

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

2022年4月27日

事 務 局

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

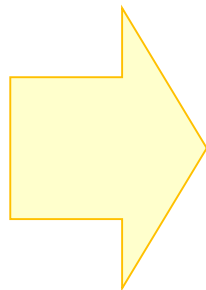
AIの倫理・ガバナンスについて、本推進会議において、AI開発ガイドライン※¹（2017年7月）やAI利活用ガイドライン（2019年8月）を策定し、G7やOECD等における国際的な議論に貢献してきた。

その後、AIの倫理・ガバナンスに関する国内外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、原則・指針の策定から社会実装の取組にシフトしつつあるものと考えられる。

本推進会議においても、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、その取組事例等を取りまとめた。



- 国内外において、多数の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注目すべきものがある。
- 欧州委員会が2021年4月に規制法案※²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求められている。



次の観点を考慮しつつ、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【前回の合同会合において(2/8)実施済】
- ② 国内における原則・指針・ガイドライン等との比較
- ③ ヒアリングにおいて示された取組事例との対応からの検討【前回の合同会合において(2/8)実施済】
- ④ AIの利活用、社会実装に関する事例との対応からの検討

このほか、EUの規制法案の動向等を踏まえて検討を行うことが重要である。

※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」のこと(以下同じ。)

国内における原則・指針・ガイドライン等との比較

○ 国内の事業者、団体において策定された原則・指針・ガイドライン等を踏まえて、現行のAI開発ガイドライン及びAI利活用ガイドラインで定められている原則等の射程、内容が妥当・適切であるか等について検討。

■ 調査対象ドキュメント

- ・ 国内の事業者、団体等が公表している22個のドキュメントを調査（14～15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
- ・ 1. ～16. までは、AIガイドライン比較表（「報告書2019」別紙2）における項目であり、17.～22.が新たに確認された項目

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	

【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。

国内における原則・指針・ガイドライン等との比較

- (株) ABEJA
 - (株) エヌ・ティ・ティ・データ
 - 沖電気工業 (株)
 - クウジツ (株)
 - KDDI (株)
 - コニカミノルタ (株)
 - (株) J.Score
 - (株) スタジアム
- ソニーグループ (株)
 - 日本電気 (株)
 - 日本ユニシス (株)
 - (株) 野村総合研究所
 - (株) 日立製作所
 - 富士通 (株)
 - 富士フイルム (株)
 - (株) 三菱総合研究所
- 三菱電機 (株)
 - (株) リクルート (リクルートワークス研究所)
 - (一社) 人工知能学会
 - (一社) 日本経済団体連合会
 - (一社) 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) 半数を超えるドキュメントに記載のある場合、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。

国内における原則・指針・ガイドライン等との比較

尊重すべき価値	ガイドラインにおいて関係があるのではないかと考えられる箇所（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、14～15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。

国内における原則・指針・ガイドライン等との比較

尊重すべき価値	ガイドラインにおいて関係があるのではないかと考えられる箇所（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】

国内における原則・指針・ガイドライン等との比較

尊重すべき価値	ガイドラインにおいて関係があるのではないかと考えられる箇所（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】

国内における原則・指針・ガイドライン等との比較

尊重すべき価値	ガイドラインにおいて関係があるのではないかと考えられる箇所（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」を求めらるれば、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】

国内における原則・指針・ガイドライン等との比較

尊重すべき価値	ガイドラインにおいて関係があるのではないかと考えられる箇所（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】

国内における原則・指針・ガイドライン等との比較

尊重すべき価値	ガイドラインにおいて関係があるのではないかと考えられる箇所（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 <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】

国内における原則・指針・ガイドライン等との比較

尊重すべき価値	ガイドラインにおいて関係があるのではないかと考えられる箇所（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】

国内における原則・指針・ガイドライン等との比較

尊重すべき価値	ガイドラインにおいて関係があるのではないかと考えられる箇所（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】

国内における原則・指針・ガイドライン等との比較

尊重すべき価値	ガイドラインにおいて関係があるのではないかと考えられる箇所（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】

国内における原則・指針・ガイドライン等との比較

尊重すべき価値	ガイドラインにおいて関係があるのではないかと考えられる箇所（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】

国内における原則・指針・ガイドライン等との比較

<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) ABEJA	AIポリシー URL < https://abejainc.com/ja/contact/ai_policy/ >
2	(株) 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)	OKIグループ AI原則 URL < https://www.oki.com/jp/press/2019/09/z19033.pdf >
4	クウジット (株)	クウジットAI倫理ガイドライン URL < https://www.koozyt.com/aiethicsguidelines >
5	KDDI (株)	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)	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) J.Score	AI活用ポリシー URL < https://www.jscore.co.jp/ai_policy/ >
8	(株) 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)	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)	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) ※	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 >

国内における原則・指針・ガイドライン等との比較

	事業者、団体	ガイドラインの名称
12	(株) 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) 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)	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)	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) 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)	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) 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	(一社) 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	(一社) 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	(一社) 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 >

AIの利活用、社会実装に関する事例との対応からの検討

○ 国内外におけるAIの利活用、社会実装に関する事例を踏まえて、現行のAI開発ガイドライン及びAI利活用ガイドラインで定められている原則等の射程、内容が妥当・適切であるか等について検討。

■ 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（19ページ参照）
 - 「生体認証」、「自動運転」、「画像識別」、「医療」等のAIの利活用や社会実装に際し、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が生じている。
 - 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の利活用や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の利活用、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

- ・ 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。【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。【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

- ・ 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の利活用、社会実装に関する事例から抽出された課題等（概要）」に記載されている番号は、19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。

AIの利活用、社会実装に関する事例との対応からの検討

AIの利活用、社会実装に関する事例から抽出された課題等（概要）

画像識別	<ul style="list-style-type: none"> 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	<ul style="list-style-type: none"> 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	<ul style="list-style-type: none"> 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	<ul style="list-style-type: none"> 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	<ul style="list-style-type: none"> 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	<ul style="list-style-type: none"> 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	<ul style="list-style-type: none"> 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	<ul style="list-style-type: none"> 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	<ul style="list-style-type: none"> 接触確認・追跡アプリ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・追跡の効果を享受できるようにする必要がある。【3番、14番】

AIの利活用、社会実装に関する事例との対応からの検討

	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	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	★	★	★			★			★			★	★	★									

AIの利活用、社会実装に関する事例との対応からの検討

	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	<ul style="list-style-type: none"> DX白書 2021 URL <https://www.ipa.go.jp/files/000093706.pdf>
2	日本	情報処理推進機構	<ul style="list-style-type: none"> 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	<ul style="list-style-type: none"> 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	<ul style="list-style-type: none"> 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	<ul style="list-style-type: none"> 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	<ul style="list-style-type: none"> 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	独国	連邦政府情報セキュリティ局	<ul style="list-style-type: none"> 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	仏国	情報処理と自由に関する 国家委員会	<ul style="list-style-type: none"> 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	<ul style="list-style-type: none"> 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	<ul style="list-style-type: none"> 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	<ul style="list-style-type: none"> 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)	<ul style="list-style-type: none"> 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)	<ul style="list-style-type: none"> Ethics and governance artificial intelligence for health URL <https://www.who.int/publications/i/item/9789240029200>
14	国際機関	世界経済フォーラム (WEF)	<ul style="list-style-type: none"> The AI Governance Journey: Development and Opportunities URL <https://www3.weforum.org/docs/WEF_The%20AI_Governance_Journey_Development_and_Opportunities_2021.pdf>

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

AIネットワーク社会推進会議・AIガバナンス検討会 合同会議（2/8）における主な御意見

前回の合同会合（2/8）において実施したレビュー①及び③の際の主な御意見（概要）は、次のとおり。

- 尊重すべき価値として新しく抽出された「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。
- 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。
- AI自体をガバナンスするという考え方があるが、他方で、AIをツールとして社会や企業をガバナンスするという考え方もある。この2つの考え方をもう少し精密に分けて考えることが今後必要になってくるのではないかと思う。
- 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。
- 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。
- 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のネットワーク化について、既にガイドラインに記載がある※が、）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。

※ AI開発ガイドライン

「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に留意する。

AI利活用ガイドライン

「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。

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

ガイドラインの見直しに関する論点（案）

- 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、これまで各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（原則・指針・ガイドライン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

【注】 見直しに当たっては、イノベーションを阻害しない、開発者等にとって過度に負担とならないものとするといった点については、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に関するレビュー

ガイドラインの見直しに関する論点（案）

- AIの開発のハードルが低くなってきていることやAIの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、このようの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の規制法案が公表さ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、求められる対応はあるか。

等

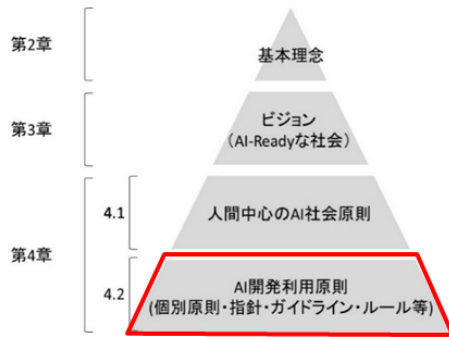
【注】 見直しに当たっては、イノベーションを阻害しない、開発者等にとって過度に負担とならないものとするといった点については、引き続き、留意することが重要である。

(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。