

# 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

2022年2月8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

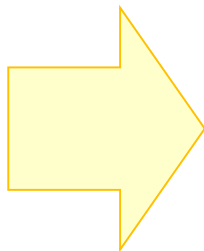
AIの倫理・ガバナンスについて、本推進会議において、AI開発ガイドライン※<sup>1</sup>（2017年7月）やAI利活用ガイドライン（2019年8月）を策定し、G7やOECD等における国際的な議論に貢献してきた。

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※<sup>2</sup>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 国内外における原則・指針・ガイドライン等との比較
- ② 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

このほか、EUの規制法案を踏まえて検討を行うことが重要である。

※<sup>1</sup> 「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。)

※<sup>2</sup> 「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。)

# 海外における原則・指針・ガイドライン等との比較

○ 海外で策定された原則・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。

## ■ 調査対象国、調査対象ドキュメント

- ・ 16か国<sup>※3</sup>（アメリカ、カナダ、イギリス、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（国連教育科学文化機関（UNESCO）、世界保健機関（WHO））
- ・ 各国政府機関、業界団体等、大手グローバル企業等が公表している67個のドキュメント<sup>※4</sup>を調査（16～23ページ参照）
  - 政府機関（40）、業界団体等（10）、グローバル企業（17）

※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研究者数のいずれかが10位以内の国を選定。

※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。

## ■ 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<sup>※5</sup>

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	

※5 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。17.～22.が新たに追加した項目。

【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。なお、現時点では、調査の途上であり、今後変更等があり得る。

# 海外における原則・指針・ガイドライン等との比較

	ドキュメント数	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	透明性・ 説明可能性	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
アメリカ	6	1	2			1	1		1	1		3	3	1	4	5	4	1	2	2	3		
カナダ	1							1								1							1
イギリス	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	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1) 政府機関、EU及び国際機関を対象として集計。

(注2) 各国等のドキュメント数に対し、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。

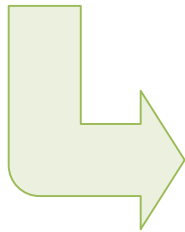
# 海外における原則・指針・ガイドライン等との比較

<地域ごとの特色>

	ドキュメント数	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	透明性・説明可能性	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) 政府機関及びEUを対象として集計。

(注2) 各地域のドキュメント数に対し、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。



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

# 海外における原則・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"> <li>○ 基本理念           <ul style="list-style-type: none"> <li>・ 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</li> </ul> </li> </ul> <p>【AI利活用ガイドライン】</p> <ul style="list-style-type: none"> <li>○ 基本理念           <ul style="list-style-type: none"> <li>・ 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</li> </ul> </li> </ul>	<ul style="list-style-type: none"> <li>・ 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】</li> <li>・ 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】</li> <li>・ 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】</li> </ul>
2.人間の尊厳	<p>【AI開発ガイドライン】</p> <ul style="list-style-type: none"> <li>○ 倫理の原則           <ul style="list-style-type: none"> <li>・ 人間の尊厳と個人の自律を尊重する</li> </ul> </li> </ul> <p>【AI利活用ガイドライン】</p> <ul style="list-style-type: none"> <li>○ 尊厳・自律の原則           <ul style="list-style-type: none"> <li>・ 人間の尊厳と個人の自律を尊重する</li> </ul> </li> </ul>	<ul style="list-style-type: none"> <li>・ 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】</li> <li>・ 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】</li> <li>・ 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】</li> </ul>

(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。

# 海外における原則・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"> <li>○ 基本理念           <ul style="list-style-type: none"> <li>・ 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</li> </ul> </li> <li>○ 利用者支援の原則           <ul style="list-style-type: none"> <li>➢ 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</li> </ul> </li> <li>○ 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"> <li>・ 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</li> <li>・ 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</li> </ul> </li> </ul> <p>【AI利活用ガイドライン】</p> <ul style="list-style-type: none"> <li>○ 基本理念           <ul style="list-style-type: none"> <li>・ 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</li> <li>・ 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</li> </ul> </li> </ul>	<ul style="list-style-type: none"> <li>・ 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】</li> <li>・ 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】</li> <li>・ 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】</li> </ul>
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"> <li>○ 基本理念           <ul style="list-style-type: none"> <li>・ 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</li> </ul> </li> </ul>	<ul style="list-style-type: none"> <li>・ 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】</li> <li>・ 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】</li> <li>・ 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】</li> </ul>

# 海外における原則・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"> <li>○ 基本理念           <ul style="list-style-type: none"> <li>・ 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</li> </ul> </li> <li>○ 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"> <li>・ 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</li> </ul> </li> </ul> <p>【AI利活用ガイドライン】</p> <ul style="list-style-type: none"> <li>○ 基本理念           <ul style="list-style-type: none"> <li>・ 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</li> </ul> </li> </ul>	<ul style="list-style-type: none"> <li>・ 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】</li> <li>・ 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】</li> <li>・ 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】</li> </ul>
6.適正な利用	<p>【AI開発ガイドライン】</p> <ul style="list-style-type: none"> <li>○ 制御可能性の原則           <ul style="list-style-type: none"> <li>➢ 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</li> </ul> </li> </ul> <p>【AI利活用ガイドライン】</p> <ul style="list-style-type: none"> <li>○ 適正利用の原則           <ul style="list-style-type: none"> <li>・ 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</li> </ul> </li> </ul>	<ul style="list-style-type: none"> <li>・ 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】</li> <li>・ 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、について自身自身でコントロールできるようにするべきである。【38番】</li> <li>・ 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】</li> <li>・ 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】</li> </ul>



# 海外における原則・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"> <li>○ 適正利用の原則(参考)           <ul style="list-style-type: none"> <li>➢ 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。</li> <li>➢ 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。</li> </ul> </li> </ul>	<ul style="list-style-type: none"> <li>・ 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】</li> <li>・ 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】</li> <li>・ 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】</li> </ul>
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"> <li>○ 制御可能性の原則           <ul style="list-style-type: none"> <li>・ AIシステムの制御可能性に留意する</li> <li>➢ 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</li> </ul> </li> </ul> <p>【AI利活用ガイドライン】</p> <ul style="list-style-type: none"> <li>○ 適正利用の原則           <ul style="list-style-type: none"> <li>➢ 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</li> </ul> </li> <li>○ 公平性の原則           <ul style="list-style-type: none"> <li>➢ 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</li> </ul> </li> </ul>	<ul style="list-style-type: none"> <li>・ 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】</li> <li>・ 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】</li> <li>・ 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】</li> </ul>

# 海外における原則・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"> <li>○ 倫理の原則           <ul style="list-style-type: none"> <li>➢ 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</li> </ul> </li> </ul> <p>【AI利活用ガイドライン】</p> <ul style="list-style-type: none"> <li>○ 適正学習の原則           <ul style="list-style-type: none"> <li>・ AIシステムの学習等に用いるデータの質に留意する</li> </ul> </li> </ul>	<ul style="list-style-type: none"> <li>・ 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】</li> <li>・ 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】</li> <li>・ 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】</li> </ul>
10.AI間の連携	<p>【AI開発ガイドライン】</p> <ul style="list-style-type: none"> <li>○ 連携の原則           <ul style="list-style-type: none"> <li>・ AIシステムの相互接続性と相互運用性に留意する</li> </ul> </li> </ul> <p>【AI利活用ガイドライン】</p> <ul style="list-style-type: none"> <li>○ 連携の原則           <ul style="list-style-type: none"> <li>・ AIシステム又はAIサービス相互間の連携に留意する</li> </ul> </li> </ul>	<ul style="list-style-type: none"> <li>・ 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】</li> </ul>
11.安全性	<p>【AI開発ガイドライン】</p> <ul style="list-style-type: none"> <li>○ 安全の原則           <ul style="list-style-type: none"> <li>・ 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。</li> </ul> </li> </ul> <p>【AI利活用ガイドライン】</p> <ul style="list-style-type: none"> <li>○ 安全の原則           <ul style="list-style-type: none"> <li>・ 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</li> </ul> </li> </ul>	<ul style="list-style-type: none"> <li>・ 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】</li> <li>・ 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】</li> <li>・ 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"> <li>(1)環境とユーザーに対する安全性</li> <li>(2)ハッキング等に対する信頼性と堅牢性</li> <li>(3)人間と機械の相互作用に関する情緒的安全性</li> </ol> </li> </ul>

# 海外における原則・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"> <li>○ セキュリティの原則           <ul style="list-style-type: none"> <li>・ AIシステムのセキュリティに留意する</li> </ul> </li> </ul> <p>【AI利活用ガイドライン】</p> <ul style="list-style-type: none"> <li>○ セキュリティの原則           <ul style="list-style-type: none"> <li>・ AIシステム又はAIサービスのセキュリティに留意する</li> </ul> </li> <li>○ 適正学習の原則           <ul style="list-style-type: none"> <li>➢ 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</li> </ul> </li> </ul>	<ul style="list-style-type: none"> <li>・ 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】</li> <li>・ 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】</li> <li>・ 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"> <li>(1)環境とユーザーに対する安全性</li> <li>(2)ハッキング等に対する信頼性と堅牢性</li> <li>(3)人間と機械の相互作用に関する情緒的安全性</li> </ol> </li> <li>・ 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】</li> </ul>
13.プライバシー	<p>【AI開発ガイドライン】</p> <ul style="list-style-type: none"> <li>○ プライバシーの原則           <ul style="list-style-type: none"> <li>・ 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</li> </ul> </li> <li>○ 利用者支援の原則           <ul style="list-style-type: none"> <li>・ 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</li> </ul> </li> </ul> <p>【AI利活用ガイドライン】</p> <ul style="list-style-type: none"> <li>○ プライバシーの原則           <ul style="list-style-type: none"> <li>・ 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</li> </ul> </li> </ul>	<ul style="list-style-type: none"> <li>・ 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ーから提供される場合に課題となる。【40番】</li> <li>・ 個人情報収集、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】</li> <li>・ 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】</li> <li>・ 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】</li> </ul>

# 海外における原則・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"> <li>○ 倫理の原則           <ul style="list-style-type: none"> <li>➢ 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</li> </ul> </li> </ul> <p>【AI利活用ガイドライン】</p> <ul style="list-style-type: none"> <li>○ 公平性の原則           <ul style="list-style-type: none"> <li>・ 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</li> </ul> </li> </ul>	<ul style="list-style-type: none"> <li>・ 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】</li> <li>・ 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】</li> <li>・ 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しなければならない。【53番】</li> <li>・ 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】</li> </ul>
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"> <li>○ 透明性の原則           <ul style="list-style-type: none"> <li>・ 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</li> </ul> </li> </ul> <p>【AI利活用ガイドライン】</p> <ul style="list-style-type: none"> <li>○ 透明性の原則           <ul style="list-style-type: none"> <li>・ 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</li> </ul> </li> <li>○ 適正利用の原則           <ul style="list-style-type: none"> <li>➢ 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</li> </ul> </li> </ul>	<ul style="list-style-type: none"> <li>・ 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】</li> <li>・ 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】</li> <li>・ 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】</li> <li>・ 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】</li> </ul>

# 海外における原則・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"> <li>○ アカウンタビリティの原則           <ul style="list-style-type: none"> <li>・ 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</li> </ul> </li> <li>○ 利用者支援の原則           <ul style="list-style-type: none"> <li>・ 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</li> </ul> </li> </ul> <p>【AI利活用ガイドライン】</p> <ul style="list-style-type: none"> <li>○ アカウンタビリティの原則           <ul style="list-style-type: none"> <li>・ 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</li> </ul> </li> <li>○ 適正利用の原則           <ul style="list-style-type: none"> <li>➢ 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</li> </ul> </li> </ul>	<ul style="list-style-type: none"> <li>・ 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】</li> <li>・ 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】</li> <li>・ 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】</li> </ul>
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"> <li>・ 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】</li> <li>・ 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】</li> <li>・ 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】</li> </ul>

# 海外における原則・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"> <li>○ アカウンタビリティの原則           <ul style="list-style-type: none"> <li>・ 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</li> </ul> </li> </ul> <p>【AI利活用ガイドライン】</p> <ul style="list-style-type: none"> <li>○ アカウンタビリティの原則           <ul style="list-style-type: none"> <li>・ 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</li> </ul> </li> </ul>	<ul style="list-style-type: none"> <li>・ 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】</li> <li>・ 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】</li> <li>・ AIシステムの設計、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】</li> <li>・ 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】</li> </ul>
19.追跡可能性	<p>【AI開発ガイドライン】</p> <ul style="list-style-type: none"> <li>○ 透明性の原則           <ul style="list-style-type: none"> <li>・ 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</li> </ul> </li> </ul> <p>【AI利活用ガイドライン】</p> <ul style="list-style-type: none"> <li>○ 透明性の原則           <ul style="list-style-type: none"> <li>・ 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</li> </ul> </li> </ul>	<ul style="list-style-type: none"> <li>・ 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】</li> <li>・ 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】</li> <li>・ 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】</li> </ul>

# 海外における原則・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
<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"> <li>○ 透明性の原則           <ul style="list-style-type: none"> <li>・ 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</li> </ul> </li> <li>○ 制御可能性の原則           <ul style="list-style-type: none"> <li>・ 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</li> </ul> </li> </ul> <p>【AI利活用ガイドライン】</p> <ul style="list-style-type: none"> <li>○ 透明性の原則           <ul style="list-style-type: none"> <li>・ 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</li> </ul> </li> </ul>	<ul style="list-style-type: none"> <li>・ 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】</li> <li>・ 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】</li> <li>・ 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】</li> </ul>
<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"> <li>・ 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】</li> <li>・ 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】</li> <li>・ 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】</li> </ul>

# 海外における原則・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
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"> <li>・ 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</li> </ul>	<ul style="list-style-type: none"> <li>・ 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】</li> </ul>
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"> <li>・ 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】</li> </ul>



# 海外における原則・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>

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

# 海外における原則・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>

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

# 海外における原則・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>

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

# 海外における原則・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>

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

# 海外における原則・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>

	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 < <a href="https://www.almega.se/app/uploads/sites/2/2020/05/ittelekomforetagen-industry-code-for-ai-eng.pdf">https://www.almega.se/app/uploads/sites/2/2020/05/ittelekomforetagen-industry-code-for-ai-eng.pdf</a> >
37	デンマーク	・財務省 ・産業・経済・金融省	National Strategy for Artificial Intelligence URL < <a href="https://en.digst.dk/media/19337/305755_gb_version_final-a.pdf">https://en.digst.dk/media/19337/305755_gb_version_final-a.pdf</a> >
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 < <a href="https://dataetiskraad.dk/sites/default/files/2021-10/Dataetik%20-%20S%C3%A5dan%20g%C3%B8r%20du.pdf">https://dataetiskraad.dk/sites/default/files/2021-10/Dataetik%20-%20S%C3%A5dan%20g%C3%B8r%20du.pdf</a> >
39	ドイツ	連邦政府	Artificial Intelligence Strategy of the German Federal Government URL < <a href="https://www.ki-strategie-deutschland.de/files/downloads/Fortschreibung_KI-Strategie_engl.pdf">https://www.ki-strategie-deutschland.de/files/downloads/Fortschreibung_KI-Strategie_engl.pdf</a> >
40	ドイツ	連邦経済・エネルギー省	GERMAN STANDARDIZATION ROADMAP ON ARTIFICIAL INTELLIGENCE URL < <a href="https://www.din.de/resource/blob/772610/e96c34dd6b12900ea75b460538805349/normungsroadmap-en-data.pdf">https://www.din.de/resource/blob/772610/e96c34dd6b12900ea75b460538805349/normungsroadmap-en-data.pdf</a> >
41	ドイツ	データ倫理委員会	Opinion of the Data Ethics Commission URL < <a href="https://www.bmjv.de/SharedDocs/Downloads/DE/Themen/Fokusthemen/Gutachten_DEK_EN.pdf?__blob=publicationFile&amp;v=2">https://www.bmjv.de/SharedDocs/Downloads/DE/Themen/Fokusthemen/Gutachten_DEK_EN.pdf?__blob=publicationFile&amp;v=2</a> >
42	ドイツ	BMW	BMW Group code of ethics for artificial intelligence URL < <a href="https://www.bmwgroup.com/content/dam/grpw/websites/bmwgroup_com/downloads/ENG_PR_CodeOfEthicsForAI_Short.pdf">https://www.bmwgroup.com/content/dam/grpw/websites/bmwgroup_com/downloads/ENG_PR_CodeOfEthicsForAI_Short.pdf</a> >
43	ドイツ	DAIMLER	Two Letters and Four Principles: How Daimler Uses Artificial Intelligence (AI) URL < <a href="https://www.daimler.com/sustainability/data/ki-guidelines.html">https://www.daimler.com/sustainability/data/ki-guidelines.html</a> >

# 海外における原則・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>

	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 < <a href="https://news.sap.com/2018/09/sap-guiding-principles-for-artificial-intelligence/">https://news.sap.com/2018/09/sap-guiding-principles-for-artificial-intelligence/</a> >
45	ノルウェー	地方政府・近代化省	The National Strategy for Artificial Intelligence URL < <a href="https://www.regjeringen.no/en/dokumenter/nasjonal-strategi-for-kunstig-intelligens/id2685594/?ch=7">https://www.regjeringen.no/en/dokumenter/nasjonal-strategi-for-kunstig-intelligens/id2685594/?ch=7</a> >
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 < <a href="https://teknologiradet.no/wp-content/uploads/sites/105/2018/11/AI-and-machine-learning-1.pdf">https://teknologiradet.no/wp-content/uploads/sites/105/2018/11/AI-and-machine-learning-1.pdf</a> >
47	フィンランド	経済・雇用省	Work in the age of artificial intelligence URL < <a href="https://julkaisut.valtioneuvosto.fi/bitstream/handle/10024/160980/TEMjul_21_2018_Work_in_the_age.pdf">https://julkaisut.valtioneuvosto.fi/bitstream/handle/10024/160980/TEMjul_21_2018_Work_in_the_age.pdf</a> >
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 < <a href="https://julkaisut.valtioneuvosto.fi/bitstream/handle/10024/161345/14-2019-Tekoaly%20viranomaistoiminnassa.pdf?sequence=1&amp;isAllowed=y">https://julkaisut.valtioneuvosto.fi/bitstream/handle/10024/161345/14-2019-Tekoaly%20viranomaistoiminnassa.pdf?sequence=1&amp;isAllowed=y</a> >
49	フィンランド	OP Financial Group	OP Financial Group's ethical guidelines for artificial intelligence URL < <a href="https://www.op.fi/op-financial-group/corporate-social-responsibility/commitments-and-principles">https://www.op.fi/op-financial-group/corporate-social-responsibility/commitments-and-principles</a> >
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 < <a href="https://www.cnil.fr/sites/default/files/atoms/files/cnil_rapport_ai_gb_web.pdf">https://www.cnil.fr/sites/default/files/atoms/files/cnil_rapport_ai_gb_web.pdf</a> >
51	フランス	ALLISTENE	Research Ethics in Machine Learning URL < <a href="http://cerna-ethics-allistene.org/digitalAssets/54/54730_cerna_2017_machine_learning.pdf">http://cerna-ethics-allistene.org/digitalAssets/54/54730_cerna_2017_machine_learning.pdf</a> >
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 < <a href="https://www.niti.gov.in/sites/default/files/2021-02/Responsible-AI-22022021.pdf">https://www.niti.gov.in/sites/default/files/2021-02/Responsible-AI-22022021.pdf</a> >

# 海外における原則・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>

	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 < <a href="https://www.msit.go.kr/bbs/view.do?sCode=user&amp;mPid=112&amp;mId=113&amp;bbsSeqNo=94&amp;nttSeqNo=3179742">https://www.msit.go.kr/bbs/view.do?sCode=user&amp;mPid=112&amp;mId=113&amp;bbsSeqNo=94&amp;nttSeqNo=3179742</a> >
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 < <a href="https://www.pdpc.gov.sg/-/media/Files/PDPC/PDF-Files/Resource-for-Organisation/AI/Discussion-Paper-on-AI-and-PD---050618.pdf">https://www.pdpc.gov.sg/-/media/Files/PDPC/PDF-Files/Resource-for-Organisation/AI/Discussion-Paper-on-AI-and-PD---050618.pdf</a> >
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 < <a href="https://www.mas.gov.sg/~media/MAS/News%20and%20Publications/Monographs%20and%20Information%20Papers/FEAT%20Principles%20Final.pdf">https://www.mas.gov.sg/~media/MAS/News%20and%20Publications/Monographs%20and%20Information%20Papers/FEAT%20Principles%20Final.pdf</a> >
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 < <a href="https://files-scs-prod.s3-ap-southeast-1.amazonaws.com/public/uploads/200721016+SCS+Booklet+6+Nov.pdf">https://files-scs-prod.s3-ap-southeast-1.amazonaws.com/public/uploads/200721016+SCS+Booklet+6+Nov.pdf</a> >
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 < <a href="http://www.most.gov.cn/kjbgz/201906/t20190617_147107.html">http://www.most.gov.cn/kjbgz/201906/t20190617_147107.html</a> >
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 < <a href="http://www.most.gov.cn/kjbgz/202109/t20210926_177063.html">http://www.most.gov.cn/kjbgz/202109/t20210926_177063.html</a> >
59	中国	北京智源人工智能研究院	Beijing AI Principles URL < <a href="https://www-pre.baai.ac.cn/news/beijing-ai-principles-en.html">https://www-pre.baai.ac.cn/news/beijing-ai-principles-en.html</a> >
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 < <a href="http://aiiaorg.cn/uploadfile/2019/0808/20190808053719487.pdf">http://aiiaorg.cn/uploadfile/2019/0808/20190808053719487.pdf</a> >

# 海外における原則・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>

	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 < <a href="https://ai.baidu.com/ai-doc/REFERENCE/xk3dwjgfe">https://ai.baidu.com/ai-doc/REFERENCE/xk3dwjgfe</a> >
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 < <a href="https://www.industry.gov.au/data-and-publications/australias-artificial-intelligence-ethics-framework/australias-ai-ethics-principles">https://www.industry.gov.au/data-and-publications/australias-artificial-intelligence-ethics-framework/australias-ai-ethics-principles</a> >
63	EU	欧州委員会	Artificial Intelligence, Robotics and 'Autonomous' Systems URL < <a href="https://op.europa.eu/en/publication-detail/-/publication/dfebe62e-4ce9-11e8-be1d-01aa75ed71a1/language-en/format-PDF/source-78120382">https://op.europa.eu/en/publication-detail/-/publication/dfebe62e-4ce9-11e8-be1d-01aa75ed71a1/language-en/format-PDF/source-78120382</a> >
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 < <a href="https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=68342">https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=68342</a> >
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 < <a href="https://www.eismd.eu/wp-content/uploads/2019/03/AI4People%e2%80%99s-Ethical-Framework-for-a-Good-AI-Society.pdf">https://www.eismd.eu/wp-content/uploads/2019/03/AI4People%e2%80%99s-Ethical-Framework-for-a-Good-AI-Society.pdf</a> >
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 < <a href="https://unesdoc.unesco.org/ark:/48223/pf0000373434?posInSet=6&amp;queryId=ed3e3a0d-e278-46a2-898f-3e24d43f75cc">https://unesdoc.unesco.org/ark:/48223/pf0000373434?posInSet=6&amp;queryId=ed3e3a0d-e278-46a2-898f-3e24d43f75cc</a> >
67	国際機関	世界保健機関 (WHO)	Ethics and governance of artificial intelligence for health URL < <a href="https://apps.who.int/iris/rest/bitstreams/1352854/retrieve">https://apps.who.int/iris/rest/bitstreams/1352854/retrieve</a> >



# 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

- 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）における各事業者の取組等を踏まえて、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、原則等の内容が妥当・適切であるか等について検討。

項目・事項	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）	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"> <li>○ 倫理の原則               <ul style="list-style-type: none"> <li>・ 人間の尊厳と個人の自律を尊重する</li> </ul> </li> </ul> <p>【AI利活用ガイドライン】</p> <ul style="list-style-type: none"> <li>○ 基本理念               <ul style="list-style-type: none"> <li>・ 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</li> </ul> </li> <li>○ 尊厳・自律の原則               <ul style="list-style-type: none"> <li>・ 人間の尊厳と個人の自律を尊重する</li> </ul> </li> <li>○ 公平性の原則               <ul style="list-style-type: none"> <li>・ 個人及び集団が不当に差別されないよう配慮する</li> </ul> </li> </ul>	<ul style="list-style-type: none"> <li>・ 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。</li> <li>・ 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。</li> <li>・ 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。</li> </ul>
セキュリティ	<p>【AI開発ガイドライン】</p> <ul style="list-style-type: none"> <li>○ セキュリティの原則               <ul style="list-style-type: none"> <li>・ AIシステムのセキュリティに留意する</li> </ul> </li> </ul> <p>【AI利活用ガイドライン】</p> <ul style="list-style-type: none"> <li>○ セキュリティの原則               <ul style="list-style-type: none"> <li>・ AIシステム又はAIサービスのセキュリティに留意する</li> </ul> </li> <li>○ 適正学習の原則               <ul style="list-style-type: none"> <li>➢ 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</li> </ul> </li> </ul>	<ul style="list-style-type: none"> <li>・ 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。</li> <li>・ AIとセキュリティの関係を考えた場合、(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。</li> </ul>

# 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

項目・事項	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）	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"> <li>○ プライバシーの原則           <ul style="list-style-type: none"> <li>・ 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</li> </ul> </li> <li>○ 利用者支援の原則           <ul style="list-style-type: none"> <li>・ 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</li> </ul> </li> <li>○ アカウンタビリティの原則           <ul style="list-style-type: none"> <li>・ 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</li> </ul> </li> </ul> <p>【AI利活用ガイドライン】</p> <ul style="list-style-type: none"> <li>○ プライバシーの原則           <ul style="list-style-type: none"> <li>・ 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</li> </ul> </li> <li>○ アカウンタビリティの原則           <ul style="list-style-type: none"> <li>・ 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</li> </ul> </li> </ul>	<ul style="list-style-type: none"> <li>・ 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。</li> <li>・ 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。</li> <li>・ 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。</li> </ul>
公平性	<p>【AI開発ガイドライン】</p> <ul style="list-style-type: none"> <li>○ 倫理の原則           <ul style="list-style-type: none"> <li>・ 人間の尊厳と個人の自律を尊重する</li> </ul> </li> </ul> <p>【AI利活用ガイドライン】</p> <ul style="list-style-type: none"> <li>○ 公平性の原則           <ul style="list-style-type: none"> <li>・ 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</li> </ul> </li> </ul>	<ul style="list-style-type: none"> <li>・ 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。</li> <li>・ 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。</li> <li>・ 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。</li> <li>・ 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。</li> <li>・ 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。</li> </ul>

# 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

項目・事項	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）	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"> <li>○ 透明性の原則               <ul style="list-style-type: none"> <li>・ 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</li> </ul> </li> <li>○ アカウンタビリティの原則               <ul style="list-style-type: none"> <li>・ 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</li> </ul> </li> </ul> <p>【AI活用ガイドライン】</p> <ul style="list-style-type: none"> <li>○ 透明性の原則               <ul style="list-style-type: none"> <li>・ 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</li> </ul> </li> <li>○ アカウンタビリティの原則               <ul style="list-style-type: none"> <li>・ 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</li> </ul> </li> <li>○ 適正利用の原則               <ul style="list-style-type: none"> <li>➢ 必要な情報を適時に提供することが期待される</li> </ul> </li> </ul>	<ul style="list-style-type: none"> <li>・ 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。</li> <li>・ 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。</li> <li>・ 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。</li> <li>・ 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。</li> <li>・ 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。</li> <li>・ 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。</li> </ul>
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"> <li>○ 制御可能性の原則               <ul style="list-style-type: none"> <li>➢ 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</li> </ul> </li> </ul> <p>【AI活用ガイドライン】</p> <ul style="list-style-type: none"> <li>○ 適正利用の原則               <ul style="list-style-type: none"> <li>・ 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</li> </ul> </li> </ul>	<ul style="list-style-type: none"> <li>・ 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。</li> <li>・ 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。</li> <li>・ 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。</li> <li>・ 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載せたり、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。</li> </ul>

# 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

項目・事項	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）	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"> <li>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。</li> <li>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。</li> <li>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。</li> </ul>
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"> <li>○ 適正利用の原則           <ul style="list-style-type: none"> <li>➢ 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</li> </ul> </li> <li>○ 適正学習の原則           <ul style="list-style-type: none"> <li>➢ 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</li> </ul> </li> </ul>	<ul style="list-style-type: none"> <li>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。</li> <li>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。</li> <li>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。</li> <li>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。</li> <li>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）</li> <li>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。</li> <li>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）。</li> <li>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。</li> <li>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。</li> </ul>

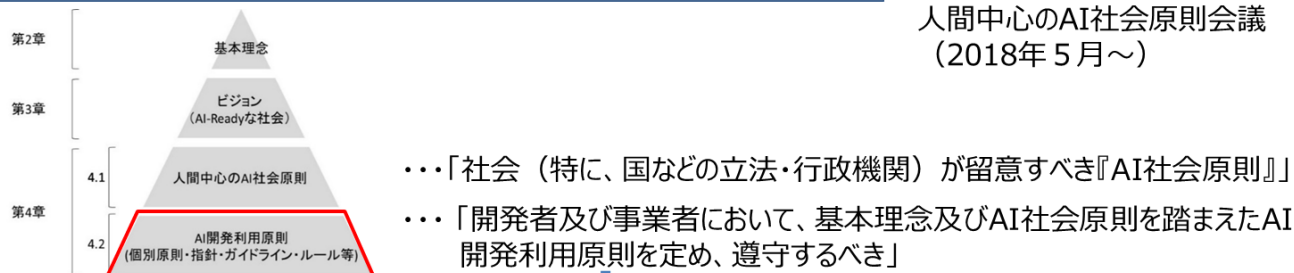
# AI開発ガイドライン及びAI利活用ガイドラインに関するレビュー

## (その他の論点)

- 「組織・体制」、「人材育成」などについても、各事業者等において取組が行われているが、これらの項目・事項を追加する必要があるかどうか。
- 有事（パンデミック、災害等）を想定した見直しを行う必要があるかどうか（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。）。
- AIの開発のハードルが低くなってきている（特段の専門的な知識・経験がなくてもAIを開発することができる環境が整ってきている）ことを踏まえて、現行のガイドラインを見直す必要があるかどうか。
- データの取扱いに関し、横断的なガイドラインを策定する必要があるかどうか。

## <参考>

「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



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。

## 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