当翻訳は仮訳であり、正確には原文を参照してください。
Please refer to the original text for accuracy.

仮訳
Provisional Translation

（Draft） AI Guidelines for Business

January 2024

Ministry of Internal Affairs and Communications
Ministry of Economy, Trade and Industry

Table of Contents

Introduction

AI(Artificial Intelligence)-related technologies are developing day by day, and the opportunities and various possibilities for using AI are expanding, and it is also being used to create innovation in industry and solve social issues. In addition, the “democratization of AI” has occurred due to the emergence of conversational generative AI in recent years, and many people have become able to easily use AI for various purposes through “dialogue”. As a result, companies are not only incorporating AI into their business processes, but also rebuilding their business models leveraging the value created by AI. In addition, individuals are also accelerating efforts to reflect their own knowledge in AI and expand their own productivity. In Japan, Society 5.0 has long been a concept of a human-centered society that achieves both economic development and the resolution of social issues through a system that highly integrates cyberspace and physical space (CPS: Cyber-Physical System). In order to realize the concept, the “Human-Centered AI Social Principles” were formulated in March 2019 with the aim of AI’s proper usage and acceptance from the society. On the other hand, risks are increasing as the scope of use of AI technology and the number of users in those fields expand. In particular, with regard to generative AI, new societal risks have arisen that were not previously associated with AI, such as invasion of intellectual property rights, generation and dissemination of disinformation and misinformation, and the risks to society brought about by AI are becoming more diverse and expanding.

Due to those background, this guideline provides unified guiding Principles for AI governance in Japan to promote the safe and secure use of AI. This will enable those who utilize AI in various business to correctly recognize the risks of AI based on international trends and stakeholder concerns, and to independently implement the necessary measures throughout the AI lifecycle. This will also support proactive co-create of a framework that both promotes innovation and mitigates risks throughout the implementation of “Common guiding Principles” and Important matters for each AI business actor, AI governance in collaboration with related parties.

Japan has led discussions at international organizations such as the G7, G20 and OECD, and has made many contributions starting with the proposal for AI development principles at the G7 Kagawa-Takamatsu Information and Communications Ministers Meeting held April 2016. On the other hand, the following points have been made regarding the practical implementation of AI principles.

- The use of AI is expected to be a means of solving social issues such as a decline in the workforce due to the declining birthrate and aging population.
- The development and enforcement of laws has not kept up with the technological development of AI and the speed and complexity of its social implementation.
- Rule-based regulations that stipulate detailed behavioral obligations may inhibit innovation.

Based on this, in order to reduce the societal risks brought by AI and promote the innovation and utilization of AI, we decided to create guidelines with a goal-based approach that leads to achieving objectives through non-binding soft law, and that will encourage voluntary efforts by AI business actors and related parties.

Based on those recognition, so far, the Ministry of Internal Affairs and Communications has led to formulate and publish the “The Draft AI R&D GUIDELINES for International Discussions”, “AI Utilization Guidelines - Practical Reference for AI Utilization”, and the Ministry of Economy, Trade and Industry has also led to formulate and publish the “Governance Guidelines for

Implementation of AI Principles Ver. 1.1”. Now, based on the three guidelines, we have recently formulated a new guideline (non-binding soft law) for AI business actors to practice social implementation of AI and governance reflecting the characteristics of AI technology that has further developed over the past few years and the discussion regarding the social implementation of AI in both domestically and internationally (see “Figure 1. Role of this guideline”). By referring to this guideline instead of the previous guidelines, businesses that utilize AI (including public institutions such as governments and local governments) can confirm the guiding principles that leads to desirable actions for the safe and secure use of AI. Furthermore, this guideline was developed with an emphasis on effectiveness through repeated deliberations in collaboration with multi-stakeholders such as academic and research institutions, civil society, and private sector companies rather than being developed by the government alone.

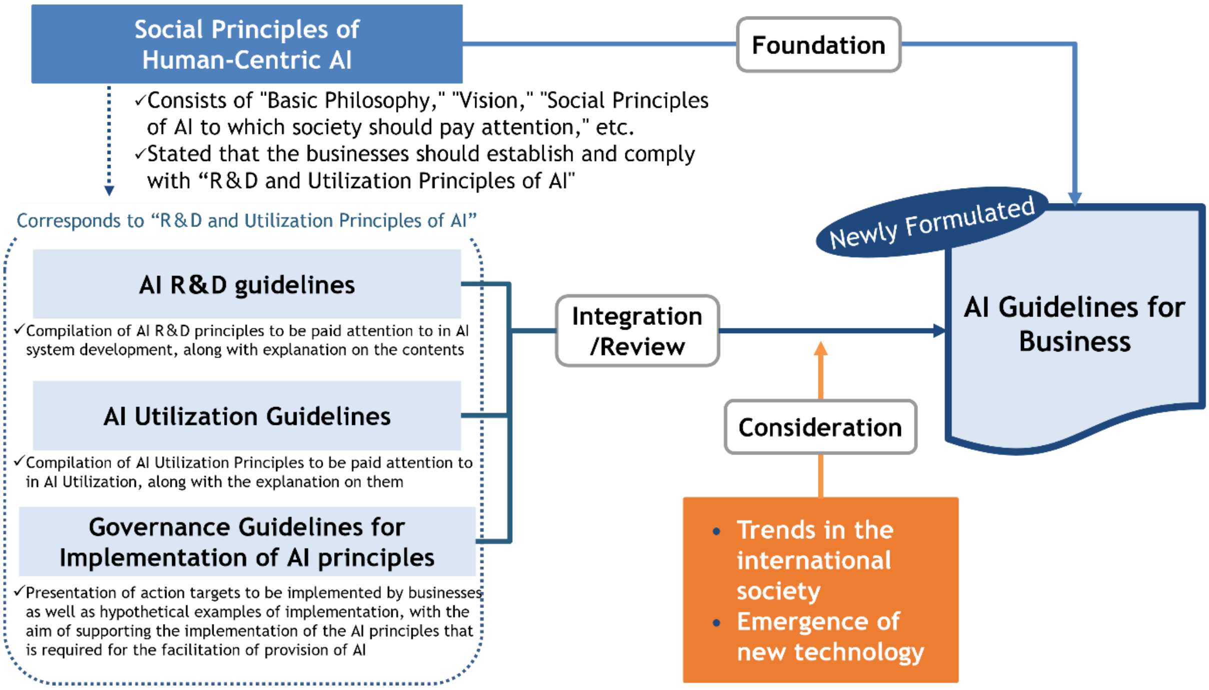
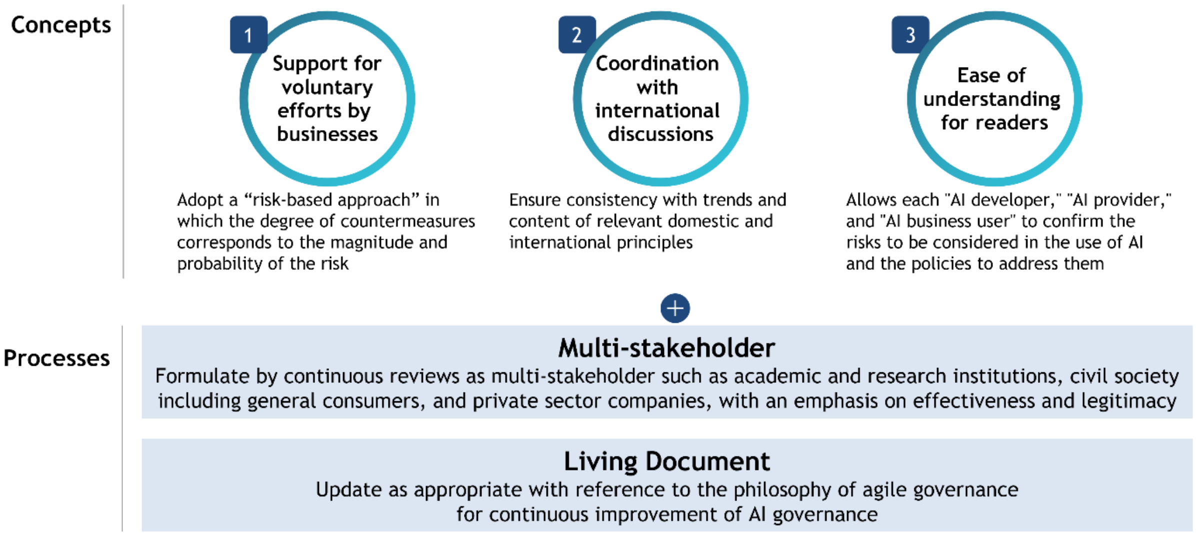

Social Principles of Human-Centric AI
✓Consists of "Basic Philosophy," "Vision," "Social Principles of AI to which society should pay attention," etc.
✓Stated that the businesses should establish and comply with “R&D and Utilization Principles of AI"
Foundation
Corresponds to “R&D and Utilization Principles of AI”
AI R&D guidelines
✓Compilation of AI R&D principles to be paid attention to in AI system development, along with explanation on the contents
AI Utilization Guidelines
✓Compilation of AI Utilization Principles to be paid attention to in AI Utilization, along with the explanation on them
Governance Guidelines for Implementation of AI principles
✓Presentation of action targets to be implemented by businesses as well as hypothetical examples of implementation, with the aim of supporting the implementation of the AI principles that is required for the facilitation of provision of AI
Integration /Review
Newly Formulated
AI Guidelines for Business
Consideration
• Trends in the international society
• Emergence of new technology

Figure 1. Role of this guideline

Depending on the field and form of use, use of AI may pose major risks to society and utilization of AI itself may be hindered due to social conflicts associated by those risks. On the other hand, taking excessive measures may similarly inhibit the use of AI itself or the benefits that can be obtained from the use of AI. Under these circumstances, it is important to understand in advance the magnitude of risks (the magnitude of harm and its probability) that may arise from the form of use in the field, and a “risk-based approach” is also important to adjust the degree of countermeasures to match the magnitude of the risk. This guideline describes the direction of measures for companies based on this “risk-based approach”. The concept of this “risk-based approach” is widely shared among AI advanced countries.

In addition, as trends surrounding AI change rapidly, referencing international discussion, this guideline will be updated as a Living Document with the involvement of multi-stakeholders, under the philosophy of agile governance, in order to continuously improve AI governance (see “Figure 2. Basic concept of this guideline”). In addition, we will consider a specific system for updating the information in the future.



Concepts

1 Support for voluntary efforts by businesses

Adopt a “risk-based approach” in which the degree of countermeasures corresponds to the magnitude and probability of the risk

2 Coordination with international discussions

Ensure consistency with trends and content of relevant domestic and international principles

3 Ease of understanding for readers

Allows each "AI developer," "AI provider," and "AI business user" to confirm the risks to be considered in the use of AI and the policies to address them

Processes

Multi-stakeholder

Formulate by continuous reviews as multi-stakeholder such as academic and research institutions, civil society including general consumers, and private sector companies, with an emphasis on effectiveness and legitimacy

Living Document

Update as appropriate with reference to the philosophy of agile governance for continuous improvement of AI governance

Figure 2. Basic concept of this guideline

This guideline indicates the basic concept of the efforts necessary for the development, provision, and business use of AI. Therefore, in the actual development, provision, and business use of AI, it is important that all AI business actors voluntarily promote specific initiatives while using this guideline as a reference. At the same time, all AI business actors should be aware of the magnitude of the impact AI will have on society and that they have responsibilities to utilize AI to improve human society. It is important to keep in mind that if such efforts are judged inappropriate or insufficient by society, there is a risk of opportunity loss in one's own business and a situation in which maintaining business value becomes difficult. By paying attention to these points, it will be possible to maximize the benefits of AI in one's business, strengthen competitiveness, and maintain and improve business value. It should also be noted that this document contains information that can be helpful for people other than AI business actors and related parties, such as those belonging to educational and research institutions and general consumers (including minors), as well as information regarding risks when utilizing AI. Therefore, it can be said to be useful.

This guideline is intended for all business actors (including public institutions such as national and local governments) who utilize AI in various business. On the other hand, those who use AI for purposes other than business, or those who do not directly use AI for business but receive benefits from, or those who in some cases suffer losses (hereinafter collectively referred to as “Non-business users”) are not included in the scope of this guideline. Additionally, training data will become essential to use AI. While there are specific corporations and individuals (hereinafter referred to as “data providers”) that provide such data, it is also possible to obtain such data from public information such as the internet. When utilizing AI, the actor receiving or acquiring the data, rather than the data provider, is responsible for handling the data, so data providers are not included in the scope of this guideline. Similarly, entities that create data such as public information on the Internet, etc. are not those who utilize AI and are not expected to promote initiatives in accordance with this guideline, so they are not included in the scope of this guideline. Based on the above, the targets of this guideline are broadly divided into three categories: “AI developers,” “AI providers,” and “AI business users”, who are responsible for AI business, and each is defined as follows. These actors are assumed to be business organizations (or departments of them), and depending on how AI is used, certain AI business actor may be an AI developer, AI provider, and AI business user at the same time (see Figure 3. “Responses of Subjects in the Flow of AI Utilization”). [1]

[1] Generative AI is also included in the scope of development, provision, and use. If the AI provider or AI business user is a public organization such as a government or local government, different thinking may be needed than in the case of a private business.

- **AI Developer**
  Business actor that develop AI systems (including business actor that research and develop AI)
- Responsible for building AI models and systems including system infrastructure of the AI through the development of AI models and algorithms, data collection (including purchasing), pre-processing, AI model training, and verification.

- **AI Provider**
  Business actor that provide AI systems to AI Business Users, or in some cases to Non-business users, as services that incorporate AI systems into applications, products, or existing systems or business processes.
  Responsible for AI system verification, implementation for connection to other systems, provision of AI systems and services, and operational support for AI business users of the AI system for normal operation, and operation of the AI service itself. When providing AI services, communication with various stakeholders may be required.

- **AI Business User**
  Business actor that use AI systems or AI services in their business.
  The role is to use AI systems or services regarding to appropriate usage intended by the AI provider and share information with them about environmental changes and to continue normal operation with provided AI system as necessary. In addition, if the use of AI is likely to have some impact on non-business users[2], AI Business User should avoid unintended disadvantages caused by AI and strive to maximize the benefits of AI.

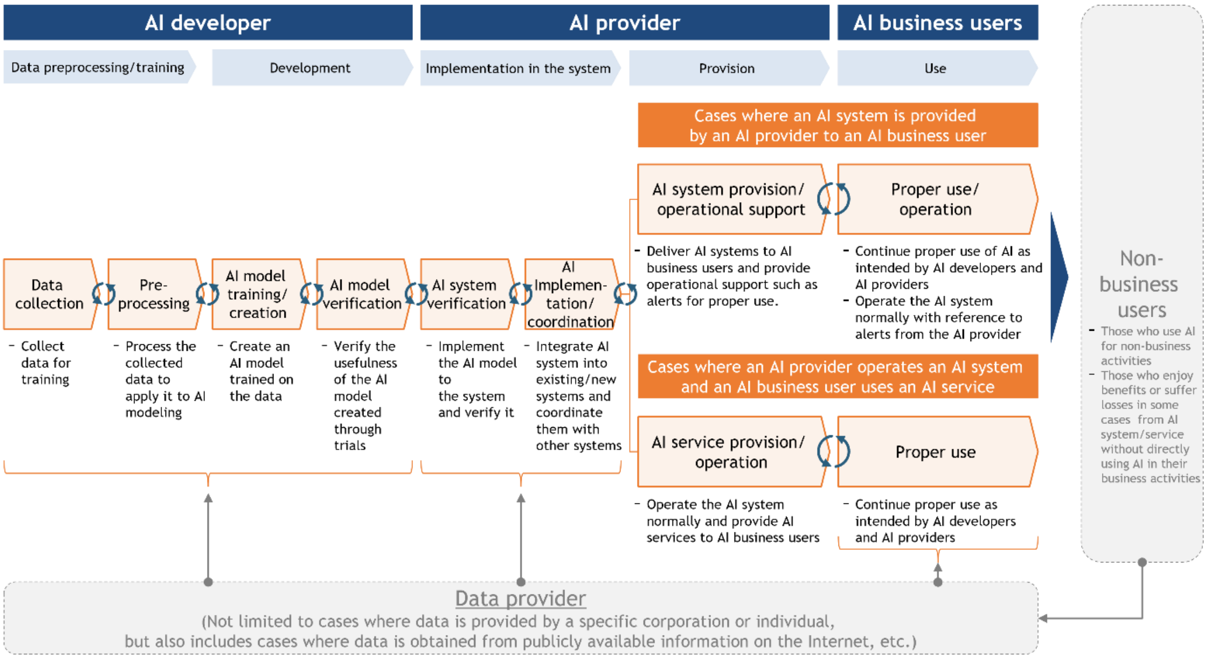

AI developer
AI provider
AI business users
Data preprocessing/training
Development
Implementation in the system
Provision
Use
Cases where an AI system is provided by an AI provider to an AI business user
AI system provision/ operational support
Proper use/ operation
- Deliver AI systems to AI business users and provide operational support such as alerts for proper use.
- Continue proper use of AI as intended by AI developers and AI providers
- Operate the AI system normally with reference to alerts from the AI provider
Data collection
Pre-processing
AI model training/ creation
AI model verification
AI system verification
AI Implementation/ coordination
- Collect data for training
- Process the collected data to apply it to AI modeling
- Create an AI model trained on the data
- Verify the usefulness of the AI model created through trials
- Implement the AI model to the system and verify it
- Integrate AI system into existing/new systems and coordinate them with other systems
Cases where an AI provider operates an AI system and an AI business user uses an AI service
AI service provision/ operation
Proper use
- Operate the AI system normally and provide AI services to AI business users
- Continue proper use as intended by AI developers and AI providers
Non-business users
- Those who use AI for non-business activities
- Those who enjoy benefits or suffer losses in some cases from AI system/service without directly using AI in their business activities
Data provider
(Not limited to cases where data is provided by a specific corporation or individual, but also includes cases where data is obtained from publicly available information on the Internet, etc.)

Figure 3. Responses of actors in the general flow of AI utilization

From the standpoints of "AI developers," "AI providers," and "AI business users", we consider "what kind of society do we aim for ("Basic Philosophy" = why) while considering the expectations of stakeholders". It is important to clarify "what kind of initiatives should be taken regarding AI (guidelines = what)", and in order to realize the guidelines, it is considered to be useful to discuss and decide "what specific approach should be taken (implementation =

[2] Non-business users need to be aware that if they do not follow the instructions and warnings of AI business users, they may suffer some damage.

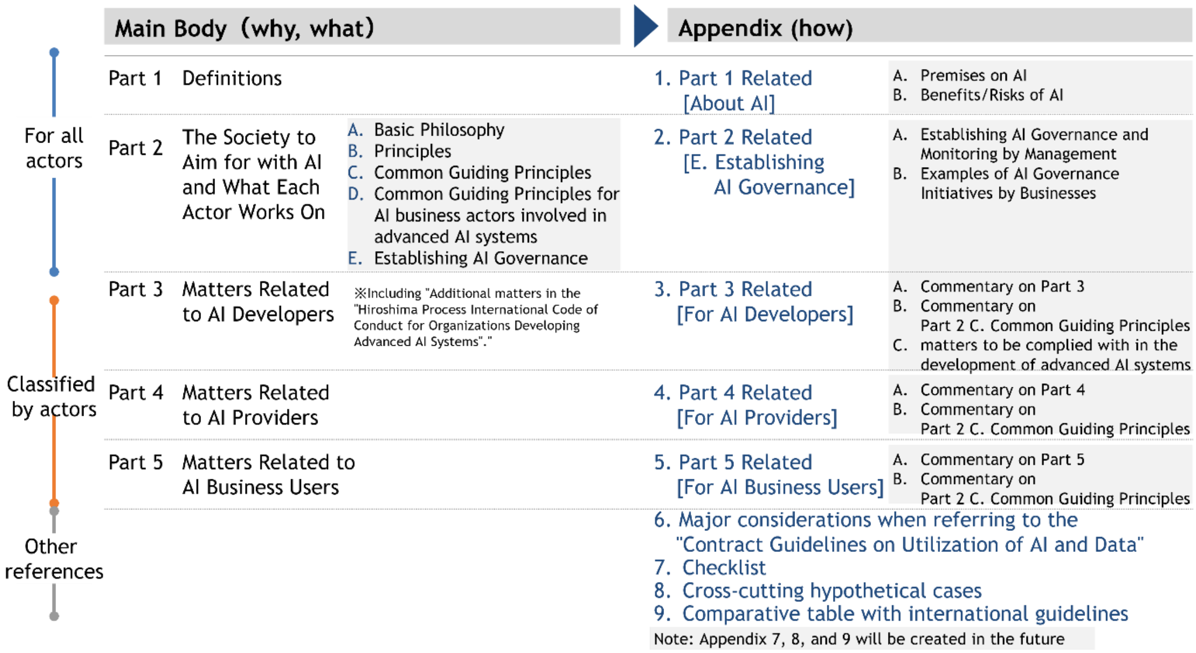
how)" for the safe and secure use of AI. Actual AI systems and services have a variety of use cases depending on the purpose, utilization technology, data, usage environment, etc., and it is important for AI developers, AI providers, and AI business users to collaborate and to take into account changes in the external environment such as technological developments, and to consider the optimal approach. For ease of reading, this guideline deals with the "Basic Philosophy" and "Guiding Principles" in the main part, and "implementation" in the appendix (attached material).

The structure of the main part of this guideline, which deals with the "Basic Philosophy" and "Guiding Principles", is described below.

- **Part 1**
  In order to help you understand the contents of this guideline, we mainly describe "definitions of terms".

- **Part 2**
  Describe the society that should be aimed at through the use of AI, the "Basic Philosophy" (why) and Principles to realize it, and the guiding Principles common to each AI business actor (what). While seeking benefits from the use of AI, we will also touch on the creation of governance that is necessary to put "Common Guiding Principles" into practice, considering the possibility that AI may pose risks to society. Part 2 explains the content that forms the basis of Part 3 and subsequent parts, so it is important for all business actors that utilize AI to review and understand the content.

- **Part 3 ~5**
  Regarding the three business actors responsible for business that utilize AI, we will describe points to keep in mind for each AI business actor that cannot be covered in Part 2. It is important for business actors that utilize AI to understand matters related to themselves, and at the same time, because there are many matters related to adjacent entities, it is important to understand matters related to other entities as well ("Figure 4. Structure of this guideline").



	Main Body（why, what)			Appendix (how)	
For all actors	Part 1	Definitions		1. Part 1 Related [About AI]	A. Premises on AI B. Benefits/Risks of AI
	Part 2	The Society to Aim for with AI and What Each Actor Works On	A. Basic Philosophy B. Principles C. Common Guiding Principles D. Common Guiding Principles for AI business actors involved in advanced AI systems E. Establishing AI Governance	2. Part 2 Related [E. Establishing AI Governance]	A. Establishing AI Governance and Monitoring by Management B. Examples of AI Governance Initiatives by Businesses
Classified by actors	Part 3	Matters Related to AI Developers	※Including "Additional matters in the "Hiroshima Process International Code of Conduct for Organizations Developing Advanced AI Systems"."	3. Part 3 Related [For AI Developers]	A. Commentary on Part 3 B. Commentary on Part 2 C. Common Guiding Principles C. matters to be complied with in the development of advanced AI systems
	Part 4	Matters Related to AI Providers		4. Part 4 Related [For AI Providers]	A. Commentary on Part 4 B. Commentary on Part 2 C. Common Guiding Principles
	Part 5	Matters Related to AI Business Users		5. Part 5 Related [For AI Business Users]	A. Commentary on Part 5 B. Commentary on Part 2 C. Common Guiding Principles
Other references				6. Major considerations when referring to the "Contract Guidelines on Utilization of AI and Data" 7. Checklist 8. Cross-cutting hypothetical cases 9. Comparative table with international guidelines Note: Appendix 7, 8, and 9 will be created in the future	

Figure 4. Structure of this guideline

"AI developers", "AI providers" and "AI business users" will be able to understand the basic concepts of risks when utilizing AI and countermeasure policies, by checking the relevant parts and attachments (attached materials) starting from Part 3, in addition to Part1 and 2. In particular, for AI business actors that have not yet decided on specific initiatives, the examples in the appendix will be helpful, so it is important to check the relevant sections in the attached appendix. Additionally, for business executive officers[3] including managers, it is important to consider and implement risk countermeasures and promote the safe and secure use of AI considering the "Basic Philosophy" (why) and guiding Principles (what) in this guideline when utilizing AI in conjunction with business strategy in order to fulfill their duties.

As the environment surrounding AI continues to evolve rapidly on a global scale, it is important for business actors that utilize AI to pay attention to international trends. In light of this current situation, Japan also leads formulating an international common understanding and guiding principles regarding AI through the Hiroshima AI Process[4], and organized a comprehensive policy framework for the Hiroshima AI Process in December 2023. This guideline is intended to contribute to the same process and has considered in light of international discussions including the process. On the other hand, since ideas and laws regarding AI differ between countries and regions, business actors that engage in cross-border activities should respond in accordance with local laws and stakeholders' expectations. In particular, some countries/regions have taken measures to ensure the effectiveness of governance, such as considering a framework for safety evaluation before introducing advanced AI systems into the market[5]. So, paying attention to this is very important.

[3] Business executive officers also includes the person responsible for business execution of public institutions such as the government and local governments.

[4] In response to the results of the G7 Hiroshima Summit in May 2023, we launched the "Hiroshima AI Process" to consider international rules regarding generative AI. After that, in December 2023, based on the "G7 Leaders' Statement on AI" issued after the "G7 Digital and Technology Ministerial Meeting" in September 2023 and the "Multi-stakeholder High-Level Meeting" at the Kyoto IGF in October, "The G7 Digital and Technology Ministerial Meeting" was held, and the results of the same year, we put together the "Hiroshima AI Process Comprehensive Policy Framework" and the "Hiroshima AI Process Promotion Work Plan".

[5] In November 2023, the UK announced plans to establish an AI Safety Institute that will develop and implement evaluations of advanced AI systems, and the US announced plans to establish the US AI Safety Institute within the National Institute of Standards and Technology (NIST), which will implement AI risk management frameworks and evaluate red teaming. In Japan as well, in cooperation with these overseas organizations, "AI Safety Institute" is scheduled to be established within the "Information-technology Promotion Agency, Japan" for developing safety evaluation standards and test tools, etc. in around January 2024.

Part 1 Definitions

AI stands for Artificial Intelligence and was first used at the Dartmouth Conference in 1956. AI does not yet have an established definition, but as the terms “artificial” and “intelligence” indicate, it refers to computer programs that operate in a manner like human thought processes, and systems that can make intelligent decisions on a computer. Expert systems, which do not perform machine learning (ML) and perform knowledge-based reasoning by inputting a large amount of expert knowledge, were originally considered as a type of AI. However, since the 2000s, with the advent of deep learning, it has become possible to perform “image recognition”, “natural language processing (translation, etc.)”, and “speech recognition” using machine learning. Then, AI has come to refer to systems that can make proposals and decisions in some specific fields. In addition, since 2021, with the rise of foundation models[6], the development of Artificial General Intelligence, rather than applied AI only in a specific field, is developing. As a result, “generative AI” that goes beyond “predictions”, “suggestions”, and “decisions” and generates completely new images and texts has become popular and is attracting attention. In this way, even though we collectively refer to “AI,” there are a wide variety of types, and it is difficult even for experts to predict the future state of AI technology.

Considering this situation, related terms in this guideline are defined as follows.

Related terms

- **AI**
  There is currently no established definition (Decision by the Integrated Innovation Strategy Promotion Council on “Human-centered AI social principles” (March 29, 2019)), and it is difficult to strictly define the scope of artificial intelligence in a broad sense. AI in this guideline is an abstract concept that includes the “AI system” itself and machine learning software and programs.
  (For reference, JIS X22989 defined AI as follows based on ISO/IEC22989)
  <discipline> research and development of mechanisms and applications of AI systems
  Note 1: Research and development can take place across any number of fields such as computer science, data science, humanities, mathematics and natural sciences.

- **AI system**
  A system that includes software as an element that has the ability to operate and learn with various levels of autonomy through the process of utilization (machines, robots, cloud systems, etc.).
  (For reference, JIS X22989 defined as follows based on ISO/IEC22989)
  Note 1. engineered system that generates outputs such as content, forecasts, recommendations or decisions for a given set of human-defined objectives
  Note 2. AI systems are designed to operate with varying levels of automation
  (For reference, it is defined as follows in the OECD AI Principles overview)
  An AI system is a machine-based system that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments. Different AI systems vary in their levels of autonomy and adaptiveness after deployment.

- **Advanced AI systems**
  Refers to the most advanced AI systems, including the most advanced foundation models

[6] Fundamental models, represented by large language models, are the core technology base that generates individual models that support various services. It has characteristics different from general AI in terms of the development of models adapted to a wide range of downstream tasks derived from the basic model and the knowledge gained from the development process itself.

and generative AI systems.
(Quoted definition from the Hiroshima AI process)

- **AI model（ML model）**
  A model included in AI systems that are obtained through machine learning using training data, and generates prediction according to input data.
  (For reference, JIS X22989 defined as follows based on ISO/IEC22989)
  Mathematical structure that generates inferences or predictions based on input data or information.
  Example: If a univariate linear function $y=\theta_0 +\theta_1 x$ is trained using linear regression, the example model is $y=3+7x$
  Note 1. A machine learning model is the result of training based on a machine learning algorithm.

- **AI services**
  Refers to services using “AI systems”. It refers to the overall provision of value to users, and the provision and operation of AI services is not limited to the configuration technology of AI systems, but also involves non-technical approaches such as human monitoring and appropriate communication with stakeholders.

- **Generative AI**
  A general term for AI based on AI models that can generate sentences, images, programs, etc.

- **AI governance**
  Design and operation of technical, organizational, and social systems by stakeholders to manage the risks arising from the use of AI at an acceptable level to them and to maximize the positive impact (benefits) resulting from it.

Part 2 The society to aim for with AI and what each AI business actor works on

In Part 2, we will first describe “A. Basic Philosophy”' for the society we aim to achieve through AI. Furthermore, in order to realize this, we will describe the “B. Principles” that each AI business actor works on, as well as the “C. Common Guiding Principles” that can be derived from these principles. In addition, “D. Common Guiding Principles for AI Business actors involved in advanced AI systems” is described for businesses involved in advanced AI systems should comply with. In addition, we also describe the “E. Establishing AI governance”, which is important for putting these “common guidelines” into practice and using AI safely and securely.

A. Basic philosophy

As mentioned in the “Introduction”, the “Principles of Human-centric AI Society” formulated by Japan in March 2019 expects AI to contribute to realize Society 5.0. It also states that it is important to utilize AI as public goods for humanity and to lead to global sustainability through qualitative changes in the way society exists and through true innovation. It states that the following three values should be respected as “Basic Philosophy” and “should be built a society that pursues realization of those.”

① **A society that has respect for human dignity (Dignity)**
A society where human dignity is respected and can live a rich life both materially and spiritually need to be built not by using AI to make humans overly dependent on or to control human behavior with too much effort to pursue efficiency and convenience, but by using AI as a tool and it enables humans to further demonstrate their various abilities, display greater creativity, and engage in more rewarding work.

② **A society where people with diverse backgrounds can pursue diverse happiness (Diversity and Inclusion)**
A society in which people with diverse backgrounds, values, and ways of thinking can pursue a variety of happiness and create new value by flexibly incorporating them is an ideal and a major challenge today. The powerful technology of AI can be an important tool that brings us closer to this ideal. We need to transform the society in this way through the appropriate development and deployment of AI.

③ **A sustainable society (Sustainability)**
We need to use AI to create businesses and solutions one after another, eliminate social disparities, and develop a sustainable society that can respond to global environmental issues and climate change. Japan, as a nation built on science and technology, has responsibilities to strengthen its scientific and technological accumulation through AI and contribute to creating such a society.

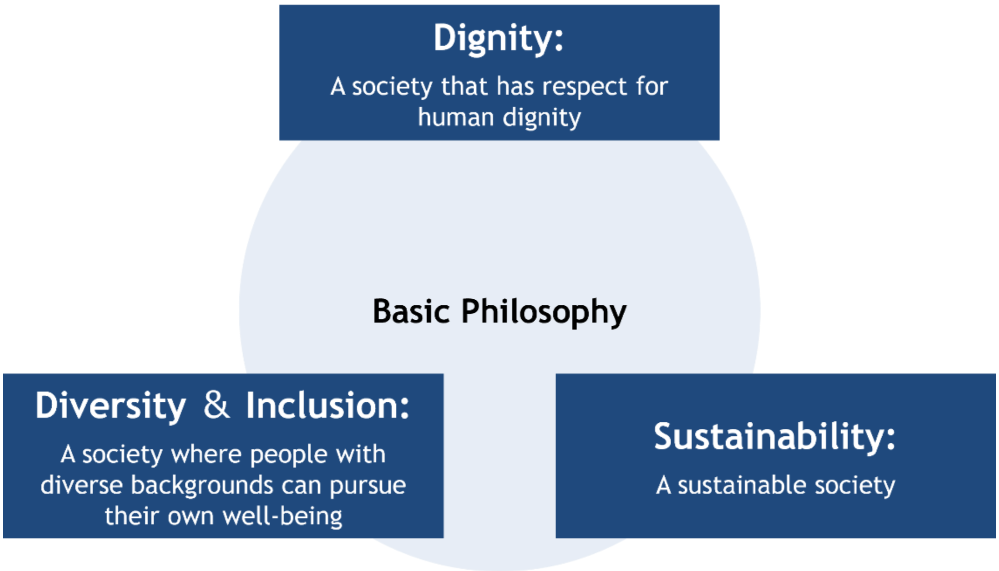

Dignity:
A society that has respect for human dignity
Basic Philosophy
Diversity & Inclusion:
A society where people with diverse backgrounds can pursue their own well-being
Sustainability:
A sustainable society

Figure 5. Basic Philosophy

This basic idea itself has not changed even with the remarkable development of technology, and it continues to be the philosophy we should aim for. Therefore, as AI develops, these “Basic Philosophy” should be respected as the direction Japan and multilateral frameworks should take.

B. Principles

To realize the “Basic Philosophy”, it is important for each AI business actor to proceed with its efforts in accordance with these “Principles”. In order to do so, we have organized the “Principles” that each AI business actor should keep in mind; the matters each AI business actor addresses and the matters that are expected to be addressed in cooperation with the society. These “Principles” are based on the “Principles of Human-centric AI Society” and have been restructured based on overseas principles such as the OECD's AI Principles.

Matters each AI business actor works on

It is important for each AI business actor to realize the purpose of AI promoting the development, provision, and use of AI systems and services based on human-centric thinking derived from the “Basic Philosophy” and create value in business and solve social issues while protecting human dignity. Therefore, it is important for each AI business actor to ensure values such as safety and fairness to reduce the societal risks associated with the use of AI. It is also important to protect privacy, including preventing inappropriate use of personal information, and to ensure security against risks such as reduced availability due to AI system vulnerabilities and external attacks. In order to realize the above, it is important for each AI business actor to improve transparency and fulfill accountability by providing appropriate information to stakeholders[7] while ensuring the verifiability of the system.

In addition, in consideration of the possibility that the role of each AI business actor will change due to changes in the value chain according to the diversification of AI architecture, it is important to strive for cooperate among each AI business actor and improve the quality of AI throughout the value chain, and continue to discuss with multi-stakeholders.

[7] All entities that may be directly or indirectly affected by the use of AI, including third parties other than AI developers, AI providers, AI business users and Non-business users. (the same applies hereinafter)

It is important that these efforts be carried out voluntarily, taking into consideration the resource constraints of each AI business actor, considering the characteristics, uses, purposes, and social context of the AI systems and services that each AI business actor develops, provides, and uses. By taking such measures, each AI business actor is expected to minimize the risks of AI while receiving maximum benefits through the development, provision, and use of AI systems and services.

Matters that are expected to be worked in cooperation with society

In order to further increase the benefits of AI to society and realize the "Basic Philosophy" that we should aim for, in addition to the efforts of each AI business actor, we are expected to collaborate actively with society (including the government, local governments, and communities). For this reason, each AI business actor is expected to work together with society to avoid social division and provide opportunities for education and literacy so that the benefits of AI can be distributed to all people. In addition, it is expected to contribute to ensuring fair competition and promoting innovation so that new business services will be created, sustainable economic growth will be maintained, and solutions to social issues will be presented.

C. Common Guiding Principles

In undertaking these initiatives, each actor should develops, provides, and uses AI systems and services with respect for the rule of law, human rights, democracy, diversity, and a fair and just society in light of the philosophy of "1) Human-Centric", and complies with the Constitution, relevant laws and regulations including the Act on the Protection of Personal Information, the Intellectual Property Protection Act, and other existing laws and regulations in individual fields related to AI. And it is important to pay attention to the status of consideration of international guiding principles. [8,9]

Specifically, the tasks that each AI business actor should work on throughout the value chain are summarized as follows.

1) human-centric

In the development, provision, and use of AI systems and services, each AI business actor must not at least violate the human rights guaranteed by the Constitution or internationally recognized, as the basis for deriving all matters to be addressed, including the matters described below. Besides that, it is important to act in a way that allows AI to expand people's abilities and enable diverse people to pursue diverse forms of well-being.

① Human dignity and individual autonomy

✧ Respect human dignity and individual autonomy, considering the social context in which AI is used.

[8] It is necessary to comply with each applicable law depending on the geographical development status of the business, the location of the AI provider or AI business user using the developed AI model, the location of the training server, etc. When complying with the domestic laws of Japan, we will handle the data in accordance with the laws and regulations applicable to personal information, intellectual property rights, confidential information, etc., depending on the type of data. In addition, when handling data, it should be noted that even if it is not stipulated by law, there are cases where use is prohibited due to contractual relationships between stakeholders.

[9] Regarding the relationship with intellectual property-related laws and regulations, discussions are underway at the Cabinet Office's Intellectual Property Strategy Promotion Secretariat, the Agency for Cultural Affairs, and other organizations, and the status of future deliberations should be kept in mind. In particular, with regard to the relationship between AI and copyright, the Legal System Subcommittee of the Copyright Subcommittee of the Cultural Affairs Council is working to clarify the scope of use of copyrighted materials without permission from the copyright holder in the development and learning of AI. We are currently organizing our ideas regarding measures to reduce the risk of copyright infringement caused by products, and it is important for each AI business actor to take this idea into account.

- ✧ In particular, when linking AI with the human brain and body, it is important to take into account information on peripheral technologies and refer to bioethics' discussions in other countries and research institutions.
- ✧ When conducting profiling using AI in fields that may have a significant impact on the rights and benefits of individuals, respect the dignity of the individual, and understand and utilize the limitations of AI predictions, recommendations, or judgments while maintaining output accuracy. And carefully consider the potential disadvantages, and consider whether or not the purpose is inappropriate. Do not use it for inappropriate purposes.

② Attention to decision-making and emotional manipulation by AI

- ✧ Do not develop, provide, or use AI systems or services that aim at or are based on the premise of improperly manipulating human decision-making, cognition, or human emotions.
- ✧ In developing, providing, and using AI systems and services, take necessary measures against the risks of over-reliance on AI, such as automation bias[10].
- ✧ Pay attention to the use of AI that promotes inclinations in information and values, as exemplified by filter bubbles[11] and that inadvertently limits the choices of humans including AI business users.
- ✧ In particular, carefully treat the output of AI in cases where it may be related to procedures that have a significant impact on society, such as elections and community decision-making.

③ Measures against disinformation, etc.

- ✧ Recognize that disinformation, misinformation, and biased information AI generated can destabilize and threaten society and take necessary measures since generative AI allows anyone to create information that appears to be true and fair.

④ Ensuring diversity and inclusion

- ✧ In addition to ensuring fairness, pay attention to making it easier for socially vulnerable people to utilize AI so that more people can enjoy the benefits of AI, without creating the so-called "informationally weak" or "technologically weak".
  - ● Universal design, ensuring accessibility, education and follow-up for related stakeholders[12], etc.

⑤ User support

- ✧ Provide information on the functions of AI systems and services and related technologies to a reasonable extent and ensure that functions which provide timely and appropriate selection opportunities are available.
  - ● Set defaults, present easy-to-understand options, provide feedback, alert in emergencies, deal with errors, etc.

⑥ Ensuring sustainability

[10] Indicates the phenomenon of excessive trust and dependence on automated systems and technology in human judgment and decision-making.

[11] A "filter bubble" is an algorithm that analyzes and learns from the search history or click history of individual internet usees, so that information that the individual wants to see is displayed preferentially, regardless of whether the individual wants to see it or not. It refers to an information environment in which people are isolated from information that does not match their viewpoint, and are isolated in a "bubble" of their own ways of thinking and values. In addition to "filter bubble", "echo chambers" can also be cited as phenomena that are said to be the result of the interaction between inherent human tendencies and the characteristics of internet media. While there are risks, AI also has the benefit of providing personalized and targeted responses to AI business users and non-business users, allowing them to make recommendations in a useful way.

[12] Entities directly or indirectly involved in the utilization of AI, including AI developers, AI providers, AI business users, and non-business users.

- ✧ In developing, providing, and using AI systems and services, consider the impact on the global environment throughout their lifecycle.

Based on all of these assumptions, each AI business actor is expected to improve the performance (usefullness) of AI to the extent possible, provide benefits and enrichment to people, and realize happiness.

2) Safety

Through the development, provision, and use of AI systems and services, each AI business actor should not harm the lives, bodies, and property of stakeholders involved with AI. In addition, it is important to avoid harming minds of them as well as the environment.

① Consideration for human life, body, property and mind as well as the environment

- ✧ The AI works satisfactorily for the requirements, including the accuracy of its output (reliability)
- ✧ Maintaining performance levels under a variety of conditions and avoiding significantly incorrect decisions regarding unrelated events (robustness)
- ✧ In light of the nature and use of the AI, including the severity of rights infringements that may occur due to the use of AI and unintended AI operations, and the possibility of infringement, human beings should ensure controllability and take measures including objective monitoring and countermeasures as necessary.
- ✧ Conduct appropriate risk analysis and take countermeasures (avoid, reduce, transfer, accept) against risks.
- ✧ If there is a possibility of harm to human life, body, property and mind as well as the environment, organize in advance the measures that should be taken and provide related information to stakeholders involved with AI.
  - ● Specify measures to be taken by non-business users and usage rules.
- ✧ Consider measures to be taken in case of a situation that compromises the safety of AI systems and services and prepare to take prompt action if such situation occurs.

② Proper use

- ✧ Develop, provide, and use AI systems and services in order to avoid harm caused by provision and use that deviate from the original purpose within the scope of the AI business actor's control.

③ Proper training[13]

- ✧ Considering the characteristics and uses of AI systems and services, ensure the accuracy of data used for training, etc., and if necessary, currency (that the data is appropriate), etc.
- ✧ Ensure transparency of data used for training appropriately, comply with legal frameworks, update AI models, etc. to a reasonable extent.

3) Fairness

In the development, provision, and use of AI systems and services, it is important for AI business actors to make efforts to do away with prejudice and discrimination not to impose unfair or harmful actions against specific individuals or groups based on diverse backgrounds such as race, gender, nationality, age, political opinion, religion, etc. In addition, it is important to recognize that some biases cannot be avoided, and evaluate

[13] When performing fine-tuning or relearning, it is important for AI providers and AI business users to strive to ensure safety in the same way as AI developers.

whether these unavoidable biases are acceptable from the perspective of respecting human rights and diverse cultures. It is important to develop, provide, and use AI systems and services based on this understanding.

① Consideration of bias included in each component technology of AI model

- ✧ There are a wide variety of factors that can cause inappropriate bias, so we have to identify points of bias that can lead to fairness issues in each technical element (training data, training process of the model, prompts input by AI business users[14], information referenced during AI model inference, linked external services, etc.) or in user behavior.
- ✧ Also consider the possibility of potential bias depending on the characteristics and usage of AI systems and services.

② Intervention of human judgment

- ✧ In order to ensure that the output results of AI do not lack fairness, consider not only having AI make decisions on its own, but also using human judgment as an intervention.
- ✧ Implement processes to analyze and address the objectives, constraints, requirements, and decisions of AI systems and services for bias clearly and transparently.
- ✧ Be mindful of unconscious and potential biases and make policy decisions after dialogue with stakeholders from diverse backgrounds, cultures, and fields.

4) Privacy protection

It is important for each AI business actors to respect and protect privacy in the development, provision, and use of AI systems and services, depending on its importance. At that time, relevant laws and regulations should be complied.

① Privacy protection in general AI systems and services

- ✧ Take appropriate measures to respect and protect stakeholder's privacy depending on its importance and take into account the social context and people's reasonable expectations by complying regarding related laws and regulations such as the Personal Information Protection Act, and formulating and publishing each AI business actor's privacy policy, etc.
- ✧ Consider measures to protect privacy, taking into account the following:
  - ● Ensuring responses based on the Personal Information Protection Act
  - ● Reference to international personal data protection principles and standards[15]

5) Ensuring security

When developing, providing, and using AI systems and services, it is important for each AI business actor to ensure security to prevent unintentional changes or suspension of AI behavior due to dishonest manipulation.

[14] In generative AI such as large language models, AI business users can use a learning method called in-context learning to identify specific parameters in response to AI business user input (called prompts) without updating learned parameters. It is possible to perform learning for the following tasks.

[15] OECD, Recommendation of the Council concerning Guidelines Governing the Protection of Privacy and Transborder Flows of Personal Data, OECD/LEGAL/0188 and ISO/IEC 29100:2011 Information technology Security techniques Privacy framework, etc. Be expected. In addition, the Global Cross-Border Privacy Rules (CBPR) Forum has been established to promote the smooth cross-border transfer of personal data over a wider range of countries and the interoperability of rules in each country, and Japan will also implement the same in April 2022. We are participating in the Global CBPR Framework and publishing the Global CBPR Framework. Regarding generative AI, the G7 Data Protection and Privacy Agencies Roundtable Meeting's "Statement on Generative AI" (June 2023) and the GPA (Global Privacy Assembly)'s "Resolution on Generative AI Systems" (October 2023) are also available. reference.

① Security measures affecting AI systems and services[16]

- ✧ To maintain the confidentiality, integrity, and availability of AI systems and services, and to always ensure the safe and secure utilization of AI, take reasonable measures in light of the current technological level.
- ✧ Understand the characteristics of AI systems and services and consider whether the connections between systems necessary for normal operation are properly established.
- ✧ Recognize that it is not possible to completely eliminate vulnerabilities in AI systems and services, given the possibility that unintended decisions will be made by non-business users and prediction targets by mixing trifling information into the data.

② Attention to the latest trends

- ✧ New methods of external attacks on AI systems and services are emerging every day, so check the points to keep in mind when dealing with these risks.

6) Transparency[17]

In the development, provision, and use of AI systems and services, it is important for AI business actor to provide appropriate information to stakeholders within a reasonable scope considering the social context in which the AI systems and services are used and ensuring the verifiability of the AI systems and services as necessary and technically possible.

① Ensuring verifiability

- ✧ In order to ensure the verifiability of AI judgments, record and save logs of AI training process, inference process, judgment basis, such as Input/output that occurs during the development process of AI systems and services etc. within a reasonable range in light of the amount and content of data.
- ✧ When recording and storing logs, consider recording methods, frequency, etc., based on the importance of investigating the cause of the accident, considering measures to prevent recurrence, and proving the requirements for liability for damages, in light of the characteristics and usage of the technology being used.

② Providing information to related stakeholders

- ✧ In light of the relationship with AI or the nature and purpose of AI, provide and explain information compiled on the following according to each person's knowledge and abilities.
  - ● AI systems and services in general
    - ➢ The fact that using AI
    - ➢ Data collection and annotation methods
    - ➢ Training and evaluation methods
    - ➢ Information about the underlying AI model
    - ➢ Capabilities, limitations, and appropriate/inappropriate usage methods for AI systems and services
    - ➢ Related laws and regulations applicable in the countries/regions where AI systems and services are provided and where AI business users are located.

[16] For detailed methods, see also the National Cyber Security Center (NCSC) "Guidelines for secure AI system development" (November 2023).https://www8.cao.go.jp/cstp/stmain/20231128ai.html

[17] Transparency has various definitions in different countries. For example, NIST,in Artificial Intelligence Risk Management Framework（January 2023）, divide into three categories: transparency (being able to answer what happened in the system), explainability (being able to answer how decisions were made in the system), and interpretability (being able to answer the meaning and context of why decisions were made) (Be able to answer). European Commission, in ETHICS GUIDELINES FOR TRUSTWORTHY AI （April 2019）, treats traceability, explainability, and communication. Furthermore, international standards (ISO/IEC JTC1/SC42) define transparency (provision of appropriate information to relevant parties). In this document, matters related to information disclosure are broadly referred to as "transparency."

- ✧ Encourage active involvement through dialogue with diverse stakeholders and collect various opinions regarding social impact and safety.
- ✧ In addition, demonstrate to related stakeholders, including each AI business actor, the advantages of providing and using AI systems and services in accordance with the actual situation.

③ Reasonable and honest response

- ✧ "② Providing information to relevant stakeholders" above is Implemented to the extent permitted and does not mean the disclosure of algorithms or source code, but respects privacy and trade secrets, and is based on social rationality in light of the characteristics and uses of the technology to be adopted. Implemented to the extent permitted.
- ✧ When using publicly available technology, comply with the respective regulations.
- ✧ Consider the social impact when making the developed AI system open source

④ Improving explainability and interpretability to related stakeholders

- ✧ Analyze and understand what kind of explanation is required and take necessary measures for the purpose of gaining a sense of satisfaction and security feeling from related stakeholders, as well as presenting evidence of AI operations.
  - ● AI provider: Share with the AI developer what kind of explanation is required.
  - ● AI business users: Share with AI developers/AI providers what kind of explanation is required.

7) Accountability[18]

In the development, provision, and use of AI systems and services, it is important for each AI business actor to fulfill accountability to its stakeholders to a reasonable extent with regard to ensuring traceability and compliance with "Common Guiding Principles".

① Improving traceability

- ✧ Ensure that data sources and decisions made during the development, provision, and use of AI systems and services can be traced to the technically possible and reasonable extent.

② Explanation of compliance status of "Common Guiding Principles"

- ✧ Regarding the status of compliance with the "Common Guiding Principles", regularly provide and explain information to stakeholders (including suppliers) based on their respective knowledge and abilities, including the following:
  - ● General
    - ➢ Assessment of the existence and extent of risks that impede the implementation of "Common Guiding Principles"
    - ➢ Progress in implementing "Common Guiding Principles"
  - ● "Human-centric" related
    - ➢ Attention to disinformation, diversity and inclusion, user support, and status of ensuring sustainability
  - ● "Safety" related
    - ➢ Known risks and countermeasures related to AI systems and services, and mechanisms to ensure safety
  - ● "Fairness" related

[18] Accountability is sometimes defined as explainability, but in this document, information disclosure is handled with transparency, and accountability refers to the assumption of de facto and legal responsibility for AI and the prerequisites for assuming that responsibility. The concept is related to the maintenance of conditions.

        - ➢ Bias may be included due to each technical element that makes up the AI model (training data, model training process, prompts input by AI business users, information referenced when inferring the AI model, linked external services, etc.)
    - ● "Privacy protection" related
        - ➢ Risks and countermeasures that AI systems and services may infringe on one's own and stakeholders' privacy, as well as measures expected to be taken in the event of privacy intrusion.
    - ● "Ensuring security" related
        - ➢ Standards compliance, when connection between AI systems and services or other systems occurs, necessary to promote such connection, etc.
        - ➢ Risks that may occur when AI systems/services connect with other AI systems/services, etc. via the Internet and countermeasures

③ Clarification of responsible person

- ✧ Establish a person responsible for fulfilling accountability in each AI business actor

④ Distribution of responsibilities between parties

- ✧ Regarding responsibilities among related parties, clarify the location of responsibility through contracts and social promises (voluntary commitments) between each AI business actor, including non-business users.

⑤ Specific responses to stakeholders

- ✧ As necessary, develop and publish policies such as AI governance policies and privacy policies for each AI business actor to manage risks and ensure safety associated with the use of AI systems and services. (share the vision with society and the general public, including social responsibility such as providing information such as social dissemination)
- ✧ If necessary, create opportunities to receive feedback from stakeholders regarding errors in AI output, etc., and conduct objective monitoring.
- ✧ In the event that a situation that harms stakeholder interests arises, formulate a policy on how to respond, steadily implement the policy, and regularly report progress to stakeholders as necessary.

⑥ Documentation

- ✧ Document and store information related to the above and make it available for reference when and where needed.

Specifically, the matters that each AI business actor is expected to address in cooperation with society are summarized as follows.

8) Education/Literacy

Each AI business actor is expected to provide the necessary education so that those involved in AI within each AI business actor can have the knowledge, literacy, and ethical sense to properly understand AI and use it socially. Furthermore, each AI business actor is expected to educate stakeholders, taking into consideration the complexity of AI, its characteristics such as misinformation, and the possibility of intentional misuse. [19]

① Ensuring AI literacy

- ✧ Take necessary measures to ensure that those involved in AI within each AI business actor have a sufficient level of AI literacy in their engagement.

② Education/Reskilling

- ✧ As the use of generative AI expands, it is expected that the division of work between AI and humans will change, so consider education and reskilling to enable new ways of working.
- ✧ Provide educational opportunities that take generational gaps into consideration so that various people can deepen their understanding of the benefits that can be obtained from AI and increase their resilience to risks.

③ Follow-up with stakeholders

- ✧ Safe use by non-business users will improve the overall safety of AI services, so provide necessary follow-up to stakeholders to ensure their education and literacy as necessary.

9) Ensuring fair competition

Each AI business actor is expected to strive to maintain a fair competitive environment surrounding AI so that new business services that utilize AI are created and sustainable economic growth is maintained, and solutions to social issues are presented.

10) Innovation

Each AI business actor is expected to strive to contribute to promoting innovation in society as a whole.

① Promotion of open innovation, etc.

- ✧ Promote internationalization and diversification, industry-academia-government collaboration, and open innovation
- ✧ Consider maintaining an environment that creates the data necessary for AI innovation

② Attention to interconnectivity and interoperability

- ✧ Ensure interconnectivity and interoperability of your own AI systems and services with other AI systems and services
- ✧ Comply with standard specifications, if any

③ Providing appropriate information

- ✧ Provide necessary information to the extent that does not impair own innovation.

In addition to the above, the matters that AI developers, AI providers, and AI business users should pay special attention to are summarized in “Table 1. Important matters for each AI business actor in addition to “Common Guiding Principles””. Places marked with a “-” in the table indicate that each AI business actor is expected to take action based on the matters

[19] The Ministry of Economy, Trade and Industry and Information-technology Promotion Agency has organized the image of human resources in the DX era as a "digital skill standard" as a guideline for individual learning and companies' recruitment and development of human resources (December 2020). In order to further promote corporate DX through the use of generative AI, in August 2020 we compiled the "Thoughts on human resources and skills necessary for promoting DX in the era of generative AI", and we have compiled "Thoughts on the human resources and skills necessary for promoting DX in the era of generative AI", and we have compiled "Thoughts on human resources and skills necessary for promoting DX in the era of generative AI", The need to ask questions, verify hypotheses, etc. is reflected in skill standards.

described in “Part 2 C. Common Guiding Principles,” and does not mean that action is not necessary.

Hereinafter, the contents (items) described in “Table 1. Important matters for each AI business actor in addition to “Common Guiding Principles”” are identified and indicated in the rules of [AI business actor-Guiding Principle’s number)Content’s number].

- Identifier of the AI business actor use the acronym of Developer, Provider, and (Business) User, and the Guiding Principle’s number and the content’s number are numbered in the same table

(e.g., D-2). i. refers to key information about training by appropriate data of AI developer.

Table 1. Important matters for each AI business actor in addition to “Common Guiding Principles”

	Part 2. “Common Guiding Principles”	Important matters for each AI business actor in addition to “Common Guiding Principles”		
		Part 3. AI developer (D)	Part 4. AI provider (P)	Part 5. AI business user (U)
1) Human-centric	① Human dignity and individual autonomy ② Attention to decision-making and emotional manipulation by AI ③ Measures against disinformation, etc ④ Ensuring diversity and inclusion ⑤ User support ⑥ Ensuring Sustainability	-	-	-
2) Safety	① Consideration for human life, body, property and mind as well as the environment ② Proper use ③ Proper training	i. Training by appropriate data ii. Development considering human life, body, property and mind, as well as the environment iii. Development contributing to proper use	i. Risk measures considering human life, body, property, and mind as well as the environment ii. Provision contributing to proper use	i. Appropriate use for safety
3) Fairness	① Consideration of bias included in each component technology of AI model ② Intervention of human judgment	i. Consideration for bias in data ii. Consideration for bias included in AI model algorithms	i. Consideration of the bias contained in AI systems and services components and data	i. Consideration of bias in input data and prompts
4) Privacy Protection	① Privacy protection in general AI systems and services	i. Training appropriate data (D-2) i. Repost)	i. Introduction of mechanisms and measures to protect privacy ii. Measures against invasion of privacy	i. Measures against inappropriate input of personal information and invasion of privacy
5) Ensuring security	① Security measures affecting AI systems and services ② Attention to the latest trends	i. Introduction of mechanisms for security measures ii. Attention to the latest trends	i. Introduction of mechanisms for security measures ii. Treatment for vulnerabilities	i. Implementation of security measures
6) Transparency	① Ensuring verifiability ② Providing information to related stakeholders ③ Reasonable and honest response ④ Improving explainability and interpretability to related stakeholders	i. Ensuring verifiability ii. Provision of information to related stakeholders	i. Documentation of the system architecture ii. Provision of information to related stakeholders	i. Provision of information to related stakeholders
7) Accountability	① Improving traceability ② Explanation of compliance status of Common Guiding Principles ③ Clarification of responsible person ④ Distribution of responsibilities between parties ⑤ Specific responses to stakeholders ⑥ Documentation	i. Explanation of compliance status to Common Guiding Principles for AI providers ii. Documentation of development-related information	i. Explanation of compliance status to Common Guiding Principles for AI business users ii. Documentation of Service Regulations, etc.	i. Explanation to related stakeholders ii. Utilization of provided documents and compliance with terms and conditions
8) Education Literacy	① Ensuring AI literacy ② Education/Reskilling ③ Follow-up with stakeholders	-	-	-
9) Ensuring fair competition	-	-	-	-

10) Innovation	① Promotion of open innovation, etc. ② Attention to interconnectivity and interoperability ③ Providing appropriate information	i. Contribution to the creation of innovation opportunities	-	-

D. Common Guiding Principles for AI Business actors involved in advanced AI systems

AI Business actors involved in advanced AI systems should comply with the following, in addition to the "C. Common Guiding Principles", based on the " Hiroshima Process International Guiding Principles for all AI actors" established through the Hiroshima AI process and the " Hiroshima Process International Code of Conduct for Organizations Developing Advanced AI Systems." [20] However, since I) to XI) are applicable only to AI developers, each AI business actor is required to adhere to the appropriate scope as described later in Parts 3 to 5.

I) Take appropriate measures throughout the development of advanced AI systems, including prior to and throughout their deployment and placement on the market, to identify, evaluate, and mitigate risks across the AI lifecycle. ("2) Safety", "6) Transparency")

- Specifically, employing diverse internal and independent external testing measures, through a combination of methods such as red-teaming[21], and implementing appropriate mitigation to address identified risks and vulnerabilities
- In support of such testing, developers should seek to enable traceability, in relation to datasets, processes, and decisions made during system development.

II) Identify and mitigate vulnerabilities, and, where appropriate, incidents and patterns of misuse, after deployment including placement on the market. ("5) Ensuring security " and "7) Accountability").

- Use, as and when appropriate commensurate to the level of risk, AI systems as intended and monitor for vulnerabilities, and take appropriate action to address these.
  - encouraged to maintain appropriate documentation of reported incidents and to mitigate the identified risks and vulnerabilities, in collaboration with other stakeholders.

III) Publicly report advanced AI systems' capabilities, limitations and domains of appropriate and inappropriate use, to support ensuring sufficient transparency, thereby contributing to increase accountability. ("6) Transparency" and "7) Accountability")

- Make a reasonable explanation of what decision was made, starting with the source of the data, and document and publish it to ensure traceability.
- Document and publish in a clear and understandable manner so that relevant stakeholders can interpret the output of the AI system and use it appropriately by AI business users and non-business users

IV) Work towards responsible information sharing and reporting of incidents among organizations developing advanced AI systems including with industry, governments, civil society, and academia. ("5) Ensuring security", "6) Transparency", "7) Accountability", "10) Innovation")

[20] For details, refer to "II. International Guideline for Hiroshima Process for All AI Participants and Organizations Developing Advanced AI Systems" of "Hiroshima AI Process Comprehensive Policy Framework" in "Hiroshima AI Process G7 Digital and Technologies Ministerial Statement" adopted at the G7 Digital and Technologies Ministerial Conference (December 2023). https://www.soumu.go.jp/menu_news/s-news/01tsushin06_02000283.html

[21] A team that verifies the effectiveness of security response system and countermeasures from the perspective of how attackers attack target organizations.

- These include reports on monitoring results and documents related to security and safety risks.

V) Develop, implement and disclose AI governance and risk management policies grounded in a risk-based approach – including privacy policies, and mitigation measures, in particular for organizations developing advanced AI systems. (see "(4) Privacy protection", "7) Accountability")
- If appropriate case, publish privacy policy
- It is expected to establish and disclose AI governance policies and practices.

VI) Invest in and implement robust security management, including physical security, cyber security and security measures against internal threats, throughout the AI lifecycle ("5) Ensuring security").
- Consider operational measures for information security and appropriate cyber/physical access control, etc.

VII) Develop and deploy reliable content authentication and provenance mechanisms, where technically feasible, such as watermarking or other techniques to enable users to identify AI-generated content ("6" Transparency").
- Specifically, it includes content authentication and usual mechanisms created by the organization's advanced AI systems where appropriate and technically feasible.
- Make effort to develop tools and APIs that allow AI business users and non-business users to determine whether or not specific content through watermarks has been created using advanced AI systems.
  - It is encouraged to introduce other mechanisms, such as labelling and disclaimer labelling, to help AI business users and non-business users know that they are interacting with the AI system.

- Prioritize research to reduce social, safety and security risks and prioritize investment in effective mitigation measures ("10) Innovation") Includes research on improving AI safety, security, reliability, and handling risks

VIII) Prioritize research to mitigate societal, safety and security risks and prioritize investment in effective mitigation measures. ("10 Innovation")
- Implement efforts to develop reliable, human-centric AI, and at the same time provide support for the improvement of literacy among non-business users.

IX) Prioritize the development of advanced AI systems to address the world's greatest challenges, notably but not limited to the climate crisis, global health and education ("10) Innovation").
- Contribute to the development of international technical standards and best practices, including watermarks, use them if appropriate, and go along with the Standard Development Organization (SDO)

X) Advance the development of and, where appropriate, adoption of international technical standards ("2) Safety" and "3) Fairness")
- Appropriate measures to manage the quality of data, such as training data and data collection, are encouraged to mitigate harmful bias
- Appropriate transparency of the training dataset should also be supported and adherence to applicable legal frameworks

XI) Implement appropriate data input measures and protections for personal data and intellectual property ("5) Ensuring security" and "8) Education and literacy")
- Provide opportunities to improve the literacy and awareness of each AI business actor and stakeholder, including issues such as how advanced AI systems increase specific risks (e.g., those related to the proliferation of fake information) and how new risks are created

It is encouraged to collaborate among AI business actors to share information to identify and handle new risks and vulnerabilities associated with advanced AI systems.

XII) Promote and contribute to trustworthy and responsible use of advanced AI systems ("5) Ensuring security" and "8) Education and literacy")

- ➢ Provide opportunities to improve their own and, where appropriate, others' digital literacy, training and awareness, including on issues such as how advanced AI systems may exacerbate certain risks (e.g. with regard to the spread of disinformation) and/or create new ones

E. Establishing AI governance

In order to implement the "Common Guiding Principles" across the value chain and to utilize AI safely and securely, it is important to establish AI governance to manage AI-related risks at an acceptable level for stakeholders and maximize the benefits derived from them. In order to realize "Society 5.0", it is indispensable to construct the appropriate AI governance while advancing the social implementation of a system (Cyber physical system: CPS) which highly fuses cyber space and physical space. The CPS-based social is complex, rapidly changing, and it is difficult to control risks. As a result, the goal of AI governance is constantly changing. For this reason, it is important to implement "Agile Governance" in a variety of governance systems, such as enterprises, legal regulations, infrastructure, market, and social norms, rather than AI governance in which rules and procedures are fixed in advance. It is important to implement "Agile Governance", which involves the continuous and high-speed cycle rotation of "environmental and risk analysis", "goal setting", "system design", "operation", and "evaluation" by multi-stakeholders.[22]

It is important to consider the degree and likelihood of the AI risks and the resource constraints of each AI business actor when planning to develop, provide, and use it.

① First, we will conduct "environmental and risk analysis" related to the target AI systems and services based on the social acceptance of the benefits/risks and development/operation of AI systems and services throughout their lifecycles, changes in the external environment, and the level of AI proficiency.

② Based on this, it is determined whether or not to develop, provide, or use AI systems and services. When developing, providing, or using AI systems and services, the "setting of AI governance goals" should be examined through the development of AI governance policies. This AI governance goal should be set in accordance with management goals, such as the meaning of each AI business actor and its philosophy and vision.[23]

③ Moreover, after "design of AI management system" for achieving this AI governance goal, it will be "operated". In doing so, each AI business actor shall ensure that the AI Goals and their operational status fulfill "transparency and accountability to stakeholders" (fairness, etc.) externally.

④ Continuously monitor whether the AI management system is functioning effectively, including risk assessment, and perform "evaluation" and continuous improvement.

⑤ Even after the start of the operation of the AI system and services, based on changes in the external environment such as changes in the social systems such as regulatory, again conduct an environmental and risk analysis and revise the goals as necessary.

[22] In addition to the detailed explanation of AI governance practices based on METI's Governance Guidelines for AI Practice Ver. 1.1, the appendix also contains "Action Objectives" as specific actions of each AI business actor, and virtual "Practical Examples" assuming each AI business actor.

[23] As AI governance goals, it is conceivable that the company's action policies, which consist of the items to be addressed to the "Common Guidelines" described in these guidelines (such as the "AI Policy" with different names depending on each AI business actor) and action policies, which include other elements while encompassing the items to be addressed to the "Common Guidelines" (such as the Data Utilization Policy) may be established. Guidelines may also be provided to enhance the benefits of AI, such as increased inclusion by utilizing AI. The name is also left to each AI business actor.

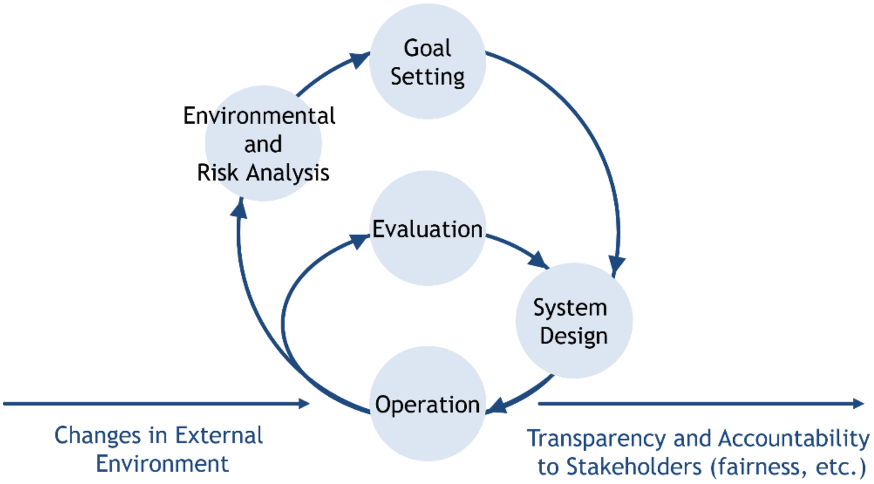

Goal Setting
Environmental and Risk Analysis
Evaluation
System Design
Operation
Changes in External Environment
Transparency and Accountability to Stakeholders (fairness, etc.)

Figure 6. Basic Model of Agile Governance

In addition, in considering AI governance, it is important to bear in mind the value chain and pay attention to the following points.

- ➢ Ensure collaboration among AI business actors from the perspective of the value chain and risk chain
  - ✧ Discussion points common to multiple AI business actors: Understanding AI risks, improving quality, creating new value by connecting AI systems and services together (System of Systems), improving the literacy of AI business users and Non-business users.
  - ✧ Points that need to be organized among AI business actors: Training data, rights-related contracts on generated AI models, etc.
- ➢ Clarify risk chains, including data distribution, and implement risk management and AI governance regimes appropriate for each stage of development, provision, and use.
  - ✧ If it is assumed that the value chain/risk chain covering AI development and service implementation will span multiple countries, ensure the state of consideration by the international community on appropriate AI governance for ensuring the free cross-border transfer of data (hereinafter referred to as "DFFT") and interoperability based on the situation (two aspects of "standards" and "interoperability between frameworks").

In order to make these measures effective, it is important for management to demonstrate leadership with their responsibility. In doing so, it is important to consider this as an up-front investment aimed at sustainable growth and medium-to long-term development of each AI business actor, rather than simply a cost from the perspective of pursuing short-term profits. It is expected that AI governance will be reflected into the strategies and corporate systems of each organization with the above-mentioned cycles turned under the leadership, and will be rooted in the culture of each organization.

Part 3. Matters Related to AI Developers

AI developers can design and modify AI models directly, so have a strong impact on AI output through whole AI systems and services. And they are expected to lead the innovation, and they have large impact on the whole society. For this reason, it is important to consider their AI's impacts when it provided and used, and to take countermeasures in advance as much as possible.

In AI development, there are sometimes conflicts between risks and ethics, such as the loss of privacy and fairness to emphasize accuracy, or the loss of transparency due to excessive emphasis on privacy. In such cases, it is important to make decisions and amendments accordingly based on the business risks and social impacts of the business operator. It is also important to keep records so that AI developers can reasonably explain what kind of involvement they have made, as parts of the value chain of the AI which may needs some explanation about the event of an unexpected accident in the AI system.

Important matters for AI developers are shown below.

- **Pre-processing of data and training**
  - ➢ D-2)i. Training by appropriate data
    - ✧ Collect data appropriately in the training phase, and handle the data throughout the lifecycle of the AI in accordance with laws and regulations through Privacy by Design, etc., when confidential information, personal data, and data protected by intellectual property rights that require attention are included. ("2) Safety", "4) Privacy protection", "5) Ensuring security")
    - ✧ Implement appropriate safeguards, such as considering the introduction of data management and restriction functions to control access to data during the entire period of pre-processing of data and training. ("2) Safety","5) Ensuring security")
  - ➢ D-3)i. Consideration for bias in data
    - ✧ Keep in mind that training data and the training process of the model can include biases (including potential biases that do not appear in the training data), and take reasonable steps to manage the quality of the data. ("3) Fairness")
    - ✧ Considering that the bias cannot be eliminated completely from the training data and the training process of the model, develop AI model in a reasonable manner based on various methods in parallel, rather than on a single method. ("3) Fairness")

- **During AI development**
  - ➢ D-2)ii. Development considering human life, body, property and mind, as well as the environment
    - ✧ Consider the followings to ensure that they do not harm the life, body, property and mind of their stakeholders, as well as the environment. (2) Safety)
      - Performance requirements that can withstand unexpected environments as well as expected performance under various conditions of use.
      - Methods for minimizing risks, such as guard rail technology. (e.g. inability to control or inappropriate output of linked robots) ("3) Fairness")
  - ➢ D-2)iii. Development contributing to proper use
    - ✧ Set safe-use ranges of the AI and develop it in the ranges to avoid harm caused by provision and use that are not envisioned at the time of development. ("2)

Safety”)
- ✧ Properly select the AI model when performing post-training for pre-trained AI models. (e.g., whether it is a commercially available licensed one, pre-trained data, specifications required for training and execution, etc.) (“2) Safety”)

- ➢ <u>D-3)ii. Consideration for bias included in AI model algorithms</u>
  - ✧ Consider the fact that bias can be included in each technical element constituting the AI model. (prompt for input by the AI business users and Non-business users, referred information when inferring by AI model, linked external services, etc.) ("3) Fairness")
  - ✧ Considering that bias cannot be completely eliminated from AI models, develop it in parallel based on a variety of methods rather than a single method. ("3) Fairness")

- ➢ <u>D-5)i. Introduction of mechanisms for security measures</u>
  - ✧ Take appropriate security measures in light of the characteristics of the technology used during the development of the AI system. ("Security by Design") ("5) Ensuring security ").

- ➢ <u>D-6)i. Ensuring verifiability</u>
  - ✧ Maintain and improve the quality of AI while preserving work records for post-verification, based on the characteristics that the predictive performance and quality of AI may significantly fluctuate from the beginning of utilization and may not reach the expected level. ("2) Safety" and "6) Transparency")

- **After AI development**

  - ➢ <u>D-5)ii. Attention to the latest trends</u>
    - ✧ Identify what to pay attention to in each development process in order to respond to the risks as new attack methods against AI systems are emerging every day. ("5) Ensuring security ").[24]

  - ➢ <u>D-6)ii. Provision of information to related stakeholders</u>
    - ✧ Give information on the AI systems they developed timely and appropriately. Examples are shown below. ("6) Transparency")
      - Possibility of output or program change due to AI system training, etc. ("1) Human-centric")
      - Safety information, such as the technical characteristics of the AI system, the security mechanisms, and the foreseeable risks that may result from the use of the AI system and mitigation measures (“2) Safety”)
      - Safe use ranges intended by AI developers to avoid harm caused by provision or use that is not envisioned at the time of development ("2) Safety")
      - Information on the operational status of the AI system and causes of defects and response status ("2) Safety")
      - Information on the AI systems update and the reason for the update (“2) Safety”)
      - Collection policy of data trained by AI model and its training method and implementation system ("3) Fairness", "4) Privacy protection", "5) Ensuring security”)

  - ➢ <u>D-7)i. Explanation of compliance status to “Common Guiding Principles” for AI providers</u>

[24] Information can be collected through the MIC's "AI-Security Information Transmission Portal （https://www.mbsd.jp/aisec_portal/)."

- ✧ Provide AI providers with information on the possible fluctuations in predictive performance and output quality from the beginning of utilization, the characteristics that may not reach the expected level, and the risks that may occur as a result. Specifically, announce the following items. (“7) Accountability”)
    - ● Response to biases that may be included in each technical element of the AI model. (training data, training process of the model, prompts assumed to be input by the AI business user and non-business user, referred information when inferring the AI model, linked external services, etc.) (“3) Fairness”)

➢ D-7)ii. Documentation of development-related information

- ✧ To improve traceability and transparency, document the development process of AI systems, and the data collection, labeling and the algorithms used that affect decision making, in a manner that allows third parties to verify them as much as possible. ("7) Accountability").
  (note) This does not mean disclosing everything documented here.

Expected matters for AI developers are shown below.

➢ D-10)i. Contribution to the creation of innovation opportunities

- ✧ Contribute to the creation of innovation opportunities by conducting followings to the extent possible. ("10) Innovation")
    - ● Research and develop the quality, reliability, and development methodologies
    - ● Contribute to maintaining sustained economic growth and solving social issues
    - ● Conduct internationalization and diversification, and industry-academic-government collaboration, such as by referring trends in international discussions on DFFT, etc., and participate in AI-developer communities and academic societies
    - ● Provide AI related information to the entire society

Additional entries in the "International Code of Conduct for the Hiroshima Process for Organizations Developing Advanced AI Systems"

In addition to the above, AI developers who develop advanced AI systems should comply with "Part 2 D. Common Guiding Principles for AI Business actors involved in advanced AI systems " and the "International Code of Conduct for Hiroshima Process for Organizations Developing Advanced AI Systems."[25]

In the comparison with "Part 2 D. Common Guiding Principles for AI Business actors involved in advanced AI systems ", the items additionally described in the "Code of Conduct" are shown below. Refer to "Appendix 3. A. Items to be Observed in Developing Advanced AI Systems" for the content of the entire Code of Conduct.

[25] G7 Leaders' Statement on the Hiroshima AI Process, "The Hiroshima Process International Code of Conduct for Organizations Developing Advanced AI Systems" (October 2023)
It should be noted that the document is a living document that is structured on the basis of existing OECD AI principles in response to trends in advanced AI-systems.
https://www.mofa.go.jp/mofaj/files/100573472.pdf

I. Take appropriate measures throughout the development of advanced AI systems, including prior to and throughout their deployment and placement on the market, to identify, evaluate, and mitigate risks across the AI lifecycle.

   - Document measures for risk mitigation and update them regularly. In addition, each AI business actor should evaluate and adopt mitigation measures against these risks in cooperation with relevant parties from across sectors.

II. Identify and mitigate vulnerabilities, and, where appropriate, incidents and patterns of misuse, after deployment including placement on the market.

   - Encourage consideration of incentives to disclose vulnerabilities through reward systems, contests, prizes, etc.

III. Publicly report advanced AI systems' capabilities, limitations and domains of appropriate and inappropriate use, to support ensuring sufficient transparency, thereby contributing to increase accountability.

   - In addition to the Transparency Report, the Instructions for Use and related technical documents should be kept up to date.

IV. Work towards responsible information sharing and reporting of incidents among organizations developing advanced AI systems including with industry, governments, civil society, and academia.

   - Develop and promote shared standards and mechanisms to ensure the safety and security of AI systems. In addition, appropriate documentation and cooperation with other AI business actors, sharing of relevant information and reporting to social should be conducted throughout the AI lifecycle.

V. Develop, implement and disclose AI governance and risk management policies, grounded in a risk-based approach – including privacy policies, and mitigation measures, in particular for organizations developing advanced AI systems.

   - If possible, the AI Governance Policy should be developed, implemented, disclosed and regularly updated to identify, evaluate, prevent and address AI risks throughout the entire AI lifecycle. In addition, an education policy should be established for business staff, etc.

VI. Invest in and implement robust security controls, including physical security, cybersecurity and insider threat safeguards across the AI lifecycle.

   - Assess the cyber security risks of advanced AI systems and require the storage of work and documents in an appropriate and secure environment. Measures to deal with unauthorized disclosure of risks, and the establishment of robust internal threat detection programs that are consistent with the protection of intellectual property and corporate secrets.

VII. Develop and deploy reliable content authentication and provenance mechanisms, where technically feasible, such as watermarking or other techniques to enable users to identify AI-generated content.

   - In addition to using watermarks and identifiers, each AI business actor should cooperate and invest in research to advance the context in this area.

VIII. Prioritize research to mitigate societal, safety and security risks and prioritize investment in effective mitigation measures.

   - Research and cooperate preferentially to handle risks, such as maintaining democratic values, respecting human rights, and protecting children and vulnerable populations. In addition, it is preferred to manage risks actively, including environmental and climate impacts, and to share risk research and best practices.

IX. Prioritize the development of advanced AI systems to address the world's greatest challenges, notably but not limited to the climate crisis, global health and education.

   - Support digital literacy initiatives that enable individuals and local communities to benefit from the use of AI and promote education and training for the public. Also develop solutions and identify issues with the public community and community groups.

X. Advance the development of and, where appropriate, adoption of international technical standards

- In addition to the development of international technology standards, the development of technology standards that can distinguish between AI-generated content and other content should be done.

XI. Implement appropriate data input measures and protections for personal data and intellectual property.

As appropriate measures for managing the quality of data, implement machine learning for transparency and privacy protection and countermeasure including testing and fine tuning of leaks of sensitive data, and the introduce appropriate safeguards to honor rights related to privacy and intellectual property, including copyright-protected content.

Part 4. Matters Related to AI Providers

AI providers are responsible for adding value to AI systems developed by AI developers and providing AI systems and services to AI business users. It is important for AI providers to realize the provision of AI systems and services based on the assumption of the appropriate use of AI because of the large impact on the social, while it greatly contributes to the socio-economic growth as well as the popularization and development of AI in the society. Therefore, it is important to note whether the AI incorporated in the AI system and service is appropriate for itself. In addition, appropriate change management, configuration management, and maintenance of services considering changes in the expectations of AI due to changes in business strategy and social environment is also important.

It is important to require AI developers to develop AI systems properly as well as implement AI systems and services within the scope intended by AI developers, and continue normal and proper operation. For AI business users, it is important to provide AI systems and continue supporting the operation or to provide AI services while operating AI systems. AI providers are expected to share relevant information, including incident cases to a reasonable extent so as not to cause unintended disadvantages, such as violation of the rights of social and other stakeholders, and to provide safer, more secure and reliable AI systems and services.

Important matters for AI providers are shown below.

- **AI system implementation**

  - P-2)i. Risk measures considering human life, body, property and mind, as well as the environment
    - Maintain performance levels under various circumstances, as well as under the conditions of use expected at the time of provision, so as not to cause harm to the life, body, property and mind of relevant stakeholders, including AI business users, and the environment. Consider methods (such as guard rail technology) to minimize risks. (such as inability to control or inappropriate output of linked robots) (“2) Safety”)

  - P-2)ii. Provision contributing to proper use
    - Correctly determine points to keep in mind when using AI systems and services. (“2) Safety”)
    - Utilize AI within the scope set by AI developers. (“2) Safety”)
    - Confirm the accuracy, and if necessary, currency (appropriateness) of the data at the time of provision. (“2) Safety”)
    - Examine whether there is any difference between AI usage environment AI developers established and the actual usage environment of AI business users. (“2) Safety”)

  - P-3)i. Consideration of the bias contained in the AI systems and services components and data
    - Confirm the fairness of data at the time of provision and consider bias in referred information and linked external services., etc. (“3) Fairness”)
    - Assess the input/output and the rationale of judgment of AI models and monitor the occurrence of bias on a regular basis. Also, if necessary, encourage AI developers to re-evaluate the bias of each technical element that constitutes the AI model, and to decide to improve the AI model based on the evaluation results. ("3) Fairness")
    - Consider the possibility that AI systems, services and user interfaces that receive AI model’s output may contain biases that arbitrarily restrict business processes and decisions by AI business users and Non-business users. ("3) Fairness")

- P-4)i. Introduction of mechanisms and measures to protect privacy
  - Take measures to protect privacy through the implementation of AI systems, such as the introduction of a mechanism to manage and restrict access to personal information appropriately in light of the characteristics of the technologies adopted. (Privacy-by-Design) (“4) Privacy protection”)

- P-5)i. Introduction of mechanisms for security measures
  - Take appropriate security measures in light of the characteristics of the technology to be adopted through the process of providing AI systems and services. (Security-by-Design) ("5) Ensuring Security")

- P-6)i. Documentation of the system architecture
  - To improve traceability and transparency, document the system architecture and data processing processes of the AI systems that will affect decision making. ("6) Transparency")

- **After provision of AI systems and services**

  - P-2)ii. Provision contributing proper use
    - Verify that AI systems and services are being used for appropriate purposes in a regular basis. ("2) Safety")

  - P-4)ii. Measures against invasion of privacy
    - Collect information properly regarding invasion of privacy in the AI systems and services, and deal with any infringement appropriately, and consider preventing a recurrence. ("4) Privacy Protection")

  - P-5)ii. Treatment for vulnerabilities
    - Since a number of attack methods against AI systems and services have been developed, check the latest risks and trends to be aware of in each process of provision in order to treat for latest risks. Also, consider eliminating vulnerabilities. ("5) Ensuring security")

  - P-6)ii. Provision of Information to Related Stakeholders
    - Give information on the AI systems and services they implemented, timely and appropriately in a clear and accessible manner. Examples are shown below. ("6) Transparency")
      - The fact that AI is being used and appropriate/inappropriate AI usage, etc. ("6) Transparency")
      - Safety information, such as the technical characteristics of the AI system, the security mechanisms, and the foreseeable risks that may result from the use of the AI system and mitigation measures (“2) Safety”)
      - Potential changes in output or programs due to training of AI systems and services ("1) Human-centric")
      - Information on the operational status of the AI system and causes of defects and response status, incident cases, etc. ("2) Safety")
      - Information on the AI systems update and the reason for the update (“2) Safety”)
      - Collection policy of data trained by AI model and its training method and implementation system ("3) Fairness", "4) Privacy protection", "5) Ensuring security)"

- P-7)i. Explanation of compliance status to “Common Guiding Principles” for AI business users
  - Promote appropriate use by AI business users and provide the following information to them. ("7) Accountability")
    - Notice regarding the use of data that are guaranteed to be accurate and, if necessary, to be currency (appropriate) (“2) Safety”)
    - Notice against the inappropriate training of AI models by in-context learning ("2) Safety")
    - Notes when entering personal information ("4) Privacy protection")
  - Notice about inappropriate entry of personal information into AI systems and services provided. ("4) Privacy protection")

- P-7)ii. Documentation of service regulations, etc.
  - Prepare service agreements for AI business users and non-business users. ("7) Accountability")
  - Specify privacy policies. ("7) Accountability")

In addition, AI providers dealing with advanced AI systems respond as follows to “Part 2 D. Common Guiding Principles for AI Business actors involved in advanced AI systems”.

- I)~XI) Should be followed to the appropriate extent
- XII) Should be observed

Part 5. Matters Related to AI Business Users

It is important for AI business users to receive safe, secure, and reliable AI systems and services from AI providers and to continuously use them properly within the scope intended by AI providers and to operate AI systems as needed. This will enable us to realize the greatest benefits of innovation through AI, such as improving operational efficiency, productivity, and creativity. By intervening human judgment, it is also possible to prevent unexpected incidents and to enrich oneself and society while protecting human dignity and autonomy.

When an AI business user is asked to explain the capabilities and output results of the AI by the social and the stakeholders, it is expected that the AI business user respond to the request and gain their understanding with support from the AI provider. It is also expected that the AI business user acquire the knowledge necessary for more effective AI use and make efforts to obtain the understanding not only from the relevant parties but also from society and the stakeholders.

Important matters for AI business users are shown below.

- **Use of AI systems and services**
  - U-2)i. Appropriate use for safety
    - Comply with the usage considerations specified by the AI provider, and use AI within the scope anticipated by the AI provider in designing the AI. ("2) Safety")
    - Input accurate and, if necessary, currency guaranteed (appropriate) data. (“2) Safety”)
    - Understand the accuracy and degree of risk of AI output and confirm various risk factors before using it. (“2) Safety”)
  - U-3)i. Consideration of bias in input data and prompts
    - Input fairness confirmed data so not to be suffered from a significant lack of fairness, and determine the use of the AI output into business responsibly, taking into account the bias included in the prompt. ("3) Fairness")
  - U-4)i. Measures against inappropriate input of personal information and invasion of privacy
    - Pay attention not to input personal information inappropriately into AI systems and services. ("4) Privacy protection")
    - Collect information on privacy violations in AI systems and services appropriately and consider preventing such violations. ("4) Privacy Protection")
  - U-5)i. Implementation of security measures
    - Comply with the security considerations by AI providers. ("5) Ensuring security")
  - U-6)i. Provision of information to related stakeholders
    - Input fairness confirmed data so not to be suffered from a significant lack of fairness, and obtain output by AI with paying attention to the bias included in the prompt, so that use the output for making business decision. Then inform the related stakeholders the result of the decision. ("3) Fairness", "6) Transparency")
  - U-7)i. Explanation to related stakeholders
    - Provide information, including how to use it appropriately, in plain and an easy-to-access manner, to reasonable extent depending on the characteristics of the related stakeholders . ("7. Accountability").

- ✧ If data provided by related stakeholders is planned to be used, provide information regarding data provision methods, formats, etc. to the related stakeholders in advance based on the characteristics and uses of AI, points of contact with the provider, and privacy policy, etc. ("7) Accountability")
- ✧ In case of referencing the output of the AI for evaluation of a particular individual or group, inform the particular individual or group that AI is used for the evaluation. And comply with processes this guideline recommends to confirm the accuracy and fairness and transparency of the output, so that fulfill explanation responsibility to the AI evaluation target individual or group with human reasonable judgment considering the automation bias. ("1) Human-centered", "6) Transparency", "7) Accountability").
- ✧ Depending on the characteristics of the AI systems and services used, establish a contact point to respond to inquiries from related stakeholders, and provide explanation and receive requests in cooperation with AI providers. ("7) Accountability").

- ➢ U-7)ii. Utilization of provided documents and compliance with terms and conditions
  - ✧ Properly store and utilize documents on AI systems and services provided by AI providers. ("7) Accountability")
  - ✧ Comply with the terms of service set by AI providers. ("7) Accountability")

In addition, AI business users dealing with advanced AI systems respond as follows to “Part 2 D. Common Guiding Principles for AI Business actors involved in advanced AI systems”.

- ● I)~XI) Should be followed to the appropriate extent
- ● XII) Should be observed