

Tentative
Translation

Report 2022

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

July 25, 2022

The Conference toward AI Network Society

Contents

Introduction.....	1
Chapter 1. Recent Trends in AI Networking.....	4
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 for Promoting the "Safe, Secure, and Trustworthy Implementation of AI in Society" (Compilation of case studies of initiatives)	28
1. Overview of interviews	
2. Key points for presentations and exchanges of opinions at interviews	
Chapter 3. Reviews of the AI R&D Guidelines and the AI Utilization Guidelines	48
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	62
1. AI ethics and governance	
2. Disseminating and sharing case studies of initiatives	
3. Human resources development	
Conclusion	68
Attachment 1: Outline of the international symposium "Global Forum on AI Network Society2022"	
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)	

Introduction

Artificial intelligence (AI) is increasingly being used in our lives. In corporate activities too, AI is increasingly used to perform tasks such as producing goods and providing services, and the implementation of AI in society is advancing.

The Ministry of Internal Affairs and Communications has held the Conference toward AI Network Society (hereinafter referred to as "this Conference") since October 2016 to deliberate on the social, economic, ethical, and legal issues related to AI networking. As a result of deliberations by this Conference, the "Draft AI R&D Guidelines for International Discussions" (hereinafter referred to as the "AI R&D Guidelines") was compiled and published in July 2017, the "Draft AI Utilization Principles" in July 2018, and the "AI Utilization Guidelines" in August 2019. Following that, under the theme "Safe, Secure, and Trustworthy Implementation of AI in Society"¹, this Conference (may also include the AI Governance Review Committee, and the same shall apply hereinafter) interviewed stakeholders² including AI developers, service providers, business users, and consumer-type users, compiled their initiatives and efforts, and published them as Report 2020 in July 2020, and Report 2021 in August 2021³.

In addition, in conjunction with the G7⁴ Information and Communication Technology Ministers' Meeting held in Takamatsu City, Kagawa Prefecture, Japan in April 2016, active international discussions took place at the G7 and the Organization for Economic Cooperation and Development (OECD)⁵. In May 2019, the OECD adopted the Recommendation of the Council on Artificial Intelligence, and in June of the same year, the G20⁶ adopted the G20 AI Principles. Since then, the OECD has been working on sharing information on initiatives related to AI and their social implementation. The Global Partnership on AI (GPAI)⁷ and the United Nations Educational, Scientific and Cultural Organization (hereinafter referred to as "UNESCO")⁸ have also been engaged in international discussions on the ethics and governance of AI.

Furthermore, in Japan, the Fifth Science and Technology Basic Plan (approved by the Cabinet

¹ Here, while it does depend on the subjective judgment of each individual, "safety" refers to a state in which the building of trust is required as a premise, and trust is built up while a socially agreed level of security is maintained. "Security" refers to a state in which it is objectively judged that there is no damage to life, body or property (described with reference to Chapter 2 of the report by the Ministry of Education, Culture, Sports, Science and Technology's Advisory Board on the Science and Technology Policy for Contributing to Building a Safety and Security based Society). In the principle of safety in the Draft AI R&D Guidelines for International Discussions and the AI Utilization Guidelines compiled by this Conference, it is stated that consideration shall be given so as not to harm the life, body or property of users or third parties.

² For information on stakeholder categories, see the "AI Utilization Guidelines."

³ In addition to summarizing initiatives based on interviews, "Report 2020" summarizes the outlook for the ecosystem that will be formed as AI networking advances, and "Report 2021" summarizes discussions on COVID-19 and AI utilization.

⁴ The G7 Summit is an annual international conference attended by the heads of state of the seven G7 countries, which are France, the United States, the United Kingdom, Germany, Japan, Italy, and Canada, and also the European Union (EU) (country names listed in order of chair country).

⁵ For recent trends, see Section 3. (1) in Chapter 1.

⁶ The G20 Summit (summit on finance and the world economy) is an annual international conference attended by the leaders of the G7 countries and Argentina, Australia, Brazil, China, India, Indonesia, Mexico, South Korea, Russia, Saudi Arabia, South Africa, Turkey and the European Union (EU) (countries listed in alphabetical order).

⁷ In order to realize the development and use of "responsible AI" based on a human-centered approach, this organization was established in June 2020 as an international public-private partnership consisting of governments, international organizations, industries, and experts that share common values. For recent trends, see Section 3. (2) in Chapter 1.

⁸ For recent trends, see Section 3. (5)-[1] in Chapter 1.

on January 22, 2016)⁹ calls for Society 5.0¹⁰, a human-centered society that achieves economic development and solutions to social issues through a system that highly integrates cyberspace and physical space, as the future society that our country should aim for, and to realize Society 5.0, the Sixth Science, Technology and Innovation Basic Plan (approved by the Cabinet on March 26, 2021)¹¹ calls for Japan to aim to transform itself into a society that creates a dynamic and ever-changing virtuous cycle, in which new value can be created by integrating cyberspace and physical space through the construction of digital twins in cyberspace using a wide range of high-quality data, and then actively using AI based on that to change the physical space and reproduce the results in cyberspace¹².

In light of this situation, even after publishing Report 2021, this Conference continued to interview stakeholders, including AI developers, service providers, and business users, in order to promote the "Safe, Secure, and Trustworthy Implementation of AI in Society," and collected and compiled case studies of initiatives related to AI implementation in society. In addition, several years have passed since the AI R&D Guidelines and the AI Utilization Guidelines were compiled, and during this time many principles, policies, and guidelines on AI ethics and governance have been formulated in Japan and overseas, and Report 2021, which goes beyond the scope of the two guidelines, shows what can be evaluated as excellent initiatives for the "Safe, Secure, and Trustworthy Implementation of AI in Society." In light of this, the AI R&D Guidelines and the AI Utilization Guidelines have been reviewed. In addition, in order to further deepen discussion on the fusion of cyberspace and physical space and digital twins, etc., interviews¹³ were held over two occasions, and in anticipation of the development and utilization of AI in the future, interviews¹⁴ were also held to deepen discussions regarding quantum computers and AI. Furthermore, three interviews¹⁵ were held to deepen research on the European Commission (EC) Proposal for a Regulation laying down harmonised rules on artificial intelligence published in April 2021.

This report provides an overview of domestic and international trends in AI networking and trends in international discussions, and introduces examples of initiatives collected through interviews with business operators, experts, and relevant organizations that are undertaking advanced or ambitious initiatives to implement AI in society. This report also reviews the AI R&D

⁹ Available at the website of the following URL.

<<https://www8.cao.go.jp/cstp/kihonkeikaku/5honbun.pdf>>

¹⁰ The Cyber-Physical System (CPS) is a system that combines cyberspace (virtual space) and physical space (real space).

¹¹ Available at the website of the following URL.

<<https://www8.cao.go.jp/cstp/kihonkeikaku/6honbun.pdf>>

¹² Regarding this point, as the ideal social image to be achieved through the advancement of AI networking, "Harmony between Physical Space and Cyberspace" is set forth as a basic philosophy to be followed in forming the "Wisdom Network Society (WINS)" proposed by this Conference. "By connecting physical space and cyberspace utilizing AI networks and achieving harmony between them, cooperation beyond the space between people, goods, and things can be realized." See pages 15 and 16 in the Interim Report (April 2016) and Note (43) in Report 2016 (June 2016) compiled by the AI Networking Study Group, the predecessor body to this Conference, and pages 7 and 8 in Report 2017 (July 2017), and pages 8 and 9 in Report 2018 (July 2018) compiled by this Conference.

¹³ Joint meeting of this Conference and AI Governance Review Committee (February 8, 2022, and April 27 of the same year).

¹⁴ Joint meeting of this Conference and AI Governance Review Committee (October 27, 2021). Presentation materials by Keisuke Fujii (Professor, Graduate School of Engineering Science, Osaka University) are available at the website of the following URL.

<https://www.soumu.go.jp/main_content/000775388.pdf>

¹⁵ Joint meeting of this Conference and AI Governance Review Committee (October 27, 2021, February 8, 2022 and April 27 of the same year). For details of the interviews, see Section 2. (Reference) in Chapter 1.

Guidelines and AI Utilization Guidelines, and summarizes issues related to the review and arranges future initiatives based on the review results.

Case studies of initiatives introduced in this report are listed in the following table.

○ 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	: Ethics 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) Names of business operators and others as of the time the interview was conducted.

In this report, we have been able to summarize multiple concrete case studies related to the social implementation of AI thanks to the cooperation of the business operators, experts, and relevant organizations listed in the table above. We expect that these case studies will be very useful and beneficial information for those engaged in or considering developing and utilizing AI.

We would like to express our gratitude to the business operators, experts, and relevant organizations for their cooperation, and hope that this report will be shared with many stakeholders to contribute to the promotion of initiatives for the development and utilization of AI, and further promote the "Safe, Secure, and Trustworthy Implementation of AI in Society."

Chapter 1. Recent Trends in AI Networking

This chapter provides an overview of domestic and international trends in AI networking and trends in international discussions that were referred to in discussions held by this Conference, with a focus on trends since the publication of Report 2021¹⁶.

1. Domestic trends

(1) AI Strategy 2022

Based on discussions by the New AI Strategy Study Group held from October 2021 to February 2022, the government decided on AI Strategy 2022¹⁷ at the 11th Integrated Innovation Strategy Promotion Council held on April 22 of the same year. 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 for handling imminent crises such as pandemics and large-scale disasters

[Targets for promoting social implementation]

- Improved trustworthiness of AI
- Enhancement of data that supports AI utilization
- Development of additional environments such as securing human resources
- Promotion of the utilization of AI by the government
- Integration of AI with fields where Japan's strengths are

[Targets for handling imminent crises]

- Establishment of "AI for National Resilience"
- Establishment of leadership in "AI for Global Resilience"
- Establishment of leadership in "Resilient and Responsible AI"

(2) Priority Policy Program for Realizing Digital Society

On June 7, 2022, the Cabinet approved the Priority Policy Program for Realizing Digital Society¹⁸. It describes initiatives for using AI in a wide range of fields in line with the spread of AI, for example, it states the following regarding "Accelerating Initiatives for the Social Implementation of AI."

- Toward further practical application of AI going forward, deep learning is positioned as an important field based on AI Strategy 2022, and efforts will be made with its implementation by companies in mind.¹⁹
- In implementing AI in society, it is necessary to promote the transformation of society to one where AI can be used effectively and safely (AI-ready society). Therefore, work will progress

¹⁶ For trends in AI networking prior to the publication of Report 2021, see Chapter 1 in Report 2021. Note that the descriptions provided in this chapter are not exhaustive, and contain the information to the extent that could be ascertained through surveys conducted by the Ministry of Internal Affairs and Communications, etc. that was used as a reference for discussions by this Conference.

¹⁷ Available at the website of the following URL.
<https://www8.cao.go.jp/cstp/ai/aistrategy2022_honbun.pdf>

¹⁸ Available at the website of the following URL.
<<https://www.digital.go.jp/policies/priority-policy-program/>>

¹⁹ Regarding this point, specific initiatives are described, including the development of technologies such as "accountable AI" for realizing "responsible AI," the development of the environment to enhance the data that supports AI utilization such as through reexamining data handling rules, and the strengthening of efforts for developing advanced AI human resources.

on the basis of the seven AI social principles concerning the social framework in an AI-ready society²⁰, which are summarized in "Social Principles of Human-Centric AI." In promoting the utilization of AI in government agencies, initiatives including promoting the adoption of general-purpose AI and strengthening the promotion system will be undertaken based on the aforementioned AI social principles.

(3) Guidebook on AI-based Cloud Services

On February 15, 2022, the Ministry of Internal Affairs and Communications published the Guidebook on AI-based Cloud Services²¹ that compiles matters that should be taken into consideration when developing and providing AI-based cloud services. This guidebook is based on the comprehensive studies of the AI Cloud Services Study Group comprised of experts such as academic experts, lawyers, business operators and others.

This guidebook systematically organizes specific procedures for the development of AI-based cloud services and compiles matters for developers to consider in each process. In addition, it proposes voluntary measures that business operators recommend users follow when providing AI-based cloud services, together with specific examples.

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

On February 15, 2022, the Ministry of Internal Affairs and Communications added the Information Disclosure Guidelines for Safety and Reliability of Cloud Services Using AI (ASP/SaaS Edition)²² to its existing Information Disclosure Guidelines for Safety and Reliability of Cloud Services²³.

In line with the increase in the provision of cloud services using AI, this guideline was newly added to the Information Disclosure Guidelines for Safety and Reliability of Cloud Services as a guideline for handling risks when cloud service providers such as ASP and SaaS provide cloud services using AI.

(5) Governance Guidelines for Implementation of AI Principles

The Ministry of Economy, Trade and Industry announced Governance Guidelines for Implementation of AI Principles Ver. 1.1²⁴ on January 28, 2022.

In order to support the implementation of the AI principles necessary to promote the social implementation of AI, these guidelines provide action targets that should be implemented by AI business operators and also provide examples of virtual actions corresponding to each

²⁰ The seven principles are [1] Human-Centric, [2] Education/Literacy, [3] Privacy Protection, [4] Ensuring Security, [5] Fair Competition, [6] Fairness, Accountability, and Transparency, and [7] Innovation.

²¹ Available at the website of the following URL.
<https://www.soumu.go.jp/main_content/000792669.pdf>

²² Available at the website of the following URL.
<https://www.soumu.go.jp/main_content/000792672.pdf>

²³ With the aim of making it easier for users to compare, evaluate, and select cloud services, the Ministry of Internal Affairs and Communications has so far formulated and published seven information disclosure guidelines, collectively referred to as the Information Disclosure Guidelines for Safety and Reliability of Cloud Services.

²⁴ Available at the website of the following URL.
<https://www.meti.go.jp/shingikai/mono_info_service/ai_shakai_jisso/pdf/20220128_1.pdf>
These guidelines were compiled based on the opinions expressed in Governance Guidelines for Implementation of AI Principles Ver. 1.0 published on July 9, 2021, and the discussions held by the Expert Group on How AI Principles Should be Implemented until September 15 of the same year.

action target and practical actions for assessing the discrepancy between the action targets and the AI governance goals.

(6) Precautions and recommendations for introducing AI ethics into corporate activities

On August 31, 2021, the AI Business Promotion Consortium, a general incorporated association, published a report titled Precautions and Recommendations for the Introduction of AI Ethics in Corporate Activities: Consideration Based on a Risk-Based Approach²⁵ that summarizes the risks involved in introducing AI ethics in corporate activities, precautions, and issues that companies face in addressing AI ethics issues.

The report makes the following recommendations on how AI ethics should be incorporated into corporate activities.

- Recommendation 1 ... We recommend that you establish a company policy for AI ethics.
- Recommendation 2 ... We recommend that you visualize both the value and the risks that AI creates.
- Recommendation 3 ... We recommend that you establish a common understanding with external stakeholders.

(7) Guidelines for Quality Assurance of AI-based Products and Services

The Consortium of Quality Assurance for Artificial-Intelligence-based Products and Services (QA4 AI Consortium) published the Guidelines for Quality Assurance of AI-based Products and Services 2021.09 version²⁶ on September 15, 2021.

These guidelines were created by revising the Guidelines for Quality Assurance of AI-based Products and Services 2020.08 version to provide common guidance for quality assurance of AI products²⁷ in order to prevent excessive expectations of AI technology by organizations and ensure appropriate use and timely release. The four main revisions are as follows:

- Chapter 3 Technical catalog, : Specific definition of fairness, etc., Added Revised
- Chapter 5 Generation : Test and design examples, Added systems, Revised
- Chapter 7 Industrial : Explanations of points of concern from the standpoint of whether there are legal or ethical issues in using AI products, Revised processes, Revised
- Chapter 8 Autonomous : Descriptions of autonomous driving-related standards, driving, Partially revised Modified

2. Overseas trends

(1) European Union (EU)

[1] High-level meeting on international initiatives in the field of AI

On September 14 and 15, 2021, the European Commission and the European Council held a high-level meeting on international initiatives in the field of AI in Slovenia, following the

²⁵ Available at the website of the following URL.

<https://aibpc.org/cp-bin/wordpress/wp-content/uploads/2021/08/AIBPC_企業活動にAI倫理を導入していく上での注意点と提言_リスクベース・アプローチを踏まえた検討_-20210831.pdf>

²⁶ Available at the website of the following URL.

<<https://www.qa4ai.jp/download/>>

The first edition was published in May 2019, and this is the third edition.

²⁷ The guidelines define "products and services that use AI technologies."

Commission's announcement of a policy package on AI (April 2021)

At the meeting, as means for translating AI policy into concrete activities, the international support initiative for a human-centered approach known as InTouchAI.eu was launched, and globalpolicy.AI was mentioned. The intention to make the European AI regulations the international standard was also indicated.

[2] Research on opportunities and challenges of AI in the cultural and creative sectors

Research on copyright and new technologies: Copyright data management and AI

On March 16, 2022, the European Commission published two reports titled Study on Opportunities and Challenges of Artificial Intelligence (AI) Technologies for the Cultural and Creative Sectors and Study on Copyright and New Technologies - Copyright Data Management and Artificial Intelligence to promote cultural and creative activities over 10 years of digitization²⁸.

The former discusses the potential for utilizing AI in the distribution of European cultural content (architecture, publishing, film, music, news) and issues that could hinder this, such as language diversity, over-dependence on AI, and changes in employment structure, and also recommends that system interoperability and education related to AI skills should be new areas of funding to address these issues.

The latter covers issues related to copyright data management and states the importance of rights metadata, while acknowledging the role of technology in rights metadata management, licensing, rewards and distribution in the cultural and creative fields, and also presents measures for increasing the interoperability of rights metadata between domains.

[3] Adoption of resolution by the European Parliament

On October 6, 2021, the European Parliament adopted a resolution banning law enforcement from using facial recognition and anti-crime technology on the assertion that human oversight and strong legal authority limits to prevent AI from being discriminatory are required, particularly in situations such as law enforcement investigations and border checkpoints, on the basis that bias exists in AI algorithms²⁹ (377 in favor, 248 against, 62 abstentions).

On May 3, 2022, it also adopted the final recommendations of the Special Committee on Artificial Intelligence in a Digital Age (AIDA)³⁰ (495 in favor, 34 against, 102 abstentions).

This 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

²⁸ Study on Opportunities and Challenges of Artificial Intelligence (AI) Technologies for the Cultural and Creative Sectors and Study on Copyright and New Technologies
Available at the website of the following URL.

<<https://digital-strategy.ec.europa.eu/en/library/study-opportunities-and-challenges-artificial-intelligence-ai-technologies-cultural-and-creative>> and
<<https://digital-strategy.ec.europa.eu/en/library/study-copyright-and-new-technologies>>

²⁹ A European Parliament press release on this matter is available at the website of the following URL.
<<https://www.europarl.europa.eu/news/en/press-room/20210930IPR13925/use-of-artificial-intelligence-by-the-police-meps-oppose-mass-surveillance>>

³⁰ A European Parliament press release on this matter is available at the website of the following URL.
<<https://www.europarl.europa.eu/news/pt/press-room/20220429IPR28228/artificial-intelligence-meps-want-the-eu-to-be-a-global-standard-setter>>.

The adopted documents are available at the website of the following URL.

<https://www.europarl.europa.eu/doceo/document/TA-9-2022-0140_EN.html>

AIDA published the draft Artificial Intelligence in a Digital Age on November 9, 2021 and adopted it as their final report on March 22, 2022.

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. Furthermore, it proposes that democratic nations of the same mindset should cooperate and form international discussions toward a global agreement on common standards for the responsible use of AI.

[4] Securing machine learning algorithms

On December 14, 2021, the European Union Agency for Cybersecurity, the Union's agency dedicated to cybersecurity, published a report titled Securing Machine Learning Algorithms that analyzes the classification of machine learning algorithms and the threats targeting machine learning systems³¹.

This report presents a taxonomy for machine learning algorithms and analyzes security management for machine learning. It also identifies six threats to machine learning and recommends 37 security management items. However, with less than half (16 items) of the items being machine-learning-specific management items, it states that security management for machine learning should be implemented along with conventional cybersecurity management.

[5] Guidelines 05/2022 on the use of facial recognition technology in the area of law enforcement

On May 12, 2022, the European Data Protection Board adopted the Guidelines 05/2022 on the use of facial recognition technology in the area of law enforcement, which are aimed at law makers and law enforcement agencies of the EU and its Member States.³² Comments were accepted from May 16 to June 27 of the same year.

The guidelines state that the use of facial recognition technology in the following specific cases should be prohibited.

- Remote biometric authentication of individuals in public places
- AI facial recognition system that classifies individuals based on ethnicity, gender, political or sexual preference, or other discriminatory factors
- Facial recognition or similar techniques for inferring feelings of natural persons
- Processing of personal data in the context of law enforcement that relies on a database of personal data aggregated in a large, non-discriminatory manner (scraping of photographs and mug shots accessible online)

(2) France

On November 8, 2021, the French government announced its National AI Strategy that focuses on fostering and attracting AI human resources as the second phase of its AI strategy announced in March 2018³³.

³¹ "Securing Machine Learning Algorithms"

Available at the website of the following URL.

<<https://www.enisa.europa.eu/publications/securing-machine-learning-algorithms>>

³² "Guidelines 05/2022 on the use of facial recognition technology in the area of law enforcement"

Available at the website of the following URL.

<https://edpb.europa.eu/our-work-tools/documents/public-consultations/2022/guidelines-052022-use-facial-recognition_en>

³³ "STRATÉGIE NATIONALE POUR L'INTELLIGENCE ARTIFICIELLE – 2e phase"

Unlike the first phase, which focused on research and development, this strategy also focuses on human resources development and the commercialization and deployment of technologies, with public investment of €1.5 billion and private investment of €500 million, for a total of €2.2 billion (development of embedded AI and trustworthy AI, etc.: approximately €1.22 billion; human resource development: approximately €780 million; research on innovative technologies: approximately €130 million, etc.). It also has three major goals: improve the national skill level, make France a leader in embedded and trustworthy AI, and accelerate the deployment of AI in the economy.

(3) Germany

[1] Artificial intelligence in the network sectors

On December 15, 2021, the Federal Network Agency published a report, titled Artificial Intelligence in the Network Sectors, on the use of AI in the network services sectors, including telecommunications, electricity, gas, postal services, and railways³⁴.

The report shows that AI has enormous potential and that it should be actively introduced into the network service sectors. It also points out that there are issues such as closing the gap between companies, the shortage of AI specialist workers, the complexity of algorithms, and the lack of a legal framework. In addition, with regard to the AI legal system, the report indicates the intention to continue discussions with related business operators in order to develop laws suitable for the business environment in Germany in accordance with the legislative framework proposed by the EU.

[2] Towards auditable AI systems

On May 23, 2022, the Federal Office for Information Security published a whitepaper on auditing AI systems titled Towards Auditable AI Systems³⁵.

The white paper proposes a Certification Readiness Matrix (CRM) that assigns scores according to the lifecycle (planning phase, data acquisition phase, training phase, operation phase, etc.) and evaluation items (security, safety, explainability, etc.) of AI to comprehensively inventorize the auditability of AI systems in various use cases and to compare the auditability of different applications and time horizons.

(4) United Kingdom

[1] National AI Strategy

On September 22, 2021, the Government of the United Kingdom published its first 10 year plan titled National AI Strategy³⁶. The main points of the strategy are described below.

Available at the website of the following URL.

<<https://www.enseignementsup-recherche.gouv.fr/sites/default/files/2021-11/dossier-de-presse---strat-gie-nationale-pour-l-intelligence-artificielle-2e-phase-14920.pdf>>

³⁴ "Künstliche Intelligenz in den Netzsektoren"

Available at the website of the following URL.

<<https://www.bundesnetzagentur.de/SharedDocs/Downloads/DE/Sachgebiete/Digitales/KI/KI-Bericht-Marktdialog.pdf>>

³⁵ "Towards Auditable AI Systems"

Available at the website of the following URL.

<https://www.bsi.bund.de/SharedDocs/Downloads/EN/BSI/KI/Towards_Auditable_AI_Systems_2022.pdf;jsessionid=64A6880D58497D957001FBEDB6D08E29.internet472?__blob=publicationFile&v=4>

³⁶ "National AI Strategy"

Available at the website of the following URL.

- Enhance the ability of the business and public sectors to enter the market by improving coordination and cooperation between researchers through the launch of the National AI Research and Innovation Programs
- Launch a joint program between the Office for Artificial Intelligence and UK Research and Innovation
- Announce a joint review on the potential and capability of computing power to drive large-scale deployment of AI technologies
- Initiate consultations on AI copyright and patents through the Intellectual Property Office
- Coordinate UK involvement in setting international rules through trialing the AI Standards Hub and update guidance on AI ethics and safety in the public sector

[2] Algorithmic Transparency Standard

On November 29, 2021, the Cabinet Office's Central Digital and Data Office compiled and published the Algorithmic Transparency Standard, in line with its commitment to seek appropriate and effective ways to increase the transparency of algorithm-supported decision-making in the public sector as set out in the National AI Strategy and the National Data Strategy³⁷.

This standard consists of the Algorithmic Transparency Data Standard composed of 37 items formulated for organizations in the public sector and the Algorithmic Transparency Recording Standard - Guidance for Public Sector Bodies, which is a usage manual. A pilot program was conducted in public organizations to verify its effectiveness.

[3] Road map to an effective AI assurance ecosystem

On December 8, 2021, the Centre for Data Ethics and Innovation (CDEI) published a roadmap outlining the steps needed to build a world-leading AI assurance ecosystem within five years³⁸.

The roadmap states that in order for AI technologies to be accepted by society and generate profit, it is necessary to ensure confidence in the effectiveness, accuracy, and safety of AI, and like assurance in other areas (financial reporting, quality management, etc.), providers of AI assurance services must meet two requirements below.

- Reliably evaluate evidence that proves trustworthiness
- Provide appropriate levels of evidence to users (e.g. adopting companies)

The next step is the implementation of the roadmap, and as part of this, on January 12, 2022, the Centre together with the Department for Digital, Culture, Media & Sport and the Office for Artificial Intelligence announced the trial operation of the AI Standards Hub as a tool development and community platform for improving AI governance, promoting investment and employment, and contributing to the development of global AI technology standards.

<<https://www.gov.uk/government/publications/national-ai-strategy>>

³⁷ "Algorithmic Transparency Standard"

Available at the website of the following URL.

<<https://www.gov.uk/government/collections/algorithmic-transparency-standard>>

³⁸ "The road map to an effective AI assurance ecosystem"

Available at the website of the following URL.

<<https://www.gov.uk/government/publications/the-roadmap-to-an-effective-ai-assurance-ecosystem>>

[4] AI Barometer (2nd Edition)

On December 17, 2021, the CDEI published its AI Barometer (2nd Edition), which analyzes opportunities, risks, and governance challenges related to AI and data use³⁹.

The report, based on a survey of around 1,000 companies, reveals how AI and data are being used across the United Kingdom. For example, the report found significant differences across industry sectors in the adoption of data-driven technologies, as well as barriers to accessing and sharing data.

(5) United States

[1] National AI Advisory Committee

On September 8, 2021, the U.S. Department of Commerce announced the creation of the National AI Advisory Committee (NAIAC) to advise the President and other federal agencies on AI issues⁴⁰.

NAIAC, which was established under the National AI Initiative Act of 2020, is composed of experts from a wide range of fields related to AI, including academia, industry, non-profit organizations, civic groups, and national laboratories. It is tasked with making recommendations on the current state of competitiveness in the field of AI in the United States, the implementation of initiatives, the state of science surrounding AI, issues related to AI and the workforce, international partnerships, and legal issues. The first meeting was held on May 4, 2022.

At the meeting, the Sub-Committee on AI and Law Enforcement (NAIAC-LE) was formally established. The subcommittee is tasked with reviewing the challenges of utilizing AI in law enforcement and advising the President on topics including bias, data security, AI safety and law enforcement adoption, and legal standards to ensure that the utilization of AI aligns with privacy rights.

[2] Invitation for comments on the deployment of AI, including facial recognition

From November 5, 2021 to December 6, the Department of Home and Security (DHS) invited comments from the public regarding their awareness of the application of artificial intelligence, including facial recognition⁴¹.

The DHS has already been using or experimenting with AI-based technologies in key operations such as customs, border security, transportation security, and surveys. However, the use of these technologies has not been fully discussed with the general public, and therefore the DHS invited comments. The DHS commented that it believes it needs to understand how the public perceives the benefits and risks of new technologies in order to successfully apply them.

³⁹ "AI Barometer 2021"

Available at the website of the following URL.

<<https://www.gov.uk/government/publications/ai-barometer-2021>>

⁴⁰ A U.S. Department of Commerce press release on this matter is available at the website of the following URL.

<<https://www.commerce.gov/news/press-releases/2021/09/department-commerce-establishes-national-artificial-intelligence>>

⁴¹ Details of the invitation for comments and the opinions received are available at the website of the following URL.

<<https://www.regulations.gov/docket/DHS-2021-0015-0001>>

[3] Responsible AI Guidelines

The Defense Innovation Unit of the DOD published the Responsible AI Guidelines on November 15, 2021⁴².

The guidelines provide a comprehensive framework for DoD AI system developers, including private industry and government employees, to determine whether their AI programs conform to DoD AI ethical principles.

In addition, the report in the guidelines explains how to apply the guidelines to technologies for developing and procuring AI systems, and provides step-by-step procedures for various stages from planning, development, and implementation.

[4] AI researcher's portal

On December 20, 2021, the National Artificial Intelligence Initiative launched the AI Researchers Portal for AI researchers at "AI.gov"⁴³ to provide easy access to datasets and testbed environments that train AI initiative applications⁴⁴.

Tools accessible through the portal include information on federal AI grants, datasets, computing resources run by the National Science Foundation and others, a directory of research programs, and 40 test beds. According to a tweet⁴⁵ announcing its launch, "It's a central connection to many Federally-supported resources for America's AI research community."

[5] Revised report on bias in AI

The National Institute of Standards and Technology (NIST) published a revised report on bias in AI on March 15, 2022⁴⁶.

Reflecting comments received in response to a draft published in the summer of 2021, the report suggests that, as AI systems do not operate independently, the focus should be on the overall context, and as a step towards improving our ability to identify and manage the harmful effects of bias, the scope for investigating its causes should be expanded to include to a broad range of societal factors that have a significant impact on how technology is developed in addition to machine learning processes and data used for AI software learning.

[6] AI risk management framework (initial draft)

On March 17, 2022, NIST announced the initial draft of the AI Risk Management Framework for the purpose of promoting the development and use of trustworthy and responsible AI⁴⁷.

The framework was created based on comments received regarding the concept paper

⁴² "Responsible AI Guidelines"

Available at the website of the following URL.

<<https://www.diu.mil/responsible-ai-guidelines>>

⁴³ See Section 2. (6)-[4] in Chapter 1 of Report 2021.

⁴⁴ The URL for the portal is provided below.

<<https://www.ai.gov/ai-researchers-portal/>>

⁴⁵ The URL of the tweet is provided below.

<<https://twitter.com/WHOSTP/status/1471877506504568834>>

⁴⁶ "Towards a Standard for Identifying and Managing Bias in Artificial Intelligence"

Available at the website of the following URL.

<<https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.1270.pdf>>

⁴⁷ "AI Risk Management Framework: Initial Draft"

Available at the website of the following URL.

<<https://www.nist.gov/system/files/documents/2022/03/17/AI-RMF-1stdraft.pdf>>

published in December 2021, etc. to deepen understanding of the risks of business operators and society related to AI systems, and to support risk management. It defines various concepts related to AI systems, such as stakeholders and lifecycles, classifies the characteristics of risks associated with AI, and organizes four functions (mapping, measurement, management, governance) necessary for risk management.

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 publish a second draft in the summer or autumn of 2022 and publish Version 1.0 of the official framework by the winter.

[7] Guidance for addressing employment discrimination created by AI

On May 12, 2022, the Department of Justice and the Equal Employment Opportunity Commission published guidance and technical assistance documents on disability discrimination when employers make employment decisions utilizing AI and other software tools⁴⁸.

The guidance warns that the use of AI tools by employers to make employment decisions may result in unlawful discrimination to persons with disabilities under the Americans with Disabilities Act. It also stipulates that employers must consider how a tool affects various disabilities when a selecting tool, and that processes should be in place to provide reasonable accommodation when using algorithmic decision-making tools.

[8] Combatting online harms through innovation

On June 16, 2022, the Federal Trade Commission published a report titled Combatting Online Harms Through Innovation, which urged caution in promoting the use of AI tools aimed at combating illegal and harmful information online⁴⁹.

The report follows a request under the 2021 Appropriations Act to investigate and report to Congress how AI can be used to detect, remove, and otherwise respond appropriately to illegal and harmful information online.

The report noted that AI has inherent design flaws and inaccuracies, the potential to reflect the developer's bias and lead to erroneous results, and the potential to motivate commercial surveillance due to the massive data requirements of technological development, and concluded that "governments, platforms, and others must exercise great caution in either mandating the use of, or over-relying on, these tools even for the important purpose of reducing harms."

(6) China

[1] White Paper on Trustworthy Artificial Intelligence, Artificial Intelligence White Paper 2022

On July 9, 2021, the China Academy of Information and Communications Technology

⁴⁸ "Algorithms, Artificial Intelligence, and Disability Discrimination in Hiring" and "The Americans with Disabilities Act and the Use of Software, Algorithms, and Artificial Intelligence to Assess Job Applicants and Employees"
Available at the website of the following URL.
<<https://beta.ada.gov/ai-guidance/>> and <<https://www.eeoc.gov/laws/guidance/americans-disabilities-act-and-use-software-algorithms-and-artificial-intelligence>>

⁴⁹ "Combatting online harms through innovation"
Available at the website of the following URL.
<[https://www.ftc.gov/system/files/ftc_gov/pdf/Combatting Online Harms Through Innovation](https://www.ftc.gov/system/files/ftc_gov/pdf/Combatting%20Online%20Harms%20Through%20Innovation.pdf); Federal Trade Commission Report to Congress.pdf>

(CAICT) published the White Paper on Trustworthy Artificial Intelligence⁵⁰.

The white paper systematically presents a comprehensive framework of trustworthy AI, describes the characteristics of trustworthy AI, and analyzes the relationship between trustworthy AI and AI science technology ethics and management. Focusing on aspects such as trustworthy AI technologies, industry and industrial practices, it analyzes pathways for achieving trustworthy AI that is controllable and trustworthy, transparent and interpretable, privacy-protecting, accountable, multidimensional and inclusive, and proposes future developments in trustworthy AI.

CAICT published its White Paper on Artificial Intelligence 2022 on April 12, 2022⁵¹.

The white paper comprehensively reviews the latest trends in global AI policy, technology, implementation, and governance, and summarizes the challenges. In addition, in light of recent developments in AI, it shows that the aim is sustainable and sound development of AI.

[2] White Paper on Standardization of Artificial Intelligence (2021 Version)

The White Paper on the Standardization of Artificial Intelligence (2021 Version), compiled by the China Electronics Standardization Institute, etc. with guidance from the National Artificial Intelligence Standardization Group and the Artificial Intelligence Subcommittee of the National Information Standardization Committee, was published on July 9, 2021⁵².

The white paper describes the current state and trends in the development of AI, the state of standardization of AI in China and abroad, and the state of establishment of a standard system, as well as and points to keep in mind.

[3] New Generation Artificial Intelligence Ethics Specifications

On September 25, 2021, the Ministry of Science and Technology's National New Generation Artificial Intelligence Governance Expert Committee published the New Generation Artificial Intelligence Ethics Specifications⁵³.

The specifications aims to ensure that AI is always under human control and sets out six basic ethical requirements together with 18 specific ethical requirements for AI management, research and development, supply and use activities.

[Basic ethical requirements]

- Advancement of human welfare
- Assurance of controllability and trustworthiness
- Promotion of fairness and justice
- Clarification of responsible person
- Protection of privacy and security
- Improved ethical awareness

⁵⁰ 可信人工智能白皮书 in Chinese

Available at the website of the following URL.

<<http://www.caict.ac.cn/kxyj/qwfb/bps/202107/P020210709319866413974.pdf>>

⁵¹ 人工智能白皮书 (2022 年) in Chinese

Available at the website of the following URL.

<<http://www.caict.ac.cn/kxyj/qwfb/bps/202204/P020220412613255124271.pdf>>

⁵² 人工智能标准化白皮书 (2021 版) in Chinese

Available at the website of the following URL.

<[http://www.cqsoft.org/ueditor/php/upload/file/20210917/人工智能标准化白皮书 \(2021 版\) .pdf](http://www.cqsoft.org/ueditor/php/upload/file/20210917/人工智能标准化白皮书 (2021 版) .pdf)>

⁵³ 新一代人工智能伦理规范 in Chinese

Available at the website of the following URL.

<http://www.most.gov.cn/kjbgz/202109/t20210926_177063.html>

[Specific ethical requirements]

< Artificial Intelligence Management Specifications >

- Promotion of the sustainable development of artificial intelligence
- Strengthening of risk prevention

< Research and Development Specifications >

- Enhancing the safety and transparency of algorithms, etc.
- Avoiding bias discrimination in data collection and algorithm development

< Supply Specifications >

- Strengthening of quality management
- Guaranteeing the interests of users

< Use Specifications >

- Avoiding misuse and abuse
- Prohibition of unauthorized use

Etc.

(Reference) Proposal for a Regulation laying down harmonised rules on artificial intelligence

On April 21, 2021, the European Commission published its Proposal for a Regulation laying down harmonised rules on artificial intelligence (hereinafter referred to as "Artificial Intelligence Act")⁵⁴. The Artificial Intelligence Act aims to create an ecosystem of trust by proposing a legal framework for trustworthy AI, which is being debated and coordinated by the European Parliament and the European Council for adoption.

This Act is expected to have a significant impact not only within the EU, but also internationally, including Japan. Therefore, this Conference is closely monitoring its developments. This Conference interviewed relevant parties about their perceptions and reactions to the Act in Japan and overseas, and an overview is provided below⁵⁵.

< Domestic >

• Opinion of KEIDANREN (Japan Business Federation)

• Basic concept

- The direction is consistent with KEIDANREN's aim of building a trustworthy, high-quality AI ecosystem.
- However, there are ambiguities and room for interpretation in the definitions of prohibited AI and high-risk AI, which may affect innovation and national security.
- Definitions should be clarified, explanations should be added, and guidelines should be provided prior to enforcement.

• Specific opinions

- Regarding real-time remote biometric authentication systems used for law enforcement purposes, the risks of each use case should be taken into account and the range of publicly accessible spaces should be clarified with specific examples.
- The scope of prohibited AI and high-risk AI should be more limited, and the scope and rationale of regulations and methods of risk measurement and assessment should be

⁵⁴ See Section 2. (1)-[3] in Chapter 1 of Report 2021.

Both the body and the overview of the Artificial Intelligence Act (Japanese translations for information purposes only) are published on the website of the Ministry of Internal Affairs and Communications as reference materials for this report. Available at the website of the following URL.

<<https://www.soumu.go.jp/iicp/research/results/ai-network.html>>

⁵⁵ Presentations followed by an exchange of opinions were conducted three times. See Note 15 above. In Japan, in addition to the KEIDANREN (Japan Business Federation), others including NEC Corporation and FUJITSU LIMITED also submitted their opinions on the Artificial Intelligence Act. The opinions of Europe and the United States are based on the Survey Research on the Trends of Regulations on Artificial Intelligence (AI) in Foreign Countries published by the Ministry of Internal Affairs and Communications.

clarified.

- Guidelines should be provided that take into account risk tradeoffs related to AI risk management.
- High risk AI providers should be granted immunity subject to close examination of certain risks.
- Appropriate penalties should be determined according to the type and nature of a violation, the magnitude of the benefit obtained, and whether or not the violation was malicious.

< Europe >

- o Opinions of researchers, law firms and companies in European countries
 - Overall view of the Artificial Intelligence Act

The following observations were made, and there were no objections to enacting the Artificial Intelligence Act.

 - It is recognized that the Act ensures EU leadership on AI, develops uniform rules for protecting markets and supports innovation, and that the Act adopts a normative, rather than just ethical, framework for the protection of human rights, society with a human-centered approach.
 - It is recognized that the Act was created based on the importance of respecting and protecting human rights (a human-centered approach) together with strengthening Europe's competitiveness.
 - It is recognized that the Act aims to develop trustworthy and secure AI in the EU and protect the values, fundamental rights and security of EU citizens.
 - Views on the definition of "AI system"
 - The definition should be further clarified.
 - One problem is that "statistical approach" is also included.
 - Could potentially cover most software.
 - Views on AI with unacceptable risk
 - Definition is vague and the scope of prohibition is narrow.
 - A clear ban on AI for social scoring and surveillance by public organizations is reasonable.
 - It is considered that the definition is normatively narrow in order to ensure predictability, but this can still be a burden on entities.
 - Views on high-risk AI
 - While the scope of high-risk AI is believed to be limited, the technical requirements need to be clarified.
 - Range of high-risk AI is wide. AI categories, use cases, and interpretation guidance should be clearly described.
 - The content in the Artificial Intelligence Act regarding the definition and obligations of high-risk AI is not expected to change significantly in future discussions, but it is desirable to have a clear standard.
 - Views on support for small-to-medium-sized enterprises
 - The current Artificial Intelligence Act could have a chilling effect on innovation.
 - Some view regulatory sandboxes favorably.
 - The provision on support for small-to-medium-sized enterprises is too vague and needs to be amended.

< United States >

o Opinions of a law firm in the United States

• Regulations in the United States

- It is expected that the United States will not create AI regulations that are as broad as the EU and have as many challenges for business to deal with. However, even without new legislation, federal agencies are working to regulate AI consistent with current (typically sectoral) regulatory powers.
 - The U.S.-EU Trade and Technology Council (TTC) has the potential to bring U.S. AI regulations and EU AI regulations at least somewhat in line with each other on discrimination, bias, fraud, transparency and other areas that the U.S. and EU generally agree on. The TTC may also exert its maximum efforts towards policies designed to limit dependence on or lessen the influence of certain countries that use technology in an anti-democratic manner or create challenges to the security of the United States or its allies in Europe and the Asia-Pacific region.
 - The United States uses a mix of hard and soft laws to control AI. In general, the United States uses soft laws whenever possible. Legislators and regulators turn to hard laws only when experience shows that soft laws are insufficient to protect the public from harm from AI. Recently, there has been some movement towards applying or enacting such hard laws.
 - U.S. regulators are very focused on applying current laws to AI because it is much easier to apply existing laws than to enact new laws. At the same time, legislation or proposals for legislation are also found at the federal and state levels.
 - The government is focusing on the risks of AI. Companies should do the same. The idea that AI is not regulated in the U.S. is incorrect. (Note: There was no objection from the law firm to the view that hard laws have been adopted or will be adopted to a wider extent than typically imagined in Japan.)
- U.S.-Japan collaboration on AI
- The United States and Japan will continue to work together on the issues of ethical, safe and trustworthy AI, either in multilateral fora (OECD, G7, G20, GPAI, etc.) or through bilateral dialogue.
 - The United States would like to cooperate with Japan in the same way that it cooperates with the EU under the TTC, with the aim of establishing technical standards for AI and requiring regulatory compliance.

3. Trends in international discussions

(1) OECD

[1] Council of Ministers

The Council met on October 5-6, 2021.

Referring to promoting the digital economy including implementing AI principles, the Ministerial Statement⁵⁶ states "We commit to working together to advance an inclusive digital economy including through implementation of the OECD AI Principles and promoting data free flow with trust (DFFT) and the possible development of high-level principles for trusted government access to personal data."

⁵⁶ OECD Ministerial Statement on Part II of the Council 2021 (Japanese translation for information purposes only)

Available at the website of the following URL.

<<https://www.mofa.go.jp/mofaj/files/100243592.pdf>>

On October 4 of the same year, a session on the implementation of AI principles was held as a side event⁵⁷.

[2] Committee on Digital Economy Policy

The AI session of the Committee on Digital Economy Policy (CDEP) was held on December 2, 2021. Regarding Japan's initiatives, at the meeting the Ministry of Internal Affairs and Communications introduced "Report 2021" by this Conference, which summarizes specific examples of initiatives and good practices based on interviews with business operators and others, and the Ministry of Economy, Trade and Industry introduced the "Governance Guidelines for Implementation AI Principles"⁵⁸ with reference cases.

Furthermore, it was agreed to transfer⁵⁹ the WG on AI Policies, one of the three working groups of the informal OECD Network of Experts on AI (ONE AI)⁶⁰, 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.

[3] Working Party on Artificial Intelligence Governance

The first meeting of the Working Party on Artificial Intelligence Governance (WP AIGO) was held on May 24-25, 2022.

The working group was established to support the implementation of OECD standards on AI, including the Recommendation of the Council on Artificial Intelligence, and to facilitate responsible stewardship of trustworthy AI by information exchange and documentation of national AI policies and further functional development of online platforms (OECD.AI Policy Observatory).

The 2022 Chair and Vice-Chair (Including Japan) were approved at the meeting. In addition, policies on AI governance in each country were introduced and shared, and the activities and states of progress of the ONE AI working groups were reported and discussed. In addition, regarding the OECD.AI Policy Observatory, a proposal was introduced to expand its functions, such as a user interface that quantitatively analyzes and visualizes AI policies in each country based on the number of AI-related documents and indicators on the status of cooperation between organizations.

As for Japan, regarding compliance with the AI Principles, the Ministry of Internal Affairs and Communications mentioned the fact that it is a soft low stance and the importance of objective evaluation through audits, and regarding language technology, introduced for example the Global Communication Plan 2025⁶¹ and the Multimodal Spoken Dialog System

⁵⁷ "Putting the OECD AI Principles into Practice: progress and future perspectives"

Available at the website of the following URL.

<<https://oecd.ai/en/mcm>>

⁵⁸ At that meeting, Governance Guidelines for Implementation AI Principles Version 1.0 (now revised to Version 1.1) was introduced.

Available at the website of the following URL.

<https://www.meti.go.jp/shingikai/mono_info_service/ai_shakai_jisso/2021070902_report.html>

For information on Governance Guidelines for Implementation of AI Principles Ver. 1.1, see 1. (5) in this chapter.

⁵⁹ See [3] Working Party on Artificial Intelligence Governance.

⁶⁰ The participants from this Conference are Chairperson Sudo and member Sugiyama.

⁶¹ Global Communications Plan 2025, Ministry of Internal Affairs and Communications

Available at the website of the following URL.

<https://www.soumu.go.jp/menu_news/s-news/01tsushin03_02000298.html>

for Elderly Care (MICSUS)⁶². In addition, in a panel discussion on the relationship between AI and efficient energy systems and communications infrastructure policy, Toshiba Corporation introduced the Virtual Power Plant (VPP)⁶³, which functions like a single power plant by remotely controlling scattered energy sources such as distributed power sources and storage batteries using IoT devices, and its application technology, AI technology for predicting energy demand and solar power generation.

(2) Global Partnership on AI (GPAI)

[1] Second Plenary Meeting

The GPAI Summit (Second Plenary Meeting) was held on November 11-12, 2021, and at the Cabinet-level Council, it was approved that Japan would assume the next presidency (autumn 2022 to autumn 2023). The participation of Belgium, Czech Republic, Denmark, Ireland, Israel, and Sweden as new members and the participation of UNESCO as an observer from 2022 onward was also approved.

Participating countries expressed the importance of promoting responsible AI based on the AI principles of the OECD, the importance of promoting multi-stakeholder participation, enhancing diversity by approaching developing countries that share democratic values, and their expectations of Japan as the first chair country from the Asia-Pacific region.

[2] GPAI Symposium

The Ministry of Internal Affairs and Communications co-hosted the GPAI Symposium with the Ministry of Economy, Trade and Industry on February 9, 2022.

AI experts from industry and academia, including experts participating in the GPAI, exchanged opinions on topics such as international trends in AI ethics, from theory to practice, and expectations of stakeholders in Japan and overseas⁶⁴. More than 400 people attended, and many said they gained a better understanding of GPAI activities and the implementation of AI principles.

[3] The Future of GPAI Work: Reports and Suggestions from the Japan Survey

Working group 3, the "Future of Work"⁶⁵ Japanese team released a report on March 15, 2022 titled GPAI Future of Work: Reports and Suggestions from the Japan Survey⁶⁶.

This report examines survey items that match the situation in Japan based on previous surveys already conducted overseas, and compiles the results of 11 interviews conducted with companies and local governments for an awareness of the impact of AI on work and an understanding of the actual situation.

The survey results revealed that the aims of AI utilization, such as lack of human resources and improvement of service quality, are categorized, and efforts are being made to build

⁶² Multimodal Spoken Dialog System for Elderly Care (MICSUS), National Institute of Information and Communications Technology (NICT)

Available at the website of the following URL.

<<https://keihanna-fair.jp/exhibition/ai/899>>

⁶³ Virtual Power Plant (VPP), Toshiba

Available at the website of the following URL.

<<https://www.global.toshiba/jp/products-solutions/renewable-energy/products-technical-services/vpp.html>>

⁶⁴ Of the members of this Conference, Chairperson Sudo and active members participated, and of the members of the AI Governance Review Committee, member Ema participated.

⁶⁵ Of the members of the AI Governance Review Committee, Ema is participating.

⁶⁶ Available at the website of the following URL.

<<https://ifi.u-tokyo.ac.jp/news/12582/>>

governance systems and develop human resources. However, issues concerning AI utilization include not only technical issues such as transparency and fairness, but also human issues such as redefining the division of roles between AI and humans, building trust between AI and humans, and excessive reliance on AI. The following two items were proposed based on the findings of the survey.

- Necessity of country/region-specific reports on the Future of Work survey
- Promotion and effectiveness verification of student-centered research methodology

(3) G20

The Group of 20 summit in Rome was held on October 30-31, 2021.

Regarding the promotion of the AI Principles, the Leaders' Declaration⁶⁷ stated "Well aware of the benefits stemming from the responsible use and development of trustworthy human-centered Artificial Intelligence (AI), we will advance the implementation of the G20 AI Principles, while considering the specific needs of MSMEs and start-ups to encourage competition and innovation, as well as diversity and inclusion, and the importance of international cooperation to promote research, development and application of AI. We welcome the G20 Policy Examples on How to Enhance the Adoption of AI by MSMEs and Start-ups."

(4) Council of Europe Ad hoc Committee on Artificial Intelligence (CAHAI)

From November 30 to December 2, 2021, the Sixth Session of the Assembly adopted a document (Possible elements of a legal framework on artificial intelligence, based on the Council of Europe's standards on human rights, democracy and the rule of law) on the legal framework for AI.⁶⁸

The document states that a legally-binding cross-sectoral legal instrument based on three values (human rights, democracy and the rule of law) is needed to prevent and mitigate risks that may arise from the use of AI systems, and notes that additional legally-binding/non-legally-binding legal instruments may be needed in the future for detailed guidance in each sector.

It is also proposed to introduce additional legally binding and non-binding instruments in the public administration sector to supplement the cross-sectoral legal framework. In particular, from the perspective of transparency, fairness, responsibility, accountability, explainability and compensation, it is recommended that mechanisms be established to ensure that the use, design, procurement, development and implementation of AI systems are supervised and that the three values are complied with.

The activities of CAHAI were taken over by the Committee on Artificial Intelligence (CAI), which held a high-level meeting on April 4, 2022 and its inaugural general meeting from the same day to April 6th⁶⁹. The General Assembly exchanged views on the outcomes of CAHAI

⁶⁷ G20 Rome Leaders' Declaration (Japanese translation for information purposes only)
Available at the website of the following URL.
<<https://www.mofa.go.jp/mofaj/files/100259275.pdf>>

⁶⁸ "Possible elements of a legal framework on artificial intelligence, based on the Council of Europe's standards on human rights, democracy and the rule of law"
Available at the website of the following URL.
<https://search.coe.int/cm/Pages/result_details.aspx?ObjectID=0900001680a4e8a5>

⁶⁹ "Inaugural Meeting of the Committee on Artificial Intelligence (CAI)"
Available at the website of the following URL.
<<https://www.coe.int/en/web/artificial-intelligence/-/inaugural-meeting-of-the-committee-on-artificial->

and examined measures for advancing the formulation of legal systems. The CAI is tasked with drafting a legal system by November 2023.

(5) United Nations

[1] United Nations Educational, Scientific and Cultural Organization (UNESCO)

The 41th Session of the Assembly was held from November 9 to 24, 2021, and the Recommendation on the Ethics of Artificial Intelligence was adopted⁷⁰.

The Recommendation specifies the following values and principles (matters to be respected by all parties concerned in the lifecycle of AI systems) and areas for policy measures (areas to be addressed by Member States based on the Recommendation, etc.) for the purposes of "providing a universal framework of values, principles, and actions consistent with international law in the formulation of legislation, policies, etc., related to AI by States" and "encouraging individuals, companies, and others to take action to ensure that ethics are embedded in every stage of the lifecycle of an AI system" and also indicates the necessity of monitoring and evaluation of AI ethics.

< Values >

- | | | |
|---|--------------------------------------|---------------------------|
| [1] Human dignity | [4] Ensuring diversity and inclusion | [5] Peace and coexistence |
| [2] Respect for human rights and fundamental freedoms | [3] Rich environment and ecosystems | |

< Principles >

- | | | |
|-------------------------------------|---------------------------------------|--|
| [1] Proportionality and Do No Harm | [5] Privacy and data protection | [9] Awareness and literacy |
| [2] Safety and security | [6] Human oversight and determination | [10] Adaptive governance by multi-stakeholders |
| [3] Fairness and non-discrimination | [7] Transparency and explainability | |
| [4] Sustainability | [8] Responsibility and accountability | |

< Scope of policy measures >

- | | | |
|---|--------------------------------|-----------------------------------|
| [1] Ethical impact assessment | [5] Environment and ecosystems | [8] Education and research |
| [2] Ethical governance and management | < Ecosystem > | [9] Communication and information |
| [3] Data policy | [6] Gender | [10] Economy and labor |
| [4] Development and international cooperation | [7] Culture | [11] Health and social welfare |

[2] Office of the High Commissioner

intelligence-cai->

⁷⁰ "UNESCO member states adopt the first ever global agreement on the Ethics of Artificial Intelligence" Available at the website of the following URL.

<<https://en.unesco.org/news/unesco-member-states-adopt-first-ever-global-agreement-ethics-artificial-intelligence>>

Although Russia, China, and other countries raised additional technical amendment opinions and discussions on the draft resolution, most countries welcomed the draft recommendation, and the Recommendation on the Ethics of Artificial Intelligence and the content of the resolution were adopted without substantial changes.

On September 15, 2021, the Office of the High Commissioner issued a report titled "The right to privacy in the digital age"⁷¹. When the report was issued, United Nations High Commissioner for Human Rights Michelle Bachelet commented that countries should explicitly ban AI applications that do not comply with international human rights law⁷².

The report analyzes the impact of AI utilization on privacy and other rights and provides examples of the impact in four key areas (law enforcement, national security, criminal justice and border control). It also provides recommendations on the design and implementation of safeguards to prevent and minimize adverse consequences and promote the enjoyment of the benefits provided by AI.

(6) APEC

From December 7 to 10, 2021, Peru hosted a workshop on AI. In the session titled Artificial Intelligence for Development: Risks and Opportunities, the Ministry of Internal Affairs and Communications introduced the initiatives of this Conference, etc. (AI R&D Guidelines, AI Utilization Guidelines, and examples in Report 2020 and Report 2021 of initiatives taken by business operators). Views on the importance of multi-stakeholder discussions and the direction of government procurement in the field of AI were also exchanged.

(7) Bilateral policy dialogue

[1] Japan-U.S. Policy Cooperation Dialogue on the Internet Economy

The Japan-U.S. Policy Cooperation Dialogue on the Internet Economy was held on November 11-12, 2021, and the AI session discussed public-private partnerships for deepening public trust in AI technologies and promoting responsible governance of trustworthy AI.

The Ministry of Internal Affairs and Communications introduced the content of this Conference's Report 2021 (summary of best practices and good practices related to AI ethics and governance, AI development and utilization, and human resources development), and stated that Japan would stick to the soft law approach while paying close attention to discussions on the EU's Artificial regulation Act, etc., and that Japan will work together with the United States to address issues.

The United States and Japan reaffirmed their support for the OECD Recommendation on AI and confirmed their continued collaboration on the responsible development and deployment of reliable AI through the OECD and GPAI. Furthermore, the United States welcomed Japan's future role as the next chair of the GPAI.

[2] Japan-EU ICT Strategy Workshop

On November 17, 2021, the Japan-EU ICT Strategy Workshop was held, and in the AI session, the EU explained the status of the Artificial Intelligence Act. In response, Japan explained its view that soft law is effective in flexibly responding to technological change, and introduced its status of participation in international discussions such as GPAI and UNESCO, its contributions so far through the formulation of the AI Utilization Guidelines, etc., and the

⁷¹ "The right to privacy in the digital age: report (2021)"

Available at the website of the following URL.

<<https://www.ohchr.org/en/calls-for-input/calls-input/2021/right-privacy-digital-age-report-2021>>

⁷² "Urgent action needed over artificial intelligence risks to human rights"

Available at the website of the following URL.

<<https://news.un.org/en/story/2021/09/1099972>>

status in industry of the implementation of the AI Principles and the Governance Guidelines for Implementation of AI Principles. Both sides pointed out the importance of Japan-EU cooperation in multilateral discussions and decided to continue the exchange of views and discussions.

On April 8, 2022, the workshop was held again, and the AI session reiterated the importance of Japan-EU cooperation in multilateral meetings, and the two sides agreed to continue exchanging views and discussions. Japan provided an explanation of the background to discussions on AI principles in the international arena, such as the G7 and the G20, and future initiatives (in particular, initiatives concerning the GPAI, which Japan will chair from autumn 2022, and initiatives for realizing human-centered AI). The EU explained the status of its initiatives regarding the Artificial Intelligence Act and AI principles, the importance of cooperation at the policy level, and the importance of understanding and sharing regarding AI.

[3] Japan-EU ICT Policy Dialogue

On February 3, 2022, the Japan-EU ICT Policy Dialogue was held, and at the AI session, both sides explained their views of the AI principles, and discussed initiatives such as international cooperation and similarities and differences between initiatives.

The Ministry of Internal Affairs and Communications explained the international situation, such as in the OECD, GPAI, and UNESCO, and also introduced the status of this Conference's initiatives and Report 2021. Both sides shared the view that they are on the same track and that they will continue to work closely together on differences in approaches. The EU commented that although Japan and the EU share the same values, soft law is insufficient and a legal framework is necessary to gain trust.

[4] Japan-Germany ICT Policy Dialogue

On March 23-24, 2022, the Japan-Germany ICT Policy Dialogue was held, and at the intergovernmental meeting on the 23rd, discussions were held on initiatives for promoting the AI network society and data utilization as part of the session on global digital governance.

The Ministry of Internal Affairs and Communications introduced the status of this Conference's initiatives and examples of initiatives by business operators described in Report 2021. Germany explained its domestic response to the EU Digital Services Act and the Data Act.

4. Trends in international standardization

While there are many aspects of AI systems to which the international standards of conventional software and hardware technologies can be applied, AI systems have unique challenges such as black box problems, ethical issues such as bias, and problems caused by learning data. In addition, due to the rapid progress of technological development related to AI, initiatives for standardization are being made in parallel with development and operation, including in fields such as quality assurance technologies, where many technologies are still in the research stage.

As for trends of major international standardization organizations, in 2016, the Institute of Electrical and Electronics Engineers (IEEE) launched a project on AI ethics (P7000 series). In addition, in 2017, the International Organization for Standardization (ISO) and International Electrotechnical Commission (IEC) established a subcommittee on AI (SC42) in the first joint technical committee (JTC1), and the working groups within SC42 have started a wide range of

projects.

In these standardization bodies, projects and working groups are active in a wide range of areas from principles to implementation, including the technical aspects of AI ethics, definitions of AI concepts and terms, definitions of AI systems and lifecycles, use cases, definitions of data, and quality management. In addition, standardization activities related to the clarification of government procurement requirements, management, and testing have recently been activated. The main initiatives undertaken since the publication of Report 2021 are described below⁷³.

(1) Institute of Electrical and Electronics Engineers (IEEE)

The IEEE Standards Board approved the project on government procurement of AI (P3119)⁷⁴ and its working group on September 23, 2021. The project is working defining terms and processes with respect to government procurement of AI for the purposes below.

- To establish uniform definitions and process requirements to address social, technical and responsible innovation challenges in the procurement of AI and Automated Decision Systems (ADS)
- To help support agencies adapt to processes for responsibly procuring AI systems for the public good
- To promote ethically aligned values and strong public participation in process models

(2) International Organization for Standardization (ISO) and International Electrotechnical Commission (IEC)

In the ISO/IEC 42001 project on AI system management, ISO/IEC JTC1/SC42 is examining requirements and guidance for organizations on establishing, implementing and maintaining AI management systems. In addition, in the ISO/IEC TS 29119 -11 project concerning test technologies and evaluation indices for verifying and validating AI systems according to their lifecycles, it is examining standard specifications.

(3) Trends in international standardization in various countries

[1] Europe

- In July 2021, the European Commission's Joint Research Centre published AI Watch: AI Standardisation Landscape⁷⁵. It picks up standards related to the Artificial Intelligence Act, examines them for consistency with the requirements of the Act, and identifies gaps for use in formulating a European standardisation roadmap.
- The European Committee for Standardization (CEN) and the European Committee for Electrical Standardization (CENELEC) analyzed the standards formulated by ISO/IEC JTC 1/SC42, and identified the standards that can be utilized in implementing the Artificial

⁷³ Trends in international standardization in this section are described mainly based on the results of presentations and exchanges of opinions, etc. in interviews (5th interview session held on March 24, 2022) conducted by the Chair of this Conference. For information on the discussions conducted by the Chair of this Conference, see the next chapter.

For information trends up to now related to international standardization, such as IEEE and ISO/IEC, see Section 2. (1) in Chapter 1 of Report 2018, Section 2. (1) in Chapter 1 of Report 2019, and Section 3. (8) in Chapter 1 of Report 2021.

⁷⁴ "Standard for the Procurement of Artificial Intelligence and Automated Decision System" Available at the website of the following URL.
<<https://standards.ieee.org/ieee/3119/10729/>>

⁷⁵ "AI Watch, AI standardisation landscape state of play and link to the EC proposal for an AI regulatory framework" Available at the website of the following URL.
<<https://op.europa.eu/en/publication-detail/-/publication/36c46b8e-e518-11eb-a1a5-01aa75ed71a1/language-en>>

Intelligence Act and the standards that JTC21 (CEN-CENELEC Joint Technical Committee on AI) should formulate as JTC21. Going forward, specific standardization activities are scheduled to start after coordination with the European Commission and other European standardization bodies.

[2] United States

- In August 2019, NIST published U.S. LEADERSHIP IN AI: A Plan for Federal Engagement in Developing Technical Standards and Related Tools⁷⁶, which identified the following nine areas of focus for standardization.
 - Concepts and terminology
 - Data and knowledge
 - Human interactions
 - Metrics
 - Networking
 - Performance testing and reporting methodology
 - Safety
 - Risk management
 - Trustworthiness
- NIST will develop its AI risk management framework⁷⁷ within 2 years from January 2021 and periodically update it, and will align it with international standards (such as ISO/IEC standards) as required.
- In October 2021, the Office of Science and Technology Policy called for the formulation of a Bill of Rights for an Automated Society, and held public hearings and workshops in November of the same year.

5. International symposium "Global Forum on AI Network Society 2022"

On March 1, 2022, the Ministry of Internal Affairs and Communications held the international symposium "Global Forum on AI Network Society 2022" with the aim of contributing to solving 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 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 they engaged in exchanges of ideas and opinions. For an overview of the Forum, see Attachment 1⁷⁸.

At the symposium, one opinion shared was that it is important to promote the sharing and distribution of data while balancing the protection of personal information with respect for data owners, and that it is necessary to improve the environment for this purpose.

In addition, with regard to AI ethics and governance, discussions were held on regulations in Japan based on overseas trends in regulations, etc. and on how AI governance should be implemented by business operators, and the following opinions were expressed⁷⁹.

- AI is a critical component in the disruptive technology trends since the beginning of the private use of the Internet, and AI principles and AI governance initiatives are attracting a lot

⁷⁶ "Plan Outlines Priorities for Federal Agency Engagement in AI Standards Development" Available at the website of the following URL.
<<https://www.nist.gov/news-events/news/2019/08/plan-outlines-priorities-federal-agency-engagement-ai-standards-development>>

⁷⁷ See Section 2. (5)-[6] in this chapter.

⁷⁸ An English version is available at the website of the following URL.
<https://www.soumu.go.jp/main_content/000811496.pdf>

⁷⁹ An overview of the lectures and panel discussions with the same content as Attachment 1 is already available on the Ministry of Internal Affairs and Communications website, however some of the descriptions in this chapter have been revised.

of attention.

- In March 2019, the Government of Japan compiled the "Social Principles of Human-Centric AI"⁸⁰ based on the discussions of the Advisory Council, for appropriate development and social implementation of AI. These principles are central to building a new society utilizing AI.
- At the Conference toward AI Network Society of the Ministry of Internal Affairs and Communications, a multifaceted study was conducted on how to utilize AI and how to distribute data, considering the ideal AI society.
- The G7 is scheduled to be held in Japan in 2023, and preparations are under way to actively propose the ideal form of AI. Amid various developments, it is desirable to create a new society in cooperation with the international community.
- Regulations are just one of many reasons to uphold AI ethics, and it is important to also consider AI ethics from the perspectives of, for example, corporate value and reputation, social justice and business opportunities. Compliance with regulations is important, but AI ethics goes beyond mere compliance, and laws are only part of getting AI right.
- AI governance is moving toward the phase for defining and implementing principles such as fairness, transparency and trustworthiness. Initiatives for solving issues related to AI governance are in the cooperation domain rather than the competition domain, and it is important to build an ecosystem by engaging various parties in discussions.
- The goal of utilizing AI is to supplement, not replace, human intelligence. The data and the findings derived from it should belong to the authors and new technologies should be transparent and explainable.
- In order to create social value through AI, together with functions and performance, it is important to gain trust and acceptance from society. While mechanisms for providing AI solutions that are trusted by society are effective for market expansion, it is essential to ensure that AI governance is not a hindrance to innovation creation or a barrier to entry.
- There are discussions on hard law and soft law in relation to the regulation of AI, but it is not a dichotomous debate, and it is important to share ideas on what are reasonable regulations for various and specific risks that will emerge in the future.
- Regarding the EU's Artificial Intelligence Act, I can understand that prevention is better than cure, but I am also concerned about the risk of slow innovation. So it is necessary to examine the details of regulations.
- AI is a developing technology, and it is important to think according to the assumption that the technology, risks, and social acceptability will change over time.
- The EU's high-level approach is important, but from the private sector's perspective, it is important to insist on the importance of the use case basis and work together to create a form that can be practically observed.
- There are differences between countries and regions, such as ways of thinking about privacy and human dignity, and ways of working, and how to handle and reconcile such differences is a very difficult task.
- AI regulations should be changed according to the situation as they are being used, and there may be a need to adapt to the strictest regulations. It is also important to ensure interoperability.
- There are cases of the utilization of AI in the public sector proceeding in the EU, and it is important to proceed with AI while gaining an understanding of the merits to the users.
- It is important to seek a cooperative line with the EU, but human resource development and

⁸⁰ See Note 20 above.

discussions are also essential.

- In a human-centered society, ELSI (ethical, legal, and social issues) reviews will function like the steering wheel and brake of a car, because the car will run out of control if AI is only utilized as the engine.
- It is necessary to clearly demonstrate to the EU and other countries that there are differences in the legal systems, industrial structures, and administrative systems of each country, and to promote international cooperation.

Going forward, it is important to continue efforts to promote the social implementation of AI based on such opinions.

Chapter 2. Initiatives for Promoting the "Safe, Secure, and Trustworthy Implementation of AI in Society" (Compilation of case studies of initiatives)

This Conference formulated the AI R&D Guidelines (2017) and the AI Utilization Guidelines (August 2019), and after compiling the AI Utilization Guidelines then conducted interviews with various stakeholders⁸¹ including AI developers, service providers, business users, and consumer-type users in order to promote the "Safe, Secure, and Trustworthy Implementation of AI in Society." These efforts are described in "Report 2020" (July 2020) and "Report 2021" (August 2021).

This chapter introduces the efforts of business operators, experts, relevant organizations, etc. (hereinafter referred to as "business operators and others") based on the interviews conducted from November 2021 to March 2022 by the Chair of this Conference (hereinafter referred to as "the interviews") after Report 2021 was published, and arranges the points to consider (points for good practice) in promoting the "Safe, Secure, and Trustworthy Implementation of AI in Society"⁸².

The initiatives of business operators and others are summarized in the supplementary volume (Collection of Good Practices), so please refer to it (the presentation materials of business operators and others are posted on the Ministry of Internal Affairs and Communications website⁸³).

1. Overview of interviews

After compiling "Report 2021," this Conference used the interviews as an opportunity to exchange opinions, based on presentations from business operators and others undertaking advanced or ambitious initiatives to implement AI in society. The main points discussed during the interviews are based on the points discussed during interviews conducted for Report 2020, from the perspective of promoting the "Safe, Secure, and Trustworthy Implementation of AI in Society" going forward⁸⁴. Specifically, discussions focused on the following topics:

- What initiatives are developers and users (AI service providers and business users) implementing 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 initiatives, 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?"

Opinions were exchanged on these topics in order to further discussion⁸⁵.

⁸¹ See Note 2 above.

⁸² In Report 2020, it was pointed out that it is necessary to continue conducting interviews and organize the contents as "shared knowledge" (see the conclusion of Report 2020).

⁸³ Presentation materials for which permission to disclose was granted are available at the following website.

<https://www.soumu.go.jp/main_sosiki/kenkyu/ai_network/02iicp01_04000232.html>

⁸⁴ The same points were used for discussion during the interviews conducted to compile the case studies provided in "Report 2021." For details on these discussion points, see the introduction to "Report 2020."

⁸⁵ More specifically, interviews were in some cases conducted from the perspectives of human resources development and the AI supply chain (in particular, during the 5th interview session conducted on March 24, 2022, opinions were exchanged to further discussion from the perspective of the AI supply chain).

Details of the interviews are summarized below.

	Presented by	Title
2021		
1st interview session (November 24)	SoftBank Corp. [SoftBank]	Examples of AI use at SoftBank
	Panasonic Corporation [Panasonic]	Panasonic's approach to AI development and application examples
	Sharp Corporation, AloT Cloud Inc. [Sharp]	AI initiatives at the Sharp Group
2nd interview session (December 22)	Prof. Keiichi Nakayama (Medical Institute of Bioregulation, Kyushu University) [Prof. Nakayama (Kyushu University)]	Application of artificial intelligence (AI) to medical biology
	Prof. Takuya Ueda (Clinical AI Human Resources Development Program, Tohoku University) [Prof. Ueda (Tohoku University)]	Clinical AI: An advanced AI R&D and human resources development center aiming to solve Global x Local medical issues
	Mercari, Inc. [Mercari]	Description of AI efforts by the Mercari Group
2022		
3rd interview session (January 25)	Japan Data Management Consortium [JDMC]	Ethics Framework for AI and Data use conceived by practitioners
	Mitsui Sumitomo Insurance Company, Limited [Mitsui Sumitomo Insurance]	Initiatives to promote digitalization at Mitsui Sumitomo Insurance
	DAIKIN INDUSTRIES, LTD. [DAIKIN INDUSTRIES]	AI human resources development efforts at Daikin Industries
4th interview session (February 15)	West Japan Railway Company [JR West]	Data analytics initiatives at JR West
	KDDI CORPORATION [KDDI]	AI use cases and AI governance initiatives at KDDI
	KPMG AZSA LLC [KPMG AZSA]	Initiatives toward verifying the appropriateness of AI applications
5th interview session (March 24)	NTT DATA Corporation [NTT DATA]	Supply chain risks associated with AI and NTT Data's responses
	FUJITSU LIMITED [FUJITSU]	Fujitsu's AI supply chain initiatives
	AI Data Consortium [AIDC]	Data use, issues, and initiatives

(Note): Names of business operators and others as of the time the interview was conducted. Names indicated in brackets ([]) represent abbreviations using in this text.

The descriptions in this chapter are based on the summary of the initiatives of business operator and others and the overview of presentations and opinions exchanged at the interviews found in Section 2 (2. Initiatives by business operators and others) of the supplementary volume (Collection of Good Practices.)⁸⁶ Section 2 of this chapter (2. Key points of presentations and exchanges of opinions at the interviews) is organized in a cross-sectional and cross-sectoral manner, such as for each principle item in the AI R&D Guidelines and the AI Utilization Guidelines, and for each field of AI utilization from the perspectives of AI ethics and governance initiatives, AI development and utilization initiatives, human resources development initiatives, and AI supply chain initiatives. In addition, initiatives that overlap these items and fields are described in the items and fields considered to be more similar in content (in some cases, they may be described redundantly in the form of repetition).

In addition, this collection of case studies consists mainly of cases considered to be particularly beneficial and useful to share, such as regarding the recognition of issues and their solutions, so that it will be of use to those who intend to start developing and utilizing AI, those who are currently developing and utilizing AI but are facing some challenges, and those who intend to further promote the development and utilization of AI, from the perspective of promoting the "Safe, Secure, and Trustworthy Implementation of AI in Society."

Note that the wording used in descriptions is based on presentations and exchanges of opinions during actual interviews.

Although the descriptions in this chapter were compiled based on presentations and exchanges of opinions held during the interviews, business operators and others are also engaged in various initiatives other than those described herein. We plan to conduct interviews on these other initiatives as necessary, and hope to play a supporting role in promoting the "Safe, Secure, and Trustworthy Implementation of AI in Society."

2. Key points of presentations and exchanges of opinions at interviews

(1) AI ethics and governance initiatives

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."

<Key points of Good Practices>

- 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 for AI developers and users, etc. to formulate and establish policies, guidelines, and principles, 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 in the interviews⁸⁷.

⁸⁶ The Collection of Good Practices supplementary volume was created as a complete collection of case studies. Therefore, Section 1 of this chapter and Section 1 of the Collection of Good Practices supplementary volume contain the same content.

⁸⁷ As Report 2021 also presents several specific initiative case studies, referring to those examples in conjunction with this report is considered helpful in formulating and establishing policies, guidelines, and principles. As well as the formulation and establishment of policies, guidelines and principles,

- 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 for AI developers and users, etc. to construct organizations and structures to implement and operate AI ethics and governance appropriately, while referring to the case studies presented in the interviews.

- Initiatives are seen prioritizing security assurance, privacy protection, assurance of fairness, and elimination of biases. In particular, 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 for AI developers and users, etc. 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 in the interviews.

- Initiatives are seen prioritizing accountability and the assurance of transparency, such as visualizing the basis used by AI to make decisions or tracing past conditions. In particular, attention should be paid to initiatives that provide accountability to end users by clearly stating on website that the purpose of using personal information is to develop technology.

It is desirable for AI developers and users, etc. 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 in the interviews.

- Many business operators and others position AI as a tool for humans to use, and implement the so-called "human in the loop"⁸⁸ operation, in which humans ultimately confirm the AI judgement and have AI as a support, rather than leaving everything to AI judgement. This 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.

It is desirable for AI developers and users, etc. to build and appropriately operate a mechanism for realizing "human in the loop" according to the mode of AI utilization and the nature of AI, while referring to the case studies presented in the interviews.

- In addition to initiatives to establish processes and rules for implementing quality assurance and evaluation in-house, attention should be given to initiatives in which third parties provide evaluation mechanisms (checklists, frameworks) and monitoring services. Such initiatives are considered a useful and effective measure for realizing security assurance, privacy protection, assurance of fairness, and the elimination of biases.

It is desirable for AI developers and users, etc. to proceed with initiatives that implement quality assurance and evaluations appropriately, while referring to the case studies presented in the interviews. 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.

- Initiatives are seen where business operators and others cooperate and collaborate with

Report 2021 introduces specific examples for other items such as organization and structure, security, and privacy, so please refer to it in conjunction with this report. See Section 2. (1) and Section 4. in Chapter 3 of Report 2021 and Attachment 3.

⁸⁸It is used here to mean human intervention in some of the decisions and controls of machines and systems that have become more automated and autonomous through artificial intelligence.

governments, relevant organizations, and outside experts and specialists, including those from overseas. This kind of 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 for AI developers and users, etc. 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 in the interviews.

a. Policies, guidelines, and principles⁸⁹

[Panasonic]

- The Company has created its own guidelines on what must never be done with AI, and has established its "AI Ethics Principles" as the minimum commitment required to deliver safety and security to customers.

[KDDI]

- Trustworthy AI is important as an element for achieving both AI utilization and governance. To this end, the Company is implementing three initiatives: formulating and disclosing principles, conducting research and development, and improving internal operations. With regard to formulating and disclosing principles, in August 2021, the Company formulated and disclosed nine principles considered important for the appropriate development and utilization of AI, having referred to the AI R&D Guidelines and AI Utilization Guidelines released by the Ministry of Internal Affairs and Communications.

b. Organization and structure

[Panasonic]

- The Company has established an internal AI Ethics Committee, and is promoting efforts to ensure thorough implementation of AI ethics and to establish a mechanism to update ethics as soon as possible.

[FUJITSU]

- In order to promote AI ethics, the Company believes that it is important to follow its own rules appropriately and to engage in continuous efforts to disseminate AI ethics both inside and outside the Company. Toward that end, the Company established its "AI Ethics and Governance Office" directly under the Company president, in February 2022.

c. Security

[NTT DATA]

- With regard to security, there is a risk of products containing backdoors as there is no way of knowing what was used as a base for creating models and programs procured externally. For this, there is no choice but to check source code and make judgments based on the trustworthiness of the provider. The Company believes that conducting thorough legal investigations/reviews would be a realistic approach for companies to take.

There is also a risk of data being stolen through backdoors. For this, the Company believes some kind of public endorsement of dominant components or products would bring peace of

⁸⁹ Of the business operators and others targeted for interviews, Fujitsu and NTT Data have already been interviewed and descriptions are provided in Report 2020. See Chapter 3 in Report 2020.

mind to users.

d. Privacy

[Panasonic]

- Autonomously updated edge devices and services process images only on the edge side and upload only the resulting metadata to the cloud, avoiding issues with privacy and transmission volume.

[JR West]

- The Company has begun developing a system that quickly detects users who need assistance (such as individuals in wheelchairs, using canes, or accompanied by guide dogs), using images from surveillance cameras installed in stations. Note that AI looks only at the skeletal framework of individuals when making a decision, and that technologies such as facial recognition or movement tracking are not used to identify individuals, so that AI retains no personal information.
- The use of security cameras is a sensitive topic and of much concern. The purpose of use of personal information is explained on the Company website, alongside information on technology development. In addition to contributing to the development of AI, the Company does not collect and analyze individual behavior logs to determine individual characteristics and does not perform facial recognition, and instead uses AI in a way that does not include personal information.

[NTT DATA]

- With regard to privacy and intellectual property, there is a risk that products may be used for purposes other than the intended purposes, and that data provided through APIs may be reused. For this, it is important to carefully check the terms of use.

e. Fairness

[SoftBank]

- It is difficult to eliminate data bias completely, and so it is important to continue to make steady efforts to collect a wide range of data in large amounts to improve accuracy. The Company is also engaged in technical efforts to prevent discrepancies based on factors such as ethnicity and gender. In areas where bias could be caused by AI, humans take over.

[NTT DATA]

- There is a risk of data bias, as bias in labels can lead AI to make inferences lacking fairness. For this, it is important to determine the extent to which parameters related to universal human attributes (such as gender) affect output.

f. Transparency and accountability

[Panasonic]

- AI is a black box, so it must be developed in a way where customers are satisfied with the causal relationship between the learning data and results. Quality assurance processes must therefore be established jointly with customers.
- The Company has introduced false detection countermeasures by visualizing the basis used by AI to make decisions, with a high level of accuracy. This allows causes of erroneous detection, which had previously been estimated based on intuition and experience, to now be

identified instantly, reducing the time required for countermeasures by half.

[Mercari]

- As for the accountability of decisions and predictions made by AI, to protect it the Company has decided to prepare experimental designs and store information so that past situations can be traced.

Experimental designs will indicate the background, content, evaluation indicators, and evaluation methods. These experiments are open within the Company, and comments from relevant departments and other engineers are welcomed, helping to prevent risks from being actualized.

- The Company provides a sensitive service called the credit scoring model, and strictly scrutinizes model performance evaluations and feature results. The Company is engaged in various efforts to establish conditions where the results of credit screening can be explained.

[JR West]

- The use of security cameras is a sensitive topic and of much concern. The purpose of use of personal information is explained on the Company website, alongside information on technology development.

[KDDI]

- With regard to accountable AI, the Company is currently focusing its research efforts on explaining reasons for decisions to users after the fact. The Company is currently engaged in research while using knowledge graphs as features for the reasons that decisions are made.

g. Appropriate use

[SoftBank]

- It is difficult to eliminate data bias completely, and so it is important to continue to make steady efforts to collect a wide range of data in large amounts to improve accuracy. The Company is also engaged in technical efforts to prevent discrepancies based on factors such as ethnicity and gender. In areas where bias could be caused by AI, humans take over.
- When used for recruitment, the first countermeasure is to use it so that it does not have any direct impact on the decision to accept or reject. The goal is to reduce the amount of time humans spend reading applications, not to have AI accept or reject applications.
- When AI rejects an application during the selection process for new graduates, this decision is always checked by a human to ensure that no application is rejected based solely on a decision by AI. This is true also for interviews. Any interview rejected by AI must be confirmed by HR personnel, so that the decision of whether the interviewee passes is not made by AI alone.
- AI is best suited to operations that would require great time, cost, and labor for humans to perform, such as work consisting of many or repetitive tasks. However, it is difficult to use AI for operations where human judgment is important or where making a decision itself is difficult. When AI is used during these kinds of operations, final confirmation must be performed by a human.
- If performing an operation or task where any problem occurring during the process could not be tolerated, a human would have to take over. This requires a lot of work, but can make up for the imperfect accuracy of AI.

[Panasonic]

- The Company believes that AI is a tool to be actively utilized in promoting the implementation of AI in business, and that problems in the field are solved by identifying necessary technologies. Under the concept of Data & AI for Co-Creation (DAICC), engineers with data and domain knowledge use AI as a tool to solve problems.

[Mercari]

- The Company needs to be able to detect transactions that violate its own rules or are fraudulent, and has therefore created a model that can identify potentially risky behavior from listings and deposit transactions. AI detects users engaging in risky behavior, and then a human manually checks for fraud.
- The Company's approach to detecting counterfeit items is to identify them from images and descriptions. However, due to the possibility of erroneous detection and omissions occurring during the selection process, a human must also manually check what the AI detects to reach a final decision.
- The Company takes a human-centered approach. Erroneous detection would result in the customer being penalized, so such cases must be checked to some degree by a human. However, there are issues of manpower and cost, and this will need to be investigated further.

[Mitsui Sumitomo Insurance]

- The Company is also considering the issue of personal data affecting insurance products or insurance amounts. For example, with an automobile accident, it is necessary to verify not only the results of the accident, but also the causal relationship between driving behavior and the accident. For this reason, driving data alone is not sufficient when an insurance company reflects conditions to insurance premiums. In Japan, rates are calculated based on laws on rates used to calculate basic insurance premiums. The Company believes that it is more important to reduce accidents all over the world.

[JR West]

- With regard to the balance between humans and AI, the Company believes that it is still difficult to introduce AI into security devices and other devices where mistakes in judgment cannot be tolerated. Although there are some issues that human inspectors may miss or some cases where rules could be interpreted differently, using AI can provide an answer around 70% of the way. The Company plans to replace tasks by obtaining data more often than humans can and increasing the number of decisions that are made.
- The Company is using AI as a tool to help humans make final decisions.

[KDDI]

- Regarding the AI watering and fertilizing system, when tasks can be automated with AI, humans can focus on other tasks such as improving quality. The Company is trying to improve both yield and quality by determining which tasks are best done by humans or by AI.

h. Quality assurance and development reviews

[JDMC]

- The ethics framework is designed for businesses that want to acquire information and make actual use of data. The organization is compiling issues focusing on acquired data, the amount of derived data increasing due to the use of AI, and the fact that use of this data is

being used in multiple stages and being provided externally.

- It is important for data to be practically useful, so the organization has been collecting case studies of actual problems occurring within Japan. Rather than deductively arriving at case studies from the concept of ethics, the organization first focused on collecting case studies inductively, abstracted these case studies, organized them by important points to check and ethical points of view, applied them to actual business cases, and repeatedly examined whether they were easy to use, in order to create a framework.
- The organization arranged the entire structure of the ethics framework into five ethics items. At the bottom of the structure is "information protection." Above this is ensuring "safety and security," realizing the use of data that in a manner that provides "fairness" and "accountability," and then using all of this to provide "propriety."
- The organization recognizes that the ethics framework is very comprehensive, and applying it to actual business would make it possible to include very detailed information. By using the framework, it should be possible to identify problems and issues that were not expected.
- In terms of actually using the framework, the organization is considering providing an overview of all risks identified through checking, determining each risk factor, and formulating how to reduce these risks.

[KDDI]

- With regard to improving internal operations, the Company continues to conduct privacy impact assessments (PIAs) on how specifically to put policies into operation and how to protect privacy. In addition to privacy-related efforts, the Company will likely begin checking items that require governance within PIAs, such as fairness and non-discrimination, accountability, transparency, and the intervention of human judgment. The Company will then begin discussing what to do with regard to areas that cannot be checked.

[KPMG AZSA]

- Conventionally, there has been no need to verify a model that has been created to any large degree, other than minor upgrades. For AI, however, the model continues to transform through continuous learning, and so the way it transforms must be verified with the issue of overfitting (over-learning) kept in mind.
- Since continuous learning is a feature of AI, the Company believes that AI model verification requires a framework that includes an audit system for continuously monitoring the AI model, rather than once a year as in conventional audits.
- The AI verification framework would be used to check whether there are any problems with input (such as what AI should be made to do and how to prepare it) or output (such as confirming what AI produces and whether the result is as expected), in light of the elements required for AI (such as accuracy, explainability, elimination of discrimination, and speed/robustness).
- It is important for the management framework to verify not only solution and data management, but also project management (upon which this relies), enterprise management (for the entire business operator), and technology management (as a technical element), in a multi-layered manner.
- The Company has defined eight perspectives for verification (fairness, explainability/interpretability, AI model accuracy, security/data protection, traceability/auditability, business continuity, data quality, and governance), but it will be necessary to confirm various elements together.

- The Company believes that the eighth perspective (governance in terms of how AI is being created and implemented) is particularly important, and also believes that checking whether there are any discrepancies between what is being developed and what is being operated will carry much weight during verification.
- The verification process would work on multiple levels, starting with strategy and governance, and then proceeding on to requirement definition, data preparation/preprocessing, model building, model verification, and then model implementation and operation/monitoring. However, that does not mean that all of these processes must be performed; instead, areas with a greater focus shall be verified with certain frequency depending on risk sensitivity.

[NTT DATA]

- With respect to quality, there is a risk regarding whether the accuracy of an externally procured product matches the required task. For this, there is currently no choice but to verify accuracy using a PoC based on the task. However, it would be preferable to establish AI quality standards to achieve certain guidelines.

[AIDC]

- Biases in the quality of AI and machine learning are often caused by data. However, if data annotation (labelling) is outsourced, one cannot rule out the possibility of a malicious operator. The organization has examined how to take responsibility and how data transaction contracts should be written, if such an issue should occur.

i. Cooperation and collaboration with external parties

[Panasonic]

- The field of AI continues to evolve at a rapid pace, so it is important to incorporate cutting-edge and necessary technologies (especially those that have a clear path to practical application), and the Company is increasing cooperation with world-class universities such as Stanford University and University of California, Berkeley.

[Mitsui Sumitomo Insurance]

- With regard to the organizational structure, insurance companies must promote innovation together with various stakeholders (such as external customers and business partners), rather than independently. The Company has therefore established its Business Innovation Division, a new department aimed at forming alliances.

[AIDC]

- The organization is engaged in work related to research on intellectual property and AI, data collection, and the construction of a data platform, and is promoting these initiatives in cooperation with various universities, research institutes, and government organizations.

(2) AI development and utilization initiatives

Of 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>

- 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.

It is desirable for AI developers and users, etc. to continue initiatives applying AI to COVID-19 measures, based on the status of COVID-19 infections, while referring to the case studies presented in the interviews⁹⁰.

- In the medical and healthcare field, attention should be paid to initiatives into accurate estimation of cancer prognoses and into drug discovery as well as initiatives⁹¹ in human resource development connected to medicine, AI, and data science where design thinking is incorporated. The advancement of such initiatives is expected to bring benefits to medical professionals, patients, and other stakeholders.

It is desirable for AI developers and users, etc. to promote AI development, utilization, and other initiatives in the medical and healthcare fields, while referring to the case studies presented in the interviews.

- Attention should be paid to the many business operators and others who take initiatives to apply AI to support the elderly 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.

In Japan, the importance and necessity of initiatives to support the elderly and people with disabilities are expected to continue to increase going forward, so it is desirable for AI developers and users, etc. to promote initiatives that utilize AI to support the elderly and people with disabilities, while referring to the case studies presented in the interviews.

a. Development and utilization of AI as a countermeasure against COVID-19

[SoftBank]

- In a vacancy count to show congestion, in order to safeguard privacy the congestion is visualized by displaying icons to indicate the presence of people but prevent them from being identified. The technology can also be used to visualize office and conference room congestion as a countermeasure against COVID-19, and the Company plans to begin offering it outside the Company.

[Panasonic]

- The Company is conducting verification experiments using small and low-speed delivery robots for a delivery service in residential areas. In addition to the growing shortage of delivery

⁹⁰ As Report 2021 also presents several specific initiative case studies, referring to those examples in conjunction with this report is considered helpful for the development and utilization of AI for COVID-19 measures. As well as the development and utilization of AI for COVID-19 measures, Report 2021 introduces specific examples of the development and utilization of AI for medical and healthcare and the development and utilization of AI for the elderly and people with disabilities, so please refer to them in conjunction with this report. See Section 2. (2) and Section 4. in Chapter 3 of Report 2021 and Attachment 3.

Examples of initiatives to support persons with disabilities utilizing AI can also be found on the website of the following URL. Please refer to them.

<<https://www.minnanosyogai.com/article1/ai> で障がい者の生活は大きく変わる？>

⁹¹ This point is common to "Initiatives aimed at developing human resources who understand AI and IoT in addition to technical expertise" described in the following section, (3) Human resources development initiatives, and it is considered that attention should be paid to it. See Note 95.

personnel, there is a need to respond to new lifestyles, such as non-face-to-face and non-contact services. The Company will accelerate its efforts to develop new delivery services by utilizing its accumulated knowledge in robot technology.

b. Development and utilization of AI for medical and healthcare

[Prof. Nakayama (Kyushu University)]

- Researchers created an accurate cancer prognosis system that utilizes AI. Researchers examined the expression of about 20,000 human genes and found that 286 genes had different prognoses depending on the amount of expression. As they continued their research, they found that 184 of these genes had an effect of prognosis regardless of factors such as ethnicity or environment.
- Researchers also discovered that they could obtain significant data if using 23 of these 184 genes, so they selected 23 genes and assigned weights to each to determine optimal parameters, in order to create a molecular prognostic score (mPS). They divided mPS into six groups (groups A to F) and then tracked the prognosis of each. The prognosis of each group matched closely with predictions.
- The ability to accurately estimate a cancer prognosis will [1] allow the patient to know exactly how many more years he or she can live, [2] help physicians to create treatment plans and eliminate unnecessary or excessive treatment (for example, patients in groups A and B do not need to receive strong anticancer drugs), and [3] allow life insurance companies to calculate accurate insurance premiums and create policies that better suit each individual patient.
- When creating a new drug, it is necessary to determine whether a protein binds to a compound and how strongly they bind. However, examining the approximately 20,000 types of human proteins and 10^{60} types of low-molecular compounds would be extremely difficult.
- Based on the fact that a one-dimensional structure (string) will form a three-dimensional structure (ball) when folded, researchers considered whether there might be a correspondence relationship even with a one-dimensional structure. With this in mind, they created the "LIGHTHOUSE" software, which is used to replace proteins and compounds with one-dimensional numerical vectors and determine relationships with binding force.
- LIGHTHOUSE is capable of making decisions with almost the same accuracy as a three-dimensional docking simulation. It can even made decisions without three-dimensional structure information, making it possible to screen an extremely wide range and large amount of information (LIGHTHOUSE: 6,000 types/min. [when using a PC], docking simulation: two to three types/min. [when using a supercomputer]).
- Researchers were able to use LIGHTHOUSE to discover candidate compounds for new anticancer drugs and drugs effective against COVID-19.

[Prof. Ueda (Tohoku University)]

- "Tohoku University Hospital AI Lab" was created as a verification space in one floor of the hospital, so that data scientists could work with medical personnel to create value in medical care together. The focus here is on research and development of AI based on a design concept that incorporates various perspectives from the medical field.
- The design concept is to take an approach toward creating innovation with a focus on humans. Even in medicine, it is important to first incorporate design concepts and consider the medical value that doctors find important in the medical field rather than prior technology, and to discover "fusion points" with AI that match this.

It is important to develop human resources who can find these "fusion points," and the organization is attempting to develop human resources who are medical practitioners who also understand data science and can discover areas where these fields meet, within a consistent education program that consists of three stages: university undergraduate education (medical school), initial hospital training, and graduate school education.

- The organization has incorporated programs to study design concepts. In addition to programs to study standard design concept processes (such as how to reduce waiting times at hospitals), there are also programs to study processes such as biodesign itself or what to remain mindful of in order to develop AI.
- Tohoku University, Hokkaido University, and Okayama University have teamed up to promote the GLOCAL (global and local) "Clinical AI" project. Most issues in the medical field (including doctor shortages and increasingly aging populations) are not in urban areas but in rural areas, so universities in rural areas are working together to develop human resources who can match AI with solutions to solve these issues. Each university and hospital has its own characteristics, with Tohoku University focusing on hospitals and Okayama University focusing on pharmaceutical department work.
- Medical personnel are aware of medical value, while data science personnel are experts at algorithms. However, there are few people that are experts in both fields. The program is designed to develop human resources capable of bridging both fields, and the hope is to develop human resources who can conduct AI research utilizing design concepts while remaining sensitive to the needs of the medical field and with the values of doctors.

c. Development and utilization of AI for the elderly and people with disabilities

[SoftBank]

- The Company is building a system that allows people to communicate using sign language and speech in real-time, so that hearing-impaired people and others can communicate naturally. This system uses AI to track body movements during video calls (from the hearing-impaired person to the other person), extract sign language characteristics and convert them into text, and automatically transcribe speech (from the other person to the hearing-impaired person), so that both people can communicate more naturally.

[Panasonic]

- In the nursing care field at facilities for the elderly, the Company is using non-contact room sensors to detect the movement, heart rate, respiration, etc. of residents, and to log and analyze life rhythms to improve life, sleep, etc., as a monitoring and peace of mind service. Walking support robots help the elderly during walking training while adjusting the optimum load according to the individual situation.

[JR West]

- The Company has begun developing a system that quickly detects users who need assistance (such as individuals in wheelchairs, using canes, or accompanied by guide dogs), using images from surveillance cameras installed in stations.

[KDDI]

- KDDI Evolva, Inc. (a group company) has installed existing applications and devices for an AI irrigation and fertilization system in Higashi Matsushima City in Miyagi Prefecture, and is working to support reconstruction and create jobs for the disabled through growing agricultural

products. Although the experience and knowledge of growers are important with regard to when and how much to water and fertilize crops, the Company is using AI to help people with disabilities establish stable crop cultivation environments without having to rely on agricultural experience.

[AIDC]

- It can be difficult for people other than supporters to understand what is being said by individuals with cerebral palsy or hearing impairment. The organization is interested in using AI to solve such issues, under the assumption that there must be some common points or correlations.

(3) Human resources development initiatives

It has been pointed out that there is a shortage of human resources related to AI, and the development and securing of human resources has become an issue⁹², so this report summarizes the initiatives⁹³ taken by business operators regarding human resource development⁹⁴.

<Key points of Good Practices>

- Many business operators are actively working to develop AI and digital human resources, and some initiatives are achieving steady results. In addition, business operators and others are engaged in initiatives⁹⁵ and working with universities to develop a 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, and it is considered that this will serve as a reference when examining the ideal form and methods of human resources development.

In particular, 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.

With regard to AI human resources, human resource development is an urgent issue for enhancing Japan's technological capabilities and international competitiveness, and it is expected that AI developers and users, etc. will strongly promote efforts to develop and secure AI human resources while referring to the case studies presented in the interviews⁹⁶.

⁹² Also, AI Strategy 2022 states that "Compared to other countries, such as the United States, which is advanced in AI, Japan has insufficient human resources to support the utilization of AI." and "The development and securing of the relevant human resources is an urgent issue, and is a long-term issue that includes primary and secondary education, higher education, recurrent education, and lifelong education. This issue is being addressed since the formulation of AI Strategy 2019."

⁹³ In addition to in-house development of human resources development by business operators and others, another important perspective is to improve social acceptability. Therefore, initiatives for improving user literacy are also covered.

⁹⁴ Terms used to refer to human resources development vary by business operator, and so the preferred terms for each business operator are used in this report.

⁹⁵ This point is common to "initiatives in human resource development connected to medicine, AI, and data science where design thinking is incorporated" described in the preceding section (development and utilization of AI for medical and healthcare), and it is considered that attention should be paid to it. See Note 91 above.

⁹⁶ As Report 2021 also presents several specific initiative case studies, referring to those examples in conjunction with this report is considered helpful for the development of human resources. See Section 2. (3) and Section 4. in Chapter 3 of Report 2021 and Attachment 3.

[SoftBank]

- The Company is working to develop a wide range of AI human resources, and offers online training courses in data analysis and AI use. It offers a variety of AI training opportunities, including rewarding individuals who pass practical qualification exams (such as statistical exams and algorithmic practical exams), through its certification support system.
- In order to maintain sufficient AI human resources within the Company, it is necessary to have a broad knowledge of computer science topics such as statistics. AI human resources are therefore mainly recruited from universities that provide education in such fields, or through mid-career recruitment. A wide range of OJT training is also provided so that employees can acquire practical knowledge and skills after joining the Company.
- In May 2022, the Company began running an education program for high school students to develop skills using AI for business purposes. It consists of an AI utilization literacy course and a practical AI utilization course. Students are expected to make use of learning data to create AI models and to incorporate AI into actual web services and the Pepper robot, in order to solve familiar problems within their schools and local communities.

[Panasonic]

- Since FY2016, the Company has been working to enhance its AI human resources. In 2021, it was able to train more than 1,000 individuals capable of using AI technology. Each operating company of the Group has established an organization to investigate AI utilization.
- The Company has introduced a system that allows its employees to also work at universities (cross-appointment system), in order to develop more expert employees. These employees continue to be adopted by notable academic societies, and continue to achieve excellent results in international competitions, and are leading the field of AI technology.
- In the area of human resources development, the Company is focusing on improving the quality of its human resources based on the concept of retaining employees. Beginning with developing human resources who are experts in using AI, the Company believes that the development of AI architects who excel in AI engineering and quality assurance will be a priority, and is therefore enhancing its development programs.

[Sharp]

- Since February 2020, the Company began engaging in AI literacy education initiatives in addition to programming education. The aim of AI literacy education is to increase the number of people who can develop AI technologies in order to realize a society where AI is increasingly utilized going forward. RoBoHoN is used to teach people about AI technologies used in everyday life, and to stimulate interest in AI technologies at an early age so that students will have an opportunity to think about AI.
- For example, one task would be to create a program that uses image recognition to distinguish between RoBoHoN and a panda. In addition to creating the program, students actually use the program to make the AI learn and make decisions, allowing them to learn not only by creating the task but also by experiencing how it works.
- During demonstration lessons, children study and experience the most efficient ways to learn as they go through a process of trial and error while creating actual programs and then having the robot learn and make decisions.
Students can be very creative and focused on their own solutions during demonstration lessons, and students have indicated after lessons that they are now more interested in AI.

[Mitsui Sumitomo Insurance]

- The Company has established a system for developing digital human resources within the Company, and is working to develop them by dividing them into those who mainly analyze data and those who mainly conduct digital business.

For the data analysis group, the Company has adopted a three-step curriculum in order to develop specialist human resources: [1] acquire basic knowledge within an information-related department at a partner university, [2] perform practical data analysis after returning to the workplace, and [3] implement analysis projects as business.

For the digital business group, the Company has established a training program in which employees learn about various new business models using the latest digital technologies in courses taught at partner universities, while working alongside their managers (as managers would also need to understand these processes).

[DAIKIN INDUSTRIES]

- The Company's medium-term management plan for 2025 calls for the development of 1,500 digital human resources, mainly through Daikin Information and Communications Technology College, with a vision of promoting business innovation and process innovation.
- Originally, there were very few system personnel to develop and utilize technologies, and in-house training alone would not have been enough to meet the Company's goal. Daikin Information and Communications Technology College was therefore established in December 2017. The Company is working to reach a level where the Company can become independent in three areas required of data scientists: topic execution, analysis, and data engineering.
- In addition to the technical expertise employees bring to the Company, the Company hopes to use this training to develop T-type human resources that are experts in two or more different areas including AI and IoT.
- The Company provides separate training to new employees, existing employees, core management, management, and executives.

For example, during the first year of training for new employees, new hires learn in a classroom setting and are taught using problem-solving exercises (on topics such as building clouds), and take exams such as the Fundamental Information Technology Engineer Examination and Japan Statistical Society Certificate. The Company's pass rate exceeds the national average. During the second year, employees engage in project based learning (PBL), and conduct debriefings for the entire company. Two advantages of this initiative are that it provides the power of collective knowledge, and that employees have time to think carefully about their career over the two-year period. However, there is a barrier between learning the basics and putting them into practice, and the challenge now is how to overcome it.

- The Company believes that it has made steady progress over the past four years. The number of students who want to join the Company after learning about these initiatives has increased significantly, and turnover is very low compared to the average, resulting in a sense of camaraderie among employees.

[JR West]

- Initiatives are divided into two fields (a technical field called "Condition Based Maintenance" [CBM] and a marketing field), and are unique in that the data scientists in charge of business and analysis work as a team overseeing projects and deciding on issues to tackle.
- The Company is currently considering introduction of AI to make determinations on HR and employee benefits. The data science field is a largely unexplored area in the railway industry,

and the Company's initiatives are truly industry-first efforts. They are worth pursuing for that reason. However, this will not be enough in the future, so wages, authorities, team organizations, and work styles/speeds will need to change.

- There are two career paths for employees with data scientist knowledge. The first is to feed technology back into railway bodies as a member of an organization responsible for advanced data analysis. However, the Company will also need to allow employees in the field to conduct data analysis in a more general manner. The other pattern is therefore to have an employee join a digital team for a certain period of time in order to acquire a background in data analysis, and then return to their business division and make improvements to work in the field.

[KDDI]

- It is often difficult to make employees aware of rules and principles that have been created. However, employees working with AI have created a community to hold short talks discussing examples of using AI, and to promote the importance of governance (such as ethical guidelines). The Company also conducts e-learning for all employees, and makes use of internal PR to disseminate information.
- Rather than developing AI human resources, the Company is engaged in developing DX human resources. One example is that the Company is creating courses on personal information protection and AI ethics. The Company is promoting initiatives in a balanced manner by making its employees aware of the importance of privacy and AI ethics before engaging in related projects.

(4) AI supply chain initiatives

The process from development to application of AI relies on numerous processes, such as data acquisition, processing, AI model construction, and integration into systems. The following section summarizes the risks, issues, and solutions for the AI supply chain⁹⁷.

<Key points of Good Practices>

- It has become clear that there are risks in various processes such as data acquisition and processing, and algorithm and model generation in the AI supply chain, and initiatives such as checks of data providers and the guaranteeing of work through contracts can be seen among AI developers and service providers. In particular, attention should be paid to initiatives for building a data distribution infrastructure. 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 users, etc. 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 in the interviews.

[NTT DATA]

- Data and software (both components of AI) are intangible assets that are neither

⁹⁷ There is no clear definition of the AI supply chain, but in this report the series of processes from data acquisition, processing, AI model creation, to integration into systems, etc. are collectively referred to as the AI supply chain.

warehoused nor transported like conventional goods, so the risk in the AI supply chain is whether the services, data, and software products the user wants can be used when and on the terms and conditions they want.

- First, the Company creates a learning/inference program similar to OSS based on published papers and innovative algorithms, processes data from the field or society (original data), and then creates data for learning (processed data). Next, the Company creates a trained model based on the learning/inference program and the processed data, and incorporates it into actual software and systems. Ultimately, the entire AI supply chain would likely operate software and systems to obtain output.
- Within such a process, risks related to intellectual property and licensing, risks related to data privacy, ethical risks such as data bias (for example, whether biased data is used), risks related to quality such as whether the expected benefits are obtained, and risks related to security such as data leaks and unintended output results (leading to inferences) would be considered risk factors in the AI supply chain.
- With regard to intellectual property and licenses, there is a risk that they may become unusable due to changes in provision or disclosure policies, or that additional fees may be required due to a change in pricing policy, even if a product was previously available for free.
- With regard to privacy and intellectual property, there is a risk that products may be used for purposes other than the intended purposes, and that data provided through APIs may be reused. For this, it is important to carefully check the terms of use. (Repeated)
- With regard to ethics, there is a risk regarding whether data is obtained legally and ethically. For this, it is important to check whether there are any problems with individual suppliers and procurement specifications according to the sensitivity of the data being handled. The Company also believes that establishing a mechanism for obtaining third party endorsements of the provider's data governance concept and methods could allow for data to be used with peace of mind.
- There is also a risk of data bias, as bias in labels can lead AI to make inferences lacking fairness. For this, it is important to determine the extent to which parameters related to universal human attributes (such as gender) affect output. (Repeated)
- With respect to quality, there is a risk regarding whether the accuracy of an externally procured product matches the required task. For this, there is currently no choice but to verify accuracy using a PoC based on the task. However, it would be preferable to establish AI quality standards to achieve certain guidelines. (Repeated)
- With regard to security, there is a risk of products containing backdoors as there is no way of knowing what was used as a base for creating models and programs procured externally. For this, there is no choice but to check source code and make judgments based on the trustworthiness of the provider. The Company believes that conducting thorough legal investigations/reviews would be a realistic approach for companies to take.

There is also a risk of data being stolen through backdoors. For this, the Company believes some kind of public endorsement of dominant components or products would bring peace of mind to users. (Repeated)

[FUJITSU]

- The Company believes that discussions of AI supply chain practices can be limited to those typical of major transactions. With regard to the distribution of algorithms, there are cases where the Company provides these to other companies, and cases where the Company

receives these from other companies. With regard to the distribution of data, in most cases data is prepared by the Company's users or joint research partners. In some cases, the Company purchases data and resells it to other companies, but this is rare.

- Legal requirements in the AI supply chain are mostly the same as those in conventional supply chains. The Company is not currently aware of any special legal requirements required for AI, due to the limited nature of transaction patterns. With regard to contractual measures, if the data subject is an individual, procedures such as obtaining consent are already being carried out within existing frameworks. However, technical limitations and ethical issues must be addressed jointly with users. When the Company obtains data from data collectors, it guarantees the data by representing and warranting the legitimate authority, etc. of the data collectors in contracts, and conducting audits as necessary.
- As an example of the Company providing an algorithm to another company (automatic analysis of measurement data), the user company collects and annotates data, shares this data with the Company based on a joint research contract, and then the Company develops an AI model. From a system development perspective, the Company uses data from the user company to develop an AI model and AI software with this model embedded, and then provides a license to the user company. The user company develops software with the provided AI software embedded, and then sells this to its end users.
- As an example of the Company receiving an algorithm provided by another company and using it to provide a service (translation service) to other users, during the product development stage the partner company uses its own data (which cannot be accessed by the Company), and end users of the service upload files to the Company's cloud and download the created translations, after which the uploaded data and translations are deleted. From a system development perspective, the Company provides its own service incorporating the service of a partner company, with quality guaranteed by the partner company.
- As an example of the Company developing an algorithm (for nursery school admissions) and providing it to other parties (local governments), data is shared in order to verify the algorithm and its effectiveness during verification experiments. However, the model does not incorporate resident data because it uses a mathematical optimization algorithm rather than a machine learning algorithm. From a system development perspective, the Company transfers algorithms (with effectiveness verified during verification experiments) from the Research Division to the Business Division, and sells the software with embedded AI functionality for use by the end user.
- The Company believes that it is important to apply technology in addressing challenges unique to the AI supply chain. As for issues such as reduced accuracy, the operator has an important role in performing monitoring and correction, and this requires supporting technology. For example, the Company is developing technology related to quality assurance in order to help prevent performance degradation during operation. The Company believes that applying this technology throughout the supply chain will be a step toward resolving the unique challenges of AI.

[AIDC]

- With regard to data distribution, the value of data often changes depending on the degree of data bias and the passage of time. For example, natural language processing requires dialect data and data on the elderly, on a daily basis. However, this data is difficult to obtain, and is therefore valuable in the data distribution market.
- Data is generally subject to general controls in information systems. However, as more

organizations begin applying AI and machine learning solutions, it will become important to ensure control and accountability.

- There is a need for a data distribution transaction platform to meet these needs, and the organization has been promoted initiatives to build a platform for data transactions and contracts. In March 2022, the organization built a platform for data transactions (AIDC Data Cloud) and began providing services. There is a wide variety of usage cases with respect to contracts and billing models, so the organization is currently working with others to implement mechanisms for tasks such as creating dynamic contract templates. The data platform can also be used to what was used as a basis for labelling data, as well as the historical trail of development data.

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

As described in the previous chapter, the AI R&D Guidelines and the AI Utilization Guidelines (hereinafter referred to as the "two guidelines") were formulated by this Conference, and the two guidelines have contributed to international discussions at the OECD, G7, and G20, etc. Following that, while following domestic and international trends and trends in international discussions, in order to promote the "Safe, Secure, and Trustworthy Implementation of AI in Society" this Conference conducted interviews with stakeholders⁹⁸, including AI developers, service providers, business users, and consumer-type users, and their initiatives are summarized in Report 2020 (July 2020) and Report 2021 (August 2021).

Five years have passed since the establishment of the AI R&D Guidelines, and three years have passed since the establishment of the AI Utilization Guidelines, and during this time, many principles, policies, and guidelines on AI ethics and governance (hereinafter referred to as "guidelines and policies") have been formulated in Japan and overseas. In addition, after formulating the AI Utilization Guidelines, this Conference and the Chair of this Conference conducted interviews (hereinafter referred to as "interviews, etc.")⁹⁹ that went beyond the scope of the two guidelines to show¹⁰⁰ what could be evaluated as excellent initiatives for the "Safe, Secure, and Trustworthy Implementation of AI in Society."

Based on this situation, this chapter reviews the AI R&D Guidelines and AI Utilization Guidelines, and organizes issues related to the reviews¹⁰¹.

1. Reviews of the AI R&D Guidelines and the AI Utilization Guidelines

(1) Overview of the Reviews

As mentioned above, several years have passed since the two guidelines were formulated, and during this period, many guidelines and policies on AI ethics and governance have been formulated overseas and by domestic business operators and others. Therefore, it is important to verify the scope and content of the two guidelines by comparing them with these guidelines and policies.

In addition, it has become clear through the interviews, etc., that excellent initiatives are being made for the "Safe, Secure, and Trustworthy Implementation of AI in Society" that go beyond the scope of the two guidelines. Therefore, it is important to examine the scope and content of the two guidelines based on the response to the initiatives presented in the interviews, etc.

Furthermore, while AI utilization and social implementation continues to be promoted in

⁹⁸ See Note 2 above.

⁹⁹ It should be kept in mind that those interviews were conducted to compile Report 2021 and are different from the interviews conducted to compile this report. For an overview of those interviews, see Section 1. in Chapter 3 of Report 2021. For an overview of the interviews conducted to compile this report, see Section 1. in the previous chapter of this report.

¹⁰⁰ Mainly refer to Section 2. (1) in Chapter 3 Report 2021. In addition, we conducted interviews with business operators regarding more specific initiatives related to AI governance, summarized the cases, and published Examples of Initiatives Related to AI Governance (September 2021). Of the initiatives covered in interviews, this document picks up the initiatives evaluated as being excellent with respect to the "Safe, Secure, and Trustworthy Implementation of AI in Society."
Available at the website of the following URL.

<https://www.soumu.go.jp/main_content/000770820.pdf> (Japanese version)

<https://www.soumu.go.jp/main_content/000785877.pdf> (English version)

¹⁰¹ Section 3. (1) [2] in Chapter 3 of Report 2021 points out that it is important to review the AI R&D Guidelines and AI Utilization Guidelines, and review their positioning, scope, and principles as necessary.

Japan and overseas, there have been cases in which risks have become apparent or incidents have occurred, so it is important to examine the scope and content of the two guidelines based on such cases.

Based on the above, the two guidelines will be reviewed from the following four perspectives¹⁰².

- Comparison with overseas guidelines and policies
- Comparison with domestic guidelines and policies
- Consistency with initiative case studies presented in the interviews, etc.
- Consistency with case studies of AI utilization and social implementation

(2) Verification through comparison with overseas guidelines and policies

The scope and content of the current AI R&D Guidelines and AI Utilization Guidelines are verified based on guidelines and policies formulated overseas¹⁰³.

Regarding verification, the survey covered 17 countries/regions (United States, Canada, United Kingdom, Italy, Netherlands, Sweden, Denmark, Germany, Norway, Finland, France, India, South Korea, Singapore, China, Australia and the EU), and two international organizations (UNESCO and WHO)¹⁰⁴. In addition, 67 guidelines and policies published by government agencies, industry associations, and global enterprises in these countries and regions were surveyed¹⁰⁵.

The following 22 "values to be respected" were confirmed through the survey of the guidelines and policies. Of these, "1. Human-centered" to "16. Accountability" are the same values shown in the AI guidelines comparison table (Attachment 2 of Report 2019), and "17. Robustness" to "22. Other (cost, effectiveness measurements)" are newly identified values¹⁰⁶.

- | | | |
|------------------------------|------------------------------------|--------------------------|
| 1. Human-centered | 9. Proper data | 17. Robustness |
| 2. Human dignity | 10. Collaboration among AI systems | 18. Responsibility |
| 3. Diversity, Inclusiveness | 11. Safety | 19. Traceability |
| 4. Sustainable society | 12. Security | 20. Monitoring, Auditing |
| 5. International cooperation | 13. Privacy | 21. Governance |

¹⁰² Since it is considered important to respond based on both domestic and international trends such as the announcement of the European Commission's Artificial Intelligence Act, as well as trends in international discussions, such points will also be taken into account in the review.

¹⁰³ 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.

¹⁰⁴ 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 each of these categories were selected. The EU and two international organizations were added to the survey, given overseas trends and directions of international discussions.

¹⁰⁵ The survey generally concentrated on guidelines and policies announced in August 2019 or later, when the AI Utilization Guidelines were announced. The breakdown is as follows: 40 government agencies (including the European Commission and the High-level expert group on artificial intelligence) and international organizations, 10 industry associations, etc., and 17 global companies. For a list of the guidelines and policies, see pages 16 to 23 in Attachment 2.

¹⁰⁶ For example, as a concept similar to "robustness," the Principles of Safety in the AI R&D Guidelines state that it is necessary to remain mindful of the reliability and robustness of AI systems. In addition, as a concept similar to "traceability," the Principles of Transparency in the AI Utilization Guidelines state that logs including input/output are expected to be recorded and stored. As such, there are cases where it is difficult to make a clear distinction between existing values and newly identified values, but they have been arranged based on the wording and context given in the guidelines and policies.

- 6. Proper utilization
- 7. Education/Literacy
- 8. Human intervention, Controllability

- 14. Fairness, Equity, removal of Discrimination
- 15. Transparency, Explainability
- 16. Accountability

- 22. Other (cost, effectiveness measurements)

Table 1 shows how many of these "values to be respected" are described in the guidelines and policies (of the 67 guidelines and policies, 40 guidelines and policies published by government agencies¹⁰⁷ and international organizations in each country are listed).

(Table 1)

	No. of guidelines	Human-centered	Human dignity	Inclusivity, 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							
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

Note: 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.

In addition, Table 2 clearly shows for North America, Europe, Asia, and Oceania¹⁰⁸ that "transparency, explainability," "fairness, equity, removal of discrimination," "accountability," and "human dignity" are emphasized in all four regions.

¹⁰⁷ See Note 105 above.

¹⁰⁸ Of the 40 guidelines and policies summarized in Table 1, those of international organizations are excluded (2) and 38 guidelines and policies are summarized.

(Table 2)

	No. of guidelines and policies	Human-centered	Human dignity	Inclusiveness, Diversity, Inclusion	Sustainable society	International cooperation	Proper utilization	Educational literacy	Human intervention, Controllability	Proper data (quality of learning data)	Collaboration among AI systems	Safety	Security	Privacy	Fairness, Equity, removal of Discrimination	Transparency, Explainability	Accountability	Robustness	Responsibility	Traceability	Monitoring, Auditing	Governance	Other
North America	7	1	2			1	1	1	1	1		3	3	1	4	6	4	1	2	2	3		1
Europe	23	1	14	7	7	1	8	2	7	7		8	6	12	18	20	14	8	8	6	8		1
Asia	7		6	4	2	3	2	2	3	1		5	3	4	7	6	5	3	1	2	3	3	
Oceania	1	1	1									1	1	1	1	1	1		1		1		
Total	38	3	23	11	9	5	11	5	11	9	0	17	13	18	30	33	24	12	12	10	15	3	2

Note: 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.

	North America	Europe	Asia
	<ul style="list-style-type: none"> [1] Transparency, Explainability [2] Fairness, Equity, Removal of discrimination [2] Accountability [4] Safety [4] Security [4] Monitoring, Auditing 	<ul style="list-style-type: none"> [1] Transparency, Explainability [2] Fairness, Equity, Removal of discrimination [3] Human dignity [3] Accountability [5] Privacy 	<ul style="list-style-type: none"> [1] Fairness, Equity, Removal of discrimination [2] Human dignity [2] Transparency, Explainability [4] Safety [4] Accountability

Furthermore, for each "values to be respected," descriptions considered to be relevant in the two guidelines and descriptions in the various surveyed guidelines and policies are summarized as shown in the following example. All 22 items are summarized in pages 5 to 15 of Attachment 2.

By arranging descriptions in this way, it became clear that while there are many instances of similar concepts being used in the two guidelines and the overseas guidelines and policies, there are also differences¹⁰⁹ in the details between the two guidelines and the overseas guidelines and policies despite sharing existing "values to be respected."

Values that should be respected	Content considered relevant in guidelines (main)	Examples of main content
1. Human-centered	<p>[AI R&D Guidelines]</p> <ul style="list-style-type: none"> ○ Basic philosophy • Realize a human-centered society in which human dignity and individual autonomy are respected <p>[AI Utilization Guidelines]</p> <ul style="list-style-type: none"> ○ Basic philosophy • Realize a human-centered society in which human dignity and individual autonomy are respected 	<ul style="list-style-type: none"> • Remain mindful of the development and use of human-centered AI to strengthen national security and trusted partnerships. • Systems must focus on those who use it or are affected by its decisions. • AI systems must emphasize human-centered values and respect human rights, diversity, and individual autonomy.
2. Human dignity	<p>[AI R&D Guidelines]</p> <ul style="list-style-type: none"> ○ Principles of ethics • Respect human dignity and individual autonomy <p>[AI Utilization Guidelines]</p> <ul style="list-style-type: none"> ○ Principles of dignity and autonomy • Respect human dignity and individual autonomy 	<ul style="list-style-type: none"> • AI systems must be designed to be consistent with international treaties protecting human dignity, rights, and freedoms. The organization takes responsibility for implementing measures to avoid reasonably predictable misuse through the use of ethics by design. • It is necessary to respect the dignity and autonomy of individuals equally, regardless of nationality, gender, national or ethnic origin, ethnicity, religion, or language. • Respect human rights in accordance with common human values, pursue human interests, and remain mindful of protecting national and regional ethics and morals.

¹⁰⁹ For example, in the two guidelines, "16. Accountability" is mainly used to mean responsibility to explain. However, in some overseas guidelines and policies, it is used in a broader sense to mean not only the responsibility to explain but also to respond and provide relief, etc. in the event of an incident. As a specific example, the ETHICS AND GOVERNANCE OF ARTIFICIAL INTELLIGENCE FOR HEALTH document published by the World Health Organization (WHO) states that if something goes wrong with an AI technology, there should be accountability. Available at the website of the following URL.

<<https://apps.who.int/iris/rest/bitstreams/1352854/retrieve>>

As described above, as a result of verification based on a comparison with overseas guidelines and policies, items including the following were confirmed.

- Guidelines and policies from other countries 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."
- There are also differences¹¹⁰ in the details between the two guidelines and the overseas guidelines and policies despite sharing existing "values to be respected."

(3) Verification through comparison with overseas guidelines and policies

The scope and content of the current AI R&D Guidelines and AI Utilization Guidelines were verified based on guidelines and policies formulated by domestic business operators and organizations, etc¹¹¹.

Regarding verification, the survey covered 22 guidelines and policies published by business operators and organizations in Japan¹¹².

The business operators and organizations, etc. that published the surveyed guidelines and policies are as follows.

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

Note: Company names at the time of the guideline announcements

Through the survey of the guidelines and policies, 22 "values to be respected" as in (2) above were confirmed, of which "17. Robustness" to "22. Cost"¹¹³ were newly confirmed.

Table 3 shows how many of these "values to be respected" are described in the guidelines and policies. It should be noted that expectations regarding which "values to be respected" should be emphasized varies depending on the industry or field of the respective business operator or organization, the purpose or application of AI, and other factors.

¹¹⁰ See Note 109 above.

¹¹¹ See Note 103 above.

¹¹² For a list of guidelines and policies, see pages 36 and 37 in Attachment 2.

¹¹³ Regarding the 22nd value, in the guidelines and policies instituted in other countries, "effectiveness measurements" were included, so the 22nd value was listed as "22. Other (cost, effectiveness measurements)" In the guidelines and policies instituted in Japan, however, there were no instances of "effectiveness measurements," so the 22nd value is listed here as "22. Cost."

(Table 3)

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

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

In addition, for each "values to be respected," descriptions considered to be relevant in the two guidelines and descriptions in the various surveyed guidelines and policies are summarized as shown in the following example. All 22 items are summarized in pages 26 to 35 of Attachment 2.

By arranging descriptions in this way, it became clear that the two guidelines and the guidelines and policies instituted in Japan generally use many of the same concepts.

Values that should be respected	Content considered relevant in guidelines (main)	Examples of main content
3. Diversity, Inclusiveness	<p>[AI R&D Guidelines]</p> <ul style="list-style-type: none"> ○ Basic philosophy <ul style="list-style-type: none"> • Realize a human-centered society in which humans and AI networks co-exist, the benefits of AI networks are widely enjoyed by all, and human dignity and individual autonomy are respected ○ Principles of user support <ul style="list-style-type: none"> ➢ Strive to engage in initiatives to promote use by the socially weak, such as implementing universal design ○ Roles expected of related stakeholders <ul style="list-style-type: none"> • In implementing and reviewing these guidelines, it is expected that organizations will strive to develop an environment that promotes dialog between diverse stakeholders, including national governments, international organizations, developers, and users (including civil society) • Organizations are expected to strive to share awareness and engage in mutual cooperation in promoting the benefits and reducing the risks of AI, while ensuring the diversity of discussions on AI <p>[AI Utilization Guidelines]</p> <ul style="list-style-type: none"> ○ Basic philosophy <ul style="list-style-type: none"> • Realize a human-centered society in which humans and AI networks co-exist, the benefits of AI networks are widely enjoyed by all, and human dignity and individual autonomy are respected • Respect the diversity of users and include people with diverse backgrounds, values, and thinking when making use of and applying AI 	<ul style="list-style-type: none"> • Strive to respect the human rights and diversity of customers and stakeholders so as not to cause unfair discrimination in the use of AI. • While working to realize safe, robust, and easy-to-use products and services, actively promote the use of AI so that more people can enjoy the benefits of AI. In particular, aim to lead the way in solving social issues in a wide range of fields including healthcare, by rapidly creating new value through the integration of advanced and proprietary technologies developed over the years and the promotion of co-creation with outside companies. • Actively develop and expand AI technologies, based on the philosophy that AI technologies that support the realization of diverse lifestyles of diverse people are the future of mankind and meant to realize the diverse dreams of diverse people.
4. Sustainable society	<p>[AI R&D Guidelines]</p> <p>(No content)</p> <p>[AI Utilization Guidelines]</p> <ul style="list-style-type: none"> ○ Basic philosophy <ul style="list-style-type: none"> • Realize a sustainable society by solving various problems faced by individuals, local communities, countries, and the international community 	<ul style="list-style-type: none"> • Strive to use the power of AI to help solve global problems and to develop a peaceful and sustainable society. • Plan to develop and utilize AI to solve diverse social issues, realize a comfortable and resilient sustainable society, and improve the quality of life of people around the world. • Aim to realize a sustainable future society by using AI, etc. to solve various social issues.

As described above, as a result of verification based on a comparison with overseas guidelines and policies, items including the following were confirmed.

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

(4) Verification of consistency with initiative case studies presented in the interviews

The scope and content of the current AI R&D Guidelines and AI Utilization Guidelines were

mainly examined based on initiatives, etc. of business operators in Japan¹¹⁴ as indicated in "Report 2021" and "Examples of Initiatives Related to AI Governance" (September 2021)¹¹⁵.

With regard to "diversity," "security," "privacy," "fairness, equity, removal of discrimination," "transparency and accountability," "appropriate use," and "quality management," which are considered to be directly and deeply related to the objectives, basic philosophies, principles, etc. of the two guidelines, descriptions and examples of initiatives by business operators considered to be related to the two guidelines are summarized as shown in the following example. All seven items are summarized in pages 38 to 41 of Attachment 2.

By arranging descriptions in this way, it became clear that various initiatives, including ensuring diversity and utilizing new technologies, are being made to ensure security and fairness, to protect privacy, and realize explainability and accountability, and that appropriate use with human intervention is being carried out through "human in the loop"¹¹⁶ operation. In addition, regarding quality management, it became clear that excellent initiatives such as the development of evaluation processes throughout the lifecycle are being carried out.

Item	Content considered relevant in guidelines (main)	Examples of specific initiatives
Diversity	<p>[AI R&D Guidelines]</p> <ul style="list-style-type: none"> ○ Principles of ethics <ul style="list-style-type: none"> • Respect human dignity and individual autonomy <p>[AI Utilization Guidelines]</p> <ul style="list-style-type: none"> ○ Basic philosophy <ul style="list-style-type: none"> • Respect the diversity of users and include people with diverse backgrounds, values, and thinking when making use of and applying AI ○ Principles of dignity and autonomy <ul style="list-style-type: none"> • Respect human dignity and individual autonomy ○ Principles of fairness <ul style="list-style-type: none"> • Make arrangements so that individuals and groups are not unfairly discriminated against 	<ul style="list-style-type: none"> • Emphasize the diversity of human resources, and focus on how extremely diverse humans providing input are. This includes both men and women, as well as scholars from different cultures. • In addition to computer science researchers, the research institute also includes social science researchers working in psychology, fieldwork, and cognitive science. It is important to include the viewpoints of psychology and cognitive science. • R&D and social implementation of safe, secure, and reliable AI requires people with various backgrounds. A team specializing in AI ethics has been formed, including researchers who have been engaged in design and social psychology.
Security	<p>[AI R&D Guidelines]</p> <ul style="list-style-type: none"> ○ Principles of security <ul style="list-style-type: none"> • Remain mindful of AI system security <p>[AI Utilization Guidelines]</p> <ul style="list-style-type: none"> ○ Principles of security <ul style="list-style-type: none"> • Remain mindful of AI system and AI service security ○ Principles of appropriate learning <ul style="list-style-type: none"> ➢ Organizations are expected to remain mindful of the risk of AI learning inaccurate or inappropriate data, leading to vulnerabilities in AI security 	<ul style="list-style-type: none"> • With regard to AI security, the spread of AI has led to attacks against AI (duping AI, stealing AI information, etc.) and threats using AI (using AI for deception, etc.). • There are four perspectives for considering the relationship between AI and security: (a) Attack using AI, (b) Attack by AI, (c) Attack to AI, and (d) Measure using AI. It is important to promote discussion that considers not only attacks against AI (c) but also other issues related to AI.

As described above, as a result of verification of consistency with the case studies of initiatives presented in the interviews, the following items were confirmed.

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

(5) Verification of consistency with case studies of AI utilization and social implementation

AI utilization and social implementation is being promoted both in Japan and overseas, but there have been cases where issues have been pointed out, such as the emergence of risks and the occurrence of incidents. The scope and content of the AI R&D Guidelines and the AI Utilization Guidelines were verified based on these cases¹¹⁷.

Regarding verification, the survey covered seven countries/regions (Japan, United States, United Kingdom, Germany, France, China, and the EU) and three international organizations

¹¹⁴ Some of the initiatives summarized in "Report 2020" are included.

¹¹⁵ See Note 100 above.

¹¹⁶ See Note 88 above.

¹¹⁷ See Note 103 above.

(UNESCO, WHO and the World Economic Forum (WEF)), and 14 documents¹¹⁸ published by governmental agencies of these countries/regions were examined.

Examples of AI utilization and social implementation are categorized into "biometric authentication," "autonomous driving," "image identification," "medical," "crime prediction," "recidivism prediction," "hiring" "admissions," "performance/HR evaluation," "advertising," and "COVID-19 measures."

For each of these categories, cases where issues have been pointed out are summarized as shown in the following example. All of these categories are summarized in pages 42 and 43 of Attachment 2.

	Issues, etc. identified from case studies of AI development, utilization, and social implementation (overview)
Biometric authentication	<ul style="list-style-type: none"> In the United States, several major tech companies decided not to offer facial recognition technology to police after a black man died from being pinned to the ground by the neck by a white police officer. Some companies have also announced that they will not provide services using facial recognition technology until national laws based on human rights have been established. It is necessary to examine legal systems, etc. concerning biometric authentication technology including facial recognition MIT scientists have revealed bias based on ethnicity and gender in the facial recognition systems of major tech companies; it is necessary to ensure fairness. The privacy policy of a Chinese face app company indicates that user face data is provided free of charge and that the rights to these images belong to the company in perpetuity; it is necessary to ensure privacy. An AI company in the United States has used 3D masks to bypass the facial recognition system of a major Chinese tech company; it is necessary to ensure security. The Face ID (facial recognition) system can be deceived by attaching a printout to a hat; it is necessary to ensure robustness.
Autonomous driving	<ul style="list-style-type: none"> Autonomous vehicles have been involved in fatal accidents. Safety must be ensured, and the locus of responsibility between AI system developers and users must be clarified. It is possible to deceive sign recognition technology by attaching stickers to signs; it is necessary to ensure robustness.

In addition, with regard to the cases where issues have been pointed out, of the "values to be respected" identified in (2) and (3) above, the values related to these cases are summarized in Table 4.

(Table 4)

	Human-centered	Human dignity	Diversity, Inclusiveness	Sustainable society	International cooperation	Proper utilization	Education, Literacy	Human intervention, Controllability	Proper data (quality of learning data)	Collaboration among AI systems	Safety	Security	Privacy	Fairness, Equity, removal of Discrimination	Transparency, Explainability	Accountability	Robustness	Responsibility	Traceability	Monitoring, Auditing	Governance	Cost	
Biometric authentication	★	★	★			★		★				★	★	★									
Autonomous driving	★	★				★				★	★	★			★			★					
Image identification												★					★						
Medical	★	★	★			★		★					★	★	★	★							
Crime prediction	★	★	★			★		★						★	★	★							
Recidivism prediction	★	★	★			★		★					★	★	★	★							
Hiring	★	★	★			★		★					★	★	★	★							
Admissions	★	★	★			★		★						★	★	★							
Performance/HR evaluation	★	★	★			★		★					★	★	★	★							
Advertising	★	★	★			★						★	★	★									
COVID-19 measures	★	★	★			★		★				★	★	★									

Note: Values mentioned in guidelines, etc. are colored in pink.

As described above, as a result of verification of the consistency with case studies of AI utilization and social implementation, the following items were confirmed.

- In many cases where risks have actualized or incidents have occurred, the values

¹¹⁸ White papers and reports, etc. issued by government agencies are called "documents." For a list of the documents, see page 45 in Attachment 2.

(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 utilization and social implementation.

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

(1) Examination of the arrangement of discussion points regarding revisions of the AI R&D Guidelines and the AI Utilization Guidelines

The review of the AI R&D Guidelines and the AI Utilization Guidelines mentioned in 1. above confirmed that many guidelines and policies in Japan and overseas include "values to be respected" that are not included in the two guidelines, and that some initiatives being taken by business operators are excellent in terms of the "Safe, Secure, and Trustworthy Implementation of AI in Society," which go beyond the scope of the two guidelines.

Furthermore, the following opinions were expressed regarding the review of the two guidelines at this Conference¹¹⁹.

- Newly identified values to be respected, such as "robustness," "responsibility," "traceability," and "monitoring, auditing" are aspects of guidelines for the operational phase, and it is important to develop these concepts.
- Regarding risk control to minimize risks, how the architecture is built and how guidelines are written becomes important.
- As AI becomes increasingly interdependent, it is necessary to establish guidelines while grasping the situation.
- At present, frameworks shared internationally, such as the OECD, the G20, and UNESCO, have been formed, and the discussion phase has progressed. Therefore, treating the AI R&D Guidelines as a document that has played a particular historical role, we believe that the time has come to review its title, structure, and content. It is necessary to think carefully about the format and kind of content to share within the country.
- The focus of the discussion should be on how to minimize the number of principles to the extent possible and how to introduce them into corporate governance.
- Given the increasing utilization of AI in a variety of situations, the scope of the AI R&D Guidelines and AI Utilization guidelines should be examined more. (...) Regarding the differences between AI and conventional technology and what the problems are, I think that if we do not narrow down the points on the limitations and decide what the AI guidelines should be, it will be difficult to use.
- In considering the review of the Guidelines, considerations should be appropriate for the current situation while also taking into account trends in the EU and other countries.

¹¹⁹ Opinions at the joint meeting of this Conference and AI Governance Review Committee (February 8, 2022, and April 27 of the same year).

- At UNESCO, culture and diversity have become key words. Going forward, I think it's necessary to create an ideal society with the aims of culture, diversity and the SDGs firmly in mind.
- It is expected that the supply chain will also be taken into consideration, so I believe that it is worthwhile to consider providing guidelines.
- From the perspective of whether end users can understand the guidelines, they should be easy to understand not only for junior and senior high school students but also for adults.

In addition, Report 2021 points out that it may be necessary to review the two guidelines from the perspective of appropriately striking a balance between improving and promoting public health and protecting privacy, based on an international comparison of COVID-19 measures that utilize AI¹²⁰.

Based on these considerations, this Conference will study revising the two guidelines and will summarize the discussion points pertaining to revisions with aim of making the two guidelines more useful and beneficial for AI developers and users and to further promote the "Safe, Secure, and Trustworthy Implementation of AI in Society."¹²¹

In considering the review, it is important to consider how to ensure the effectiveness of each respective principle in the two guidelines. 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 standardization¹²² will lead to the promotion of the "Safe, Secure, and Trustworthy Implementation of AI in Society." In light of this, in order to realize a human-centered society, this Conference will study revising the two guidelines from the perspective of what kind of guidelines are desirable as the embodiment of the "Social Principles of Human-Centric AI."¹²³

(2) Arrangement of discussion points regarding revisions of the AI R&D Guidelines and the AI Utilization Guidelines

Based on the review of the two guidelines, the opinions of this Conference on the review of the two guidelines, and also domestic and international trends and trends in international discussions, etc., in studying the review of the two guidelines, four issues are summarized as follows.

¹²⁰ See Section 3. (2) in Chapter 2 of Report 2021. Regarding this, Report 2020 points out that it is important to consider the priority and importance of each principle in order to implement emergency responses to avoid serious damage even if some of the principles are relaxed (see Note 117 in Report 2020).

¹²¹ A basic philosophy of each of the two guidelines is to constantly review them and flexibly make revisions as necessary.

¹²² For example, regarding the purposes of AI R&D Guidelines, with these two guidelines, of the matters that are expected to be considered regarding the development of AI, this Conference will stipulate matters that are expected to be considered in terms of matters common to fields where AI is utilized and in relation to collaboration between fields. For matters that are expected to be considered according to the circumstances in each field, separate from these two guidelines, it is expected that the relevant stakeholders, including international organizations in each field, will discuss how guidelines should be formulated in each field, including whether or not they should be formulated.

¹²³ See Note 20 above.

- Whether or not to add or revise the purposes, basic philosophies, scope (definitions) of the AI covered, principles, and explanations, etc. In addition, whether or not to review the structures, including adding new content, etc., based on advanced cases, etc.

For example,

- > 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 government organizations? Etc.

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

For 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 and 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? Etc.

- 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?

For example,

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

For example,

- > 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 (government organizations), or the establishment of check sheets or certification systems¹²⁴?

- > What actions will be necessary in relation to international governance frameworks, given the EU's Artificial Intelligence Act is published and being discussed in this area?

At this Conference, in addition to the opinions expressed in (1) above, for example, the following opinions were also expressed regarding these issues¹²⁵.

- Regarding monitoring and auditing, it is assumed that there will be cases where it will not be possible for humans to be involved in time. It would be good if Japan's uniqueness could be brought out in incorporating perspectives such as how human intervention is desirable in the so-called "human in the loop."¹²⁶
- Consideration should be given to the different concepts of monitoring and auditing, depending on whether responding after obtaining the results of an AI operation or responding by stopping AI while it is running, etc. Furthermore, human intervention is not fast enough for online auditing so it will be done using AI technologies, etc., but off-line auditing can be performed by humans taking the time required. I think this difference should be clarified to make it easier to understand for the reader.
- The EU's Artificial Intelligence Act also details human intervention for high-risk AI operators. Although no conclusion has been reached as to how and in what ways intervention is desirable, I think it is meaningful to provide a direction¹²⁷. It would also be good to include a perspective on the human-machine interface.
- I think it's time to put more emphasis on the perspective of AI networking.
- In a society where AI is used, I think it would be helpful as a guideline for AI developers to have some direction on how to interpret trust.
- From a security perspective, I think it is important to clarify how blockchain should be positioned in the guidelines.
- Currently, like with autonomous driving, individual AIs communicate with each other and operate in a connected manner. In terms of security, it is important to prevent AI that has

¹²⁴ In considering this point, for example, the Information Disclosure Guidelines for Safety and Reliability of Cloud Services Using AI (ASP/SaaS Edition) and the Information Disclosure Guidelines for Safety and Reliability of ASP / SaaS (AI Cloud Services) may be helpful.

See Section 1. (4) in Chapter 1 for details on the former information disclosure guidelines. In order to certify that information that confirms an AI cloud service is safe and secure is being disclosed, on April 1, 2022, the Japan Cloud Industry Association (ASPIC, formerly the ASP-SaaS-AI-IoT Cloud Industry Association) established a new information disclosure certification system for the safety and reliability of ASP and SaaS (AI cloud services). These kinds of initiatives may be helpful.

¹²⁵ Opinions at the joint meeting of this Conference and AI Governance Review Committee (February 8, 2022, April 27 and June 22 of the same year).

¹²⁶ Matters concerning the intervention of human judgment are described in the principle of proper utilization in the AI Utilization Guidelines, and it is considered important that this point be brushed up.

¹²⁷ Attachment 1 (Annexed document), Detailed Explanation on Key Points Concerning AI Utilization Principles, of Report 2019 provides five perspectives that can be considered criteria (nature of end users rights, benefits and intention affected by AI's decision, reliability of AI's decision (compared with that of human decisions), allowable time necessary for human decisions, expected ability of users making decisions, necessity for protecting target for decision (for example, whether it is a response to an individual application by a human or response to a mass application by an AI)) and it is considered important going forward to undertake consideration based on them.

some problems from spreading to other AI, and system resilience may be necessary in a society where AI is used going forward.

- The more practical the interface between humans and AI systems becomes, the more important that interface becomes, so I think it's time to add something to the guidelines.
- It is important to consider the interface between humans and AI in light of progress in the development and utilization of AI. In addition, it is important to consider what should be considered in terms of the inability to clearly separate the development and utilization of AI.
- Products with embedded AI systems may be difficult for end users to use due to, for example, complex setup. Consideration needs to be undertaken from the perspective of ensuring that end users can properly use AI and not use it incorrectly while using it.¹²⁸
- AI is increasingly being used in practical applications. Business models are not mentioned in existing guidelines, but there may be some situations where AI can be used in business modeling.
- It is particularly important to indicate our stance on the EU's Artificial Intelligence Act when revising the guidelines. Although there is a question as to whether this should be incorporated into the main body of the guidelines, it is important for this Conference to indicate how it views overseas trends and how it interprets them.

Going forward, consideration for revising the two guidelines will proceed based on such issues, the opinions and recommendations of this Conference, and the following two perspectives¹²⁹.

- Whether or not to add to or change the purposes, basic philosophies, or the principles on the basis of the current approach of the two guidelines.
- Whether or not 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.

In considering the review, it is important to ensure the guidelines continue to be technologically neutral, while also not impeding innovation, and not placing excessive loads on developers and users. In addition, when considering the guidelines themselves, as mentioned above, it is necessary to keep in mind during consideration that the two guidelines embody the "Social Principles of Human-Centric AI"¹³⁰ formulated by the government as a whole¹³¹.

In addition, the purpose of the two guidelines is to promote the benefits of AI and reduce risks through the sound development of AI networking. Furthermore, since "To ensure an

¹²⁸ The principle of proper utilization in the AI Utilization Guidelines states that AI service providers are expected to provide necessary information, and it is considered important to study how this point can be brushed up.

¹²⁹ The two mentioned perspectives are not necessarily contradictory, and it is possible that objectives, basic philosophies, principles, etc. stipulated in the current guidelines will be added or changed while reviewing the guidelines themselves.

¹³⁰ See Note 20 above.

¹³¹ In this regard, one opinion was that if the two guidelines are to be reviewed, it would be better to limit the contents to the characteristic items after clarifying the structure that there are guidelines to be formulated by this Conference (Ministry of Internal Affairs and Communications) based on the "Social Principles of Human-Centric AI" formulated by the government as a whole.

appropriate balance between the benefits and risks of AI networks, so as to: (a) promote the benefits from AI networks through innovative and open R&D activities and fair competition; and (b) mitigate the risk that AI systems might infringe rights or interests, while fully respecting the value of the democratic society such as academic freedom and freedom of expression." is stipulated as a basic philosophy in the AI R&D Guidelines and "To respect the diversity of people utilizing AI (users) and include people with diverse backgrounds, values, and ideas." is stipulated as a basic philosophy in the AI Utilization Guidelines, it is necessary to keep these things in mind when reviewing the guidelines.

Chapter 4. Future Initiatives

Chapter 1 provides an overview of domestic and international trends and trends in international discussions. Chapter 2¹³² introduces the initiatives of business operators and others based on the interviews, and summarizes the points to consider (key points of good practices) in promoting the "Safe, Secure, and Trustworthy Implementation of AI in Society." In addition, the previous chapter reviewed the AI R&D Guidelines and the AI Utilization Guidelines. Based on the results, it was decided that a review of the two guidelines would be conducted in the future, and issues related to the review were summarized.

This chapter summarizes future initiatives in light of previous discussions.

1. AI ethics and governance

(1) Consideration for reviewing the AI R&D Guidelines and AI Utilization Guidelines

In addition, the previous chapter reviewed the AI R&D Guidelines and the AI Utilization Guidelines. Based on the results, it was decided to conduct a review of the two guidelines in the future, and issues related to the review were summarized. Also, as summarized in the previous chapter, in addition to the initiatives presented in the interviews that can be evaluated as excellent initiatives for responding to the principles in the two guidelines, there are noteworthy initiatives that go beyond the scope of the two guidelines with respect to the "Safe, Secure, and Trustworthy Implementation of AI in Society." In light of these and other factors, a review of the two guidelines will be conducted focusing on the issues summarized in the previous chapter.

In considering the review, it is important to do so without being bound by existing frameworks, taking into account the opinion that the guidelines should be reviewed regarding their positioning, modalities, scope, etc.¹³³. Consideration will be given to the differences between the two guidelines discussed in the previous chapter and the content of guidelines and policies of other countries.

Furthermore, consideration needs to be given to ensuring the guidelines are technologically neutral, while also not impeding innovation, and not placing excessive loads on developers and users. Furthermore, in light of the basic philosophies in each guideline, consideration needs to be given to perspectives such as "securing a competitive ecosystem" and "protecting the interests of users."¹³⁴

In addition, a wide range of stakeholders, including developers, users, experts, and relevant organization organizations, will engage in exchanging views as the review progresses.

If it is decided to review the two guidelines, after clarifying the target audience¹³⁵ and legal nature¹³⁶ of the two current guidelines again, in consideration of disseminating information overseas and how the guidelines are viewed from overseas, it is considered important to clarify the differences between the revised guidelines and consider appropriate names.

¹³² Includes the supplementary volume (Compilation of Case Studies for "Safe, Secure, and Trustworthy Implementation of AI in Society" - Collection of Good Practices.)

¹³³ In this regard, AI Guidelines, Trust and Governance - Towards Optimal Rule Formulation - by Toshiya Jitsuzumi in Nextcom Vol. 50 (KDDI Research Institute, Inc., 2022) serves as a reference.

¹³⁴ See Section 2. (2) in the previous chapter.

¹³⁵ This includes not only private business operators, but also individuals, government organizations, and local governments.

¹³⁶ The basic philosophies of the two guidelines call for sharing guidelines and best practices internationally as non-binding soft laws.

(2) Disseminating and sharing the AI R&D Guidelines and the AI Utilization Guidelines

Report 2020 pointed out that while the establishment of AI principles, etc. in itself does not directly lead to corporate profits, the AI principles, etc. do set forth a basic corporate policy for AI R&D, etc., and this helps to dispel the concerns of corporate stakeholders regarding AI development, etc. and foster confidence in initiatives for AI development, etc.¹³⁷. However, the previous chapter covers the status of the formulation of guidelines and policies in Japan, but only some business operators and organizations, etc. have formulated and published them, so it is important to work to spread initiatives on formulating guidelines and policies. This Conference and the Ministry of Internal Affairs and Communications will continue to promote activities to disseminate and share the AI R&D Guidelines and the AI Utilization Guidelines¹³⁸.

(3) Monitor trends in Japan and overseas and directions of international discussions, and provide information

As described in Chapter 1, various initiatives are being made in Japan and overseas on AI ethics and governance, as well as active international discussions. In considering AI ethics and governance, it is important to take into account domestic and international trends and trends in international discussions, and this Conference will continue to follow up on developments in these areas.

It is also important to not only follow up on developments, but in international discussions, it is also important for Japan to disseminate information as well as lead and contribute to international discussions. It is important to disseminate information on the case studies of initiatives by individual business operators and others compiled in this report, taking into account Japan's industrial structure, such as the large number of Business to Business (B2B) transactions, in international forums, including multilateral forums such as the OECD and GPAI, as well as in bilateral policy dialogues.

Japan is expected to play a significant role in international discussions, especially in light of the fact that the OECD is scheduled to hold a ministerial meeting on digital issues at the end of this year (2022), that Japan will chair the GPAI from autumn this year (2022) to autumn next year (2023), and that the G7 will be held in Japan next year (2023).

In addition, the EU's Artificial Intelligence Act states that providers and users in third countries who provide AI systems or products created by AI systems within the EU are also subject to the regulations on high-risk AI. Therefore, once the Act is adopted and applied, it is likely to have a major impact on the rest of the world, including Japan. It is important for this Conference to continue to conduct research on the Act taking into account trends in international discussions and its reception and reactions in other countries.

(4) Initiatives to ensure the effectiveness of principles

Following the adoption of the Council Recommendations including AI Principles by the OECD in May 2019 and the adoption of G20 AI Principles by the G20 in June of the same year, the focus on national efforts and international discussions on AI is shifting to how to

¹³⁷ See Section 9. in Chapter 3 of Report 2020.

¹³⁸ We will continue to examine revising the two guidelines going forward, but even the current guidelines are considered to be helpful to business operators and others in developing guidelines and policies. In addition, after the guidelines are reviewed, it is naturally our intention to disseminate and share the revised guidelines.

comply with these principles and promote the social implementation of trustworthy AI.

The interviews showed that many business operators and others are implementing initiatives to respond to the principles in the AI R&D Guidelines and the AI Utilization Guidelines¹³⁹. As mentioned in the previous chapter, from the perspective of ensuring the effectiveness of the principles in the guidelines, in addition to the initiatives of the business operators and others themselves, it is important to ensure the effectiveness of the principles in the guidelines as a whole by combining various initiatives such as common/cross-sectoral governance, sector-specific and by sector governance, and rule formation through standardization, so that trustworthiness toward AI will increase and AI will be increasingly accepted by society. In addition, initiatives may include industry-wide initiatives, collaboration with external organizations, and initiatives involving the government (government organizations), and it is desirable that these initiatives be implemented appropriately. In relation to this, it should be noted that in addition to initiatives by the business operators and others themselves, the interviews showed that initiatives are being implemented to ensure the effectiveness of the principles in the guidelines, such as by establishing frameworks for identifying risks and developing mechanisms for monitoring by external organizations¹⁴⁰. Also, initiatives¹⁴¹ such as formulating check sheets and certification systems can also be considered, and based on this it is important to consider initiatives for ensuring the effectiveness of the principles in conjunction with reviewing the guidelines¹⁴².

In conducting this review, it is considered that a comprehensive review should be conducted while bearing in mind the importance of establishing a mechanism and developing an environment for promoting and disseminating initiatives that ensure the effectiveness of these principles.

2. Disseminating and sharing case studies of initiatives

(1) AI ethics and governance initiatives

The interviews showed that many business operators and others are formulating principles, policies, guidelines, etc. related to AI ethics and governance and developing organizations and systems, as well as implementing initiatives to develop mechanisms for quality assurance and for collaborating and cooperating with external stakeholders. These initiatives can be evaluated as excellent initiatives that respond to the principles in the AI R&D Guidelines and the AI Utilization Guidelines, such as security, privacy, fairness, and transparency/accountability.

It should be noted that the interviews showed that initiatives are being implemented to ensure the effectiveness of the principles in the guidelines, such as by establishing frameworks for identifying risks and developing mechanisms for monitoring by external organizations, or alternatively that initiatives are being implemented to comply with the principles not only at the companies themselves but also including their business partners, such as identifying risks in the supply chain and taking measures to prevent the risks from materializing. These initiatives can be evaluated as initiatives that further promote the "Safe, Secure, and Trustworthy Implementation of AI in Society."

¹³⁹ Report 2021 also covers various initiatives implemented by many business operators to respond to the principles (see Section 2. (1) and Section 4. in Chapter 3 of Report 2021 and Attachment 3).

¹⁴⁰ See Section 2. (1) in Chapter 2 and the supplementary volume (Collection of Good Practices.)

¹⁴¹ See Note 124 above. Such initiatives by industry and others are considered to be helpful.

¹⁴² These points are also addressed in the previous chapter's discussion on reviewing the two guidelines.

It is considered that AI developers and users, etc. who are engaged in such initiatives can be evaluated as fulfilling their social responsibilities, and it is expected that they will be highly praised by society. Furthermore, it is hoped that the spread and penetration of such initiatives throughout society will improve society's acceptance of AI¹⁴³.

It is necessary to keep in mind that, depending on the types of AI and services developed and utilized and the scale and industry, etc. of the business operator, the initiatives discussed in the interviews may not necessarily be useful or beneficial for all business operators, but the initiatives that many business operators and others are implementing in common are considered to be standard content that will serve as a reference for AI developers and users, etc. who intend to implement initiatives going forward. In addition, the establishment of frameworks for identifying risks, the development of mechanisms for monitoring by external organizations, and initiatives related to the supply chain, etc. are considered to be useful and beneficial for developers and users, etc. who have already implemented certain initiatives. Furthermore, it is expected that informing the general public and users that AI developers and users, etc. are implementing AI ethics and governance related initiatives will enhance the acceptance of AI in society.

In light of these points, it is important to disseminate and share the case studies of initiatives implemented by the various business operators and others compiled in this report, and this Conference and the Ministry of Internal Affairs and Communications will promote activities for disseminating and sharing the case studies of initiatives in cooperation with external stakeholders¹⁴⁴.

In promoting activities for disseminating and sharing case studies, regarding AI ethics and governance, it is important to increase trust in AI and improve social acceptance by promoting multi-stakeholder initiatives, and in particular, we will work to expand to users, user groups, etc. and will continue to exchange opinions with a wide range of stakeholders, including users, user groups and experts from the perspective that it is important to incorporate the opinions of users and experts, etc.¹⁴⁵.

In addition, AI Strategy 2022 states, "As people's lifestyles and ways of working are changing, it is necessary to consider more active use of AI than ever before in order to respond to these changes and to further improve the efficiency of work of national government organizations and provide high-quality administrative services. When national government organizations use AI, they need to understand that it is particularly important to ensure transparency, fairness and explainability, and to promote the introduction of AI. To this end, we

¹⁴³ The opinion was expressed that as a measure for promoting the dissemination and penetration of such initiatives throughout society, preferential tax incentives and preferential treatment in procuring government contracts, etc. could be offered to business operators that are sincerely committed to AI ethics and governance. However, it should be noted that these kind of measures require careful consideration and coordination.

¹⁴⁴ For example, the Ministry of Internal Affairs and Communications, in collaboration with the ELSI Center of Chuo University, disseminated and shared the initiatives of this Conference at the community activity that the center conducted in June 2022.

¹⁴⁵ For example, the "Ensuring explainability" section of the "Principle of transparency" in the AI Utilization Guidelines states, "AI service providers and business users are expected to ensure the explainability of the decisions of AI for the purpose of ensuring the trust of users and to present evidence of AI behavior with consideration of the social context." In addition, the Detailed Explanation on Key Points Concerning AI Utilization Principles document annexed to these guidelines states, "With consideration of the needs and opinions of consumer-type users, clarify parts in which an explanation is lacking, and collaborate with developers to find out what kind of explanation is necessary. "

will compile and implement comprehensive measures such as organizing the basic concept of AI introduction in national government organizations and establishing AI introduction guidelines. In addition, AI will be actively used in the work of government organizations." Therefore, going forward it is assumed that the utilization of AI in government organizations will progress, and we will exchange opinions with the relevant ministries and agencies.

(2) AI development and utilization initiatives

The interviews showed that many business operators and others are engaged in initiatives to develop and utilize AI in various fields, including COVID-19 measures, medical and healthcare, and support for the elderly and people with disabilities. This can be evaluated as showing the usefulness of utilizing AI in a wide variety of fields. In addition, it is expected that new technologies will be developed, new services will be provided, and the convenience of existing services will be improved with reference to these case studies.

In light of these points, it is important to disseminate and share the case studies of initiatives implemented by the various business operators and others compiled in this report, and this Conference and the Ministry of Internal Affairs and Communications will promote activities for disseminating and sharing the case studies of initiatives in cooperation with external stakeholders.

As a developed country facing challenges, it is also important for Japan to demonstrate to the world that the development and utilization of AI will contribute to solving social issues. It is important to disseminate the information contained in the case studies of the initiatives implemented by the various business operators and others summarized in this report in international discussions.

3. Human resources development

The interviews showed that many business operators and others are implementing initiatives utilizing various means to develop human resources.

However, with regard to AI human resources, the shortage of human resources is a common issue worldwide, and human resources development is an urgent issue in Japan as well. It is important for AI developers and users, etc. to continue to promote initiatives for developing human resources while referring to the case studies of initiatives implemented by business operators and others compiled in this report. With regard to AI human resources, in addition to the expectation to develop human resources with excellent technical skills, initiatives to improve AI literacy are also being implemented¹⁴⁶, and in this regard, there is an expectation that human resources who can contribute to identifying and responding to ethical, legal, and social issues (ELSI) related to the development and utilization of AI will also be developed¹⁴⁷. In addition to initiatives for improving AI literacy so that end users can make full use of products and services incorporating AI systems, the development of human resources that support end users and initiatives that improve literacy are

¹⁴⁶ The case studies compiled in this report also provide examples of technology education for elementary school students, provision of educational programs for high schools, and collaboration between companies and universities in human resource development that are considered to be helpful. For details, see Section 2. (3) in Chapter 2 and the supplementary volume (Collection of Good Practices.)

¹⁴⁷ AI Strategy 2022 states, "efforts to realize "Responsible AI" are also expected through initiatives related to the ELSI of AI, such as designing AI with ethical considerations in the first place and conducting audits in the AI utilization cycle."

It has also been pointed out that it is important to identify areas in which humanities and social sciences can play an active role, and it is considered important to develop such human resources.

also important.

With regard to the development of human resources, collaboration with external educational and research institutes is expected to deepen the initiatives of AI developers and users, etc. themselves and contribute to raising the level of society as a whole. It is expected that AI developers and users, etc. will promote their initiatives in cooperation with external educational and research institutes rather than by themselves¹⁴⁸.

Furthermore, 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.

¹⁴⁸ Regarding ELSI for example, Osaka University established the Research Center on Ethical, Legal and Social Issues (ELSI Center) in April 2020, and Chuo University established the ELSI Center in April 2021, and initiatives for promoting industry-academia collaboration are being implemented.

Conclusion

This report provides an overview of trends in Japan and overseas and trends in international discussions, etc., with a focus on AI ethics and governance, introduces examples of initiatives for promoting the social implementation of AI collected through interviews with business operators, experts, and relevant organizations, reviews the AI R&D Guidelines and the AI Utilization Guidelines based on the fact that many principles, policies, and guidelines, etc. concerning AI ethics and governance have been formulated in Japan and overseas, arranges issues related to reviewing the two guidelines, and arranges future initiatives.

As described in Chapter 1, various AI-related initiatives are being implemented in Japan and overseas, international discussions are also being advanced at the OECD, GPAI, etc., and it is important to contribute to international discussions while continuing to follow trends in Japan and overseas. In particular, the OECD is scheduled to hold a ministerial-level meeting on digital at the end of this year (2022), and Japan will chair the GPAI from autumn this year (2022) to autumn next year (2023). Furthermore, a meeting of the G7 is scheduled to be held in Japan next year (2023). It is important that Japan actively disseminates information and contributes to international discussions.

Chapter 2 compiles many specific case studies of initiatives related to the social implementation of AI based on interviews with business operators, experts, and relevant organizations that are implementing advanced or ambitious initiatives to implement AI in society¹⁴⁹. Following on from Report 2021, we believe that no other document contains such a large number of specific case studies of initiatives, and that the information provided is very useful and beneficial. We would like to express our gratitude once again to the business operators, experts, and relevant organizations for their cooperation, and it is hoped that this report will be widely shared with those engaged in or considering developing and utilizing AI, and that the implementation of AI in society will be further promoted.

In Chapter 3, the AI R&D Guidelines and the AI Utilization Guidelines are reviewed, and issues concerning reviewing the two guidelines are summarized, and it is important to examine making the two guidelines more useful and beneficial to AI developers and users, etc. Although there are a wide range of issues, it is important to examine these issues without being bound by existing framework through careful discussions with multiple stakeholders.

In the AI R&D Guidelines and the AI Utilization Guidelines compiled by this Conference, "To achieve a human-centered society where all human beings across all of society enjoy the benefits while in harmony with AI networks, and human dignity and individual autonomy are respected" is stated as a basic philosophy.

Realizing a human-centered society is a common challenge for all humankind. In order to realize an inclusive and sustainable human-centered society, it is important to promote initiatives that enable everyone to safely and securely utilize AI under appropriate conditions. It is also important to form a global network of wisdom and co-create wisdom in order to solve the problems common to all humankind.

Based on the above, this Conference will continue to undertake initiatives to further promote the "Safe, Secure, and Trustworthy Implementation of AI in Society" by multi-stakeholders, based on the initiatives described in Chapter 4 to be implemented going forward.

¹⁴⁹ Also see the Compilation of Case Studies for "Safe, Secure, and Trustworthy Implementation of AI in Society" - Collection of Good Practices supplementary volume.

With the participation of experts from Japan and abroad, this forum was held to contribute to the resolution of various social issues related to AI through discussion on the use and distribution of data, which is the key to the dissemination and use of AI, as well as regulatory issues and governance related to AI, taking into account international trends.

○ Date and time: Tuesday, March 1, 2022, 13:00 to 18:00 (online)

13:00-13:05	Opening Remarks (5)	KANEKO Yasushi, Minister for Internal Affairs and Communications
13:05-13:25	Keynote Lecture (20)	"How to Value Data in a World with AI" Laura VELDKAMP, Professor, Columbia University
13:25-14:05	Keynote Dialogue (40)	The Industrial Revolution by AI and Data Laura VELDKAMP, Professor, Columbia University IWATA Kazumasa, President, Japan Center for Economic Research
14:05-15:25	Panel Discussion (80)	"Tomorrow's World Spurred by Data and AI" « Moderator » OHASHI Hiroshi, Professor of Economics and Dean, Graduate School of Public Policy, The University of Tokyo « Panelists » TAKI Toshio, Executive Director, Head of Sustainability and CoPA, Head of the Money Forward Fintech Institute, Money Forward, Inc. TAMAKI Emi, Professor, Computer Science and Intelligent Systems Program, Faculty of Engineering, University of the Ryukyus; CEO, H2L, Inc. MUTO Shinsuke, President, Tetsuyu Institute Medical Corporation; Chairman, Integrity Healthcare
15:25-15:55	Special Lecture (30)	"AI, Globalisation, and the Future of Work" Richard BALDWIN, Professor, International Economics, The Graduate Institute, Geneva Moderator: SEKIGUCHI Waichi, President, MM Research Institute; Former Editorial Writer, Nikkei, Inc.
15:55-16:00		- Break -

※ Affiliations and titles of participants as of the time of the event.

Global Forum on AI Network Society 2022

16:00-16:15	Honor Lecture (15)	SUDO Osamu, Professor, Faculty of Global Informatics, Chuo University; Director, ELSI Center, Chuo University; Project Professor, Graduate School of Interdisciplinary Information Studies, The University of Tokyo
16:15-16:35	Special Lecture (20)	<p>"AI Ethics : Translating Principles into Governance and Regulation"</p> <p>Francesca ROSSI, IBM Fellow and the IBM AI Ethics Global Leader, IBM Corporation</p>
16:35-17:55	Panel Discussion (80)	<p>"Governance for the World Living in Harmony with AI"</p> <p><< Moderator >> SUDO Osamu, Professor, Faculty of Global Informatics, Chuo University, Director, ELSI Center, Chuo University; Project Professor, Graduate School of Interdisciplinary Information Studies, The University of Tokyo</p> <p><< Panelists >> EMA Arisa, Associate Professor, Institute for Future Initiatives, The University of Tokyo HIRANO Susumu, Professor and Dean, Faculty of Global Informatics, Chuo University FUKUDA Takeshi, Director of IBM Research – Tokyo, IBM Japan MOCHIZUKI Yasunori, NEC Fellow, NEC Corporation</p>
17:55-18:00	Closing Remarks (5)	NAKANISHI Yusuke, State Minister for Internal Affairs and Communications

※ Affiliations and titles of participants as of the time of the event

Sessions Overview ①

Keynote Lecture

Laura VELDKAMP

(Professor, Columbia University)



Honor Lecture

SUDO Osamu

(Professor, Faculty of Global Informatics, Chuo University; Director, ELSI Center, Chuo University; Project Professor, Graduate School of Interdisciplinary Information Studies, The University of Tokyo)



"How to Value Data in a World with AI"

- More value is created by leveraging AI to produce knowledge from data. As the volume of data grows exponentially, the value of a company is determined by the amount of data it possesses. The reason behind for the growth of large U.S. corporations is not their tangible assets, but the vast amount of intangible assets they possess: data. Maximizing data value through AI is a key to maximize corporate value.
- Data utilization will have different outcomes depending on data-intensity and labor-intensity. It is essential to employ an optimal ratio of workers who manage data and workers with older technological skills, and then to increase employment of workers who use AI to generate knowledge for the future. In practice, the growth rate of AI-skilled workers has outpaced that of old-tech workers and data managers over the past few years. Wages are also higher for AI-skilled workers, old-tech workers, and data managers, in that order.
- AI-skilled workers currently make up a small percentage of the workforce, and the contribution to revenue from data utilization is higher for old-tech workers. Even during the Industrial Revolution, only 5-13% of the workforce used new technologies; there is no doubt that data utilization through AI is a source of value creation, and the labor share in firms is to change.
- In the midst of disruptive technological trends since the start of private-sector use of the Internet, AI is a critical element and its principles and governance initiatives have received significant attention.
- For the appropriate development and implementation of AI, in March 2019, the Japanese government compiled "Human-Centric AI Social Principles" (① Human-Centric Principles, ② Principles for Education and Literacy, ③ Privacy Principles, ④ Principles for Ensuring Security, ⑤ Principles for Securing Fair Competition, ⑥ Principles for Fairness, Accountability and Transparency, and ⑦ Principles for Innovation). These principles are the core of building a new society through the use of AI.
- The Conference toward AI Network Society of the Ministry of Internal Affairs and Communications, which I chair, has been discussing what the ideal AI society is, how AI should be utilized and how data should be distributed.
- The G7 meeting is scheduled to be held in Japan in 2023, and preparations are underway to actively make proposals on how AI should be used. In the midst of various developments, we are expected to cooperate with international communities to create a new society.

Sessions Overview ②

Special Lecture

Richard BALDWIN

(Professor, International Economics,
The Graduate Institute, Geneva)



"AI, Globalization, and the Future of Work"

<Moderator> SEKIGUCHI Waichi, President, MM Research Institute; Former Editorial Writer, Nikkei, Inc.

- In my 2019 book, "GLOBOTICS," I stated that digital technology will simultaneously automate and globalize white-collar and professional jobs. As computers acquire new cognitive skills, many jobs that once required human intervention, such as translation, editing, and graphics, can now be accomplished through machine learning.
- Digital technology grows slowly at first, and then extremely fast at some point, where disruption (disruptive innovation) occurs.
- In the future, jobs that can be automated by software robots will leave human hands. Jobs that can be done by cheap labor in emerging markets will be offshored. But collecting big data about ethics, creativity, curiosity, motivation, and other human and complex things will be difficult for remote intelligence and AI.
- Jobs that can be automated, including bookkeeping, accounting, and certain legal work, will be replaced relatively quickly by robots, while competitive workers will have more opportunities.
- Governments are required to have proactive labor market policies, such as retraining and relocation supports to help workers readjust.

Special Lecture

Francesca ROSSI

(IBM Fellow and the IBM AI Ethics
Global Leader, IBM Corporation)



"AI Ethics : Translating Principles into Governance and Regulation"

- AI ethics has passed through the phase of awareness and the phase of principles, and is now in the operational phase called "AI Ethics 3.0".
- The draft regulation called "EU AI ACT" published in 2021 is one form of AI ethics, and is characterized by the fact that it regulates applications of AI rather than AI itself. The draft regulation defines four risk levels, with obligations (for providers and users of AI systems) and restrictions on applications according to the levels. It also defines seven requirements for "trusted AI. Companies must build AI solutions in accordance with these regulations.
- IBM has established principles that AI is an extension of human intelligence which requires transparency and accountability, and under these principles it sets out what trustworthy AI should look like. IBM's principles are similar to the seven requirements for "trustworthy AI" in the proposed EU regulation, and the company has put the principles into operation with effective internal governance.
- Regulation is only one of many reasons to adhere to AI ethics. While compliance with regulations is important, AI ethics is about more than mere regulatory compliance.

Sessions Overview ③

Keynote Dialogue

Laura VELDKAMP

(Professor, Columbia University)



IWATA Kazumasa

(President, Japan Center for Economic Research)



“The Industry Revolution by AI and Data”

- AI technologies include "prediction algorithms" and "invention algorithms". Which do you think has a more direct impact on productivity and leads to a higher growth rate? Also, do you see any difference compared to the industrial revolution? (Iwata).
- In the same way that technology changed the way goods were produced in the Industrial Revolution, AI technology will change the way knowledge is generated. Capital can be divided into those used for capital investment and those used for research and development, but we believe that it is the "invention algorithms," which are research and development, that generate higher growth. (Veldkamp)
- How do you see the AI revolution impacting the U.S. economy? (Iwata)
- While there will certainly be a significant impact, it will take decades for AI technology to become a general-purpose technology and to be adopted in a variety of industries. Also the technology alone is not enough to keep growing. (Veldkamp)
- In a study conducted by the Japan Center for Economic Research, it was found that if AI and IoT were fully utilized, the projected growth rate of the Japanese economy would change from less than 1% to 5%. In the long run, technological advances will converge, but this impact will be significant. (Iwata)
- Companies know that AI technology is far more productive and profitable than older technologies. That is why they are promoting the use of AI and are working to develop human resources. (Veldkamp)
- How do you think the free distribution of data across different industries, sectors, or countries will affect the productivity of companies? (Iwata)
- Data utilization has made our lives more convenient, efficient, and productive. On the other hand, increased data sharing may transform corporate behavior. Care must be taken in the free distribution of data, including privacy issues. (Veldkamp)
- Japan lags behind in the utilization of AI due to the lack of a data sharing mechanism. In China, more than 80% of companies are already utilizing AI, but in Japan, only less than 40% are utilizing AI. (Iwata)
- The utilization of AI has just begun, and it will be necessary to closely monitor changes in the labor participation rate and intangible assets resulting from the utilization of AI. (Veldkamp)
- It has been pointed out that the flow of data is the flow of money. If a world in which data and currency are combined is created, what do you think the outcome will be? (Iwata.)
- While money is a form of payment for services and products, data is knowledge. Knowledge must be shared and we can grow by sharing knowledge. It is important to share knowledge and data while striking a balance between privacy protection and respect for the owner of the information. (Veldkamp)

Sessions Overview ④

Panel Discussion

"Tomorrow's World Spurred by Data and AI"

【 Moderator 】 OHASHI Hiroshi

(Professor of Economics and Dean, Graduate School of Public Policy, The University of Tokyo)



【 Panelists 】 TAKI Toshio

(Executive Director, Head of Sustainability and CoPA, Head of the Money Forward Fintech Institute, Money Forward, Inc)



TAMAKI Emi

(Professor, Computer Science and Intelligent Systems Program, Faculty of Engineering, University of the Ryukyus; CEO, H2L, Inc.)



MUTO Shinsuke

(President, Tetsuyu Institute Medical Corporation; Chairman, Integrity Healthcare)



- As data adds economic value at an accelerating pace, it is necessary to conduct deeper, sector-specific analysis in areas such as healthcare and finance. (Ohashi)
- The Japanese financial industry is increasingly using application programming interfaces (APIs) to link programs together, and clear consent for information transfer is required at the time of linkage. However, the concept of rights pertaining to access and ownership of data and related legislation are insufficient. (Taki)
- In order to realize personalized medicine with quality data, it is necessary to build a platform that people can perceive as valuable, but it is important to reduce the burden on the healthcare providers. Good AI will be meaningless if it does not help the healthcare providers. (Muto).
- I have been challenging to digitize the "sensation of a person acting on an object to obtain a sense of presence" and its data output. If this is realized, it will create a lot of added value, such as working in a virtual space (metaverse) while in a room, or gaining shared experiences in sports and sightseeing. On the other hand, there are many issues to be considered, and the researchers are already discussing the rule-making for practical use. (Tamaki)
- The key to financial digitalization is not to cause incidents such as information leaks. Security standards are necessary for fintech businesses, even if they are ventures. (Taki)
- Online medical care is growing, but a new digital divide is emerging. It is important to reduce the load that has been endured unknowingly. (Muto.)
- It is necessary to investigate how to integrate data and return results to users, as well as the movements and transformations of users. It is important to control self-information. (Tamaki)
- What kind of environment needs to be developed in Japan to realize the DFTT (trusted and free data distribution)? (Ohashi)
- The transfer of personal information is protected by law and easy to discuss, but comparatively statistical and processed information is difficult to discuss. Companies individually need to take actions such as issuing data statements. (Taki)
- The goal of democratization of healthcare is to have the right to self-determination and to create a well-governed framework. In preventive medicine, it is necessary to share and utilize data from multiple fields. (Muto)
- The three challenges to improve the environment are data ownership, user experience of consent, and data security and sharing across disciplines. (Tamaki)

Sessions Overview ⑤

Panel Discussion

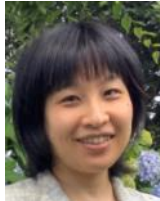
【 Moderator 】 SUDO Osamu

(Professor, Faculty of Global Informatics,
Chuo University, Director, ELSI Center,
Chuo University
Project Professor, Graduate School of
Interdisciplinary Information Studies,
The University of Tokyo)



【 Panelists 】 EMA Arisa

(Associate Professor, Institute
for Future Initiatives, The
University of Tokyo)



HIRANO Susumu

(Professor and Dean, Faculty
of Global Informatics, Chuo
University)



FUKUDA Takeshi

(Director of IBM Research –
Tokyo, IBM Japan)



MOCHIZUKI Yasunori
(NEC Fellow, NEC Corporation)



"Governance for the World Living in Harmony with AI"

- AI governance is moving toward the stage of defining and implementing principles of fairness, transparency, and trustworthiness. Efforts to solve issues related to AI governance are not in the "competitive domain" but in the "collaborative domain," and it is important to involve various stakeholders in discussions and to build an ecosystem. (Ema)
- The purpose of AI utilization is to augment human intelligence, and not to replace it. New technologies must also be transparent and explainable; IBM has clarified the principles for trust and transparency and has stated that precision regulation of AI will be determined on a use-case basis, rather than regulating the technology itself. (Fukuda)
- In order to create social value through AI, it is important to gain society's trust and acceptance as well as functionality and performance; it is essential to ensure that AI governance does not become an impediment to innovation creation or a barrier to entry. (Mochizuki)
- Although there are discussions on "hard law" and "soft law" for the regulation of AI, it is important to share wisdom on what are reasonable regulations for the diverse and specific risks that will emerge in the future, rather than a dichotomous discussion. (Hirano)
- Regarding the EU regulatory bill, while I understand the idea that it is better to be on the safe side, but there is the risk of delayed innovation. The contents of the regulations need to be scrutinized. (Hirano).
- AI is a developing technology, and it is important to assume that technology, risk, and social acceptance will change over time. (Fukuda)
- While the high-level thinking of the EU is important, from the view of the private sector, I would like to stress on the significance of a use case-based thinking and to make it realistic and compliant while cooperating. (Mochizuki)
- There are differences among countries and regions in terms of attitudes towards privacy, human dignity, and work style. We believe that the functions of academia, such as information sharing and dissemination, will be useful. (Ema)
- The regulations for AI should be changed according to the situation as it is moved around, and there may be a need to adapt to the most stringent regulations. It is important to ensure interoperability. (Fukuda)
- It is important to proceed with AI while gaining an understanding of the benefits and other aspects of its use. (Mochizuki)
- It is important to seek a cooperative path with the EU. Human resource development and discussion are essential. (Ema)
- In a human-centered society, it is important to discuss "ELSI" (ethical, legal, and social issues) so that AI is under control. (Hirano)
- It should be noted that there are differences in legislation, industrial structures, administrative systems, etc. in each country, International cooperation should be promoted after clearly indicating this diversity to the EU and other countries. (Sudo)

Tentative
Translation

Attachment 2

Details of the Reviews of the AI R&D Guidelines and the AI Utilization Guidelines

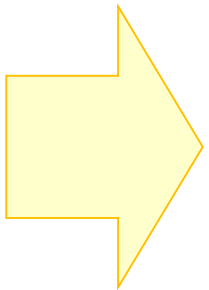
(1) Outline of the Reviews

In the area of AI ethics and governance, this Conference established the *AI R&D Guidelines*^{*1} (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*,^{*2} so the reviews took this matter into account as well.

^{*1} Formally known as the Draft AI R&D Guidelines for International Discussion.

^{*2} *Proposal for a Regulation laying down harmonised rules on Artificial Intelligence* (April 2021)

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

○ The scope and content of the current AI R&D Guidelines and AI Utilization Guidelines were examined based on principles, policies, guidelines, etc. established overseas ("guidelines and policies" hereafter).

■ Surveyed countries and surveyed guidelines and policies

- The survey covered 16 countries^{*3} (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^{*4} announced by government agencies, industry associations, and major global companies (refer to pages 16-23)
 - Government agencies^{*5} and international organizations (40), industry associations (10), global companies (17)

^{*3} 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.

^{*4} The survey generally concentrated on guidelines announced after August 2019, when the *AI Utilization Guidelines* were announced.

^{*5} 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^{*6}

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

^{*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.

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.

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

	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	Transparency, Explainability, Fairness, Equity, removal of Discrimination	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							
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					
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

Note 1: Statistics cover government agencies (including the EU's European Commission and High-level expert group on artificial intelligence) and international organizations.

Note 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.

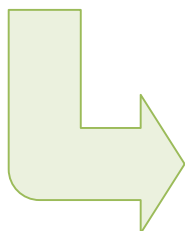
(2) Verification through comparisons with principles, policies, and guidelines instituted in other countries

<Comparison by region>

	No. of guidelines and policies	Human-centered	Human dignity	Diversity, Inclusiveness	Sustainable society	International cooperation	Proper utilization	Education/Literacy	Human intervention, Controllability	Proper data (quality of learning data)	Collaboration among AI systems	Safety	Security	Privacy	Fairness, Equity, removal of Discrimination	Transparency, Explainability	Accountability	Robustness	Responsibility	Traceability	Monitoring, Auditing	Governance	Other
North America	7	1	2			1	1	1	1	1		3	3	1	4	6	4	1	2	2	3		1
Europe	23	1	14	7	7	1	8	2	7	7		8	6	12	18	20	14	8	8	6	8		1
Asia	7		6	4	2	3	2	2	3	1		5	3	4	7	6	5	3	1	2	3	3	
Oceania	1	1	1									1	1	1	1	1	1		1		1		
Total	38	3	23	11	9	5	11	5	11	9	0	17	13	18	30	33	24	12	12	10	15	3	2

Note 1: Statistics cover government agencies (including the EU's European Commission and High-level expert group on artificial intelligence).

Note 2: Values mentioned in guidelines, etc. are colored in pink. Figures indicate the number of guidelines and policies mentioning the particular value. Figures in red indicate that the value is mentioned in more than half of the country or organization's guidelines and policies.



North America	Europe	Asia
<ul style="list-style-type: none"> [1] Transparency, Explainability [2] Fairness, Equity, Removal of discrimination [2] Accountability [4] Safety [4] Security [4] Monitoring, Auditing 	<ul style="list-style-type: none"> [1] Transparency, Explainability [2] Fairness, Equity, Removal of discrimination [3] Human dignity [3] Accountability [5] Privacy 	<ul style="list-style-type: none"> [1] Fairness, Equity, Removal of discrimination [2] Human dignity [2] Transparency, Explainability [4] Safety [4] Accountability

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

Values that should be respected	Content considered relevant in guidelines (main)	Examples of main content
1. Human-centered	<p>[AI R&D Guidelines]</p> <ul style="list-style-type: none"> ○ Basic philosophy <ul style="list-style-type: none"> ▪ Realize a human-centered society in which human dignity and individual autonomy are respected <p>[AI Utilization Guidelines]</p> <ul style="list-style-type: none"> ○ Basic philosophy <ul style="list-style-type: none"> ▪ Realize a human-centered society in which human dignity and individual autonomy are respected 	<ul style="list-style-type: none"> ▪ Remain mindful of the development and use of human-centered AI to strengthen national security and trusted partnerships. [No. 3] ▪ Systems must focus on those who use it or are affected by its decisions. [No. 41] ▪ AI systems must emphasize human-centered values and respect human rights, diversity, and individual autonomy. [No. 62]
2. Human dignity	<p>[AI R&D Guidelines]</p> <ul style="list-style-type: none"> ○ Principles of ethics <ul style="list-style-type: none"> ▪ Respect human dignity and individual autonomy <p>[AI Utilization Guidelines]</p> <ul style="list-style-type: none"> ○ Principles of dignity and autonomy <ul style="list-style-type: none"> ▪ Respect human dignity and individual autonomy 	<ul style="list-style-type: none"> ▪ AI systems must be designed to be consistent with international treaties protecting human dignity, rights, and freedoms. The organization takes responsibility for implementing measures to avoid reasonably predictable misuse through the use of ethics by design. [No. 7] ▪ It is necessary to respect the dignity and autonomy of individuals equally, regardless of nationality, gender, national or ethnic origin, ethnicity, religion, or language. [No. 8] ▪ Respect human rights in accordance with common human values, pursue human interests, and remain mindful of protecting national and regional ethics and morals. [No. 58]

Note: The numbers indicated in "Examples of main content" refer to principles, policies, guidelines, etc. (corresponding to the numbers in the list on pages 16-23).

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

Values that should be respected	Content considered relevant in guidelines (main)	Examples of main content
3. Diversity, Inclusiveness	<p>[AI R&D Guidelines]</p> <ul style="list-style-type: none"> ○ Basic philosophy <ul style="list-style-type: none"> ▪ Realize a human-centered society in which humans and AI networks co-exist, the benefits of AI networks are widely enjoyed by all, and human dignity and individual autonomy are respected ○ Principles of user support <ul style="list-style-type: none"> ➢ Strive to engage in initiatives to promote use by the socially weak, such as implementing universal design ○ Roles expected of related stakeholders <ul style="list-style-type: none"> ▪ In implementing and reviewing these guidelines, it is expected that organizations will strive to develop an environment that promotes dialog between diverse stakeholders, including national governments, international organizations, developers, and users (including civil society) ▪ Organizations are expected to strive to share awareness and engage in mutual cooperation in promoting the benefits and reducing the risks of AI, while ensuring the diversity of discussions on AI <p>[AI Utilization Guidelines]</p> <ul style="list-style-type: none"> ○ Basic philosophy <ul style="list-style-type: none"> ▪ Realize a human-centered society in which humans and AI networks co-exist, the benefits of AI networks are widely enjoyed by all, and human dignity and individual autonomy are respected ▪ Respect the diversity of users and include people with diverse backgrounds, values, and thinking when making use of and applying AI 	<ul style="list-style-type: none"> ▪ Data processing must not result in prejudice or discrimination that marginalizes or stigmatizes a particular group. Targeted efforts must be made to ensure that disadvantaged and vulnerable people benefit from technology development. Diversity must be ensured in the development and use of new technologies, by involving the relevant expert groups, user groups, and organizations. [No. 38] ▪ The diversity and representation of users must be reflected in the early stages of AI development and use, bias and discrimination based on individual characteristics (such as gender, age, disability, region, ethnicity, religion, or nation) must be minimized, and commercialized AI must be applied fairly to all. Efforts must be made to ensure that the socially weak and vulnerable have access to AI technologies and services, and that the benefits of AI are shared with all equally and not only a particular group. [No. 53] ▪ Ensure the diversity of AI system developers, strive to make use of comprehensive learning data, and remain mindful that algorithms are continuously tested and verified without discriminating against users based on ethnicity, gender, nationality, age, or religious beliefs. [No. 60]
4. Sustainable society	<p>[AI R&D Guidelines] (No content)</p> <p>[AI Utilization Guidelines]</p> <ul style="list-style-type: none"> ○ Basic philosophy <ul style="list-style-type: none"> ▪ Realize a sustainable society by solving various problems faced by individuals, local communities, countries, and the international community 	<ul style="list-style-type: none"> ▪ AI must be developed and used to ensure the sustainability of the environment. [No. 19] ▪ AI technology must be in line with the responsibility of humans to create the environment for life on Earth, sustain human prosperity, and maintain the environment for future generations. [No. 63] ▪ All actors involved in the AI system lifecycle must comply with international laws, national laws, and standards and practices (preventive measures, etc.) designed for the protection, restoration, and sustainable development of the environment and ecosystems. [No. 66]

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

Values that should be respected	Content considered relevant in guidelines (main)	Examples of main content
5. International cooperation	<p>[AI R&D Guidelines]</p> <ul style="list-style-type: none"> ○ Basic philosophy <ul style="list-style-type: none"> ▪ Share policies and best practices that will serve as non-binding soft law internationally between stakeholders, with regard to how AI system research and development should be conducted ○ Roles expected of related stakeholders <ul style="list-style-type: none"> ▪ In implementing and reviewing these guidelines, it is expected that national governments and international organizations will strive to develop an environment that promotes dialogue between diverse stakeholders, including national governments, international organizations, developers, and users (including civil society) <p>[AI Utilization Guidelines]</p> <ul style="list-style-type: none"> ○ Basic philosophy <ul style="list-style-type: none"> ▪ Share policies and best practices that will serve as non-binding soft law internationally, with regard to how AI should be used and applied 	<ul style="list-style-type: none"> ▪ Assuming that the principles and practices of AI governance in each country will be fully respected, it is necessary to promote the development of an international AI governance framework and standards based on broad consensus. [No. 57] ▪ It is necessary to work together globally to provide equal access to "autonomous" technologies and achieve a fair distribution of benefits and equal opportunities across and within societies. [No. 63] ▪ The participation of various stakeholders throughout the AI system lifecycle is required, in order to provide comprehensive AI governance, share the benefits of AI, develop fair technologies, and contribute toward development goals. Stakeholders include governments, intergovernmental organizations, the technical community, civil society, researchers, academia, the media, education, policy makers, private companies, human rights institutions, institutions working to promote equality, institutions monitoring discrimination, and youth and children's organizations. [No. 66]
6. Proper utilization	<p>[AI R&D Guidelines]</p> <ul style="list-style-type: none"> ○ Principles of controllability <ul style="list-style-type: none"> ➢ To ensure controllability, it is desirable to remain mindful of the effectiveness of human or other reliable AI supervision (monitoring, alerts, etc.) and response (AI system outages, network disconnections, repairs, etc.), to the extent possible in light of the characteristics of the technology employed <p>[AI Utilization Guidelines]</p> <ul style="list-style-type: none"> ○ Principles of appropriate use <ul style="list-style-type: none"> ▪ Strive to use AI systems and AI services within an appropriate scope and in an appropriate manner, based on an appropriate division of roles between humans and AI systems and between users 	<ul style="list-style-type: none"> ▪ It is necessary to confirm that situations where AI is used are consistent with the use cases in which that AI was trained. [No. 3] ▪ Individuals must themselves have control over what data is collected, what it is used for, and in what circumstances it is used. [No. 38] ▪ Informed consent must be obtained from stakeholders with regard to the impact of AI systems on their rights and interests. A mechanism must be established to allow data and services to be destroyed so that rights and interests are not violated if an unexpected event should occur. [No. 59] ▪ Humans must choose whether, when, and how to entrust decisions and actions to autonomous systems, and must use them appropriately. [No. 63]

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

Values that should be respected	Content considered relevant in guidelines (main)	Examples of main content
7. Education/Literacy	<p>[AI R&D Guidelines] (No content)</p> <p>[AI Utilization Guidelines]</p> <ul style="list-style-type: none"> ○ Principles of appropriate use (reference) <ul style="list-style-type: none"> ➢ If it is deemed appropriate for consumer-type users to make final judgments on AI decisions, consumer-type users should acquire the necessary skills and knowledge to make appropriate judgments. ➢ When using/applying AI operated through actuators, etc., consumer-type users should be aware of the locus of responsibility in each state (before, during, and after transitioning) in cases where operation is expected to switch to human operation under certain conditions. These users should also receive explanations from the AI service provider on conditions and methods for switching, and should acquire necessary skills and knowledge. 	<ul style="list-style-type: none"> • All citizens must have the right to receive training so that they can thrive mentally, emotionally, and economically with AI. [No. 20] • In order to resolve public concerns about AI technology, raise safety awareness, and address existing and future workforce challenges, remain mindful of efforts with regard to AI training for the general public, moral and ethical training for the relevant practitioners, and re-training of digital work skills to individuals whose jobs have been replaced. [No. 60] • Training in digital skills and AI ethics is necessary so that all members of society can make informed decisions about the use of AI systems and be protected from undue influence. [No. 66]
8. Human intervention, Controllability	<p>[AI R&D Guidelines]</p> <ul style="list-style-type: none"> ○ Principles of controllability <ul style="list-style-type: none"> • Remain mindful of AI system controllability ➢ To ensure controllability, it is desirable to remain mindful of the effectiveness of human or other reliable AI supervision (monitoring, alerts, etc.) and response (AI system outages, network disconnections, repairs, etc.), to the extent possible in light of the characteristics of the technology employed <p>[AI Utilization Guidelines]</p> <ul style="list-style-type: none"> ○ Principles of appropriate use <ul style="list-style-type: none"> ➢ When necessary and possible, it is expected that human decision-making will be involving in decisions made by AI, with regard to whether and how to use those decisions. In such cases, it is expected that standard examples will be taken into account and considerations will be made depending on the area of use and its applications, with regard to whether the intervention of human decision-making is required ○ Principles of fairness <ul style="list-style-type: none"> ➢ In order to ensure the fairness of decisions made by AI, it is expected that the social context when using/applying AI will be taken into consideration along with the reasonable expectations of people, and that human decision-making will be involved with regard to deciding whether to use these decisions or how to use them 	<ul style="list-style-type: none"> • AI systems must be designed to perform their intended functions while having the ability to detect and avoid unintended harm or disruption. They should also be designed to be stopped automatically or manually through human operation if they exhibit unintended escalation or other behavior. [No. 4] • The safety of users and third parties must be prioritized, and AI systems must include safeguards to allow humans to control AI as it operates. [No. 7] • Ensure that humans have the right to make fully autonomous decisions, the right to choose whether to receive services provided by AI, the right to withdraw from dialogue with AI at any time, and the right to stop operating AI systems at any time, and remain mindful that AI is placed under human control. [No. 58]

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

Values that should be respected	Content considered relevant in guidelines (main)	Examples of main content
9. Proper data	<p>[AI R&D Guidelines]</p> <ul style="list-style-type: none"> ○ Principles of ethics <ul style="list-style-type: none"> ➢ It is best to strive to implement necessary measures so that unfair discrimination does not occur due to bias and other factors contained in the learning data of AI systems, to the extent possible in light of the characteristics of the technology employed <p>[AI Utilization Guidelines]</p> <ul style="list-style-type: none"> ○ Principles of appropriate learning <ul style="list-style-type: none"> ▪ Remain mindful of the quality of data used for AI system learning, etc. 	<ul style="list-style-type: none"> ▪ To promote the responsible use of data at all stages and ensure its integrity, industry is responsible for understanding the parameters and characteristics of data, recognizing the potential for harmful biases, and testing for bias before and during the deployment of AI systems. [No. 7] ▪ We recommend documenting the overall image of all datasets used and detailing each data set and its source. [No. 33] ▪ Machine learning system trainers are responsible for checking the existence of bias in the data used for AI system learning, and in the continuous learning performed while the system is operating. [No. 51]
10. Collaboration among AI systems	<p>[AI R&D Guidelines]</p> <ul style="list-style-type: none"> ○ Principles of collaboration <ul style="list-style-type: none"> ▪ Remain mindful of the interconnectivity and interoperability of AI systems <p>[AI Utilization Guidelines]</p> <ul style="list-style-type: none"> ○ Principles of collaboration <ul style="list-style-type: none"> ▪ Remain mindful of collaboration among AI systems and AI services 	<ul style="list-style-type: none"> ▪ With regard to interoperability, measures must be taken to enable intervention due to technological changes, the emergence of new stakeholders, and groups marginalized by society. [No. 66]
11. Safety	<p>[AI R&D Guidelines]</p> <ul style="list-style-type: none"> ○ Principles of safety <ul style="list-style-type: none"> ▪ Make arrangements so that AI systems do not harm the lives, bodies, and property of users and third parties through actuators, etc. <p>[AI Utilization Guidelines]</p> <ul style="list-style-type: none"> ○ Principles of safety <ul style="list-style-type: none"> ▪ Make arrangements so that the use of an AI system and AI service does not harm the lives, bodies, and property of users and third parties through actuators, etc. 	<ul style="list-style-type: none"> ▪ AI systems must have an explicit and well-defined domain of use, and the safety, security, and robustness of such systems must be tested and ensured throughout the lifecycle of the domain. [No. 4] ▪ AI must be built on a system that is secure and technologically robust, prevents harm, operates as intended, and is technically robust. The risk of unintended or unexpected harm must also be minimized. Technical robustness is also important for system accuracy, reliability, and repeatability. [No. 45] ▪ Efforts must be made to prevent potential risks and ensure safety throughout the entire lifecycle of AI development and use. Efforts must be made to provide users with the ability to control what happens when there is an obvious error or breach in the process of using AI. [No. 53]

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

Values that should be respected	Content considered relevant in guidelines (main)	Examples of main content
12. Security	<p>[AI R&D Guidelines]</p> <ul style="list-style-type: none"> ○ Principles of security <ul style="list-style-type: none"> ▪ Remain mindful of AI system security <p>[AI Utilization Guidelines]</p> <ul style="list-style-type: none"> ○ Principles of security <ul style="list-style-type: none"> ▪ Remain mindful of AI system and AI service security ○ Principles of appropriate learning <ul style="list-style-type: none"> ➢ Organizations are expected to remain mindful of the risk of AI learning inaccurate or inappropriate data, leading to vulnerabilities in AI security 	<ul style="list-style-type: none"> ▪ AI systems must have an explicit and well-defined domain of use, and the safety, security, and robustness of such systems must be tested and ensured throughout the lifecycle of the domain. [No. 4] ▪ In terms of security, it is important to ensure confidentiality, integrity, and availability throughout the lifecycle of an AI system. When developing an AI system, risk assessment including attack detection is necessary. [No. 39] ▪ The principles of security concern the physical and emotional safety of humans and the protection of the environment, and include the protection of critical assets. Security assurance requires compliance with strict requirements such as human-machine interaction and system resilience against attacks and abuse. [No. 41] ▪ Autonomous systems must meet the following safety and security requirements. [No. 63] <ol style="list-style-type: none"> (1) Safety for the environment and users (2) Reliability and robustness against hacking, etc. (3) Emotional safety of human-machine interactions ▪ To ensure the safety and security of humans, the environment, and ecosystems, safety and security risks must be avoided throughout the AI system lifecycle. [No. 66]
13. Privacy	<p>[AI R&D Guidelines]</p> <ul style="list-style-type: none"> ○ Principles of privacy <ul style="list-style-type: none"> ▪ Make arrangements so that the privacy of users and third parties is not infringed upon by AI systems ○ Principles of user support <ul style="list-style-type: none"> ▪ Make arrangements so that it is possible for AI systems to support users and provide them with appropriate opportunities to make choices <p>[AI Utilization Guidelines]</p> <ul style="list-style-type: none"> ○ Principles of privacy <ul style="list-style-type: none"> ▪ Make arrangements so that the privacy of users and others is not violated when utilizing an AI system or AI service 	<ul style="list-style-type: none"> ▪ The challenge in standardizing privacy and data governance is not only the management of data protection for AI, but also ensuring the quality of the data as a whole. This is particularly true when data for machine learning is provided from an external provider. [No. 40] ▪ It is necessary to develop norms for the collection, storage, processing, and use of personal information, and to protect the right of individuals to know and choose them. [No. 57] ▪ It is necessary to regulate data generated in or transiting through a country's territory, and to take steps towards effectively regulating data based on respecting privacy rights and other human rights. [No. 66] ▪ Privacy must be protected through an appropriate legal framework for data protection, and effective informed consent must be obtained. [No. 67]

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

Values that should be respected	Content considered relevant in guidelines (main)	Examples of main content
14. Fairness, Equity, removal of Discrimination	<p>[AI R&D Guidelines]</p> <ul style="list-style-type: none"> ○ Principles of ethics <ul style="list-style-type: none"> ➢ It is best to strive to implement necessary measures so that unfair discrimination does not occur due to bias and other factors contained in the learning data of AI systems, to the extent possible in light of the characteristics of the technology employed <p>[AI Utilization Guidelines]</p> <ul style="list-style-type: none"> ○ Principles of fairness <ul style="list-style-type: none"> ▪ Remain mindful of the fact that there may be bias in decisions made by AI systems and AI services, and make arrangements so that individuals and groups are not unfairly discriminated against through the decisions of AI systems and AI services 	<ul style="list-style-type: none"> ▪ It is important to eliminate the possibility of unintended discriminatory effects on individuals and groups. [No. 21] ▪ Fairness is a model learned and tested with appropriate, accurate, and generalizable datasets, and AI systems must be operated by users who are trained to implement responsibly and without bias. [No. 23] ▪ The diversity and representation of users must be reflected in the early stages of AI development and use, bias and discrimination based on individual characteristics (such as gender, age, disability, region, ethnicity, religion, or nation) must be minimized, and commercialized AI must be applied fairly to all. Efforts must be made to ensure that the socially weak and vulnerable have access to AI technologies and services, and that the benefits of AI are shared with all equally and not only a particular group. [No. 53] ▪ Decision-making AI algorithms, models, and learning datasets must ensure fairness. [No. 54]
15. Transparency, Explainability	<p>[AI R&D Guidelines]</p> <ul style="list-style-type: none"> ○ Principles of transparency <ul style="list-style-type: none"> ▪ Remain mindful of the verifiability of AI system input/output and the explainability of decisions <p>[AI Utilization Guidelines]</p> <ul style="list-style-type: none"> ○ Principles of transparency <ul style="list-style-type: none"> ▪ Remain mindful of the verifiability of AI system and AI service input/output, etc. and the explainability of decisions ○ Principles of appropriate use <ul style="list-style-type: none"> ➢ Organizations are expected to ensure the use of AI services under fair conditions, and to provide necessary information in a timely manner 	<ul style="list-style-type: none"> ▪ Ministries/agencies must disclose information related to the use of AI to appropriate stakeholders (including legislative assemblies and private citizens) to the extent possible in accordance with applicable laws and policies with regard to privacy protection, law enforcement, national security, and other protected confidential information, and must ensure transparency. [No. 1] ▪ AI systems must provide explanations that individual users can understand. [No. 2] ▪ Transparency refers to openness related to (1) what data is collected and for what purposes (to support AI-driven decision-making), and (2) the purpose for supporting algorithms and making decisions. This must be kept in mind when using AI to make decisions. Another important feature related to transparency is having the ability to trace data and the basis for making decisions that have influenced AI-driven decision-making. [No. 47] ▪ There is a tradeoff between transparency/explainability and other principles such as safety and security, so the level of transparency and explainability must always be appropriate. [No. 66]

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

Values that should be respected	Content considered relevant in guidelines (main)	Examples of main content
16. Accountability	<p>[AI R&D Guidelines]</p> <ul style="list-style-type: none"> ○ Principles of accountability <ul style="list-style-type: none"> ▪ Strive to ensure accountability to stakeholders, including users ○ Principles of user support <ul style="list-style-type: none"> ▪ It is best to strive to provide appropriate information to users in light of the possibility of changes being made to output or a program due to AI system learning, etc. <p>[AI Utilization Guidelines]</p> <ul style="list-style-type: none"> ○ With regard to accountability, it is best to strive to provide appropriate information to users in light of the possibility of changes being made to a program As a general rule, strive to ensure accountability ○ Principles of appropriate use <ul style="list-style-type: none"> ➢ Fair conditions for AI services <ul style="list-style-type: none"> ▪ Organizations are expected to ensure use by stakeholders, and to provide necessary information in a timely manner 	<ul style="list-style-type: none"> ▪ It is necessary to be able to explain actions and processes taken when designing and deploying AI models to others within an organization, external agencies such as regulators, and individuals affected by AI decisions. [No. 24] ▪ Researchers must explain the limitations they are aware of, such as the extent to which system decisions and behaviors are a result of source code and learning processes. In the absence or delay of such an explanation, the designer may be further responsible. [No. 51] ▪ Mechanisms must be introduced to ensure responsibility for the development, deployment, and use of AI systems. Accountability is closely related to risk management and requires identifying and mitigating risks in a transparent manner that can be explained to third parties and audited. [No. 64]
17. Robustness	<p>[AI R&D Guidelines] (No content)</p> <p>[AI Utilization Guidelines] (No content)</p>	<ul style="list-style-type: none"> ▪ When designing a robust and safe AI system, remain mindful of protecting humans and the environment from possible adverse effects from the system, in addition to protecting the system from external threats. [No. 41] ▪ AI must prevent harm, operate as intended, and be built on a technically robust system. The risk of unintended or unexpected harm must also be minimized. Technical robustness is also important for system accuracy, reliability, and repeatability. [No. 45] ▪ Reliability (the ability to provide legitimate and reliable services) and resilience (robustness in the face of change) are key requirements for reliable AI systems. Technically robust AI systems must be developed with a precautionary approach to risk, and must operate as intended while minimizing unintended and unexpected harm to the extent that this is possible. [No. 64]

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

Values that should be respected	Content considered relevant in guidelines (main)	Examples of main content
18. Responsibility	<p>[AI R&D Guidelines]</p> <ul style="list-style-type: none"> ○ Principles of accountability <ul style="list-style-type: none"> ▪ Strive to ensure accountability to stakeholders, including users <p>[AI Utilization Guidelines]</p> <ul style="list-style-type: none"> ○ Principles of accountability <ul style="list-style-type: none"> ▪ Strive to ensure accountability to stakeholders 	<ul style="list-style-type: none"> ▪ AI systems must be designed to be consistent with international treaties protecting human dignity, rights, and freedoms. The organization takes responsibility for implementing measures to avoid reasonably predictable misuse through the use of ethics by design. [No. 7] ▪ Responsibility means that AI-driven decision-making poses no threat to anyone's health or safety. This requirement applies not only to the physical and mental health of individuals, but also to data protection and privacy protection. As a requirement, responsibility means that decision-making should not exacerbate structural inequalities in society or otherwise cause injustice, injury, or suffering to individuals or groups. One way to influence this is to promote diversity among AI developers and other experts. [No. 47] ▪ All stakeholders involved in the design, development, and deployment of AI systems must take responsibility for their actions. Stakeholders must conduct risk and impact assessments to assess the direct and indirect potential impacts of AI systems on end users, establish audit processes (internal, and external when needed) to monitor compliance with principles, and establish mechanisms to resolve complaints in the event of adverse impacts. [No. 52] ▪ There is a responsibility to stakeholders to ensure that AI is used under appropriate conditions and by people with appropriate training. Accountability must be ensured in the event that AI technology causes an issue. [No. 67]
19. Traceability	<p>[AI R&D Guidelines]</p> <ul style="list-style-type: none"> ○ Principles of transparency <ul style="list-style-type: none"> ▪ Remain mindful of the verifiability of AI system input/output and the explainability of decisions <p>[AI Utilization Guidelines]</p> <ul style="list-style-type: none"> ○ Principles of transparency <ul style="list-style-type: none"> ▪ Remain mindful of the verifiability of AI system and AI service input/output, etc. and the explainability of decisions 	<ul style="list-style-type: none"> ▪ The design, development, acquisition, and use of AI, and the inputs and outputs associated with a particular AI application, must be well documented and traceable to the extent that this is appropriate and feasible. [No. 1] ▪ Decisions made by AI systems must be traceable, explainable, and transparent. This means that individuals and corporations must be given the opportunity to gain insight into how decisions affecting them were made. Traceability promotes auditability as well as explainability. [No. 45] ▪ Transparency refers to openness related to (1) what data is collected and for what purposes (to support AI-driven decision-making), and (2) the purpose for supporting algorithms and making decisions. This must be kept in mind when using AI to make decisions. Another important feature related to transparency is having the ability to trace data and the basis for making decisions that have influenced AI-driven decision-making. [No. 47]

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

Values that should be respected	Content considered relevant in guidelines (main)	Examples of main content
<p>20. Monitoring, Auditing</p>	<p>[AI R&D Guidelines]</p> <ul style="list-style-type: none"> ○ Principles of transparency <ul style="list-style-type: none"> ▪ Remain mindful of the verifiability of AI system input/output and the explainability of decisions ○ Principles of controllability <ul style="list-style-type: none"> ▪ To ensure controllability, one should remain mindful of the effectiveness of human or other reliable AI supervision (monitoring, alerts, etc.) and response (AI system outages, network disconnections, repairs, etc.), to the extent possible in light of the characteristics of the technology employed <p>[AI Utilization Guidelines]</p> <ul style="list-style-type: none"> ○ Principles of transparency <ul style="list-style-type: none"> ▪ Remain mindful of the verifiability of AI system and AI service input/output, etc. and the explainability of decisions 	<ul style="list-style-type: none"> ▪ AI systems must be designed to be fully answerable and auditable, and must implement activity monitoring to enable monitoring and reviewing throughout the entire project. [No. 22] ▪ Decisions made by AI systems must be traceable, explainable, and transparent. This means that individuals and corporations must be given the opportunity to gain insight into how decisions affecting them were made. Traceability promotes auditability as well as explainability. [No. 45] ▪ Appropriate monitoring, impact assessment, and due diligence mechanisms must be developed to ensure the accountability of AI systems and their impacts throughout their lifecycles. When designing technologies and systems, auditability and traceability of AI systems must be ensured to address threats, particularly with regard to human rights conflicts and the environment and ecosystems. [No. 66]
<p>21. Governance</p>	<p>[AI R&D Guidelines] (No content)</p> <p>[AI Utilization Guidelines] (No content)</p>	<ul style="list-style-type: none"> ▪ It is necessary to continually improve intelligent technical tools, optimize management mechanisms, improve governance systems, and promote governance throughout the lifecycles of AI products and services. [No. 57] ▪ Remain mindful to make use of AI governance practices in order to provide timely feedback on issues such as technical safety and policy/regulation gaps that will emerge when using AI products and services. [No. 58] ▪ Governance revisions must be considered so that AI principles, policies, and regulations suit AI developments. AI governance measures must be commensurate with AI development, not only to ensure that they do not interfere with proper use, but also to ensure that they are beneficial to society and nature. [No. 59]

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

Values that should be respected	Content considered relevant in guidelines (main)	Examples of main content
22. Other (cost)	<p>[AI R&D Guidelines] (No content)</p> <p>[AI Utilization Guidelines] ○ Principles of appropriate learning</p> <ul style="list-style-type: none"> • It is expected that a standard on the accuracy of decisions made by AI will be determined in advance, taking into account the scale of possible rights infringements, the frequency of rights infringements, the level of technology, the cost of maintaining accuracy, etc., as accuracy is expected to be impaired or degraded after the fact. 	<ul style="list-style-type: none"> • Organizations must consider how much it costs to build, operate, and maintain AI infrastructures; train and educate staff; and whether the work of implementing AI outweighs potential savings. [No. 22]
22. Other (effectiveness measurements)	<p>[AI R&D Guidelines] (No content)</p> <p>[AI Utilization Guidelines] (No content)</p>	<ul style="list-style-type: none"> • The benefits of using AI must be understood and measured. [No. 18]

(2) Verification through comparisons with principles, policies, guidelines, etc. instituted in other countries

<List of principles, policies, guidelines, etc.>

	Country, etc.	Organization, etc.	Name of principles, policies, guidelines, etc.
1	U.S.	The White House	Executive Order on Promoting the Use of Trustworthy Artificial Intelligence in the Federal Government URL< https://trumpwhitehouse.archives.gov/presidential-actions/executive-order-promoting-use-trustworthy-artificial-intelligence-federal-government/ >
2	U.S.	National Institute of Standards and Technology	Four Principles of Explainable Artificial Intelligence URL< https://www.nist.gov/system/files/documents/2020/08/17/NIST%20Explainable%20AI%20Draft%20NISTIR8312%20%281%29.pdf >
3	U.S.	Office of the Director of National Intelligence	Principles of Artificial Intelligence Ethics for the Intelligence Community URL< https://www.dni.gov/files/ODNI/documents/Principles_of_AI_Ethics_for_the_Intelligence_Community.pdf >
4	U.S.	Defense Innovation Board	AI Principles: Recommendations on the Ethical Use of Artificial Intelligence by the Department of Defense URL< https://media.defense.gov/2019/Oct/31/2002204458/-1/-1/0/DIB_AI_PRINCIPLES_PRIMARY_DOCUMENT.PDF >
5	U.S.	Federal Trade Commission	Using Artificial Intelligence and Algorithms URL< https://www.ftc.gov/news-events/blogs/business-blog/2020/04/using-artificial-intelligence-algorithms >
6	U.S.	Federal Trade Commission	Aiming for truth, fairness, and equity in your company's use of AI URL< https://www.ftc.gov/news-events/blogs/business-blog/2021/04/aiming-truth-fairness-equity-your-companys-use-ai >
7	U.S.	Information Technology Industry Council	AI POLICY PRINCIPLES URL< https://www.itic.org/public-policy/ITIAIPolicyPrinciplesFINAL.pdf >
8	U.S.	Software & Information Industry Association	Ethical Principles for Artificial Intelligence and Data Analytics URL< https://history.siiia.net/Portals/0/pdf/Policy/Ethical%20Principles%20for%20Artificial%20Intelligence%20and%20Data%20Analytics%20SIIA%20Issue%20Brief.pdf?ver=2017-11-06-160346-990 >

(2) Verification through comparisons with principles, policies, guidelines, etc. instituted in other countries

<List of principles, policies, guidelines, etc.>

	Country, etc.	Organization, etc.	Name of principles, policies, guidelines, etc.
9	U.S.	IBM	Everyday Ethics for Artificial Intelligence URL< https://www.ibm.com/watson/assets/duo/pdf/everydayethics.pdf >
10	U.S.	IBM	IBM's Principles for Trust and Transparency URL< https://www.ibm.com/blogs/policy/wp-content/uploads/2018/06/IBM_Principles_SHORT.V4.3.pdf >
11	U.S.	IBM	Our fundamental properties for trustworthy AI URL< https://www.ibm.com/artificial-intelligence/ai-ethics-focus-areas >
12	U.S.	intel	Intel's AI Privacy Policy White Paper URL< https://www.intel.com/content/dam/www/public/us/en/ai/documents/Intels-AI-Privacy-Policy-White-Paper-2018.pdf >
13	U.S.	Microsoft	Microsoft's Basic Principles of AI URL< https://www.microsoft.com/ja-jp/ai/responsible-ai?activetab=pivot1:primaryr6 >
14	U.S.	Microsoft	Responsible bots: 10 guidelines for developers of conversational AI URL< https://www.microsoft.com/en-us/research/uploads/prod/2018/11/Bot_Guidelines_Nov_2018.pdf >
15	U.S.	Google	Artificial Intelligence at Google: Our Principles URL< https://ai.google/principles/ >
16	U.S.	Meta (formerly Facebook)	Facebook's five pillars of Responsible AI URL< https://ai.facebook.com/blog/facebooks-five-pillars-of-responsible-ai/ >
17	U.S.	BSR	Artificial Intelligence: A Rights-Based Blueprint for Business Paper 1: Why a Rights-Based Approach? URL< https://www.bsr.org/reports/BSR-Artificial-Intelligence-A-Rights-Based-Blueprint-for-Business-Paper-01.pdf >

(2) Verification through comparisons with principles, policies, guidelines, etc. instituted in other countries

<List of principles, policies, guidelines, etc.>

	Country, etc.	Organization, etc.	Name of principles, policies, guidelines, etc.
18	Canada	Treasury Board of Canada Secretariat	Our guiding principles URL< https://www.canada.ca/en/government/system/digital-government/digital-government-innovations/responsible-use-ai.html#toc1 >
19	Canada	University of Montreal	Montreal Declaration for a Responsible Development of AI URL< https://www.montrealdeclaration-responsibleai.com/the-declaration >
20	UK	Artificial Intelligence Committee AI	Artificial Intelligence Committee AI in the UK: ready, willing and able? URL< https://publications.parliament.uk/pa/ld201719/ldselect/ldai/100/100.pdf >
21	UK	Central Digital & Data Office	Data Ethics Framework URL< https://www.gov.uk/government/publications/data-ethics-framework/data-ethics-framework-2020#overarching-principles >
22	UK	<ul style="list-style-type: none"> • Central Digital & Data Office • Office for Artificial Intelligence 	Understanding artificial intelligence ethics and safety URL< https://www.gov.uk/guidance/understanding-artificial-intelligence-ethics-and-safety >
23	UK	<ul style="list-style-type: none"> • Central Digital & Data Office • Office for Artificial Intelligence 	A guide to using artificial intelligence in the public sector URL< https://www.gov.uk/government/publications/understanding-artificial-intelligence/a-guide-to-using-artificial-intelligence-in-the-public-sector >
24	UK	Information Commissioner's Office	Explaining decisions made with AI URL< https://ico.org.uk/for-organisations/guide-to-data-protection/key-dp-themes/explaining-decisions-made-with-artificial-intelligence/ >
25	UK	Information Commissioner's Office	Guidance on AI and data protection URL< https://ico.org.uk/for-organisations/guide-to-data-protection/key-dp-themes/guidance-on-ai-and-data-protection/ >
26	UK	UK Finance	ETHICAL PRINCIPLES FOR ADVANCED ANALYTICS AND ARTIFICIAL INTELLIGENCE IN FINANCIAL SERVICES URL< https://www.ukfinance.org.uk/system/files/AAAI-Principles-FINAL.pdf >

(2) Verification through comparisons with principles, policies, guidelines, etc. instituted in other countries

<List of principles, policies, guidelines, etc.>

	Country, etc.	Organization, etc.	Name of principles, policies, guidelines, etc.
27	UK	Sage Group	The Ethics of Code: Developing AI for Business with Five Core Principles URL< https://www.sage.com/~/media/group/files/business-builders/business-builders-ethics-of-code.pdf?la=en >
28	UK	randstad	artificial intelligence principles URL< https://www.randstad.co.uk/s3fs-media/uk/public/2020-01/Randstad-Global-AI-principles.pdf >
29	UK	Astrazeneca	AstraZeneca's principles for ethical data and AI URL< https://www.astrazeneca.com/sustainability/ethics-and-transparency/data-and-ai-ethics.html >
30	Italy	Agenzia per l'Italia Digitale	White paper Artificial Intelligence at the service of the citizen URL< https://libro-bianco-ia.readthedocs.io/en/latest/ >
31	Italy	Rome Call for AI ethics	Rome Call for AI ethics URL< https://www.romecall.org/wp-content/uploads/2021/02/AI-Rome-Call-x-firma_DEF_DEF_con-firme_.pdf >
32	The Netherlands	Ministry of Economic Affairs and Climate Policy	Strategic Action Plan for Artificial Intelligence URL< https://www.government.nl/binaries/government/documents/reports/2019/10/09/strategic-action-plan-for-artificial-intelligence/Strategic+Action+Plan+for+Artificial+Intelligence.pdf >
33	The Netherlands	City of Amsterdam	Public AI Registers URL< https://algoritmeregister.amsterdam.nl/wp-content/uploads/White-Paper.pdf >
34	The Netherlands	PHILIPS	Five guiding principles for responsible use of AI in healthcare and healthy living URL< https://www.philips.com/a-w/about/news/archive/blogs/innovation-matters/2020/20200121-five-guiding-principles-for-responsible-use-of-ai-in-healthcare-and-healthy-living.html >
35	Sweden	Agency for Digital Government	Främja den offentliga förvaltningens förmåga att använda AI (<i>Promoting the Use of AI through Government Policy</i>) URL< https://www.digg.se/4a3a73/globalassets/dokument/publicerat/publikationer/framja-den-offentliga-forvaltningens-formaga-att-anvanda-ai.pdf >

(2) Verification through comparisons with principles, policies, guidelines, etc. instituted in other countries

<List of principles, policies, guidelines, etc.>

	Country, etc.	Organization, etc.	Name of principles, policies, guidelines, etc.
36	Sweden	Swedish Employment Agency, etc.	Together we are taking responsibility for a sustainable use of AI URL< https://www.almega.se/app/uploads/sites/2/2020/05/ittelekomforetagen-industry-code-for-ai-eng.pdf >
37	Denmark	<ul style="list-style-type: none"> • Ministry of Finance • Ministry of Industry, Business and Financial Affairs 	National Strategy for Artificial Intelligence URL< https://en.digst.dk/media/19337/305755_gb_version_final-a.pdf >
38	Denmark	Data Ethics Assessment Committee	Dataetik – Sådan gør du (<i>Data Ethics - What Must Be Done?</i>) URL< https://dataetiskraad.dk/sites/default/files/2021-10/Dataetik%20-%20S%C3%A5dan%20g%C3%B8r%20du.pdf >
39	Germany	Federal Government	Artificial Intelligence Strategy of the German Federal Government URL< https://www.ki-strategie-deutschland.de/files/downloads/Fortschreibung_KI-Strategie_engl.pdf >
40	Germany	Federal Ministry for Economic Affairs and Climate Action	GERMAN STANDARDIZATION ROADMAP ON ARTIFICIAL INTELLIGENCE URL< https://www.din.de/resource/blob/772610/e96c34dd6b12900ea75b460538805349/normungsroadmap-en-data.pdf >
41	Germany	Data Ethics Commission	Opinion of the Data Ethics Commission URL< https://www.bmju.de/SharedDocs/Downloads/DE/Themen/Fokusthemen/Gutachten_DEK_EN.pdf?__blob=publicationFile&v=2 >
42	Germany	BMW	BMW Group code of ethics for artificial intelligence URL< https://www.bmwgroup.com/content/dam/grpw/websites/bmwgroup_com/downloads/ENG_PR_CodeOfEthicsForAI_Short.pdf >
43	Germany	DAIMLER	Two Letters and Four Principles: How Daimler Uses Artificial Intelligence (AI) URL< https://www.daimler.com/sustainability/data/ki-guidelines.html >

(2) Verification through comparisons with principles, policies, guidelines, etc. instituted in other countries

<List of principles, policies, guidelines, etc.>

	Country, etc.	Organization, etc.	Name of principles, policies, guidelines, etc.
44	Germany	SAP	SAP's Guiding Principles for Artificial Intelligence URL< https://news.sap.com/2018/09/sap-guiding-principles-for-artificial-intelligence/ >
45	Norway	Ministry of Local Government and Modernisation	The National Strategy for Artificial Intelligence URL< https://www.regjeringen.no/en/dokumenter/nasjonal-strategi-for-kunstig-intelligens/id2685594/?ch=7 >
46	Norway	Norwegian Board of Technology	ARTIFICIAL INTELLIGENCE: OPPORTUNITIES, CHALLENGES AND A PLAN FOR NORWAY URL< https://teknologiradet.no/wp-content/uploads/sites/105/2018/11/AI-and-machine-learning-1.pdf >
47	Finland	Ministry of Economic Affairs and Employment	Work in the age of artificial intelligence URL< https://julkaisut.valtioneuvosto.fi/bitstream/handle/10024/160980/TEMjul_21_2018_Work_in_the_age.pdf >
48	Finland	Cabinet	Tekoäly viranomaistoiminnassa - eettiset kysymykset ja yhteiskunnallinen hyväksyttävyyys (<i>AI in Government Policy - Ethical Issues and Social Acceptance</i>) URL< https://julkaisut.valtioneuvosto.fi/bitstream/handle/10024/161345/14-2019-Tekoaly%20viranomaistoiminnassa.pdf?sequence=1&isAllowed=y >
49	Finland	OP Financial Group	OP Financial Group's ethical guidelines for artificial intelligence URL< https://www.op.fi/op-financial-group/corporate-social-responsibility/commitments-and-principles >
50	France	Commission Nationale de l'Informatique et des Libertés	HOW CAN HUMANS KEEP THE UPPER HAND? URL< https://www.cnil.fr/sites/default/files/atoms/files/cnil_rapport_ai_gb_web.pdf >
51	France	ALLISTENE	Research Ethics in Machine Learning URL< http://cerna-ethics-allistene.org/digitalAssets/54/54730_cerna_2017_machine_learning.pdf >
52	India	Indian Policy Commission	RESPONSIBLE AI #AIFORALL Approach Document for India Part 1 – Principles for Responsible AI URL< https://www.niti.gov.in/sites/default/files/2021-02/Responsible-AI-22022021.pdf >

(2) Verification through comparisons with principles, policies, guidelines, etc. instituted in other countries

<List of principles, policies, guidelines, etc.>

	Country, etc.	Organization, etc.	Name of principles, policies, guidelines, etc.
53	South Korea	Ministry of Science and ICT	사람이 중심이 되는「인공지능(AI) 윤리기준」(<i>Human-Centered "Artificial Intelligence (AI) Ethical Standards</i>) URL< https://www.msit.go.kr/bbs/view.do?sCode=user&mPid=112&mPid=113&bbsSeqNo=94&nttSeqNo=3179742 >
54	Singapore	Personal Data Protection Commission	DISCUSSION PAPER ON ARTIFICIAL INTELLIGENCE (AI) AND PERSONAL DATA – FOSTERING RESPONSIBLE DEVELOPMENT AND ADOPTION OF AI URL< https://www.pdpc.gov.sg/-/media/Files/PDPC/PDF-Files/Resource-for-Organisation/AI/Discussion-Paper-on-AI-and-PD---050618.pdf >
55	Singapore	Monetary Authority of Singapore	Principles to Promote Fairness, Ethics, Accountability and Transparency (FEAT) in the Use of Artificial Intelligence and Data Analytics in Singapore's Financial Sector URL< https://www.mas.gov.sg/~media/MAS/News%20and%20Publications/Monographs%20and%20Information%20Papers/FEAT%20Principles%20Final.pdf >
56	Singapore	Singapore Computer Society	AI ETHICS AND GOVERNANCE TOOLKIT URL< https://files-scs-prod.s3-ap-southeast-1.amazonaws.com/public/uploads/200721016+SCS+Booklet+6+Nov.pdf >
57	China	National New Generation Artificial Intelligence Governance Expert Committee	新一代人工智能治理原则—发展负责任的人工智能 (<i>Next-Generation AI Governance Principles - Responsible AI Development</i>) URL< http://www.most.gov.cn/kjbgz/201906/t20190617_147107.html >
58	China	National New Generation Artificial Intelligence Governance Expert Committee	新一代人工智能伦理规范 (<i>New-Generation Artificial Intelligence Ethics Rules</i>) URL< http://www.most.gov.cn/kjbgz/202109/t20210926_177063.html >
59	China	Beijing Academy of Artificial Intelligence	Beijing AI Principles URL< https://www-pre.baai.ac.cn/news/beijing-ai-principles-en.html >
60	China	Artificial Intelligence Industry Alliance	人工智能行业自律公约 (<i>Artificial Intelligence Industry Self-Regulation Pledge</i>) URL< http://aiaaorg.cn/uploadfile/2019/0808/20190808053719487.pdf >

(2) Verification through comparisons with principles, policies, guidelines, etc. instituted in other countries

<List of principles, policies, guidelines, etc.>

	Country, etc.	Organization, etc.	Name of principles, policies, guidelines, etc.
61	China	Baidu	AI伦理四原则 (<i>Four Principles of AI Ethics</i>) URL< https://ai.baidu.com/ai-doc/REFERENCE/xk3dwjgfe >
62	Australia	Department of Industry, Science, Energy and Resources	Australia's Artificial Intelligence Ethics Framework URL< https://www.industry.gov.au/data-and-publications/australias-artificial-intelligence-ethics-framework/australias-ai-ethics-principles >
63	EU	European Commission	Artificial Intelligence, Robotics and 'Autonomous' Systems URL< https://op.europa.eu/en/publication-detail/-/publication/dfebe62e-4ce9-11e8-be1d-01aa75ed71a1/language-en/format-PDF/source-78120382 >
64	EU	High-Level Expert Group on Artificial Intelligence	THE ASSESSMENT LIST FOR TRUSTWORTHY ARTIFICIAL INTELLIGENCE (ALTAI) for self assessment URL< https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=68342 >
65	EU	AI4People* * Organized as an industry association planned by companies	AI4People's Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations URL< https://www.eismd.eu/wp-content/uploads/2019/03/AI4People%e2%80%99s-Ethical-Framework-for-a-Good-AI-Society.pdf >
66	International organizations	United Nations Educational, Scientific and Cultural Organization (UNESCO)	FIRST DRAFT OF THE RECOMMENDATION ON THE ETHICS OF ARTIFICIAL INTELLIGENCE URL< https://unesdoc.unesco.org/ark:/48223/pf0000373434?posInSet=6&queryId=ed3e3a0d-e278-46a2-898f-3e24d43f75cc >
67	International organizations	World Health Organization (WHO)	ETHICS AND GOVERNANCE OF ARTIFICIAL INTELLIGENCE FOR HEALTH URL< https://apps.who.int/iris/rest/bitstreams/1352854/retrieve >

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

○ The scope and content of the current AI R&D Guidelines and AI Utilization Guidelines were examined based on guidelines and policies established by domestic business operators and groups

■ Surveyed guidelines and policies

- The survey covered 22 guidelines and other policies announced by business operators and organizations in Japan (refer to pages 36-37).

■ Values that should be respected

- The survey identified 22 values that should be respected^{*7}

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

^{*7} 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.

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

- 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.

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

Values that should be respected	Content considered relevant in guidelines (main)	Examples of main content
1. Human-centered	<p>[AI R&D Guidelines]</p> <ul style="list-style-type: none"> ○ Basic philosophy <ul style="list-style-type: none"> ▪ Realize a human-centered society in which human dignity and individual autonomy are respected <p>[AI Utilization Guidelines]</p> <ul style="list-style-type: none"> ○ Basic philosophy <ul style="list-style-type: none"> ▪ Realize a human-centered society in which human dignity and individual autonomy are respected 	<ul style="list-style-type: none"> ▪ In accordance with the principles of freedom, fairness, and justice, implement AI in society from a human-centered perspective and strive to verify that AI functions as expected. [No. 13] ▪ Based on a human-centric philosophy, define AI as a tool to expand the potential of all people and help them pursue happiness and contribute to society in whichever way they desire. Aim to develop AI that people can use with peace of mind by protecting their privacy and implementing it with fairness and safety in mind, so that people will not be discriminated against or harmed. [No. 14] ▪ Aim to realize a human-centered society in which the benefits of using AI, etc. are enjoyed by many people, and in which human dignity and individual autonomy are respected. [No. 16]
2. Human dignity	<p>[AI R&D Guidelines]</p> <ul style="list-style-type: none"> ○ Principles of ethics <ul style="list-style-type: none"> ▪ Respect human dignity and individual autonomy <p>[AI Utilization Guidelines]</p> <ul style="list-style-type: none"> ○ Principles of dignity and autonomy <ul style="list-style-type: none"> ▪ Respect human dignity and individual autonomy 	<ul style="list-style-type: none"> ▪ In order to respect human dignity, excessive reliance on AI must be avoided. Strive to develop and provide AI services that take into account the appropriate division of roles between humans and AI. [No. 8] ▪ Respect human dignity and individual autonomy, remain mindful of the quality of data used in learning, and ensure accountability to stakeholders. [No. 11] ▪ Design and develop AI that can be used by humans at all times with peace of mind. Implement and enhance security measures throughout the lifecycle of AI design, development, and operation, and strive to prevent abuses that threaten human dignity. In the unlikely event that an unexpected event occurs, strive to take appropriate measures. [No. 12]

Note: The numbers indicated in "Examples of main content" refer to principles, policies, guidelines, etc. corresponding to the numbers in the list on pages 36-37.

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

Values that should be respected	Content considered relevant in guidelines (main)	Examples of main content
3. Diversity, Inclusiveness	<p>[AI R&D Guidelines]</p> <ul style="list-style-type: none"> ○ Basic philosophy <ul style="list-style-type: none"> ▪ Realize a human-centered society in which humans and AI networks co-exist, the benefits of AI networks are widely enjoyed by all, and human dignity and individual autonomy are respected ○ Principles of user support <ul style="list-style-type: none"> ➢ Strive to engage in initiatives to promote use by the socially weak, such as implementing universal design ○ Roles expected of related stakeholders <ul style="list-style-type: none"> ▪ In implementing and reviewing these guidelines, it is expected that organizations will strive to develop an environment that promotes dialog between diverse stakeholders, including national governments, international organizations, developers, and users (including civil society) ▪ Organizations are expected to strive to share awareness and engage in mutual cooperation in promoting the benefits and reducing the risks of AI, while ensuring the diversity of discussions on AI <p>[AI Utilization Guidelines]</p> <ul style="list-style-type: none"> ○ Basic philosophy <ul style="list-style-type: none"> ▪ Realize a human-centered society in which humans and AI networks co-exist, the benefits of AI networks are widely enjoyed by all, and human dignity and individual autonomy are respected ▪ Respect the diversity of users and include people with diverse backgrounds, values, and thinking when making use of and applying AI 	<ul style="list-style-type: none"> ▪ Strive to respect the human rights and diversity of customers and stakeholders so as not to cause unfair discrimination in the use of AI. [No. 4] ▪ While working to realize safe, robust, and easy-to-use products and services, actively promote the use of AI so that more people can enjoy the benefits of AI. In particular, aim to lead the way in solving social issues in a wide range of fields including healthcare, by rapidly creating new value through the integration of advanced and proprietary technologies developed over the years and the promotion of co-creation with outside companies. [No. 15] ▪ Actively develop and expand AI technologies, based on the philosophy that AI technologies that support the realization of diverse lifestyles of diverse people are the future of mankind and meant to realize the diverse dreams of diverse people. [No. 20]
4. Sustainable society	<p>[AI R&D Guidelines] (No content)</p> <p>[AI Utilization Guidelines]</p> <ul style="list-style-type: none"> ○ Basic philosophy <ul style="list-style-type: none"> ▪ Realize a sustainable society by solving various problems faced by individuals, local communities, countries, and the international community 	<ul style="list-style-type: none"> ▪ Strive to use the power of AI to help solve global problems and to develop a peaceful and sustainable society. [No. 9] ▪ Plan to develop and utilize AI to solve diverse social issues, realize a comfortable and resilient sustainable society, and improve the quality of life of people around the world. [No. 13] ▪ Aim to realize a sustainable future society by using AI, etc. to solve various social issues. [No. 16]

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

Values that should be respected	Content considered relevant in guidelines (main)	Examples of main content
5. International cooperation	<p>[AI R&D Guidelines]</p> <ul style="list-style-type: none"> ○ Basic philosophy <ul style="list-style-type: none"> ▪ Share policies and best practices that will serve as non-binding soft law internationally between stakeholders, with regard to how AI system research and development should be conducted ○ Roles expected of related stakeholders <ul style="list-style-type: none"> ▪ In implementing and reviewing these guidelines, it is expected that national governments and international organizations will strive to develop an environment that promotes dialogue between diverse stakeholders, including national governments, international organizations, developers, and users (including civil society) <p>[AI Utilization Guidelines]</p> <ul style="list-style-type: none"> ○ Basic philosophy <ul style="list-style-type: none"> ▪ Share policies and best practices that will serve as non-binding soft law internationally, with regard to how AI should be used and applied 	<ul style="list-style-type: none"> ▪ Remain aware of changes in the environment and values concerning AI, and to changes in how business is done (such as expanding partnerships). In promoting the use and application of AI, engage in dialogue and deepen mutual understanding with diverse stakeholders (including customers and partners in Japan and overseas), comply with the laws and regulations of each country and region, and pursue fair and harmonious use and application of AI. [No. 6]
6. Proper utilization	<p>[AI R&D Guidelines]</p> <ul style="list-style-type: none"> ○ Principles of controllability <ul style="list-style-type: none"> ➢ To ensure controllability, it is desirable to remain mindful of the effectiveness of human or other reliable AI supervision (monitoring, alerts, etc.) and response (AI system outages, network disconnections, repairs, etc.), to the extent possible in light of the characteristics of the technology employed <p>[AI Utilization Guidelines]</p> <ul style="list-style-type: none"> ○ Principles of appropriate use <ul style="list-style-type: none"> ▪ Strive to use AI systems and AI services within an appropriate scope and in an appropriate manner, based on an appropriate division of roles between humans and AI systems and between users 	<ul style="list-style-type: none"> ▪ When providing products and services, strive to ensure that they are used properly in a manner that respects human rights. [No. 10] ▪ In providing businesses using AI systems, strive to maintain a division of duties with individuals within an appropriate scope and in an appropriate manner. [No. 11] ▪ Engage in development while taking the potential for risk in the actual use and application of AI (use cases) into consideration, and strive to use and apply AI appropriately, so that usage does not diverge from the intended use or operation conditions of AI. In addition, strive to ensure the proper operation of AI by stipulating usage policies and conditions to users and operators. [No. 13]

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

Values that should be respected	Content considered relevant in guidelines (main)	Examples of main content
7. Education/Literacy	<p>[AI R&D Guidelines] (No content)</p> <p>[AI Utilization Guidelines]</p> <ul style="list-style-type: none"> ○ Principles of appropriate use (reference) <ul style="list-style-type: none"> ➤ If it is deemed appropriate for consumer-type users to make final judgments on AI decisions, consumer-type users should acquire the necessary skills and knowledge to make appropriate judgments. ➤ When using/applying AI operated through actuators, etc., consumer-type users should be aware of the locus of responsibility in each state (before, during, and after transitioning) in cases where operation is expected to switch to human operation under certain conditions. These users should also receive explanations from the AI service provider on conditions and methods for switching, and should acquire necessary skills and knowledge. 	<ul style="list-style-type: none"> • Acquire the latest knowledge on these issues by actively participating in organizations such as industry associations and public research groups, and by actively surveying the latest academic findings. [No. 1] • Strive to gather information on the technological development of AI, its impact on society, and discussions on issues. Then, strive to expand training for executives and employees on AI knowledge and ethics, in order to respond to the increasing sophistication and complexity of AI and to prevent its misuse. [No. 7] • Having recognized the impact on society of products and services using AI, help develop AI that will help realize a better society, and actively work to develop human resources who can shape a bright future through the use of AI and research and development. [No. 9]
8. Human intervention, Controllability	<p>[AI R&D Guidelines]</p> <ul style="list-style-type: none"> ○ Principles of controllability <ul style="list-style-type: none"> • Remain mindful of AI system controllability ➤ To ensure controllability, it is desirable to remain mindful of the effectiveness of human or other reliable AI supervision (monitoring, alerts, etc.) and response (AI system outages, network disconnections, repairs, etc.), to the extent possible in light of the characteristics of the technology employed <p>[AI Utilization Guidelines]</p> <ul style="list-style-type: none"> ○ Principles of appropriate use <ul style="list-style-type: none"> ➤ When necessary and possible, it is expected that human decision-making will be involving in decisions made by AI, with regard to whether and how to use those decisions. In such cases, it is expected that standard examples will be taken into account and considerations will be made depending on the area of use and its applications, with regard to whether the intervention of human decision-making is required ○ Principles of fairness <ul style="list-style-type: none"> ➤ In order to ensure the fairness of decisions made by AI, it is expected that the social context when using/applying AI will be taken into consideration along with the reasonable expectations of people, and that human decision-making will be involved with regard to deciding whether to use these decisions or how to use them 	<ul style="list-style-type: none"> • AI developers must remain mindful of controllability of AI system, etc. When necessary and possible, it is expected of AI users that human decision-making will be involving in decisions made by AI, with regard to whether and how to use those decisions. [No. 5] • It is important for people to examine and make decisions on suggestions and results derived by AI. Even when promoting automation through AI, aim to provide mechanisms to accurately show the basis for suggestions and results, for the system as a whole. [No. 14] • Machine learning is just a tool, and humans decide how to use it. Machine learning has the potential to contribute significantly to the prosperity of human society, but may be against the interests of human society if used inappropriately. Machine learning predicts the future based on past examples, so predicting the future based on a biased past could result in bias. If we are seeking a "better future" different from the past, we must be careful to intervene in machine learning so that machine learning predictions and judgments are not unfair. [No. 19]

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

Values that should be respected	Content considered relevant in guidelines (main)	Examples of main content
9. Proper data	<p>[AI R&D Guidelines]</p> <ul style="list-style-type: none"> ○ Principles of ethics <ul style="list-style-type: none"> ➢ It is best to strive to implement necessary measures so that unfair discrimination does not occur due to bias and other factors contained in the learning data of AI systems, to the extent possible in light of the characteristics of the technology employed <p>[AI Utilization Guidelines]</p> <ul style="list-style-type: none"> ○ Principles of appropriate learning <ul style="list-style-type: none"> • Remain mindful of the quality of data used for AI system learning, etc. 	<ul style="list-style-type: none"> • Respect human dignity and individual autonomy, remain mindful of the quality of data used in learning, and ensure accountability to stakeholders. [No. 11] • Remain mindful of the quality and quantity of data used for AI, etc., and strive to evaluate AI, etc. appropriately. Also remain mindful of the risks and security associated with collaboration among systems and networks that constitute AI, etc. [No. 16] • When carrying out processing such as profiling personnel data, business operators must strive to ensure the accuracy and currency of both the original data and the processing results. [No. 21]
10. Collaboration among AI systems	<p>[AI R&D Guidelines]</p> <ul style="list-style-type: none"> ○ Principles of collaboration <ul style="list-style-type: none"> • Remain mindful of the interconnectivity and interoperability of AI systems <p>[AI Utilization Guidelines]</p> <ul style="list-style-type: none"> ○ Principles of collaboration <ul style="list-style-type: none"> • Remain mindful of collaboration among AI systems and AI services 	<ul style="list-style-type: none"> • Collaboration between humans and AI, and between AIs, is a challenge for a better human-centered society, and we will continue to investigate how to accomplish this. [No. 3] • AI developers, AI users, and AI data providers must remain mindful of the possibility of risks emerging and being amplified through the collaboration and networking of AI systems, etc. [No. 5] • Remain mindful of risks and security associated with collaboration among systems and networks that constitute AI, etc. [No. 16] • Open source software can be subject to sudden fixes and updates. Some changes can even break compatibility. It is necessary to determine how to change systems if there is a change to open source software. [No. 22]
11. Safety	<p>[AI R&D Guidelines]</p> <ul style="list-style-type: none"> ○ Principles of safety <ul style="list-style-type: none"> • Make arrangements so that AI systems do not harm the lives, bodies, and property of users and third parties through actuators, etc. <p>[AI Utilization Guidelines]</p> <ul style="list-style-type: none"> ○ Principles of safety <ul style="list-style-type: none"> • Make arrangements so that the use of an AI system and AI service does not harm the lives, bodies, and property of users and third parties through actuators, etc. 	<ul style="list-style-type: none"> • When using and applying AI, prioritize ensuring safety, and strive to apply and operate AI in a risk-conscious manner, so as not to adversely affect society and people through unintended AI actions. [No. 6] • Strive not to harm human life, body, or property, including the fundamental rights of human beings. [No. 8] • Make arrangements so that there is no dependence on AI, and that AI does not limit or deny human dignity, skill or potential or endanger the human body or life. Develop and provide products and services that respect basic human rights, so that AI is not maliciously or mistakenly used for intended or unintended purposes that may cause this. [No. 15]

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

Values that should be respected	Content considered relevant in guidelines (main)	Examples of main content
12. Security	<p>[AI R&D Guidelines]</p> <ul style="list-style-type: none"> ○ Principles of security <ul style="list-style-type: none"> ▪ Remain mindful of AI system security <p>[AI Utilization Guidelines]</p> <ul style="list-style-type: none"> ○ Principles of security <ul style="list-style-type: none"> ▪ Remain mindful of AI system and AI service security ○ Principles of appropriate learning <ul style="list-style-type: none"> ➢ Organizations are expected to remain mindful of the risk of AI learning inaccurate or inappropriate data, leading to vulnerabilities in AI security 	<ul style="list-style-type: none"> ▪ In addition to protecting personal information and respecting privacy, strive to ensure the appropriate acquisition, use, management, and security of data, so as not to cause undue damage to customers and other stakeholders. [No. 3] ▪ Given the importance and sensitivity of the data we handle, we will attempt to enhance security in AI services. [No. 8] ▪ Verify the quality of AI-equipped products and services so that they operate as expected, and strive to ensure safety so that they do not threaten life and freedom. Also, continue to implement security measures such as protections against unauthorized access by third parties. [No. 17]
13. Privacy	<p>[AI R&D Guidelines]</p> <ul style="list-style-type: none"> ○ Principles of privacy <ul style="list-style-type: none"> ▪ Make arrangements so that the privacy of users and third parties is not infringed by AI systems ○ Principles of user support <ul style="list-style-type: none"> ▪ Make arrangements so that it is possible for AI systems to support users and provide them with appropriate opportunities to make choices <p>[AI Utilization Guidelines]</p> <ul style="list-style-type: none"> ○ Principles of privacy <ul style="list-style-type: none"> ▪ Make arrangements so that the privacy of users and others is not violated when utilizing an AI system or AI service 	<ul style="list-style-type: none"> ▪ Comply with laws, regulations, and the relevant internal rules for products and services that use AI. Protect and handle personal information in accordance with the wishes of customers, and strive to protect privacy. [No. 4] ▪ AI developers, AI users, and AI data providers make arrangements so that the privacy of the user and others is not violated when utilizing an AI system, etc. [No. 5] ▪ Prior to collecting data such as personal information, HR provides the individual providing the data with an explanation of the purpose of the data collection, the type of data collected, the rights of the individual, etc., and obtains their consent. HR must be responsible for making the best effort to ensure that consent is a "substantial agreement" with the understanding and satisfaction of the individual providing the data. The individual has the right to withdraw consent. [No. 18]

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

Values that should be respected	Content considered relevant in guidelines (main)	Examples of main content
14. Fairness, Equity, removal of Discrimination	<p>[AI R&D Guidelines]</p> <ul style="list-style-type: none"> ○ Principles of ethics <ul style="list-style-type: none"> ➢ It is best to strive to implement necessary measures so that unfair discrimination does not occur due to bias and other factors contained in the learning data of AI systems, to the extent possible in light of the characteristics of the technology employed <p>[AI Utilization Guidelines]</p> <ul style="list-style-type: none"> ○ Principles of fairness <ul style="list-style-type: none"> ▪ Remain mindful of the fact that there may be bias in decisions made by AI systems and AI services, and make arrangements so that individuals and groups are not unfairly discriminated against through the decisions of AI systems and AI services 	<ul style="list-style-type: none"> ▪ AI developers, AI users, and AI data providers remain mindful of the fact that there may be bias in decisions made by AI systems, etc. and make arrangements so that individuals are not unfairly discriminated against through the decisions of AI systems, etc. [No. 5] ▪ Strive to implement and use AI so that AI decisions will serve the interests of diverse stakeholders and not cause or promote discrimination or prejudice based on ethnicity, gender, nationality, etc. [No. 13] ▪ HR constantly verifies the fairness of the design philosophy to ensure that people are not unfairly discriminated against for their diverse backgrounds, including ethnicity, gender, nationality, age, political beliefs, and religion, when using AI. [No. 18]
15. Transparency, Explainability	<p>[AI R&D Guidelines]</p> <ul style="list-style-type: none"> ○ Principles of transparency <ul style="list-style-type: none"> ▪ Remain mindful of the verifiability of AI system input/output and the explainability of decisions <p>[AI Utilization Guidelines]</p> <ul style="list-style-type: none"> ○ Principles of transparency <ul style="list-style-type: none"> ▪ Remain mindful of the verifiability of AI system and AI service input/output, etc. and the explainability of decisions ○ Principles of appropriate use <ul style="list-style-type: none"> ➢ Organizations are expected to ensure the use of AI services under fair conditions, and to provide necessary information in a timely manner 	<ul style="list-style-type: none"> ▪ In the AI business, strive to explain to customers and other stakeholders the intended purposes and methods of use, the effects and impact of use, and limitations, in accordance with the nature of the AI products and the situations in which they are used. In particular, strive to maintain transparency for AI decisions made in individual AI products, and to provide information to help others understand what AI decisions are. [No. 3] ▪ Strive to provide information and explanations that are easy for customers to understand by seeking the possibility of introducing mechanisms to explain the reasons for decisions made by AI in products and services, from the planning and design stages of products and services. [No. 4] ▪ Verify the validity of AI service input/output and decisions, and strive to provide transparent explanations. [No. 16]

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

Values that should be respected	Content considered relevant in guidelines (main)	Examples of main content
16. Accountability	<p>[AI R&D Guidelines]</p> <ul style="list-style-type: none"> ○ Principles of accountability <ul style="list-style-type: none"> ▪ Strive to ensure accountability to stakeholders, including users ○ Principles of user support <ul style="list-style-type: none"> ▪ It is best to strive to provide appropriate information to users in light of the possibility of changes being made to output or a program due to AI system learning, etc. <p>[AI Utilization Guidelines]</p> <ul style="list-style-type: none"> ○ Principles of accountability <ul style="list-style-type: none"> ▪ Strive to ensure accountability to stakeholders ○ Principles of appropriate use <ul style="list-style-type: none"> ➢ Organizations are expected to ensure the use of AI services under fair conditions, and to provide necessary information in a timely manner 	<ul style="list-style-type: none"> ▪ We recognize that there is an inherent opacity in the decision-making processes of AI, and strive to increase transparency in its use and application, and to ensure accountability to society honestly and openly. [No. 6] ▪ Provide appropriate explanations on the effects, value, and impact of AI utilization, and strive to gain the understanding of all stakeholders. [No. 10] ▪ Strive to ensure accountability by engaging in dialogue with diverse stakeholders regarding the purposes and methods of using AI and the various potential impacts of using AI-equipped products and services. [No. 17]
17. Robustness	<p>[AI R&D Guidelines] (No content)</p> <p>[AI Utilization Guidelines] (No content)</p>	<ul style="list-style-type: none"> ▪ While working to realize safe, robust, and easy-to-use products and services, actively promote the use of AI so that more people can enjoy the benefits of AI. [No. 15] ▪ It is necessary to develop high-quality AI technologies that ensure reliability, privacy, security, and dependability by ensuring fairness, accountability, transparency, etc., and to establish operational standards and systems. [No. 20] ▪ It is also necessary to evaluate the "generalization performance" of how accurate an answer can be given for data not used for learning, and the "robustness" of whether an accurate answer can be given even if the data contains noise. [No. 22]

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

Values that should be respected	Content considered relevant in guidelines (main)	Examples of main content
18. Responsibility	<p>[AI R&D Guidelines]</p> <ul style="list-style-type: none"> ○ Principles of accountability <ul style="list-style-type: none"> ▪ Strive to ensure accountability to stakeholders, including users <p>[AI Utilization Guidelines]</p> <ul style="list-style-type: none"> ○ Principles of accountability <ul style="list-style-type: none"> ▪ Strive to ensure accountability to stakeholders 	<ul style="list-style-type: none"> ▪ HR takes moral and technical responsibility for the collection and storage of personal information, and for the security of AI systems themselves and their resulting data. [No. 18] ▪ When handling personnel data, establish an organizational structure that clarifies the locus of responsibility by establishing a department specializing in people analytics and appointing an officer (such as a data protection officer) who is responsible for the company-wide protection of personnel data, from the viewpoint of accurately grasping the global and multipolar changing situation and realizing a balance between legitimate and appropriate protection of the rights and interests of individuals and utilization. [No. 21] ▪ It is also necessary to be aware of the division of responsibility in the event of damage due to a defect in collected data. <ul style="list-style-type: none"> [1] When purchasing data from a data provider, remain mindful of the rules for sharing responsibility in the purchase agreement. This is usually a disclaimer for data providers. [2] When collecting open data via the Internet, etc., remain mindful that the company is also responsible for data defects. [No. 22]
19. Traceability	<p>[AI R&D Guidelines]</p> <ul style="list-style-type: none"> ○ Principles of transparency <ul style="list-style-type: none"> ▪ Remain mindful of the verifiability of AI system input/output and the explainability of decisions <p>[AI Utilization Guidelines]</p> <ul style="list-style-type: none"> ○ Principles of transparency <ul style="list-style-type: none"> ▪ Remain mindful of the verifiability of AI system and AI service input/output, etc. and the explainability of decisions 	<ul style="list-style-type: none"> ▪ Prevent misuse of data and provide customers and users with peace of mind and safety by striving to create a system that allows for the appropriate acquisition, use, and provision of personal information; ensure data traceability; and prevent unauthorized access. [No. 2] ▪ AI developers and AI users must remain mindful of the verifiability of AI system, etc. input/output, etc. and the explainability of decisions. [No. 5] ▪ Recognize the importance of social responsibility borne by information and communications technology (which serves as the foundation of society for all people), strive to improve the quality of AI so that it does not produce undesirable or unforeseen results, implement a mechanism to investigate causes as needed, and aim for a safe and secure social system. [No. 14]

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

Values that should be respected	Content considered relevant in guidelines (main)	Examples of main content
20. Monitoring, Auditing	<p>[AI R&D Guidelines]</p> <ul style="list-style-type: none"> ○ Principles of transparency <ul style="list-style-type: none"> ▪ Remain mindful of the verifiability of AI system input/output and the explainability of decisions ○ Principles of controllability <ul style="list-style-type: none"> ▪ To ensure controllability, one should remain mindful of the effectiveness of human or other reliable AI supervision (monitoring, alerts, etc.) and response (AI system outages, network disconnections, repairs, etc.), to the extent possible in light of the characteristics of the technology employed <p>[AI Utilization Guidelines]</p> <ul style="list-style-type: none"> ○ Principles of transparency <ul style="list-style-type: none"> ▪ Remain mindful of the verifiability of AI system and AI service input/output, etc. and the explainability of decisions 	<ul style="list-style-type: none"> ▪ In order to ensure fairness when building and providing all services utilizing AI, continue to have the results of AI utilization monitored by internal and external experts. [No. 7] ▪ HR considers potential problems that could occur when using AI, enhances their ability to take technical and non-technical measures both in advance and after the fact, and conducts ongoing inspections [No. 18] ▪ During the operation phase of AI cloud services, output results may change due to changes in input data or additional learning, and business operators will need to monitor the operation status. [No. 22]
21. Governance	<p>[AI R&D Guidelines] (No content)</p> <p>[AI Utilization Guidelines] (No content)</p>	<ul style="list-style-type: none"> ▪ In addition to acquiring knowledge and awareness of the above issues, management conducts company-wide management to respond to these issues. [No. 1] ▪ Establish a committee composed of outside experts, etc. to share objective evaluations with the Board of Directors, in order to enhance corporate governance with regard to AI ethics. [No. 14]
22. Cost	<p>[AI R&D Guidelines] (No content)</p> <p>[AI Utilization Guidelines]</p> <ul style="list-style-type: none"> ○ Principles of appropriate learning <ul style="list-style-type: none"> ▪ It is expected that a standard on the accuracy of decisions made by AI will be determined in advance, taking into account the scale of possible rights infringements, the frequency of rights infringements, the level of technology, the cost of maintaining accuracy, etc., as accuracy is expected to be impaired or degraded after the fact. 	<ul style="list-style-type: none"> ▪ The data used for learning is extremely important when developing AI, and collection methods and scopes must be examined at an early stage. Data collection and processing require a lot of work and costs, so it is necessary to allow sufficient time to prepare data when creating schedules. [No. 22]

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

<List of principles, policies, guidelines, etc.>

	Business operator, organization	Name of guideline
1	ABEJA, Inc.	AI Policy URL< https://abejainc.com/ja/contact/ai_policy/ >
2	NTT DATA Corporation	NTT Data Group AI Guidelines URL< https://www.nttdata.com/jp/ja/-/media/nttdatajapan/files/news/release/2019/nttdata_ai_guidelines.pdf >
3	Ok Electric Industry Co., Ltd.	OKI Group AI Principles URL< https://www.oki.com/jp/press/2019/09/z19033.pdf >
4	Koozyt, Inc.	Koozyt AI Ethics Guidelines URL< https://www.koozyt.com/aiethicsguidelines >
5	KDDI CORPORATION	AI R&D and Utilization Principles for KDDI Group URL< https://www.kddi.com/corporate/kddi/public/ai_principles/ >
6	KONICA MINOLTA, INC.	KONICA MINOLTA Use of AI Basic Policy URL< https://www.konicaminolta.jp/about/csr/pdf/use-of-ai-basic-policy.pdf >
7	J.Score Co., LTD.	AI Utilization Policy URL< https://www.jscore.co.jp/ai_policy/ >
8	STADIUM Co., Ltd.	Basic Stance on AI URL< https://stadium.co.jp/policy/ai >
9	Sony Group Corporation	Sony Group AI Ethics Guidelines URL< https://www.sony.com/ja/SonyInfo/csr_report/humanrights/AI_Engagement_within_Sony_Group_Ja.pdf >
10	NEC Corporation	NEC Group AI & Human Rights Policies URL< https://jpn.nec.com/press/201904/images/0201-01-01.pdf >
11	Nihon Unisys, Ltd.*	BIPROGY Group AI Ethics Principles ("Nihon Unisys Group AI Ethics Principles" at the time of publication) * Company name at the time of publication. The company changed its name to BIPROGY Inc. on April 1, 2022. URL< https://www.biprogy.com/com/ai_ethics_principles_nihonunisys_group.pdf >

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

	Business operator, organization	Name of guideline
12	Nomura Research Institute, Ltd.	NRI Group AI Ethics Guidelines URL< https://www.nri.com/jp/sustainability/social/policies >
13	Hitachi, Ltd.	AI Ethics Principles for Utilizing AI in Social Innovation Business URL< https://www.hitachi.co.jp/products/it/lumada/about/ai/ldsl/document/ai_document_jp.pdf >
14	FUJITSU LIMITED	Fujitsu Group AI Commitment URL< https://www.fujitsu.com/jp/documents/about/csr/humanrights/fujitsu-group-ai-commitment-201903_ja.pdf >
15	FUJIFILM Holdings Corporation	FUJIFILM Group AI Basic Policy URL< https://holdings.fujifilm.com/ja/sustainability/vision/policy/ai >
16	Mitsubishi Research Institute, Inc.	Mitsubishi Research Institute AI Business Promotion Guidelines URL< https://www.mri.co.jp/sustainability/governance/dia6ou000001oqqk-att/ai-guideline20191023.pdf >
17	Mitsubishi Electric Corporation	Mitsubishi Electric Group "AI Ethics Policy" URL< https://www.mitsubishielectric.co.jp/corporate/sustainability/social/humanrights/aipolicy/index.html >
18	Recruit Co., Ltd. (Recruit Works Institute) *	AI Principles for Human Resources ("Tentative Version") * Draft prepared by Works Editorial Department and released after discussion with experts, etc. in a special feature in issue 156 of the "Works" publication. URL< https://www.works-i.com/works/no156/ >
19	The Japanese Society for Artificial Intelligence	Statement on Machine Learning and Fairness URL< http://ai-elsi.org/archives/888 >
20	KEIDANREN (Japan Business Federation)	AI Utilization Principles URL< https://www.keidanren.or.jp/policy/2019/013_honbun.pdf >
21	People Analytics & HR Technology Association	Principles for Utilizing HR Data URL< https://peopleanalytics.or.jp/media/HRDataUtilizationPrinciples.pdf >
22	AI Cloud Services Study Group *	Guidebook on AI-based Cloud Services * MIC announced a guideline based largely on the examinations of this Study Group. URL< https://www.soumu.go.jp/main_content/000792669.pdf >

(4) Verification of consistency with initiative case studies presented in the hearings

- The scope and content of the current AI R&D Guidelines and AI Utilization Guidelines were mainly examined based on initiatives, etc. of business operators as indicated in "Report 2021" (August 2021) and "Examples of Initiatives Related to AI Governance" (September 2021)*8.
- These were organized into several categories thought to be directly and deeply related to the objectives, basic philosophies, principles, etc. of both guideline documents: "diversity", "security", "privacy", "fairness", "transparency and accountability", "appropriate use", and "quality management".

Item	Content considered relevant in guidelines (main)	Examples of specific initiatives
Diversity	<p>[AI R&D Guidelines]</p> <ul style="list-style-type: none"> ○ Principles of ethics <ul style="list-style-type: none"> ▪ Respect human dignity and individual autonomy <p>[AI Utilization Guidelines]</p> <ul style="list-style-type: none"> ○ Basic philosophy <ul style="list-style-type: none"> ▪ Respect the diversity of users and include people with diverse backgrounds, values, and thinking when making use of and applying AI ○ Principles of dignity and autonomy <ul style="list-style-type: none"> ▪ Respect human dignity and individual autonomy ○ Principles of fairness <ul style="list-style-type: none"> ▪ Make arrangements so that individuals and groups are not unfairly discriminated against 	<ul style="list-style-type: none"> ▪ Emphasize the diversity of human resources, and focus on how extremely diverse humans providing input are. This includes both men and women, as well as scholars from different cultures. ▪ In addition to computer science researchers, the research institute also includes social science researchers working in psychology, fieldwork, and cognitive science. It is important to include the viewpoints of psychology and cognitive science. ▪ R&D and social implementation of safe, secure, and reliable AI requires people with various backgrounds. A team specializing in AI ethics has been formed, including researchers who have been engaged in design and social psychology.
Security	<p>[AI R&D Guidelines]</p> <ul style="list-style-type: none"> ○ Principles of security <ul style="list-style-type: none"> ▪ Remain mindful of AI system security <p>[AI Utilization Guidelines]</p> <ul style="list-style-type: none"> ○ Principles of security <ul style="list-style-type: none"> ▪ Remain mindful of AI system and AI service security ○ Principles of appropriate learning <ul style="list-style-type: none"> ➢ Organizations are expected to remain mindful of the risk of AI learning inaccurate or inappropriate data, leading to vulnerabilities in AI security 	<ul style="list-style-type: none"> ▪ With regard to AI security, the spread of AI has led to attacks against AI (duping AI, stealing AI information, etc.) and threats using AI (using AI for deception, etc.). ▪ There are four perspectives for considering the relationship between AI and security: (a) Attack using AI, (b) Attack by AI, (c) Attack to AI, and (d) Measure using AI. It is important to promote discussion that considers not only attacks against AI (c) but also other issues related to AI.

*8 Some of the initiatives summarized in "Report 2020" are included.

(4) Verification of consistency with initiative case studies presented in the hearings

Item	Content considered relevant in guidelines (main)	Examples of specific initiatives
Privacy	<p>[AI R&D Guidelines]</p> <ul style="list-style-type: none"> ○ Principles of privacy <ul style="list-style-type: none"> ▪ Make arrangements so that the privacy of users and third parties is not infringed by AI systems ○ Principles of user support <ul style="list-style-type: none"> ▪ Make arrangements so that it is possible for AI systems to support users and provide them with appropriate opportunities to make choices ○ Principles of accountability <ul style="list-style-type: none"> ▪ Strive to ensure accountability to stakeholders, including users <p>[AI Utilization Guidelines]</p> <ul style="list-style-type: none"> ○ Principles of privacy <ul style="list-style-type: none"> ▪ Make arrangements so that the privacy of users and others is not violated when utilizing an AI system or AI service ○ Principles of accountability <ul style="list-style-type: none"> ▪ Strive to ensure accountability to stakeholders 	<ul style="list-style-type: none"> ▪ Airport services are opt-in and will only be used if consent to provide facial information is obtained. ▪ Blockchain is used to protect personal information. Data can be tracked, students can choose who (companies) they want to disclose information to and the scope of this information, and companies cannot extract information without permission. ▪ Data use is opt-in, and services are provided with the consent of the user. It is important to convince citizens to use services utilizing AI.
Fairness, Equity, removal of Discrimination	<p>[AI R&D Guidelines]</p> <ul style="list-style-type: none"> ○ Principles of ethics <ul style="list-style-type: none"> ▪ Respect human dignity and individual autonomy <p>[AI Utilization Guidelines]</p> <ul style="list-style-type: none"> ○ Principles of fairness <ul style="list-style-type: none"> ▪ Remain mindful of the fact that there may be bias in decisions made by AI systems and AI services, and make arrangements so that individuals and groups are not unfairly discriminated against through the decisions of AI systems and AI services 	<ul style="list-style-type: none"> ▪ It starts with thinking about how to check for bias against sensitive attributes and what the sensitive attributes are in the first place, but making a decision is difficult. ▪ With regard to considering diverse stakeholders, for example, we check whether there are any minorities (including disabled persons and children) among stakeholders. ▪ We hold workshops with stakeholders such as designers, operators, and users involved in AI, extract and quantify values that should be emphasized from the perspective of fairness, and reflect these in the development of AI. ▪ We extract cross-biases in which nonlinear differences occur due to overlapping attributes such as pluralistic fairness, gender, or ethnicity where values differ in Japan, the U.S., and Europe; consider the proper balance with AI accuracy; and attempt to reduce bias. ▪ Facial recognition AI may become a problem from the viewpoint of racial discrimination, so we decided not to use it in public places, and announced this fact.

(4) Verification of consistency with initiative case studies presented in the hearings

Item	Content considered relevant in guidelines (main)	Examples of specific initiatives
<p>Transparency and accountability</p>	<p>[AI R&D Guidelines]</p> <ul style="list-style-type: none"> ○ Principles of transparency <ul style="list-style-type: none"> ▪ Remain mindful of the verifiability of AI system input/output and the explainability of decisions ○ Principles of accountability <ul style="list-style-type: none"> ▪ Strive to ensure accountability to stakeholders, including users <p>[AI Utilization Guidelines]</p> <ul style="list-style-type: none"> ○ Principles of transparency <ul style="list-style-type: none"> ▪ Remain mindful of the verifiability of AI system and AI service input/output, etc. and the explainability of decisions ○ Principles of accountability <ul style="list-style-type: none"> ▪ Strive to ensure accountability to stakeholders ○ Principles of appropriate use <ul style="list-style-type: none"> ➢ Expected to provide necessary information in a timely manner 	<ul style="list-style-type: none"> ▪ For medical, explainable AI will create a path for creating knowledge graphs from previous medical papers and knowledge databases and establishing medical and pharmaceutical basis, and arriving at output from input combining past medical papers and knowledge databases. ▪ If you want to make a factory more efficient, you could use AI and not necessarily need to understand it, as long as it works. ▪ When predicting demand, it is AI that suggests that some demand will grow because of such conditions. ▪ There is now a global code of ethics saying that white AI should be used in human resources and education. The use of advanced AI makes it impossible to provide ex post accountability for why HR decisions or educational evaluations are made, which eliminates accountability. ▪ Data use is opt-in, and services are provided with the consent of the user. It is important to convince citizens to use services utilizing AI. ▪ Fair decisions are required to use AI to screen housing loans, so we use XAI (Explainable AI), which can explain the basis used to make decisions.
<p>Appropriate use</p> <p><Continued on next page></p>	<p>[AI R&D Guidelines]</p> <ul style="list-style-type: none"> ○ Principles of controllability <ul style="list-style-type: none"> ➢ To ensure controllability, it is desirable to remain mindful of the effectiveness of human or other reliable AI supervision (monitoring, alerts, etc.) and response (AI system outages, network disconnections, repairs, etc.), to the extent possible in light of the characteristics of the technology employed <p>[AI Utilization Guidelines]</p> <ul style="list-style-type: none"> ○ Principles of appropriate use <ul style="list-style-type: none"> ▪ Strive to use AI systems and AI services within an appropriate scope and in an appropriate manner, based on an appropriate division of roles between humans and AI systems and between users 	<ul style="list-style-type: none"> ▪ It is dangerous to leave things entirely to AI, and AI should be used as an auxiliary tool for humans that make the final decision, especially in the fields of education and human resources. ▪ It might be best to divide roles, such as organizing routine tasks such as pre-processing applications and making plans, leaving the parts that require objective evaluation without subjectivity to AI, and using this as reference when humans conduct evaluations. ▪ AI cannot do everything when reviewing contracts, so it is not safe to leave decisions to AI. A human shall make the final decision. ▪ We're still in the experimental stages for the system with regard to AI-based automatic manuscript summary. The plan is to automatically generate a summary with the system first, and then have a human check the results, make changes, and then post it on a website or social media. We aren't planning to release AI summaries as is, but instead a final check will be performed by a human. This is our policy for any AI.

(4) Verification of consistency with initiative case studies presented in the hearings

Item	Content considered relevant in guidelines (main)	Examples of specific initiatives
Appropriate use <Continued>	<ul style="list-style-type: none"> ➢ When necessary and possible, it is expected that human decision-making will be involving in decisions made by AI, with regard to whether and how to use those decisions. In such cases, it is expected that standard examples will be taken into account and considerations will be made depending on the area of use and its applications, with regard to whether the intervention of human decision-making is required 	<ul style="list-style-type: none"> ▪ For the automatic manuscript creation system, (...) a human ultimately determines whether a manuscript created by the system requires editing and creates the manuscript. ▪ When creating subtitles using voice recognition, a human will ultimately revise any errors prior to broadcast. ▪ Diagnostic imaging AI may have a higher detection rate than specialists under certain conditions. However, instead of relying on AI to make decisions, we see AI as a means to support doctors.
Quality management (Lifecycle)	<p>[AI Utilization Guidelines]</p> <ul style="list-style-type: none"> ○ Principles of appropriate use <ul style="list-style-type: none"> ➢ Organizations are expected to provide AI software updates and AI inspections and repairs to improve AI functions and reduce risks throughout the utilization process ○ Principles of appropriate learning <ul style="list-style-type: none"> ➢ It is expected that a standard on the accuracy of decisions made by AI will be determined in advance, taking into account the scale of possible rights infringements, the frequency of rights infringements, the level of technology, the cost of maintaining accuracy, etc., as accuracy is expected to be impaired or degraded after the fact. If accuracy falls below the relevant standard, it is expected that data quality will be taken into consideration and learning will be repeated 	<ul style="list-style-type: none"> ▪ When operating AI, after the service is provided, the data that has been used until now can become unusable, such as due to the characteristics of the data changing (...). How to operate it is a major issue. ▪ The lifecycle is very important for quality management. (...) How to conduct quality management throughout the lifecycle is a very important research topic. ▪ We perform some assessments aiming for AI Ethics by Design, by confirming AI ethics prior to design. ▪ We continue to monitor and check for incidents in the market even after release. ▪ Instead of building an application and being done with it, we've established a cycle of incorporating new requirements and growing it. Within this cycle, new data is also created whenever the application is updated. It is necessary to continuously grow AI models using newly generated data (...) ▪ Even after the AI system is put into operation, we manage the behavior of the AI model by providing development and operation support infrastructures. ▪ We use development and operation support infrastructures, etc. to monitor the operation of AI, in order to introduce (or develop for introduction) a mechanism that can monitor reductions in AI accuracy and can automatically detect and repair problems. ▪ When operating and maintaining an AI system, it's difficult to guarantee it will always stay accurate. We have therefore implemented procedures for managing, monitoring, and updating AI with agreement from users. ▪ We organize the impact of AI systems on customer businesses based on the risks, and then decide on countermeasures based on these risks. For each project, we reach consensus on the amount of risk with the customer on factors such as what needs to be done quickly and when the model can be revised over time, and consider how and how often to perform monitoring.

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

○ AI is being developed, used, and implemented throughout society all over the world, and there are cases in which issues such as risks emerging or incidents occurring are identified. The scope and content of the AI R&D Guidelines and AI Utilization Guidelines were examined based on these examples.

■ Documents to be Investigated

- A total of 14 documents published by domestic and foreign government agencies were surveyed (refer to page 45)
 - Examples of developing, using, and implementing AI throughout society were organized into several categories based on examples in which issues were indicated from perspectives such as fairness (bias), privacy, and security: "biometric authentication", "autonomous driving", "image identification", "medical", "crime prediction", "recidivism prediction", "hiring", "school admissions", "performance/HR evaluation", "advertising", and "COVID-19 measures."
 - In many cases, these are covered by the values (principles) described in AI R&D Guidelines and AI Utilization Guidelines, and it is unlikely that risks will materialize far beyond the scope of both guidelines at the present time. What is important is studying measures to limit risks (preventing them from actualizing) while monitoring future trends in AI development, application, and social implementation.

	Issues, etc. identified from case studies of AI development, utilization, and social implementation (overview)
Biometric authentication	<ul style="list-style-type: none"> • In the United States, several major tech companies decided not to offer facial recognition technology to police after a black man died from being pinned to the ground by the neck by a white police officer. Some companies have also announced that they will not provide services using facial recognition technology until national laws based on human rights have been established. It is necessary to examine legal systems, etc. concerning biometric authentication technology including facial recognition. [No. 3, 4, 5, 9] • MIT scientists have revealed bias based on ethnicity and gender in the facial recognition systems of major tech companies; it is necessary to ensure fairness. [No. 12] • The privacy policy of a Chinese face app company indicates that user face data is provided free of charge and that the rights to these images belong to the company in perpetuity; it is necessary to ensure privacy. [No. 11] • An AI company in the United States has used 3D masks to bypass the facial recognition system of a major Chinese tech company; it is necessary to ensure security. [No. 11] • The Face ID (facial recognition) system can be deceived by attaching a printout to a hat; it is necessary to ensure robustness. [No. 7]
Autonomous driving	<ul style="list-style-type: none"> • Autonomous vehicles have been involved in fatal accidents. Safety must be ensured, and the locus of responsibility between AI system developers and users must be clarified. [No. 10, 11] • It is possible to deceive sign recognition technology by attaching stickers to signs; it is necessary to ensure robustness. [No. 7]

Note: The numbers indicated in "Issues, etc. identified from case studies of AI development, utilization, and social implementation (overview)" refer to the numbers in the document list on page 45.

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.

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

	Issues, etc. identified from case studies of AI development, utilization, and social implementation (overview)
Image identification	<ul style="list-style-type: none"> An image of a panda could be misrecognized as a miniature poodle, etc. due to noise to the image; it is necessary to ensure security and robustness. [No. 7]
Medical	<ul style="list-style-type: none"> AI for selecting a treatment program based on the patient's condition was not able to select an appropriate program because it tended to determine that blacks have a lower risk of disease than whites. AI for determining whether a person is a doctor or a nurse based on photos tends to judge a man as a doctor and a woman as a nurse; it is necessary to ensure data and algorithm fairness. [No. 1, 3, 13]
Crime prediction	<ul style="list-style-type: none"> Crime prediction systems use data reported by victims, which can lead to false predictions in situations where victims tend not to provide reports. Victims must be protected from crime by ensuring that they are treated fairly, whether they provide a report or not. [No. 8]
Recidivism prediction	<ul style="list-style-type: none"> Recidivism prediction systems tend to predict that blacks are at higher risk of recidivism than whites; it is necessary to ensure data and algorithm fairness. [No. 5, 8, 11]
Hiring	<ul style="list-style-type: none"> AI used by a major U.S. tech company for hiring provides discriminatory evaluations of women; it is necessary to ensure data and algorithm fairness and transparency. [No. 3, 5, 8, 10, 12, 14]
Admissions	<ul style="list-style-type: none"> The screening system for a university in France has become a black box and it is impossible to determine whether fair screening is being conducted; it is necessary to ensure transparency and fairness. [No. 8]
Performance/HR evaluation	<ul style="list-style-type: none"> The performance prediction evaluation system used by Ofqual (a qualification and testing control organization in the UK) was shown to provide unfavorable evaluations of minority students, etc.; it is necessary to ensure fairness. [No. 5] A system for evaluating the performance of teachers in the state of Texas was introduced, but the system has become a black box; it is necessary to ensure transparency and explainability. [No. 11]
Advertising	<ul style="list-style-type: none"> Personal information from users collected by a major tech company in the U.S. was provided to a business operator in the UK without the consent of users and used for political advertisements; it is necessary to ensure privacy. [No. 2, 8, 13] A major tech company in the U.S. allows advertisers to exclude individuals based on ethnicity, age, and gender from receiving advertisements for housing, employment, etc.; it is necessary to ensure fairness. [No. 6]
COVID-19 measures	<ul style="list-style-type: none"> When collecting and analyzing data using a contact verification and tracing app^{*9}, it is necessary to disclose how data is used and obtain consent from users. It is also necessary to allow users to enjoy the benefits of contact verification and tracing, regardless of factors such as ethnicity and income level. [No. 3, 14]

^{*9} Applications with such functions might not be using an AI system, but are being used to perform analysis and judgment based on collected data, and are considered to be important model cases for verification. Applications that do not directly use an AI system should therefore be included. Note that in Japan, the supported functions and specifications vary by application (COCOA does not include tracing functionality).

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

	Human-centered	Human dignity	Diversity, Inclusiveness	Sustainable society	International cooperation	Proper utilization	Education, Literacy	Human intervention, Controllability	Proper data (quality of learning data)	Collaboration among AI systems	Safety	Security	Privacy	Fairness, Equity, removal of Discrimination	Transparency, Explainability	Accountability	Robustness	Responsibility	Traceability	Monitoring, Auditing	Governance	Cost	
Biometric authentication	★	★	★			★			★			★	★	★									
Autonomous driving	★	★				★				★	★	★			★			★					
Image identification												★					★						
Medical	★	★	★			★			★				★	★	★	★							
Crime prediction	★	★	★			★			★					★	★	★							
Recidivism prediction	★	★	★			★			★				★	★	★	★							
Hiring	★	★	★			★			★				★	★	★	★							
Admissions	★	★	★			★			★					★	★	★							
Performance/HR evaluation	★	★	★			★			★				★	★	★	★							
Advertising	★	★	★			★						★	★	★									
COVID-19 measures	★	★	★			★			★			★	★	★									

Note: Values mentioned in guidelines, etc. are colored in pink.

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

	Country, etc.	Organization, etc.	Document name
1	Japan	Information-Technology Promotion Agency	• DX White Paper 2021 URL< https://www.ipa.go.jp/files/000093706.pdf >
2	Japan	Information-Technology Promotion Agency	• AI White Paper 2020 (Note: Refers to a book, so there is no URL)
3	U.S.	National Security Commission on Artificial Intelligence	• The Final Report URL< https://www.nscai.gov/wp-content/uploads/2021/03/Full-Report-Digital-1.pdf >
4	U.S.	Stanford University Institute for Human-Centered Artificial Intelligence	• 2021 AI Index Report URL< https://aiindex.stanford.edu/report/ >
5	UK	Centre for Data Ethics and Innovation	• Review into bias in algorithmic decision-making URL< https://www.gov.uk/government/publications/cdei-publishes-review-into-bias-in-algorithmic-decision-making/main-report-cdei-review-into-bias-in-algorithmic-decision-making#preface >
6	UK	Centre for Data Ethics and Innovation	• Online targeting: Final report and recommendations URL< https://www.gov.uk/government/publications/cdei-review-of-online-targeting/online-targeting-final-report-and-recommendations >
7	Germany	Bundesamt für Sicherheit in der Informationstechnik	• Towards Auditable AI Systems URL< https://www.bsi.bund.de/SharedDocs/Downloads/EN/BSI/KI/Towards_Auditable_AI_Systems.pdf?__blob=publicationFile&v=4 >
8	France	Commission Nationale de l'Informatique et des Libertés	• HOW CAN HUMANS KEEP THE UPPER HAND? URL< https://www.cnil.fr/sites/default/files/atoms/files/cnil_rapport_ai_gb_web.pdf >
9	EU	European Commission	• Proposal for a REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL URL< https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:52021PC0206&from=EN >
10	EU	European Commission	White Paper On Artificial Intelligence URL< https://ec.europa.eu/info/sites/default/files/commission-white-paper-artificial-intelligence-feb2020_en.pdf >
11	China	China Academy of Information and Communications Technology	• White Paper on Trustworthy Artificial Intelligence URL< http://www.caict.ac.cn/english/research/whitepapers/202110/P020211014399666967457.pdf >
12	International organizations	United Nations Educational, Scientific and Cultural Organization (UNESCO)	• ARTIFICIAL INTELLIGENCE and GENDER EQUALITY URL< https://unesdoc.unesco.org/ark:/48223/pf0000374174 >
13	International organizations	World Health Organization (WHO)	• ETHICS AND GOVERNANCE OF ARTIFICIAL INTELLIGENCE FOR HEALTH URL< https://apps.who.int/iris/rest/bitstreams/1352854/retrieve >
14	International organizations	World Economic Forum (WEF)	• The AI Governance Journey: Development and Opportunities URL< https://www3.weforum.org/docs/WEF_The%20AI_Governance_Journey_Development_and_Opportunities_2021.pdf >

(6) 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^{*10} 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.

^{*10} These differences should be taken into consideration when examining guideline revisions.

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.

(These two perspectives are not necessarily contradictory and can be compatible with each other.)

- 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.

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.

Discussion points on guideline revisions

- 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 than 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?

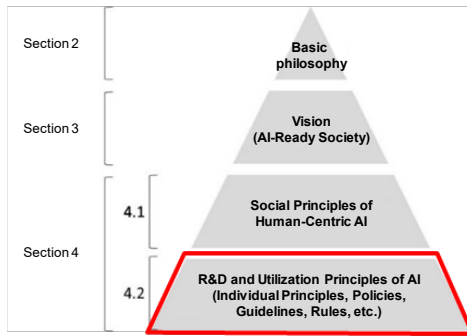
Note: It is important that revision examinations ensure the guidelines continue to be technologically neutral, to not impede innovation, and to not place excessive costs on developers or users.

[Reference] Social Principles of Human-Centric AI

Social Principles of Human-Centric AI

- The Human-Centric Principle
- The Principle of Education/Literacy
- The Principle of Privacy Protection
- The Principle of Ensuring Security
- The Principle of Fair Competition
- The Principle of Fairness, Accountability, and Transparency
- The Principle of Innovation

From "Social Principles of Human-Centric AI" (decided by the Integrated Innovation Strategy Promotion Council in March 2019)



Council for Social Principles of Human-centric AI (from May 2018)

- ... "AI social principles" that must be taken into consideration by society (especially by national legislative and government organizations)
- ... Developers and business operators must establish R&D and Utilization Principles of AI based on the basic philosophy and AI social principles and comply with it

Each developer and business operator is expected to establish R&D and Utilization Principles of AI

Guidelines required to serve as reference

Note: Investigated as the "AI Network Investigation Council" from Feb. through Sept. 2016

MIC initiatives

The Conference toward AI Network Society (from Oct. 2016)

AI R&D Guidelines*

- Points developers must take into consideration, and explanations

* "Draft AI R&D Guidelines for International Discussion"

AI Utilization Guidelines

- Points users must take into consideration, and explanations

Shared with relevant ministries and agencies, and then provided to developers and business operators. Voluntary response supported.