

A I の利活用において 留意することが期待される事項

平成30年4月26日
事 務 局

AIの利活用において利用者等が留意することが期待される事項について

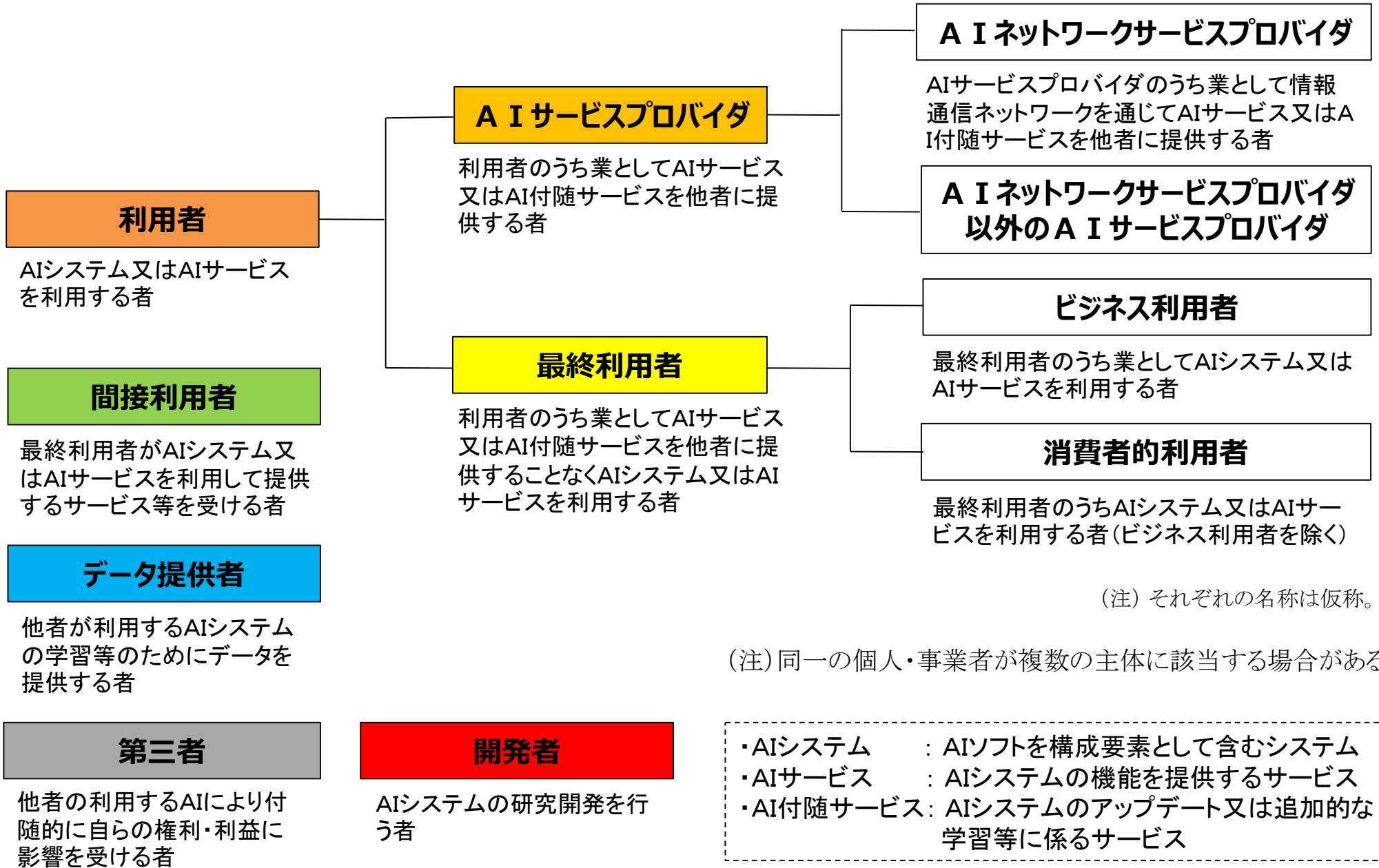
- AIの開発・利活用の加速化、AIネットワーク化の健全な進展等には、A I ネットワーク化の便益の増進・リスクの抑制のための取組が必要。
- これまでの取組として、本推進会議において、A I の開発者が留意することが期待される事項として「国際的な議論のためのA I 開発ガイドライン案」（以下「A I 開発ガイドライン案」という。）を取りまとめ。
また、海外においては、主に開発者の視点からA I に関する原則・指針・ベストプラクティス・標準等に関する取組が進行。
（例）米国電気電子学会（IEEE）「倫理的に調整された設計 第2版」（2017年12月12日）
- 他方、A I は、利活用の過程でデータの学習等により自らの出力やプログラムを継続的に変化させる可能性があることから、A I の利用者、間接利用者及びデータ提供者（以下「利用者等」という。）においても留意することが期待される事項があるものと想定される。
→ これらの事項を整理することは、開発者との関係からも重要であるものと考えられる。
- これらを踏まえ、本推進会議において、「分野別評価」や「A I ネットワーク化の進展に伴い形成されるエコシステムの展望」において抽出された課題等をもとに、A I の利活用において利用者等が留意することが期待される事項について検討することとする。
→ 『報告書2018（仮）』において、「A I 利活用原則案」を取りまとめるとともに、それぞれの原則の内容に関する論点を整理することとする。その後、整理された論点を踏まえ、引き続き、最終的なアウトプットの取りまとめに向けた検討を行うこととする。

関係する主体の整理 (1/7)

<基本的な考え方>

- A I 開発ガイドライン案において、「利用者」とは、“A I システムを利用する者（最終利用者（エンドユーザ）のほか、他者が開発したA I ネットワークサービスを第三者に提供するプロバイダを含む。）をいう。”と定義されているのみであり、利用者には様々な立場の者が含まれる。
また、A I の利活用に当たっては、A I システム又はA I サービス（以下単に「A I」ということがある。）を直接的に利用しない場合でも、A I を用いたサービス（例：医療サービスや金融サービス）を利用（A I を間接的に利用）する場合が想定される。
- A I 開発ガイドライン案において、「A I システム」とは、“A I ソフトを構成要素として含むシステムをいう。例えば、A I ソフトを実装したロボットやクラウドシステムはこれに含まれる。”とされている。他方、「A I サービス」については定義されていないところ、「A I サービス」を“A I システムの機能を提供するサービス”とし、また、「A I 付随サービス」を“A I 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”としてはどうか。
- これらを踏まえると、「分野別評価」や「A I ネットワーク化の進展に伴い形成されるエコシステムの展望」において抽出された課題等に対して、A I の利活用において利用者等が留意することが期待される事項を検討するに当たっては、利用者等について一定の分類を試み、分類された利用者の類型に沿って検討を行うことが適当であるものと考えられる。
- このため、A I の提供について外形的に観察した上で、利用者の分類を試みることとする（次ページ参照）。
- なお、利用者の分類について、
 - ・ 特に消費者的利用者及び間接利用者について、A I であることを認識している者と認識していない者とを区別すべきではないか。
 - ・ 消費者的利用者について、A I システムを保有・管理している場合やA I の入出力を調整することができる場合と、こうした事情が認められない場合とを区別すべきではないか。
 - ・ 子ども、高齢者、障害者など社会的弱者については、特別に留意することが期待される事項が考えられるのではないか。といった意見があることを踏まえると、今後「A I 利活用原則案」に掲げられた各原則の内容に関し整理された論点を検討する上では、利用者の主観面や立場等も考慮する必要があるものと考えられる。
- A I の提供及び利活用の形態については、様々なケースが想定されることから、必要に応じて、今後の検討において関係する主体の整理を見直すこともあり得る。

関係する主体の整理 (2/7)



関係する主体の整理 (3/7)

例① <医療用A Iクラウド>

開発者/A Iサービスプロバイダ
(例: 医療用A Iクラウド開発者)



自ら開発したA Iシステムを利用したA Iサービス
(例: 医療用A Iクラウドサービス)を他者
(例: 医師)に提供※しており、「開発者」である
とともに、「A Iサービスプロバイダ」にも該当



A Iサービス (例: 医療用A I
クラウドサービス)の提供※

※ AI付随サービス(アップデート・追加的な学習等)も提供

最終利用者
(例: 医師)



A Iサービス (例: 医療用A Iクラウドサービス)
を利用する者であり、「最終利用者」に該当



A Iサービス (例: 医療用A I
クラウドサービス)を利用したサービ
ス (例: 医療サービス)の提供

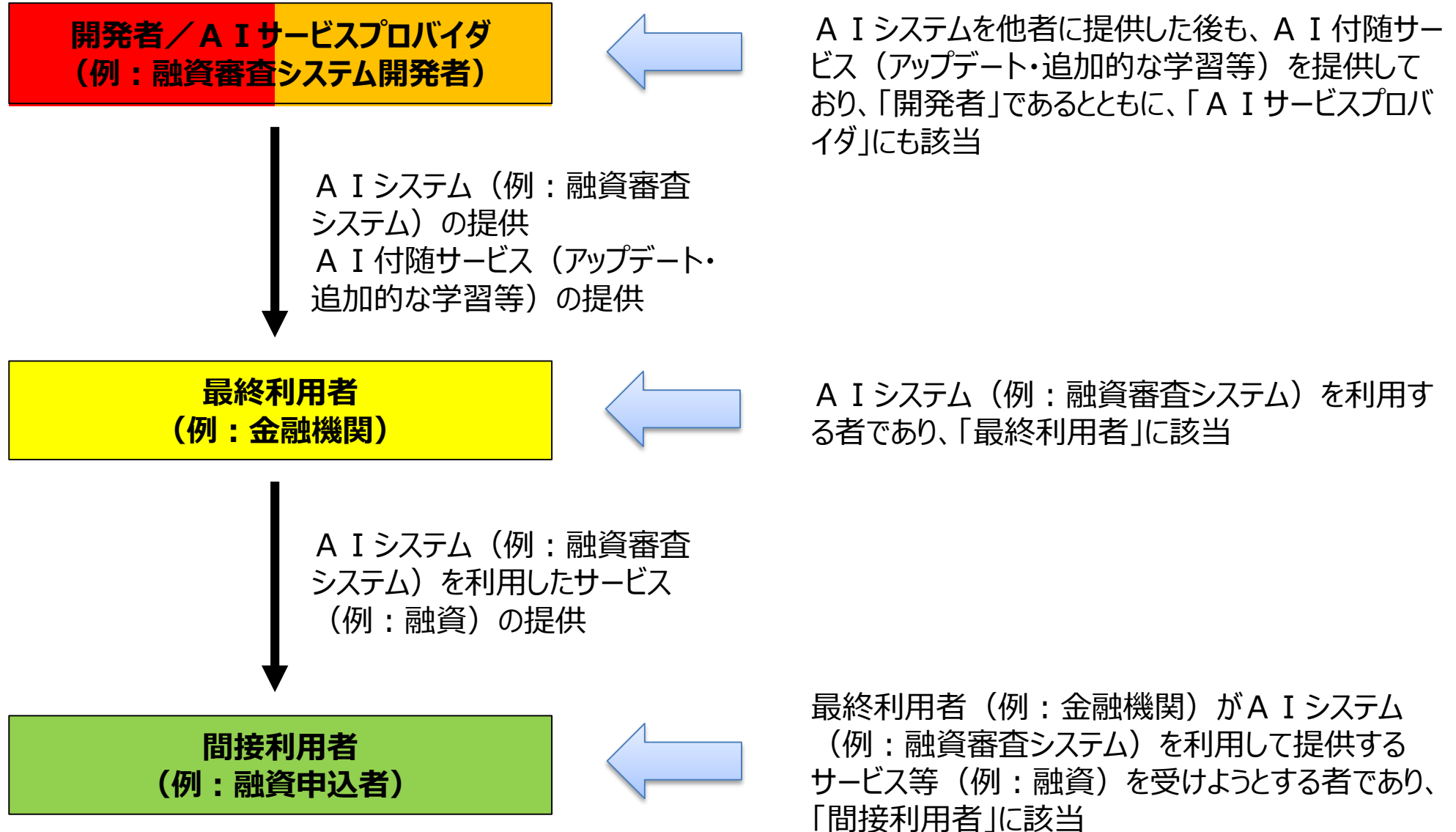
間接利用者
(例: 患者)



最終利用者 (例: 医師)がA Iサービス (例:
医療用A Iクラウドサービス)を利用して提供する
サービス (例: 医療サービス)を受ける者であり、
「間接利用者」に該当

関係する主体の整理 (4/7)

例② <融資審査>



関係する主体の整理 (5/7)

例③ <自動運転タクシー>

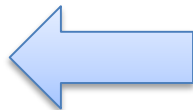
開発者/AIサービスプロバイダ
(例: 自動運転車製造者)



AIシステムを他者(例: 自動運転タクシー業者)に提供した後も、AI付随サービス(アップデート・追加的な学習等)を提供しており、「開発者」であるとともに「AIサービスプロバイダ」にも該当

↓
AIシステム(例: 自動運転車)の提供
AI付随サービス(アップデート・追加的な学習等)の提供

AIサービスプロバイダ
(例: 自動運転タクシー業者)



AIシステム(例: 自動運転車)を利用する者であるが、AIシステムを用いたAIサービス(例: 自動運転タクシー)を他者(例: 乗客)に提供する者であり、「AIサービスプロバイダ」に該当

↓
AIサービス(例: 自動運転タクシー)の提供

最終利用者
(例: 乗客)



AIサービス(例: 自動運転タクシー)を利用する者であり、「最終利用者」に該当

↓
事故発生時の被害等

第三者
(例: 通行人)

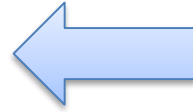


利用者には該当しないが、事故発生時の被害等権利・利益への影響に配慮する必要

関係する主体の整理 (6/7)

例④ <生産ラインの異常検知>

開発者/AIサービスプロバイダ
(例: 異常検知システム開発者)

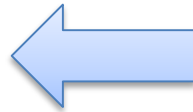


自ら開発したAIシステムを利用したAIサービス
(例: 異常検知サービス)を他者(例: 製造メーカー等)に提供※しており、「開発者」であるとともに、「AIサービスプロバイダ」にも該当

AIサービス(例: 異常検知サービス)の提供※

※ AI付随サービス(アップデート・追加的な学習等)も提供

最終利用者
(例: 製造メーカー等)

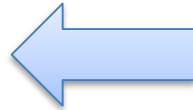


AIサービス(例: 異常検知サービス)を利用する者であり、「最終利用者」に該当

関係する主体の整理 (7/7)

例⑤ <A I アシスタント>

開発者/A I サービスプロバイダ
(例: A I アシスタント開発者)



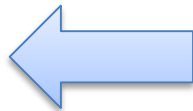
自ら開発したA I システムを利用したA I サービス (例: A I アシスタントサービス) を他者 (例: A I アシスタント対応のA I スピーカーを利用する個人) に提供※しており、「開発者」であるとともに、「A I サービスプロバイダ」にも該当

※ AI付随サービス(アップデート・追加的な学習等)も提供

A I サービス (例: A I アシスタントサービス) の提供※



最終利用者
(例: A I アシスタント対応のA I スピーカーを利用する個人)

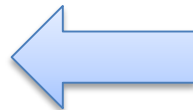


A I サービス (例: A I アシスタントサービス) を利用する者であり、「最終利用者」に該当

プライバシーへの影響等



第三者
(例: 最終利用者と同居する家族、最終利用者宅を訪問する友人等)



利用者には該当しないが、プライバシーへの影響等権利・利益への影響に配慮する必要

AI利活用原則案 (1/2)

○ 「分野別評価」や「AIネットワーク化の進展に伴い形成されるエコシステムの展望」において抽出された課題等を踏まえ、AIの利活用の促進やAIネットワーク化の健全な進展に向けて、AIシステムの便益の増進とリスクの抑制を図り、AIシステムに対する信頼を醸成するため、AIシステムの利活用において留意することが期待される事項を「AI利活用原則案」として取りまとめることとしたい。

⇒ AIの便益及びリスクは、国境を越えて波及することが見込まれることから、**国際的な議論のためのもの**として取りまとめることが適当であるものと考えられる。また、「AI利活用原則案」は、AI開発ガイドライン案と同様に、**非規制的かつ非拘束的なもの**として取りまとめることが適当であるものと考えられる。

① 適正利用の原則	[安全][役割分担]	… 主に <u>便益の増進</u> に関係
② 積極利用の原則		
③ 適正学習の原則	[データ][正当性・公平性]	
④ 連携の原則	[連携]	
⑤ 安全の原則	[安全]	… 主に <u>リスクの抑制</u> に関係
⑥ セキュリティの原則	[セキュリティ]	
⑦ プライバシーの原則	[プライバシー]	
⑧ 尊厳・自律の原則	[正当性・公平性]	
⑨ 公平性の原則	[正当性・公平性]	… 主に <u>信頼の醸成</u> に関係
⑩ 透明性の原則	[ブラックボックス化]	
⑪ アカウンタビリティの原則	[受容性]	
⑫ 協力の原則	[安全][役割分担]	

(注1) 上記の項目は、あくまでも現時点における「AI利活用原則案」であり、最終的なアウトプットにおいては、利用者の類型ごとに、上記の項目の一部のみが盛り込まれる可能性や上記の項目以外の項目が盛り込まれる可能性が考えられる。

(注2) 各項目に関係する便益及びリスクには、利用者本人に影響を与えるもののみならず、他者や社会全体に波及するものも含まれる。

AI活用原則案 (2/2)

① 適正利用の原則

利用者は、人間とAIとの適切な役割分担のもと、適正な範囲及び方法でAIシステム又はAIサービスを利用するよう努める。

② 積極利用の原則

AIサービスプロバイダ及びビジネス利用者は、AIのリスクを過大に評価することなく、AIシステム又はAIサービス並びにデータを積極的に利活用するよう努める。

③ 適正学習の原則

利用者及びデータ提供者は、AIの学習等に用いるデータの正確性、バイアス等に留意する。

④ 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 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。

⑥ 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。

⑦ 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。

⑧ 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。

⑨ 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り差別的な取扱いがなされないよう配慮する。

⑩ 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。

⑪ アカウンタビリティの原則

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タビリティを果たすよう努める。

⑫ 協力の原則

利用者は、関係者と協力して事故の未然防止及び事故の発生時等の対応に取り組むよう努める。

それぞれの原則の内容に関する論点 (1/12)

① 適正利用の原則

利用者は、人間とAIとの適切な役割分担のもと、適正な範囲及び方法でAIシステム又はAIサービスを利用するよう努める。

【主な論点】

A) 適正な範囲・方法での利用

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、リスクを認識し、正しい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。

B) AIソフトのアップデート及びAIの点検・修理等

利用者は、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・修理等を行うよう努めることが期待されるのではないかと。

C) 人間の判断の介在

AIの自動的な判断について、必要かつ可能な場合(例: AIを用いた医療診断等)には、人間の判断を介在させることが期待されるのではないかと。その場合、人間の判断の介在の要否の基準をどのように考えるか。

また、AIの判断から人間の判断に移行する場合において、どのような事項に留意することを期待することが適当か。

[基準として考えられる観点(例)]

- ・ AIの判断に影響を受ける間接利用者等の権利・利益の性質及び意向
- ・ AIの判断の信頼性の程度(人間の判断の信頼性との優劣)
- ・ 人間の判断に必要な時間的猶予
- ・ 利用者に期待される能力

など

② 積極利用の原則

AIサービスプロバイダ及びビジネス利用者は、AIのリスクを過大に評価することなく、AIシステム又はAIサービス並びにデータを積極的に利活用するよう努める。

【主な論点】

A) 積極的な利活用

AIサービスプロバイダ及びビジネス利用者は、AIのリスクを過大に評価することなく、生産性の向上や業務の効率化のために、現に提供され又は提供が見込まれるAIの特性等に鑑み、AIやデータを積極的に利活用することが期待されるのではないか。

③ 適正学習の原則

利用者及びデータ提供者は、AIの学習等に用いるデータの正確性、バイアス等に留意する。

【主な論点】

A) AIの学習等に用いるデータの正確性への留意

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正確性に留意することが期待されるのではないか。また、学習したデータの正確性が損なわれた場合には、データの正確性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。

また、どのような場合に、どの程度、学習等に用いるデータの正確性に留意することを期待することが適当か。

B) AIの学習等に用いられるデータのバイアスへの留意

AIの判断結果により、個人が人種・信条・性別等によって差別的に取り扱われることのないよう、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の学習等に用いられるデータのバイアスに留意することが期待されるのではないか。

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バイアスに留意することを期待することが適当か。

C) 不正確又は不適切なデータの学習等によるAIのセキュリティ脆弱性への留意

利用者及びデータ提供者は、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ずるリスクに留意することが期待されるのではないか。

④ 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がネットワーク化することによってリスクが惹起・増幅される可能性があることに留意する。

【主な論点】

A) 相互接続性と相互運用性への留意

AIネットワークサービスプロバイダは、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。

B) データ形式やプロトコル等の標準化への対応

AIサービスプロバイダ及びビジネス利用者は、AI相互間及びAIと他のシステム等との連携を促進するため、データ形式やプロトコル等の標準化に対応することが期待されるのではないか。また、データ提供者についても、データ形式の標準化に対応することが期待されるのではないか。

C) AIネットワーク化により惹起・増幅される課題への留意

AIが連携することによって便益が増進することが期待されるが、利用者は、自ら利用するAIがインターネット等を通じて他のAI等と接続・連携することにより制御不能となる等、AI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ことが期待されるのではないか。

⑤ 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に配慮する。

【主な論点】

A) 適正な範囲・方法での利用

利用者は、開発者等からの情報提供や説明を踏まえ、AIを安全に利活用するために、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。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、リスクを認識し、正しい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。

B) 人の生命・身体・財産への配慮

医療診断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、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。

⑥ セキュリティの原則

利用者及びデータ提供者は、AIシステム又はAIサービスのセキュリティに留意する。

【主な論点】

A) セキュリティ対策の実施

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。

B) セキュリティ対策のためのサービス提供等

AIネットワークサービスプロバイダは、自ら提供するAIネットワークサービスについて、最終利用者にセキュリティ対策のためのサービスを提供するとともに、インシデント情報の共有を図ることが期待されるのではないか。

C) 不正確又は不適切なデータの学習によるAIのセキュリティ脆弱性への留意

利用者及びデータ提供者は、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ずるリスクに留意することが期待されるのではないか。

⑦ プライバシーの原則

利用者及びデータ提供者は、AIシステム又はAIサービスの利活用において、他者又は自己のプライバシーが侵害されないよう配慮する。

【主な論点】

A) 他者のプライバシーの尊重

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、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。

B) パーソナルデータの収集・分析・提供等におけるプライバシーの尊重

利用者及びデータ提供者は、AIの学習等に用いられるパーソナルデータの収集・分析・提供等において、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。

C) AIを利用したプロファイリングを行う場合におけるプライバシー等の配慮

AIサービスプロバイダ及びビジネス利用者は、人事評価・採用や融資など個人の権利・利益に重要な影響を及ぼす可能性のある分野においてAIを利用したプロファイリングを行う場合には、対象者のプライバシー等に慎重に配慮することが期待されるのではないか。

D) 自己のプライバシー侵害への留意

消費者的利用者は、ペットロボットなどのAIに過度に感情移入すること等により、特に秘匿性の高い情報をむやみにAIに与えることのないよう留意することが期待されるのではないか。

⑧ 尊厳・自律の原則

利用者は、AIシステム又はAIサービスの利活用において、人間の尊厳と個人の自律を尊重する。

【主な論点】

A) 他者の尊厳と自律の尊重

利用者は、AIを利活用する際の社会的文脈を踏まえ、他者の尊厳と自律を尊重することが期待されるのではないか。

B) AIによる意思決定・感情の操作等への留意

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。

また、このようなリスクについて誰がいかなる役割を果たすべきか。

C) AIと人間の脳・身体を連携する際の生命倫理等の議論の参照

利用者は、AIを人間の脳・身体と連携させる場合には、生命倫理の議論等を参照し、人間の尊厳と自律が侵害されないよう特に慎重な配慮が求められるのではないか。

⑨ 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り差別的な取扱いがなされないよう配慮する。

【主な論点】

A) AIの学習等に用いられるデータのバイアスへの留意

AIの判断結果により、個人が人種・信条・性別等によって差別的に取り扱われることのないよう、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の学習等に用いられるデータのバイアスに留意することが期待されるのではないか。

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バイアスに留意することを期待することが適当か。

B) 人間の判断の介在

AIの判断結果により、個人が人種・信条・性別等によって差別的に取り扱われることのないよう、AIサービスプロバイダ及びビジネス利用者は、評価対象者の意向に応じて、人間の判断を介在させることが期待されるのではなか。

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、人間の判断を介在させることを期待することが適当か。

それぞれの原則の内容に関する論点（10／12）

⑩ 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。

【主な論点】

A) AIの入出力の記録・保存

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。

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、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。

B) 説明可能性の確保

AIサービスプロバイダ及びビジネス利用者は、AIの判断結果の説明可能性を確保することが期待されるのではないか。

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、説明可能性を確保することを期待することが適当か。特に、医療診断、人事評価・採用、融資など個人の権利・利益に重大な影響を及ぼす可能性のある分野において利活用する場合には、AIの判断結果の説明可能性を確保することが期待されるのではないか。

（例えば、深層学習は予測精度が高いが、判断結果の説明が困難であるという現状等にも留意が必要）。

⑪ アカウンタビリティの原則

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タビリティを果たすよう努める。

【主な論点】

A) アカウンタビリティを果たす努力

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、相応のアカウンタ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。

B) AIに関する利用方針の通知・公表

AIサービスプロバイダ及びビジネス利用者は、消費者的利用者や間接利用者らがAIの利活用について、適切に認識することのできるよう、AIに関する利用方針（AIを利用している旨、適正な利活用の範囲及び方法、利活用に伴うリスク、相談窓口など）を通知又は公表することが期待されるのではないか。

利用する技術の性質及び用途に照らして、どのような場合に、どのような内容の利用方針を通知又は公表することを期待することが適切か。

⑫ 協力の原則

利用者は、関係者と協力して事故の未然防止及び事故の発生時等の対応に取り組むよう努める。

【主な論点】

A) 事故等による被害に関する対策の実施

利用者は、AIの利活用により生じ得る又は生じた事故、セキュリティ侵害、プライバシー侵害等による被害の性質・態様等に応じて、関係者と協力して予防措置及び事後対応（情報共有、停止・復旧、原因解明、再発防止措置等）に取り組むことが期待されるのではないか。また、その実効性を確保するために利用者にどのようなことを期待することが適切か。

なお、事故等が発生した場合に、原因解明を進め、被害者を救済し、早期に適切な再発防止対策を行うことができるようにするために、捜査に協力した者については、刑事責任を免除するなどの措置を講ずることが適切か否か。

B) 能力・知識の多寡等の考慮

利用者は、予防措置及び事後対応を行うに当たり、それぞれの主体がAIに関して有していると期待される能力及び知識の多寡並びに必要な措置を実施する容易性を踏まえ、適切な対応を取ることが期待されるのではないか。

【参考①】 国際的な議論のためのAI開発ガイドライン案の概要

目的

- AIネットワーク化の進展による**便益は広範で多大** その一方、**不透明化等のリスク**への懸念も存在
- **拘束的ではないソフトロー**としてガイドラインを**国際的に共有**することが必要
- 本ガイドラインの目的

AIネットワーク化の健全な進展を通じたAIシステムの**便益の増進**と**リスクの抑制**

- ・ **利用者の利益を保護**
- ・ **リスクの波及を抑止**

人間中心の「**智連社会**」を実現

基本理念

- ・ AIネットワーク化の**便益がすべての人によりあまねく享受**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
- ・ **拘束的ではないソフトロー**として**国際的に共有**
- ・ **イノベティブでオープンな研究開発**と**公正な競争**、学問の自由等を尊重するとともに、**便益とリスクの適正なバランス**を確保
- ・ **技術的中立性**を確保し、開発者に**過度の負担**を課さないよう配慮
- ・ **継続的な見直し**、必要に応じた**柔軟な改定**、広範で柔軟な議論

AI開発原則

- ① **連携の原則**
 - ・ AIシステムの**相互接続性**と**相互運用性**に留意
- ② **透明性の原則**
 - ・ 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
- ③ **制御可能性の原則**
 - ・ AIシステムの**制御可能性**に留意
- ④ **安全の原則**
 - ・ 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
- ⑤ **セキュリティの原則**
 - ・ AIシステムの**セキュリティ**に留意
- ⑥ **プライバシーの原則**
 - ・ 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
- ⑦ **倫理の原則**
 - ・ AIシステムの開発において、**人間の尊厳**と**個人の自律**を尊重
- ⑧ **利用者支援の原則**
 - ・ AIシステムが利用者を支援し、**利用者**に**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
- ⑨ **アカウントビリティの原則**
 - ・ 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

関係者に期待される役割

- ・ 各国政府及び国際機関： 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
- ・ 開発者、利用者等ステークホルダ： 対話やベストプラクティスの共有、相互協力
- ・ 各国政府： AIの開発者コミュニティの支援、AIに関する研究開発を支援する政策の積極的な推進

【参考②】 国際的な議論のためのAI開発ガイドライン案における定義

- 「AI」とは、「AIソフト及びAIシステムを総称する概念」をいう。
- 「AIソフト」とは、データ・情報・知識の学習等により、利活用の過程を通じて自らの出力やプログラムを変化させる機能を有するソフトウェアをいう。例えば、機械学習ソフトウェアはこれに含まれる。
- 「AIシステム」とは、AIソフトを構成要素として含むシステムをいう。例えば、AIソフトを実装したロボットやクラウドシステムはこれに含まれる。
- AIシステムの「開発者」とは、AIシステムの研究開発（AIシステムを利用しながら行う研究開発を含む。）を行う者（自らが開発したAIシステムを用いてAIネットワークサービスを他者に提供するプロバイダを含む。）をいう。
- AIシステムの「利用者」とは、AIシステムを利用する者（最終利用者（エンドユーザ）のほか、他者が開発したAIネットワークサービスを第三者に提供するプロバイダを含む。）をいう。

今後の検討の方向性

<「AI利活用原則案」に関する本推進会議による検討のアウトプットについて>

- 『報告書2018(仮)』において、「AI利活用原則案」を取りまとめるとともに、それぞれの原則の内容に関する論点を整理することとする。その後、整理された論点を踏まえ、最終的なアウトプットの取りまとめに向けた検討を行うこととする。
なお、この取りまとめに向けた検討においては、次に掲げるアウトプットの在り方に関する事項についても、併せて検討することとする。
 - i それぞれの原則の内容に関して整理した論点について、「ガイドライン」という形や名称で取りまとめることとするか。
 - ii 分野ごとの問題の性質等、国ごとの法制度等に相違があることに鑑み、分野共通のものとして取りまとめるのではなく、各分野のステークホルダ等が分野ごとのガイドライン等を策定する際に参照されることが期待される基本的な指針(メタガイドライン)として取りまとめることとし、そのための論点・用語の定義等を提示することが適当か。
 - iii 項目ごとに整理するか、さらに利用者等の類型ごとに整理するか。
 - iv 利用者等が参照するガイドライン等を策定する者が参照するものとするに加え、利用者等が直接に参照するものまで含めるか。(利用者(特に消費者的利用者)向けに、「ハンドブック」や「マニュアル」などリテラシー教材(利用者の手引き)を作成したり、それらに基づいてワークショップ等を行ったりすること等も考えられる。)