

Outline of “AI Guidelines for Business Ver1.0”

Ministry of Internal Affairs and Communications
Ministry of Economy, Trade and Industry
(April 19, 2024)

Background and Purpose of Formulating “AI Guidelines for Business”

- Technologies related to AI (Artificial Intelligence), as represented by generative AI, are constantly evolving, opportunities to use AI and its various possibilities have been increasing continuously. AI is being used for making industrial innovations and solving social challenges as well.
- In Japan, it is expected that AI will contribute to the realization of "Society 5.0."
- Japan led discussions held at international forums, such as the G7, G20, and OECD, and made a lot of contributions, starting with the proposal for AI R&D Principles at the G7.
- Given this situation, the aim is to actively and cooperatively develop a framework that achieves both promotion of innovation and reduction of risk across the lifecycle by providing a unified guiding principle in AI governance in Japan.

Innovation

Realization of
"Society 5.0"

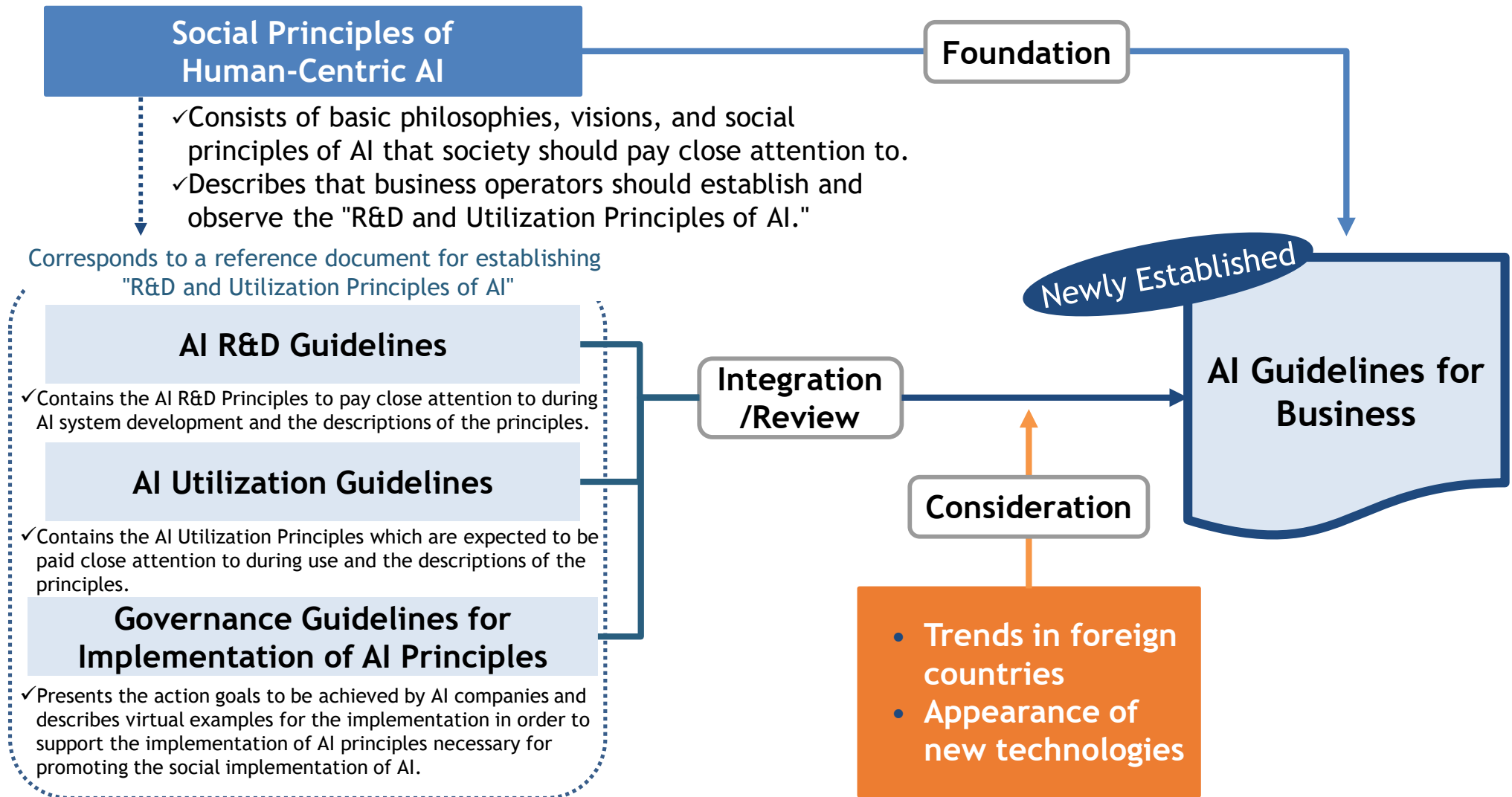
International Discussion

Formulation of “AI Guidelines for Business”

- Aim to actively and cooperatively develop a framework that achieves both promotion of innovation and reduction of risks across the lifecycle by encouraging those using AI in various businesses to fully recognize AI risks based on international trends and stakeholders' concerns.
- Aim to voluntarily take the necessary countermeasures across the entire lifecycle.

Policy for Formulating “AI Guidelines for Business”

- “AI Guidelines for Business” is established based on the “Social Principles of Human-Centric AI,” integrating three guidelines in Japan, reflecting the features of AI technologies that had further advanced in recent years, and the international discussions about social implementation of AI.
- Ensuring consistency with previous guidelines, continuous development as a governance mechanism that supports business activities is expected.



Basic Concept of “AI Guidelines for Business”

- The basic concepts of "AI Guidelines for Business" are **1** Support for voluntary efforts by business operators, **2** Coordination with international discussions, and **3** Understandability for readers.
- In addition, the Guidelines will continue to be updated as a "Living Document" through continuous "multiple stakeholder" reviews and with an emphasis on effectiveness and legitimacy.

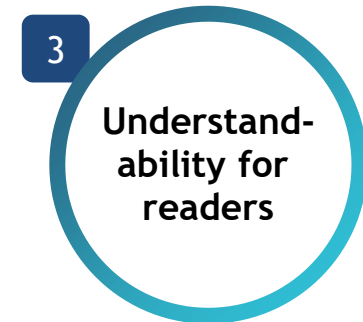
Concepts



Show directions for AI business actors founded on the risk-based approach where the degree of measures should be proportionate to the level and probability of risks.



Ensure consistency with trends and contents of domestic and overseas relevant principles.



Readers can check risks and handling policies that should be considered regarding AI, for each of AI developers, AI providers, and AI business users.

Processes



Multiple stakeholders

Established through studies conducted by multiple stakeholders that consisted of academic and research institutions, civil societies including general consumers, private sector companies, and the like, to prioritize effectiveness and validity.

Living Document

To continuously improve AI governance, updated as needed while reflecting the agile governance philosophy.

Reference) Multiple stakeholder Collaboration

- The Guidelines are established through studies conducted by multiple stakeholders that consisted of academic and research institutions, civil societies including general consumers, private sector companies, and others, rather than having the government take the initiative alone, to prioritize effectiveness and validity.

Collaborators



Collaboration Method

Numerous opportunities to exchange opinions and engage in discussion

- Review committee composed of the collaborators listed on the left
- Working groups primarily composed of practitioners
- Discussions with private sector companies



Gathering a wide range of knowledge through inquiries of opinions

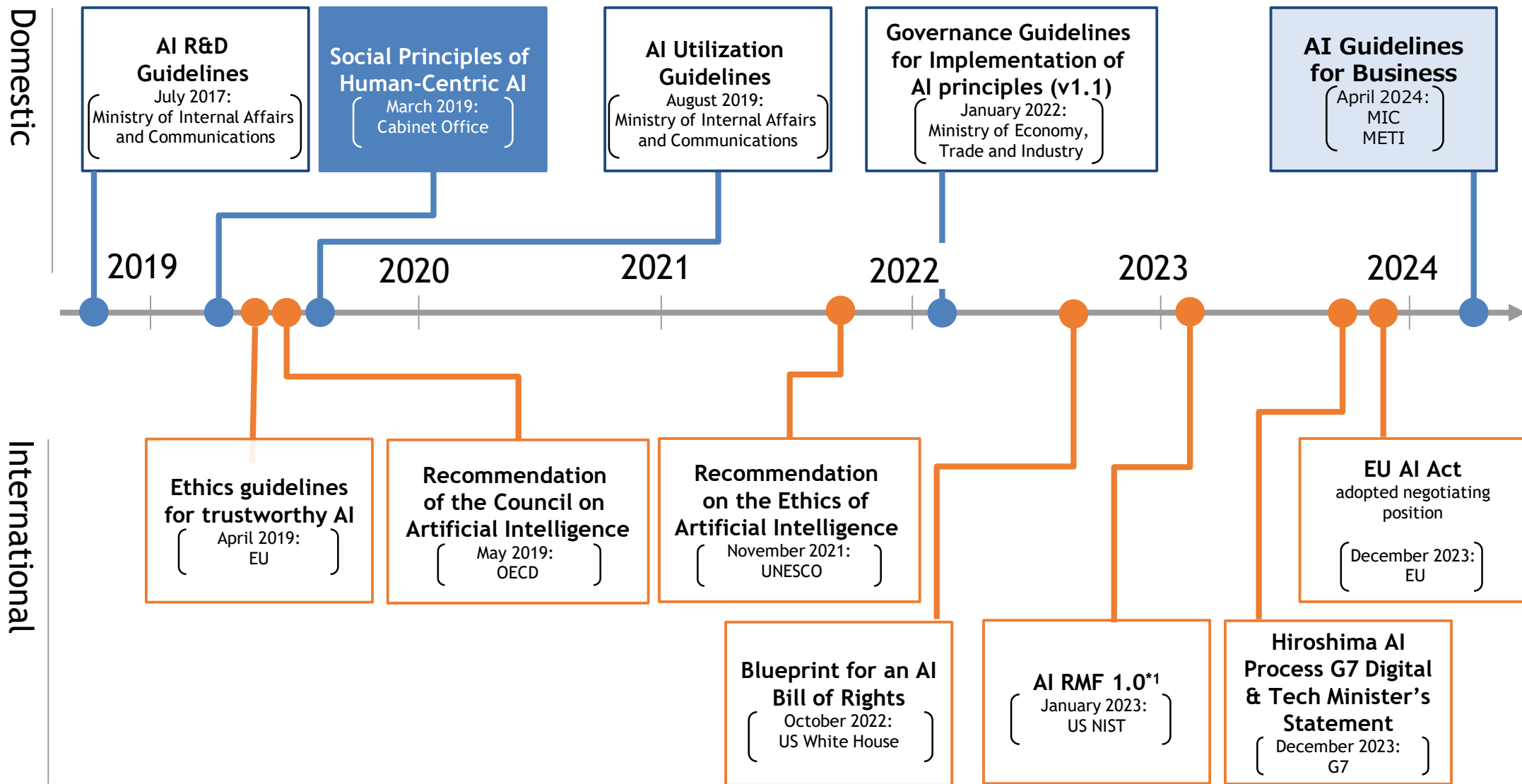
- Approximately 100 experts
 - Private company representatives
 - Specialists, researchers
 - Civil societies, consumer groups, etc.



Gathering a wide range of opinions through public comments

Reference) Major Principles Related to AI, etc.

- As the formulation of various regulations and guidelines is actively discussed in other countries, "AI Guidelines for Business" will also be closely aligned with various principles and regulatory trends.

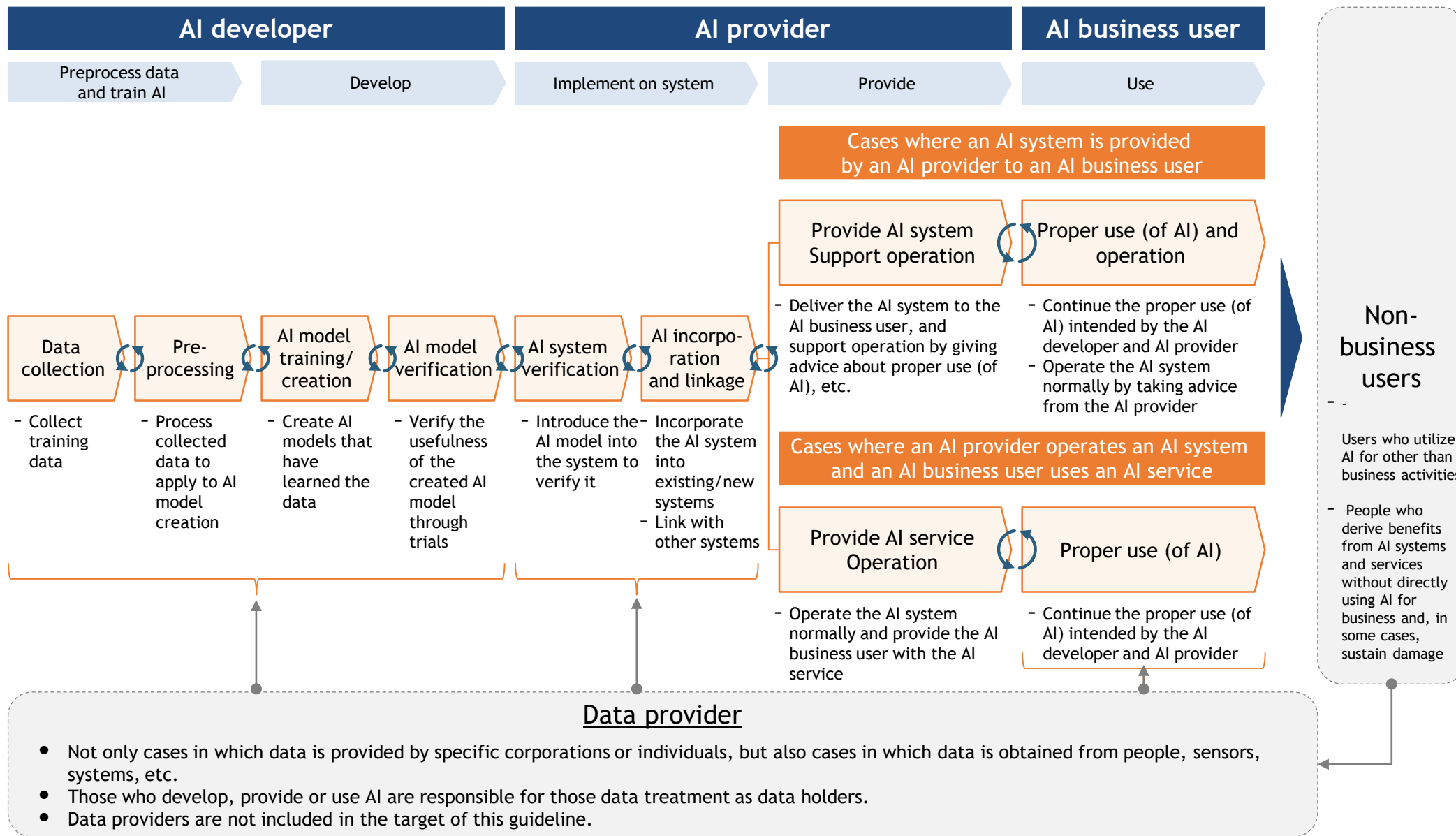


*1: AI Risk Management Framework 1.0

to the General AI Utilization Process

- Considering their specific roles in the AI lifecycle, the parties that the Guidelines are intended for are roughly grouped under the three categories: "AI developers," "AI providers," and "AI business users."

*"Data providers" and "non-business users" are excluded.



Positioning of main part and appendix of AI Guidelines for Business

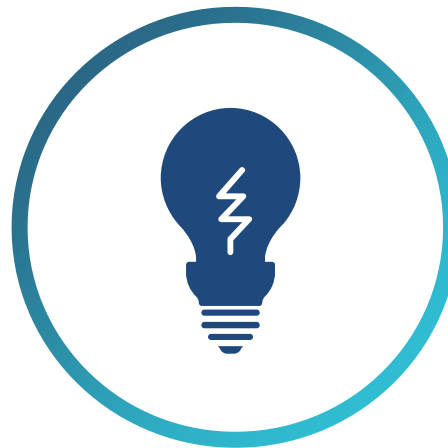
- The main part covers "the efforts to be made regarding AI (guiding principles = what)" based on "the ideal society while considering stakeholders' expectations (basic philosophies = why)" that are important for using AI safely and securely to maximize the benefits of AI.
- The appendix covers "the specific approach to be adopted (implementation = how)" to lead AI business actors to take actual implementation of the principles.

Main part (why, what)

Appendix (how)



The ideal society while
considering stakeholders'
expectations
(basic philosophies = why)



The efforts to be made
regarding AI
(guiding principles = what)



The specific approach to
be adopted
(implementation = how)

Structure of “AI Guidelines for Business”

- The descriptions in the Appendix correspond to those in the main part and serve as a supporting document for the reading of the main part and considerations and actions based on the main part.

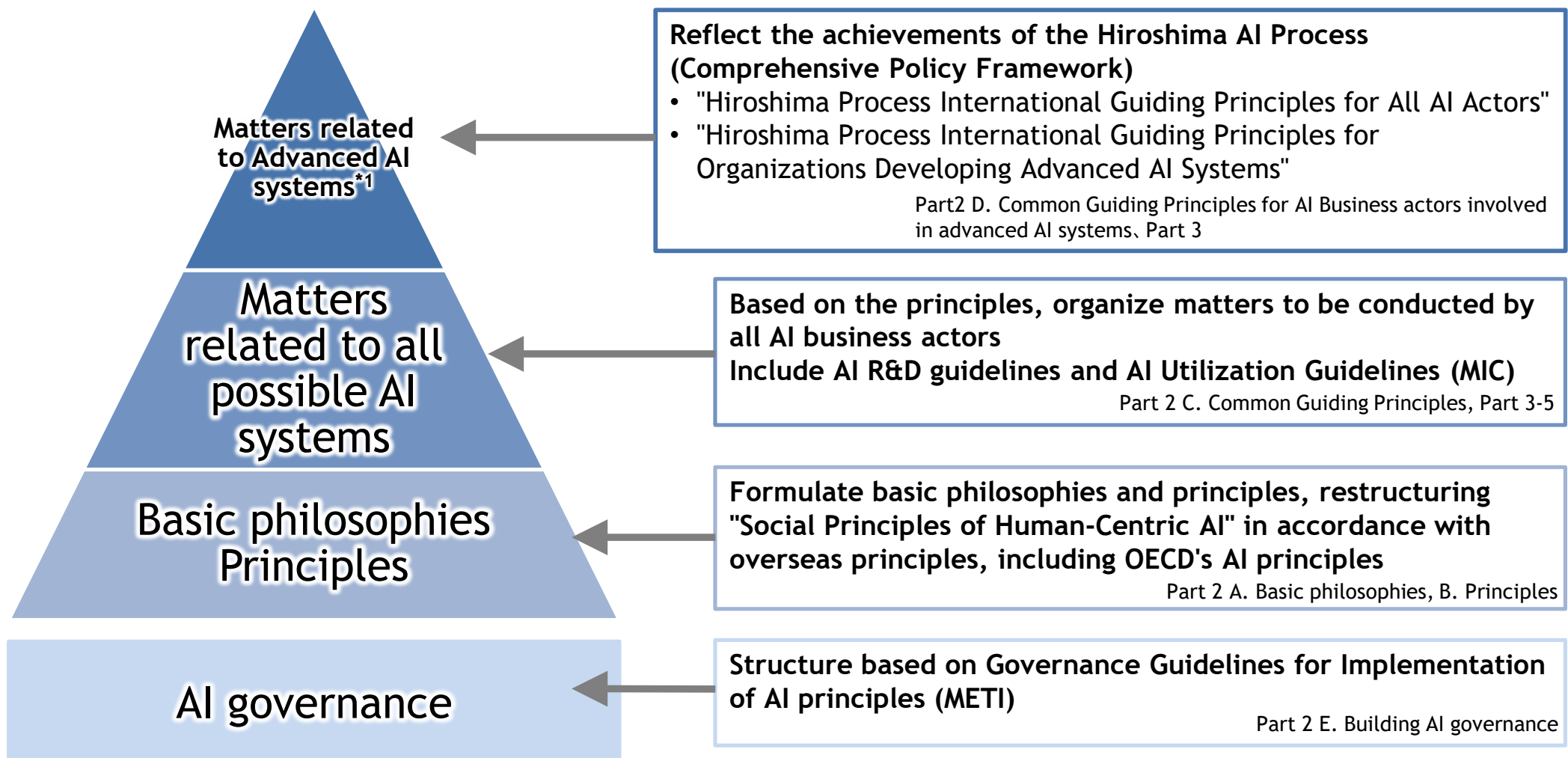
	Main part (why, what)	Appendix (how)
For all AI business actors	Part 1 Definitions	1. Relevant to Part 1 [About AI]
	Part 2 Society to aim for with AI, and matters each AI business actor works on	2. Relevant to Part 2 [E.Building AI Governance]
For each AI business actor	Part 3 Matters Related to AI Developers <small>* Includes additional matters described in "Hiroshima Process International Code of Conduct for Organizations Developing Advanced AI Systems" as well</small>	3. Relevant to Part 3 [For AI Developers]
	Part 4 Matters Related to AI Providers	4. Relevant to Part 4 [For AI Providers]
	Part 5 Matters Related to AI Business Users	5. Relevant to Part 5 [For AI Business users]
Other references		6. Major precautions for referring to "Contract Guidelines on Utilization of AI and Data"
		7. Checklist
		8. Cross-actor virtual cases
		9. References for overseas guidelines, etc.

- A. Preconditions for AI
- B. AI's benefits and risks
- A. Building of AI governance and monitoring by management
- B. Examples of business operator's efforts at AI governance
- A. Descriptions of Part 3 "Matters Related to AI Developers"
- B. Descriptions of "Common Guiding Principles" in Part 2
- C. Matters to be observed in developing advanced AI systems
- A. Descriptions of Part 4 "Matters Related to AI Providers"
- B. Descriptions of "Common guiding principles" in Part 2
- A. Descriptions of Part 5 "Matters Related to AI Business Users"
- B. Descriptions of "Common Guiding Principles" in Part 2

The appendixes 7, 8, and 9 are Japanese only.

The parties the Guidelines are intended for

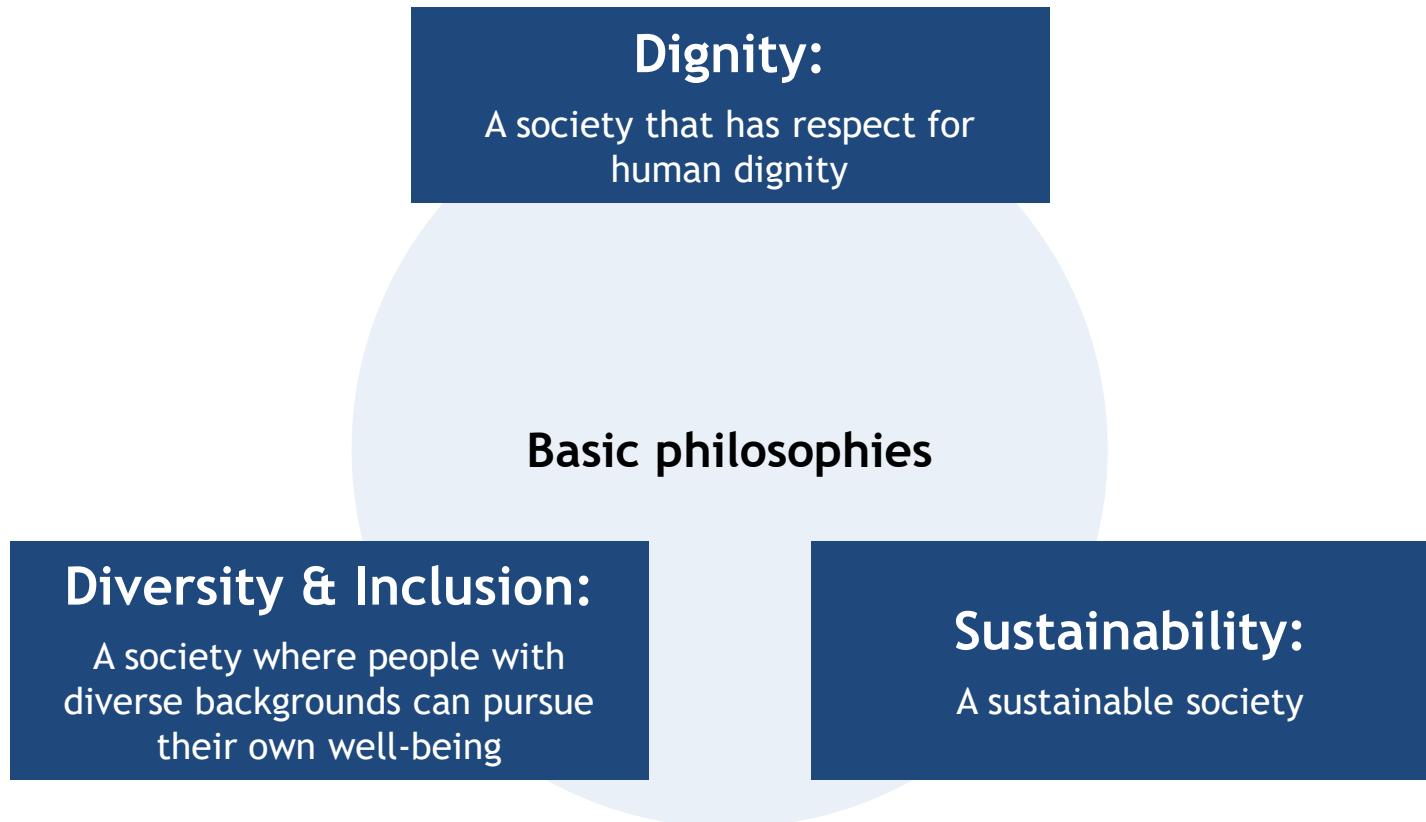
- **Broadly covers all possible AI system and services, including general AI, while reflecting The Hiroshima AI Process Guiding Principles as well as Code of Conduct regarding advanced AI systems.**
- **For the actual development, provision, and use of AI, it is important that all business operators who intend to use AI will voluntarily promote specific efforts using the Guidelines as one of their references, such as properly building AI governance for compliance with the guidelines.**



*1: The most advanced AI systems including the cutting-edge foundation models and generative AI systems

Basic Philosophies

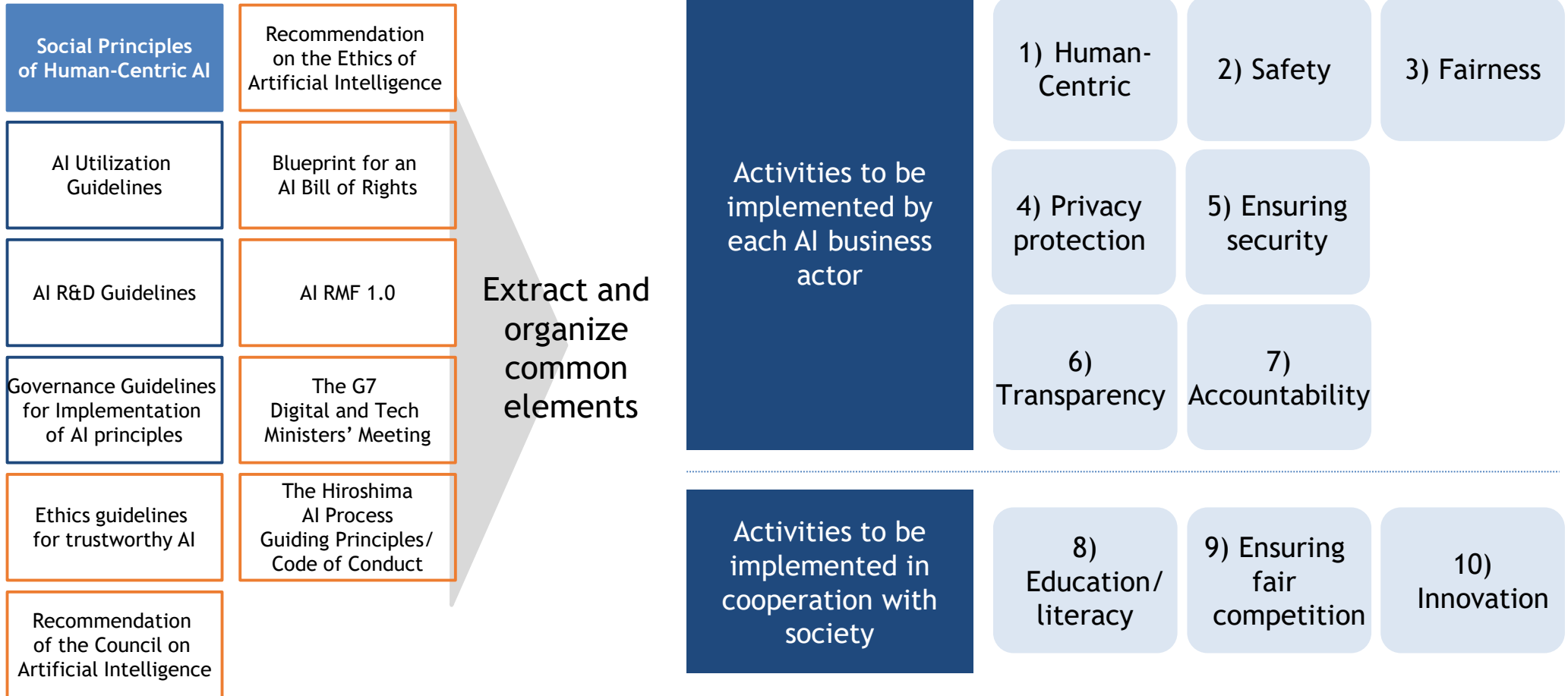
- "Social Principles of Human-Centric AI" states that it is expected that AI will contribute to the creation of Society 5.0, and that it is important to use AI as a public asset of humans that can lead to the achievement of global sustainability through qualitative changes of the ideal society as well as true innovations.
- In addition, the document also states that the following three values should be respected as basic philosophies in order to build a society that upholds such philosophies..



Common Guiding Principles for All Actors

Part 2

- The "Common Guiding Principles" outlines what each actor works on in collaboration to achieve the society aimed for through AI.
- The "Common Guiding Principles" are formulated based on the "Social Principles of Human-Centric AI," integrating three guidelines in Japan with consideration of trends in other countries and the emergence of new technologies.
- As a result, it is organized into what each actor works on and what is expected to be worked on in collaboration with society.



Common Guiding Principles for All Actors [1/2]

- Each AI business actor should develop, provide, or use AI systems and services respecting the rule of law, human rights, democracy, diversity, and fair and just society in accordance with the first principle: "1) Human-Centric."
- Relevant laws, including the Constitution of Japan, Intellectual Property Basic Act and relevant laws, and Act on the Protection of Personal Information as well as existing laws and regulations in individual fields pertaining to AI should be observed

Guiding Principles

Content (excerpts of main points)

Activities to be implemented by each AI business actor

1) Human-Centric

- ✓ Act so that AI expands human abilities and enables diverse people to seek diverse well-being.
- ✓ Recognize the increasing risk of destabilizing and confusing the society through **disinformation, misinformation, and biased information**, and take necessary countermeasures.
- ✓ Pay attention to make it easy for **socially vulnerable people to use AI** to allow more people to enjoy the benefits

2) Safety

- ✓ Conduct appropriate risk analyses to take **countermeasures against risks**.
- ✓ Prevent damage due to a provision or use that deviates from the intended purpose the range in which the AI business actor can control.
- ✓ In accordance with the characteristics and purposes of AI systems and services, ensure the accuracy and recency of the data to be used for training, properly take actions such as **the securement of transparency of data used for training, compliance with the legal framework**, and update AI models to a reasonable extent.

3) Fairness

- ✓ Attempt to **eliminate unfair and harmful bias and discrimination** against any specific individuals or groups based on race, gender, national origin, age, political opinion, religion, and so forth.
- ✓ To prevent AI from generating unfair results, consider implementing timely human interventions, rather than letting AI make the decisions alone, and **be careful of unconscious and potential biases**.

4) Privacy protection

- ✓ **Observe relevant laws**, including the Act on the Protection of Personal Information, and **formulate and announce the privacy policy** of each AI business actor, to take measures to respect and protect the privacy of stakeholders, in accordance with its importance, based on the social contexts and legitimate expectations of people.

5) Ensuring security

- ✓ **Maintain the confidentiality, integrity, and availability** of AI systems and services and ensure safe and secure AI use constantly, take reasonable measures based on the technological level at the time.
- ✓ New methods for attacking AI systems and services from the outside are increasing on a daily basis. In order to **address those risks, check the matters to be noted**.

Common Guiding Principles for All Actors [2/2]

Part 2

- Each AI business actor should develop, provide, or use AI systems and services respecting the rule of law, human rights, democracy, diversity, and fair and just society in accordance with the first principle: "1) Human-Centric."
- Relevant laws, including the Constitution of Japan, Intellectual Property Basic Act and relevant laws, and Act on the Protection of Personal Information as well as existing laws and regulations in individual fields pertaining to AI should be observed

Guiding Principles

Content (excerpts of main points)

Activities to be implemented by each AI business actor

6) Transparency

- ✓ **Provide stakeholders with information to the reasonable extent necessary** and technically possible while ensuring the verifiability of the AI system or service based on the social context when the AI system or service is used. (e.g., fact that AI is used and its scope, methods for data collection and annotation, capabilities and limitations of the AI system or service, proper/improper use by AI business users, etc.)

7) Accountability

- ✓ Provide and explain information to stakeholders within reasonable extent for ensuring traceability, conforming to common guiding principles, and the like.
- ✓ **Establish and publicly report policies** on AI governance or privacy.
- ✓ Document and store relevant information and keep them available for a prescribed period whenever and wherever required and able to be referenced in a manner appropriate for their use.

8) Education /literacy

- ✓ Take the necessary steps to ensure that the persons engaged in AI in each AI business actor **acquire AI literacy of the level sufficient for the engagement**.
- ✓ It is expected to **provide stakeholders with education** in consideration of the characteristics of AI including its complexity and the misinformation that it may provide, and possibilities of intentional misuse of AI.

9) Ensuring fair competition

- ✓ Each AI business actor is expected to **maintain the fair competitive environment surrounding AI** so that new businesses and services using AI are created, the sustainable economic growth is maintained, and solutions for social challenges are provided.

10) Innovation

- ✓ Promote internationalization, diversification, **collaboration among industry, academia, and government sectors**, and open innovation.
- ✓ Ensure the interconnectivity and interoperability between your AI systems/services and other AI systems/services.
- ✓ When there are standard specifications, comply with them.

Activities to be implemented in cooperation with society

- The business actors involved in advanced AI systems should comply with the following in addition to the common guiding principles. Note that some items from I) to XI) are applicable only to AI developers of advanced AI systems, so AI providers and AI business users are required to comply with items within the appropriate scope.
 - 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.
 - II. Identify and mitigate vulnerabilities, and, where appropriate, incidents and patterns of misuse, after deployment including placement on the market.
 - 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.
 - IV. Work towards responsible information sharing and reporting of incidents among organizations developing advanced AI systems including with industry, governments, civil society, and academia.
 - 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.
 - VI. Invest in and implement robust security management, including physical security, cyber security and security measures against internal threats, throughout the AI lifecycle.
 - 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.
 - VIII. Prioritize research to mitigate societal, safety and security risks and prioritize investment in effective mitigation measures.
 - 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.
 - X. X) Advance the development of and, where appropriate, adoption of international technical standards.
 - XI. Implement appropriate data input measures and protections for personal data and intellectual property.
 - XII. Promote and contribute to trustworthy and responsible use of advanced AI systems.

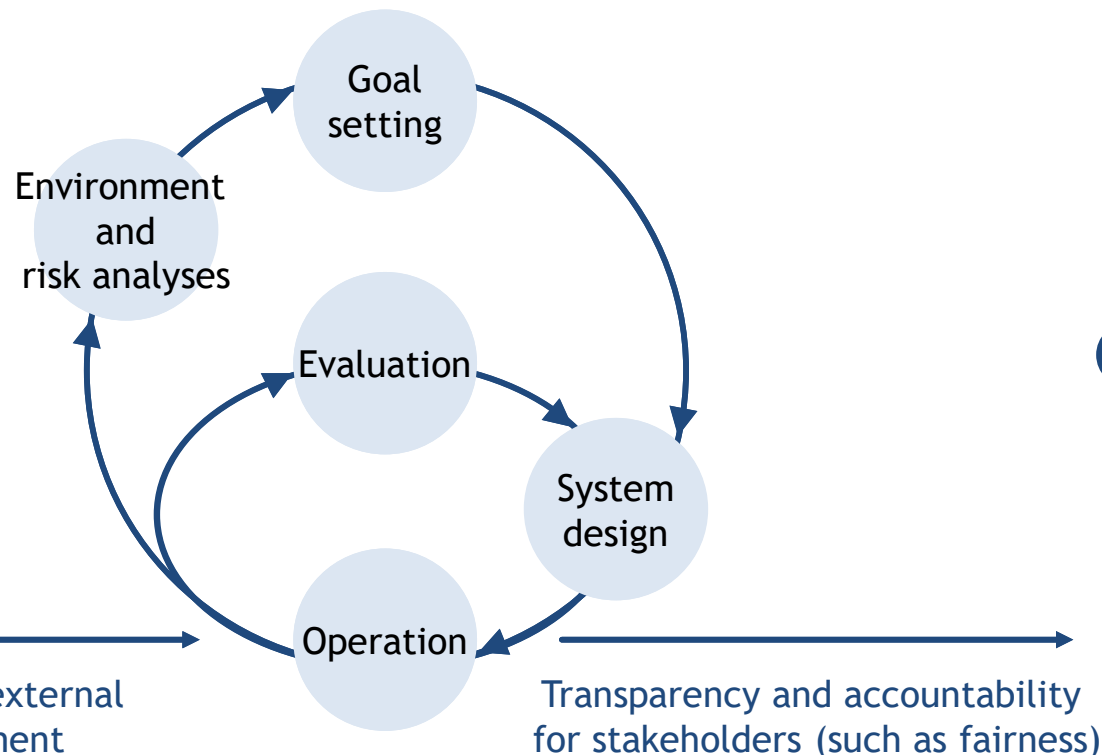
*1: For details, see the "Hiroshima Process International Guiding Principles for All AI Actors" in the "Hiroshima AI Process G7 Digital & Tech Ministers' Statement" adopted on December 1, 2023, at the G7 Digital and Technology Ministerial Meeting.

Note that advanced AI systems are defined as the most advanced AI systems, including the most advanced foundation models and generative AI systems.

Building AI governance

- In order to utilize AI safely and securely, it is important to manage risks by establishing appropriate governance under the leadership of management, paying attention to the following:
 - Secure the cooperation for issues among multiple AI business actors from the viewpoints of value chain and risk chain
 - If the value and/or risk chains are expected to span multiple countries, understand how the international society is studying AI governance suitable for ensuring free distribution of data
 - Fit AI governance into the strategy of each organization as well as the company system expecting the agile governance cycle to take hold in each organization as its culture through management's leadership

Building Appropriate Governance



+

Secure cooperation between multiple actors

Secure cooperation between actors from the value chain/risk chain perspective



Appropriate data transfer

Appropriate risk management/governance implementation in case of multi-country scenarios



Management leadership

Fit into strategies/company systems and become a part of corporate culture

- It is especially important for AI developers to study in advance as much as possible the impacts that the AI they develop may pose when it is provided or used and take necessary measures against the impacts as they can directly design and modify AI models

During data preprocessing/ training

- D-2) i. Proper data training
- Properly collect training data through privacy-by-design, etc., and if it contains third-parties' personal data, data requiring attention to intellectual property rights, etc., ensure that it is properly handled in compliance with laws and regulations.
 - Implement proper protective measures**, for example, considering the deployment of any data management and restriction function that controls access to data.
- D-3) i. Consideration for bias in data
- Take reasonable measures to control the quality of the data**, noting that there may be biases on the learning process of training data and AI models.
 - Based on the fact that biases cannot be completely eliminated, make sure AI models are trained with properly represented data sets and check AI systems assume no bias.

When developing AI

- D-2) ii. Development that takes into consideration the lives, bodies, properties, and minds of humans and the environment
- Consider requirements for the performance achievable under the use in an unexpected environment, and methods for **minimizing risks**.
- D-2) iii. Development contributing to proper use (of AI)
- Establish clear policies and guidance on the safe use of AI** and **select a proper pre-trained AI model** when giving a post-training.
- D-3) ii. Consideration for bias in algorithms, etc., of AI models
- Consider the possibility that **bias can be included by each technical element** that makes up the AI model.
 - Make sure AI models are trained with properly represented data sets and AI systems assume no bias based on the fact
- D-5) i. Deployment of mechanisms for security measures
- Take security measures** appropriately based on the characteristics of the adopted technologies (secure by design).
- D-6) i. Ensuring verifiability
- Note that the prediction performance and output quality of AI may significantly change or may fail to attain the expected precision after the use of AI is started. **Preserve work records for follow-up verification** and take measures to maintain and improve the AI quality.

- It is especially important for AI developers to study in advance as much as possible the impacts that the AI they develop may pose when it is provided or used and take necessary measures against the impacts as they can directly design and modify AI models

After developing AI

- | | | |
|----------|--|---|
| D-5) ii. | Consideration for the latest trends | <ul style="list-style-type: none"> - New attack methods to AI systems are increasing on a daily basis. In order to address those risks, <u>considerations to be noted in each step of development should be identified.</u> |
| D-6) ii. | Providing relevant stakeholders with information | <ul style="list-style-type: none"> - Provide information on safety, including technical characteristics of AI systems, mechanisms for ensuring safety, foreseeable risks, remedies against them, causes of failures, and status of actions against them. |
| D-7) i. | Explanation to AI providers of conformity to common guiding principles | <ul style="list-style-type: none"> - <u>Explain</u> to AI providers that the output quality of AI may significantly change and that <u>risks may arise as a result of this characteristic.</u> |
| D-7) ii. | Documentation of development-related information | <ul style="list-style-type: none"> - <u>Prepare documents</u> on your AI system development processes, data collection and labeling affecting decision-makings, algorithms you have used, and related matters. |
| D-10) i. | Contribution to creation of opportunities for innovation | <ul style="list-style-type: none"> - <u>Research and develop quality, reliability, and development methodologies.</u> - Contribute to the <u>maintenance of the sustainable economic growth and the provision of solutions for social challenges.</u> - Promote internationalization, diversification, and collaboration among industry, academia, and government sectors, including watching trends in international arguments, such as DFFT, and joining AI developer communities and academic societies. - <u>Provide all of society with information</u> about AI. |

See "Hiroshima Process International Code of Conduct for Organizations Developing Advanced AI Systems" for AI developers who develop advanced AI systems, including the most advanced foundation models and generative AI systems.

- It is important for AI providers to provide AI systems and services on the precondition that AI is operated and proper use

When implementing an AI system

- | | | |
|----------|--|---|
| P-2) i. | Actions against risks that consider the lives, bodies, properties, and minds of humans and the environment | - Take measures to ensure proper performance under usage conditions, enabling the AI system to maintain those performances in various situations, and <u>minimizing risks</u> . |
| P-2) ii. | Provision contributing to proper use (of AI) | - Use AI within the expected scope of use set by AI developers
- Guarantee the accuracy and while examining how AI usage environments of the AI system or service differ from those that AI developers expect. |
| P-3) i. | Considerations for bias in configurations or data of AI systems and services | - Guarantee fairness of data and <u>examine bias</u> contained in referenced information and collaborating external services.
- <u>Regularly evaluate</u> inputs/outputs of AI models and <u>rationales of decisions</u> made by AI models to monitor for any bias generated.
- Examine the possibility where bias may be introduced that arbitrarily restricts decisions made by users receiving AI output results. |
| P-4) i. | Deployment of mechanisms and measures for protecting privacy | - <u>Take privacy protection measures</u> by, for example, introducing a mechanism that appropriately manages and restricts access to personal data based on the characteristics of the adopted technologies (privacy by design). |
| P-5) i. | Deployment of mechanisms for security measures | - <u>Take security measures appropriately</u> based on the characteristics of the adopted technologies (security by design). |
| P-6) i. | Documentation of systems architectures and the like | - <u>Prepare documents</u> describing the system architecture and data processing of the provided AI system or service that influences the decision-making. |

- It is important for AI providers to provide AI systems and services on the precondition that AI is operated and proper use

After an AI system or service starts to be provided

P-2) ii. Provision contributing to proper use (of AI)	- Periodically verify whether the AI system or service is used for <u>proper purposes</u> .
P-4) ii. Countermeasures against privacy violation	- <u>Collect information on privacy violation in AI systems and services</u> , properly handle violations if there are any, and <u>take the necessary steps to prevent a recurrence</u> .
P-5) ii. Handling vulnerabilities	- <u>Identify trends in the latest risks and matters requiring attention</u> in each provision step and make efforts to <u>eliminate vulnerabilities as well</u> .
P-6) ii. Providing relevant stakeholders with information	<ul style="list-style-type: none"> - Provide information on safety, including technical characteristics of AI systems, mechanisms for ensuring safety, foreseeable risks, remedies against them, possibility of changes in output or programs, causes of failures, and status of actions against them, incidents, policies on collecting data learned by AI models, how AI models learn the data, and the system for implementing the learning. - <u>Provide information and explanations on the fact that AI is used, appropriate/inappropriate use methods, and the details of and information on reasons for the update</u> in light of the nature of AI and the purpose of its use, etc.
P-7) i. Explanation to AI business users of conformity to common guiding principles	<ul style="list-style-type: none"> - <u>Encourage AI business users to use AI properly</u>, and call attention to the use of data for which accuracy, and recency as necessary are guaranteed, and to the learning of inappropriate AI models during in-context learning, as well as providing <u>precautions for when inputting personal data</u>. - Call attention to inappropriate input of personal data into the AI systems and services.
P-7) ii. Documentation of service agreements or the like	- <u>Compile service agreements</u> for AI users and <u>present privacy policies</u> .

AI providers handling advanced AI systems should comply with the “D. Common Guiding Principles for AI business actors involved in advanced AI systems” in Part 2, I. to XI. within the appropriate scope and comply with XII.

- It is important that AI business users always use properly within the scope of use set by the AI providers and, as necessary, operate the AI systems, and expected to learn the necessary insights to use AI more effectively

When using AI systems and services

- | | | |
|----------|---|--|
| U-2) i. | Proper use (of AI) that considers safety | <ul style="list-style-type: none"> - Conform to instructions for use specified by AI providers and use <u>within the expected scope of use set by AI providers during the design process.</u> - Understand the degrees of precision and risks of AI output and use AI output <u>after confirming various risk factors,</u> |
| U-3) i. | Consideration for bias in input data or prompt | <ul style="list-style-type: none"> - Input data for which fairness is guaranteed, pay attention to bias in prompts, and be responsible for <u>determining whether to use AI output results for business.</u> |
| U-4) i. | Countermeasures against inappropriate input of personal data and privacy violations | <ul style="list-style-type: none"> - Refrain from improperly inputting personal data to AI systems and services. - <u>Collect information on privacy violations in AI systems and services properly</u> and take the necessary steps to prevent violations. |
| U-5) i. | Implementation of security measures | <ul style="list-style-type: none"> - <u>Conform to instructions for security</u> specified by AI providers. |
| U-6) i. | Providing relevant stakeholders with information | <ul style="list-style-type: none"> - Input data for which fairness is guaranteed and pay attention to bias in prompts when obtaining the output. <u>When using the output result for business decision-making, inform the relevant stakeholders about this.</u> |
| U-7) i. | Explanation to relevant stakeholders | <ul style="list-style-type: none"> - <u>Provide information in a plain and accessible manner to relevant stakeholders</u> in advance how to provide the data and its format based on the characteristics and use purposes of AI, contact points with them, privacy policies, and the like. - If intending to use the AI output result as a reference for an evaluation of a specific individual or group, make a reasonable judgment by humans. - <u>Set up a help desk that handles inquiries from relevant stakeholders</u> to give explanations and receive requests in cooperation with the AI providers. |
| U-7) ii. | Effective use of provided documents and conformity to agreements | <ul style="list-style-type: none"> - Properly <u>store and use the documents</u> about the AI systems and services provided by the AI providers. - <u>Conform to the service agreements</u> specified by the AI providers. |

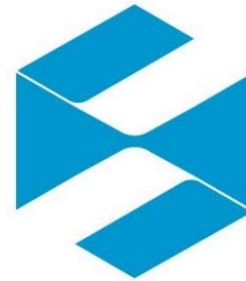
AI business users handling advanced AI systems should comply with the “D. Common Guiding Principles for businesses involved in advanced AI systems” in Part 2, I. to XI. within the appropriate scope and comply with XII.



MIC

総務省

Ministry of Internal Affairs
and Communications



経済産業省

Ministry of Economy, Trade and Industry