Tentative Translation

Report 2019 Appendix 1

AI Utilization Guidelines

Practical Reference for AI utilization

9 August 2019

The Conference toward AI Network Society

Table of Contents

(Attachment)

Preface

Research and development (R&D) and utilization of artificial intelligence (AI) are expected to progress dramatically in the years to come, and discussions are ongoing internationally to foster trust in AI. In Japan, from the viewpoint of promoting the transformation to an "AI-ready society," the government started studying common AI social principles in May 2018 and formulated the "Social Principles of Human-Centric AI"[1] in March 2019.

At the G7 ICT Ministers' Meeting in April 2016, Japan as the host nation introduced the principles of AI development, which triggered a study on AI-related principles. At the G7 ICT Ministers' Meeting, discussions were held by relevant ministers, and the G7 countries agreed to continue having a central role and to discuss formulating the "AI R&D Guidelines", which consist of "AI R&D principles" and explanations for them, with the cooperation of international organizations, such as the Organization for Economic Co-operation and Development (OECD). For the promotion of the benefits of AI with risk mitigation, the Conference toward AI Network Society (hereinafter referred to as the "Conference") studied items that should be taken into consideration at the time of research and development. In July 2017, the Conference released a document entitled "AI R&D Guidelines for International Discussions" (hereinafter referred to as the "Draft AI R&D Guidelines") to serve as the basis of international discussions in the G7 and the OECD.

On the other hand, AI may change its implementation and output continuously by learning from data in the process of its utilization. Therefore, it is assumed that there are not only matters that developers are expected to take into consideration, but also matters that users[2] are expected to take into consideration when using AI. Furthermore, it is considered important to itemize matters to be taken into consideration in the utilization of AI from the viewpoint of studying roles expected from various stakeholders, including developers, users, and data providers. For this reason, the Conference released the "Draft AI Utilization Principles" in July 2018 as rules to which AI service providers, end

[1] "Human-centered AI Social Principles" (decision by the Council for Integrated Innovation Strategy In Japan in March 2019)

[2] Refer chapter 3 regarding classification of related entities such as "developers" and "users", etc.

users, and data providers are expected to pay attention.

Japan has been actively disseminating the study results to the international community. In particular, at the meeting of the AI expert group at the OECD (AIGO), which was established by the OECD in September 2018 to draft the recommendations of the OECD Board, Japan made a significant contribution that included the introduction of Japan's "Social Principles of Human-Centric AI", "Draft AI R&D Guidelines", and "Draft AI Utilization Principles" . Japanese experts introduced not only the contents of these principles and guidelines, but also the background of the study and situation of discussion that took place on them. For this reason, the Recommendation on Artificial Intelligence of the OECD Board, published in May 2019, is consistent with the above principles and guidelines that have been discussed in Japan.

Furthermore, the "G20 AI Principles", which adopted the recommendations of the OECD Board, were compiled at the G20 Trade and Digital Economy Ministers Meeting (June 2019), at which Japan served as the host country.

As described above, Japan has led international discussions on the principles of AI. As a result, international consensus has been built on the concept of the principles. Meanwhile, progress continues on studying specific measures that are expected to be needed in the future, such as measures to realize the principles and what steps each stakeholder should take for that purpose. The "Social Principles of Human-Centric AI" expect that developers and business operators establish and comply with their own AI development and utilization principles based on these basic ideas and the AI social principles outlined in them. For that reason, a specific commentary is required as a reference

The guidelines, entitled "AI Utilization Guidelines" (hereinafter referred to as the "Guidelines"), consist of the AI Utilization Principles and the commentary on them. The AI Utilization Principles have been arranged based on a draft of what is expected to be taken into consideration for the promotion of the benefits of AI with risk mitigation[3,4] proposed by the Conference in 2018. In explaining each

[3] The Guidelines explain specific measures based on the AI Utilization Principles. In the case of realizing each principle in the OECD Council Recommendation on Artificial Intelligence, the AI Utilization Guidelines also make it possible to grasp concrete measures by referring to the corresponding description in the AI Utilization Guidelines.

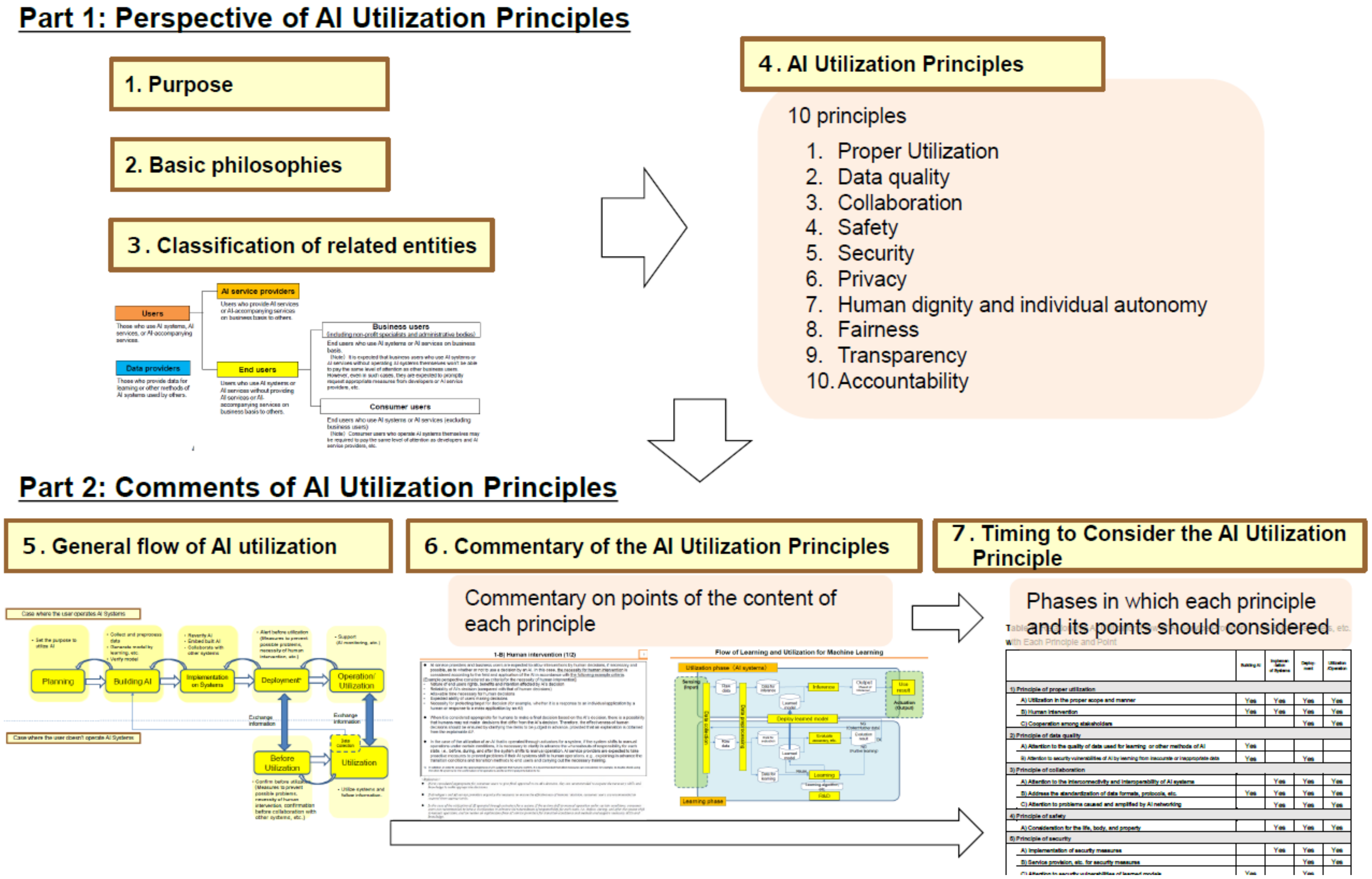
[4] The Guidelines consist of the AI Utilization Principles compiled as a soft law and its

principle, the Guidelines attempt to give specific descriptions for measures to be taken to realize each principle. Since the Guidelines are formulated with the participation of multiple stakeholders, it can be used as a common reference by stakeholders at all levels for AI utilization.

The Guidelines are intended to encourage AI users (especially AI service providers who provide AI services, etc. to others and business users who use AI systems, etc. on a business basis) to recognize the proper consideration needed in relation to AI utilization and to take action voluntarily. This can be done by referring to the Guidelines when they establish their own AI development and utilization principles based on the "Social Principles of Human-Centric AI". Furthermore, it may be possible for AI service providers and business users to add value to their AI services and business utilizing AI by undertaking such voluntary efforts.

If AI users refer to the Guidelines and voluntarily take proper measures, according to the purpose of using AI and the social context in using AI, it is expected that the related stakeholders' interests will be protected and the risk diffusion will be curbed. Furthermore, voluntary efforts by AI service providers and business users might enable them to earn the trust of consumers and third parties, which will lead to an increasing demand for the provision of AI systems and AI services. This will lend itself to the promotion of AI's social implementation.

explanations concerning the viewpoint of realizing them. Unlike the terms and conditions of laws and regulations, these Guidelines do not have the nature of applying described matters literally, but a compilation of matters that AI users are expected to take voluntarily when using AI.



Part 1: Perspective of AI Utilization Principles

1. Purpose

2. Basic philosophies

3 . Classification of related entities

Users
Those who use AI systems, AI services, or AI-accompanying services.
Data providers
Those who provide data for learning or other methods of AI systems used by others.
AI service providers
Users who provide AI services or AI-accompanying services on business basis to others.
End users
Users who use AI systems or AI services without providing AI services or AI-accompanying services on business basis to others.
Business users
(including non-profit specialists and administrative bodies)
End users who use AI systems or AI services on business basis.
(Note) It is expected that business users who use AI systems or AI services without operating AI systems themselves won't be able to pay the same level of attention as other business users. However, even in such cases, they are expected to promptly request appropriate measures from developers or AI service providers, etc.
Consumer users
End users who use AI systems or AI services (excluding business users)
(Note) Consumer users who operate AI systems themselves may be required to pay the same level of attention as developers and AI service providers, etc.

4 . AI Utilization Principles

10 principles

1. Proper Utilization
2. Data quality
3. Collaboration
4. Safety
5. Security
6. Privacy
7. Human dignity and individual autonomy
8. Fairness
9. Transparency
10. Accountability

Part 2: Comments of AI Utilization Principles

5 . General flow of AI utilization

Case where the user operates AI Systems
• Set the purpose to utilize AI
Planning
• Collect and preprocess data
• Generate model by learning, etc.
• Verify model
Building AI
• Reverify AI
• Embed built AI
• Collaborate with other systems
Implementation on Systems
• Alert before utilization (Measures to prevent possible problems, necessity of human intervention, etc.)
Deployment*
• Support (AI monitoring, etc.)
Operation/ Utilization
Exchange information
Exchange information
Case where the user doesn't operate AI Systems
Before Utilization
• Confirm before utilization (Measures to prevent possible problems, necessity of human intervention, confirmation before collaboration with other systems, etc.)
Data collection
Utilization
• Utilize systems and follow information.

6 . Commentary of the AI Utilization Principles

Commentary on points of the content of each principle

1-B) Human intervention (12)
Flow of Learning and Utilization for Machine Learning
Utilization phase (AI systems)
Inference
Deploy learned model
Learning phase
Learning

7 . Timing to Consider the AI Utilization Principle

Phases in which each principle and its points should considered

	Building AI	Implementation of Systems	Deployment	Utilization /Operation
1) Principle of proper utilization				
A) Utilization in the proper scope and manner	Yes	Yes	Yes	Yes
B) Human intervention	Yes	Yes	Yes	Yes
C) Cooperation among stakeholders			Yes	Yes
2) Principle of data quality				
A) Attention to the quality of data used for learning or other methods of AI	Yes			
B) Attention to security vulnerabilities of AI by learning from inaccurate or inappropriate data	Yes		Yes	
3) Principle of collaboration				
A) Attention to the interconnectivity and interoperability of AI systems		Yes	Yes	Yes
B) Address the standardization of data formats, protocols, etc.	Yes	Yes	Yes	Yes
C) Attention to problems caused and amplified by AI networking		Yes	Yes	Yes
4) Principle of safety				
A) Consideration for the life, body, and property		Yes	Yes	Yes
5) Principle of security				
A) Implementation of security measures		Yes	Yes	Yes
B) Service provision, etc. for security measures			Yes	Yes
C) Attention to security vulnerabilities of learned models	Yes		Yes	

Fig. 1: Structure of AI Utilization Guidelines

Definition of AI and Scope

(1) Definition of AI

The terms related to “AI” used in the Guidelines are defined as follows:

- “**AI**” refers to a concept that collectively refers to AI software and AI systems[5].
- “**AI software**” refers to software that has the function to change its own implementation or output in the process of utilization, by learning from data, information, and knowledge; or by other methods[6]. For example, machine learning software is classified into this category.
- “**AI system**” refers to systems that incorporate AI software as a component. For example, robots and cloud systems that implement AI software are classified under this category.

(2) Scope

The Guidelines cover **AI systems** that can be networked (i.e. connected to networks), since they can be used across national borders via networks, thereby widely bringing about benefits and risks to humans and society.

[5] This definition of AI in the Guidelines to apply mainly to Narrow AI which has already been put into practical application. In anticipation of rapid technological progress related to AI such as autonomous AI and artificial general intelligence (AGI), however, it will also be able to cover various types of AI to be developed in the future if they have functions to change their own outputs or programs by learning. In the Guidelines, the definition of AI as described above may apply to a variety of AI to be developed in the future depending on their functions. How to define AI in the Guidelines needs to be continuously discussed based on the trends of technological progress of AI.

[6] Methods other than learning, which might cause AI software to change its own implementation or output, include inferences based on data, information, and knowledge; and interactions with the environment through sensors, and actuators, etc.

1. Purpose

R&D and utilization of AI are expected to progress rapidly in the years to come. In the process of the evolution of AI networking[7], enormous benefits for humans as well as society and the economy come into being, for example, by making significant contributions to solving various problems that individuals, local communities, countries, and international community[8] are confronted with. The R&D and utilization of AI should be accelerated in such a direction.

As part of this, from the viewpoint of promoting benefits from AI to society and the economy, as well as mitigating any connecting risks such as a lack of transparency and loss of control, it becomes necessary to address relevant social, economic, ethical, and legal issues. In particular, services utilizing AI, like other information-and-communication ones, will be provided beyond national borders via networks; therefore, it is essential to promote the benefits of AI while mitigating risks by fostering an international consensus through open discussions between diverse stakeholders (e.g., developers, service providers, and users, including civil society, governments, and international organizations).

In view of this, the Guidelines aim to facilitate AI utilization and social implementation by way of increasing the benefits of AI, and mitigating the risks[9] of AI, as well as fostering trust in AI through the sound progress of AI networks.

From the viewpoint of achieving the above purposes, the Guidelines have formulated the matters that should be taken into consideration in AI utilization as the "AI Utilization Principles," which are themselves based on the draft proposed by the Conference in 2018, and they provide explanations on concrete measures expected to be taken for the achievement of them[10]. The Guidelines are regarded as practical guidance for AI service providers and business users, so as to establish their own AI development and utilization guidelines, which are

[7] AI networking refers to a formation of networks where AIs are connected, over the Internet or other information-and-communication networks, to each other or to other types of software, systems (hereinafter referred to as “AI networks”).

[8] For details on challenges faced by the international community, refer to the United Nations’ Sustainable Development Goals (SDGs) (http://www.un.org/ga/search/view_doc.asp?symbol=A/70/L.1) as an example.

[9] “Risks” mean “a possibility of causing disadvantages, such as damages.”

[10] See footnote 3.

based on the Social Principles for Human-centric AI.

There are various purposes and applications for AI utilization. Therefore, not all of the ten principles organized as the AI Utilization Principles will necessarily need to be taken into consideration in AI utilization. AI users are assumed to select the necessary principles that should be taken into consideration from the ten principles according to the purpose and social context of AI utilization, and then to consider what kind of measures should be taken voluntarily by referring to the commentaries, which are described later, for each of the selected principles.[11,12].

In addition, the matters expected to be taken into consideration in AI R&D were formulated as “Draft AI R&D Guidelines” in July 2017. It is difficult to distinguish AI development from AI utilization in several cases. Therefore, it is best to refer to both of them.

[11] It is recommended that attention is paid to the trade-offs described in the Appendix 2 in the Reference”.

[12] The number of companies formulating and publishing AI principles voluntarily have increased. In addition there have been cases of obtaining approval from internal ethics committees by maintaining in-house handling policies.

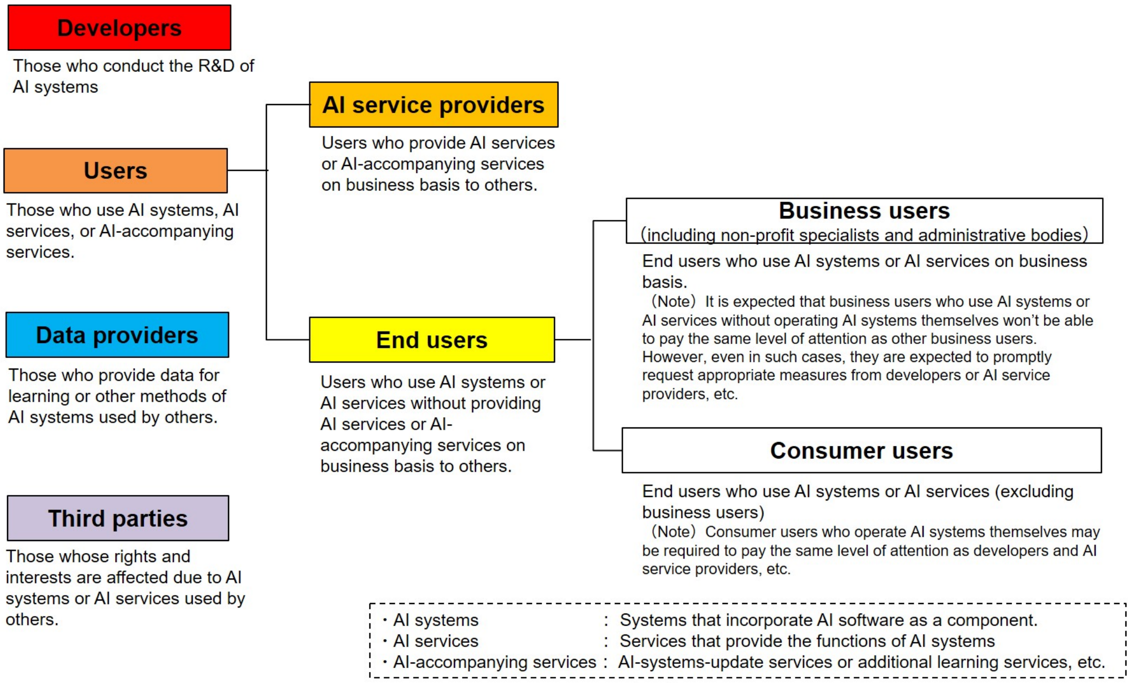
2. Basic Philosophies

Recognizing the purpose of the Guidelines mentioned above, seven basic philosophies have been formulated as follows:

- To **achieve a human-centered society** where all human beings across all of society enjoy the benefits while in harmony with AI networks, and human dignity and individual autonomy are respected.
- To **respect the diversity of people utilizing AI (users)** and **include** people with diverse backgrounds, values, and ideas.
- To achieve **a sustainable society through AI networking** to **solve various problems** faced by individuals, local communities, countries, and the international community.
- To **ensure an appropriate balance between the benefits and risks** of AI to enhance the benefits of AI networking, while at the same time respecting the values of democratic society and controlling the risk of infringement of rights and interests.
- To **realize appropriate role assignment among stakeholders** with consideration for the **abilities and knowledge** of AIs that each user should have.
- To **share the Guidelines as non-binding soft law and as best practices** on how to use AI.
- To constantly review the Guidelines through **continuous** international discussions, and **flexibly revise** them as necessary with consideration for the progress of AI networking.

3. Classification of Related Entities

This chapter shows stakeholders assumed to be involved in utilizing AI[13].



Those who conduct the R&D of AI systems

Those who use AI systems, AI services, or AI-accompanying services.

Those who provide data for learning or other methods of AI systems used by others.

Those whose rights and interests are affected due to AI systems or AI services used by others.

Users who provide AI services or AI-accompanying services on business basis to others.

Users who use AI systems or AI services without providing AI services or AI-accompanying services on business basis to others.

End users who use AI systems or AI services on business basis.

（Note）It is expected that business users who use AI systems or AI services without operating AI systems themselves won't be able to pay the same level of attention as other business users. However, even in such cases, they are expected to promptly request appropriate measures from developers or AI service providers, etc.

End users who use AI systems or AI services (excluding business users)

（Note）Consumer users who operate AI systems themselves may be required to pay the same level of attention as developers and AI service providers, etc.

・AI systems	:	Systems that incorporate AI software as a component.
・AI services	:	Services that provide the functions of AI systems
・AI-accompanying services	:	AI-systems-update services or additional learning services, etc.

（Note）One individual or enterprise may be included in multiple entities.

Fig. 2: Classification of Related Entities

- Developers:

  Those who conduct the R&D of AI systems.
- Users:

  Those who use AI systems, AI services, or AI-accompanying services.
- AI service providers:

  Users who provide AI services or AI-accompanying services on a business basis to others.
- End users:

  Those who use AI systems or AI services without providing AI services or AI-accompanying services on a business basis to others.
- Business users (including non-profit specialists and administrative bodies):

[13] Here, a typical classification based on the difference in the position in providing an AI system or AI service is described. Therefore, if a company develops an AI service in-house and provides the service while also utilizing the service for the company's own business, the company has a character as both developers and business users. In that case, it is recommended that they refer to both the Draft AI Development Guidelines and the AI Utilization Guidelines.

End users who use AI systems or AI services on a business basis.
It is assumed that business users who use AI systems or AI services, without operating AI systems themselves, won’t be able to pay the same level of attention as other business users.
However, even in such cases, they are expected to promptly request appropriate measures from developers or AI service providers, etc.

- Consumer users:
  End users who use AI systems or AI services (excluding business users)
  However, consumer users who operate AI systems themselves may be required to pay the same level of attention as developers or AI service providers, etc.
- Data providers:
  Those who provide data for learning or other methods of AI systems used by others.
- Third parties:
  Those whose rights and interests are affected due to AI systems or AI services used by others.

4. AI Utilization Principles

With consideration of the above-mentioned purpose and basic philosophies, this chapter organizes the matters to be taken into consideration by AI users into ten principles as follows:

1) Principle of proper utilization

Users should make efforts to utilize AI systems or AI services in proper scope and manner, under the proper assignment of roles between humans and AI systems, or among users.

2) Principle of data quality

Users and data providers should pay attention to the quality of data used for learning and other methods of AI systems.

3) Principle of collaboration

AI service providers, business users, and data providers should pay attention to the collaboration of AI systems or AI services. Users should take into consideration that risks might occur and even be amplified when AI systems are to be networked.

4) Principle of safety

Users should take into consideration that AI systems or AI services in use will not harm the life, body, or property of users or third parties through actuators or other devices.

5) Principle of security

Users and data providers should pay attention to the security of AI systems or AI services.

6) Principle of privacy

Users and data providers should take into consideration that the utilization of AI systems or AI services will not infringe on the privacy of users or others.

7) Principle of human dignity and individual autonomy

Users should respect human dignity and individual autonomy in the utilization of AI systems or AI services.

8) Principle of fairness[14]

AI service providers, business users, and data providers should pay attention to the possibility of bias inherent in the judgements of AI systems or AI services, and take into consideration that individuals and groups will not be unfairly discriminated against by their judgements.

9) Principle of transparency[15]

AI service providers and business users should pay attention to the verifiability of inputs/outputs of AI systems or AI services and the explainability of their judgments.

10) Principle of accountability[16]

Users should make efforts to fulfill their accountability to stakeholders.

[14] Note that there are several criteria for “fairness” such as group fairness and individual fairness.

[15] This principle is not intended to ask for the disclosure of algorithms, source codes, or learning data. In interpreting this principle, privacy of individuals and trade secrets of enterprises are also taken into account.

[16] “Accountability” means the possibility to take appropriate measures, such as to proving the explanation behind the meaning and reason for the judgment, along with compensation as needed, after clarifying with the person responsible, in order to gain the understanding of the person who is affected by the result of the judgment.

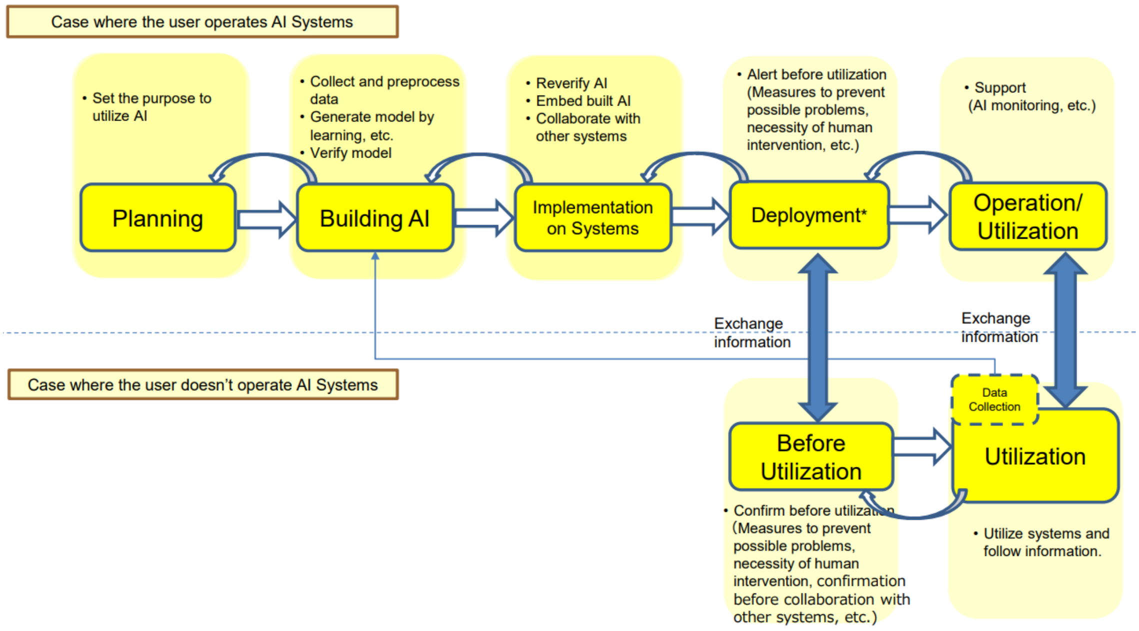
5. General Flow of AI Utilization

This chapter organizes the general flow of AI utilization as the following two classes in order to clarify in which phase of utilization of AI system or services the above-mentioned AI Utilization Principles are taken into consideration:

(i) Users operating AI systems or services (including learning) by themselves.

(ii) Users using AI systems or AI services without operating them.

The flow of AI utilization in each case is as follows:



Case where the user operates AI Systems
• Set the purpose to utilize AI
Planning
• Collect and preprocess data
• Generate model by learning, etc.
• Verify model
Building AI
• Reverify AI
• Embed built AI
• Collaborate with other systems
Implementation on Systems
• Alert before utilization (Measures to prevent possible problems, necessity of human intervention, etc.)
Deployment*
• Support (AI monitoring, etc.)
Operation/ Utilization
Exchange information
Exchange information
Case where the user doesn't operate AI Systems
Before Utilization
• Confirm before utilization (Measures to prevent possible problems, necessity of human intervention, confirmation before collaboration with other systems, etc.)
Data Collection
Utilization
• Utilize systems and follow information.

*Deployment: To make AI software/systems available

Figure 3: General flow of AI utilization for each entity[17]

(i) Operating AI systems or services by themselves

- A) Planning phase:

  This phase sets the purpose of using AI, and a study is made on roughly what kind of data should be applied.

- B) AI building phase:

[17] This general flow of AI utilization describes a typical case in order to clarify the phase in which each principle is to be considered in Chapter 7. On the other hand, AI utilization is not limited to cases where the phases are arranged in chronological order as shown in the above figure, and various cases are assumed such as in the case that development and operation are performed simultaneously (ex. DevOps). Therefore, when applying the Guidelines, it is expected to replace each description according to the flow of each utilization.

This phase is to build AI software and perform verification through trials. In this phase, data collection, the preparation of models by preprocessing and learning, and tasks such as verification, are performed.

- C) System Implementation phase:
  This phase is where AI software created in phase B is introduced into systems to perform verification. The system referred to here is assumed to be both existing and new cases. Furthermore, verification with other systems is also expected.
- D) Deployment phase:
  This phase is where users themselves, including consumer users, can use AI systems created in phase C. it is assumed that AI system users themselves, including consumer users, will be provided with information when making the AI systems available.
- E) Operation/Use phase:
  This phase is where deployed AI is operated for users, including consumer users. The monitoring of the AI is assumed in this phase with consideration of the autonomous change of AI according to data given to AI, as well as response to inquiries from users, including consumer users.

(ii) Not operating AI systems or services by themselves

- A) Pre-utilization phase:
  This phase occurs before using AI. It is assumed that information provided by developers and AI service providers is grasped before use.
- B) Utilization phase:
  This phase is where AI is utilized. It is expected that AI is utilized and updated, if necessary, based on information provided by the developers and AI service providers.

The above describes the flow of general AI utilization. However, exceptions exist, and it is expected that a necessary reading is to be performed. For example, no system will be implemented in the case of providing a single piece of software that makes some judgment through a machine learning model as an AI service to consumer users. Therefore, phase C in (i) should be read as that only linking with other systems is to be verified.

6. Commentary on the AI Utilization Principles

Concerning the ten principles organized as AI utilization principles, this chapter explains the matters to be taken into consideration for each principle by AI service providers, business users, and data providers (points that consumer users are recommended to take into consideration are also described for <*Reference*>). Refer to the attached references for specific examples of the underlined parts, if necessary.

1) Principle of proper utilization

Users should make efforts to utilize AI systems or AI services in a proper scope and manner, under the proper assignment of roles between humans and AI systems, or among users.

[A. Utilization in the proper scope and manner]

- AI service providers and business users are expected to use AI in a proper scope and manner based on information and explanations provided by developers, and after properly recognizing the utilization purpose, usage, nature, and capability of the AI according to its social context. Furthermore, AI service providers are expected to provide necessary information in a timely manner.
- AI service providers are expected to provide AI software updates[18] and AI inspection/repair, etc. services to improve AI functions and mitigate risks through the process of utilization. In particular, if it is assumed that the update affects[19] other linked AI systems, AI service providers are expected to provide information on these risks.

[18] From the time a problem is discovered to the time an update information is provided, AI service providers are expected to give information on the problem to end users in a timely and appropriate manner and alert them.

[19] It is assumed that the operation of AI to which the update is applied affects the other AIs. For example, if AI software incorporated into a home electrical appliance is updated, it is assumed that a conflict will result between the judgment of a home administration robot in control of the whole house and that of other home electrical appliances incorporating AI unless they respond to the updating. (An example case in *Appendix 3 Risk of AI in Unexpected Operation* in “Report 2018”).

- Depending on the nature and usage mode of AI systems or AI services to be provided, AI service providers are expected to confirm the reliability of users in advance if cases where the use of AI is likely to harm human lives, bodies, and property. Furthermore, after an AI service is provided, there may be a necessity for recording and saving input and output logs on the service in order to make sure that no end users misuse or make malicious use of the AI service or AI system.

<Reference>

- *Consumer users are recommended to use AI in a proper range and method, with consideration for information and explanations provided from developers and AI service providers, along with the social context (Matters that should be taken into consideration).*

[B. Human intervention]

- AI service providers and business users are expected to allow the intervention of human decisions, if necessary and possible, as to whether or not to use the AI's decision. In that case, the necessity for human intervention is considered according to the field and application of the AI in accordance with the example criteria.
- When it is considered appropriate for humans to make a final decision based on the AI's decision, there is a possibility that humans may not make decisions that differ from the AI's decision. Therefore, the effectiveness of humans' decision should be ensured by clarifying the items to be judged in advance, provided that an explanation is obtained from the explainable AI[20].
- In the case of the utilization of an AI that is operated through actuators for a system, if the system shifts to manual operations under certain conditions, it is necessary to clarify in advance the whereabouts of responsibility for each state, i.e., before, during, and after the system shifts to manual operations. AI service providers are expected to take proactive measures to prevent problems if their AI systems shift to human operations, e.g., explaining in advance the transition conditions and transition methods to end users and carrying out the necessary training.

<Reference>

[20] In addition, in order to ensure the appropriateness of AI's judgment that humans confirm, it is recommended that other measures are considered for example, to double-check using the other AI systems for the confirmation of AI operations and to do the input perturbation to AI.

- *If it is considered appropriate for consumer users to give final approval to AI's decision, they are recommended to acquire the necessary skills and knowledge to make appropriate decisions.*
- *They are recommended to respond appropriately based on the measures organized to ensure the effectiveness of humans' decision if developers and AI service providers have the measures.*
- *In the case of the utilization of AI operated through actuators for a system, if the system shift to manual operation under certain conditions, consumer users are recommended to have a clarification in advance the whereabouts of responsibility for each state, i.e., before, during, and after the system shift to manual operation, and to receive an explanation from AI service providers for transition conditions and methods and acquire necessary skills and knowledge.*

[C. Cooperation among stakeholders]

- AI service providers, business users and data providers are expected to cooperate with the related stakeholders and to work on preventive or remedial measures (including information sharing, stopping and restoration of AI, elucidation of causes, and measures to prevent recurrence, etc.) in accordance with the nature, conditions, etc. of accidents through the use of AI or damages caused by security breaches and privacy infringement, etc. that may occur in the future or have occurred.

<Reference>

- *With consideration of information that developers or AI service providers provide, consumer users are recommended to cooperate with the related stakeholders and to work on preventive or remedial measures (including information sharing, stopping and restoration of AI, elucidation of causes, and measures to prevent recurrence, etc.) in accordance with the nature, conditions, etc. of accidents through the use of AI or damages caused by security breaches and privacy infringement, etc. that may occur in the future or have occurred.*

2) Principle of data quality

Users and data providers should pay attention to the quality of data used for learning and other methods of AI systems.

[A. Attention to the quality of data used for the learning or other methods of AI]

- AI service providers, business users, and data providers are expected to pay attention to the quality of data (e.g. data integrity) used for learning or other methods of AI, with consideration of the characteristics of AI to be used and its usage (measures for ensuring the quality of data used in machine learning).

- It is assumed that the accuracy[21] of an AI's judgment can become impaired or decline afterwards. Therefore, AI service providers, business users, and data providers are expected to define reference levels concerning accuracy in advance based on the assumed magnitude and frequency of occurrence of the infringement of rights, the technology level available, and the cost[22] to maintain accuracy, etc. If the accuracy falls below a reference level, they are expected to put the AI through relearning with consideration for data quality.
- If it is planned to use data provided by consumer users, they are expected to provide consumer users with information on the means and format of data provision in advance, taking into consideration the characteristics and usage of the AI.

<Reference>

- *If consumer users plan to collect data by themselves and make AI learn the collected data, it is recommended that the data format and content[23] is based on the information provided by developers and AI service providers.*

[B. Attention to security vulnerabilities of AI by learning inaccurate or inappropriate data]

- AI service providers, business users, and data providers are expected to pay attention to the risk that AI security might become vulnerable by learning inaccurate or inappropriate data. They are also expected to inform consumer users in advance of the existence of such risks.

<Reference>

- *Consumer users are recommended to pay attention to the risk of vulnerability to AI security by learning inaccurate or inappropriate data with consideration for information from developers, AI service providers, and data providers.*
- *Furthermore, if they have doubts in security when using the AI, they are recommended to report them to developers, AI service providers, and data providers.*

3) Principle of collaboration

AI service providers, business users, and data providers should pay attention to the collaboration of AI systems or AI services. Users should

[21] The term "accuracy" includes checks on whether AI is making right judgments, for example, checks on whether AI is not using a violent expression or making hate speech.

[22] For example, since AI based on machine learning is an inductive approach, the corresponding AI alone cannot in principle guarantee 100% accuracy.

[23] It is recommended to confirm whether the data is not wrong or due to malicious input.

take into consideration that risks might occur and even be amplified when AI systems are to be networked.

[A. Attention to the interconnectivity and interoperability of AI]

- AI service providers are expected to pay attention to the interconnectivity and interoperability of AI, with consideration for the characteristics of AI to be used and its usage, in order to promote the benefits of AI through the sound progress of AI networking.

[B. Address the standardization of data formats, protocols, etc.]

- AI service providers and business users are expected to comply with the following standards in order to promote collaboration among AIs, and between AI and other systems: Data format (with syntax and semantics[24]) for AI input and output and connection methods for collaboration (especially protocols in each layer when using a network for collaboration).
- Data providers are also expected to comply with data format standards (with syntax and semantics) to promote collaboration among AIs, and between AI and other systems.

<Reference>

- *If consumer users plan to collect data by themselves and make AI learn from the collected data, it is are recommended that the data format is based on the information provided by developers and AI service providers.*

[C. Attention to problems caused and amplified by AI networking]

- Although it is expected that benefits will be promoted through collaboration between AIs, AI service providers, business users, and data providers should pay attention to the possibility that risks (e.g. loss of control by interconnecting or collaborating their AIs with other AIs, etc. through the Internet or other network) might be caused and amplified by AI networking. Therefore, AI service providers, business users, and data providers are expected to analyze possible risks with consideration for information from developers, share the risks with the cooperating parties, organize preventive measures and countermeasures for problems, if any, and provide necessary information to consumer users.

[24] Collaboration will not work correctly unless the meaning is shown even if the syntax of data is shown.

<Reference>

- *Although it is expected that benefits will be enhanced through the interaction of AI systems, consumer users are recommended to pay attention that risks (e.g. loss of control by interconnecting or collaborating their AIs with other AIs, etc. through the Internet or other network) might be caused and amplified by AI networking. Furthermore, if information on preventive measures in advance or countermeasures for problems that have occurred are provided from the developers and AI service providers, they are recommended to be careful when using AIs.*

4) Principle of safety

Users should take into consideration that AI systems or AI services in use will not harm the life, body, or property of users or third parties through actuators or other devices.

[A. Consideration for the life, body, and property]

- In cases where AI is used in fields where AI may harm human life, body, or property, AI service providers and business users are expected to take into consideration so that AI will not harm them through actuators or other devices by taking measures as necessary, based on information from the developers, and with consideration of the nature, and conditions, etc. of the assumed damage.
- Furthermore, AI service providers and business users are expected to organize in advance the measures to be taken if an AI damages a human life, body, or property through actuators or other devices. Also, they are expected to provide necessary information on such countermeasures to consumer users.

<Reference>

- *In case where AI is used in fields where AI may harm a human life, body, or property, consumer users are recommended to take into consideration that AI will not harm the human life, body, or property through the actuators or other devices by taking measures (e.g. checking AI, updating AI software, etc.) as necessary, based on information from the developers and AI service providers and with consideration of the nature, and conditions, etc. of assumed damages.*
- *Furthermore, if developers and AI service providers provide information on measures to be taken to prevent AI damage to human life, body, and property through actuators or other devices, consumer users are recommended to keep the information in mind when using the AI.*

5) Principle of security

Users and data providers should pay attention to the security of AI systems or AI services.

[A. Implementation of security measures]

- AI service providers and business users are expected to pay attention to the security of AI and take reasonable measures corresponding to the technology level at that time to ensure the confidentiality, integrity and availability (CIA) of AI systems.
- Furthermore, they are expected to organize measures to be taken against security infringement in advance, taking into consideration the usage and characteristics of the AI and the magnitude of the influence of the infringement.

<Reference>

- *If consumer users are supposed to implement security measures (on their side), they are recommended to pay attention to the security of the AI and take necessary measures based on the provision of information from developers and AI service providers.*

[B. Service provision, etc. for security measures]

- AI service providers are expected, with regard to their AI services, to provide end users services with security measures for AI services and to share past accident and incident information.
- Furthermore, AI service providers and business users are expected to provide consumer users with the necessary information on measures in cases of security infringements.

<Reference>

- *If developers and AI service providers provide information on measures to be taken against security infringement, consumer users are recommended to pay attention to this information when using the AI.*
- *Furthermore, if there are security concerns when using the AI, they are recommended to report them to developers, AI service providers, and data providers.*

[C. Attention to security vulnerabilities of learned models]

- AI service providers, business users, and data providers are expected to pay attention to the risk that learning models in AI might be vulnerable in their generation and management. They are also expected to inform consumer users in advance the existence of such risks.

<Reference>

- *Consumer users are recommended to pay attention to the risk that AI might become vulnerable in the generation and management of learning models, with consideration of information provided by developers, AI service providers, and data providers.*
- *Furthermore, if there are security concerns when using the AI, they are recommended to report them to developers, AI service providers, and data providers.*

6) Principle of privacy

Users and data providers should take into consideration that the utilization of AI systems or AI services[25] will not infringe on the privacy of users or others.

[A. Respect for the privacy of end users and others]

- AI service providers and business users should respect the privacy of end users and third parties in the utilization of AI, based on the social context and reasonable expectations of people in the utilization of AI.
- Furthermore, they are expected to consider measures to be taken against privacy infringement of end users and third parties caused by AI.
- Besides, they are expected to provide the necessary information on such countermeasures to end users and third parties.

<Reference>

- *Consumer users should respect the privacy of third parties in the utilization of AI, based on the social context and reasonable expectations of people in the utilization of AI.*
- *Furthermore, if developers and AI service providers provide information on measures to be taken against privacy infringement of third parties, they are recommended to pay attention to the information when using the AI.*

[B. Respect for the privacy of others in the collection, preprocessing, and provision, etc. of personal data]

- AI service providers, business users, and data providers should respect the privacy of end users and third parties in the collection, preprocessing, and

[25] In Japan, it is precondition to comply with the Act on the Protection of Personal Information.

provision[26,27] etc. of personal data used for AI learning and in the provision of learning models generated through them.

<Reference>

- *Consumer users should respect the privacy of third parties in the collection of data if they plan to collect data on their own for AI learning.*

[C. Attention to the infringement of the privacy of users or others, and prevention of personal data leakage]

- AI service providers, business users, and data providers are expected to take appropriate measures, including the prevention of unconsented data being made available to third parties, in their systems so that personal data is not provided under the judgement of an AI to third parties without the consent of those persons.

<Reference>

- *Consumer users are recommended to be careful not to give particularly confidential information (e.g., others' personal information as well as their own information) unnecessarily to AI as a result of being overly emotional toward AI, including pet robots.*

7) Principle of human dignity and individual autonomy

Users should respect human dignity and individual autonomy in the utilization of AI systems or AI services.

[A. Respect for human dignity and individual autonomy]

- AI service providers and business users are expected to respect human dignity and individual autonomy based on the social context in the AI utilization[28].

<Reference>

- *Consumer users are recommended to respect human dignity and personal autonomy based on the social context in the AI utilization.*

[B. Attention to the manipulation of human decision making, and

[26] With regard to the handling of personal data provided to others, the deletion of the personal data, for example, is expected.

[27] AI service providers, business users, and data providers are required to understand by whom and how the data they provide is used if the data contains personal information.

[28] For example, to recognize that AI supports human activities, on the assumption of the heterogeneity of humans and AI. The heterogeneity of humans and AI means that humans and AI have different natures. By this assumption, it can be recognized that AI should not be treated like a human (i.e., respect human dignity and individual autonomy).

emotions, etc. by AI]

- AI service providers and business users are expected to take necessary measures with consideration for the possibility[29] that consumer users' decision or emotions are manipulated by AI and the risk of their overdependence on AI.

<Reference>

- *Consumer users are recommended to recognize the possibility that their decisions or emotions are manipulated by AI and the risk of over-dependence on AI, with consideration for information from developers and AI service providers[30].*

[C. Reference to the discussion of bioethics, etc. in the case of linking AI systems with the human brain and body]

- If AI is linked to the human brain and/or body, especially in pursuit of human enhancement (in pursuit of enhancements or improvements in the capabilities of humans that transcends maintaining or recovering health), AI service providers and business users are expected to particularly take into consideration that human dignity and autonomy are not violated, in light of the discussion of bioethics and information from developers about the surrounding technologies.
- Furthermore, they are expected to provide information on the function and peripheral technology of AI to be provided to consumer users.

<Reference>

- *If consumer users use AI that links to the human brain and body, they are recommended to pay attention to the possibility of the AI affecting the autonomy of humans and use the AI with consideration of information on functions and peripheral technologies of the AI from developers and AI service providers.*

[D. Consideration for prejudice against the subject in profiling which uses AI]

- In the case of profiling by using AI in fields that might have a significant influences on individual rights and interests, AI service providers and

[29] The term "possibility" is used here because the decision-making and emotional manipulation of consumer users by AI is not always risk-taking when AI performs nudging (i.e., support for a rational choice). In addition, when AI performs nudging, consumer users are expected to refer to the Principle of user assistance (make it possible to provide selection opportunities to the users) in the AI R&D Guidelines.

[30] This includes not only information obtained directly from the developer and AI service provider, but also information obtained at educational sites, etc.

business users are expected to carefully consider[31,32] all disadvantages that may occur to the target individuals.

<Reference>

- *Consumer users are recommended to be aware of the proper use of their information and, if necessary, check with AI service providers and business users considering that profiling might take place by AI.*

8) Principle of fairness

AI service providers, business users, and data providers should pay attention to the possibility of bias[33] inherent in the judgements of AI systems or AI services, and take into consideration that individuals and groups will not be unfairly discriminated against by their judgments.[34]

[A. Attention to the representativeness of data used for learning or other methods of AI]

- AI service providers, business users, and data providers are expected to pay attention to the representativeness[35] of data used for AI learning or other methods and the social bias inherent in the data according to the social context in utilizing AI, with consideration for how the results of AI judgements may be determined by learning data.

<Reference>

- *If consumer users have any doubts in the decision made by an AI, they are recommended to contact developers, AI service providers, and business users as required.*

[B. Attention to unfair discrimination by learning algorithms]

[31] Article 22 of the EU's General Data Protection Regulation guarantees that the data subject shall have the right not to be subject to a decision based solely on automated processing.

[32] Refer to B) Human Intervention under 1) Principle of proper utilization.

[33] The term "bias" has various possible interpretations as follows and is used as all-inclusive term in the Guidelines:

- Statistic terms (sampling bias, and deviation, etc.)
- Psychological terms (cognitive bias (due to delusion, including social bias due to conventional wisdom etc. for each group), emotional bias (due to human emotion and opportunity) etc.)

[34] Note that there are several criteria for fairness such as group fairness and individual fairness.

[35] The "representativeness" of data means the state in which data extracted as a sample and subjected to utilization does not distort the nature of the statistical population.

- AI service providers and business users are expected to pay attention to the possibility of bias inherent in AI judgements due to the algorithm used in it. In machine learning in particular, the majority tends to be adopted, and the minority is less likely to be done (bandwagon effect). Therefore, there are several measures to avoid this effect.

<Reference>

- *If consumer users have any doubts in the decision made by an AI, they are recommended to contact developers, AI service providers, and business users as required.*

[C. Human intervention (viewpoint of ensuring fairness)]

- AI service providers and business users are expected to intervene with human judgment on whether to adopt, or how to use, the judgement of AI, as well as with consideration for the social context and the reasonable expectations of people when utilizing the AI, to ensure the fairness[36] of the judgement result from it.
- They are expected to consider the necessity of human intervention based on examples of criteria from a viewpoint of fairness while referring to the content of [1)- B].

9) Principle of transparency

AI service providers and business users should pay attention to the verifiability of inputs/outputs[37] of AI systems or AI services and the explainability of their judgments.

[A. Recording and preserving logs such as inputs/outputs, etc. of AI]

- AI service providers and business users are expected to record and preserve logs, including those on inputs/outputs, to ensure the verifiability[38] of inputs/outputs of AI. When recording and preserving logs, they are expected to consider the purpose of log recording and preservation, and the frequency of log acquisition and recording, etc. with consideration of the characteristics of technology to be used and its usage.

[36] It is premised that the social bias inherent in data used in AI learning can affect the fairness of the decision made by an AI.

[37] This principle is not intended to ask the disclosure of the algorithm, source code, or training data. In interpreting this principle, the privacy of individuals and trade secrets of enterprises are also taken into account. .

[38] Scenarios that are expected to ensure input and output verifiability assume the case that make sure that end users are not using AI wrongly or with malicious intentions, in addition to the case to clarify the causes of accidents, if any.

[B. Ensuring explainability]

- AI service providers and business users are expected to ensure the explainability of the judgment results of AI for the purpose of ensuring the trust of users and to present evidence of AI behavior with consideration of the social context, in case of utilizing AI in a field that has a significant impact on individual's rights and interests.

  At that time, with consideration of the social context in the AI utilization, they are expected to ensure the explainability of the decision results made by AI, by analyzing and understanding what kind of explanation is required and taking necessary measures,.

[C. Ensuring transparency when AI is used in administrative bodies]

- When administrative bodies use AI, they are expected to ensure the explainability of decision results made by AI, according to the social context in the AI utilization with consideration of the reign of law, while ensuring administrative transparency, and keeping within the requirement of proper procedures (Examples of measures to improve explainability).

10) Principle of accountability

Users should make efforts to fulfill their accountability to stakeholders.

[A. Efforts to fulfill accountability]

- With consideration for the purpose of the Utilization Principles (1) to (9) described in these Guidelines in order to earn the trust of AI from people and societies, AI service providers and business users are expected to strive to fulfill the corresponding accountability to consumer users and third parties affected by AI utilization. Therefore, based on the nature and purpose of the AI to be used, they are expected to provide, and further explain, information on the characteristics of the AI system, and communicate with various stakeholders according to their knowledge and capability.

<Reference>

- *Consumer users are recommended to strive to fulfill their accountability according to their knowledge and ability.*
- *If consumer users have any doubts in the decision made by an AI, they are recommended to contact developers, AI service providers, and business users as required.*

[B. Notification and publication of usage policy on AI systems or AI services]

- AI service providers and business users are expected to create, publish and notify AI usage policies as described below so that consumer users and others can appropriately recognize the utilization of AI.
  - (i) To create and publish an AI usage policy so that consumer users and third parties are aware of the use of AI in case the judgment of the AI could directly affect them, and to provide notifications to them when asked.
  - (ii) Regarding (i), to proactively provide notifications to consumers and third parties[39] in case their rights and interests may have been seriously affected concerning (i),
- They are expected to publish or notify them, not only before use of AI is started, but after AI behavior is changed or the use of AI is terminated (especially when assumed risks are changed due to a change of AI behavior).

<Reference>

- *If consumer users have any doubts in the decision made by an AI, they are recommended to contact developers, AI service providers, and business users as required.*

7. Timing to Consider the AI Utilization Principles

Table 1 and Table 2 summarize in which phase each principle and its point listed in Chapter 6 should be considered according to the flow of AI utilization[40].

This chapter summarizes the relationship with each utilization phase on the assumption that AI service providers and business users operate AI and that consumer users do not[41]. For this reason, in the case of business users

[39] It is considered that AI service providers and business users are required to publish usage policies related to AI if the judgment of AI to be used directly affects consumer users and third parties. In other words, if AI is only used as an analytical tool for human thinking, or if AI is making a draft, but it is practically guaranteed that humans will ultimately judge, it is not always required to announce the usage policy regarding AI. (However, even in such a case, it is recommended that the announcement should be voluntarily published.)

[40] Tables 1 and 2 describe the phases for considering each point of each principle. However, it should be noted that the necessity and degree of consideration of each point should be considered according to the purpose and social context of AI utilization.

[41] Data providers are organized based on the utilization phase of AI service providers and business users who operate AI.

who do not operate AI and consumer users who operate AI, it is expected that the description in the tables be replaced as necessary.

Table 1: Relationship AI Utilization Flow of AI Service Providers, and Business Users, etc. with Each Principle and Point

	Building AI	Implementation of Systems	Deployment	Utilization /Operation
1) Principle of proper utilization				
A) Utilization in the proper scope and manner	Yes	Yes	Yes	Yes
B) Human intervention	Yes	Yes	Yes	Yes
C) Cooperation among stakeholders			Yes	Yes
2) Principle of data quality				
A) Attention to the quality of data used for learning or other methods of AI	Yes			
B) Attention to security vulnerabilities of AI by learning from inaccurate or inappropriate data	Yes		Yes	
3) Principle of collaboration				
A) Attention to the interconnectivity and interoperability of AI systems		Yes	Yes	Yes
B) Address the standardization of data formats, protocols, etc.	Yes	Yes	Yes	Yes
C) Attention to problems caused and amplified by AI networking		Yes	Yes	Yes
4) Principle of safety				
A) Consideration for the life, body, and property		Yes	Yes	Yes
5) Principle of security				
A) Implementation of security measures		Yes	Yes	Yes
B) Service provision, etc. for security measures			Yes	Yes
C) Attention to security vulnerabilities of learned models	Yes		Yes	
6) Principle of privacy				
A) Respect for the privacy of end users and others		Yes	Yes	Yes
B) Respect for the privacy of others in the collection, preprocessing, provision, etc. of personal	Yes		Yes	
C) Attention to the infringement of the privacy of users' or others and prevention of personal data leakage		Yes		
7) Principle of human dignity and individual autonomy				
A) Respect for human dignity and individual autonomy	Yes	Yes	Yes	Yes
B) Attention to the manipulation of human decision making, emotions, etc. by AI			Yes	Yes
C) Reference to the discussion of bioethics, etc. in the case of linking AI systems with the human brain and body		Yes	Yes	Yes
D) Consideration for prejudice against the subject in profiling which uses AI	Yes	Yes	Yes	Yes
8) Principle of fairness				
A) Attention to the representativeness of data used for learning or other methods of AI	Yes	Yes	Yes	Yes
B) Attention to unfair discrimination by learned algorithm	Yes	Yes	Yes	Yes
C) Human intervention (viewpoint of ensuring fairness)	Yes	Yes	Yes	Yes
9) Principle of transparency				
A) Recording and preserving logs such as inputs/outputs, etc. of AI		Yes	Yes	Yes
B) Ensuring explainability	Yes			
C) Ensuring transparency when AI is used in administrative bodies	Yes	Yes	Yes	Yes
10) Principle of accountability				
A) Efforts to fulfill accountability	Yes	Yes	Yes	Yes
B) Notification and publication of usage policy on AI systems or AI services	Yes	Yes	Yes	Yes

Table 2: Relationship of AI Utilization Flow of Consumer Users with Each Principle and Point

	Before use	During use	Data collection
1) Principle of proper utilization			
A) Utilization in the proper scope and manner	Yes	Yes	
B) Human intervention	Yes	Yes	
C) Cooperation among stakeholders	Yes	Yes	
2) Principle of data quality			
A) Attention to the Quality of Data Used for the Learning of AI			Yes
B) Attention to security vulnerabilities of AI by learning inaccurate or inappropriate data		Yes	Yes
3) Principle of collaboration			
A) Attention to the interconnectivity and interoperability of AI systems	Yes	Yes	
B) Address the standardization of data formats, protocols, etc.	Yes	Yes	Yes
C) Attention to problems caused and amplified by AI networking	Yes	Yes	
4) Principle of safety			
A) Consideration for life, body, and property	Yes	Yes	
5) Principle of security			
A) Implementation of security measures	Yes	Yes	
B) Service provision, etc. for security measures	Yes	Yes	
C) Attention to security vulnerabilities of learned models		Yes	Yes
6) Principle of privacy			
A) Respect for the privacy of end users and others	Yes	Yes	
B) Respect for the privacy of others in the collection, preprocessing, provision, etc. of personal data	Yes		Yes
C) Attention to the infringement of the privacy of users' or others and prevention of personal data leakage		Yes	
7) Principle of human dignity and individual autonomy			
A) Respect for human dignity and individual autonomy	Yes	Yes	
B) Attention to the manipulation of human decision making, emotions, etc. by AI	Yes	Yes	
C) Reference to the discussion of bioethics, etc. in the case of linking AI systems with the human brain and body	Yes	Yes	
D) Consideration for prejudice against the subject in profiling which uses AI	Yes	Yes	
8) Principle of fairness			
A) Attention to the representativeness of data used for learning or other methods of AI	Yes	Yes	Yes
B) Attention to unfair discrimination by algorithm	Yes	Yes	Yes
C) Human intervention (viewpoint of ensuring fairness)			
9) Principle of transparency			
A) Recording and preserving logs such as inputs/outputs, etc. of AI			
B) Ensuring explainability			
C) Ensuring transparency when AI is used in administrative bodies			
10) Principle of accountability			
A) Efforts to fulfill accountability	Yes	Yes	Yes
B) Notification and publication of usage policy on AI systems or AI services	Yes	Yes	Yes