別紙１

AI利活用ガイドライン案

令和元年●月●日

AIネットワーク社会推進会議

目次

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いる。こうした中、平成 28 年（2016 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は AI 開発原則のたたき台を紹介し、各国関係閣僚による議論が行われた。その結果、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 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。本推進会議では、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、平成 29 年（2017 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。

本 AI 利活用ガイドライン案は、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事項を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。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。AI の利用者が、AI を利活用するにあたって本ガイドラインを参照し、それぞれの立場に応じた必要な措置を講じることにより、関連するステークホルダの利益を保護するとともにリスクの波及を抑止し、これにより AI の利活用や社会実装が進展することが期待される。

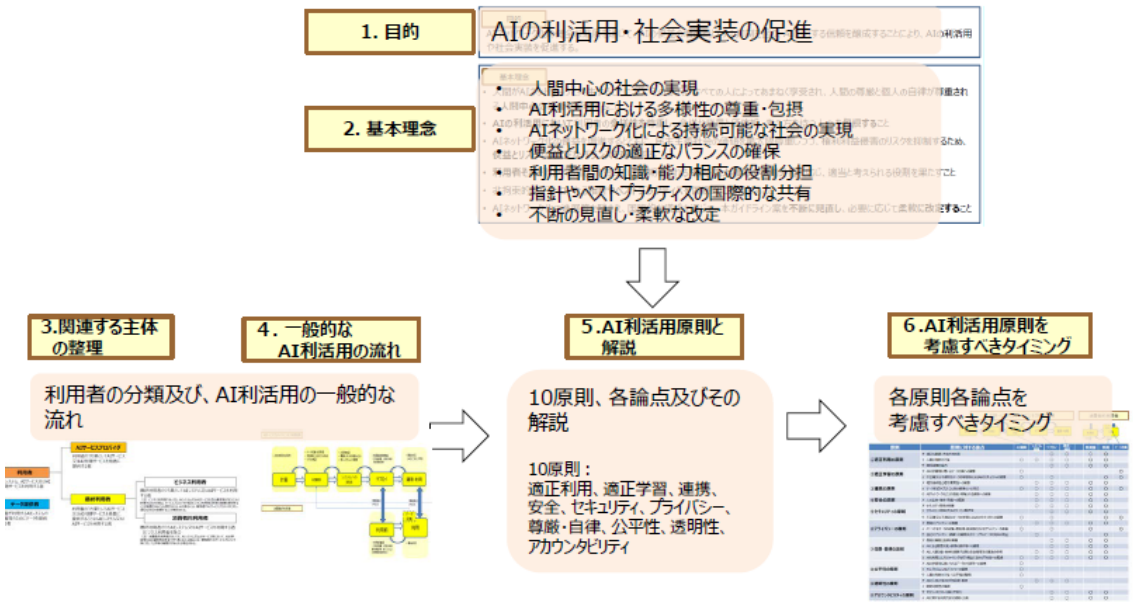

図 1：AI 利活用ガイドライン案の構成

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[1]が進展していく過程で、個人、地域社会、各国、国際社会[2]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、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[3]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を促進することを目的とする。

なお、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、すでに「AI 開発ガイドライン案」が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るものではなく、両ガイドラインをセットとして参照されることが望ましい。

---

[1] 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。以下同じ。

[2] 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）などを参照。

[3] 「リスク」とは「損害をもたらす可能性があるもの」を意味する。以下同様。

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本ガイドラインの基本理念とする。

- 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 ネットワーク化の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、権利利益侵害の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
- **利用者それぞれが** AI に関して有していると期待される**知識・能力相応の役割分担**に応じ、適当と考えられる役割を果たすこと
- 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、本ガイドライン案を**不断に見直し**、必要に応じて**柔軟に改定する**こと

3. 関連する主体の整理

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。

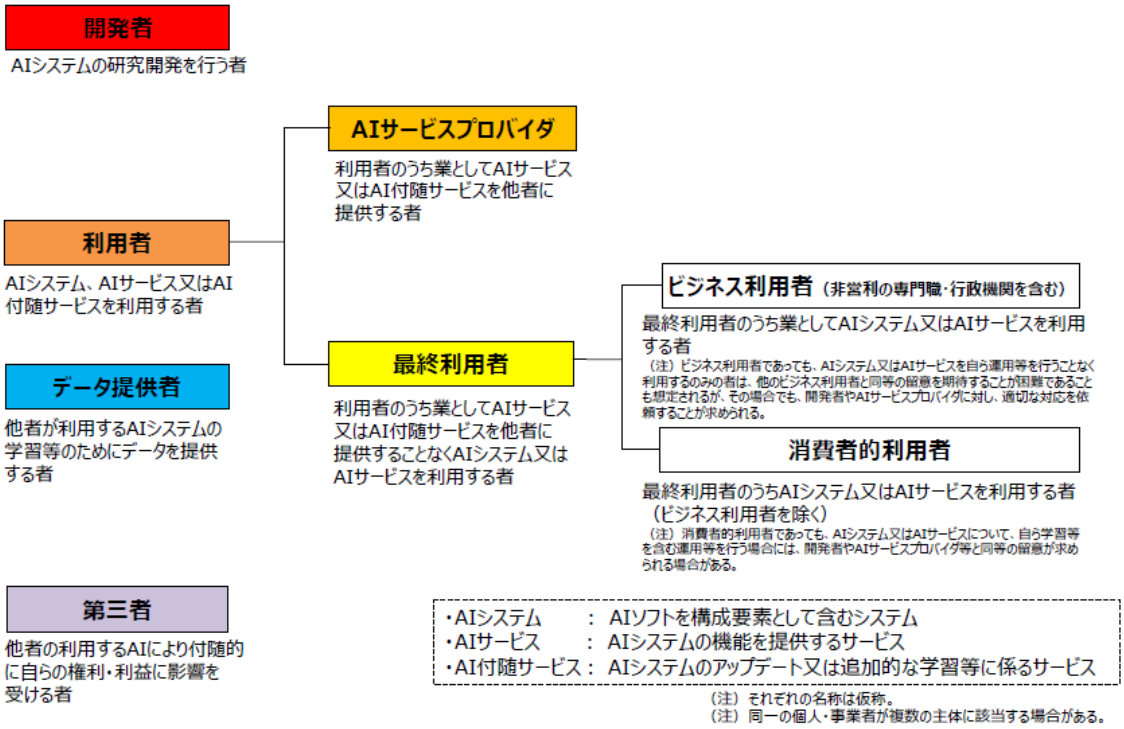

図 2：　関連する主体の整理

- 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：
  AI 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）：
  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。
  ただし、ビジネス利用者であっても、AI システム又は AI サービスを自ら運用等を行うことなく利用するのみの者は、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、適切な対応を依頼することが求められる。

- 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、自ら学習等を含む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 により付随的に自らの権利・利益に影響を受ける者

4. 一般的な AI 利活用の流れ

後述の AI 利活用原則について、AI 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 システムや AI サービスの運用等（学習等を含む）を行う場合

(ii) 自ら AI システムや 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 利活用の際の流れは下図の通りである。

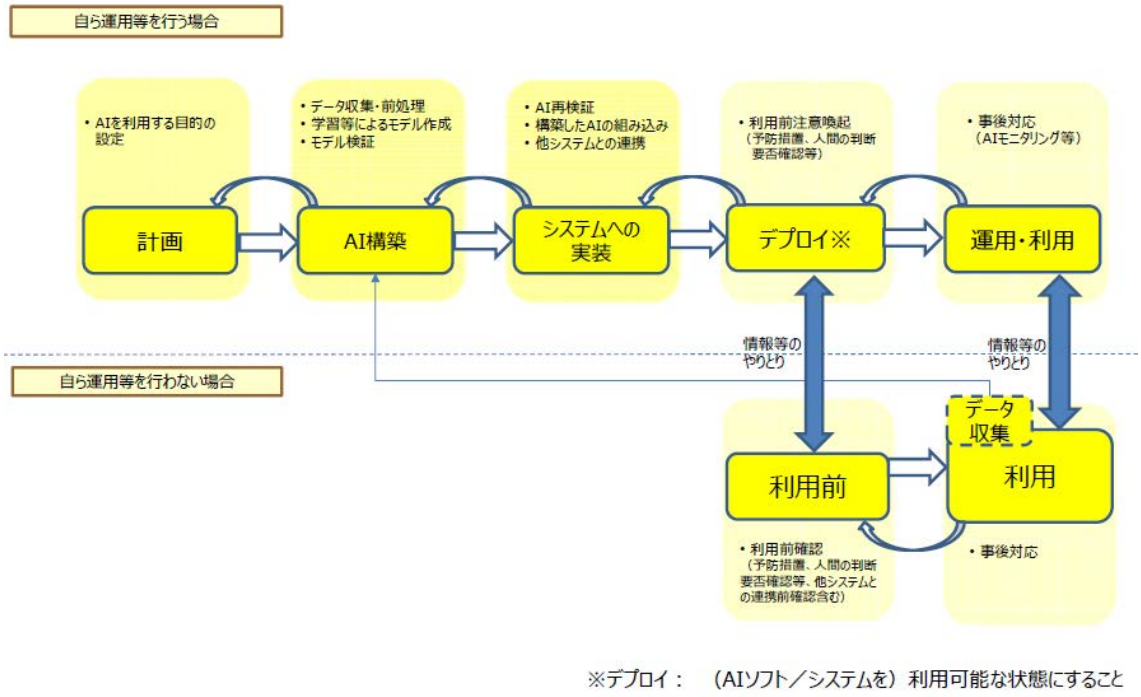

図 3：主体ごとの一般的な AI 利活用の流れ

(i)、(ii)の各場合における AI 利活用の一般的な流れとそれぞれのフェーズは以下の通りである。

(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（AI）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 ソフト／システム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 ソフト／システム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 ソフト／システム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 ソフト／システムを利用する前のフェーズである。利用に当たり、AI サービスプロバイダ・ビジネス利用者等から与えら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 ソフト／システムを利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 ソフト／システム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。

以上、一般的な AI 利活用の流れを記載したが、例外も存在するため、相応の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。

5. AI 利活用原則とその解説

以下では、AI 利活用原則として 10 原則を掲げるとともに、 (i)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（以下「AI サービスプロバイダ、ビジネス利用者等」という。）と、(ii)消費者的利用者のそれぞれの立場から、各原則についての解説を行う。なお、下線部 に対する具体例につい

ては附属資料を必要に応じ参照されたい。

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、ビジネス利用者等）

- AI サービスプロバイダ、ビジネス利用者は、消費者的利用者に AI サービス等を提供し、運用等を行うにあたり、開発者からの情報を踏まえ、関連する情報を適時適切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、アップデートのための機能を提供する際に、他の AI との関係でリスクが想定される場合は、当該リスク情報を提示した上で提供することが望ましい。
- また、提供される AI の性質、利用の態様等によっては、提供対象となる利用者が当該 AI を提供するにふさわしい者であるか（信頼性）について事前に考慮することが期待される場合も想定される。

（消費者的利用者）

- 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。開発者、AI サービスプロバイダ、ビジネス利用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。（実施すべき内容）

［イ　人間の判断の介在］

（AI サービスプロバイダ、ビジネス利用者等）

- AI サービスプロバイダ、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（判断基準）、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（の基準）については、利用分野、用途等に応じて検討されることが期待される。
- また、AI の判断に対し、人間が最終判断をすることが適当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。こうした場合に人間が行うべき判断について その項目、手段

など を明確化することにより、人間の判断の実効性を確保することが考えられる。

- 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、移行前、移行中、移行後等の各状態に伴い、予め責任の所在が明確になっている必要がある。また、前述の移行条件、移行方法等を利用者に事前に告知し、必要な訓練などを前もって実施するなど、人間による稼働に移行した場合に問題が起こらないよう、注意喚起をしておくことが期待される。

（消費者的利用者）

- AI の判断に対し、人間が最終判断をすることが適当とされている場合に、適切に判断ができるよう能力を習得しておくことが期待される。

［ウ　関係者間の協力］

（AI サービスプロバイダ、ビジネス利用者等）

- 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 を提供または利用するに当たり、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。

（消費者的利用者）

- AI 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。

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、ビジネス利用者等）

- 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、提供もしくは利用する 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。（機械学習の各タイミングにおけるデータの質の担保方法 ）
- また、AI によりなされた判断の精度が損われたり、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、実装

コスト、及び、技術水準等を踏まえ、精度に関する基準を予め定めておくことが期待される。他方、精度が当該基準を下回った場合には、データセットの質に留意して改めて学習させることが期待される。

- 加えて、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、提供もしくは利用する 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、消費者的利用者に情報を提供することが期待される。

（消費者的利用者）

- 利用する AI 等の学習に用いるデータを自ら収集することが予定されている場合には、（データの提供に関する）手段、形式等について、開発者、AI サービスプロバイダ、ビジネス利用者等からの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、ビジネス利用者等）

- 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し、予め周知することが期待される。

（消費者的利用者）

- 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等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が存在することを認識しておくことが期待される。

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、ビジネス利用者等）

- 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、ビジネス利用者等）

- 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、以下 AI の入出力等におけるデータ形式（構文（syntax）及び意味（semantics））や、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。

（消費者的利用者）

- 消費者的利用者は、利用する AI 等の学習に用いるデータを自ら収集する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、ビジネス利用者等から提供された情報を踏まえた上で収集、保存を行うことが期待される。

［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 サービスプロバイダ、ビジネス利用者等）

- 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、他者に提供し、または、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した上で、提供・利用する AI の設計を行うことが期待される。また、開発者等からの情報を基に考えられるリスクを分析し、当該情報を連携の相手方と共有するとともに、問題が生じた場合の対応策等を作成の上、消費者的利用者に情報提供することが期待される。

（消費者的利用者）

- 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期待される。また、問題が生じた場合の対応策等について、開発者、AI サービスプロバイダ及びビジネス利用者等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期待される。

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。

［ア　人の生命・身体・財産への配慮］

（AI サービスプロバイダ、ビジネス利用者等）

- 人の生命・身体・財産に危害を及ぼし得る分野で AI サービスプロバイダ及びビジネス利用者は AI を利活用する場合には、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基に、必要に応じて対応策を講じることにより、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。

（消費者的利用者）

- 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、想定される被害の性質・態様等を踏まえ、開発者、AI サービスプロバイダ及びビジネス利用者からの情報を基に、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。
- また、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、AI サービスプロバイダ及びビジネス利用者から提供された情報に留意することが期待される。

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。

［ア　セキュリティ対策の実施］

（AI サービスプロバイダ、ビジネス利用者等）

- AI サービスプロバイダ及びビジネス利用者は、AI は fragile であるという認識を持つと同時に 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。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、あらかじめ整理しておくことが期待される。また、その措置の内容について、開発者等から

の情報を踏まえ、当該 AI の用途や特性、侵害の影響の大きさ等社会的文脈に応じたものとすることが期待される。

（消費者的利用者）

- （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、AI のセキュリティに留意し、必要な対策を講ずることが期待される。

［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、ビジネス利用者等）

- 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。

（消費者的利用者）

- 人の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、AI サービスプロバイダ及びビジネス利用者から提供された情報に留意することが期待される。

［ウ　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］

（AI サービスプロバイダ、ビジネス利用者等）

- 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し、予め周知することが期待される。

（消費者的利用者）

- 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等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が存在することを認識しておくことが期待される。

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。

［ア　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、ビジネス利用者等）

- 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、最終利用者に対し、必要な情報提供を行うことが期待される。

（消費者的利用者）

- 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、AI サービスプロバイダ及びビジネス利用者等から提供された情報に留意することが期待される。

［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、ビジネス利用者等）

- 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・提供等において、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、第三者のプライバシーを尊重する。
- また、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。

（消費者的利用者）

- 消費者的利用者は、利用する AI 等の学習に用いるデータを自ら収集する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、ビジネス利用者等）

- 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、同意がないデータはシステム上提供できないこととするなど適切な措置を講ずる。

（消費者的利用者）

- ペットロボットなど 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期待される。

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。

[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、ビジネス利用者等）

- 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。その際には、人間と AI の異質性を前提とするとともに、人間の活動を支援するものであるとの認識が重要となる。

（消費者的利用者）

- 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。その際には、人間と AI の異質性を前提とするとともに、人間の活動を支援するものであるとの認識が重要となる。

[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、ビジネス利用者）

- 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や、AI に過度に依存するリスクが存在することを踏まえ、対策を講じる必要がある。

（消費者的利用者）

- AI サービスプロバイダ及びビジネス利用者からの情報等を踏まえ、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自覚することが期待される。

[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、ビジネス利用者）

- 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及）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な配慮が求められる。
- また、上記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。

（消費者的利用者）

- AI を人間の脳・身体と連携させた AI を用いる場合には、当該機能及びその周辺技術に関する開発者、AI サービスプロバイダ及びビジネス利用

者からの情報を基に、自らの自律性が侵害されないよう留意して利用することが求められる。

[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、ビジネス利用者）

- 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する。

（消費者的利用者）

- 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、確認することが期待される。

⑧ 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。

（注）「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。

[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、ビジネス利用者等）

- 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、利用時の社会的文脈及び用途に照らして、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。（考慮すべき点）

[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、ビジネス利用者等）

- 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、それによる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（バンドワゴン効果）ため、この課題を回避する方法が検討されている。）
- なお、公平性基準を定義し、それをアルゴリズムに組み込むことにより、公平性を満足する判断は可能となるため、公平性基準を選択・決定

するための意思決定に関する制度設計（メカニズムデザイン）が重要となることに留意が必要である。メカニズムデザインにあたっては、どのようなメカニズムとすべきか（特定の者が不利益を被らないルールなど）、どのようにメカニズムを確保するのか（契約、不文律、世論など）等について検討を行う。

- **［ウ　人間の判断の介在（公平性の確保）］**

（AI サービスプロバイダ、ビジネス利用者等）

- AI サービスプロバイダ及びビジネス利用者は、人の 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、人間の判断を介在させることが期待される。
- 人間の判断の介在の要否については、[①ーイ] に掲げる内容を基に、利用する技術の特性及び用途に照らして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。

［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、ビジネス利用者等）

- AI サービスプロバイダ及びビジネス利用者は、AI の提供にあたり、問題が生じた場合の原因究明（トレース）等を目的として、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、考慮すべき事項）

［イ　説明可能性の確保］

（AI サービスプロバイダ、ビジネス利用者等）

- 利用者の納得感や安心感の獲得、及びそのための 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（特に、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提供・利用する場合）。
- ただし、説明可能性の確保に当たっては、その目的に鑑み、どのような説明が必要かを分析・把握し、対処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］

（AI サービスプロバイダ、ビジネス利用者等）

- 行政機関が AI を利用する場合には、「法の支配」の原理に基づく行政の透明性確保の要請と、行政手続法に基づく適正手続の要請を踏まえ、行政運営における公正の確保と透明性の向上の要請を踏まえ、AI の判断結

果の説明可能性を確保するため、必要に応じ、措置を講じることを検討すべきである。

⑩ 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、消費者的利用者等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。

[ア　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、ビジネス利用者等）

- 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、消費者的利用者等 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他、多様なステークホルダとの対話を通じて様々な意見を聴取する等、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。

（消費者的利用者）

- AI の判断結果について疑義を感じた場合は、必要に応じて、同サービスを提供した開発者、AI サービスプロバイダ、及びビジネス利用者に問い合わせを行うことが期待される。

[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、ビジネス利用者等）

- AI サービスプロバイダ及びビジネス利用者は、自ら AI を利活用する場合には、以下の場合に、消費者的利用者や第三者がその 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の上公表し、問い合わせがあった場合には通知を行うこと、加えて、個人の権利・利益に重大な影響を及ぼす可能性のある場合には自発的に通知することが期待される。
- また、通知または公表は、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実施すること（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）が期待される。

（消費者的利用者）

- AI の判断結果について疑義を感じた場合は、必要に応じて、同サービスを提供した開発者、AI サービスプロバイダ、及びビジネス利用者に問い合わせを行うことが期待される。

6. AI 利活用原則を考慮すべきタイミング

５．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。

なお、ここでは、AI サービスプロバイダ、ビジネス利用者等については、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。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、読み替えを行う必要がある。

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
① 適正利用の原則				
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			○	○
② 適正学習の原則				
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	○		○	
③ 連携の原則				
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		○	○	○
④ 安全の原則				
ア 人の生命・身体・財産への配慮		○	○	○
⑤ セキュリティの原則				
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 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	○		○	
⑥ プライバシーの原則				
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		○		
⑦ 尊厳・自律の原則				
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	○	○	○	○
⑧ 公平性の原則				
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）	○			
⑨ 透明性の原則				
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	○	○	○	○
⑩ アカウンタビリティの原則				
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
① 適正利用の原則			
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	○	○	
② 適正学習の原則			
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		○	○
③ 連携の原則			
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	○	○	
④ 安全の原則			
ア 人の生命・身体・財産への配慮	○	○	
⑤ セキュリティの原則			
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 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		○	○
⑥ プライバシーの原則			
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		○	
⑦ 尊厳・自律の原則			
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	○	○	
⑧ 公平性の原則			
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）			
⑨ 透明性の原則			
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			
⑩ アカウンタビリティの原則			
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	○	○	

7. AIの定義及び対象範囲

７－１　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[4]。

- 「**AIソフト**」とは、データ・情報・知識の学習等[5]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。

７－２　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。

---

[4] 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。

[5] 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。