステムの開発促進のための技術基準策定、AI人材育成、米国の優位性保護のための行動計画等から構成されており、特に、信頼性・安全性・相互運用性等を備えたAIシステムの開発促進のための技術基準策定については、AIに係る規制・非規制手法の開発を行う機関に対する覚書の発出を行政管理予算局（OMB）に、AI技術を用いた安全・堅牢なシステムのための技術基準に係る計画の策定を国立標準技術院（NIST）にそれぞれ要求したものとなっている。

加えて、米・国防総省は、「我々の安全と繁栄を進化させるためのAIの活用」と題した新たな人工知能戦略を同年2月12日に発表した²⁴。米国は、同盟国やパートナーとともにAI導入を進めていかねばならないとしており、同国防総省内のAIパイロットプロジェクトは、その運用、トレーニング、維持、人材採用、サイバー攻撃対策等を含めて「共同人工知能センター（JAIC）」が主導・連携するとしている。また、軍事倫理とAIによる安全性をリードするとした上で、これらに対するビジョンや指針の明確化、及び、堅牢性、信頼性、および安全性を高めるための研究開発への投資をしていくことをうたっている。

²³ “Executive Order on Maintaining American Leadership in Artificial Intelligence”

次に掲げる URL のウェブサイトにも掲げられている。

<<https://www.whitehouse.gov/presidential-actions/executive-order-maintaining-american-leadership-artificial-intelligence/>>

²⁴ “Harnessing AI to Advance Our Security and Prosperity”

次に掲げる URL のウェブサイトにも掲げられている。

<<https://dod.defense.gov/News/News-Releases/News-Release-View/Article/1755388/new-strategy-outlines-path-forward-for-artificial-intelligence/>>

(9) 中国

中国・次世代 AI（人工知能）発展計画推進弁公室²⁵（注）は 2019 年（平成 31 年）2 月 15 日ワークショップを開催した。

その中で、科学技術部の王志剛部長は、AI に関連する法律、倫理、標準及び社会問題の研究を更に強化し、AI 関連のガバナンスに関する国際的な交流・協力を深く参画していくため、同弁公室が、次世代 AI 戦略諮問委員会の下に、次世代 AI ガバナンス特別委員会の設置を決定したと発表した²⁶。同特別委員会は産業界・学術界の専門家から構成され、清華大学公共管理学院の薛澜（Xue Wei）教授が委員長を務める。

3. 国際的な議論の動向

(1) 経済協力開発機構（OECD）

経済協力開発機構（OECD）は、2018 年（平成 30 年）11 月 14～16 日に CDEP 会合を開催し、事務局が作成した分析レポートの概要説明、後に述べる理事会勧告策定に向けた取り組みの紹介等が行われた。前者については、各国から内容を歓迎する発言があり、成果物は後日秘匿解除することが合意された。

これに先立ち、OECD は、2019 年（平成 31 年）中の AI に関する理事会勧告策定を視野に入れ、AI の信頼構築と社会実装を促すための原則（To Foster Trust in and adoption of AI）の内容について検討を実施するため、デジタル経済政策委員会（CDEP）の下に AI に関する専門家会合（AIGO: AI expert Group at the OECD）を設置し、2018 年（平成 30 年）9 月に第 1 回会合を開催した。AIGO は OECD の加盟国から派遣された産学民官の専門家等で構成され、我が国からは須藤東京大学教授（本推進会議議長）、平野中央大学教授（本推進会議幹事、AI ガバナンス検討会座長）が参加した（議長は Wonki Min デジタル経済政策委員会議長）。以後、2019 年（平成 31 年）2 月 9 日までに計 4 回の会合が開催され、同日の第 4 回 AIGO 会合にて、専門家による理事会勧告案が取りまとめられた。なお、第 4 回 AIGO 会合開催前の同年 2 月 1 日には、東京にて、AI ガバナンス検討会構成員と OECD 事務局代表との間でワークショップを開催し、AIGO とりまとめ案についてディスカッションを行い、本推進会議における議論のインプットを行うなど、本推進会議は上記理事会勧告案の作成において積極的な貢献を行った。

OECD は、2019 年（平成 31 年）3 月 14～15 日に CDEP アドホック会合を開催し、

²⁵ 2017 年（平成 29 年）11 月に設立。中国科学技術部、国家発展・改革委員会、工業・情報化部、財政部、中国科学院、中国工程院、共産党中央軍事委員会科学技術委員会、中国科学技術協会等 15 部門から構成され、次世代人工知能発展計画及び重大科学技術プロジェクトの組織・運営の推進を担っている。

²⁶ 次に掲げる URL のウェブサイトに掲載。

<http://www.most.gov.cn/kjbgz/201902/t20190220_145130.htm>

同会合では、専門家会合の結果を踏まえた理事会勧告案について審議が行われ、議論の内容を反映した上で同案が承認された。その上で、同案は、同年5月22日に開催された閣僚理事会において承認され、同日公表された²⁷。

同理事会勧告は、「人間中心」の考え方を基本に、AIに関わる全ての人に適用される実用的な指針として、「信頼できる AI のための責任あるスチュワードシップに関する原則」と「信頼できる AI のための国内政策と国際協力」の2部から構成されている：

- 信頼できる AI のための責任あるスチュワードシップに関する原則

- ① 包摂的な成長、持続可能な開発及び幸福
- ② 人間中心の価値及び公平性
- ③ 透明性及び説明可能性
- ④ 頑健性、セキュリティ及び安全性
- ⑤ アカウンタビリティ

- 信頼できる AI のための国内政策と国際協力

- ① AI の研究開発への投資
- ② AI のためのデジタル・エコシステムの整備
- ③ AI を推進するための政策環境の整備
- ④ 人的能力の強化と労働市場の変化への準備
- ⑤ 信頼できる AI のための国際協力

なお、理事会勧告で規定された各原則を実現するために具体的に講じるべき措置等については、2019年7月に開催される CDEP 会合で検討が開始され、年内に取りまとめられる予定である。

(2) G7

2018年のG7議長国であるカナダ政府は、同年（平成30年）12月6日、カナダ・モントリオールにおいて、AIに関するG7マルチステークホルダ会合を開催した²⁸。同会合は、同年3月のG7イノベーション大臣会合における議長サマリの附属書（AIに関するG7イノベーション大臣の声明）において、同年秋にカナダ主催でAIに関するマルチステークホルダ会合を開催することにつき合意したのを受け開催されたものでG7各国の産学官・市民社会におけるAI専門家等が参加した。

G7各国は、ブレイクアウトセッションにて、『①社会のためのAI』、『②イノベーションの解放』、『③AIにおけるアカウンタビリティ』、『④仕事の未来』のうち一つのテーマを担当し、それぞれのテーマごとに同セッションへの参加者（マルチステークホル

²⁷ 次に掲げる URL のウェブサイトに掲載。

<<https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449>>

²⁸ 次に掲げる URL のウェブサイトに掲載。

<<http://www.ic.gc.ca/eic/site/133.nsf/eng/00001.html>>

ダ) とディスカッションを実施した。

我が国は、カナダとともに『③AI におけるアカウントビリティ』を担当し、共同議長は、日本側は堀東京大学教授（本推進会議幹事）、カナダ側は Ian Kerr オタワ大学教授、Jason Millar オタワ大学准教授が務めた。日加の共同議長、総務省情報通信政策研究所、カナダ CIFAR (Canadian Institute for Advanced Research) は共同で、事前にディスカッションペーパーを作成し、当日は同ペーパーを基に議論が行われた。

同セッションでは、「アカウントビリティを信頼に結びつける社会システムが重要」、「抽象的な原則やガイドラインと具体策との橋渡しが重要」、「プライバシーや公平性などについて様々なトレードオフを考慮していく必要」、「AI のプロセスのアカウントビリティと結果のアカウントビリティの両方についてバランスよく考えていくべき」、「AI に対する過剰な信頼も過小な信頼も望ましくなく、適切なレベルの信頼を醸成することが大切」等の意見がなされ、全体会合にて共同議長より報告された²⁹。

なお、会合にはトルドー首相が出席され、「ケベック州に拠点を置く”the AI-Powered Supply Chains Supercluster (SCALE.AI)”に最大 2 億 3 千万ドルを投資し、1 万 6 千人以上の雇用創出や 10 年間で 160 億ドル以上の経済効果をもたらすことを見込む」、「カナダとフランスは、人工知能の研究結果を理解し共有するための国際的なパネル (IPAI: a joint International Panel on Artificial Intelligence) を創設する」等を表明した。

(3) G20

2019 年の議長国である我が国は、2019 年（令和元年）5 月 10 日、G20 デジタル経済マルチステークホルダ会議を開催した。同会議は、同年 6 月 8 日、9 日に茨城・つくばで開催される「G20 茨城つくば貿易・デジタル経済大臣会合」のデジタル経済部分にかかる主要テーマについて、各ステークホルダの立場から議論を行い、6 月の大臣会合へ貢献することを目的としたものである。

同会議においては、東大・五神総長らの基調講演に加え、① “Society 5.0 for SDGs”、② “Data Free Flow with Trust”、③ “Human-centric AI” に関するパネルディスカッションが行われた。

うち③では、東大・須藤教授（本推進会議議長）をモデレータとして Wyckoff 局長 (OECD)、Chakchouk 局長 (ユネスコ)、Osius 副社長 (Google)、及び Jacques CRO (カナダ・Mnubo 社) が参加し、人間中心の AI について議論が行われた。須藤教授より日本の人間中心の AI 社会原則・AI 利活用ガイドライン等が紹介された。

²⁹ セッション③での議論の内容、及びディスカッションペーパーは次に掲げる URL のウェブサイトに掲載。

<<http://www.ic.gc.ca/eic/site/133.nsf/eng/00005.html>>

(4) ユネスコ

ユネスコは2018年(平成30年)9月11日、パリの同本部において、AIに関するラウンドテーブル(パネルディスカッション)を開催し、オランダ・トゥウエンテ大学のVerbeek教授がモデレータとなり、同教授、及び、堀東京大学教授(本推進会議幹事)を含む全5人がスピーチした³⁰。従来、ユネスコにおけるAIの話題は労働が奪われるなどのネガティブなものが多かったとのことだが、本ラウンドテーブルでは、AIを社会(特に文化や教育)に役立たせるべき、社会の方でもその準備をすべき、といったポジティブな議論が多くなされた。

また、2018年(平成30年)12月13日、モロッコ、ベンゲリル、Mohammed VI ポリテクニック(Polytechnic University)において、AIに関するフォーラムを開催し、アフリカにおけるAIの課題や可能性、将来性、キャパシティ・ビルディング等幅広いテーマを設定して議論が行われ、日本から参加した各専門家³¹のプレゼンテーションにはアフリカの聴衆、及びユネスコ関係者からも高く評価されプレゼンスが示された。

加えて、2019年(平成31年)3月4日、パリの同本部において、AIに関するハイレベル会合「Principles for AI: Towards a Humanistic Approach? A Global Conference」が開催され、日本からは、鈴木外務大臣政務官、須藤東京大学教授(本推進会議議長)および、江村 NEC 執行役員常務兼 CTO(本推進会議 AI ガバナンス検討会構成員)が参加した³²。同会議においては、世界中から様々な学識者、政府関係者、民間企業関係者、NGO 関係者が一堂に会し、AI の開発・発展に伴う倫理的・社会的・法的課題について議論を行い、将来的な倫理規範の策定に向け、マルチステークホルダーでグローバルに検討していくべきことなどが確認された。開会式ではアズレー事務局長が挨拶を述べた後、OECD のグリア事務総長も登壇して挨拶し、両機関が今後協力して、AI についての議論を行っていくことが確認された。

4. (参考) 各ガイドライン等の比較

本推進会議でとりまとめた「国際的な議論のための AI 開発ガイドライン案」、次章に述べる「AI 利活用ガイドライン」、及び本推進会議報告書で述べた各国、各主

³⁰ 次に掲げる URL のウェブサイトに掲載。

<<https://en.unesco.org/news/artificial-intelligence-reflection-its-complexity-and-impact-society>>

³¹ 日本からは、武田 日立製作所理事、内藤 JICA 国際協力専門員、および、江川 NEC 技術イノベーション戦略本部標準化推進部シニアエキスパートが参加した。

³² 次に掲げる URL のウェブサイトに掲載。

<https://www.unesco.emb-japan.go.jp/itpr_ja/AI2019.html>

体により提示されているガイドライン等の一部（下記）に対し、それぞれの構成、及び、尊重すべき価値を軸にどのようなことが述べられているかを列挙したものを別紙2に示す。

〔比較したガイドライン（案）〕

- 総務省 AI ネットワーク社会推進会議「AI 開発ガイドライン案」
- 総務省 AI ネットワーク社会推進会議「AI 利活用ガイドライン」
- 統合イノベーション戦略推進会議「人間中心の AI 社会原則」
- 人工知能学会「人工知能学会倫理指針」³³
- 欧州委員会ハイレベル専門家会合「信頼できる AI のための倫理ガイドライン」
- OECD 理事会勧告
- 米国電気電子学会（IEEE）「倫理的に調整された設計」（第1エディション）
- Future of Life Institute「アシロマ原則」³⁴
- Partnership on AI「信条」³⁵

³³ 報告書 2017,13 頁参照。

³⁴ 報告書 2017,20 頁参照。

³⁵ 報告書 2017,17 頁参照。

第2章 AI 利活用ガイドライン策定の考え方

1. 背景・経緯

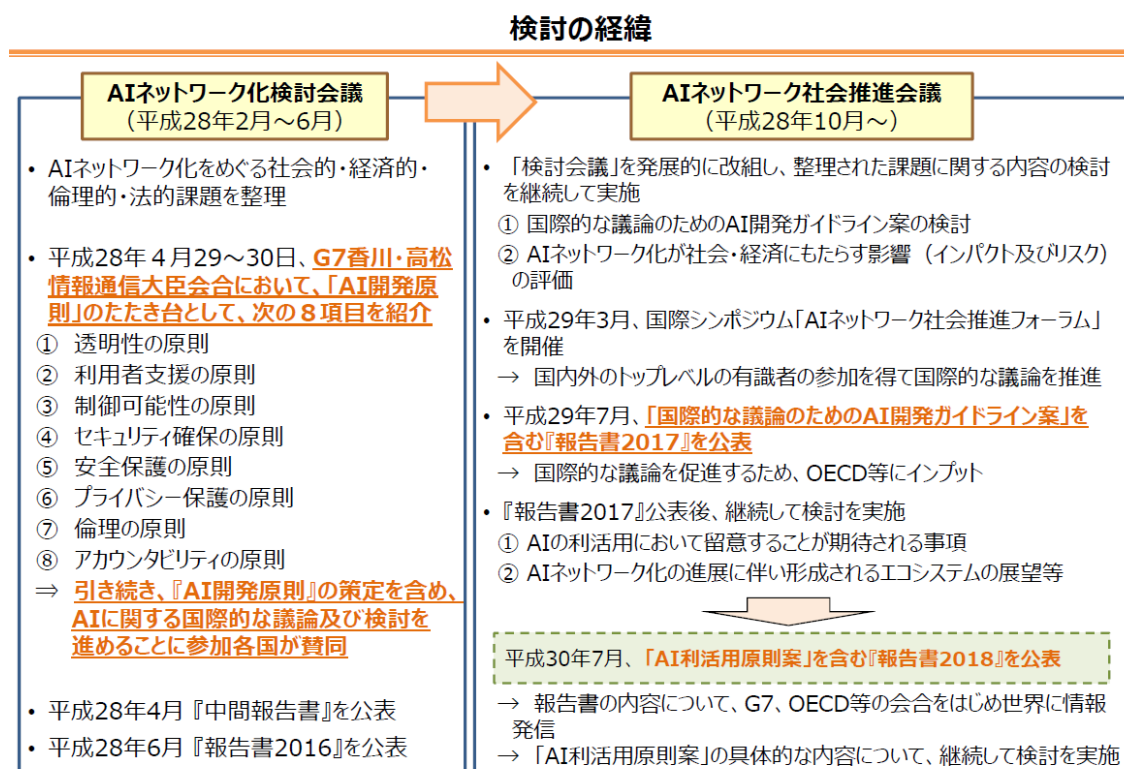


図 1：AI 開発ガイドライン案／AI 利活用ガイドライン検討の経緯

(1) 背景・経緯

2016年（平成28年）4月29日及び30日に開催されたG7香川・高松情報通信大臣会合において、高市早苗総務大臣（当時）から、G7各国が中心となり、OECD等国際機関の協力も得て、AIネットワーク化が社会・経済に与える影響やAIの開発原則の策定などAIネットワーク化をめぐる社会的・経済的・倫理的課題に関し、産学民官の関係ステークホルダの参画を得て、国際的な議論を進めることの提案がなされ、各国から賛同が得られた。

同年6月20日に、AIネットワーク化検討会議の報告書2016が公表され、今後の課題のうち第一の課題として「研究開発の原則・指針の策定」が提言された。

同年10月31日に、検討会議を発展的に改組した本推進会議が発足し、第1回会合が開催された。本推進会議には、親会の下に開発原則分科会及び影響評価分科会が置かれ、このうち主に開発原則分科会において、AI開発ガイドライン案の作成に向けた検討を行った。同年12月28日から2017年（平成29年）1月31日にかけて行ったAI開発ガイドライン

案の論点に関する意見募集に寄せられた意見、同年3月13日及び14日に開催されたAIネットワーク社会推進フォーラム（国際シンポジウム）における議論及び同年6月14日から同年7月7日にかけて行った報告書2017（案）に関する意見募集に寄せられた意見等を踏まえ、同年7月28日に、AI開発ガイドライン案を含む報告書2017が公表された。

他方、AIは、利活用の過程でデータの学習等により自らの出力やプログラムを継続的に変化させる可能性があることから、AIの利活用においても留意することが期待される事項があるものと想定される。また、AIの利活用においても留意することが期待される事項を整理することは、開発者と利用者、データ提供者などステークホルダ間の関係を考える上でも重要であるものと考えられる。

このような問題意識を踏まえ、上記報告書2017の公表後、本推進会議において、AIの利活用において留意することが期待される事項に関する検討を行うとともに、その検討に資するよう、AIネットワーク化が進化した社会の将来像を描きAIの利活用における課題等を抽出するために、AIネットワーク化が社会・経済にもたら影響（分野別評価）やAI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に関する検討、AIネットワーク化の進展に関する考察等を行った。また、これらの事項は密接に関連するものと考えられることから、主として親会の下に置かれた環境整備分科会及び影響評価分科会の合同分科会において検討を行った。その上で、2018年（平成30年）6月12日から同月26日までにかけて行った報告書2018（案）に関する意見募集に寄せられた意見を踏まえ、同年7月17日、AI利活用原則案を含む報告書2018が公表された。

報告書2018では、G7やOECD等において、AI開発ガイドライン（仮称）と併せてAI利活用ガイドライン（AIの利活用に関する指針）の策定に向けた国際的な議論が進められることが期待される旨が提言され、また、AI利活用ガイドラインの策定を優先課題の1つと位置付けている。AI利活用ガイドラインの策定にあたっては、AIの利活用に関する国内外の議論の動向を踏まえて、報告書2018でとりまとめたAI利活用原則案を必要に応じて改定するとともに、各原則の内容に関し整理された論点を踏まえて、最終的なアウトプットの取りまとめに向けた検討を進めることが必要である旨があわせて提言されている。

2018年5月には、内閣府において、AIを有効に活用して社会に便益をもたらしつつ、ネガティブな側面を事前に回避又は低減することを目的に、人、社会システム、産業構造、イノベーションシステム、ガバナンス等、あらゆる面で社会をリデザインし、AIを有効かつ安全に利用できる社会、すなわち「AI-Readyな社会」への変革を推進する観点から、AI社会原則の策定に向けた検討が開始され、2019年（平成30年）3月に「人間中心のAI社会原則」がとりまとめられた。

このように、我が国におけるAIの原則についての検討は、開発から利活用へ、そして社会へと発展的に検討が進められてきた。

また、我が国は、国際社会においてもこれらの検討成果を積極的に発信し、特に、OECD

が理事会勧告案を策定するために2018年9月に設置した専門家会合（AI expert Group at the OECD）においては、我が国から参加した有識者より、AI開発ガイドライン案、AI利活用原則案、AI社会原則のそれぞれについて、その内容のみならず、検討の背景や検討にあたり行われた議論の状況についても紹介するなど、OECD理事会勧告の原案策定にあたり大きな貢献を行った。このため、2019年（令和元年）5月に公表されたOECD理事会勧告（Recommendation on Artificial Intelligence）は、我が国において検討してきた上述の各原則、ガイドラインと整合のとれたものとなっている。

このように、AIに関する原則については、我が国が国際的な議論を主導してきた結果、その概念については、概ね国際的にコンセンサスが得られつつある。他方で、今後は、この「原則」をどのように実現していくべきか、そのためには、各ステークホルダがどのような措置を講ずるべきかといった、具体的な措置について検討が進められることが期待されている。

（２）本推進会議における検討

報告書2018における提言や国際的な議論の動向を踏まえて、2018年（平成30年）10月、本推進会議の下に「AIガバナンス検討会」を設置し、AI利活用ガイドラインの策定に向け、集中的に検討を行ってきた。報告書2018では、AI利活用原則案として整理された10の原則のそれぞれについて論点が掲げられており、その論点についての理解を深める観点から、各原則に知見を有する様々な専門家に講演をしていただき、原則を実現するための具体策の検討を行った

また、上述のとおり、AIに関する原則についてはOECD、G7及びEUなどでも精力的に議論が行われており、国際的にも共有しうる実践的なガイダンスを提言することを目指し検討を実施した。

2. AI利活用ガイドラインの位置づけ

（１）民間等で原則等を議論する際の参照

上記の検討は、内閣府の「人間中心のAI社会原則」の検討と並行して行われた。内閣府の検討は、AI開発ガイドライン案やAI利活用原則案をはじめ、国内における倫理的な検討の成果を踏まえた上で行われており、本年3月に公表されたAI社会原則は、AI開発ガイドライン案やAI利活用原則案とも整合性の取れたものとなっている。内閣府がとりまとめた「人間中心のAI社会原則」では、下図に示すとおり、基本理念、ビジョンを述べた上で社会（特に、国などの立法・行政機関）が留意すべき「AI社会原則」と、AIの研究開発と社会実装に従事する開発・事業者側が留意すべき「AI開発利用原則」に体系化されており、AI開発利用原則は、開発者及び事業者において、基本理念及び上記のAI社会原則を踏まえたAI開発利

用原則を定め、遵守すべきものと整理された。本推進会議で検討を行ってきたAI開発ガイドライン案やAI利活用ガイドラインは、開発者及び事業者がAI開発利用原則を策定する際の参考とすべきものと位置づけることができ、今後、事業者等が原則等を検討する際に同ガイドラインを参照することが期待される。

なお、AIの利活用は多様な分野に及ぶものであり、分野ごとにAIのもたらす便益やリスクが異なる可能性がある。AI利活用ガイドラインでは、AIを利用する用途（ドメイン）を限定せず、分野に共通する事項について定めたものであり、分野ごとの事情に応じて留意されることが期待される事項については、同ガイドラインを参照しつつ、策定の可否を含め検討が行われることが期待される。

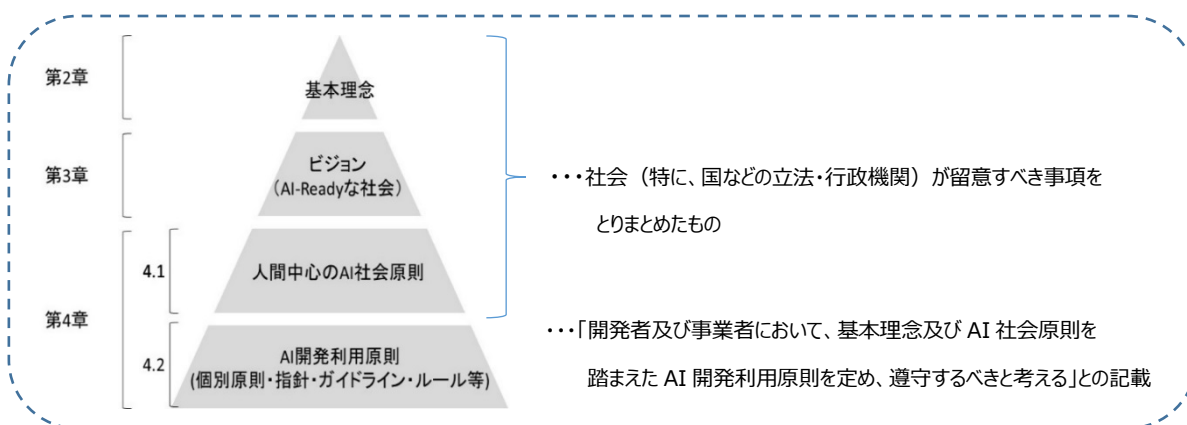


図 2 : 「人間中心の AI 社会原則」 構成³⁶

(2) 国際的な議論への継続的な貢献

第1章にも記載のとおり、AI原則の各項目については、国際的にほぼコンセンサスが得られつつあり、今後は、原則の実効性を確保するための具体的手段についての議論に移行することが見込まれる。欧州委員会が2019年（平成31年）4月に公表した「信頼できるAIのための倫理ガイドライン」においても、AIの原則に相当する7つの要求条件を実現するためのチェックリストが設けられており、また、OECDも、同年5月に公表した理事会勧告を踏まえ、その内容を具体化するプラクティカルガイダンスの検討を実施する旨の表明を行っている。本推進会議で検討してきたAI利活用ガイドラインは、こうした国際的な議論の流れと軌を一にするものであり、今後、国際的な議論において認識の共有を図っていくことが期待される。

³⁶ 本図は「人間中心のAI社会原則」（第1章1.（2）参照）より引用した全体構成図に注釈を加えたものである。

3. AI 利活用ガイドラインの概要

本推進会議における検討を踏まえ、AI 利活用において留意されることが期待される事項として AI 利活用原則を定め、その解説を記載した AI 利活用ガイドラインを作成した（別紙1）。AI は様々な分野で利用され、分野ごとに AI システムのもたらす便益やリスクは異なることが想定されるところ、同ガイドラインは、各分野に共通する項目を整理したものである。

なお、AI 利活用ガイドラインでは、目的・基本理念について述べた後、関係する主体について整理し、その上で、主体ごとの一般的な AI 利活用の流れについて述べている。続いて、10 の AI 利活用原則とそれぞれに関する主な論点について述べた後で、前述の AI 利活用の一般的な流れの中で考慮することが期待されるタイミングについて述べている。最後に、前述の各論点に対する詳説をまとめている。以下、AI 利活用ガイドラインの概要を記すとともに、必要に応じて、趣旨を解説する。


パート1：AI利活用原則の考え方

1. 目的
AIの利活用・社会実装の促進

2. 基本理念

- ・ 人間中心の社会の実現
- ・ AI利活用における多様性の尊重・包摂
- ・ AIネットワーク化による持続可能な社会の実現
- ・ 便益とリスクの適正なバランスの確保
- ・ 利用者間の知識・能力相応の役割分担
- ・ 指針やベストプラクティスの国際的な共有
- ・ 不測の見直し・柔軟な改定


3. 関係する主体の整理
「利用者」の分類



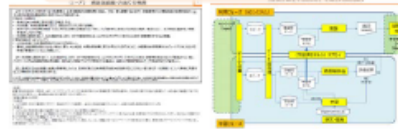
4. AI利活用原則
10原則
適正利用、適正学習、連携、安全、セキュリティ、プライバシー、
尊厳・自律、公平性、透明性、アカウントビリティ

パート2：AI利活用原則の解説

5. 一般的なAI利活用の流れ
AI利活用の一般的な流れ



6. AI利活用原則の解説
10原則の各論点に関する解説
(およびその詳説 (附属資料))



7. AI利活用原則を考慮すべきタイミング
各原則各論点を考慮すべきタイミング




図 3：AI 利活用ガイドラインの構成

(1) 目的

AI 利活用ガイドラインは、AI の利活用や社会実装を促進することを目的とし、AI ネットワーク化による国境を越えた便益及びリスクの波及などを踏まえ、AI に対する信

頼を醸成するために留意することが期待される事項を整理したものである。

以上を踏まえ、AI利活用ガイドラインの目的を「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り、AIに対する信頼を醸成することにより、AIの利活用や社会実装を促進する。」としている。

なお、AIの利活用の目的や用途は様々であることから、必ずしもあらゆるAIの利活用において利活用原則として整理された10の原則の全てに留意することが期待されるわけではない。AIの利用者は、AIの利用目的や利活用する際の社会的文脈を踏まえ、10の原則の中から自らが留意すべき原則を選定し、その上で、選定したそれぞれの原則について、後に記載する解説を参照し、どのような措置をどの程度講ずべきか自主的に検討することが想定されている³⁷。

(2) 基本理念

① 基本理念の内容

上記(1)の目的に鑑み、次に掲げる理念を一体的に基本理念として掲げている(概要のみ)。

- 人間中心の社会の実現
- AIの利活用における多様性の尊重・包摂
- AIネットワーク化による持続可能な社会の実現
- 便益とリスクの適正なバランスの確保
- 利用者間の知識・能力相応の役割分担
- 指針やベストプラクティスの国際的な共有
- 不断の見直し・柔軟な改定

② 「人間中心のAI社会原則」との関係

上述した基本理念について、「人間中心のAI社会原則」における「2. 基本理念」、「3. ビジョン」との関係を整理しておく。本推進会議で検討しているAI利活用ガイドラインは、AIの利用者の視点からAIの便益の増進とリスクの抑制を図ることを目的としている一方、人間中心のAI社会原則は、AIが社会に受け入れられ適正に利用されるため、社会(特に国などの立法・行政機関)が留意すべき事項を整理したものであり、自ずと議論の対象が異なる。すなわち、AI社会原則における「公正競争確保の原則」や「イノベーションの原則」は一利用者としての取組みを超えた社会として留意すべき事項であり、また、これらの原則

³⁷ AIの研究開発において留意することが期待される事項については、2017年(2019年)7月に「AI開発ガイドライン案」としてとりまとめられているところである。開発と利活用とは、必ずしも明確に区分できない場合も想定され、AI開発ガイドライン案とセットとして参照されることが望ましい。

と関連の深い「産業構造」や「イノベーションシステム」といったビジョンは、AI 利活用ガイドラインの検討の対象外である。以上を踏まえ、AI 利活用ガイドラインにおける基本理念と、人間中心の AI 社会原則における「2. 基本理念」、「3. ビジョン」との関係を整理した結果は図4のとおりとなり、両者は概ね一致していることが確認できる。なお、以下の表1は、「4.1 AI 社会原則」、「5. おわりに」における関連する記述も含め対応関係を整理したものである。

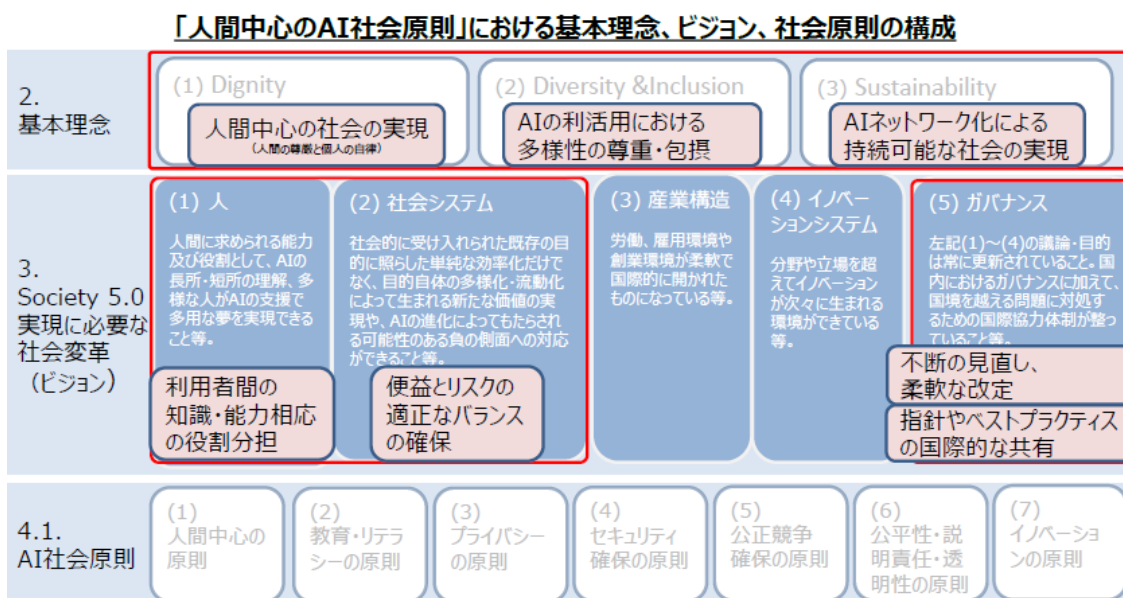


図4：「人間中心の AI 社会原則」の基本理念、ビジョン、AI 社会原則と AI 利活用ガイドラインの基本理念の対応

表1：基本理念と「人間中心の AI 社会原則」における言及との関係

基本理念	「人間中心の AI 社会原則」における言及
人間中心の社会の実現	• 2.基本理念－(1) 人間の尊厳が尊重される社会 (Dignity)
AI の利活用における多様性の尊重・包摂	• 2.基本理念－(2) 多様な背景を持つ人々が多様な幸せを追求 できる社会 (Diversity & Inclusion)
AI ネットワーク化による持続可能な社会の実現	• 2.基本理念－(3) 持続性ある社会 (Sustainability)

利用者間の知識・能力 相応の役割分担	<ul style="list-style-type: none"> 3.ビジョン-(1) 人 : 人間に求められる能力及び役割として、AIの長所・短所の理解、多様な人がAIの支援で多様な夢を実現できること等。 4.1.AI 社会原則-(1) 人間中心の原則 : AI の開発・提供・利用に関わった種々のステークホルダが適切に分担して責任を負うべきである。
便益とリスクの適正な バランスの確保	<ul style="list-style-type: none"> 3.ビジョン-(2)社会システム : 社会的に受け入れられた既存の目的に照らした単純な効率化だけでなく、目的自体の多様化・流動化によって生まれる新たな価値の実現や、AI の進化によってもたらされる可能性のある負の側面への対応ができること等。 4.1.AI 社会原則-(3) プライバシー確保の原則 : その (パーソナルデータの) 利活用と保護のバランスについては、文化的背景や社会の共通理解をもとにきめ細やかに検討される必要 4.1.AI 社会原則-(4) セキュリティ確保の原則 : 社会は、常にベネフィットとリスクのバランスに留意
指針やベストプラクティスの 国際的な共有	<ul style="list-style-type: none"> 3.ビジョン-(5)ガバナンス : 国境を越える問題に対処するための国際協力体制が整っていることが求められる。 5. おわりに : 国際的な議論の場において、我が国は、本原則を世界各国と共有した上で、国際的な議論のリーダーシップをとり、コンセンサスの形成を目指すべき
不断の見直し、 柔軟な改定	<ul style="list-style-type: none"> 3. ビジョン-(5)ガバナンス : 社会情勢の変化や技術の進展に伴い (中略) 常に更新し続ける必要がある。 5. おわりに : 本原則は、今後、AI 関連技術の進展、社会の変化、世界の情勢等に応じて、今後柔軟に進化・発展させるものである。

(3) 関係する主体の整理

① 基本的な考え方

AI 開発ガイドライン案では、開発者、利用者をそれぞれ以下のように定義していた。

- 開発者 : 「自らの開発した AI システムを用いて AI ネットワークサービスを他者に提供するプロバイダを含め、AI システムの研究開発を行う者
- 利用者 : 最終利用者 (エンドユーザ) のほか、他者が開発した AI システムを用いて AI ネットワークサービス (情報通信ネットワークを通じて AI システムの機能を提供するサービスをいう。以下同じ。) を第三者に提供するプロバイダを含む

報告書 2018 で AI 利活用原則案をとりまとめるにあたっては、AI の利活用において様々な立場の利用者の関与が想定されることを踏まえ、利用者についてさらに細分化し、「AI サービスプロバイダ」と「最終利用者」に大別した上で、「最終利用者」を「ビジネス利用者」と「消費者的利用者」に分類した。

報告書 2018 では、「ビジネス利用者」を、「最終利用者のうち業として AI システム又は AI サービスを利用する者」と定義しているが、その後の AI ガバナンス検討会における議論の過程で、ビジネス利用者には

- (A) AI の学習等を含む運用等を自ら行う者
- (B) サービスプロバイダ等から提供された AI システム又は AI サービスを業として自ら運用等を行うことなく利用するのみの者

の両者が含まれ、それぞれに対し期待されることは異なるものとなるのではないかと、特に後者は「消費者的利用者」に近いのではないかと指摘があった。

また、「消費者的利用者」についても、一部のアーリーアダプターについては、むしろ開発者等に類似するのではないかと指摘があった。

こうした指摘を踏まえ、「ビジネス利用者」と「消費者的利用者」について、それぞれ上記(A)、(B)の通り再定義することについて検討を行った。しかし、(B)のように、AI に関する十分な知見を有していないビジネス利用者であっても、社会生活上の地位に基づき反復継続して活動（業務）を行っており、消費者と同様に消費者的利用者と位置づけるのは適切ではないとの指摘があった。このため、「ビジネス利用者」と「消費者的利用者」については、報告書 2018 の定義を踏襲することとし、それぞれ以下の注記を付すこととした。

- 「ビジネス利用者」の注記：
「ビジネス利用者であっても、AI システム又は AI サービスについて、自ら運用等を行うことなく利用するのみの者については、AI サービスプロバイダや他のビジネス利用者と同等の留意を期待することが困難であることも想定されるが、その場合でも、開発者や AI サービスプロバイダに対し、適切な措置を依頼することが求められる。」
- 「消費者的利用者」の注記：
「消費者的利用者であっても、AI システム又は AI サービスについて、自ら学習等を含む運用等を行う場合には、開発者や AI サービスプロバイダと同等の留意が求められる場合がある。」

② 関係する主体の整理

以上のような検討を踏まえ、AI 利活用ガイドラインでは、AI の利活用において関与が想定される者を次のように整理した。

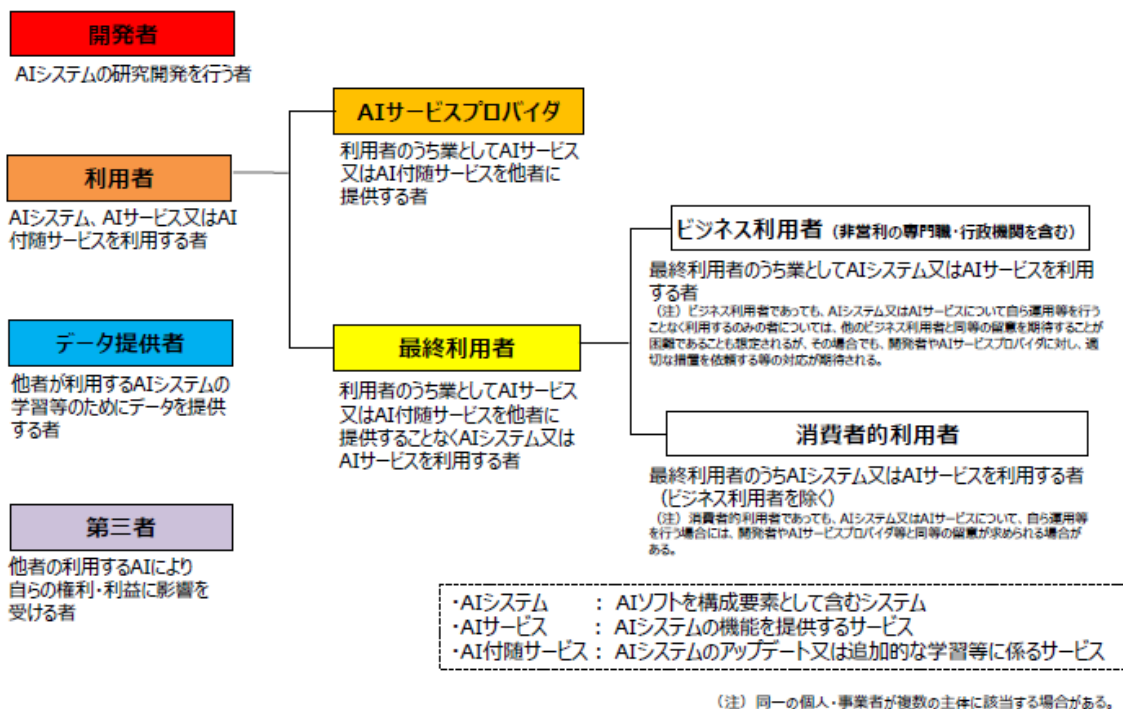


図 5：関係する主体の整理

- 開発者：
AI システムの研究開発を行う者
- 利用者：
AI システム、AI サービス又は AI 付随サービスを利用する者
- AI サービスプロバイダ³⁸：
AI 利用者のうち業として AI サービス又は AI 付随サービスを他者に提供する者
サービスプロバイダ
- 最終利用者：
利用者のうち業として AI サービス又は AI 付随サービスを他者に提供することなく AI システム又は AI サービスを利用する者
- ビジネス利用者（非営利の専門職、行政機関を含む）：
最終利用者のうち業として AI システム又は AI サービスを利用する者。

³⁸ 報告書 2018 では、AI サービスプロバイダを AI ネットワークサービスプロバイダとオフライン AI サービスプロバイダに分類していたが、AI のネットワーク化が今後さらに進展していくことや、ガイドラインの記載が煩雑になることを避けるべきこと、また、オフライン AI サービスプロバイダも AI ネットワークサービスプロバイダに関する留意事項をもとに取捨選択すれば足りることなどを踏まえ、報告書 2019 では、両者の分類を行わないこととした。

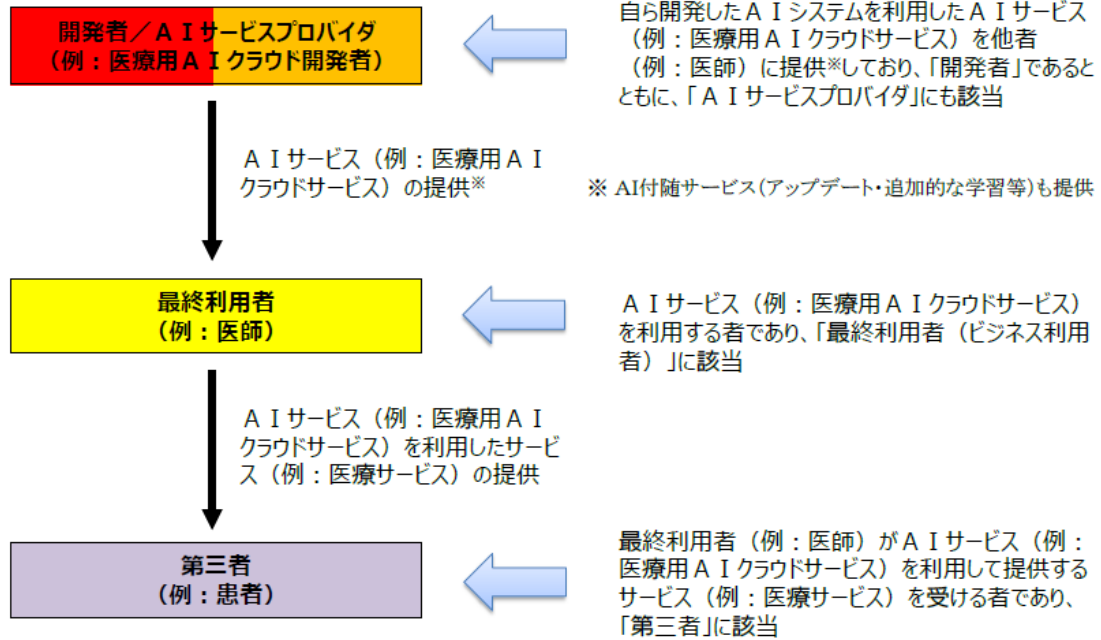
なお、ビジネス利用者であっても、AI システム又は AI サービスについて自ら運用等を行うことなく利用するのみの者については、他のビジネス利用者と同等の留意を期待することが困難であることも想定されるが、その場合でも、開発者や AI サービスプロバイダ等に対し、適切な措置を依頼する等の対応が期待される。

- 消費者的利用者：
最終利用者のうち AI システム又は AI サービスを利用する者（ビジネス利用者を除く）。
ただし、消費者的利用者であっても、AI システム又は AI サービスについて、自ら運用等を行う場合には、開発者や AI サービスプロバイダ等と同等の留意が求められる場合がある。
- データ提供者：
他者が利用する AI システムの学習等のためにデータを提供する者
- 第三者³⁹：
他者の利用する AI により自らの権利・利益に影響を受ける者（報告書 2018 で整理した「間接利用者」（最終利用者が AI システム又は AI サービスを利用して提供するサービス等を受ける者）を含む）

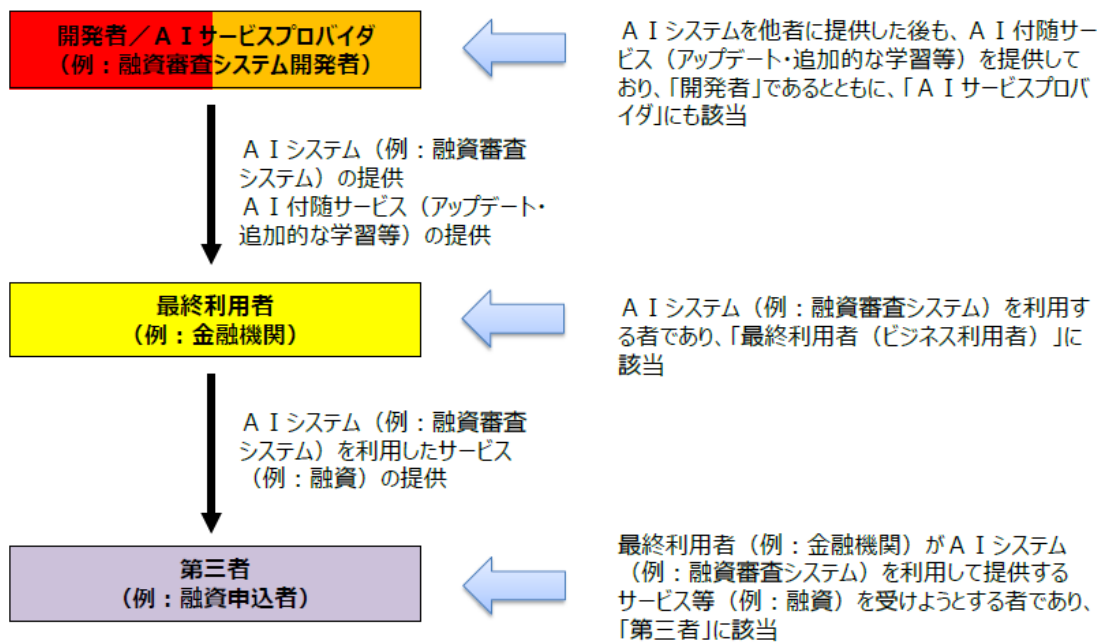
このような整理を踏まえて、AI の利活用の事例に即した具体例は次のとおりである。

³⁹ 報告書 2018 では、「最終利用者が AI システム又は AI サービスを利用して提供するサービス等を受ける者」を「間接利用者」として分類していたが、「第三者」と「間接利用者」をあえて区分して間接利用者に留意すべき事項を整理するまでの必要性も認められないことから、報告書 2019 では間接利用者を主体として明示せず、第三者の中に入れて考えることとした。

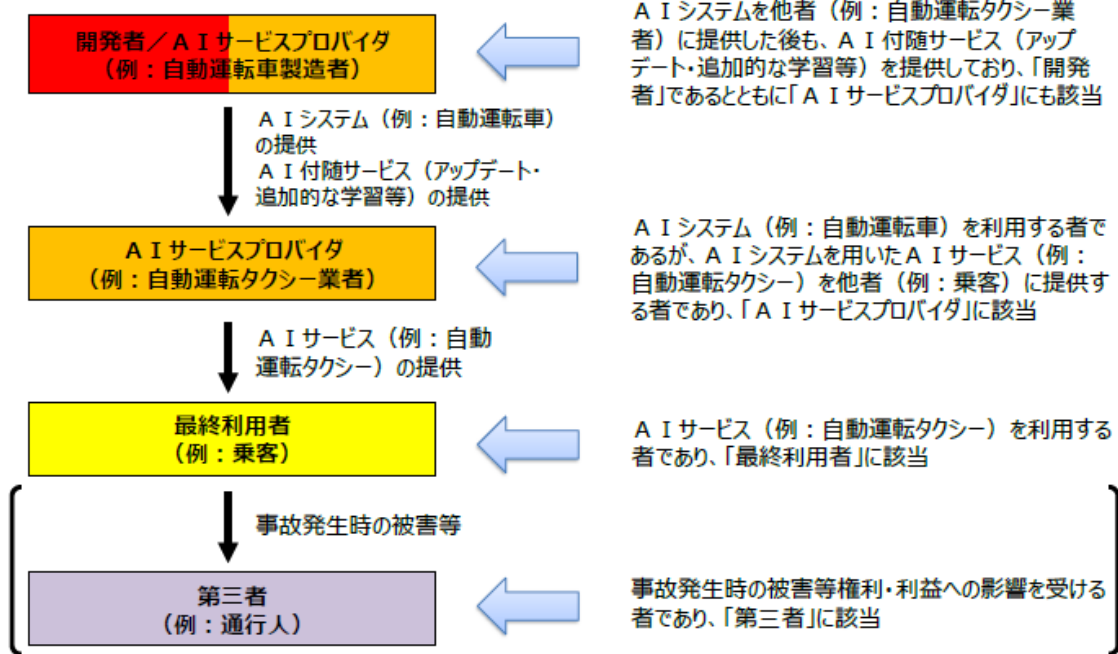
<例①：医療用クラウド>



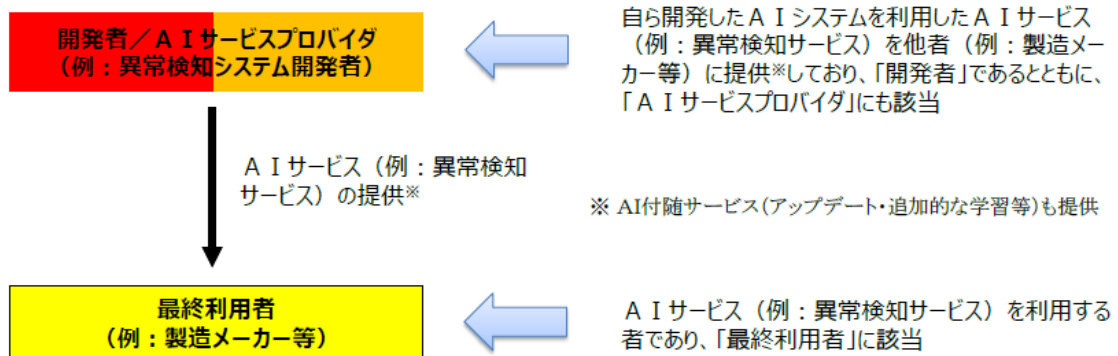
<例②：融資審査>



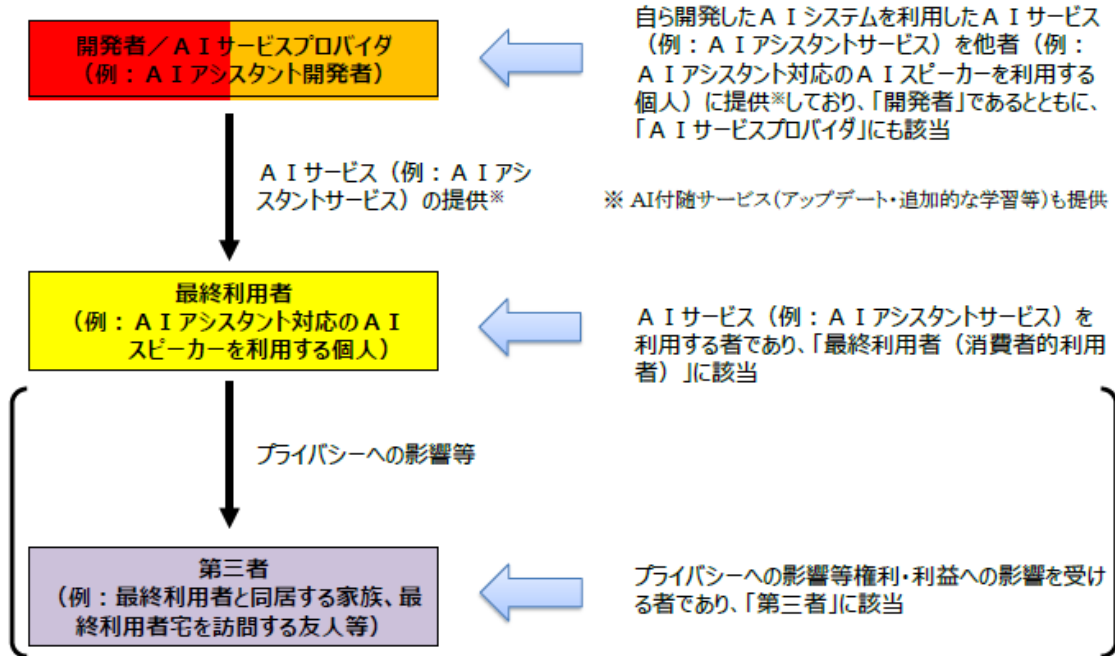
<例③：自動運転タクシー>



<例④：生産ラインの異常検知>



<例⑤ : AI アシスタント (AI スピーカー) >



(4) AI 利活用原則と論点

報告書 2018 においては AI の利活用について留意することが期待される事項を「AI 利活用原則案」として 10 原則を掲げるとともに、原則ごとに論点の整理を行った。その後の AI ガバナンス検討会及び本推進会議での検討の結果、AI 利活用原則案で掲げた 10 原則については、「AI 利活用原則」としてとりまとめることが適当であるとの結論に至るとともに、原則ごとの論点についても必要な見直しを行い、基本理念との関係を整理したほか、新たな論点の追加を行うなどの修正を行った上で下記のとおり整理することとした。

1. **適正利用の原則** : 利用者は、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。

(主な論点)

- ア 適正な範囲・方法での利用
- イ 人間の判断の介在
- ウ 関係者間の協力

2. **適正学習の原則**：利用者及びデータ提供者は、AIシステムの学習等に用いるデータの質に留意する。
 (主な論点)
 ア AIの学習等に用いるデータの質への留意
 イ 不正確又は不適切なデータの学習等によるAIのセキュリティ脆弱性への留意
3. **連携の原則**：AIサービスプロバイダ、ビジネス利用者及びデータ提供者は、AIシステム又はAIサービス相互間の連携に留意する。また、利用者は、AIシステム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。
 (主な論点)
 ア 相互接続性と相互運用性への留意
 イ データ形式やプロトコル等の標準化への対応
 ウ AIネットワーク化により惹起・増幅される課題への留意
4. **安全の原則**：利用者は、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、利用者等及び第三者の生命・身体・財産に危害を及ぼすことがないよう配慮する。
 (主な論点)
 ア 人の生命・身体・財産への配慮
5. **セキュリティの原則**：利用者及びデータ提供者は、AIシステム又はAIサービスのセキュリティに留意する。
 (主な論点)
 ア セキュリティ対策の実施
 イ セキュリティ対策のためのサービス提供等
 ウ AIの学習モデルに対するセキュリティ脆弱性への留意
6. **プライバシーの原則**：利用者及びデータ提供者は、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。
 (主な論点)
 ア AIの利活用における最終利用者及び第三者のプライバシーの尊重
 イ パーソナルデータの収集・前処理・提供等におけるプライバシーの尊重
 ウ 自己等のプライバシー侵害への留意及びパーソナルデータ流出の防止
7. **尊厳・自律の原則**：利用者は、AIシステム又はAIサービスの利活用において、人間の尊厳と個人の自律を尊重する。
 (主な論点)
 ア 人間の尊厳と個人の自律の尊重
 イ AIによる意思決定・感情の操作等への留意
 ウ AIと人間の脳・身体を連携する際の生命倫理等の議論の参照

エ AI を利用したプロファイリングを行う場合における不利益への配慮

8. **公平性⁴⁰の原則**：AI サービスプロバイダ、ビジネス利用者及びデータ提供者は、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が不当に差別されないよう配慮する。

(主な論点)

ア AI の学習等に用いられるデータの代表性への留意

イ アルゴリズムによるバイアスへの留意

ウ 人間の判断の介在（公平性の確保）

9. **透明性の原則⁴¹**：AI サービスプロバイダ及びビジネス利用者は、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。

(主な論点)

ア AI の入出力等のログの記録・保存

イ 説明可能性の確保

ウ 行政機関が利用する際の透明性の確保

10. **アカウントビリティ⁴²の原則**：利用者は、ステークホルダに対しアカウントビリティを果たすよう努める。⁴³⁴⁴

(主な論点)

ア アカウントビリティを果たす努力

イ AI に関する利用方針の通知・公表

40 「公平性」には複数の定義・基準があることに留意する必要がある。

41 本原則は、アルゴリズム、ソースコード、学習データの開示を想定するものではない。また、本原則の解釈に当たっては、プライバシーや営業秘密への配慮も求められる。

42 判断の結果についてその判断により影響を受ける者の理解を得るため、判断に関する正当な意味・理由を説明したり、(必要に応じて) 賠償・補償したりする等の措置をとること。

43 AI 利活用ガイドラインは、AI を利活用している者を対象にしているが、将来的には、AI を利用していない者であっても、社会通念に照らし AI を利用していることが期待されるにもかかわらず、AI を利用していない場合には、なぜ AI を利用しないのかにつき、消費者的利用者等に対して説明を行うことが求められる場合も想定される。

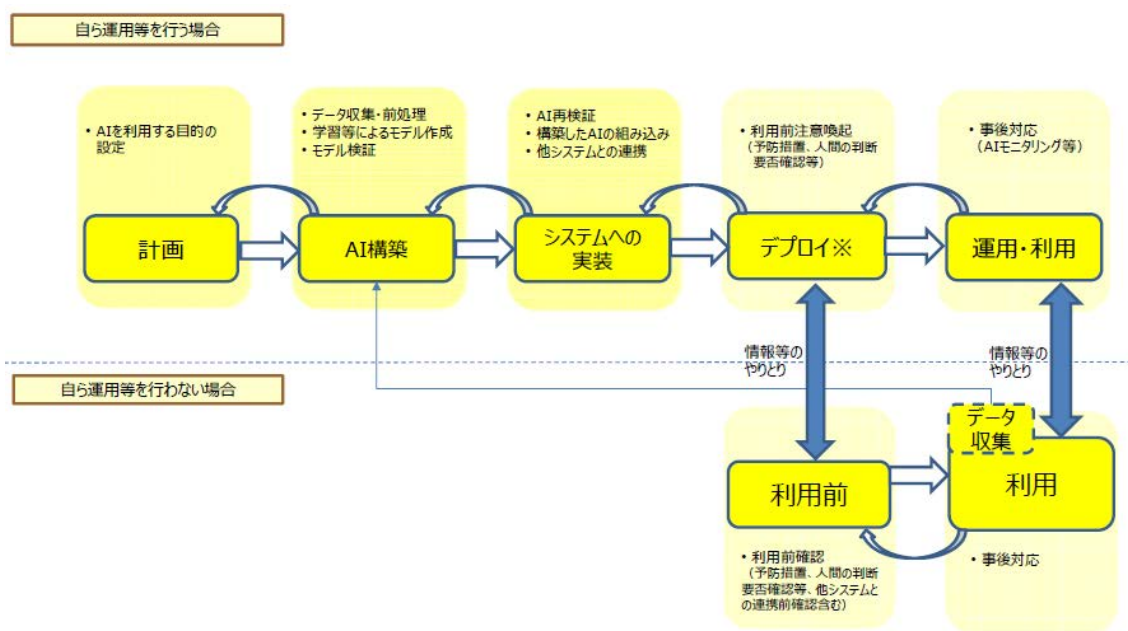
44 AI サービスプロバイダ及びビジネス利用者が、AI に関する利用方針を公表することが求められるのは、利活用する AI の判断が、消費者的利用者や第三者に直接の影響を及ぼす場合であると考えられる。すなわち、人間の思考に供するための分析道具として AI を利活用するにとどまる場合や、AI が原案を作成しつつも、最終的に人間が判断することが実質的に担保されている場合には、AI に関する利用方針の公表が必ずしも求められるわけではない。(もっとも、そうした場合であっても、自主的に公表されることが望ましい。)

(5) 一般的な AI 利活用の流れ

AI の利活用は、AI サービスプロバイダによるサービスの提供から、最終利用者による AI サービスの受領までの様々なプロセスまでを対象とするものであり、報告書 2018 で整理した AI 利活用原則案の各論点の理解を深めるためには、AI システム又は AI サービスを利活用するに当たり、どのような利活用プロセスを経ることになるのかを「フェーズ」として整理した上で、どのフェーズでどの論点が問題となるのかを明らかにしておくことが有益である。このため、一般的な AI 利活用の流れについて、自ら AI システムや AI サービスの運用等を行う場合、「計画」「AI 構築」「システムへの実装」「デプロイ」「運用・利用」の 5 つのフェーズに整理し、自ら運用等を行わず利用するのみの場合については、「利用前」「利用」の 2 つのフェーズに整理することとした。

- ① 自ら AI システム又は AI サービスの運用等（学習等を含む）を行う場合
- ② 自ら AI システム又は AI サービスの運用等を行うことなく利用する（自ら運用等を行わない場合）

それぞれの場合における AI を利活用の流れを図示したのが図 6 である。



※デプロイ：（AI ソフト／システムを）利用可能な状態にすること

図 6：主体ごとの一般的な AI 利活用の流れ⁴⁵

⁴⁵ ここで図示する AI 利活用の一般的な流れは、第 7 章において、利活用のどのフェーズで AI 利活用原則のどの論点を考慮すべきかにかんしての整理を行うために、典型的な事例を記載したものである。他方、AI の利活用は、本図にあるとおり各フェーズが時系列で整理される場合だけではなく、DevOps のように開発と運用が一体のものとして検討される場合など、多様なケースが想定される。そのため、AI 利活用ガイドラインの活用にあたっては、それぞれの利活用の流れに応じて必要な読み替えを行うことが期待される。

①、②の各主体における AI 利活用の一般的な流れに沿った「フェーズ」の内容の説明は以下の通りである。

① 自ら運用等を行う場合

- A) 計画フェーズ：
本フェーズは AI を利活用する目的を設定し、大まかにどのようなデータを利用するかなどを検討するフェーズである。
- B) AI 構築フェーズ：
本フェーズは AI ソフトを構築し、トライアルを通じて検証を行うフェーズである。データの収集、前処理及び学習によるモデルの作成、検証などの作業等が行われる。
- C) システム実装フェーズ：
本フェーズは上記フェーズ B で作成された AI ソフトをシステムに導入し、検証を行うフェーズである。ここで言うシステムには既存、新設の両方のケースが想定される。また、他のシステムとの連携を検証することも想定される。
- D) デプロイフェーズ：
本フェーズはフェーズ C で作成された AI を消費者的利用者等（自身を含む）が利用可能な状態にするフェーズである。利用可能な状態にするのに際し、消費者的利用者等に情報提供する等が想定される。
- E) 運用／利用フェーズ：
本フェーズは、消費者的利用者等に対し、デプロイされた AI を運用するフェーズである。AI に与えられたデータに基づき自律的に変化すること等を踏まえ AI をモニタリングする、また、消費者的利用者等からの問い合わせに対応する等が想定される。

② 自ら運用等を行わない場合

- A) 利用前フェーズ：
本フェーズは AI を利用する前のフェーズである。利用に当たり、開発者、AI サービスプロバイダ等から提供された情報を把握する等が想定される。
- B) 利用フェーズ：
本フェーズは、AI を利用するフェーズである。開発者、AI サービスプロバイダ等から提供された情報に基づき、利用し、必要なアップデート等を行うなどが想定される。

以上、一般的な AI 利活用の流れを記載したが、例外も存在し、その場合には必要な読み替えを行うことが期待される。例えば、AI サービスとして、機械学習モデルを通

じて何らかの判断を行うような単体のソフトウェアを消費者的利用者等に提供する場合は、システムの実装を行っていないことから、①のフェーズ C については、他のシステムとの連携のみを検証すると読み替えることとなる。

(6) AI 利活用原則の解説

報告書 2018 で整理した原則ごとに掲げられた論点について検討を行い、論点ごとに利用者が留意すべき事項に関する解説を行った。その際、報告書 2018 において「利用者の分類（AI サービスプロバイダ、最終利用者等）によって有している能力や知識の多寡、必要な措置を実施する容易性等が異なることを踏まえて、利用者の分類を意識しつつ、留意することが期待される事項等の具体的な内容を検討する必要がある」との記載があることを踏まえ、AI 利活用ガイドラインでは、各原則の論点ごとに、「AI サービスプロバイダ、ビジネス利用者及びデータ提供者」が留意すべき事項についての詳細な解説をとりまとめることとした。また、参考として「消費者的利用者」留意することが望ましい事項についてもあわせて解説を記載した。

②-ア) AIの学習等に用いるデータの質への留意（1/2）

9

<p>AIサービスプロバイダ、ビジネス利用者及びデータ提供者は、利用するAIの特性及び用途を踏まえ、AIの学習等に用いるデータの質（正確性及び完全性など）に留意することが期待される。特に機械学習においては、以下の方法によりデータの質を確保することが考えられる。</p> <p>[データ収集時の対策（例）]</p> <ul style="list-style-type: none"> • 収集するデータがAIの利用目的に適ったものかを確認する。 • 社会的に信用の高い者が公開するデータを活用する。 • データの作成履歴を確認した上で収集する。 • 自らデータを収集する際には、データに付随する権利に留意する。 <p>[データ前処理時の対策（例）]</p> <ul style="list-style-type: none"> • 人間でも判定が困難と考えられるデータは、学習等の対象から除外する¹⁾。 • 機械（学習器）が誤認識しやすいと考えられるデータは積極的に学習の対象とする²⁾。 • （特に教師あり学習等で）アノテーション（ラベル付与）を行う際には、誤って行わないよう留意する。 • 利用時に利用（入力）されるデータの形式を意識してデータセットを作成する。 • 前処理をどのように行ったのか（データ前処理に関する来歴）について、ログを取得し保存する。 <p>[学習時の対策（例）]</p> <ul style="list-style-type: none"> • 既存の学習モデルを利用して転移学習³⁾等を行う。 • 学習の精度を上げるため、特定のデータを拡張した上で学習を行う。 • 一過性のある時系列データを学習する場合などは、学習対象とすべきデータの範囲を適切に画定する。 <p>1) 例えば、画像認識などで、対象となるオブジェクトが人間の目でも見ても同定できない場合など。 2) 例えば、画像認識などで、対象となるオブジェクトが稀にあるなど。 3) 転移学習(Transfer Learning)とは、深層学習を含む機械学習で用いられる技術の1つで、特定の領域（ドメイン）で学習させたモデルを別の領域に適用する技術である。少ないデータで精度の高い学習結果を得ることが出来る可能性がある点がメリットである。 4) 「データの拡張」(Data Augmentation)とは、特定の学習データが少ない際に、汎化性能（未知のデータに対する性能）を高めることにより、データの正確性を確保するためにとられる手段の1つである。学習に用いるデータを拡張し（例えば画像データであれば、反転、拡大、縮小を適用し）それぞれを別のソースに基づくデータとして用いることにより、汎化性能が改善されることがある。</p> <p>また、AIよりなされる判断は、事後的に精度が損なわれたり、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、権利侵害の生じる頻度、技術水準、精度を維持するためのコスト等¹⁾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。</p> <p>なお、消費者的利用者から提供されるデータを用いることが予定されている場合には、AIの特性及び用途を踏まえ、データ提供の手段、形式等について、あらかじめ消費者的利用者に情報を提供することが期待される。</p> <p>1) 例えば、機械学習を中心としたAIは帰納的な処理を行うため、当該AI単体では、判断結果につき原理的に100%の精度を担保できないことが挙げられる。</p>
--

<参考>

消費者的利用者は、自らデータを収集し、利用するAIの学習等を行うことが予定されている場合には、データの形式について、開発者、AIサービスプロバイダ等から提供された情報を踏まえた上でデータの収集、保存を行うことが望ましい。

図 7： 詳説の例

(AI サービスプロバイダ、ビジネス利用者及びデータ提供者に対する事項を中心に、参考として消費者的利用者に対する事項を記載)

また、今回解説を行うにあたっては、定期的に見直すことを前提として、近年 AI 技術として利活用が進められている「機械学習」の例を中心に図 8 のような形で図解を行うこととした。

(機械学習を中心とした) 学習と利用の流れ

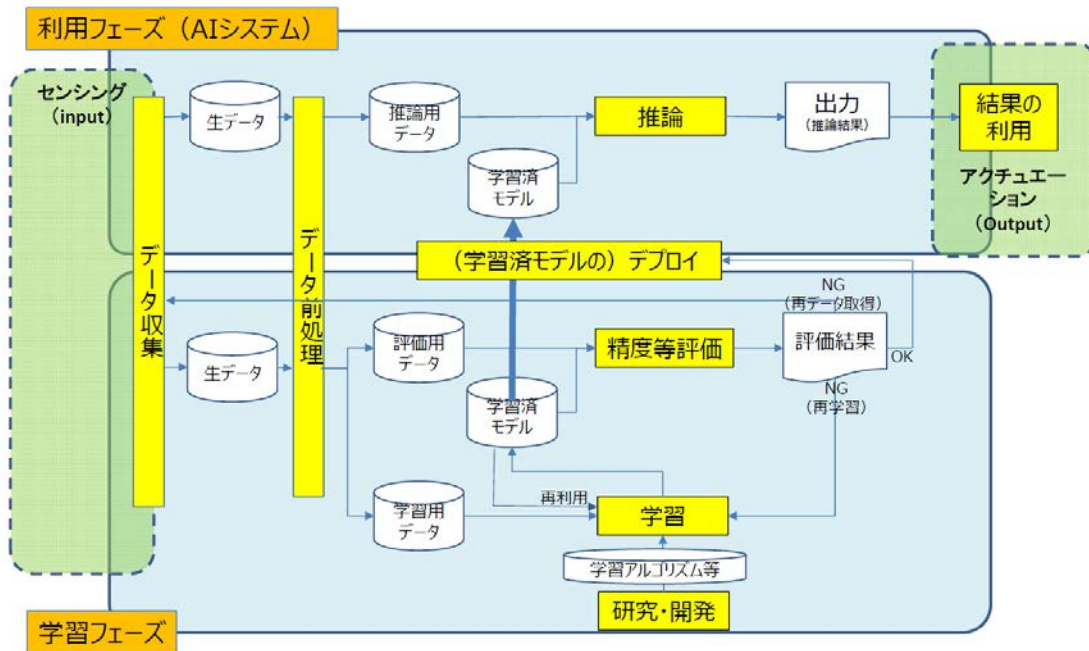


図 8 : 流れ (フロー) に関する図解例

(7) AI 利活用原則を考慮すべきタイミング

AI 利活用ガイドラインを参照する AI サービスプロバイダ、ビジネス利用者、消費者的利用者等の理解を容易にする観点から、(4) で示した各原則の各論点が (5) で示したどのフェーズで問題となるかを明示することとし、表 2 及び表 3 のとおり両者の対応関係の整理を行った。

ここでは、AI サービスプロバイダ、ビジネス利用者については、自ら AI の運用等を行う場合を想定し、また、消費者的利用者については、自ら運用等を行わない場合を想定した上で、利活用のフェーズとの関係を整理している (なお、データ提供者については、自ら運用等を行う AI サービスプロバイダ及びビジネス利用者のフェーズをもとに整理を行っている。)。このため、自ら運用等を行わないビジネス利用者や、自ら運用等を行う消費者的利用者については、読み替えを行う必要がある旨を注記している。

表 2：AI サービスプロバイダ、ビジネス利用者等の利活用の流れと各原則・各論点との関係

	AI 構築	システム実装	デプロイ	利用・運用
① 適正利用の原則				
ア 適正な範囲・方法での利用	○	○	○	○
イ 人間の判断の介入	○	○	○	○
ウ 関係者間の協力			○	○
② 適正学習の原則				
ア AIの学習等に用いるデータの質への留意	○			
イ 不正確又は不適切なデータの学習等によるAIのセキュリティ脆弱性への留意	○		○	
③ 連携の原則				
ア 相互接続性と相互運用性への留意		○	○	○
イ データ形式やプロトコル等の標準化への対応	○	○	○	○
ウ AIネットワーク化により惹起・増幅される課題への留意		○	○	○
④ 安全の原則				
ア 人の生命・身体・財産への配慮		○	○	○
⑤ セキュリティの原則				
ア セキュリティ対策の実施		○	○	○
イ セキュリティ対策のためのサービス提供等			○	○
ウ AIの学習モデルに対するセキュリティ脆弱性への留意	○		○	
⑥ プライバシーの原則				
ア 他者のプライバシーの尊重		○	○	○
イ パーソナルデータの収集・前処理・提供等におけるプライバシーの尊重	○		○	
ウ 自己のプライバシー侵害への留意及びパーソナルデータの流出の防止		○		
⑦ 尊厳・自律の原則				
ア 他者の尊厳と自律の尊重			○	○
イ AIによる意思決定・感情の操作等への留意			○	○
ウ AIと人間の脳・身体を連携する際の生命倫理等の議論の参照		○	○	○
エ AIを利用したプロファイリングを行う場合における不利益への配慮	○	○	○	○
⑧ 公平性の原則				
ア AIの学習等に用いられるデータの代表性への留意	○			
イ アルゴリズムによるバイアスへの留意	○			
ウ 人間の判断の介入(公平性の確保)	○			
⑨ 透明性の原則				
ア AIの入出力等のログの記録・保存		○	○	○
イ 説明可能性の確保	○			
ウ 行政機関が利用する際の透明性の確保	○	○	○	○
⑩ アカウンタビリティの原則				
ア アカウンタビリティを果たす努力			○	○
イ AIに関する利用方針の通知・公表			○	○

表 3：消費者的利用者の利活用の流れと各原則・各論点の関係

	利用前	利用	データ 収集
① 適正利用の原則			
ア 適正な範囲・方法での利用	○	○	
イ 人間の判断の介在	○	○	
ウ 関係者間の協力	○	○	
② 適正学習の原則			
ア AIの学習等に用いるデータの質への留意			○
イ 不正確又は不適切なデータの学習等によるAIのセキュリティ脆弱性への留意		○	○
③ 連携の原則			
ア 相互接続性と相互運用性への留意	○	○	
イ データ形式やプロトコル等の標準化への対応	○	○	○
ウ AIネットワーク化により惹起・増幅される課題への留意	○	○	
④ 安全の原則			
ア 人の生命・身体・財産への配慮	○	○	
⑤ セキュリティの原則			
ア セキュリティ対策の実施	○	○	
イ セキュリティ対策のためのサービス提供等	○	○	
ウ AIの学習モデルに対するセキュリティ脆弱性への留意		○	○
⑥ プライバシーの原則			
ア 他者のプライバシーの尊重	○	○	
イ パーソナルデータの収集・前処理・提供等におけるプライバシーの尊重	○		○
ウ 自己のプライバシー侵害への留意及びパーソナルデータの流出の防止		○	
⑦ 尊厳・自律の原則			
ア 他者の尊厳と自律の尊重	○	○	
イ AIによる意思決定・感情の操作等への留意	○	○	
ウ AIと人間の脳・身体を連携する際の生命倫理等の議論の参照	○	○	
エ AIを利用したプロファイリングを行う場合における不利益への配慮	○	○	
⑧ 公平性の原則			
ア AIの学習等に用いられるデータの代表性への留意	○	○	
イ アルゴリズムによるバイアスへの留意	○	○	
ウ 人間の判断の介在(公平性の確保)			
⑨ 透明性の原則			
ア AIの入出力等のログの記録・保存			
イ 説明可能性の確保			
ウ 行政機関が利用する際の透明性の確保			
⑩ アカウンタビリティの原則			
ア アカウンタビリティを果たす努力	○	○	
イ AIに関する利用方針の通知・公表	○	○	

(8) AI の定義

① 定義

本報告書及び AI 利活用ガイドラインにおける「AI」の概念を AI 開発ガイドライン案同様、「AI ソフト及び AI システムを総称する概念」と位置付けた上で、「AI ソフト」を学習等により自らの出力やプログラムを変化させるソフトウェアの機能に着目して定義し、その例として、機械学習ソフトウェアを挙げる一方で、「AI システム」を「AI ソフトを構成要素として含むシステム」と定義し、その例として、AI ソフトを実装したロボットやクラウドシステムを挙げている。学習等により自らの出力やプログラムを変化させる機能に着目して AI ソフトを定義する理由は、このように定義された AI ソフト及びそれを構成要素として含む AI システムについては、学習等を通じた変化によって開発者が予見し抑制することが困難なリスクが生ずるおそれがあることなどから、従来の ICT とは異なる特別の留意が求められることによる。

② 対象範囲

AI 利活用ガイドラインの対象とする AI システムの範囲は、AI システムがネットワークを通じて国境を越えて利用され、広く人間及び社会に便益やリスクをもたらす可能性があることから、AI 開発ガイドライン案同様、ネットワーク化され得る AI システム（ネットワークに接続可能な AI システム）としている。ネットワーク化され得る AI システムを AI 利活用ガイドラインの対象とする理由は、ネットワーク化され得る AI システムについては、ネットワークを通じて国境を越えて即時に便益及びリスクが波及する蓋然性が高い一方で、ネットワーク化され得ない AI システムについては、国境を越えて即時に便益及びリスクが波及する蓋然性は低いと考えられることから、前者については、後者と異なり、留意することが期待される事項を国際的に共有する必要性が高いものと考えられることによる。

4. 今後の展開

第 1 章の海外の動向、国際的な議論の動向に記載したとおり、IEEE、欧州委員会、OECD 等では、AI の開発や利活用にあたり留意すべき事項を原則として掲げるだけでなく、掲げた原則をどのように実現するか、原則に則っているかをどのように評価するかという点に議論の焦点が移行しつつある。例えば、2019 年（令和元年）7 月及び同年 11 月に開催される OECD デジタル経済政策委員会（CDEP）においては、同年 5 月に取りまとめられた理事会勧告で掲げられた原則を実現するための具体策についての議論が行われることが予定されている。引き続き、国内のみならず国際的な会議の場において、AI 利活用ガイドラインの周知を図りつつ、必要に応じフィードバックを得ながら議論を推進することにより、原則を実現するための具体策について国際的なコンセンサスが得られるよう一層の取組みを行っていくことが重要である。

また、国際的な動向を踏まえつつ、こうした原則を実現するための具体的な措置について、開発者、サービスプロバイダやビジネス利用者が自己点検、自己評価を容易にする仕組みを整備することに加え、第三者による審査・評価を受けうる仕組みについても、今後検討を進めて行くことが望ましい。

さらに、AIシステムの利活用は多様な分野に及び、分野ごとにAIシステムのもたらす便益やリスクはその軽重を含め異なる可能性があることに鑑みると、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、AI利活用ガイドラインを参考としつつ、各分野のステークホルダによる議論が進められることが期待される。

なお、分野ごとの具体的事例への適用を円滑に進めていくためには、これまで本推進会議で行ってきたシナリオを通じた分野横断的な分析や、分野別の評価を継続して実施するとともに、AIサービスプロバイダやビジネス利用者の利益にも配慮しつつ、ベストプラクティス情報の共有を図ることにより、AI利活用ガイドラインについて、具体的事例を意識した形での体系化を図っていくことが期待される。今後、AI利活用原則として掲げた10の原則について、具体的事例を念頭に、どの原則をどの程度の措置を講じることが期待されるかといった具体的な議論を積み重ねていくことが望ましい。

第3章 今後の課題

報告書 2018 に掲げられている「今後の課題」並びに本推進会議における AI 開発ガイドライン案、AI 利活用ガイドラインに関する検討及び政府 AI 戦略等国内での議論の状況等を踏まえ、今後の課題を次のとおり整理する。

1. AI の開発及び利活用の促進並びに AI ネットワーク化の健全な進展に関する事項

(1) AI 開発ガイドライン案及び AI 利活用ガイドラインの周知・展開

本推進会議においては、AI の便益を増進しリスクを抑制する観点から、国際的な議論の動向を踏まえつつ、開発者及び利用者が留意すべき事項について検討を行い、AI 開発ガイドライン案及び AI 利活用ガイドラインのとりまとめを行った。今後は、国内、国際の両面において、これらの周知・展開を進めていくことが重要である。

2017 年（平成 29 年）3 月には、総務省の主催により、東京大学において、日米欧等の産学民官のステークホルダが参加した国際シンポジウム「AI ネットワーク社会推進フォーラム」が開催された。それから 2 年が経過し、AI 技術の急激な進展に加え、各国、国際機関、企業・団体等でも AI の原則・ガイドライン策定の動きが広がっていることから、改めて AI ネットワーク化に関する社会的・経済的・倫理的・法的課題を議論するためのシンポジウムを開催し、AI 開発ガイドライン案及び AI 利活用ガイドラインのとりまとめにあたり本推進会議で行われた議論を紹介しつつ、これらの課題とその対応策等についての認識の共有を図っていくことが期待される。

国際的な枠組みにおいても、例えば、OECD では、2019 年（令和元年）夏以降、同年 5 月に取りまとめた理事会勧告の内容を実現するための実践的なガイダンス策定に向けた検討に着手する旨を表明しており、また、G7 においても、カナダ、フランスのイニシアティブにより、各国の有識者が AI の様々な課題について検討を行う International Panel on AI の設置が議論されている。こうした国際的な議論の場において、特に AI 利活用ガイドラインで整理した各原則を実現するための「詳説」について紹介し、AI の原則をどのように実現していくべきか、その具体策について、引き続き我が国が主導して国際的なコンセンサスを形成していくことが重要である。

また、AI 開発ガイドライン案及び AI 利活用ガイドラインについて、国内における周知を進め、普及を図ることも重要である。これらのガイドラインにおいてとりまとめられた具体的な内容について、開発者、AI サービスプロバイダ及びビジネス利用者による実施状況の自己点検や自己評価を促す仕組みについて検討を進めるとともに、第三者による審査・評価を受けうる仕組みについても検討を進めて行くことが期待される。

さらに、消費者的利用者向けにも分かりやすいメッセージを発信することが重要であり、AI 利活用ガイドラインにおける記載等をもとに、「ハンドブック」や「マニュアル

ル」などリテラシー教材（利用者の手引き）を作成し、それらに基づいてワークショップを実施すること等についても検討することが望ましい。

（２） AI の開発及び利活用に関する原則・ガイドラインについての議論のフォローアップ

AI の技術の進歩は著しく、また利活用の裾野も急速に広まっている。こうした状況を適時に把握しつつ、AI に関する原則・ガイドライン等の国際的な議論のフォローアップを行い、その動向を踏まえ、継続的に AI 開発ガイドライン案、AI 利活用ガイドラインの見直しを行っていくことが必要である。その際、引き続きマルチステークホルダ会合の枠組みにて対話・情報共有を行うことが重要である。

また、民間企業が自らの AI 原則を策定する際の参照として AI 利活用ガイドラインを活用していく上では、企業、団体等における具体的なベストプラクティス事例に基づき解説を行うことが効果的であり、マルチステークホルダによる議論を継続した上で、具体的な事例を意識した形で AI 利活用ガイドラインの体系化を図っていくことが望ましい。

（３） 関係するステークホルダが取り組む環境整備に関する課題

AI の便益を人々や社会が享受するためには、AI の開発をさらに推し進めるとともに、利活用の裾野をより一層広げていくことが必要である。そのためには、人々の AI に対する不安を取り除き、信頼を醸成する取組を進めることが必要である。AI の開発者や利用者が、それぞれ AI の開発や利活用において、AI の便益を増進させリスクを抑制する観点から期待される取組を行うとともに、開発者や利用者に加えて、政府や業界団体等を含めた関係するステークホルダが AI の開発及び利活用の促進や AI ネットワーク化の健全な進展に向けた環境整備のための取組を行うことにより、AI に対する社会的な受容性の向上がもたらされる。

これまで述べたように、AI に対し開発者、利用者等が留意することが期待される事項については、それぞれ AI 開発ガイドライン案、AI 利活用ガイドラインを取りまとめたところである。

これらの取組に加えて、開発者、利用者等を含む関係するステークホルダ（政府、業界団体等）が AI の開発及び利活用の促進や AI ネットワーク化の健全な進展に向けた環境整備のために取り組むことが期待される事項を整理することが必要であり、例えば次に掲げる課題が挙げられる⁴⁶。

- すべてのステークホルダが取り組むことが期待される課題
 - AI の開発及び利活用の促進並びに AI ネットワーク化の健全な進展に関するス

⁴⁶ これらの課題について、他で掲げられる課題と重複するものがあり得る。

テークホルダ間の協力、ベストプラクティスの共有の在り方の検討

- AI の利活用におけるデータ（学習用データを含む。）の確保⁴⁷
- 主として業界団体等が中心となって取り組むことが期待される課題
 - AI の学習等に用いられるデータの形式や AI 間の自動交渉のプロトコル等の標準化の在り方の検討
 - AI の機能や性能などの品質保証及びそれらに関する情報提供の在り方の検討
 - AI の利活用、AI ネットワーク化に必要な知識・技能等に関する啓発・教育
 - AI の開発及び利活用並びに AI ネットワーク化を推進するための人材育成
 - AI の利活用、AI ネットワーク化に係る市場競争環境整備、利用者の選択肢の多様性の確保の在り方の検討
 - AI の事故等に関する被害者の救済（保険等）及び被害発生防止の在り方の検討
- 主として政府が中心となって取り組むことが期待される課題
 - 開発者、利用者及びデータ提供者を含む関係するステークホルダ間の対話や情報共有が促進されるための方策（対話の場の設置等）の在り方の検討
 - AI の事故等が発生した場合の原因究明や再発防止等に関して関係者の協力が得られる仕組み（インセンティブの付与など）の検討
 - AI の便益及びリスクを検証するための実験環境（特区、仮想実験空間等）の整備の在り方の検討
 - AI の開発及び利活用並びに AI ネットワーク化を推進するための経済政策や産業政策、国内産業の活性化のための対外政策の検討
 - AI の開発及び利活用並びに AI ネットワーク化を推進するための法制度等の在り方の検討
 - ア AI の事故等に関する法的責任の所在の在り方
 - イ 各分野における AI の利活用と業法等との関係の整理
 - ウ AI の学習等に用いられるデータ、AI により作成されたモデル、AI 創作物等に関する知的財産制度上の保護と利活用の在り方
 - エ AI の利活用に関する個人情報の保護と利活用の在り方
 - オ AI ネットワークに関する通信の秘密の保護（AI 間の通信の秘密の保護の範囲及び程度など）の在り方

⁴⁷ 収集するデータが、AI を活用する目的に適っていることが極めて重要であり、その上でデータが不足している場合に、データの確保以外にも様々な技術（Generative Adversarial Network、Semi-Supervised Learning、Transfer Learning（転移学習）等）の活用が期待されるという意見がある。

(4) AI システム又は AI サービス相互間の円滑な連携の確保

AI 開発ガイドライン案・AI 利活用ガイドラインでは、下記の通り、AI 相互間の連携、AI の相互接続性と相互運用性に留意する事が挙げられている。AI 相互間の円滑な連携の確保に関し、AI 開発ガイドライン案及び AI 利活用ガイドラインにおける関係する記載を踏まえると、例えば次に掲げる課題が挙げられる。

- 開発者や AI サービスプロバイダが他の開発者や AI サービスプロバイダ等と共有することが期待される関連情報の範囲、共有の方法等の検討
- インターフェース、プロトコル、データ形式等の標準化又はオープン化の在り方の検討⁴⁸
- 標準必須特許など相互接続性と相互運用性の確保に資する知的財産権のライセンス契約及びその条件の在り方等に関する検討
- 連携及びこれに伴う責任の分配に関する当事者間の契約の在り方の検討
- 当事者間の紛争解決の在り方の検討

なお、これらの課題に取り組むに当たっては、ネットワーク上における様々な AI システムの混在(性質や性能が異なる多様な AI システムがネットワーク上に混在したり、新旧の AI システムがネットワーク上に混在したりすることや目的が異なったり競合したりする AI システムがネットワーク上に混在して相互間の交渉や調整が必要となり得ること⁴⁹等)の可能性に留意することが必要である。

[AI 開発ガイドライン案及び AI 利活用ガイドラインにおける関係する記載]

- AI 開発原則⁵⁰「連携の原則」：開発者は、AI システムの相互接続性と相互運用性⁵¹に留意すること。
- AI 利活用原則⁵²「連携の原則」：AI サービスプロバイダ、ビジネス利用者及びデータ提供者は、AI システム又は AI サービス相互間の連携に留意すること。

⁴⁸ データ形式やプロトコル等の標準化への対応については、国際的議論の動向を踏まえて検討することが重要である。他方、必ずしも標準化が必要となるものばかりではないことに留意する必要がある。

⁴⁹ AI システム相互間の連携・協調をもたらす便益を指摘するとともに、目的が競合し又は対立する AI システム相互間の交渉・協調の在り方を今後の課題として指摘するものとして、長尾顧問(京都大学名誉教授)「AI ネットワーク化の本質と将来」(AI ネットワーク社会推進フォーラム(特別講演)配付資料)2-3頁(平成29年)参照。

⁵⁰ AI 開発原則については、報告書2017「別紙1」4. 参照。以下本章における AI 開発原則に関する記述において同じ。

⁵¹ ここで相互接続性と相互運用性としては、自らの開発する AI システムが情報通信ネットワークと接続され、他の AI システム等と適切に協調して運用することが可能であることを一体的に捉えて念頭に置いている。

⁵² AI 利活用原則案については、前章3. 参照。以下本章における AI 利活用原則に関する記述において同じ。

(5) 競争的なエコシステムの確保

AI ネットワーク化を進展し、AI 利活用ガイドラインの基本理念に掲げた社会を実現するためには、

- AI システム相互間又は AI システムと他の種類のシステムとの間のネットワークの形成
- AI システムや AI サービスの提供
- ネットワーク化された AI により創造・流通・連結されるデータ・情報・知識の利活用

のそれぞれのエコシステムにおいて公正な競争が確保されることが必要である。

これらのエコシステムは、今後の AI や情報通信ネットワークに関する技術の高度化、市場における AI システムや AI サービスに対するニーズの変化、AI システムの利活用の進展等に応じて変遷していくものであり、予断を抱くべきものではないことから、競争的なエコシステムの確保に関する規制の創設の検討は将来に向けて謙抑的であるべきである。その一方で、競争的なエコシステムの確保に関し必要に応じて時宜に応じた適切な対応ができるよう備えておく見地からは、報告書 2016 において指摘されているように⁵³、関連する市場の動向を継続的に注視することが必要であるものと考えられる。

(6) 利用者の利益の保護

AI ネットワーク化が進展して、様々な物、サービスや社会のシステムがネットワーク化された AI の利用可能性を前提として提供されるようになった社会における人間の「包摂」(inclusiveness) の見地に加えて、ネットワークに接続された AI に関するいわゆるネットワーク効果の見地からは、あらゆる個人又は団体が、利用者として、関連する AI を適切な条件で⁵⁴安心・安全に利用することができるよう、その利益が保護されることが求められる。

このため、AI ネットワーク化のガバナンスにおいては、AI システムの開発のガバナンス及び利活用のガバナンスのいずれについても、その全体を通じて直接的又は間接的に利用者の利益の保護に資するものであることが期待される。AI 開発ガイドライン案においては、このような問題意識の下、その目的として利用者の利益の保護を掲げているとともに、同ガイドライン案に掲げる AI 開発原則の個々の項目は、いずれも利用

⁵³ 報告書 2016・50 頁参照。

⁵⁴ 適切な条件とは、AI を利用するために負担することを要する料金など金銭的なコストのほか、AI を利用するためにプロバイダ等に与えることを要する自らのパーソナルデータの範囲、そのプロバイダ等による当該パーソナルデータの取扱方法等を含む AI を利用するための条件が、AI の提供に要するコストや用途等に照らして総体として相応かつ非差別的なものであることをいう。これは公正な競争を通じて実現されることが基本である。

者の利益の保護に資するものとしている⁵⁵。

また、AI 利活用原則においても、個々の原則は、いずれも利用者の利益の保護に資するものとしている⁵⁶。

これらのことを踏まえると、特に利用者の利益の保護を直接的に志向する課題としては、例えば次に掲げる課題が挙げられる。

- 開発者や AI サービスプロバイダ等から利用者に対する
 - AI システム又は AI サービスの技術的特性に関する情報⁵⁷
 - 継続的なアップデートへの対応の必要性に関する情報⁵⁸
 - その他リスクの顕在化又は波及の抑止に資する情報の自発的な提供⁵⁹の在り方の検討
- AI サービスプロバイダ等がその提供の業務を通じて知り得た利用者又はその利活用に関するデータ等の取扱いの在り方の検討
- AI システムのリスクにより利用者等に被害が及ばないようにする方策の検討、リスクが顕在化した場合（事故の発生時等）における責任の分配や利用者等を保護する仕組み（保険等）等の在り方の検討

2. AI ネットワーク化が社会・経済にもたらす影響に関する事項

(1) AI ネットワーク化が社会・経済にもたらす影響に関するシナリオ分析

報告書 2018 において、AI ネットワーク化が社会・経済にもたらす影響に関するシナリオ分析の継続的な実施及び国際的な共有が掲げられているところ、引き続き、シナリオ分析を行うとともに、その成果を国際的に共有することが有益である。また、エコシ

⁵⁵ AI 開発原則のうち「連携の原則」は、連携に係る AI システムと他の AI システム等との円滑な連携を通じて、連携に係る AI システムの便益を増進するとともにリスクを抑制することにより利用者の利益の保護に資するものである。「透明性の原則」、「制御可能性の原則」、「安全の原則」、「セキュリティの原則」、「プライバシーの原則」及び「倫理の原則」は、それぞれ関連するリスクを抑制することにより利用者の利益の保護に資するものである。「利用者支援の原則」及び「アカウンタビリティの原則」は、それぞれ利用者の受容性を向上させることにより、利用者の利益の保護に資するものである。

⁵⁶ AI 利活用原則のうち「適正利用の原則」、「適正学習の原則」及び「連携の原則」は、それぞれ AI の便益を増進するとともにリスクを抑制することにより利用者の利益の保護に資するものである。「安全の原則」、「セキュリティの原則」及び「プライバシーの原則」は、それぞれ関連するリスクを抑制することにより利用者の利益の保護に資するものである。「尊厳・自律の原則」、「公平性の原則」、「透明性の原則」及び「アカウンタビリティの原則」は、それぞれ利用者の受容性を向上させるとともにリスクを抑制することにより、利用者の利益の保護に資するものである。

⁵⁷ AI 開発原則に掲げる事項に関する情報を含む。

⁵⁸ 報告書 2016・54 頁参照。

⁵⁹ これらの情報提供のうち開発者によるものについては、AI 開発原則のうち「アカウンタビリティの原則」に関連するものとして、また、利用者によるものについては、AI 利活用原則案のうち「アカウンタビリティの原則」に関連するものとして検討することが期待されよう。

システムの展望について、これまでの検討を踏まえ、AI ネットワーク上を流通するデータやAI ネットワークを通じて提供されるサービス等に関するエコシステムを展望することが期待される。これらのエコシステムの展望に当たっては、AI の利活用の在り方は単一ではなく、複数のものが想定され得ることを念頭に置いて検討を行うことが望ましい。

- AI ネットワーク化が社会・経済にもたらす影響に関するシナリオ分析の継続的な実施及び国際的な共有
- AI ネットワーク上を流通するデータに関するエコシステムやAI ネットワークを通じて提供されるサービスに関するエコシステム等についての継続的な検討及び国際的な共有

(2) AI ネットワーク化の進展に伴う影響の評価指標及び豊かさや幸せに関する評価指標の設定

報告書 2018 においては、AI ネットワーク化に関する指標について、特に次に掲げる2つの指標に関する検討を行った⁶⁰。

- AI ネットワーク化の前提となる環境整備に関するもの：AI ネットワーク化事前指標
 - AI ネットワーク化の初期段階に関するもの：AI ネットワーク化初期指標
- 今後も、引き続き、国際比較にも用いることができるような定量的に測定し、又は客観的に把握することができる指標の設定に向けた検討を行うことが期待される。

(3) AI の利活用に関する社会的受容性の醸成

AI 開発ガイドライン案及びAI 利活用ガイドラインでは、下記の通り、開発者、AI サービスプロバイダ及びビジネス利用者はアカウントビリティを果たすよう努めること挙げられている。AI の開発及び利活用の促進並びにAI ネットワーク化の健全な進展に向けて、AI 開発ガイドライン案及びAI 利活用ガイドラインの関係する記載を踏まえ、社会的な受容性を注視するとともに、受容性を高める方策の在り方等について検討することが必要である。

- 社会のAI に関する受容度の継続的注視、受容性を高める方策の在り方の検討
- AI の研究開発や利活用への投資に対する経営者の理解の促進

[AI 開発ガイドライン案及びAI 利活用ガイドラインにおける関係する記載]

- AI 開発原則「アカウントビリティの原則」：開発者は、利用者を含むステークホルダに対しアカウントビリティを果たすよう努めること。

⁶⁰ 報告書 2018 第2章5. 参照。

- AI 利活用原則「アカウントビリティの原則」:AI サービスプロバイダ及びビジネス利用者は、消費者的利用者及び間接利用者を含むステークホルダに対しアカウントビリティを果たすよう努めること。

3. AI ネットワーク化が進展する社会における人間をめぐる課題に関する事項

(1) 人間と AI との関係の在り方に関する検討

AI 開発ガイドライン案及び AI 利活用ガイドラインでは、下記の通り、人間の尊厳と個人の自律を尊重すること等が挙げられている。人間と AI との関係の在り方に関し、AI 開発ガイドライン案及び AI 利活用ガイドラインの関係する記載を踏まえ、人間と AI との役割分担、AI システムに委ねる判断の範囲の妥当性、AI ネットワーク化の進展による人間の能力の拡張の在り方等について検討することが必要である。

- 専門職（医師、弁護士、会計士等）と AI システムとの役割分担の在り方や重要な判断（診断、法令の解釈・適用、採用、人事評価、融資の審査等）に当たり AI システムの機能に委ねてもよい事項の範囲に関する検討
- AI・ロボットによる人間の知的・身体的能力の拡張（エンハンスメント）に関する倫理的問題の検討
- AI システムや AI サービスに対する過度な依存を回避する方策の検討
- 「AI 依存」など人間の心理や子供の発育への影響等に関する継続的注視

[AI 開発ガイドライン案及び AI 利活用ガイドラインにおける関係する記載]

- AI 開発原則「倫理の原則」：開発者は、AI システムの開発において人間の尊厳と個人の自律を尊重すること。
- AI 利活用原則「適正利用の原則」：利用者は、人間と AI との適切な役割分担のもと、利用者間における適切な役割分担を踏まえ、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こと
- AI 利活用原則「尊厳・自律の原則」：利用者は、AI システム又は AI サービスの利活用において、人間の尊厳と個人の自律を尊重すること
- AI 利活用原則「公平性の原則」：AI サービスプロバイダ、ビジネス利用者及びデータ提供者は、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が不当に差別されないよう配慮すること。

(2) ステークホルダ間関係の在り方に関する検討

報告書 2017 におけるシナリオ分析に基づく留意事項として、「事故の発生等 AI シス

テムのリスクが顕在化した場合の責任の分配の在り方」が挙げられている⁶¹。

今後、AIネットワーク化の進展を見据え、AI利活用ガイドラインの関係する記載を踏まえ、上記（１）人間とAIとの関係の在り方に関する検討とともに、AIのリスクの抑制等を図ることにより利用者及び第三者の利益を保護するために、ステークホルダの役割分担、権利義務の帰属の在り方、事故発生時等の責任の分配の在り方について検討することが必要である。

- AIのリスクが顕在化しないようにするためのステークホルダ間の役割分担の在り方の検討
- AIを利活用した取引における権利義務の帰属の在り方の検討
- AIのリスクが顕在化した場合のステークホルダ間の責任の分配の在り方の検討⁶²
- 特に最終利用者の利益を保護するためのアカウントビリティの在り方の検討
- AIの利活用をめぐる紛争の解決の在り方の検討

[AI開発ガイドライン案及びAI利活用ガイドラインにおける関係する記載]

- AI利活用原則「適正利用の原則」：利用者は、人間とAIとの適切な役割分担のもと、利用者間における適切な役割分担を踏まえ、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こと。
- AI利活用原則「連携の原則」：利用者は、関係者と協力して事故の未然防止及び事故の発生時等の対応に取り組むよう努めること。
- AI利活用原則「アカウントビリティの原則」：AIサービスプロバイダ及びビジネス利用者は、消費者的利用者及び間接利用者を含むステークホルダに対しアカウントビリティを果たすよう努めること。

（３）セーフティネットの整備

報告書2018において、大規模な失業が発生しないような方策の在り方等に関する課題が掲げられている⁶³。また、分野別評価において、AIネットワーク化の健全な進展に向けた今後の検討に当たって留意すべき点として、プロファイリング結果により、就職や転職、結婚などにおいて、不当に不利な立場に陥ることのないようにセーフティネットの検討が必要である旨が挙げられている⁶⁴。

今後、労働市場の動向を注視するとともに、大規模な失業が発生しないような方策の在り方、AIによるプロファイリングの結果により、不当に不利な立場に陥らないよう

⁶¹ 報告書2017第3章5.ウ参照。

⁶² AIシステムのリスクが顕在化した場合の責任の分配の在り方の検討の一環として、AIの利活用等に伴う事故等による損害の賠償等に充てるための資財が乏しい個人、中小・零細事業者等によるAIの開発及び利活用を支援するためのファンドの設立や保険の在り方等についても併せて検討することが必要である。

⁶³ 報告書2017第4章(17)参照。

⁶⁴ 報告書2018第2章1.(2)③参照。

な方策の在り方等について検討すること、セーフティネットを整備すること等が必要である。

- 労働市場の動向の継続的注視、AI ネットワーク化の進展に伴う所得の再分配等格差防止の在り方の検討
- AI によるプロファイリングの結果により、不当に不利な立場に陥らないような方策の在り方の検討

4. その他の課題

報告書 2018 において、「その他の課題」として掲げられている事項⁶⁵についても、引き続き、注視し、又は検討することが必要である。

- 情報通信インフラの高度化の加速
- 経済発展・イノベーションの促進に向けた課題
- 地球規模課題の解決を通じた人類の幸福への貢献
- AI ネットワーク化のガバナンスの在り方

⁶⁵ これらの課題は、報告書 2016 において、「今後の課題」として掲げられたもののうち、報告書 2017、2018 においては、「今後の課題」として具体的に取り上げられなかった課題である。

(参考) 報告書 2018 に掲げられている「今後の課題」

- **AI の開発及び利活用並びに AI ネットワーク化の健全な進展に関する事項**
 - AI 開発ガイドライン（仮称）の策定：
AI 開発ガイドライン（仮称）の策定に向けた国際的な議論のフォローアップ
 - AI の利活用に関する指針の策定：
AI 利活用原則案の各原則の内容に関し整理された論点を踏まえた最終アウトプットの取りまとめに向けた検討
 - 関係するステークホルダが取り組む環境整備に関する課題：
ステークホルダ間の協力・ベストプラクティスの共有、法制度等の在り方の検討等
 - AI システム又は AI サービス相互間の円滑な連携の確保：
関係ステークホルダ間で共有することが期待される関連情報の範囲等の検討
 - 競争的なエコシステムの確保：
関連する市場の動向の継続的注視
 - 利用者の利益の保護：
利用者に対する開発者等からの自発的な情報提供の在り方の検討、利用者を保護する仕組み（保険等）の在り方の検討等
 - 技術開発に関する課題：
AI ネットワークのセキュリティを高めるための技術やセキュリティ攻撃の際の早期復旧のための技術等の研究開発の推進

- **AI ネットワーク上を流通する情報・データに関する事項**
 - セキュリティ対策：
セキュリティの実装の在り方の検討、AI の研究者とセキュリティの研究者との間の連携の促進に向けた検討等
 - プライバシー及びパーソナルデータの保護：
本人同意等の在り方、データ・情報の加工（匿名化、暗号化等）に関する検討等
 - コンテンツに関する制度的課題：
学習用データの作成の促進に関する環境整備、AI 生成物の知的財産制度上の在り方の検討等

- **AI ネットワーク化が社会・経済にもたらす影響の評価に関する事項**
 - AI ネットワーク化が社会・経済にもたらす影響に関するシナリオ分析：
シナリオ分析の継続的な実施・国際的な共有等
 - AI ネットワーク化の進展に伴う影響の評価指標及び豊かさや幸せに関する評価指標の設定：

指標の設定に向けた検討

- AIシステムの利活用に関する社会的受容性の醸成：
社会におけるAIの利活用に関する受容度の継続的注視等
- **AIネットワーク化が進展する社会における人間をめぐる課題に関する事項**
- 人間とAIとの関係の在り方に関する検討：
専門職（医師、弁護士、会計士等）とAIシステムとの役割分担の在り方等の検討
- ステークホルダ間の関係の在り方に関する検討：
AIのリスクが顕在化した場合の責任の分配の在り等の検討
- AIネットワーク化に対応した教育・人材育成及び就労環境の整備：
AIネットワーク化の進展に応じた教育の実施・人材の育成等
- AIに関するリテラシーの涵養及びAIネットワーク・ディバイドの形成の防止：
特に高齢者など情報弱者のリテラシーの向上を図るための方策の検討等
- セーフティネットの整備：
労働市場の動向の継続的注視、AIネットワーク化の進展に伴う所得の再分配等格差防止の在り方の検討等

結びに代えて

本報告書は、AI ネットワーク化をめぐる国内外の議論を概観した上で、AI の利活用において留意することが期待される事項を「AI 利活用原則」として掲げ、その解説を「AI 利活用ガイドライン」として取りまとめ、さらに今後の課題を整理した。

第1章で見たとおり、G7 香川・高松情報通信大臣会合（平成28年4月）を契機として、諸外国やG7、OECDにおける国際的な枠組みでの議論が活発に行われた結果、AIの原則の概念レベルでは、国際的に概ねコンセンサスが得られつつある。他方で、今後は、これらの原則をどのように実現し、社会に実装していくのかについての検討に焦点が移行していくものと見込まれる。「AI利活用ガイドライン」を社会に普及させることにより、AIに対する人々の信頼が醸成され、AIの利活用が進展することが期待されるとともに、これを積極的に海外に情報発信することで、国際的な認識の共有が図られることが期待される。

また、第3章で整理した今後の課題は、多岐にわたっており、継続的かつ多面的な検討が必要である。これらの課題の検討に当たっては、産学民官の幅広い分野から関係ステークホルダの参画を得て国内外において検討を進めていくことが必要である。今後、緊急性、重要性等を勘案してプライオリティをつけ、継続的に検討が進められていくことを期待したい。