Tentative
Translation

Report 2022 (Overview)

Further Promotion of
“Safe, Secure, and Trustworthy Implementation of AI in Society”

25 July 2022

The Conference toward AI Network Society

Structure of Report 2022

Introduction

Chapter 1. Recent Trends in AI Networking

1. Domestic trends
2. Overseas trends
3. Trends in international discussions
4. Trends in international standardization
5. International symposium “Global Forum on AI Network Society 2022”

Chapter 2. Initiatives to Promote “Safe, Secure, and Trustworthy Implementation of AI in Society” (Compilation of case studies of initiatives)

1. Overview of hearings
2. Key points for presentations and exchanges of opinions at hearings

Chapter 3. Reviews of the *AI R&D Guidelines and the *AI Utilization Guidelines***

1. Reviews of the *AI R&D Guidelines* and the *AI Utilization Guidelines*
2. Arrangement of discussion points regarding revisions of the *AI R&D Guidelines* and the *AI Utilization Guidelines*

Chapter 4. Future Initiatives

1. AI ethics and governance
2. Disseminating and sharing case studies of initiatives
3. Human resources development

*Formally known as the *Draft AI R&D Guidelines for International Discussion*

Conclusion

Attachment 1: Outline of the international symposium “Global Forum on AI Network Society 2022”

Attachment 2: Details of the Reviews of the *AI R&D Guidelines* and the *AI Utilization Guidelines*

Attachment 3: List of Members of the Conference toward AI Network Society and the AI Governance Review Committee

Supplementary Volume: Collection of Good Practices for “Safe, Secure, and Trustworthy Implementation of AI in Society” (Compilation of case studies)

Chapter 1. Recent Trends in AI Networking (1/2)

Domestic trends

Note: List of key items

○ ***AI Strategy 2022*** **(April 22, 2022 decision by the Integrated Innovation Strategy Promotion Council)**

- Major changes from the previous year's *AI Strategy 2021*:
  - ➢ Sets out new targets and initiatives to enhance AI implementation in society
  - ➢ Specifies concrete initiatives to cope with imminent crises such as pandemics and large-scale disasters

○ ***Guidebook on AI-based Cloud Services*** **(MIC, February 15, 2022)**

- The Guidebook compiles matters that should be taken into consideration when developing and providing AI-based cloud services. The Guidebook is based on the comprehensive studies of the AI Cloud Services Study Group, which comprised academic experts, lawyers, business operators, and other knowledgeable people.
  - ➢ Systematically organizes specific procedures for the development of AI-based cloud services and compiles matters for developers to consider in each process
  - ➢ Proposes voluntary measures, together with specific examples, that business operators are recommended to follow when providing AI-based cloud services to users

Overseas trends

○ **European Parliament adopts the final recommendations of the Special committee on artificial intelligence in a Digital Age (May 3, 2022)**

- The final report recommends that AI should not be regulated as a technology and that the type, intensity, and timing of regulatory intervention should be examined in accordance with the level of risk associated with specific applications of AI systems. It further recommends that public debate on how to explore the enormous potential of AI should be promoted based on the principle that AI and robotics are human-centric and complementary to humans, along with fundamental European values such as transparency, explainability and fairness.
- It proposes that democratic nations of the same mindset should cooperate and form an international discussion toward a global agreement on common standards for the responsible use of AI.

○ **U.S. announces the *AI Risk Management Framework*: Initial Draft (March 17, 2022)**

- The National Institute of Standards and Technology (NIST) announced the initial draft of the *AI Risk Management Framework* for the purpose of promoting the development and use of trustworthy and responsible AI. NIST is developing the Framework with comments and input from AI-related communities and took feedback on the initial draft until April 29, 2022. It plans to release a second draft in the summer or fall of 2022 and publish Version 1.0 of the official Framework by the winter.
  - ➢ Created to help business operators and society better understand the risks associated with AI systems and to support risk management
  - ➢ Categories the characteristics of risks associated with AI and organizes four functions (mapping, measurement, management, and governance) necessary for risk management

Chapter 1. Recent Trends in AI Networking (2/2)

Trends in international discussions

Note: List of key items

○ OECD Committee on Digital Economy Policy (CDEP) (December 2, 2021) and first meeting of the Working Party on Artificial Intelligence Governance (May 24 – 25, 2022)

- At the meetings, Japan's initiatives were given by MIC, which presented *Report 2021* by this Conference, a summary of specific case studies of initiatives and good practices based on hearings with business operators and others, and by METI, which presented the *Governance Guidelines for Implementation of AI Principles*, which includes reference case studies.
- It was agreed to transfer the Working Group on AI Policies, one of the three working groups of the informal OECD Network of Experts on AI (ONE AI), to the Working Party on Artificial Intelligence Governance (WP AIGO), a CDEP working party, and to extend the operations of ONE AI by one year.

→ At the first meeting of WP AIGO, the Chair and Vice-Chairs (including a representative from Japan) for FY2022 were approved, AI governance policies from member countries were presented and shared, and reports were given on the activities and progress of each of the ONE AI working groups, followed by discussions on these topics.

○ Global Partnership on Artificial Intelligence (GPAI) Second Plenary Meeting (November 11 – 12, 2021)

- At the Ministerial Council meeting, Japan's appointment as the organization's incoming council co-chair (for the fall 2022–fall 2023 term) was approved.
- Approved Belgium, Czech Republic, Denmark, Ireland, Israel, and Sweden as new members and the participation of UNESCO as an observer from 2022 onward.

○ United Nations Educational, Scientific and Cultural Organization (UNESCO) adopted the Recommendation on the Ethics of Artificial Intelligence (November 9 – 24, 2021)

- The Recommendation defines values and principles (to be respected by all actors in AI system life cycles) and areas of policy action (areas where Member States are to put in place effective measures based on the Recommendation) and presents the need for monitoring and evaluation of AI ethics.

Values: 1. Respect, protection, and promotion of human rights and fundamental freedoms and human dignity; 2. Environment and ecosystem flourishing; 3. Ensuring diversity and inclusiveness; 4. Living in peaceful, just and interconnected societies

Principles: 1. Proportionality and Do No Harm; 2. Safety and security; 3. Fairness and non-discrimination; 4. Sustainability; 5. Right to Privacy, and Data Protection; 6. Human oversight and determination; 7. Transparency and explainability; 8. Responsibility and accountability; 9. Awareness and literacy; 10. Multi-stakeholder and adaptive governance and collaboration

Global Forum on AI Network Society 2022

○ International symposium "Global Forum on AI Network Society 2022" (March 1, 2022)

MIC held the Global Forum on AI Network Society 2022 with the aim of contributing to the solution of social issues related to AI through discussions on the use and distribution of data, which is key to the widespread adoption and use of AI, and through discussions on approaches to AI regulation and governance in light of international trends. Participants in the Forum included members of this Conference and the AI Governance Review Committee as well as experts and business leaders from a broad range of fields in Japan and overseas, and engaged in exchanges of ideas and opinions. (See Attachment 1 for details)

Chapter 2. Initiatives to Promote "Safe, Secure, and Trustworthy Implementation of AI in Society"

(Compilation of case studies of initiatives) (1/5)

Opinions were exchanged through presentations by business operators who are making advanced or ambitious efforts to implement AI in society. The main issues were:

● What measures are developers and users (AI service providers and business users) taking to promote "Safe, Secure, and Trustworthy Implementation of AI in Society" or to increase acceptance of AI by society?

● What issues do business operators and others face in attempting to advance these measures, and what should be done to resolve these issues?

● What kind of environment should be created to increase acceptance of AI by society and promote "Safe, Secure, and Trustworthy Implementation of AI in Society"?

Some of the hearings looked more specifically at human resources development and AI supply chains.

○ SoftBank Corp.	: Examples of AI use at SoftBank
○ Panasonic Corporation	: Panasonic's approach to AI development and application examples
○ Sharp Corporation, AIoT Cloud Inc.	: AI initiatives at the Sharp Group
○ Prof. Keiichi Nakayama (Medical Institute of Bioregulation, Kyushu University)	: Application of artificial intelligence (AI) to medical biology
○ Prof. Takuya Ueda (Clinical AI Human Resources Development Program, Tohoku University)	: Clinical AI: An advanced AI R&D and human resources development center aiming to solve Global x Local medical issues
○ Mercari, Inc.	: Description of AI efforts by the Mercari Group
○ Japan Data Management Consortium	: Ethical framework for AI and data use conceived by practitioners
○ Mitsui Sumitomo Insurance Company, Limited	: Initiatives to promote digitalization at Mitsui Sumitomo Insurance
○ DAIKIN INDUSTRIES, LTD.	: AI human resources development efforts at Daikin Industries
○ West Japan Railway Company	: Data analytics initiatives at JR West
○ KDDI CORPORATION	: AI use cases and AI governance initiatives at KDDI
○ KPMG AZSA LLC	: Initiatives toward verifying the appropriateness of AI applications
○ NTT DATA Corporation	: Supply chain risks associated with AI and NTT Data's responses
○ FUJITSU LIMITED	: Fujitsu's AI supply chain initiatives
○ AI Data Consortium	: Data use, issues, and initiatives

Note: Company or organization names at the time of the hearings

Chapter 2. Initiatives to Promote "Safe, Secure, and Trustworthy Implementation of AI in Society"

(Compilation of case studies of initiatives) (2/5)

Initiatives Related to AI Ethics and Governance

Initiatives related to AI ethics and governance by each business operator, etc. are organized from the perspectives of "guidelines and principles," "organization and structure," "security," "privacy," "fairness," "transparency and accountability," "appropriate use," "quality assurance and development review," and "cooperation and collaboration with external parties".

Good Practices

	Standard initiatives that can be used as references	Particularly noteworthy initiatives
Policies, guidelines and principles	AI developers and service providers are seen taking initiatives to formulate and establish principles targeting AI ethics and governance with reference to the *AI R&D Guidelines*, *AI Utilization Guidelines*, and other guides.	
	⇒ It is desirable to formulate and establish guidelines, principles, and other precepts, and to implement and operate them appropriately, in accordance with the mode of AI usage and the nature of AI, while referring to the case studies presented at the hearings.	
Organizations and structures	AI developers and service providers are seen taking initiatives to set up ethics committees or to establish new dedicated departments reporting directly to the president in order to implement and operate AI ethics and governance.	
	⇒ It is desirable to construct organizations and structures to implement and operate AI ethics and governance appropriately while referring to the case studies presented at the hearings.	
Security Privacy Fairness	Initiatives are seen prioritizing security assurance, privacy protection, assurance of fairness, and elimination of biases.	In the area of image recognition, attention should be paid to initiatives that identify individuals in a form that does not leave any personal information on the AI side, such as performing the image processing on the edge side and uploading only the recognition results to the cloud or estimating just the skeletal structure of the individual.
	⇒ It is desirable to undertake appropriate initiatives that perform security assurance, privacy protection, assurance of fairness, and elimination of biases, in accordance with the mode of AI usage and the nature of AI, while referring to the case studies presented at the hearings.	
Transparency and accountability	Initiatives are seen prioritizing accountability and the assurance of transparency, such as making AI decision logic explainable or tracing past conditions.	Attention should be paid to initiatives to provide accountability to end users by clearly stating that a website's purpose of using personal information is to develop technology.
	⇒ It is desirable to undertake appropriate initiatives that ensure transparency and provide accountability, in accordance with the mode of AI usage and the nature of AI, while referring to the case studies presented at the hearings.	

Note: As *Report 2021* also presents several specific initiative case studies, referring to those examples with the examples presented here is considered helpful.

Chapter 2. Initiatives to Promote "Safe, Secure, and Trustworthy Implementation of AI in Society"

(Compilation of case studies of initiatives) (3/5)

Initiatives Related to AI Ethics and Governance

Key points of Good Practices, continued

	Standard initiatives that can be used as references	Particularly noteworthy initiatives
Appropriate use	Many business operators and others position AI as a tool for humans to use, and implement the so-called "human in the loop" operation, in which humans ultimately confirm the AI judgement and have AI as a support, rather than leaving everything to AI judgement.	
	⇒ "Human in the loop" is closely related to privacy, fairness, transparency, and accountability, and is a very important point of view when considering the relationship between humans and AI. While referring to the case studies presented in the hearings, it is desirable to appropriately utilize the mechanism to realize "human in the loop" depending on the mode of AI utilization and the nature of AI.	
Quality assurance and development reviews	Business operators and others are taking initiatives to specify in-house processes and rules to implement AI quality assurance and evaluations.	Attention should be paid to initiatives in which third parties provide evaluation mechanisms (checklists, frameworks) and monitoring services.
	⇒ Initiatives that use third-party evaluations and monitoring are considered useful and effective measures to implement security assurance, privacy protection, assurance of fairness, and elimination of biases. It is desirable to proceed with initiatives that implement quality assurance and evaluations appropriately, while referring to the case studies presented at the hearings. In cases where it is difficult for a business operator to build such mechanisms on its own, it is expected business operators will make use of mechanisms and services provided by external parties.	
Cooperation and collaboration with external parties	Initiatives are seen where business operators and others cooperate and collaborate with governments, relevant organizations, and outside experts and specialists, including those from overseas.	
	⇒ Cooperation and collaboration with external stakeholders is believed to help further business operators and others' own initiatives and promote implementation of AI in society. It is desirable to actively strive for cooperation and collaboration with external parties and to promote multi-stakeholder implementation of AI in society, while referring to the case studies presented at the hearings.	

Note: As *Report 2021* also presents several specific initiative case studies, referring to those examples with the examples presented here is considered helpful.

Chapter 2. Initiatives to Promote "Safe, Secure, and Trustworthy Implementation of AI in Society"
(Compilation of case studies of initiatives) (4/5)

Initiatives Related to AI Development and Utilization

Among the initiatives for AI development and utilization taken by each business operator, we have organized efforts related to "countermeasures against COVID-19," "medical and healthcare," and "the elderly and people with disabilities" as fields that require special attention, given that the spread of COVID-19 continues, and that it is important to disseminate information internationally as a country that has faced issues in advance.

Key points of Good Practices

<table>
<tr><th></th><th>Standard initiatives that can be used as references</th><th>Particularly noteworthy initiatives</th></tr>
<tr><td rowspan="2">COVID-19 measures</td><td>Initiatives are seen by business operators and others to apply AI to COVID-19 measures, such as visualizing overcrowded conditions and implementing non-face-to-face / non-contact deliveries.</td><td></td></tr>
<tr><td colspan="2">⇒ It is desirable for AI developers, users, and others to continue initiatives applying AI to COVID-19 measures, based on the status of COVID-19 infections, while referring to the case studies presented at the hearings.</td></tr>
<tr><td rowspan="2">Medicine and healthcare</td><td></td><td>Attention should be paid to initiatives into accurate estimation of cancer prognoses and into drug discovery as well as initiatives in human resource development connected to medicine, AI, and data science where design thinking is incorporated.</td></tr>
<tr><td colspan="2">⇒ It is desirable to promote AI development, applications, and other initiatives in the medicine and healthcare fields, while referring to the case studies presented at the hearings. The advancement of such initiatives is expected to bring benefits to medical professionals, patients, and other stakeholders.</td></tr>
<tr><td rowspan="2">The elderly and people with disabilities</td><td></td><td>Attention should be paid to the many business operators and others who take initiatives to apply AI to support elderly people and people with disabilities, such as communication with the hearing impaired, monitoring and providing walking assistance at elderly care facilities, and early detection of people requiring assistance at railroad stations.</td></tr>
<tr><td colspan="2">⇒ It is desirable to promote initiatives to use AI to support the elderly and people with disabilities, while referring to the case studies presented in the hearings.</td></tr>
</table>

Note: As *Report 2021* also presents several specific initiative case studies, referring to those examples with the examples presented here is considered helpful.

Chapter 2. Initiatives to Promote "Safe, Secure, and Trustworthy Implementation of AI in Society"
(Compilation of case studies of initiatives) (5/5)

Initiatives Related to Human Resources Development

As the shortage of human resources related to AI has been pointed out, and the training and securing of human resources has become an issue, we summarize the initiatives taken by each business operator regarding human resource training.

Key points of Good Practices

Standard initiatives that can be used as references	Particularly noteworthy initiatives
Business operators and others are engaged in initiatives and working with universities to develop graded curriculum and provide training divided into finely differentiated levels with the aim of developing human resources who understand AI and IoT in addition to their technical specialties.	Attention should be paid to initiatives by business operators that provide AI literacy education to people outside the company (for high school and elementary school students). The advancement of such initiatives is expected to raise AI literacy levels.
⇒ Given that human resources development is a pressing issue, it is desirable that initiatives that train and secure AI talent be given priority, while referring to the case studies presented at the hearings.	

Initiatives Related to AI Supply Chains

AI development and application relies on numerous processes, such as data acquisition and processing, AI model construction, and system incorporation. The table below summarizes the risks, issues, and solutions for these AI supply chains.

Key points of Good Practices

Standard initiatives that can be used as references	Particularly noteworthy initiatives
AI developers and service providers are seen taking initiatives to check the veracity of data suppliers and to guarantee services through contracts.	Attention should be paid to initiatives building data distribution platforms. The establishment of such platforms will make it possible to visualize the source and provenance of data, which is expected to lead to trustworthy data distribution.
⇒ It is desirable for AI developers and service providers to take initiatives that identify AI supply-chain risks and prevent the actualization of the risks or, in the event the risks are actualized, to minimize any consequential damages, while referring to the case studies presented at the hearings.	

Note: As *Report 2021* also presents several specific initiative case studies, referring to those examples with the examples presented here is considered helpful.

Chapter 3. Reviews of the *AI R&D Guidelines* and the *AI Utilization Guidelines* (1/8)

Outline of the Reviews

In the area of AI ethics and governance, this Conference established the *AI R&D Guidelines* (July 2017) and the *AI Utilization Guidelines* (August 2019) and contributed to international discussions at the OECD, the G7, the G20, and other venues.

Following this, the Conference continued to follow domestic and international trends and directions of international discussions on AI ethics and governance while conducting hearings with specialists, developers, users (AI service providers, business users, and consumer-type users), and other stakeholders. The Conference compiled the initiatives taken by these stakeholders in *Report 2020* (July 2020) and *Report 2021* (August 2021).

- Both in Japan and abroad, many principles, policies, and guidelines have been established on AI ethics and governance (monitoring is especially needed of developments that have come after the establishment of the *AI R&D Guidelines* and the *AI Utilization Guidelines*).
- Some stakeholder initiatives presented at the hearings can be deemed outstanding initiatives toward "Safe, Secure, and Trustworthy Implementation of AI in Society", that go beyond the scope of the *AI R&D Guidelines* and the *AI Utilization Guidelines*.
- With the increasing application and social implementation of AI in Japan and abroad, there have been cases of risks materializing and incidents occurring.

The *AI R&D Guidelines* and the *AI Utilization Guidelines* were reviewed from the following perspectives.

1. Comparison with principles, policies, and guidelines instituted in other countries
2. Comparison with principles, policies, and guidelines instituted in Japan
3. Consistency with initiative case studies presented in the hearings
4. Consistency with case studies of AI development, application, and social implementation

It is important that the guidelines are consistent with such developments as the EU's *Proposed Artificial Intelligence Act*,[*1] so the reviews took this matter into account as well.

[*1] *Proposal for a Regulation laying down harmonised rules on Artificial Intelligence* (April 2021)

Chapter 3. Reviews of the *AI R&D Guidelines* and the *AI Utilization Guidelines* (2/8)

Verification through comparisons with principles, policies, and guidelines instituted in other countries (1/2)

- Surveyed countries and surveyed guidelines and policies
  - † The survey covered 16 countries[*2] (U.S., Canada, UK, Italy, the Netherlands, Sweden, Denmark, Germany, Norway, Finland, France, India, South Korea, Singapore, China, and Australia), the EU, and international organizations (the United Nations Educational, Scientific and Cultural Organization (UNESCO) and the World Health Organization (WHO))
  - † Surveyed 67 guidelines and other policies[*3] announced by government agencies, industry associations, and major global companies
  - ➢ Government agencies[*4] and international organizations (40), industry associations (10), global companies (17)

[*2] In selecting countries for the survey, the *Government AI Readiness Index*, the number of AI publications, the number of AI patent applications, and the number of AI researchers were considered, and countries ranked in the top 10 in any of these categories were selected. The EU and two international organizations were added to the survey, given overseas trends and directions of international discussions.
[*3] The survey generally concentrated on guidelines announced after August 2019, when the *AI Utilization Guidelines* were announced.
[*4] This includes the EU's European Commission and High-level expert group on artificial intelligence.

- Values that should be respected
  - † The survey identified 22 values that should be respected[*5]

1. Human-centered
2. Human dignity
3. Diversity, Inclusiveness
4. Sustainable society
5. International cooperation
6. Proper utilization
7. Education/Literacy
8. Human intervention, Controllability
9. Proper data
10. Collaboration among AI systems
11. Safety
12. Security
13. Privacy
14. Fairness, Equity, removal of Discrimination
15. Transparency, Explainability
16. Accountability
17. Robustness
18. Responsibility
19. Traceability
20. Monitoring, Auditing
21. Governance
22. Other (cost, effectiveness measurements)

[*5] Values 1 through 16 are the same values as those in the AI Guideline Comparison Table (Attachment 2, *Report 2019*). Values 17 to 22 are newly identified values.

Note: Taken from *Research Survey on the Implementation Status of Principles, Guidelines, etc. Associated with AI and Regulated Matters*, MIC. The values have been arranged here based on the survey, but this is not an exhaustive list.

Chapter 3. Reviews of the *AI R&D Guidelines* and the *AI Utilization Guidelines* (3/8)

Verification through comparisons with principles, policies, and guidelines instituted in other countries (2/2)

	No. of guidelines	Human-centered	Human dignity	Diversity, Inclusiveness	Sustainable society	International Cooperation	Proper utilization	Education / Literacy	Human intervention, Controllability	Proper data	Collaboration among AI systems	Safety	Security	Privacy	Fairness, Equity, removal of Discrimination	Transparency, Explainability	Accountability	Robustness	Responsibility	Traceability	Monitoring, Auditing	Governance	Other
U.S.	6	1	2			1	1		1	1		3	3	1	4	5	4	1	2	2	3		
Canada	1							1								1							1
UK	6		2		1			1		2		1	1	2	5	5	5	1	2		2		1
Italy	1									1				1		1							
The Netherlands	2		1				1		1	2				1	2	2	1		1	1			
Sweden	2		1	1			1	1					1	1	1	2							
Denmark	2		2	2			1		1			1	1	1	2	2	1	1			1		
Germany	3	1	2	2	3				1	2		2	2	2	2	3	2	3	1	1	1		
Norway	2		2		1		2		1			1		1	2	1	1	1		1	1		
Finland	2		2	1			1		1			1		1	1	2	1		2	2	1		
France	1								1						1	1	1				1		
India	1		1	1								1	1	1	1	1	1		1				
South Korea	1		1	1		1						1		1	1	1	1	1					
Singapore	2		1												2	2	1						
China	3		3	2	2	2	2	2	3	1		3	2	2	3	2	2	2		2	3	3	
Australia	1	1	1									1	1	1	1	1	1		1		1		
EU	2		2	1	2	1	2		1			2	1	2	2	1	2	2	2	1	1		
International organizations	2		2	2	2	1	2	1	1		1	2	1	2	1	2	1		1	1	1		
Totals	40	3	25	13	11	6	13	6	12	9	1	19	14	20	31	35	25	12	13	11	16	3	2

1. Statistics cover government agencies (including the EU's European Commission and High-level expert group on artificial intelligence) and international organizations.
2. Values mentioned in guidelines, etc. are colored in pink. Figures indicate the number of guidelines mentioning the particular value. Figures in red indicate that the value is mentioned in more than half of the country or organization's guidelines.

Chapter 3. Reviews of the *AI R&D Guidelines* and the *AI Utilization Guidelines* (4/8)

Verification through comparisons with principles, policies, and guidelines instituted in Japan (1/2)

- Surveyed guidelines and policies
  † The survey covered 22 guidelines and other policies announced by business operators and organizations in Japan.
- Values that should be respected
  † The survey identified 22 values that should be respected[*6]

1. Human-centered
2. Human dignity
3. Diversity, Inclusiveness
4. Sustainable society
5. International cooperation
6. Proper utilization
7. Education/Literacy
8. Human intervention, Controllability
9. Proper data
10. Collaboration among AI systems
11. Safety
12. Security
13. Privacy
14. Fairness, Equity, removal of Discrimination
15. Transparency, Explainability
16. Accountability
17. Robustness
18. Responsibility
19. Traceability
20. Monitoring, Auditing
21. Governance
22. Other (cost, effectiveness measurements)

[*6] Values 1 through 16 are the same values as those in the AI Guideline Comparison Table (Attachment 2, *Report 2019*). Values 17 to 22 are newly identified values. Regarding the 22nd value, in the comparison with guidelines instituted in other countries, ”effectiveness measurements” were identified, so the 22nd value was listed as Other (cost, effectiveness measurements). The comparison with guidelines instituted in Japan, however, did not find any instances of “effectiveness measurements”, so the 22nd value is listed here as Cost.

Note: Taken from *Research Survey on the Implementation Status of Principles, Guidelines, etc. Associated with AI and Regulated Matters*, MIC. The values have been arranged here based on the survey, but this is not an exhaustive list.

Chapter 3. Reviews of the *AI R&D Guidelines* and the *AI Utilization Guidelines* (5/8)

Verification through comparisons with principles, policies, and guidelines instituted in Japan (2/2)

- ABEJA, Inc.
- NTT DATA Corporation
- Oki Electric Industry Co., Ltd.
- Koozyt, Inc.
- KDDI CORPORATION
- KONICA MINOLTA, INC.
- J.Score Co., LTD.
- STADUIM Co., Ltd.
- Sony Group Corporation
- NEC Corporation
- Nihon Unisys, Ltd.
- Nomura Research Institute, Ltd.
- Hitachi, Ltd.
- FUJITSU LIMITED
- FUJIFILM Holdings Corporation
- Mitsubishi Research Institute, Inc.
- Mitsubishi Electric Corporation
- Recruit Co., Ltd. (Recruit Works Institute)
- The Japanese Society for Artificial Intelligence
- KEIDANREN (Japan Business Federation)
- People Analytics & HR Technology Association
- AI Cloud Services Study Group*

  *MIC announced a guideline based largely on the examinations of this Study Group.

Note 1: Company names at the time of the guideline announcements

1. Human-centered	5
2. Human dignity	17
3. Diversity, Inclusiveness	5
4. Sustainable society	9
5. International cooperation	1
6. Proper utilization	17
7. Education/ Literacy	14
8. Human intervention, Controllability	6
9. Proper data	7
10. Collaboration among AI systems	4
11. Safety	14

12. Security	19
13. Privacy	21
14. Fairness, Equity, Removal of Discrimination	21
15. Transparency, Explainability	20
16. Accountability	20
17. Robustness	3
18. Responsibility	3
19. Traceability	4
20. Monitoring, Auditing	3
21. Governance	2
22. Cost	1

Note 2: Values mentioned in a majority of the surveyed guidelines are displayed in red.

Note 3: It must be noted that expectations for which values should be emphasized, among the values that should be respected, vary depending on the industry or field of the respective business or organization, the purpose or application of AI, and other factors.

Chapter 3. Reviews of the *AI R&D Guidelines* and the *AI Utilization Guidelines* (6/8)

Review Results

The following matters were confirmed as a result of the verifications from each perspective.

Verification through comparisons with principles, policies, and guidelines instituted in other countries

- Other countries' guidelines incorporate "robustness", "responsibility", "traceability", and "monitoring, auditing" as new values to be respected.
- All countries and regions emphasized the values of "transparency, explainability", "fairness, equity, removal of discrimination", "accountability", and "human dignity".
- Among the existing values to be respected, differences[*7] were found in the details between the two guidelines under review and guidelines instituted in other countries.

Verification through comparisons with principles, policies, and guidelines instituted in Japan

- Guidelines instituted in Japan incorporate "robustness", "responsibility", "traceability", and "monitoring, auditing" as new values to be respected.
- The two guidelines and guidelines instituted in Japan generally use many of the same concepts.
- Nearly all businesses and organizations emphasized "privacy", "fairness, equity, removal of discrimination", "transparency, explainability", "accountability", and "security".

Verification of consistency with initiative case studies presented in the hearings

- Many business operators are taking action to address the principles in the two guidelines.
- In addition to these efforts, some outstanding initiatives toward "Safe, Secure, and Trustworthy Implementation of AI in Society", go beyond the scope of the two guidelines.

Verification of consistency with case studies of AI development, utilization, and social implementation

- In many cases where risks have actualized or incidents have occurred, the values (principles) given in the two guidelines in general cover the values that should have been respected. The problem is one of effectiveness — that is, how to implement the values.
- It is believed that, at the present time, there is no significant likelihood of risks actualizing or incidents occurring that are well beyond the scope of the two guidelines. What is important is ensuring the effectiveness of the values (principles) given in the current guidelines and studying measures to limit risks (preventing them from actualizing) while monitoring future trends in AI development, application, and social implementation.

[*7] These differences should be taken into consideration when examining guideline revisions.

Chapter 3. Reviews of the *AI R&D Guidelines* and the *AI Utilization Guidelines* (7/8)

Discussion points regarding revisions to the *AI R&D Guidelines* and the *AI Utilization Guidelines* (1/2)

- Based on the reviews so far and the opinions and recommendations of this Conference, revisions of the *AI R&D Guidelines* and the *AI Utilization Guidelines* will be studied with aim of making the guidelines more useful and beneficial for AI developers and users and to further promote "Safe, Secure, and Trustworthy Implementation of AI in Society". Discussion points pertaining to the revisions are summarized here.

  Four discussion points have been arranged to aid guideline revision considerations, based on the opinions of this Conference, AI trends around the world, and directions of international discussions.

  The revision considerations will be based on the importance of ensuring the overall effectiveness of the principles in the guidelines by combining various initiatives such as cross-sectorial governance, sector-specific governance, and rule-making through standardization.

  ⇒ Guideline revision considerations will proceed based on the discussion points arranged here, the opinions and recommendations of this Conference, and the following two perspectives.
  - Whether to add to or change the purpose, basic philosophies, or the principles on the basis of the current approach of the two guidelines.
  - Whether to examine the very approach of the two guidelines without regard for their current approach, and whether to unify the guidelines with measures to ensure the effectiveness of the principles instead of having standalone guidelines.

Note: These two perspectives are not necessarily contradictory and can be compatible with each other.

Discussion points on guideline revisions

- Should there be additions to or revisions of the guidelines' purpose, basic philosophies, the target scope (definition) of AI, the principles, explanations, etc.? Should the guidelines' structure be revised to incorporate new material in line with progressive case studies and other circumstances?

Specific examples:

➢ Should the concepts of diversity and sustainability be added to the basic philosophies of the *AI R&D Guidelines*?

➢ Should "robustness", "responsibility", "traceability", and "monitoring, auditing" be added as new principles? If these principles are added, how can their addition be squared with the argument made during the previous process of establishing the guidelines that the principles be kept to the smallest number possible?

➢ Should the explanations of the principles already specified describe specific development / utilization situations or scenarios?

➢ Should revisions be premised on potential emergencies (such as pandemics or natural disasters)? (For example, how should the protection and promotion of public health be balanced with privacy protections?)

➢ In addition, should a guideline be compiled that organizes matters expected to be considered from the perspectives of quality assurance and management, supply chains, organizations and structures, human resources development, data handling, and balancing AI use with costs at administrative agencies?

Note: It is important that revision examinations ensure the guidelines continue to be technologically neutral, to not impede innovation, and to not place excessive costs on developers or users.

Chapter 3. Reviews of the *AI R&D Guidelines* and the *AI Utilization Guidelines* (8/8)

Discussion points regarding revisions to the *AI R&D Guidelines* and the *AI Utilization Guidelines* (2/2)

- Should the positioning or the names of the guidelines be changed?

Example:

➢ Should it be made clear that the guidelines are designed to be used both in Japan, to encourage appropriate initiatives by business operators and others in Japan (references when establishing guidelines or policies), and overseas, to engage in international discussions. In this case, should the name of the *AI R&D Guidelines* (*Draft AI R&D Guidelines for International Discussions*) be changed?

- What kinds of initiatives should be promoted, given that AI development hurdles are becoming lower, AI is becoming more familiar to users (there are more applications where users do not (or can not) recognize AI has been implemented), and the boundaries between AI development and utilization are becoming increasingly vague?

Examples:

➢ Should something be done (such as adding simple explanations) to address developers who are less specialized that previously assumed and users who are not aware of AI's usage (especially consumer-like users)?

➢ Should the *AI R&D Guidelines* and the *AI Utilization Guidelines* be integrated?

- What kinds of initiatives should be promoted in the overall governance framework, which includes the guidelines, to ensure the effectiveness of the guidelines moving forward?

Examples:

➢ What kinds of initiatives should be promoted in the overall governance framework that includes measures to ensure the effectiveness of the principles in the guidelines — some conceivable initiatives are initiatives by business operators and industries, initiatives that involve the government (administrative agencies), or the establishment of check sheets or certification systems?

References: *Information Disclosure Guidelines for Safety and Reliability of Cloud Services Using AI (ASP/SaaS Edition)*

Certification system for information disclosures pertaining to the safety and trustworthiness of ASP and SaaS (AI cloud services)

➢ What actions will be necessary in relation to international governance frameworks, given the EU's announcement of the *Proposed Artificial Intelligence Act* and the directions of discussions in this area?

Note: It is important that revision examinations ensure the guidelines continue to be technologically neutral, to not impede innovation, and to not place excessive costs on developers or users.

Chapter 4. Future Initiatives

AI Ethics and Governance

- Revisions to the *AI R&D Guidelines* and the *AI Utilization Guidelines*
  - † Study revisions to the *AI R&D Guidelines* and the *AI Utilization Guidelines,* centered on the discussion points arranged in this Report
- Disseminating and sharing the *AI R&D Guidelines* and the *AI Utilization Guidelines*
  - † Disseminate and share case studies of initiatives by business operators and others while continuing to promote activities to disseminate and share the *AI R&D Guidelines* and the *AI Utilization Guidelines*
- Monitor trends in Japan and overseas and directions of international discussions, and provide information
  - † It is important to monitor trends in Japan and overseas and directions of international discussions and to provide information about case studies of initiatives by business operators and others at international discussion forums, such as multilateral forums and binational policy dialogues
    → Japan has a significant role to play, as Japan will assume the GPAI council co-chair from fall 2022 to fall 2023 and Japan is scheduled to host the G7 in 2023
- Initiatives to ensure the effectiveness of principles
  - † It is important to ensure the overall effectiveness of the principles in the guidelines by combining various initiatives such as initiatives by business operators themselves, cross-sectorial governance, sector-specific governance, and rule-making through standardization
    → In conjunction with studying guideline revisions, it is important to study initiatives to ensure the effectiveness of the principles, such as establishing check sheets or certification systems

Disseminating and sharing case studies of initiatives

- AI ethics and governance initiatives
  - † Given the importance of disseminating and sharing case studies of initiatives by business operators and others, promote activities to disseminate and share such case studies in cooperation with external stakeholders
- AI development and use initiatives
  - † Given the importance of disseminating and sharing case studies of initiatives by business operators and others, promote activities to disseminate and share such case studies in cooperation with external stakeholders

Human resources development

- It is important to continue promoting human resources development initiatives, with reference to the case studies presented in the hearings
- It is important to promote initiatives in cooperation with external educational and research institutions, etc., in order to deepen the initiatives of businesses and contribute to raising the level of society as a whole
- It is important to pursue initiatives on human resources development not only in the private sector, but also for employees of ministries and agencies and local governments