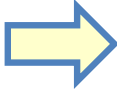


A I 利活用原則案

平成30年7月31日

総務省情報通信政策研究所

背景

- 今後、A Iの研究開発・利活用の進展、A Iの相互連携・ネットワーク形成（A Iネットワーク化）が想定
 - A Iは様々な分野で利活用され、そのサービスはネットワークを通じた（国境を越えた）提供が想定
 - A Iは人間・社会に多大な便益を広範にもたらすことが期待されるとともに、リスクの抑制も図ることが重要。
-  ・ A Iの便益の増進、リスクの抑制のための取組について**中長期的な視点**で検討が必要
- ・ **産学民官の幅広い関係者の参画を得て**、国際的にも議論することが重要

A Iネットワーク社会推進会議

目的・検討事項

A Iネットワーク化に関して、社会的・経済的・倫理的・法的課題に関する事項を検討。具体的には、国際的な議論のためのA I開発ガイドライン案の検討・フォローアップ

A Iの利活用について考慮すべき事項（論点整理）

A Iネットワーク化が社会・経済にもたらす影響

A Iネットワーク化の進展に伴う将来社会の展望 等

検討体制

- 【議長】 須藤修（東京大学大学院情報学環教授・東京大学総合教育研究センター長）
- 【構成員】 産学民の有識者（関係学会の会長経験者、関係企業の会長又は社長等）
- 【オブザーバ】 関係行政機関、関係国立研究開発法人 等

A I ネットワーク社会推進会議 構成員（親会のみ）

議長	須藤 修	（東京大学大学院情報学環教授・東京大学総合教育研究センター長）		
副議長	三友 仁志	（早稲田大学国際学術院大学院アジア太平洋研究科教授）		
構成員	岩本 敏男	（株式会社エヌ・ティ・ティ・データ相談役）	中川 裕志	（理化学研究所革新知能統合研究センターグループディレクター）
	遠藤 信博	（日本電気株式会社代表取締役会長）	長田 三紀	（全国地域婦人団体連絡協議会事務局長）
	大橋 弘	（東京大学大学院公共政策大学院・経済学研究科教授）	中西 崇文	（武蔵野大学工学部数理工学科准教授）
	大屋 雄裕	（慶應義塾大学法学部教授）	西田 豊明	（京都大学大学院情報学研究科教授）
	金井 良太	（株式会社アラヤ代表取締役CEO）	萩田 紀博	（株式会社国際電気通信基礎技術研究所 知能ロボティクス研究所長）
	北野 宏明	（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）	橋元 良明	（東京大学大学院情報学環教授）
	喜連川 優	（東京大学生産技術研究所教授、国立情報学研究所所長）	林 秀弥	（名古屋大学大学院法学研究科教授）
	エリー キーナン	（日本アイ・ビー・エム株式会社代表取締役社長執行役員）	東原 敏昭	（株式会社日立製作所代表執行役 執行役社長兼CEO）
	木村 たま代	（主婦連合会消費者相談室長）	平野 晋	（中央大学総合政策学部教授・ 中央大学大学院総合政策研究科委員長）
	小塚 莊一郎	（学習院大学法学部法学科教授）	平野 拓也	（日本マイクロソフト株式会社代表取締役社長）
	近藤 則子	（老テク研究会事務局長）	福田 雅樹	（大阪大学大学院法学研究科教授）
	穴戸 常寿	（東京大学大学院法学政治学研究科教授）	堀 浩一	（東京大学大学院工学系研究科教授）
	実積 寿也	（中央大学総合政策学部教授）	松尾 豊	（東京大学大学院工学系研究科特任准教授）
	城山 英明	（東京大学大学院法学政治学研究科教授）	村井 純	（慶應義塾大学環境情報学部長・教授）
	新保 史生	（慶應義塾大学総合政策学部教授）	村上 憲郎	（株式会社村上憲郎事務所代表取締役）
	杉原 佳堯	（グーグル合同会社執行役員 公共政策担当）	森川 博之	（東京大学大学院工学系研究科教授）
	杉山 将	（理化学研究所革新知能統合研究センター長、 東京大学新領域創成科学研究科教授）	柳川 範之	（東京大学大学院経済学研究科教授）
	鈴木 晶子	（京都大学大学院教育学研究科教授）	山川 宏	（株式会社ドワンゴ ドワンゴ人工知能研究所長）
	高橋 恒一	（理化学研究所生命機能科学研究センターチームリーダー）	山本 正巳	（富士通株式会社取締役会長）
	谷崎 勝教	（株式会社三井住友銀行取締役専務執行役員グループCIO）		（敬称略。議長及び副議長を除き、五十音順）
顧問	安西 祐一郎	（慶應義塾大学名誉教授）	オブザーバー	内閣府、内閣官房情報通信技術（IT）総合戦略室、 個人情報保護委員会事務局、消費者庁、文部科学省、経済産業省、 情報通信研究機構、科学技術振興機構、理化学研究所、 産業技術総合研究所
	長尾 真	（京都大学名誉教授）		
	西尾 章治郎	（大阪大学総長）		
	濱田 純一	（東京大学名誉教授）（敬称略。五十音順）		

検討の経緯

AIネットワーク化検討会議 (平成28年2月～6月)

- ・AIネットワーク化をめぐる社会的・経済的・倫理的・法的課題を整理
- ・平成28年4月29日～30日、G7香川・高松情報通信大臣会合において、「AI開発原則」のたたき台として、次の8項目を紹介
 - 透明性の原則
 - 利用者支援の原則
 - 制御可能性の原則
 - セキュリティ確保の原則
 - 安全保護の原則
 - プライバシー保護の原則
 - 倫理の原則
 - アカウントビリティの原則
- 引き続き、『AI開発原則』の策定を含め、AIに関する国際的な議論及び検討を進めることに参加各国が賛同
- ・平成28年4月『中間報告書』を公表
平成28年6月『報告書2016』を公表

AIネットワーク社会推進会議 (平成28年10月～)

- ・「検討会議」を発展的に改組し、整理された課題に関する内容の検討を継続して実施
 - 国際的な議論のためのAI開発ガイドライン案の検討
 - AIネットワーク化が社会・経済にもたらす影響(インパクト及びリスク)の評価
- ・平成28年12月28日から約1か月間、『AI開発ガイドライン』の策定に向けて整理した論点に関し、意見募集を実施
 - 提出意見:26件
- ・平成29年3月13日・14日に、国際シンポジウム「AIネットワーク社会推進フォーラム」を開催
 - ・米、伊、独、仏、EUの政府関係者
 - ・OECDの代表者
 - ・国際的な議論をリードする主要団体の代表者
 を含む国内外のトップレベルの有識者の参加を得て国際的な議論を推進
- ・平成29年6月14日から約1ヶ月間、意見募集を実施し、同年7月、国際的な議論のためのAI開発ガイドライン案を含む「報告書2017」を公表

AIネットワーク社会推進会議 (平成29年10月～)

- ・AIネットワーク社会推進会議を開催し、以下について検討
 - 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ネットワーク化が社会・経済にもたらす影響の評価(継続)
 - AI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(「AI利活用原則案」及び各原則に関する論点整理) 等
- ・平成30年6月12日から約2週間、意見募集を実施し、同年7月、AI利活用原則案を含む「報告書2018」を公表
- ・報告書の内容を「人間中心のAI社会原則検討会議」に紹介するとともに、国際的な議論の用に供する予定
- ・推進会議においても継続して検討を実施

序章

第1章 AIネットワーク化をめぐる最近の動向

第2章 AIネットワーク化が社会・経済にもたらす影響

分野別評価

AIシステムを利活用する分野ごとにインパクト・リスクを評価(AIネットワーク化による産業構造の変化に着目(主として利用者側の視点))

<対象とする分野>

- 公共: まちづくり、**行政**、危機管理
- 個人: **健康**、**移動**、**居住**、教育・学び、仕事、財産、趣味・娯楽
- 産業: **モノ**、**カネ**

報告書2018においては、下線の3分野について評価を実施

エコシステムの展望

<利用シーンにおけるAI利活用の展望>

- ・都市部における利用: 移動、介護など
- ・地方部における利用: 移動、医療など
- ・家庭内における利用: 医療、家事など
- ・企業における利用: ヒト、モノなど

<ケーススタディ>

- ・ケース: 完全自動運転
- ・ケース: 医療診断
- ・ケース: 人事評価・採用
- ・ケース: 金融(保険)
- ・ケース: 行政

ネットワーク化の観点からの課題

- ・個別の事業者のトラブル等がシステム全体に波及するおそれ
- ・AIシステム間の連携・調整が成立しないなどのおそれ
- ・システム全体としてAIの判断・意思決定を検証できないおそれ
- ・少数のAIの影響力が強くなりすぎるなどのおそれ
- ・基盤的なAIへの情報の集中等によるプライバシー侵害のおそれ
- ・AIが想定外の動作を行うなどのおそれ

教育・人材育成

- ・人間とAIとの関係・役割分担
- ・社会の変化及び人間の技能・能力
- ・学校教育
- ・職業スキルの習得
- ・生涯教育
- ・AIの利活用

第3章 AIの利活用において留意することが期待される事項

AI利活用原則案

関係する主体の整理

「AI利活用原則案」(10原則)

各原則の内容に関する論点整理

【国内の議論及び国際的な議論の推進】

- ・人間中心のAI社会原則検討会議
- ・OECD AIに係る専門家会合
- ・G7マルチステークホルダーカンファレンス 等

第4章 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の利活用に関する指針の策定 等
- 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、プライバシー及びパーソナルデータの保護 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: シナリオ分析、評価指標の設定、AIの利活用に関する社会的受容性の醸成
- 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 等

シナリオ分析(分野別評価)

AIネットワーク化が社会・経済にもたらすインパクト(主に良い影響、便益)及びリスクに関し、AIシステムの具体的な利活用の場面(ユースケース)を想定したシナリオに基づき評価(シナリオ分析)を実施

➤ AIシステムを利活用する分野ごとに利用者の視点から評価(本報告書においては、下線の3分野について評価を実施)

<対象とするユースケース>

「まちづくり」、「健康」、「モノ」については、報告書2017で評価を実施

公共：まちづくり、**行政**、危機管理

個人：**健康**、**移動**、**居住**、教育・学び、仕事、財産、趣味・娯楽

産業：モノ、カネ

シナリオ分析から得られた示唆

- ・ 自動運転により取得される移動履歴やカメラ映像等について、プライバシーに配慮しつつ、データの積極的な利活用が期待される。
- ・ 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。
- ・ 特に高齢者に有益なAIシステムの利活用と考えられるため、高齢者のリテラシーを向上させる方策が求められる。 等

エコシステムの展望

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を展望し、AIの利活用における便益や課題を整理

<利用シーンにおけるAIの利活用の展望>

都市部における利用：移動、介護、観光・旅行、教育・人材育成

地方部における利用：移動、医療、仕事、コミュニティ

家庭内における利用：医療、家事、安全・快適な居住環境、豊かな人生

企業における利用：ヒト、モノ、カネ、情報

<ケーススタディ>

ケース：完全自動運転

ケース：医療診断

ケース：人事評価・採用

ケース：金融(保険)

ケース：行政

AIネットワーク化の進展に関する考察

主としてAIネットワーク化の健全な進展を阻害し得る要因(AIがネットワーク化することにより生ずる又は増幅する課題)の考察

個別の事業者のトラブル等がシステム全体に波及するおそれ

AIシステム間の連携・調整が成立しないなどのおそれ

AIの判断・意思決定を検証できないおそれ(システム間の相互作用が複雑となり解析が困難になるおそれ)

少数のAIの影響力が強くなりすぎるなどのおそれ(少数のAIの判断によって企業や個人が不利な立場になるなどのおそれ)

領域横断での情報の共有と特定の基盤的なAIへの情報の集中によるプライバシー侵害のおそれ

AIが想定外の動作を行うなどのおそれ



< エコシステムの展望、A I ネットワーク化の進展に関する考察において整理された課題 >

【主として生命・身体の安全、権利・利益等を守るための課題】

- ・ 生命・身体・財産の安全に関する課題(事故の防止など) : 安全
- ・ AIによる判断の正当性や公平性に関する課題(差別、生命倫理との関係など) : 正当性・公平性
- ・ プライバシーに関する課題(プライバシーの尊重、プロファイリングなど) : プライバシー

【主として人間とA I との関係等に関する課題】

- ・ 人間とAIとの役割分担等に関する課題(人間の判断の介在、関係者間の協力など) : 役割分担
- ・ AIに対する受容性に関する課題(利用者に対する説明責任など) : 受容性

【主として技術的な観点からの解決が求められる課題】

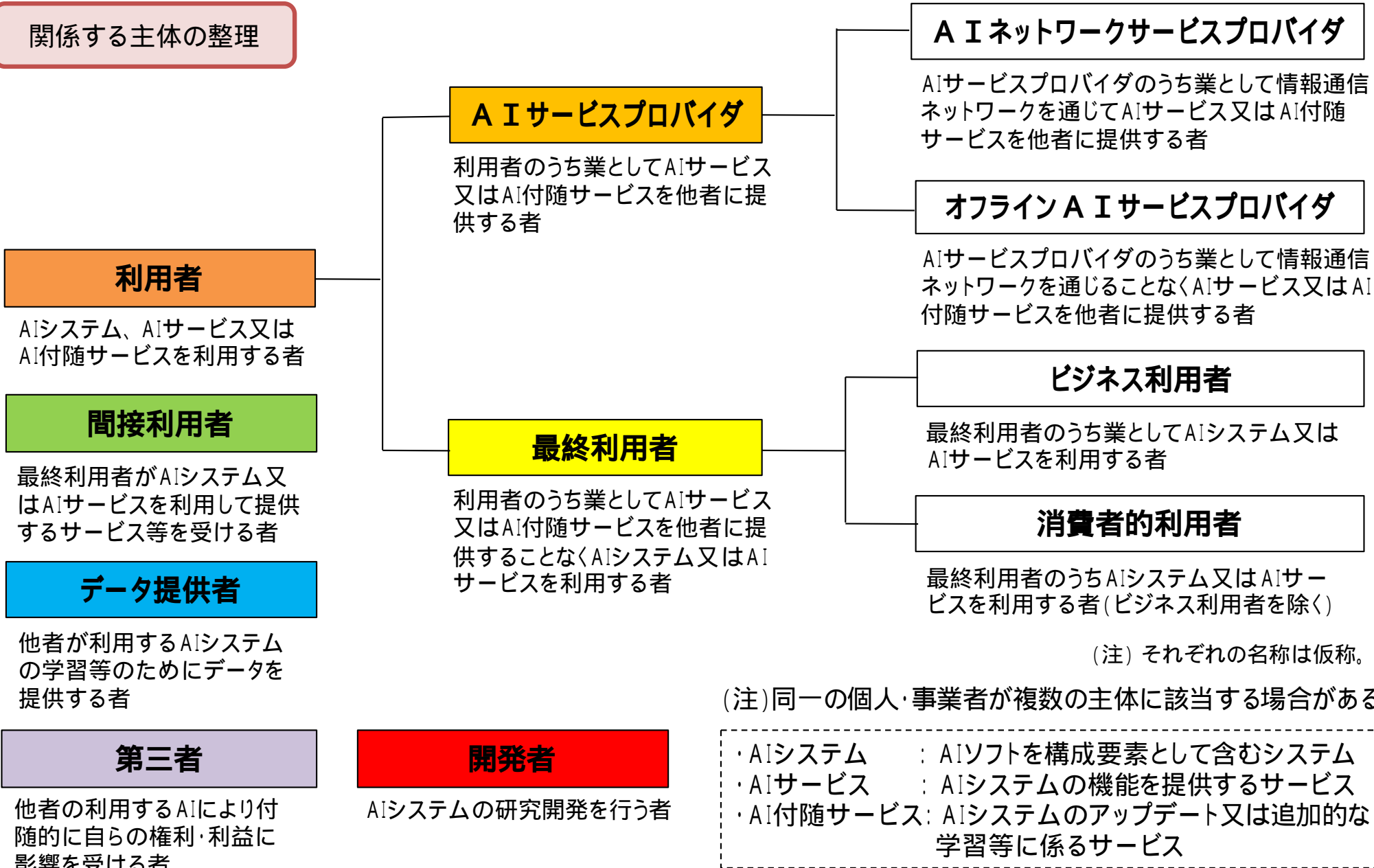
- ・ AIの判断のブラックボックス化に関する課題(事故が発生した場合の原因究明など) : ブラックボックス化
- ・ セキュリティに関する課題(ハッキング対策など) : セキュリティ
- ・ AI間の連携に関する課題(AI間の交渉・調整など) : 連携

【主としてデータに関する課題】

- ・ AIが学習するデータに関する課題(データの正確性など) : データ

これらの課題を踏まえ、
「AI利活用原則案」
を取りまとめ

関係する主体の整理



(注) それぞれの名称は仮称。

(注) 同一の個人・事業者が複数の主体に該当する場合がある。

- ・AIシステム : AIソフトを構成要素として含むシステム
- ・AIサービス : AIシステムの機能を提供するサービス
- ・AI付随サービス: AIシステムのアップデート又は追加的な学習等に係るサービス

AI利活用原則案

国際的な議論のためのものとして、また、非規制的かつ非拘束的なもの(いわゆるソフトロー)として取りまとめ

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。

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。

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。

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。

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。

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。

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。

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サービスの判断によって個人が不当に差別されないよう配慮する。

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サービスの入出力の検証可能性及び判断結果の説明可能性に留意する。

アカウントビリティの原則

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。

主に
便益の増進
に関係

主に
リスクの抑制
に関係

主に
信頼の醸成
に関係

AI利活用原則案の各原則の内容に関する主な論点

AI利活用原則案について、各原則の内容について論点を整理。これらの論点に関して、継続して検討。

AI利活用原則案	各原則の内容に関する主な論点(例)
<p>適正利用の原則</p>	<ul style="list-style-type: none"> ・ 利用者は、開発者等からの情報提供や説明を踏まえ、社会的文脈や状況にも配慮して、AIを適正な範囲・方法で利用することが期待されるのではないかと。また、利用者は、AIの性質、利用の態様等に応じて、利用する前に、便益及びリスクを認識し、適正な用途を理解するとともに、必要な知識・技能を習得すること等が期待されるのではないかと。さらに、利用者は、自らのAIの利活用が適正な範囲・方法で行われているか定期的に確認することが期待されるのではないかと。 ・ AIサービスプロバイダ及びビジネス利用者は、AIのリスクを適切に評価した上で、生産性の向上や業務の効率化のためにAIの積極的な利活用を検討するなど、AIの便益とリスクの適正なバランスに配慮することが期待されるのではないかと。
<p>適正学習の原則</p>	<ul style="list-style-type: none"> ・ 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(正確性や完全性など)に留意することが期待されるのではないかと。また、AIによりなされた判断の精度が損なわれたり、低下した場合には、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のではないかと。どのような場合に、どの程度、学習等に用いるデータの質に留意することを期待することが適当か。
<p>連携の原則</p>	<ul style="list-style-type: none"> ・ AIネットワークサービスプロバイダは、利用するAIの特性及び用途を踏まえ、AIネットワーク化の健全な進展を通じて、AIの便益を増進するため、AIの相互接続性と相互運用性に留意することが期待されるのではないかと。
<p>安全の原則</p>	<ul style="list-style-type: none"> ・ 医療や自動運転など人の生命・身体・財産に危害を及ぼし得る分野でAIを利活用する場合には、利用者は、想定される被害の性質・態様等を踏まえ、必要に応じてAIの点検・修理及びAIソフトのアップデートを行うことなどにより、AIがアクチュエータ等を通じて人の生命・身体・財産に危害を及ぼすことのないよう配慮することが期待されるのではないかと。また、利用者は、AIがアクチュエータ等を通じて人の生命・身体・財産に危害を及ぼした場合に講ずるべき措置について、あらかじめ整理しておくことが期待されるのではないかと。
<p>セキュリティの原則</p>	<ul style="list-style-type: none"> ・ 利用者は、AIのセキュリティに留意し、その時点での技術水準に照らして合理的な対策を講ずることが期待されるのではないかと。また、利用者は、セキュリティが侵害された場合に講ずるべき措置について、あらかじめ整理しておくことが期待されるのではないかと。
<p>プライバシーの原則</p>	<ul style="list-style-type: none"> ・ 利用者は、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、AIの利活用において他者のプライバシーを尊重することが期待されるのではないかと。また、他者のプライバシーを侵害した場合に講ずるべき措置について、あらかじめ整理しておくことが期待されるのではないかと。
<p>尊厳・自律の原則</p>	<ul style="list-style-type: none"> ・ 利用者は、AIを利活用する際の社会的文脈を踏まえ、人間の尊厳と個人の自律を尊重することが期待されるのではないかと。 ・ 利用者には、AIにより意思決定や感情が操作されるリスクや、AIに過度に依存するリスクに留意することを期待することは適当か否か。
<p>公平性の原則</p>	<ul style="list-style-type: none"> ・ 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の判断結果により、人種・信条・性別等によって個人が不当に差別されないよう、AIの学習等に用いられるデータの代表性やデータに内在する社会的なバイアスに留意することが期待されるのではないかと。 ・ 利用する技術の特性及び用途に照らして、どのような場合に、どの程度、学習等に用いるデータの代表性やデータに内在する社会的なバイアスに留意することを期待することが適当か。
<p>透明性の原則</p>	<ul style="list-style-type: none"> ・ AIサービスプロバイダ及びビジネス利用者は、AIの入出力の検証可能性を確保するため、入出力を記録・保存することが期待されるのではないかと。 ・ 利用する技術の特性及び用途に照らして、どのような場合に、どの程度、入出力を記録・保存することを期待することが適当か。例えば、自動運転車など人の生命・身体・財産に危害を及ぼし得る分野で利活用する場合には、事故の原因究明や再発防止に必要な範囲において、AIの入出力を記録・保存することが期待されるのではないかと。
<p>アカウントビリティの原則</p>	<ul style="list-style-type: none"> ・ AIサービスプロバイダ及びビジネス利用者は、人々と社会からAIへの信頼を獲得することができるよう、消費者的利用者、間接利用者、AIの利活用により影響を受ける第三者等に対し、利用するAIの性質及び目的等に照らして、相応のアカウントビリティを果たすよう努めることが期待されるのではないかと。

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々の検討

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

(参考) 海外における主な取組

