

別添 2

報告書（案）  
別紙 1

# AI利活用ガイドライン

令和元年●月●日

AIネットワーク社会推進会議

## 目次

序文 .....	2
AI の定義及び対象範囲 .....	5
1. 目的 .....	6
2. 基本理念 .....	8
3. 関係する主体の整理 .....	9
4. AI 利活用原則 .....	10
5. 一般的な AI 利活用の流れ .....	12
6. AI 利活用原則の解説 .....	14
7. AI 利活用原則を考慮すべきタイミング .....	27

### (附属資料)

AI 利活用原則の各論点に対する詳説

## 序文

AI に関する研究開発や利活用は今後飛躍的に発展することが期待されており、AI に対する信頼を醸成するための議論が国際的に実施されている。我が国では、政府において、「AI-Ready な社会」への変革を推進する観点から、2018 年 5 月より、政府統一の AI 社会原則に関する検討を開始し、2019 年（平成 31 年）3 月に「人間中心の AI 社会原則」<sup>1</sup>を策定した。

こうした AI に関する原則についての検討は、2016 年（平成 28 年）4 月に日本で開催された G7 情報通信大臣会合において、ホスト国である日本が AI 開発原則のたたき台を紹介したことが契機となった。同会合では、各国関係閣僚による議論が行われ、G7 において「AI 開発原則」及びその内容の解説からなる「AI 開発ガイドライン」の策定に向け、引き続き G7 各国が中心となり、OECD（経済協力開発機構）等国際機関の協力も得て議論していくことで合意がなされた。AI ネットワーク社会推進会議（以下、「本推進会議」という。）では、AI の便益の増進及びリスクの抑制のため研究開発において留意することが期待される事項について検討を行い、G7 や OECD における国際的な議論の基礎となる文書として、2017 年（平成 29 年）7 月、「国際的な議論のための AI 開発ガイドライン案」をとりまとめ、公表した。

他方、AI は、利活用の過程でデータの学習等により自らの出力やプログラムを継続的に変化させる可能性があることから、開発者が留意することが期待される事項のみならず、利用者が AI の利活用において留意することが期待される事項も想定される。また、AI の利活用において留意することが期待される事項を整理することは、開発者と利用者、データ提供者など様々なステークホルダに期待される役割を検討していく上でも重要であるものと考えられる。このため、本推進会議では、AI サービスプロバイダ、最終利用者及びデータ提供者が留意することが期待される原則として、2018 年（平成 30 年）7 月、「AI 利活用原則案」をとりまとめた。

また、我が国は、国際社会においてもこれらの検討成果を積極的に発信し、特に、OECD が理事会勧告案を策定するために 2018 年（平成 30 年）9 月に設置した専門家会合（AI expert Group at the OECD）においては、我が国から参加した有

---

<sup>1</sup> 「人間中心の AI 社会原則」（2019 年（平成 31 年）3 月 統合イノベーション戦略推進会議決定）

識者より、人間の中心の AI 社会原則、AI 開発ガイドライン案、AI 利活用原則案のそれぞれについて、その内容のみならず、検討の背景や検討にあたり行われた議論の状況についても紹介するなど、OECD 理事会勧告の原案策定にあたり大きな貢献を行った。このため、2019 年（令和元年）5 月に公表された OECD 理事会勧告（Recommendation on Artificial Intelligence）は、我が国において検討してきた上述の各原則、ガイドラインと整合のとれたものとなっている。

さらに、日本が議長国を努めた G20 貿易・デジタル経済大臣会合（2019 年 6 月）においては、OECD 理事会勧告を採用した「G20AI 原則」が取りまとめられている。

このように、AI に関する原則については、我が国が国際的な議論を主導してきた結果、その概念については、概ね国際的にコンセンサスが得られつつある。他方で、今後は、この「原則」をどのように実現していくべきか、そのためには各ステークホルダがどのような措置を講ずるべきかといった、具体的な措置について検討が進められることが期待されている。「人間中心の AI 社会原則」では、開発者及び事業者において、同原則で定める基本理念及び AI 社会原則を踏まえた AI 開発利用原則を定め、遵守することが期待されており、そのためには、参照すべき具体的な解説書が必要である。

本ガイドラインは、AI の便益の増進及びリスクの抑制のため、利活用において留意することが期待される原則、すなわち、2018 年（平成 30 年）に本推進会議が提案した「AI 利活用原則案」をベースに、これを「AI 利活用原則」として整理した上で、それぞれの原則についての解説をとりまとめたものである<sup>2,3</sup>。原則の解説にあたっては、その原則を実現するために講ずべき措置について可能な限り具体的な記載を試みている。本ガイドラインは、マルチステークホルダの参画を得て取りまとめられたものであり、AI の利活用に向けて様々な立場のステークホルダが議論をする際に共通に参照することが可能な文書である。

本ガイドラインは、AI の利用者、特に業として AI サービス等を他者に提供

---

<sup>2</sup> 本ガイドラインでは、「AI 利活用原則」に基づきその具体策の解説を行っているが、OECD 理事会勧告の各原則を実現しようとする場合でも、それに対応する本ガイドライン中の記載を参照することで、その具体策を把握することができる。

<sup>3</sup> 本ガイドラインは、ソフトローとしてとりまとめた AI 利活用原則に、その原則を実現する観点からの解説を加えたものである。本ガイドラインは、法令のように記載された事項をその文言とおりに適用をしていくという性質のものではなく、AI の利用者が AI を利用するに当たって自主的な対応を講ずることが期待される事項をとりまとめたものである。

する者（AI サービスプロバイダ）や業として AI システム等を利用する者（ビジネス利用者）が、人間中心の AI 社会原則を踏まえた AI 開発利用原則を定める際に本ガイドラインを参照することにより、AI の利活用に関して留意すべき事項を適切に認識し、それら留意事項への対応について自主的に検討することを促そうとするものである。さらに言えば、AI サービスプロバイダやビジネス利用者がこうした自主的な取組を実施することで、提供する AI サービスや AI を活用した業務に対して付加価値を与えることも可能となりうる。

AI の利用者が、本ガイドラインを参照し、AI の利用目的や利活用する際の社会的文脈に応じた適切な措置を自主的に講じることにより、関連するステークホルダの利益の保護が図られるとともにリスクの波及が抑止されることが期待される。さらに、AI サービスプロバイダ、ビジネス利用者等の自主的な取組によって、消費者的利用者や第三者からの信頼が醸成され、それが AI システムや AI サービスの提供に対する需要につながり、AI の利活用や社会実装が促進されることが期待される。


### パート1：AI利活用原則の考え方

**1. 目的**  
AIの利活用・社会実装の促進

**2. 基本理念**

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

**3. 関係する主体の整理**  
「利用者」の分類



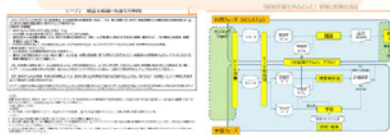
**4. AI利活用原則**  
10原則  
適正利用、適正学習、連携、安全、セキュリティ、プライバシー、  
尊厳・自律、公平性、透明性、アカウンタビリティ

### パート2：AI利活用原則の解説

**5. 一般的なAI利活用の流れ**  
AI利活用の一般的な流れ



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



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




図 1：AI 利活用ガイドラインの構成

## AI の定義及び対象範囲

### (1) AI の定義

本ガイドラインにおける「AI」については、以下のとおり定義する。

「AI」とは、「AI ソフト及び AI システムを総称する概念」をいう<sup>4</sup>。

- 「AI ソフト」とは、データ・情報・知識の学習等<sup>5</sup>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。

### (2) 対象範囲

本ガイドラインの対象とする**AI システムの範囲**は、AI システムがネットワークを通じて国境を越えて利用され、広く人間及び社会に便益やリスクをもたらす可能性があることから、ネットワーク化され得る AI システム（ネットワークに接続可能な AI システム）とする。

---

<sup>4</sup> 本ガイドラインにおける AI の定義は、現在既に実用化されている特化型 AI を主たる対象として想定しているが、自律性を有する AI や汎用 AI (Artificial General Intelligence) の開発など今後予想される AI に関する急速な技術発展を見据え、今後開発される多種多様な AI についても、学習等により自らの出力やプログラムを変化させる機能を有するものである場合には、含み得るものとしている。

本ガイドラインにおいては、上述のとおり AI を定義しており、今後開発される多種多様な AI についてもその機能次第で含み得るものとしている。なお、本ガイドラインにおける AI の定義の在り方については、AI の技術発展の動向等を踏まえ、今後継続的に議論を行っていくことが必要である。

<sup>5</sup> 学習以外の方法により AI ソフトが自らの出力やプログラムを変化させる要因としては、例えば、データ・情報・知識に基づく推論や、センサやアクチュエータ等を通じた環境とのインタラクションなどが考えられる。

## 1. 目的

AIの研究開発や利活用は、今後急速に進展することが期待されているところであり、AIネットワーク化<sup>6</sup>が進展していく過程で、個人、地域社会、各国、国際社会<sup>7</sup>の抱える様々な課題の解決に大きく貢献するなど、人間及びその社会や経済に多大な便益を広範にもたらすことが期待される。このような方向に向けて、我が国としても、諸外国に遅れをとらないよう、AIの研究開発や利活用を加速化していくことが求められる。

その一環として、AIが社会や経済にもたらす便益の増進を図るとともに、不透明化や制御喪失などAIに関するリスクの抑制を図る観点から、関連する社会的・経済的・倫理的・法的な課題に対応することが必要となる。特に、AIを利活用するサービスは、他の情報通信サービス同様、ネットワークを通じて国境を越えて提供されるものであることから、多様なステークホルダ（開発者、サービス提供者、市民社会を含む利用者、各国政府、国際機関など）によるオープンな議論を通じ、国際的なコンセンサスを醸成することにより、AIの便益の増進とリスクの抑制を図ることが求められる。

以上の問題意識に鑑み、本ガイドラインは、AIネットワーク化の健全な進展を通じて、AIの便益の増進とリスク<sup>8</sup>の抑制を図り、AIに対する信頼を醸成することにより、AIの利活用や社会実装が促進することを目的とする。

そこで、本ガイドラインでは、上記の目的を達成する観点から、2018年（平成30年）に本推進会議が提案した「AI利活用原則案」をベースに、AIの利用者がAIを利活用するにあたり留意すべき事項を「AI利活用原則」としてとりまとめ、その実現のために講ずることが期待される具体的な措置についての解説を行った<sup>9</sup>。本ガイドラインは、AIサービスプロバイダやビジネス利用者等がAI社会原則を踏まえたAI開発利用原則を定める際の解説書と位置づけられる

---

<sup>6</sup> AIがインターネットその他の情報通信ネットワークと接続され、AI相互間又はAIと他の種類のソフトウェアもしくはシステムとの間のネットワーク（以下において「AIネットワーク」という場合がある。）が形成されるようになることをいう。

<sup>7</sup> 国際社会の抱える課題については、国連の「持続可能な開発目標」（SDGs）（[http://www.un.org/ga/search/view\\_doc.asp?symbol=A/70/L.1](http://www.un.org/ga/search/view_doc.asp?symbol=A/70/L.1)）などを参照。

<sup>8</sup> 「リスク」とは「損害等の不利益をもたらす可能性」を意味する。

<sup>9</sup> 注2参照。

ものである。

なお、AIの利活用の目的や用途は様々であることから、必ずしもあらゆるAIの利活用においてAI利活用原則として整理された10の原則の全てに留意することが期待されるわけではない。AIの利用者は、AIの利用目的や利活用する際の社会的文脈を踏まえ、10の原則の中から自らが留意すべき原則を選定し、その上で、選定したそれぞれの原則について、後に記載する解説を参照し、どのような措置をどの程度講ずべきか自主的に検討することが想定されている<sup>10,11</sup>。

また、AIの研究開発において留意することが期待される事項については、2017年（平成29年）7月に「AI開発ガイドライン案」としてとりまとめられているところである。開発と利活用とは、必ずしも明確に区分できない場合も想定され、AI開発ガイドライン案とセットとして参照されることが望ましい。

---

<sup>10</sup> その際には、附属資料（参考2）に記載したトレードオフに留意しつつ検討がなされることが望ましい。

<sup>11</sup> 自主的にAI原則を策定し、公表する企業が増えてきているほか、AIの利活用にあたり社内におけるAIの取扱い方針等を整備することにより、社内の倫理委員会の了承を得た事例などが生じてきている。



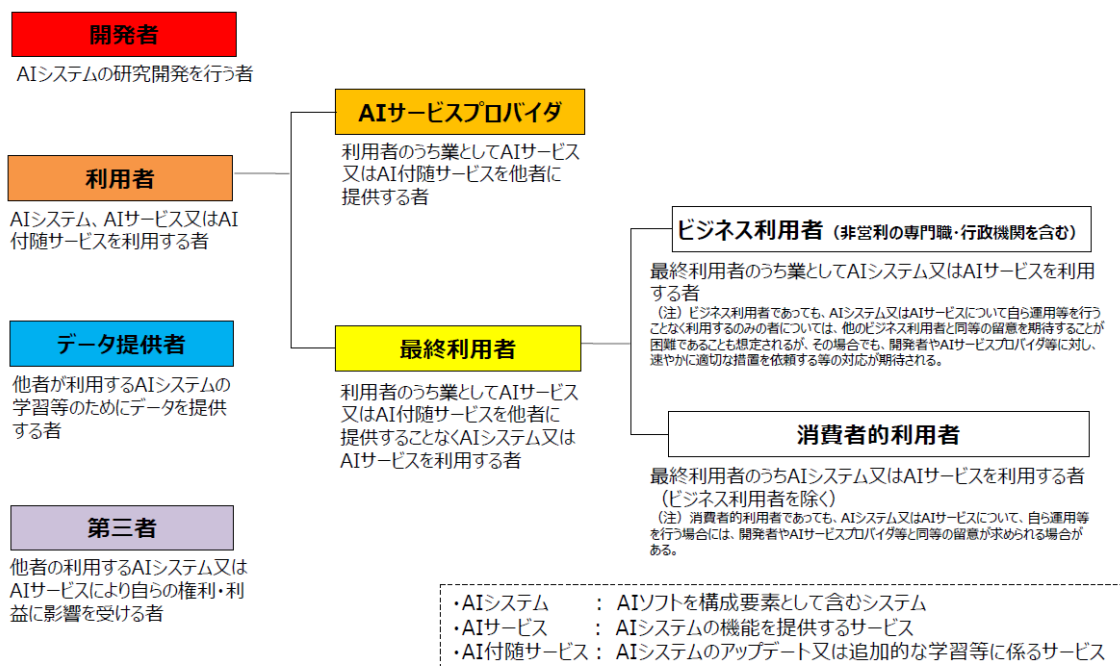
## 2. 基本理念

本ガイドラインの目的に鑑み、次に掲げる理念を一体的なものとして後に述べる AI 利活用原則の基本理念とする。

- 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
- **AI ネットワーク化により個人、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
- AI ネットワーク化による便益を増進するとともに、民主主義社会の価値を最大限尊重しつつ、権利利益が侵害されるリスクを抑制するため、**便益とリスクの適正なバランスを確保**すること
- AI に関して有していると期待される**能力や知識等に応じ**、**ステークホルダ間における適切な役割分担を実現**すること
- 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
- AI ネットワーク化の進展等を踏まえ、国際的な議論を通じて、本ガイドラインを**不断に見直し**、必要に応じて**柔軟に改定**すること

### 3. 関係する主体の整理

AI の利活用において関与が想定される者は以下の通りである<sup>12</sup>。



(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。

図 2： 関係する主体の整理

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

<sup>12</sup> ここでは、AI システム又は AI サービスの提供における立場の違いに基づき行った典型的な分類を記載している。そのため、自社で AI サービスを開発・提供し、当該サービスを自社事業に利活用する場合には、開発者としての性格と、ビジネス利用者としての性格を合わせもつこととなるが、その場合には、「AI 開発ガイドライン案」と本ガイドラインとをセットとして参照されることが望ましい。

最終利用者のうち業として AI システム又は AI サービスを利用する者。

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

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

#### 4. AI 利活用原則

以上の目的及び基本理念を踏まえ、AI の利用者が留意すべき事項を以下のとおり 10 の原則に整理する。

##### ① 適正利用の原則

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

##### ② 適正学習の原則

利用者及びデータ提供者は、AI システムの学習等に用いるデータの質に留意する。

##### ③ 連携の原則

AI サービスプロバイダ、ビジネス利用者及びデータ提供者は、AI システム又は AI サービス相互間の連携に留意する。また、利用者は、AI システムがネットワーク化することによってリスクが惹起・増幅される可能性があることに留意する。

#### ④ 安全の原則

利用者は、AI システム又は AI サービスの利活用により、アクチュエータ等を通じて、利用者及び第三者の生命・身体・財産に危害を及ぼすことがないよう配慮する。

#### ⑤ セキュリティの原則

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

#### ⑥ プライバシーの原則

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

#### ⑦ 尊厳・自律の原則

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

#### ⑧ 公平性<sup>13</sup>の原則

AI サービスプロバイダ、ビジネス利用者及びデータ提供者は、AI システム又は AI サービスの判断にバイアスが含まれる可能性があることに留意し、また、AI システム又は AI サービスの判断によって個人が不当に差別されないよう配慮する。

#### ⑨ 透明性の原則<sup>14</sup>

AI サービスプロバイダ及びビジネス利用者は、AI システム又は AI サービスの入出力等の検証可能性及び判断結果の説明可能性に留意する。

#### ⑩ アカウンタビリティ<sup>15</sup>の原則

利用者は、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。

---

<sup>13</sup> 「公平性」には複数の定義・基準があることに留意する必要がある。

<sup>14</sup> 本原則は、アルゴリズム、ソースコード、学習データの開示を想定するものではない。また、本原則の解釈に当たっては、プライバシーや営業秘密への配慮も求められる。

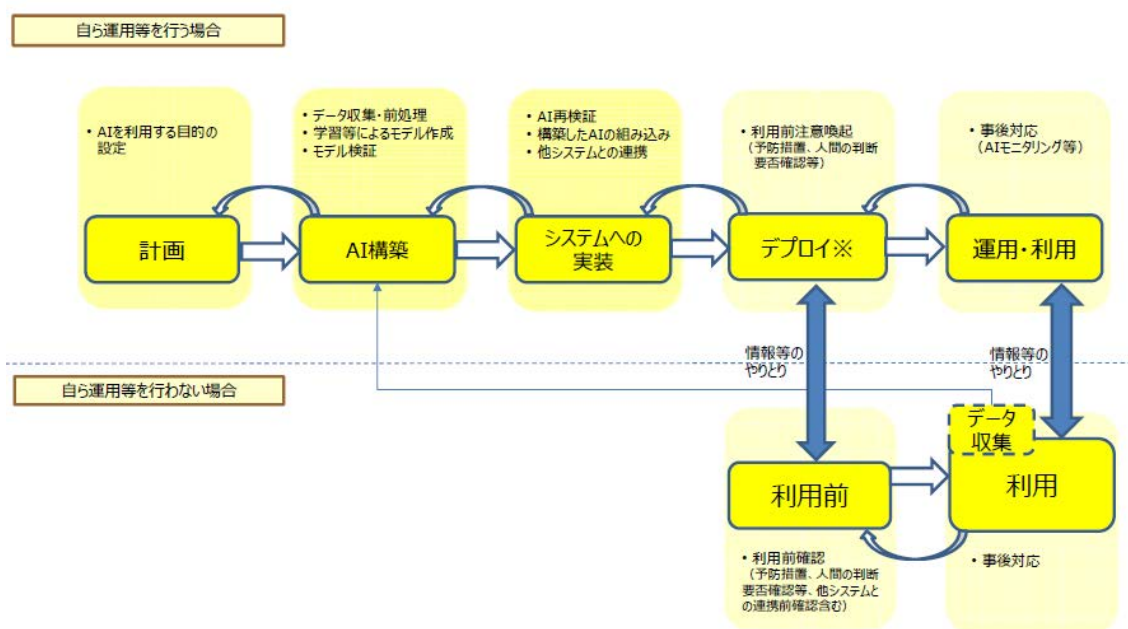
<sup>15</sup> 判断の結果についてその判断により影響を受ける者の理解を得るため、判断に関する正当な意味・理由を説明したり、(必要に応じて)賠償・補償したりする等の措置をとること。

## 5. 一般的な AI 利活用の流れ

前述の AI 利活用原則について、AI システム又は AI サービスの利活用のどのフェーズで留意すべきかを明確にするため、一般的な AI 利活用の流れを以下の 2 つに分類して整理する。

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

それぞれの場合における AI 利活用の流れは以下の通りである。



※デプロイ：（AIソフト／システムを）利用可能な状態にすること

図 3：主体ごとの一般的な AI 利活用の流れ<sup>16</sup>

<sup>16</sup> ここで図示する AI 利活用の一般的な流れは、第 7 章において、利活用のどのフェーズで AI 利活用原則のどの論点を考慮すべきかにかについての整理を行うために、典型的な事例を記載したものである。他方、AI の利活用は、本図にあるとおり各フェーズが時系列で整理される場合だけでなく、DevOps のように開発と運用が一体のものとして検討される場合など、多様なケースが想定される。そのため、本ガイドラインの活用にあたっては、それぞれの利活用の流れに応じて必要な読み替えを行うことが期待される。

(i) 自ら運用等を行う場合

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

(ii) 自ら運用等を行わない場合

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

以上、一般的な AI 利活用の流れを記載したが、例外も存在し、その場合には必要な読み替えを行うことが期待される。例えば、AI サービスとして、機械学習モデルを通じて何らかの判断を行うような単体のソフトウェアを

消費者的利用者等に提供する場合は、システムの実装を行っていないことから、(i)のフェーズ C については、他のシステムとの連携のみを検証すると読み替えることとなる。

## 6. AI 利活用原則の解説

以下では、AI 利活用原則として整理した 10 原則について、原則ごとに AI サービスプロバイダ、ビジネス利用者及びデータ提供者が留意すべき事項についての解説を行う（参考として、消費者的利用者が留意することが望ましい事項についてもあわせて記載している）。下線部に関する具体例については必要に応じ附属資料を参照されたい。

### ① 適正利用の原則

**利用者は、人間と AI システムとの間及び利用者間における適切な役割分担のもと、適正な範囲及び方法で AI システム又は AI サービスを利用するよう努める。**

#### **[ア 適正な範囲・方法での利用]**

- AI サービスプロバイダ及びビジネス利用者は、開発者等からの情報提供や説明を踏まえ、AI を利活用する際の社会的文脈に応じ、AI を利用する目的、用途と AI の性質、能力等を適切に認識した上で、AI を適正な範囲・方法で利用することが期待される。また、AI サービスプロバイダは、必要な情報を 適時に提供することが期待される。
- また、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<sup>17</sup>及び AI の点検・修理等を提供することが期待される。特に、アップデートにより、連携する他の AI に影響を及ぼすこと<sup>18</sup>が想定される場合は、そのリスクに関する情報を提供することが期待される。

---

<sup>17</sup> 問題が発見されてからアップデートが提供されるまでの間は、問題点を最終利用者に適時適切に情報提供するとともに注意喚起することが期待される。

<sup>18</sup> アップデートを適用する AI の動作が周囲の AI に影響を及ぼすことが想定される。例えば、家庭内の家電に含まれる AI ソフトがアップデートにより更新された場合、全体を統括する家庭内執事ロボットや周辺の AI を含む家電が当該アップデートに適応していないと、（家電同士、または家電とロボットの）相互の判断に齟齬が生じうる（「報告書 2018」別紙 3 「AI が想定外の動作を行うなどのおそれ」の事例）。

- また、AIの利用により人の生命・身体・財産に危害を及ぼすことが想定される場合など、提供するAIシステム又はAIサービスの性質、利用の態様等によっては、提供を受ける利用者の信頼性等について事前に確認することが期待される場合も想定される。さらに提供をした後であっても、最終利用者がAIシステム又はAIサービスを誤って使用していないか、悪意をもって使用していないかについて、入出力等のログを記録・保存すること等により確認することが期待される場面も想定される。

<以下参考>

- 消費者的利用者は、開発者及びAIサービスプロバイダからの情報提供や説明を踏まえ、AIを利活用する際の社会的文脈にも配慮して、AIを適正な範囲・方法で利用することが望ましい（留意することが望ましい事項）。

## [イ 人間の判断の介在]

- AIサービスプロバイダ及びビジネス利用者は、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。
- また、AIによりなされた判断について人間が最終判断をすることが適切とされている場合に、人間がAIと異なる判断をすることが期待できなくなることも想定されることから、説明可能性を有するAIから得られる説明を前提として、人間が判断すべき項目を事前に明確化しておくこと等により、人間の判断の実効性を確保することが期待される<sup>19</sup>。
- また、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、移行前、移行中、移行後等の各状態における責任の所在があらかじめ明確化されている必要がある。また、AIサービスプロバイダは、移行条件、移行方法等を最終利用者に事前に説明し、必要な訓練を実施するなど、人間による稼働に移行した場合に問題が生じないための事前対策を講じることが期待される。

<以下参考>

---

<sup>19</sup> 加えて、人間が確認するAIの判断の適正性を確保するため、他のAIを利用したダブルチェック、AIへの入力を摂動させることによるAI動作の確認などの措置を検討することが望ましい。



- 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。
- また、開発者及びAIサービスプロバイダにより人間の判断の実効性を確保するための対応が整理されている場合は、それに基づき適切に対応することが望ましい。
- また、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。

#### [ウ 関係者間の協力]

- 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サービスを提供し又はAIシステムを利用するに当たり、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りもたらされる又はもたらされた被害の性質・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。

<以下参考>

- 消費者的利用者は、AIを利用するに当たり、開発者又はAIサービスプロバイダからの情報提供を踏まえ、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りもたらされる又はもたらされた被害の性質・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）に取り組むことが望ましい。

## ② 適正学習の原則

**利用者及びデータ提供者は、AIシステムの学習等に用いるデータの質に留意する。**

#### [ア AIの学習等に用いるデータの質への留意]

- AIサービスプロバイダ、ビジネス利用者及びデータ提供者は、利用するAIの特性及び用途を踏まえ、AIの学習等に用いるデータの質（正確性及完全性など）に留意することが期待される（機械学習におけるデータの質を確保するための方法）。

- また、AIによりなされる判断は、事後的に精度<sup>20</sup>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<sup>21</sup>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。
- なお、消費者的利用者から提供されるデータを用いることが予定されている場合には、AIの特性及び用途を踏まえ、データ提供の手段、形式等について、あらかじめ消費者的利用者に情報を提供することが期待される。

<以下参考>

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

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

- 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。また、消費者的利用者に対し、そのようなリスクが存在することを予め周知することが期待される。

<以下参考>

- 消費者的利用者は、開発者、AIサービスプロバイダ及びデータ提供者からの情報を踏まえ、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に留意することが望ましい。
- また、AIを利用するに当たり、セキュリティ上の疑問を感じた場合は、開発者、AIサービスプロバイダ及びデータ提供者にその旨を報告することが望ましい。

<sup>20</sup> ここで言う「精度」には、AIが正しい判断を行っているか、例えばAIが暴力的な表現を行っていないか、ヘイトスピーチなどを行っていないか等も含まれる。

<sup>21</sup> 例えば、機械学習を中心としたAIは帰納的な処理を行うため、当該AI単体では、判断結果につき原理的に100%の精度を担保できないこと等が挙げられる。

### ③ 連携の原則

AI サービスプロバイダ、ビジネス利用者及びデータ提供者は、AI システム又は AI サービス相互間の連携に留意する。また、利用者は、AI システムがネットワーク化することによってリスクが惹起・増幅される可能性があることに留意する。

#### [ア 相互接続性と相互運用性への留意]

- AI 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。

#### [イ 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、AI の入出力等におけるデータ形式（構文（syntax）及び意味（semantics）<sup>22</sup>）、連携のための接続方式（特にネットワークを介す場合は各レイヤにおけるプロトコル）等の標準に準拠することが期待される。
- また、データ提供者についても、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（構文（syntax）及び意味（semantics））の標準に準拠することが期待される。

<参考>

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

#### [ウ AI ネットワーク化により惹起・増幅される課題への留意]

- 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、AI サービスプロバイダ及びビジネス利用者は、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。このため、開発者等からの情報を踏まえ、考えられるリスクを分析し、当該リスクを連携の相手方

---

<sup>22</sup> データの構文だけが示されていても、意味が示されていないと連携は正しく動作しない。

と共有するとともに、予防策や問題が生じた場合の対応策等を整理し、消費者的利用者等に対し、必要な情報提供を行うことが期待される。

<参考>

- 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、消費者的利用者は、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望ましい。また、事前の予防策や問題が生じた場合の対応策等について、開発者及びAIサービスプロバイダから情報提供があった場合には、利用にあたり留意することが望ましい。

#### ④ 安全の原則

**利用者は、AIシステム又はAIサービスの利活用により、アクチュエータ等を通じて、利用者及び第三者の生命・身体・財産に危害を及ぼすことがないよう配慮する。**

##### [ア 人の生命・身体・財産への配慮]

- 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、AIサービスプロバイダ及びビジネス利用者は、想定される被害の性質・態様等を踏まえ、開発者等からの情報をもとに、必要に応じた対応策を講ずることにより、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。
- また、AIサービスプロバイダ及びビジネス利用者は、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。加えて、当該措置について、消費者的利用者に対し、必要な情報提供を行うことが期待される。

<参考>

- 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、消費者的利用者は、想定される被害の性質・態様等を踏まえ、開発者及びAIサービスプロバイダからの情報をもとに、必要に応じてAIの点検及び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望ましい。
- また、消費者的利用者は、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、開発者及びAIサービスプロバイダから情報提供があった場合には、利用にあたり留意することが望ましい。

## ⑤ セキュリティの原則

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

### [ア セキュリティ対策の実施]

- AI サービスプロバイダ及びビジネス利用者は、AI のセキュリティに留意し、AI システムの機密性・安全性・可用性を確保するため、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。
- また、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、当該 AI の用途や特性、侵害の影響の大きさ等を踏まえ、あらかじめ整理しておくことが期待される。

<以下参考>

- 消費者的利用者は、（消費者的利用者側で）セキュリティ対策を実施することが想定されている場合には、開発者及び AI サービスプロバイダからの情報提供を踏まえ、AI のセキュリティに留意し、必要な対策を講ずることが望ましい。

### [イ セキュリティ対策のためのサービス提供等]

- AI サービスプロバイダは、自ら提供する AI 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、過去のアクシデントやインシデント情報の共有を図ることが期待される。
- また、AI サービスプロバイダ及びビジネス利用者はセキュリティが侵害された場合の措置について、消費者的利用者に対し必要な情報提供を行うことが期待される。

<以下参考>

- 消費者的利用者は、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、開発者及び AI サービスプロバイダから情報提供があった場合には、利用にあたり留意することが望ましい。
- また、AI を利用するに当たり、セキュリティ上の疑問を感じた場合は、開発者、AI サービスプロバイダ、データ提供者等にその旨を報告することが望ましい。

### [ウ AI の学習モデルに対するセキュリティ脆弱性への留意]

- 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、学習モデルの生成及びその管理において、セキュリティに脆弱性が存在する リス

クに留意することが期待される。また、消費者的利用者に対し、そのようなリスクが存在することを予め周知することが期待される。

<以下参考>

- 消費者的利用者は、開発者、AI サービスプロバイダ及びデータ提供者からの情報を踏まえ、学習モデルの生成及びその管理において、セキュリティに脆弱性が存在するリスクに留意することが望ましい。
- また、AI を利用するに当たり、セキュリティ上の疑問を感じた場合は、開発者、AI サービスプロバイダ、データ提供者等にその旨を報告することが望ましい。

## ⑥ プライバシーの原則

利用者及びデータ提供者は、AI システム又はAI サービスの利活用において、他者又は自己のプライバシーが侵害されないよう配慮する<sup>23</sup>。

### [ア AI の利活用における最終利用者及び第三者のプライバシーの尊重]

- AI サービスプロバイダ及びビジネス利用者は、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 の利活用において最終利用者及び第三者のプライバシーを尊重する。
- また、最終利用者及び第三者の 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。
- 加えて、当該措置について、最終利用者及び第三者に対し、必要な情報提供を行うことが期待される。

<以下参考>

- 消費者的利用者は、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 の利活用において第三者のプライバシーを尊重する。
- 加えて、第三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、開発者及びAI サービスプロバイダから情報提供があった場合には、利用にあたり留意することが望ましい。

### [イ パーソナルデータの収集・前処理・提供等におけるプライバシーの尊重]

- 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パーソナルデータの収集・前処理・提供等<sup>24 25</sup>におい

<sup>23</sup> 日本においては、前提として、個人情報保護法を遵守することが必要である。

<sup>24</sup> 他者に提供後のパーソナルデータの取り扱いについても例えば消去を行う等の留意が期待される。

<sup>25</sup> 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、自ら提供したデータに個人情報が含まれる場合、当該データを誰がどのように利用しているのかを把握しておくことが求められる。

て、また、それらを通じて生成された学習モデルの提供等において、最終利用者及び第三者のプライバシーを尊重する。

<以下参考>

- 消費者的利用者は、自らデータを収集し、利用する AI の学習等を行うことが予定されている場合には、収集等において第三者のプライバシーを尊重する。

#### [ウ 自己等のプライバシー侵害への留意及びパーソナルデータ流出の防止]

- 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 の判断により本人同意なくパーソナルデータが第三者に提供されることがないよう、同意が得られていないデータはシステム上第三者に提供できないこととするなど、適切な措置を講ずることが期待される。

<以下参考>

- 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(自己の情報のみならず他者の情報を含む。)をむやみに AI に与えることのないよう留意することが望ましい。

### ⑦ 尊厳・自律の原則

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

#### [ア 人間の尊厳と個人の自律の尊重]

- AI サービスプロバイダ及びビジネス利用者は、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<sup>26</sup>。

<以下参考>

- 消費者的利用者は、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望ましい。

#### [イ AI による意思決定・感情の操作等への留意]

- AI サービスプロバイダ及びビジネス利用者は、消費者的利用者には AI により意思決定や感情が操作される可能性<sup>27</sup>や、AI に過度に依存するリスクが存在することを踏まえ、必要な 対策 を講じることが期待される。

---

<sup>26</sup> 例えば、人間と AI の異質性を前提に、AI は人間の活動を支援するものであるとの認識を持つこと等が考えられる。なお、人間と AI の異質性とは、人間と AI が異なる性質を持つことを言い、この前提が成り立つことにより、AI を人間と同様に扱うべきではないと考える(すなわち、人間の尊厳と個人の自律を尊重する)ことが可能となる。

<sup>27</sup> AI によりナッジ(合理的選択のための支援)を行う場合など、AI による消費者的利用

<以下参考>

- 消費者的利用者は、開発者及び AI サービスプロバイダからの情報等<sup>28</sup>を踏まえ、AIにより意思決定や感情が操作される可能性や、AIに過度に依存するリスクがあることを認識することが望ましい。

#### [ウ AI と人間の脳・身体を連携する際の生命倫理等の議論の参照]

- AI サービスプロバイダ及びビジネス利用者は、AIを人間の脳・身体と連携させる場合、特に、エンハンスメント（健康の維持や回復を超えた人間の能力の増進の追求）を行う場合には、その周辺技術に関する開発者等からの情報を踏まえつつ、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に配慮することが期待される。
- また、提供する AI の機能及びその周辺技術に関する情報を消費者的利用者に提供することが期待される。

<以下参考>

- 消費者的利用者は、AIを人間の脳・身体と連携させた AI を用いる場合には、AIの機能及びその周辺技術に関する開発者及び AI サービスプロバイダからの情報を踏まえ、自律性に影響を及ぼす可能性が生じうることに留意して、利用することが望ましい。

#### [エ AI を利用したプロファイリングを行う場合における不利益への配慮]

- AI サービスプロバイダ及びビジネス利用者は、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、対象者に生じうる 不利益等 に慎重に配慮<sup>29,30</sup>する。

<以下参考>

- 消費者的利用者は、AIによるプロファイリングが行われている可能性があることを踏まえ、自らの情報が正しく利用されているかを意識し、必要に応じ、AI サービスプロバイダ及びビジネス利用者に確認することが望ましい。

## ⑧ 公平性の原則

者の意思決定・感情の操作はすべてがリスクにつながるとは限らないことから、ここでは「可能性」という語を用いている。なお、AIでナッジを行う際には、AI開発ガイドライン案における「⑧利用者支援の原則」（利用者に選択の機会を適切に提供する）を参照することが期待される。

<sup>28</sup> 開発者及び AI サービスプロバイダから直接入手する情報のみならず、教育現場等において得られる情報等を含む。

<sup>29</sup> GDPR においては、同 22 条において、人間が介在せずに最終的な決定を自動処理のみにより行われたい権利が保障されている。

<sup>30</sup> 「①適正利用の原則」の「論点イ 人間の判断の介在」参照。



AI サービスプロバイダ、ビジネス利用者及びデータ提供者は、AI システム又はAI サービスの判断にバイアスが含まれる可能性があることに留意し、また、AI システム又はAI サービスの判断によって個人が不当に差別されないよう配慮する<sup>31</sup>。

#### [ア AI の学習等に用いられるデータの代表性への留意]

- 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 の判断が学習時のデータによって決定づけられる可能性があることを踏まえ、AI を利活用する際の社会的文脈に応じて、AI の学習等に用いられるデータの代表性<sup>32</sup>やデータに内在する社会的なバイアス等に留意することが期待される。

<以下参考>

- 消費者的利用者は、AI の判断結果について疑義を感じた場合には、必要に応じて、開発者、AI サービスプロバイダ、ビジネス利用者等に問い合わせを行うことが望ましい。

#### [イ アルゴリズムによるバイアスへの留意]

- AI サービスプロバイダ及びビジネス利用者は、AI に用いられるアルゴリズムにより、AI の判断にバイアスが生じる可能性があることに留意することが期待される。特に、機械学習においては、一般的に、多数派がより尊重され、少数派が反映されにくい傾向にあり（バンドワゴン効果）、この課題を回避するための方法が検討されている。

<以下参考>

- 消費者的利用者は、AI の判断結果について疑義を感じた場合には、必要に応じて、開発者、AI サービスプロバイダ、ビジネス利用者等に問い合わせを行うことが望ましい。

#### [ウ 人間の判断の介入（公平性の確保）]

- AI サービスプロバイダ及びビジネス利用者は、AI によりなされた判断結果の公平性<sup>33</sup>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。

<sup>31</sup> 「公平性」には複数の定義・基準があることに留意する必要がある。

<sup>32</sup> データの「代表性」とは、サンプルとして抽出され利活用に使われているデータが、その母集団の性質を歪めていない状態のことをいう。

<sup>33</sup> AI の学習に用いるデータに内在する社会的バイアス等が、AI による判断結果の公平性に影響を及ぼしうることを前提としている。

- 人間の判断の介入の要否については、[①ーイ]に掲げる内容を参照しつつ、公平性の観点からの基準例も踏まえ、検討することが期待される。

## ⑨ 透明性の原則

AI サービスプロバイダ及びビジネス利用者は、AI システム又は AI サービスの入出力等の検証可能性及び判断結果の説明可能性に留意する<sup>34</sup>。

### [ア AI の入出力等のログの記録・保存]

- AI サービスプロバイダ及びビジネス利用者は、AI の入出力等の検証可能性を確保するため<sup>35</sup>、入出力等のログを記録・保存することが期待される。ログの記録・保存に当たっては、利用する技術の特性及び用途に照らして、ログの記録・保存の目的、ログの取得・記録の頻度等について考慮することが期待される。

### [イ 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、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 AI を利用する場合など、AI を利活用する際の社会的文脈を踏まえ、利用者の納得感や安心感の獲得、また、そのための AI の動作に対する証拠の提示等を目的として、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。その際、AI を利活用する際の社会的文脈を踏まえ、どのような説明が求められるかを分析・把握し、必要な対応を講じることにより、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。

### [ウ 行政機関が利用する際の透明性の確保]

- 行政機関が AI を利用する場合には、法の支配、行政の透明性確保、適正手続等の要請を踏まえ、AI を利活用する際の社会的文脈に応じ、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（説明可能性を向上させるための方法の例）。

## ⑩ アカウンタビリティの原則

利用者は、ステークホルダに対しアカウンタビリティを果たすよう努める。

### [ア アカウンタビリティを果たす努力]

<sup>34</sup> 本原則は、アルゴリズム、ソースコード、学習データの開示を想定するものではない。また、本原則の解釈に当たっては、プライバシーや営業秘密への配慮も求められる。

<sup>35</sup> 入出力等の検証可能性を確保することが期待される場面としては、事故発生時において原因解明を行う場合に加え、最終利用者が AI システム又は AI サービスを誤って使用していないか、悪意をもって使用していないかについて確認を行う場合等も想定される。

- 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、本ガイドラインの掲げる利活用原則①～⑨の趣旨に鑑み、消費者的利用者、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、それぞれが有する知識や能力の多寡に応じ、AI システムの特性について情報提供と説明を行うことや、多様なステークホルダとの対話を行うこと等により、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。

<以下参考>

- 消費者的利用者は、それぞれが有する知識や能力の多寡に応じ、相応のアカウントビリティを果たすよう努めることが望ましい。
- また、AI の判断結果について疑義を感じた場合には、必要に応じて、開発者、AI サービスプロバイダ、ビジネス利用者等に問い合わせを行うことが望ましい。

#### [イ 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、消費者的利用者等が AI の利活用について適切に認識できるよう、以下のとおり、AI に関する利用方針を作成・公表し、通知を行うことが期待される。
  - (i) AI の判断が直接に消費者的利用者や第三者に対して影響を及ぼす態様により AI を利活用する場合には、消費者的利用者や第三者が AI の利活用について適切に認識することができるよう、AI に関する利用方針を作成・公表し、問い合わせがあった場合には通知を行う。
  - (ii) (i)について、消費者的利用者や第三者の権利・利益に重大な影響を及ぼす可能性のある場合には積極的に通知を行う<sup>36</sup>。
- また、通知又は公表は、利用開始前だけではなく、AI の動作に変更が生じたときや利用終了時も含め（特に AI の動作変更に伴い想定されるリスクに変更が生じる場合など）実施することが期待される。

<以下参考>

- 消費者的利用者は、AI の判断結果について疑義を感じた場合には、必要に応じて、開発者、AI サービスプロバイダ、ビジネス利用者等に問い合わせを行うことが望ましい。

---

<sup>36</sup> AI サービスプロバイダ及びビジネス利用者が、AI に関する利用方針を公表することが求められるのは、利活用する AI の判断が、消費者的利用者や第三者に直接の影響を及ぼす場合であると考えられる。すなわち、人間の思考に供するための分析道具として AI を利活用するにとどまる場合や、AI が原案を作成しつつも、最終的に人間が判断することが実質的に担保されている場合には、AI に関する利用方針の公表が必ずしも求められるわけではない。（もっとも、そうした場合であっても、自主的に公表されることが望ましい。）

## 7. AI 利活用原則を考慮すべきタイミング

前章に示した各原則の各論点が、利活用の流れに応じ、どのフェーズで考慮されるべきかにつき、表1及び表2のとおり整理する<sup>37</sup>。

なお、ここでは、AI サービスプロバイダ、ビジネス利用者については、自らAIの運用等を行う場合を想定し、また、消費者的利用者については、自ら運用等を行わない場合を想定した上で、利活用のフェーズとの関係を整理している<sup>38</sup>。このため、自ら運用等を行わないビジネス利用者や、自ら運用等を行う消費者的利用者については、必要な読み替えを行うことが期待される。

---

<sup>37</sup> 表1、表2は、それぞれの論点を考慮すべきフェーズを記載したものであるが、そもそも、これらの論点を考慮すべきか否か、考慮するとしてどの程度考慮すべきかについては、AIの利用目的や利活用する際の社会的文脈に応じて検討されるべきものであることに留意が必要である。

<sup>38</sup> データ提供者については、自ら運用等を行うAIサービスプロバイダ及びビジネス利用者の利活用フェーズをもとに整理を行っている。

表 1 : AI サービスプロバイダ、ビジネス利用者等の利活用の流れと各原則・各論点との関係

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

表 2：消費者的利用者の利活用の流れと各原則・各論点の関係

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