Tentative

Translation

Report 2020

Toward “the safe, secure, and trustworthy implementation of AI in society”

21 July 2020

The Conference toward AI Network Society

Table of Contents

Introduction....1

Chapter 1 Recent Trends of AI Networking....5

1. AI and COVID-19 control measures....5
2. Trends of discussions in Japan, overseas and international organizations.... 12

Chapter 2 Prospect of an Ecosystem Formed with the Progress of AI Networking....18

1. Background and analysis policy.... 18
2. AI Utilization Prospect.... 20
3. Case studies on the social implementation of AI.... 23

Chapter 3 Initiatives by Developers and AI Service Providers....25

1. Starting point for discussion.... 25
2. ABEJA Co., Ltd. (approaches in "Ethical Approach to AI (EAA)").... 26
3. FUJITSU LIMITED ("AI Governance for AI Developers").... 28
4. IBM Japan ("IBM's Approach to AI").... 29
5. NTT DATA Corporation ("NTT DATA Group AI Governance Initiatives").... 33
6. Oki Electric Industry Co., Ltd. ("Improvement of Foundations for the implementation of AI -Establishment of "OKI Group AI Principles"-").... 36
7. Microsoft ("Business, Responsibilities, and Challenges Surrounding AI -Ethics and the Potential of AI-").... 39
8. Anonymous ("Report on Support Tools for the use of AI by Private Enterprise Volunteers").... 43
9. Summary.... 45

Chapter 4 Initiatives by Business Users....49

1. Starting point for discussion.... 49
2. Sumitomo Mitsui Financial Group, Inc. ("The SMBC Group's Digitalization Initiatives").... 49
3. Tokyo ("Tokyo Metropolitan ICT-related measures").... 54
4. Yamaha Corporation ("Yamaha's AI Singing Synthesis -Initiatives to Revive Hibari Misora-").... 56
5. Anonymous.... 61
6. Mr. Hiroyuki Sanbe, a member of the Conference ("Rikunabi Case Reviewed from the Perspective of AI Utilization").... 63
7. Summary.... 71

Chapter 5 Initiatives for Consumer Users....73

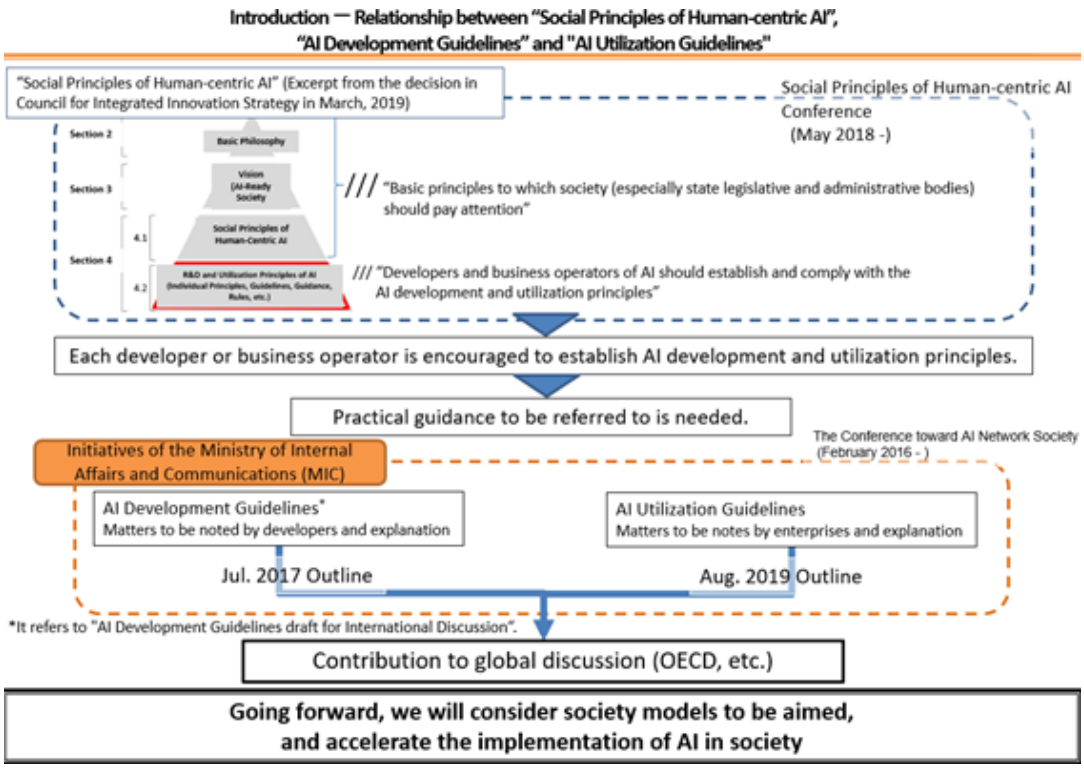
1. Starting point for discussion.... 73

2. Initiatives for Consumer Users 74
3. Initiatives for the elderly and persons with disabilities 76
4. Summary 79
Chapter 6 Initiatives for Security 81
1. Starting point for discussion 81
2. Specified non-profit organization Japan Network Security Association (JNSA) 82
3. Supplementary theory 86
4. Summary 87
Chapter 7 Initiatives Related to Insurance 88
1. Starting point for discussion 88
2. Tokio Marine & Nichido Fire Insurance Co., Ltd. (About "Insurance to help spread AI") 88
3. Sompo Japan Insurance Inc. ("Utilization of Insurance in Smart Factories") 93
4. Summary 96
In Place of Conclusion 97
(Reference) "Future Issues" as listed in Report 2019 99

【Appendix 1】Members of the Conference toward AI Network Society
Members of Committee on AI Governance
【Appendix 2】Prospect of the Ecosystem on AI Utilization
【Appendix 2-1】Prospect of AI Utilization Scenes
【Appendix 2-2】Case Studies on Social Implementation of AI

Introduction

The Conference toward AI Network Society (hereinafter referred to as "the Conference") issued a report entitled Report 2019 including the "AI Utilization Guidelines" in August 2019. The "AI Utilization Guidelines" summarizes items that AI users are expected to take into consideration. This is a counterpart to the "AI R&D Guidelines"[1], which was issued at the Conference held in July 2017, and summarizes items that developers are expected to take into consideration. In addition, developers and AI service providers are expected to formulate and comply with the AI development and utilization principles based on the basic ideas and AI social principles outlined in the "Social Principles of Human-Centric AI" (decision made by Integrated Innovation Strategy Promotion Council in March 2019). In this context, both Guidelines are intended to serve as a reference for developers and AI service providers when they formulate their own AI principles, etc.[2]



In the first place, the purpose of developers and AI service providers formulating their own principles is to eliminate people's anxiety about AI and promote efforts to build trust. While there are high expectations for AI utilization, there are concerns that people's anxiety about AI might hinder the

[1] This refers to the "Draft AI R&D Guidelines for International Discussions", which summarizes the items that AI developers are expected to take into account in the Report 2017 summarized by the Conference held in July 2017.
[2] Refer to 'Introduction' in the Report 2019.

promotion of development and utilization of AI, as well as the excellent progress of AI networking. In light of the original purpose of the developers and AI service providers formulating their own principles, both Guidelines are expected to serve as a reference tool for actively promoting "safe, secure, and trustworthy social implementation of AI".

Based on this recognition, the secretariat of the Conference made efforts to disseminate both Guidelines[3], and in line with the contents described in **"Chapter 2: Concept of formulating the AI Utilization Guidelines, 4. Future Development"** and **"Chapter 3: Future Issues"** in the Report 2019, exchanged opinions on "safe, secure, and trustworthy social implementation of AI" (including the formulation of principles) with various stakeholders. While considering exchanged opinions, the issues necessary to be discussed for "safe, secure, and trustworthy social implementation of AI" were presented. Interviews by the chairperson of the Conference (hereinafter referred to as "the interviews") were then carried out with people who were enthusiastic in the AI field, and a series of open and lively discussions took place. The main points of discussion were as follows:

(1) About developers and AI service providers (hereinafter referred to as "developers, etc.")

a. When developers, etc. are formulating their own AI principles[4] with their own characteristics, "safety and security" and "trustworthiness" are considered to be important. Therefore, from the viewpoint of promoting the AI principle formulation by developers, etc., it is necessary to consider:

- what is the significance of its formulation? and
- how can the AI principle be applied to the development and utilization of safe, secure, and trustworthy AI?

b. What kind of governance system (self-inspection/self-evaluation mechanism and external evaluation mechanism) can be considered to develop and utilize AI which is safe, secure, and trustworthy (including management of AI principles)?

(2) About business users

With the fact that there are many different types of business users in mind, what are the challenges in promoting the AI utilization? What kind of measures can be taken to solve the problem?

(3) About consumer users

Consumer users are described in the AI Utilization Guidelines as a reference only. However, from

[3] As part of the dissemination activities, a special feature article: "How to use AI with Peace of Mind? AI Utilization Guidelines" was published in the December 2019 issue of the Ministry of Internal Affairs and Communications (MIC) public relations magazine. The article is available on the website at the following URL. <https://www.soumu.go.jp/main_content/000656829.pdf>

[4] Different AI service providers use different terms, such as AI principles, AI commitment and AI guidelines, however the "AI principles" is used as a generic term in this report.

the viewpoint of actively promoting the “safe, secure and trustworthy social implementation of AI”, it is necessary to take measures to enable consumer users to use AI with peace of mind and enjoy its benefits. Therefore, we need to consider these questions:

a. How can we proceed with the efforts related to consumer users?
b. It is one of the essential efforts to realize a human-centric AI society where older people and people with disabilities as consumer users can utilize AI to eliminate the inconvenience associated with aging or disability so that everyone can achieve self-fulfillment equally. So, what can be considered as measures necessary for the social implementation of AI that is safe, secure, and trustworthy for the elderly and people with disabilities?

(4) Creating an environment for the safe, secure, and trustworthy social implementation of AI.

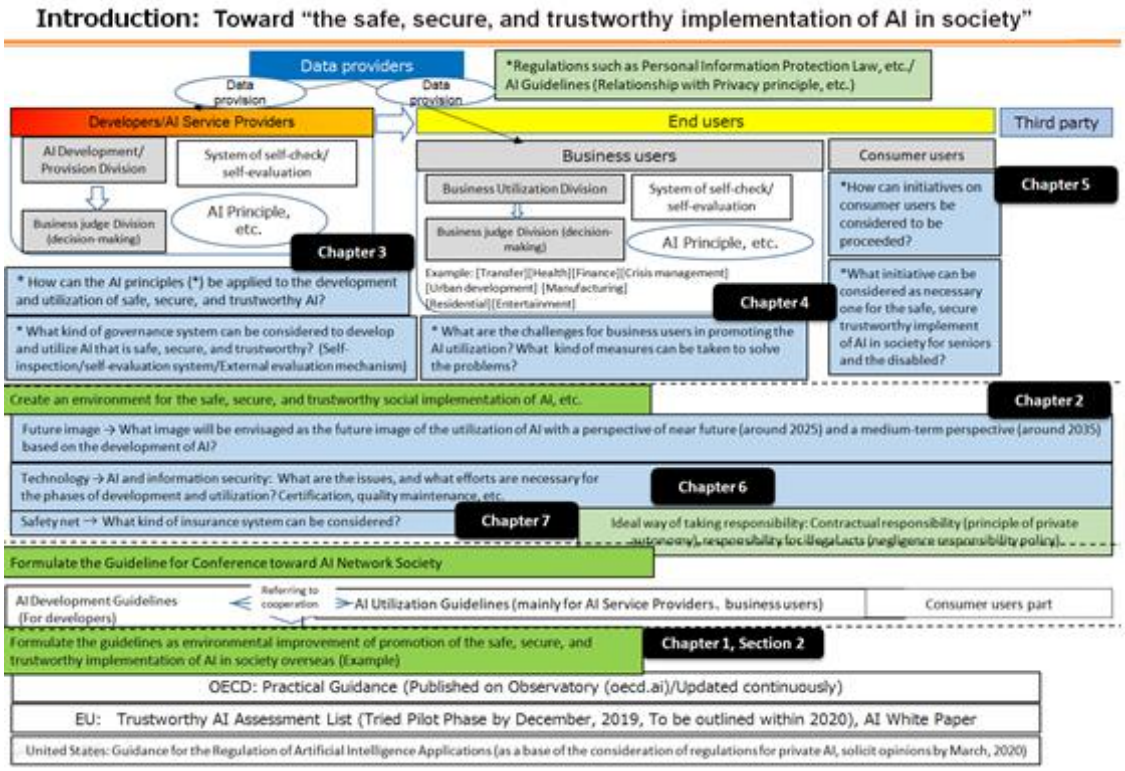
Apart from the viewpoint of the efforts made by related entities, to create the environment:

a. From a technology point of view, regarding “AI and information security”, what are the issues, and what efforts are necessary for the phases of AI development and utilization?
b. From a safety net point of view, what kind of insurance system can be considered?

The above issues were the focus of the interviews. With the permission of the interviewees, and within its scope, the contents of the interviews and discussions have been summarized and compiled as necessary.

For promoting the “safe, secure, and trustworthy social implementation of AI”, it is beneficial to analyze and present a scenario of the future social image of the AI social implementation to share a concrete image. Thus, while keeping the progress of AI utilization in mind, sharing a concrete image was attempted by analyzing and presenting the future image of AI utilization from the perspective of the near future (around 2025) and the medium-term future (around 2035).

In addition, international collaboration is essential when discussing efforts for the “safe, secure and trustworthy social implementation of AI” due to the nature of AI networking. Therefore, it is important to find out about the efforts being made by the Organization for Economic Co-operation and Development (OECD), the European Union (EU), and the United States, etc., and to take actions as necessary. From this viewpoint, the interviewees were also asked about global trends as well as trends in international discussions, and the information necessary to discuss efforts for “safe, secure and trustworthy social implementation of AI” was summarized.



The concept of the Report 2020 is as described above. It is based on the interviews regarding specific and enthusiastic efforts, etc. and does not necessarily cover all the efforts necessary for the "safe, secure and trustworthy social implementation of AI". Nevertheless, this information is expected to be very beneficial for stakeholders who are sincerely considering the development and utilization of AI, if these specific and enthusiastic efforts are widely introduced and shared as helpful efforts for the "safe, secure and trustworthy social implementation of AI". It is hoped that the Report 2020 will become widely used as a reference and that the stakeholders will be able to make necessary efforts to contribute to the promotion of "safe, secure and trustworthy social implementation of AI" in Japan.

Chapter 1 Recent Trends of AI Networking

This chapter provides an overview of the AI networking trends mainly after the publication of the Report 2019 (in particular, trends relating to the "safe, secure and trustworthy implementation of AI in society")[5].

1. AI and COVID-19[6] control measures

Various efforts are now being made to address the expansion of COVID-19, and many of these are related to AI. For example, the European AI Alliance introduced their initiative entitled "Join the AI-ROBOTICS vs COVID-19 initiative"[7]. The OECD also presented a classification[8] of base technologies and applied fields of AI that can be utilized to address COVID-19. This section provides an overview of the different initiatives being taken in each country/region[9] based on said classification as shown in the figure below[10].

---

[5] For the trends in AI networking prior to the publication of the "Report 2019", see Chapter 1 of the Report 2019.

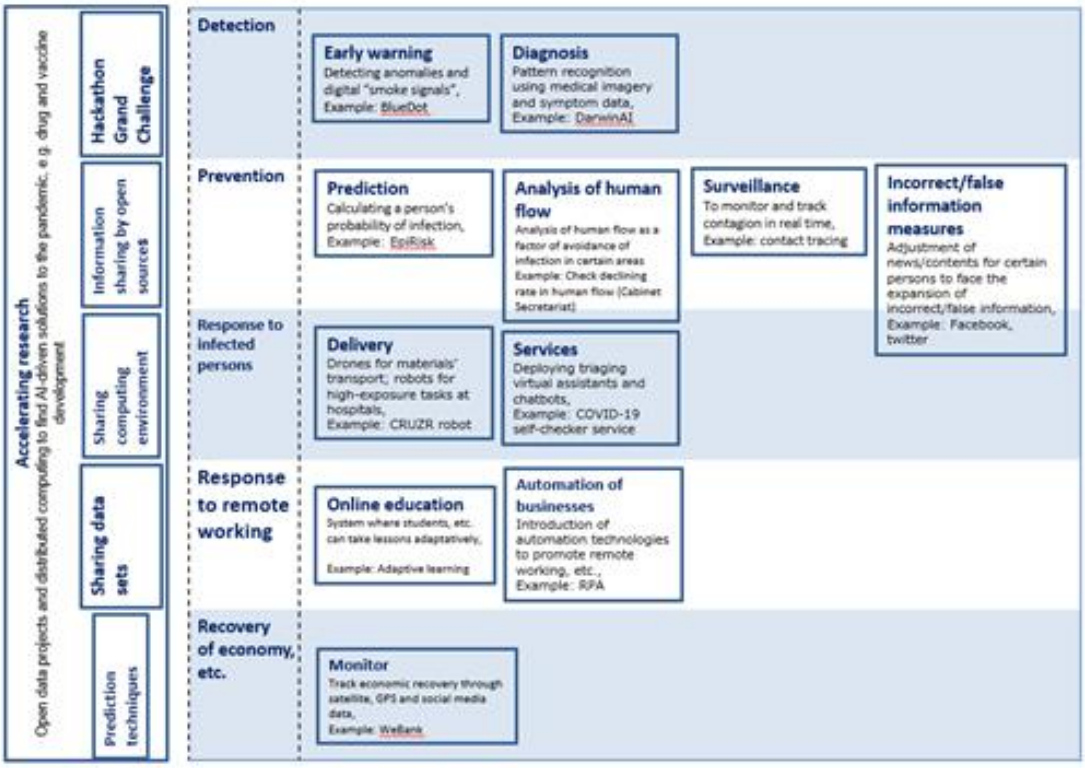
[6] COVID-19 means a novel coronavirus infectious disease. "COVID-19" was named by the World Health Organization (WHO) in February 2020 as an abbreviation for the coronavirus disease that occurred in 2019
<https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200211-sitrep-22-ncov.pdf>

[7] Join the AI-ROBOTICS vs COVID-19 initiative of the European AI Alliance
<https://ec.europa.eu/digital-single-market/en/news/join-ai-robotics-vs-covid-19-initiative-european-ai-alliance>

[8] This classification is a tentative translation (from English into Japanese) of the content in the following OECD document. This translation was not created by the OECD and should not be considered an official OECD translation. The OECD shall not be liable for any content or error in this translation.
OECD/Using artificial intelligence to help combat COVID-19,
<https://read.oecd-ilibrary.org/view/?ref=130_130771-3jtyra9uoh>

[9] For the rest, see below as examples.
Gartner: Top five areas CIOs can use AI to combat COVID-19
<https://remoteworkertech.asia/story/gartner-top-five-areas-cios-can-use-ai-to-combat-covid-19>
Ledge.ai: AI technology to fight against COVID-19 – Introduction of 23 cases including detection of infected patients and behavioral analysis<https://ledge.ai/aicompany-fighting-covid19/>
AInow: [Case Study!] How AI can face the spread of COVID-19
<https://ainow.ai/2020/05/14/222586/>
Anti-Covid-19 Tech Team (Second time)
List of current projects by the Tech Team
<https://cio.go.jp/sites/default/files/uploads/documents/techteam_20200421_01.pdf>
Tech Team for COVID-19 infection control measures
<https://cio.go.jp/techteam>
Open collaboration on COVID-19
<https://github.blog/jp/2020-04-02-open-collaboration-on-covid-19/>

[10] The efforts listed here, including the classification, are examples only and are not exhaustive.



(1) Accelerating research

In addressing COVID-19 by utilizing AI, as described in 2) to 6) below, possible initiatives and measures to accelerate research are described in the following paragraphs. Regarding information on the initiatives of universities and public institutions, mainly research institutes in Japan, refer to the AI Japan R&D Network website for a detailed summary[11].

a. Predictive technologies (structure and drug discovery, etc.)

AI is being used to identify structures associated with viruses and to accelerate drug discovery. For example, DeepMind reported that they predicted the protein structures associated with COVID-19 using their latest Alphafold system[12].

b. (Open) dataset sharing

There are government initiatives to share COVID-19 datasets in collaboration with the private sector, universities and other organizations. For example, the U.S. government is running the COVID-

[11] AI Japan R&D Network: AI-enabled research activities for COVID-19 <https://www.ai-japan.go.jp/COVID19>

[12] DeepMind: Computational predictions of protein structures associated with COVID-19 <https://deepmind.com/research/open-source/computational-predictions-of-protein-structures-associated-with-COVID-19>

19 Open Research Dataset Challenge on Kaggle[13]. The EU has also set up a portal to hold datasets[14]. In Japan, the Cabinet Secretariat[15], local governments, and other organizations are making various types of data available to the public. Furthermore, for the data of each local government, there is a movement to define items of data that enables analyses and visualizations using AI in a unified manner[16].

c. Shared computational environment

In order to enable researchers to perform calculations using AI, etc., there are initiatives to share computational environments. There is a distributed computing project by Folding@home[17] overseas, and another project to provide a computational environment free of charge, such as COVID-19 High Performance Computing Consortium[18]. In Japan, RIKEN Center for Computational Science[19], National Institute of Advanced Information and Communications Technology (AIST)[20], and Research Organization for Information Science & Technology (RIST)[21], etc. also provide computational environments free of charge.

d. Information sharing by open source

There is a movement to share source code, etc. of various applications as described below. For example, DarwinAI (Canada) provides an open source of the mechanism for COVID-19 detection by chest radiography[22]. In Japan, the Tokyo Metropolitan Government has released the source code of the COVID-19 Information Website on GitHub[23]. By sharing the data as open source, it is expected

---

13 Kaggle: COVID-19 Open Research Dataset Challenge
<https://www.kaggle.com/allen-institute-for-ai/CORD-19-research-challenge>

14 EU data portal <https://www.europeandataportal.eu/en/about/european-data-portal>

15 Matters in regard to COVID-19 infection control: status of the healthcare system of medical institutions in Japan (G-MIS data) and open data have been released to the public (beta version).
<https://cio.go.jp/hosp_monitoring_c19>

16 Data release support for the COVID-19 infection control websites
<https://www.code4japan.org/activity/stopcovid1>

17 <https://foldingathome.org/>

18 The COVID-19 High Performance Computing Consortium
<https://covid19-hpc-consortium.org/>

19 RIKEN: Research and development aimed at COVID-19 infection control will be carried out on "Fugaku".
<https://www.r-ccs.riken.jp/library/topics/fugaku-coronavirus.html>

20 AIST: ABCI, a cloud-based computational system for AI, is offered free of charge to respond to COVID-19
<https://abci.ai/ja/link/covid-19.html>

21 RIST: HPCI supercomputer resources is offered free of charge to support "Research on the response to COVID-19"
<https://www.hpci-office.jp/materials/press_20200407.pdf>

22 COVID-Net Open Source Initiative
<https://github.com/lindawangg/COVID-Net/>

23 <https://github.com/tokyo-metropolitan-gov/covid19>

that the data format will be standardized as described in (b) of this section. The open-sourcing of products (hardware) such as face masks and shields is also being promoted, contributing to the sharing and standardization of product specifications.

e. Hackathons and Grand Challenges

Hackathons and Grand Challenges focused on COVID-19 infection control have been held both in Japan and abroad. AI-focused examples overseas are the CoronaHack - AI vs. Covid-19 in London[24], The COVID-19 Detect and Protect Challenge by the United Nations Development Programme (UNDP)[25], the aforementioned COVID-19 Open Research Dataset COVID-19 Open Research Dataset Challenge, and SPACE APPS COVID-19 CHALLENGE hosted by NASA[26]. There are also examples of initiative such as the INNO-vation[27] and a project by Mitou[28] in Japan.

(2) Detection

a. Early warning

There was a movement for using AI to analyze if there is a warning sign of the infectious disease outbreak. The early warning system using AI developed by BlueDot (Canada)[29] successfully detected epidemiological patterns by data-mining major news, online contents and other information channels in multiple languages, and warned of the outbreak as early as December 2019.

b. Diagnostic support

Rapid diagnosis is a key to limit the spread of inflectional diseases. The application of AI to imaging and clinical case data can be a useful aid for the rapid diagnosis of COVID-19. The aforementioned DarwinAI initiative to detect COVID-19 on chest radiography is an example.

(3) Prevention of infection

a. Prediction

AI is useful for identifying the chain of infection and monitoring the wider economic impact. For example, Johns Hopkins University developed an interactive dashboard that can track the spread of

[24] CoronaHack - AI vs. Covid-19 - Hackathon in London
<https://www.hackathon.com/event/coronahack---ai-vs-covid-19-99337559314>
[25] <https://www.covid19detectprotect.org/>
A project calling on open source hardware that makes COVID-19 detect and protect possible
[26] NASA: SPACE APPS COVID-19 CHALLENGE
<https://www.spaceappschallenge.org/>
This is a special version of the annual Grand Challenge event, and JAXA has also contributed data.
[27] INNO-vation
<https://www.inno.go.jp/>
[28] Mitou 2nd Stage AI Frontier Program (AI utilization special program for After/With COVID-19).
[29] BlueDot: Infectious disease surveillance automated and personalized to what's relevant for you.
<https://bluedot.global/products/>

the virus through live news and real-time data regarding confirmed COVID-19 cases, recoveries and deaths[30].

Northeastern University (U.S.) also developed a tool called EpiRisk.[31]-[32] This estimates the probability of spreading the disease to other areas of the world through travel via infected individuals.

b. People flow analysis

In Japan, in order to find out about the increase or decrease in the number of people in a particular area, the visualization of people flow data mainly in urban areas is regarded as being important[33]. For example, the Cabinet Secretariat's COVID-19 Information and Resources website[34] highlights this as "the decrease rate in people flow".

c. Contact confirmation

Many countries and regions are utilizing contact confirming applications to track contact with infected people. However, the methods vary from the contact tracking using Bluetooth and location data, etc., to the infected persons' data management (central server-based or distributed data management, etc.)[35]. The purpose also varies, for example, limiting or isolating people according to the degree of contact, or assisting public health authorities to identify people who have been in close contact, or preventing the spread of infection by changing the behavior of users who received a notification (the authorities have not grasped of the details)[36]. In order for the application to work

---

[30] <https://github.com/CSSEGISandData/COVID-19>
(Dashboard)
<https://www.arcgis.com/apps/opsdashboard/index.html#/bda7594740fd40299423467b48e9ecf6>

[31] <http://epirisk.net/>

[32] The U.S. Northeastern University: EpiRisk
<https://xtech.nikkei.com/atcl/nxt/column/18/051900003/>

[33] "Statistical data" for the novel coronavirus infection control: an overview of the differences between the companies that offer it.
<https://xtech.nikkei.com/atcl/nxt/column/18/01304/051900003/>

[34] Cabinet Secretariat: "the decrease rate in people flow"
<https://corona.go.jp/#area-transition>

[35] Although not directly related to AI, in response to this trend, both Google and Apple have published Exposure Notification APIs for creating the said applications on smartphones, and as of 20 May 2020, it is reported that 22 countries and several states in the U.S. are planning to use them.
Apple's announcement:
<https://developer.apple.com/documentation/exposurenotification/building_an_app_to_notify_users_of_covid-19_exposure>
Google official blog:
<https://blog.google/inside-google/company-announcements/apple-google-exposure-notification-api-launches/>
In June of the same year, Japan also released a contact-confirming application that utilizes the above API:
(COCOA) COVID-19 Contact-Confirming Application (Ministry of Health, Labour and Welfare website):
< https://www.mhlw.go.jp/stf/seisakunitsuite/bunya/cocoa_00138.html >

[36] COVID-19 Infection Control Tech Team: Trends in countries regarding the introduction of contact-confirmation applications (3rd reference document 1-2)

effectively, the key should be to increase the number of installations, and implementing measures to achieve this is also important. In particular, there have been discussions about the need for taking security and privacy into consideration[37], and bills are being submitted to deal with this matter[38].

In Japan, various discussions are also taking place from the viewpoint of balancing the app utilization with the safety, security and privacy of citizens[39].

d. Measures against misinformation and disinformation[40]

The spread of misinformation and disinformation (the 'infodemic' of COVID-19[41]) has become a major problem[42], and in order to combat this, various social networking sites such as Twitter[43] and Facebook[44] are utilizing AI to detect and remove problematic materials from their platforms. Refer to **Chapter 6.4 (2)** as this is also associated with the issue.

(4) Management of infected persons and others

<https://cio.go.jp/sites/default/files/uploads/documents/techteam_20200508_02.pdf>

[37] For example, Coronavirus: a common approach for safe and efficient mobile tracing apps across the EU :
<https://ec.europa.eu/digital-single-market/en/news/coronavirus-common-approach-safe-and-efficient-mobile-tracing-apps-across-eu>

[38] For example, COVID-19 Consumer Protection Data Act of 2020
<https://www.commerce.senate.gov/services/files/A377AEEB-464E-4D5E-BFB8-11003149B6E0>

[39] For example, as follows:
Kishimoto and Kudo (Research Center on Ethical, Legal and Social Issues, Osaka University), "Ten Perspectives and Three Recommendations on Contact-confirming Applications and ELSI, Ver. 0.9"
<https://elsi.osaka-u.ac.jp/research_category/elsi_note/>
Mamoriai note (Referring to the status of discussions on how the applications should be etc., through the consideration of the contact-confirming application "Mamoriai Japan" by Civic tech).
<https://note.com/hal_sk/m/m53cefeea1340>In accordance with the "final report of the study group on platform services" published by the Ministry of Internal Affairs and Communications (MIC) in February 2020, "misinformation" refers to simply incorrect information, and "disinformation" refers to information with false intention. Referenced at the following URL:
< https://www.soumu.go.jp/main_content/000668595.pdf >

[40] In accordance with the "final report of the study group on platform services" published by the Ministry of Internal Affairs and Communications (MIC) in February 2020, "misinformation" refers to simply incorrect information, and "disinformation" refers to information with false intention. Referenced at the following URL:
< https://www.soumu.go.jp/main_content/000668595.pdf >

[41] A coined word consisting of "information" and "epidemic". It refers to the phenomenon of the spread of uncertain information via social networking services.

[42] The MIC reported on the reality of the problematic COVID-19 "infodemics" in the following report: "Information distribution survey on COVID-19":
< https://www.soumu.go.jp/menu_news/s-news/01kiban18_01000082.html >

[43] Twitter: Coronavirus: Staying safe and informed on Twitter
<https://blog.twitter.com/en_us/topics/company/2020/covid-19.html>

[44] Facebook: An Update on Our Work to Keep People Informed and Limit Misinformation About COVID-19
<https://about.fb.com/news/2020/04/covid-19-misinfo-update/>

a. Delivery of goods (using robots, etc.)

In order to meet the urgent needs of hospitals and other institutions, various types of deliveries are being made using robots, etc. For example, in China, drones and robots are used to deliver food and medicine, as well as for aerial spray and disinfection[45]. In other countries, robots are also being used to supplement the shortage of medical staff and to prevent hospital-acquired infections, for example, disinfecting hospitals or screening patients at the entrance of hospitals[46].

b. Services

Virtual assistants and chatbots are increasingly being introduced around the world to support healthcare providers. These tools are useful in sorting out patients according to the presence of symptoms. For example, Centers for Disease Control and Prevention in the U.S. and Microsoft have developed a coronavirus self-checker[47] that allows users to do COVID-19 self-assessment and propose a set of possible actions.

(5) Management of the online environment

a. Online education

As part of the promotion of remote learning, adaptive learning is being introduced to allow students to take classes adaptively according to their own learning progress[48].

b. Automation of work

As a means of promoting telework etc., automation of various tasks, such as the use of RPA and chatbots, is being promoted.

(6) Recovery of the economy, etc.

a. People flow analysis

There is a movement to monitor economic recovery through analysis using AI. For example, WeBank in China analyzes satellite photos, SNS and other data (such as Google's Community Mobility Report), which is considered to be useful for monitoring economic crisis and recovery[49].

---

[45] 3 ways China is using drones to fight coronavirus
<https://www.weforum.org/agenda/2020/03/three-ways-china-is-using-drones-to-fight-coronavirus/>

[46] Carrying medicines and disinfecting hospital rooms: robots fight infectious diseases
<https://project.nikkeibp.co.jp/mirakoto/atcl/robotics/h_vol35/>

[47] Testing for COVID-19
<https://www.cdc.gov/coronavirus/2019-ncov/symptoms-testing/testing.html>

[48] What is adaptive learning? An in-depth explanation of the main learning tools, their advantages and disadvantages!
<https://coeteco.jp/articles/10634>

[49] IEEE: Satellites and AI Monitor Chinese Economy's Reaction to Coronavirus
<https://spectrum.ieee.org/view-from-the-valley/artificial-intelligence/machine-learning/satellites-and-ai-

As described later, ABEJA also presented a study about the impact of COVID-19 on sales and customer numbers by using their own AI-powered retail analysis system[50].
Through such analysis, it may be possible to develop an early warning system for future outbreaks of other infections (in addition to serving as an indicator of economic recovery).

2. Trends of discussions in Japan, overseas and international organizations

(1) Ministry of Education, Culture, Sports, Science and Technology (MEXT): Trusted Quality AI

Based on the importance of developing "Trusted Quality AI" as stated in the "AI Strategy 2019"[51], MEXT decided to set the development of fundamental technologies to ensure the future progress and trustworthiness of AI as their strategic goal for FY 2020, and announced this on 9 March 2020[52].

Regarding deep learning, which is the core of current AI technology, various measures are urgently needed. These measures started being discussed by a consortium, etc. in the industry. However, it is necessary to develop and innovate AI technology itself beyond its limitations, and also to ensure its fundamental trustworthiness to meet the demands of society. Therefore, the strategic goal is being set to promote research and development for the creation of "Trusted Quality AI" based on the "Human-Centered AI Social Principles". Specifically, the following three goals are to be achieved:

- Creating new technologies that overcome the limitations of current AI technologies
- Creating technologies that ensure the trustworthiness and safety of AI systems
- Ensuring data trustworthiness and creating technologies to support decision-making and consensus building

(2) The EU

The European Commission, the executive branch of the EU, published a "White Paper on Artificial Intelligence" for shaping Europe's digital future on 19 February 2020[53]. The paper presents policy options that enable the trustworthy and safe development of AI, while promoting its widespread adoption in terms of excellence and trust.

From the viewpoint of excellence, it describes about mobilization of resources across the value

---

monitor-chinese-economys-reaction-to-coronavirus>

[50] 【ABEJA Insight for Retail】 Signs of recovery in clothing and general stores from the second week of March? Our second investigation into the COVID-19 impact
<https://abejainc.com/ja/news/article/20200325-2681>

[51] "AI Strategy 2019" (decision made by Integrated Innovation Strategy Promotion Council in June 2019).
<https://www.kantei.go.jp/jp/singi/ai_senryaku/pdf/aistratagy2019.pdf>

[52] Referenced at the following URL:
<https://www.mext.go.jp/b_menu/houdou/2020/mext_00487.html>

[53] "WHITE PAPER On Artificial Intelligence - A European approach to excellence and trust"
Referenced at the following URL: <https://ec.europa.eu/info/sites/info/files/commission-white-paper-artificial-intelligence-feb2020_en.pdf>

chain and creation of incentives to accelerate the deployment of AI by SMEs and others. It also mentions about collaboration with member states, strengthening the efforts of the research and innovation community (e.g. establishing testing facilities), skills, and collaboration between SMEs, and partnership with the private sector.

On the other hand, from the viewpoint of trust, it clarifies the scope of high-risk AI applications and suggests how future regulations should be made according to the risk level, so that they can deal with high-risk AI systems without placing excessive burden on low-risk systems. It also mentions about ensuring consumer protection, addressing unfair commercial practices, and continuing to apply strict EU rules to protect personal data and privacy.

For certain applications, such as for healthcare, transport and the public sector, which are considered to be high-risk, AI systems need to be transparent, traceable and follow the conditions set out in the EU's Ethics Guidelines for Trustworthy Artificial Intelligence, for example, ensuring human oversight (and in addition, training to ensure proper functioning and the use of unbiased data, etc.). In particular, the paper mentions that they will initiate a wide-ranging discussion on the use of face recognition for the purpose of remote biometric identification, which is currently prohibited in principle in the EU, except under certain conditions, and can only be used if justified as an exception under the EU or national law.

In addition, for low-risk AI systems, they will consider a voluntary labelling scheme through objective benchmarks in the EU, that is to say a certification system, etc., in order to build trust.

They invite comments on the proposals set out in the paper until 19 May 2020[54]. In response to this call for opinions, voluntary members of the Conference and the Committee on AI Governance have submitted their opinions[55].

The Council of the European Union released its statement of support for the digital strategy including the "AI White Paper" mentioned above on 9 June 2020[56].

(3) The U.S.

The U.S. Office of Management and Budget (OMB) published a request for comments on "Draft Memorandum to the Heads of Executive Departments and Agencies: Guidance for Regulation of Artificial Intelligence Applications" on 13 January 2020[57].

---

[54] The deadline has been extended to 14 June.

[55] See the following web page for the submitted opinions.
<https://www.soumu.go.jp/main_sosiki/kenkyu/ai_network/02iicp01_04000232.html>

[56] Referenced at the following URL:
< https://www.consilium.europa.eu/en/press/press-releases/2020/06/09/shaping-europe-s-digital-future-council-adopts-conclusions/ >

[57] Draft to the Heads of Executive Departments and Agencies, "Guidance for Regulation of Artificial Intelligence Applications"
Referenced at the following URL:
<https://www.federalregister.gov/documents/2020/01/13/2020-00261/request-for-comments-on-a-draft-

This draft memorandum describes the principles that executive departments and agencies should follow when developing rules for the introduction of AI into the private sector, in response to the Executive Order (AI Initiative[58]) issued in February 2019.

The ten principles are: 1) public trust in AI, 2) public participation in the rulemaking process, 3) scientific integrity and information quality, 4) risk assessment and management, 5) cost-effectiveness analysis, 6) flexibility, 7) fairness and non-discrimination, 8) disclosure and transparency, 9) safety and security, and 10) interagency collaboration. The call for comments was open until 13 March 2020 and it is said that a total of 81 comments were received.

According to the aforementioned Executive Order, a formal memorandum will be issued by OMB following this request for comments, and within 180 days thereafter, the heads of each executive department and agency are to submit an implementation plan to ensure consistency with the memorandum.

In February 2020, the Office of Science and Technology Policy (OSTP) also released an annual report in response to the aforementioned Executive Order, which is summarized from the following perspectives: investing in AI research and development, releasing AI resources, removing AI innovation barriers, improving the AI-ready workforce, promoting an international environment that supports AI innovation by the U.S., and adopting trustworthy AI for government services and missions[59].

The U.S. Department of Defense (DoD) also officially adopted ethical principles for AI on 24 February 2020[60]. The principles are based on the contents presented by the said department's Defense Innovation Board in October 2019, and include five areas as follows: 1) being responsible for the development, deployment, and use of AI; 2) minimizing of bias; 3) being transparent regarding AI capabilities, etc.; 4) safety and security measures being put in place; and 5) being governable to completely stop unintended behavior. In addition, the report states that guidance is needed to put these into practice.

---

memorandum-to-the-heads-of-executive-departments-and-agencies>

[58] "Executive Order on Maintaining American Leadership in Artificial Intelligence"
Referenced at the following URL:
<https://www.whitehouse.gov/presidential-actions/executive-order-maintaining-american-leadership-artificial-intelligence/>

[59] AMERICAN ARTIFICIAL INTELLIGENCE INITIATIVE: YEAR ONE ANNUAL REPORT
Referenced at the following URL:
<https://www.whitehouse.gov/wp-content/uploads/2020/02/American-AI-Initiative-One-Year-Annual-Report.pdf>

[60] DOD Adopts Ethical Principles for Artificial Intelligence
Referenced at the following URL:
<https://www.defense.gov/Newsroom/Releases/Release/Article/2091996/dod-adopts-ethical-principles-for-artificial-intelligence/>

(4) Organization for Economic Co-operation and Development (OECD)

The OECD Committee on Digital Economy Policy (CDEP) held a meeting on 21-22 November 2019. Regarding AI, there was a discussion of the draft practical guidance on the implementation of the OECD Principles on Artificial Intelligence, which was adopted in May 2019, a demonstration of the OECD AI Policy Observatory (OECD.AI) and presentations by the Secretariat on the objectives and operational policies of the informal network of experts (described later), as well as the (draft) criteria for those who participate and contribute as partners of the OECD.AI. France and Canada also presented at the Global Partnership on AI (GPAI), which is described later.

Furthermore, the OECD held the first meeting of the OECD Network of Experts on AI (ONE AI) on 26-27 February 2020. The ONE AI is an informal advisory group that offers to the OECD their expert opinions to help advance their analytical work on AI. It is composed of multi-stakeholder and interdisciplinary experts[61]. At the first meeting, the group presented specifically the classification methods in AI and their initiative for realizing human-centered AI.

Following the ONE AI meeting, the OECD also held a launch event for the OECD.AI[62] on the same day (27 February 2020). The OECD.AI is a platform (live database) for OECD-related committees and non-OECD policymakers to address AI policy issues, solutions and measurement methods, and to promote information sharing about AI initiatives. It consists of the following four core activities, which provide measures to share information about AI, take advantage of policy opportunities, and solve issues.

- **OECD AI principles**: Publish the OECD's AI principles and guidance (practical guidance) for practitioners.
- **AI Policy areas**: Provide access to a variety of contents, including AI policy news and publications on AI research, for each public policy area.
- **Trends and data**: Post data regarding AI research. Regional comparisons and changes over time are available to watch.
- **Countries and initiatives**: This is a database of national strategies, policies and initiatives relating to AI, allowing people to share and compare AI policies of each country.

The contents are expected to be updated on an ongoing basis.

(5) GPAI

On 15 June 2020, the “Global Partnership on AI (GPAI)” was established as an international initiative to address the responsible development and use of AI based on “human-centered” ideas. A joint declaration was made by the G7 countries, including Japan, as well as Australia, India, Mexico,

[61] Dr. OSAMU Sudoh, Chair of the Conference, was selected from Japan.
[62] <https://oecd.ai/>

New Zealand, South Korea, Singapore, Slovenia and the EU (going through a formal procedure)[63]. The Joint Declaration states the following:

> As the founding members, we will support the responsible and human-centric development and use of AI in a manner consistent with human rights, fundamental freedoms and our shared democratic values, as elaborated in the OECD Recommendation on AI. To this end, we look forward to working with other interested countries and partners. GPAI is an international and multi-stakeholder initiative to guide the responsible development and use of AI, grounded in human rights, inclusion, diversity, innovation and economic growth. In order to achieve this goal, the initiative will look to bridge the gap between theory and practice on AI by supporting cutting-edge research and applied activities on AI-related priorities. In collaboration with partners and international organizations, GPAI will bring together leading experts from industry, civil society, governments and academia to collaborate across four Working Group themes:
>
> 1) Responsible AI
> 2) Data Governance
> 3) The future of work
> 4) Innovation & commercialization
>
> Critically, in the short term, GPAI's experts will also investigate how AI can be leveraged to better respond to and recover from COVID-19.

Prior to this Joint Declaration, a G7 Science and Technology Ministers' meeting was held online on 28 May of the same year, and the Ministers' Declaration on COVID-19[64] was adopted. In the Declaration, the establishment of GPAI is described as follows:

> **Launch the Global Partnership on AI (GPAI)**, envisioned under the 2018 and 2019 G7 Presidencies of Canada and France, to enhance multi-stakeholder cooperation in the advancement of AI that reflects our shared democratic values and address to shared global challenges, with an initial focus that includes responding to and recovering from COVID-19. Commit to the responsible and human-centric development and use of AI in a manner consistent with human rights, fundamental freedoms, and our shared democratic values.

[63] Referenced at the following URL:
< https://www.soumu.go.jp/menu_news/s-news/01tsushin06_02000204.html >

[64] Referenced at the following URL:
< https://www.state.gov/g7-science-and-technology-ministers-declaration-on-covid-19/ >

(6) The United Nations Educational, Scientific and Cultural Organization (UNESCO)

At the 40th Session of the General Conference in November 2019, UNESCO made a resolution to develop a global recommendation on the ethics of AI before the General Conference in 2021, and has commenced work on the preparation.

UNESCO also appointed 24 experts to form an Ad Hoc Expert Group (AHEG) in order to prepare the draft recommendation on the ethics of AI on 10 March 2020[65]. The first meeting of the group was held online on 20-24 April and they discussed the first draft[66]. At the meeting, Director-General Audrey Azoulay noted that: "the use of various digital technologies has increased due to COVID-19. This highlighted the existing ethical challenges about the development of AI. It was therefore important for AHEG to start preparing the draft of the normative document".

Further discussions will be held with several stakeholders before July of the same year.

(7) G20 Digital Economy Ministerial Meeting

Saudi Arabia, the chair of the G20 in 2020, held an extraordinary G20 Ministers of Digital Economy meeting on 30 April 2020 and adopted a COVID-19 Response Statement[67].

In the Statement, the following part is in a section entitled "Exchange of data in a secure manner".

> "Acknowledging the uncertainty associated with COVID-19 and the power of data and Artificial Intelligence (AI) to accelerate pattern recognition and enable evidence-based policy-making, we encourage collaboration to collect, pool, process, and share reliable and accurate non-personal information that can contribute to the monitoring, understanding, and prevention of the further spread of COVID-19 as well as other infectious diseases. COVID-19-related data should be collected and processed in an ethical, transparent, safe, interoperable, and secure manner that protects the privacy and data security of individuals, in line with the International Health Regulations (IHR) 2005, and national laws and regulations. We acknowledge the need to ensure that potential biases in the data or algorithms are appropriately addressed".

Another section entitled "Research and development of digital technologies for health" also mentions of acknowledging the potential of digital technologies, including AI, to contribute to the fight against and prevention of pandemics, as well as increasing investment in the research of those technologies.

---

[65] Dr. OSAMU Sudoh, Chair of the Conference, was selected from Japan.

[66] Referenced at the following URL:
<https://en.unesco.org/news/unescos-international-expert-group-begins-work-drafting-first-global-recommendation-ethics-ai>

[67] Referenced at the following URL:
<https://www.soumu.go.jp/menu_news/s-news/01tsushin06_02000202.html>

Chapter 2 Prospect of an Ecosystem Formed with the Progress of AI Networking

1. Background and analysis policy

(1) Background

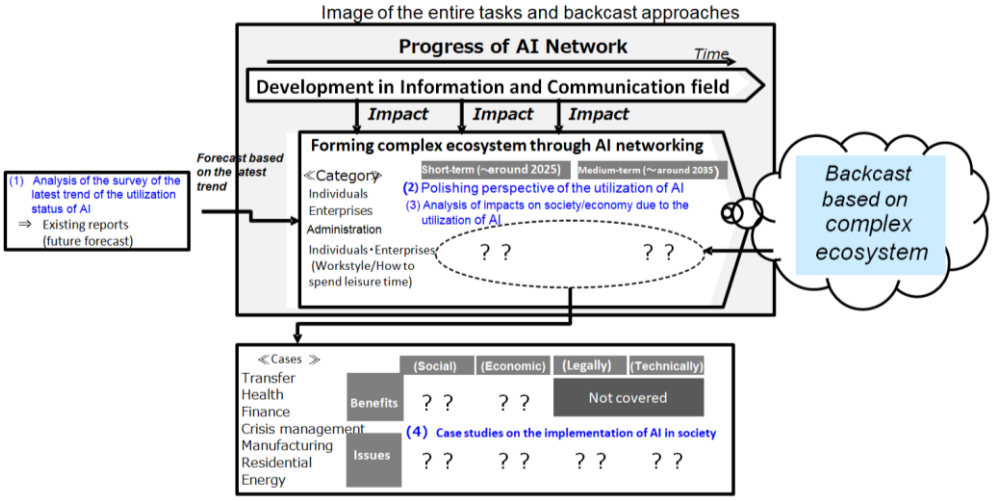
In the Report 2018, the Conference presented the prospect of an ecosystem that was expected to be formed with the progress of AI networking (hereinafter referred to as the "ecosystem prospect"), prior to discussing ways of governance regarding the utilization of AI.

Based on the envisaged ecosystem prospect, analysis of the benefits and risks associated with the utilization of AI was conducted, and the AI Utilization Guidelines, including AI principles, were formulated while taking issues common to each field into consideration.

In this context, the technologies and the AI utilization based on these technologies outlined in the Report 2018 rapidly developed. Therefore, the Report 2020 specifically provides an update in light of various information published after the publication of the Report 2018.

In a society that is undergoing a period of transformation of its industry structure and people's lifestyles due to the rapid progress of ICT, which is represented by AI, Big Data, and IoT, etc., it is necessary to consider the "AI Utilization Prospect" from an overview and assumptions of how the entire social ecosystem should be. Moreover, when considering measures to ensure the healthy development of AI networking, it is necessary to evaluate the benefits and issues based on AI utilization scenes, which are extracted from various backgrounds (including backcasts as well as forecasts) on the assumption that the ecosystem will become more complex in the future. This is shown in the figure below.

Big picture of tasks and image of the backcasting approach



(2) Analysis policy

a. Prospect of AI utilization scenes

In order to work on the prospect, as with the Report 2018, classification of AI utilization scenes was carried out from the **perspective of end-users** who use AI, both consumer users and business users. Regarding the usage scenes, the perspective of "Government use" was added to the "Personal use" and "Corporate use" that were already included in the Report 2018.

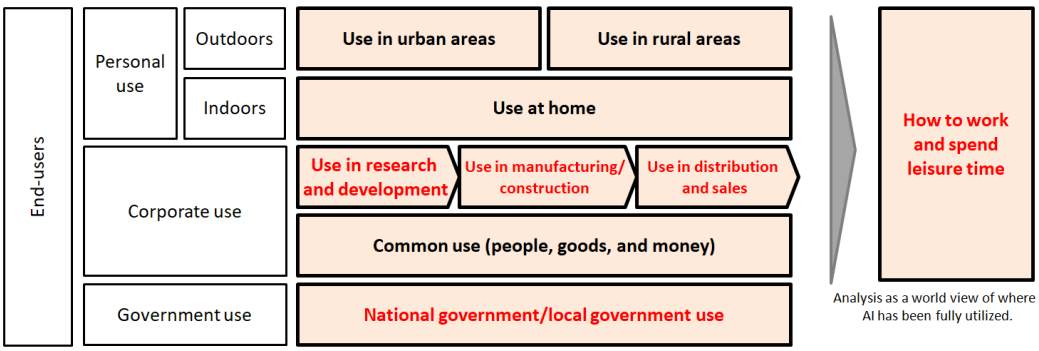
As with the Report 2018, "Personal use" was divided into "Outdoors" and "Indoors (use at home)", and "Outdoors" was further divided into "Use in urban areas" and "Use in rural areas".

Regarding "Corporate use", as with the Report 2018, analysis was also conducted on "Common use". Furthermore, "Corporate use" was divided into "Use in research and development", "Use in manufacturing/construction" and "Use in distribution and sales" by focusing on the processes within companies, and the analyses were conducted.

In addition, when AI becomes sufficiently used in these scenes, the work style and the amount of leisure time will possibly change significantly. Therefore, in order to analyze the worldview where AI has been fully utilized, the area of "How to work and spend leisure time" was defined, and the analysis was also conducted.

Refer to the chart below.

Classification of AI utilization scenes from the perspective of end-users



(Created based on "Prospect of the Ecosystem Formed with the Progress of AI Networking" in Appendix 2 of the Report 2018 (The areas with red text were added to the part that was mentioned in the Report 2018)

Also, as part of the efforts to prevent the spread of COVID-19, the new normal is being expected, and the lifestyle of people utilizing AI is changing along with this. Some of the examples of AI utilization in response to COVID-19 have been described in Chapter 1.1, and will also be discussed in this chapter.

b. Case studies on the social implementation of AI

In consideration of above utilization scenes, case studies on the following seven cases were conducted and the benefits and issues of AI utilization were summarized.

- Finance
- Manufacturing
- Residential
- Health (medical care / nursing)
- Crisis management (crime prevention, public infrastructure, and disaster prevention)
- Transport (fully autonomous driving)
- Energy

2. AI Utilization Prospect

Based on the classification of AI utilization scenes described in section 1 of this chapter, AI utilization prospect was envisaged along with the assumption of the main usage scenarios in each classification, as shown below.

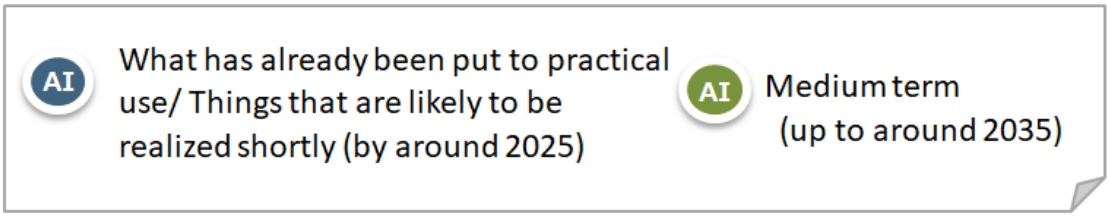
Utilization Scenes			**Usage Scenarios**
Personal use	Indoors	Use in urban areas	Transport, nursing, tourism/travel, and human resource development
		Use in rural areas	Transport, medical care, work, and the living environment
	Outdoors	Use at home	Healthcare, housework, safe and comfortable living environment, and lifestyle
Corporate use		Use in research and development	Research, development, common use (efficiency), and common use (integration of findings)
		Use in manufacturing/construction	Design, production planning, manufacturing and stocking
		Use in distribution and sales	Logistics, advertisement, after-sales service and sales
		Common use	People, goods, and money
Government use		National government/local government use	Policy making (government), administrative affairs and execution (government) Policy making (prefectures and municipalities)

	Administrative affairs and execution (prefectures and municipalities)
How to work and spend leisure time	How to work and spend leisure time

For each of these utilization scenes, the prospect of AI utilization was envisaged as shown in the following example of utilization in urban areas. Details of the AI utilization prospect in each utilization scene are shown in Appendix 2-1[68]. The following points have been taken into account in the analysis.

- **Timing for realization of the usage scenarios**

When analyzing the usage scenarios, the realization timing was classified into "already in use/likely to be realized in the near future (around 2025)" and "mid-term (around 2035)". The classification is based on the timing of the launch of the service, without taking into account the service's penetration rate or the level of functionality.



- **Positioning of "AI use in service provision" in companies**

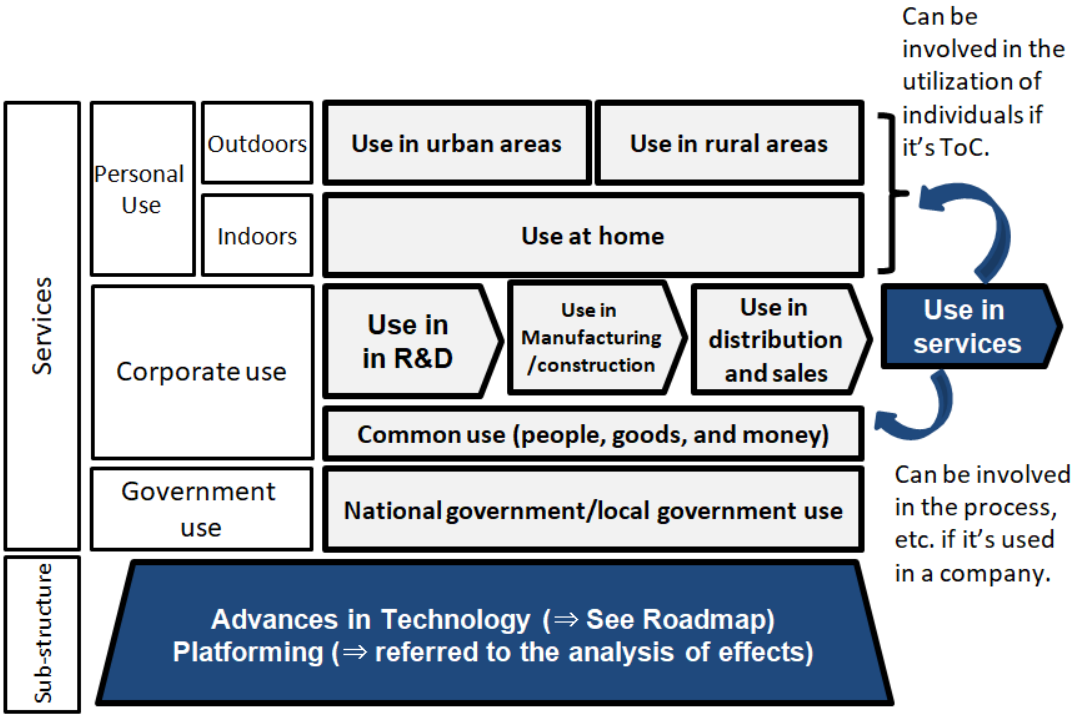
Because the utilization scenes are classified from the end-user perspective, if the service is for individuals, the use of AI in service provision by companies is included in the personal usage, and if the AI is used within the company to provide the service, it is included in the process or common use, so the scene of a "service provision" is not defined in the classification of corporate use.

- **Substructure analysis**

Technological advancements and changes in business models (e.g. platform business) are considered to be the substructure of AI utilization scenes. For this reason, the technological advancements are taken into account by extracting the usage scenarios with reference to the roadmap below, while the business model changes are referred to in the impact analysis section rather than as usage scenarios.

[68] Even if it is expected to be difficult to realize based on the current laws and regulations and the level of technologies in practical use or under research, the possibility of utilization in the future was envisaged and described here. Moreover, it is necessary to keep in mind that economic costs and other factors will be taken into account in the realization.

Points to keep in mind when analyzing the ecosystem prospect



The roadmap used as a reference (example)

Classification		発表元	Roadmaps/Strategies, etc.	Announced date
Entire domains		Cabinet Office	AI Strategies 2019	Jun. 2019
		NEDO	Roadmap of R&D Goals of AI and Industrialization	Mar. 2017
		MIC	Subcommittee for Making the Future of the IoT New Era, Tech Strategies to Catch the Future	Aug. 2018
		MIC	Interim Report from Information and Communication Council in response to Inquiry about Ideal State of New Information and Communications Policies for the IoT and Big Data Era (No. 23 of 2015)	Jul. 2017
		MIC	Interim Report from Information and Communication Council in response to Inquiry about Ideal State of New Information and Communications Policies for the IoT and Big Data Era (No. 23 of 2015)	Jul. 2016
		MIC	Roadmap of Promotion of Implementation of Regional IoT (Revised)	Apr. 2018
		MEXT	The Vision of Future in society through the development in Science and Technology	Nov. 2019
Domain of individuals	Administration	MIC	Standardization in the operating process/system in local governments and Society on Utilization of AI/Robotics (Smart Local Government Society)	May 2019
	Medical care/ Nursing	MHLW	AI Development Acceleration Consortium in Health/Medical Care Field Document Reference 3 Arrangement of AI Development Acceleration Consortium in Health/Medical Care Field Discussion and Future Directivity and Future Directivity"	Mar. 2019
	Healthcare	METI	Future Ideal Medical Care/Welfare/Nursing Fields in 2040 and survey on formulating the roadmap, etc.	Oct. 2019
	Transfer	Cabinet Office	ITS Concept for Public and Private Sectors/Roadmap 2019	Jun. 2019
	Manufacturing	METI	Smart Factory Roadmap	May 2017
	Construction	MLIT	The Study Group for preparing the establishment of AI Development Support Platform	Aug. 2019

3. Case studies on the social implementation of AI

Based on the utilization scenes of AI described in section 2 of this chapter, case studies were conducted on the seven cases described in section 1 of this chapter: finance, manufacturing, residential, health (medical care and nursing), crisis management (crime prevention, public infrastructure, and disaster prevention), transport (fully autonomous driving) and energy. Some examples of the benefits and issues that could be expected from the utilization of AI were then organized.

In particular, examples of the issues were classified into social, economic, technological and legal issues, and these were further classified into those which would arise before and after the introduction of AI as shown in the table below.

Details of each case study are listed in **Appendix 2-2**[69].

Case: Transfer (Fully autonomous driving)

Expected benefits (example)

- Humans will not need to drive, and travel time can be effectively utilized when traveling by car.
- Older people and people with disabilities will be provided with a convenient means of transportation, which will allow them to go to the hospital or go shopping smoothly.
- People will not need to drive long-distance trucks or long-distance buses at midnight or early morning, and they will be able to review their workstyle and work-life balance.
- In particular, problems such as a shortage of drivers for route buses in rural areas can be improved, and the abolition and reduction of routes can be avoided.

Expected issues (example)

	Before realization	After realization
Society	• It is unclear whether autonomous driving is technically secure or who is responsible for accidents if any. Therefore, there is a possibility that the service will not be accepted due to people's feelings of resistance to autonomous driving.	• The flow of using the time devoted to commuting and attending school for other purposes will be created. As a result, there is a possibility that the places and lifestyles of individuals will change significantly.
Economy	• In the case of infrastructure-coordinated autonomous driving, infrastructure may not be developed in local governments due to tight budgets, and there may be regional disparities in the spread of autonomous driving. • Employees related to delivery and transportation services may be reduced, making it difficult for them to find other jobs. Implementation of AI cannot be accepted socio-economic as a whole.	• Automakers, which are becoming increasingly popular, can secure a lot of post-sales learning data, which may hinder the entry of newer automakers. • The lack of driving by humans could significantly reduce the accidents that have previously occurred as a result of human error and would require a significant change in automobile pricing for insurance.
Technology	• AI may behave unpredictably with respect to data not belong to the training data. In addition, even with using accurately trained AI (model), we may not avoid erroneous recognition and derecognition.	• AI may not be able to respond to changes in the world after their deployment. • There is a possibility that proper operation cannot be performed because negotiations and adjustments cannot be made between cars. • If the AI system is hacked, not only will the autonomous vehicle not function properly, but it may also affect other autonomous vehicles via the network one after another, resulting in accidents and traffic disruptions.
Law	• The black-boxing of AI may make it difficult to establish the legal responsibility for autonomous driving, which may make it difficult to form a consensus with automobile manufacturers and users	• In addition to domestic legislation, coordination with other countries will be necessary, and it will not be possible to deal with the current legal system alone. Each automobile manufacturer may be forced to take new measures.

(Note) Some examples of expected benefits and issues are listed.

Note that the issues (examples) envisaged here are <u>not</u> described meaning that:

- (Regarding the "before" description), nothing can be realized until all issues have been resolved.
- (Regarding the "after" description), nothing should be implemented until all issues have been resolved.

[69] With regard to the case studies on social implementation in each field summarized in chapter 2 of the Report 2020 (**Appendix 2-2**), there was an opinion that it may be necessary to carefully analyze the intermediate step of promoting social implementation through individual interviews and to address many individual issues that will arise. There was also the opinion that it may be necessary to discuss how to address individual issues as well.

Rather, these issues imply opportunities for business development and research & development, and if they can be resolved through the efforts of companies, etc. great benefits will be generated in the future. This was written with a belief that resolving and responding to these issues one by one will lead to added value for the related entities, and is in line with the statement below taken from the "AI Utilization Guidelines" (page 4) published by the Conference in August 2019.

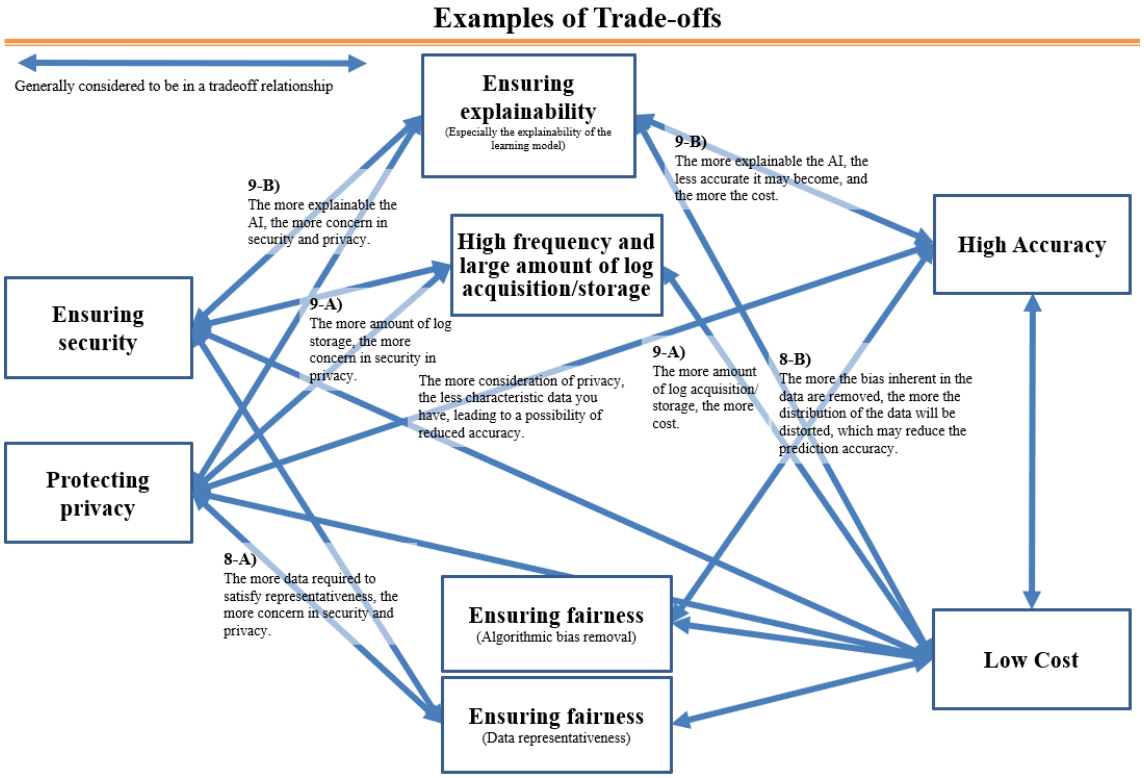
It may be possible for AI service providers and business users to add value to their AI services and businesses which utilize AI by undertaking such voluntary efforts.

Chapter 3 Initiatives by Developers and AI Service Providers

1. Starting point for discussion

Although developers and AI service providers are making efforts to establish AI principles, the situation shows that the approaches are not spreading sufficiently[70]. Of course, it can be said that the establishment of AI principles by developers and AI service providers is only one of the voluntary efforts toward promoting the "safe, secure and trustworthy implementation of AI in society". However, by sharing and deepening mutual understanding of the significance of AI principles through the development of AI and the provision of services, initiatives to establish AI principles will expand, and "safe, secure and trustworthy implementation of AI in society" will be facilitated. As stated in the "Introduction" section, both guidelines established by the Conference intend to serve as a reference for efforts to form voluntary AI principles. Further, as a specific function of the AI principles, it is conceivable that they will play a role as a business decision tool in the development of AI and the provision of services, taking into consideration the trade-off factors shown in the figure below. Therefore, deepening the understanding of what kind of governance is considered necessary for the establishment and utilization of AI, including the formulation of AI principles, is also considered to be an effort necessary to promote "safe, secure and trustworthy implementation of AI in society". Based on the awareness of these issues and related descriptions within "Chapter 2: Concept of Formulating the AI Utilization Guidelines, 4. Future Development" and "Chapter 3: Future Issues" in the Report 2019, we have presented the points at issue in "Introduction" points "(1) (a) and (b)" and decided to conduct interviews focusing on these points from enterprises that are making ambitious initiatives.

[70] Intensive surveys of initiatives to establish AI principles in Japan and overseas: Ministry of Internal Affairs and Communications, Information and Communication Policy Research Institute, Information and Communication Laws Study Group, AI Section Meeting FY2019 1st meeting, Member SHIMPO Fumio (Professor, Faculty of Policy Management, Keio University) presentation material “Do AI principles work?” https://www.soumu.go.jp/main_content/000660996.pdf See p6 and beyond.



Cited from 'Report 2019' Exhibit 1 (Appendix) "Detailed discussion of each issue of the AI utilization principle"

2. ABEJA Co., Ltd. (approaches in "Ethical Approach to AI (EAA)")

(1) Interview outline[71]

a. Service overview

Provides the ABEJA Platform, a platform that streamlines the implementation and operation process of deep learning, and ABEJA Insight for Retail, a store analysis service utilizing AI based on the acquisition and analysis of customer behavior data.

b. Establishment of Ethical Approach to AI (EAA) as a governance system

With ABEJA Insight for Retail, cameras are installed in retail stores to conduct customer flow and repeat analysis by taking facial images of customers. Due to the customers' facial images acquired in this way, the company made efforts such as complying with laws and regulations and notifying the purpose of use based on the Act on the Protection of Personal Information and the Guidebook for

[71] For interview materials (excerpt), refer to <https://www.soumu.go.jp/main_sosiki/kenkyu/ai_network/02iicp01_04000232.html>.

Utilization of Camera Images Version 2.0[72] and such. However, partly since ethical issues such as privacy became a concern to many, it established the Ethics Approach to AI (EAA) consisting of members all from outside the company (executives and employees participate as observers) in July 2019.

c. History of the establishment of EAA

The trigger was a proposal from a global member of the company suggesting that a verification organization is necessary because business verification from an ethical perspective is globally essential. There was also a need from a legal perspective regarding ethical issues related to points specified in law. Upon selecting committee members, the global members' proposal to consider balance of nationality and gender was reflected.

d. Position of the EAA

Positioned as an advisory body for conducting discussions in line with specific cases (Not a system where EAA itself makes decisions as a whole).

(2) Discussion

[Approach of sharing information with other startup companies]

Q. This is an excellent approach as a startup company. Do you plan to share this approach with other startup companies?

A. Since the EAA has only just launched, we have not introduced it to other companies yet.

[Self-assessment of the significance of the EAA]

Q. Given the limited capacity of a startup company, do you find it difficult to establish an ethics committee like the EAA?

A. As a startup company, there are some difficulties in terms of corporate vitality, but the fact that the establishment of an ethics committee itself is making positive effects is a common understanding among executives, and we recognize its importance.

[72] See IoT Promotion Consortium Camera Image Utilization SWG <http://www.iotac.jp/wg/data/camera/>.

3. FUJITSU LIMITED ("AI Governance for AI Developers")

(1) Interview outline[73]

a. Establishment and outline of AI ethics guidelines

Announced as the Fujitsu Group AI Commitment in March 2019. The Fujitsu Group AI Commitment includes the three features of "Human Centric vision for AI", "Objective AI Ethical Guidelines" and "Establishment of the Fujitsu Group External Advisory Committee on AI Ethics", as well as declarations on five commitments.

b. Objective AI Ethical Guidelines

Worked with experts in science and technology ethics through collaboration with AI4People in Europe to understand the requirements for AI ethics without excesses or deficiencies. Fujitsu formulated the AI commitment based on AI4People's five principles of AI ethics, and by converting it into a message for society as a whole, including the company's stakeholders.

c. Establishment of the Fujitsu Group External Advisory Committee on AI Ethics

Recognizing the importance of autonomous objectivity in AI ethics initiatives, the company established a committee of outside experts in AI and other fields to receive evaluations from this perspective. At present, an external committee has been set up to ensure autonomous objectivity first, rather than system establishment for auditing by a third-party independent organization, due to the following three reasons. (1) AI technology itself is in its infancy and its direction is not fixed and is quick to become out of date; (2) discussions on the standardization of AI ethics, quality assurance, etc., have only just started, and assessment items have not yet been finalized; and (3) Although AI is expected to be implemented in fields unrelated to ethics, setting up auditing before its development seems to be premature because the whole would be shrinked.

d. Features of the Fujitsu Group External Advisory Committee AI Ethics

Owns a mechanism to incorporate AI ethics into the corporate governance system. Specifically, the External AI Ethics Committee shares information with the Board of Directors. The Board of Directors is an organization that makes decisions on business execution and supervises business executive directors, so a key feature is that the management itself is linked to AI ethics.

e. Direction of discussion on AI governance in an AI network society

- The focus of discussions will be broadly defined as "Inconveniences that society may incur through the use of not only AI but also ICT in general," and in addition it is realistic to balance

[73] For interview materials (excerpt), refer to <https://www.soumu.go.jp/main_sosiki/kenkyu/ai_network/02iicp01_04000232.html>.

AI with other ICT regulations.

- Discussions tend to focus on risk control, but it is not easy at this point to envision long-term and universally effective risk control measures for AI technology, which will continue to evolve intensely into the future. Rather, while taking practical measures against currently feared risks, rules that excessively shrink research and services should be avoided.

(2) Discussion

[Activities besidesthe Fujitsu Group External Advisory Committee on AI Ethics]

Q. Are there any activities in addition to the Fujitsu Group External Advisory Committee on AI Ethics?

A. We are building a system to grow awareness amongst employees through educational activities and the preparation of manuals for the workplace when customers make inquiries.

[Mechanisms for ensuring the effectiveness of AI commitments]

Q. Regarding the AI Commitments, at which stages of AI development, services, and product supply are mechanisms in place to ensure the effectiveness of these commitments?

A. Of the main business flow of B-to-B-to-C, the concept of AI Commitments is to deliver safe and secure AI to customers, the middle B, and C. We are keeping that philosophy in mind from the point of AI development, and working with the hope that it will reach front-line customers.

4. IBM Japan ("IBM's Approach to AI")

(1) Interview outline[74]

a. Principles and ethical guidelines

Published "IBM's Principles for Trust and Transparency" in May 2018. Compiled the company's principles of trust and transparency.

A strong message that IBM is not the only company that will make efforts to ensure explainability and transparency, and that companies using IBM's technology must also make such efforts.

In the form of guidelines for this purpose, "A Practical Guide to AI Ethics" was published in September 2018. In 2018, there were cases where certain biases occurred in AI judgments especially in the United States that caused disadvantages to society. In light of this, the focus was not only on explainability and transparency but also on how to support unbiased AI judgments.

b. Approach to providing technical products based on principles and ethical guidelines

[74] For interview materials (excerpt), refer to <https://www.soumu.go.jp/main_sosiki/kenkyu/ai_network/02iicp01_04000232.html>.

Provide products by adding the elements of the above principles to the software product lineup. The Watson API and the Watson Open Scale reflect the elements very much.

c. Watson API Overview

(i) Overview

Provides a set of functions that can be used for various AI use cases, such as speech recognition, image recognition, and natural language understanding as parts, in the form of API calls. The parts have several learning patterns, and the APIs being provided have three patterns: (1) models trained by the company, (2) partly customizable by training with some customer data, and (3) customer-trained models.

(ii) Key features

- In the cases of (i)(2) and (3), regarding how the company will handle the business-specific information provided by customers and for concerns whether that information will be spread to other companies were clarified as a mechanism in the guideline. Gave the customers the option of whether they will provide the data they have fed to teach the Watson API for training the base model.
- Clarify attribution of insights gained using training data or the Watson API after training, or training data itself, in the contract at the start of using the service.
- For customers concerned about using the Watson API to feed their learning data into the cloud, an option of using an on-premises form of the Watson API is provided.

d. Ensuring Trust and Transparency of Customer-Developed AI Models with Watson Open Scale

- Released to support the monitoring of how the customized AI model for each customer is working (for example, accuracy, bias, and performance) in 2018. The tool works to monitor AI models to ensure fairness and explainability.
- Open Scale allows developers to monitor AI models developed by competitors or other open-source libraries. The system is intended to be used as an integrated monitoring system for various AI models used by companies.

e. Ensuring Explainability and Transparency

Many customers are showing interest, and there is an impression that there is growing momentum for implementing while properly securing the principles of AI within the company. With that said, the number of companies using Open Scale to monitor AI is very few, at least in the Japanese market. At present, customers are interested in what it is like to utilize AI in a company. Many say that after they

achieve this to some extent, they then want to monitor the AI, and then step in further to secure explainability and trustworthiness.

However, it is also a fact that there are many inquiries, and it is necessary for companies to have such a system when considering the future.

f. Requests

- For companies to work on ethics, explainability, and transparency will most probably take a little longer, but if we start thinking about how to deal with them when the time comes to ensure explainability and transparency, we feel it will only be a reactive approach. In that sense, we are repeatedly inputting what is needed as the next step to customers who are currently practicing AI development, but we also request the government to actively send out messages on the use of AI and the explainability and transparency beyond it.
- For that purpose, it is important to promote not only explainability and transparency, but actual utilization in companies. We request for activities where the momentum to promote AI and the support for its promotion can be seen when looking at Japan as a whole, such as the best practices for utilizing AI in the private sector be collected and made public.
- We also request to encourage the initiatives of Japanese society as a whole regarding the clarification of data rights as well.

(2) Discussion

[About the concept of fairness]

Q. There was talk about data bias and algorithm bias. Regarding the definition of fairness, there are fairness in the sense that the starting point shall be all equal and fair, and fairness in the sense that the final output shall be fair. How do you plan to apply these principles?

A. For example, fairness considered in Open Scale requires the customer to judge which items of the AI model created by the customer shall be monitored, and then what the tool can do is only raise an alert. When the alert is raised, our basic stance is to leave it to the customer company operation to decide whether this alert is actually a bias or whether it is an appropriate offset that can be ignored. Furthermore, in terms of fairness, it is not only related to data bias, input fairness, and output fairness, but also to social norms and environmental backgrounds. Therefore, in providing such a system, we basically support companies to set their own fairness, visualize it using tools, and monitor it.

[Transparency principles: input/output verifiability (monitoring other AI models)]

Q. It is often reported that inappropriate systems may be created depending on customer data or customer preferences. Regarding this, as a company, I think that the position of whether to say it is

the customer's responsibility or to take ethical responsibility for the product will be reflected in the price of the product. I believe Open Scale manages models created by the customers, but if it is your own product, you can apply the concept of Ethics by Design, but for other company products, usually, this cannot be done because the contents are mostly a black box. With that said, how can management be achieved?

A. Open Scale focuses on seeing how things are working in production. Regarding the monitoring of other company AI models, we collect a large amount of input-output data to see if there is any offset.

C. These initiatives of the Company may fall under the practice of "input-output verifiability and explainability" in the "principle of transparency" of the "AI Development Guidelines."

[Spread of the guidelines]

Q. Are customers aware of your company's guidelines or the "AI Development Guidelines" when you conduct discussions with your customers?

A. It is rare for customers to be aware of the guidelines issued by our company, and it is also rare for them to directly quote the guidelines issued by the government. However, customers make general statements such as the increasing awareness of risks and challenges associated with trust or explainability in the general public. So, I have a feeling that they may be studying the guidelines issued by the government.

[Tradeoff between explainability and performance]

Q. Explainable AI is becoming popular, but it is recognized that there is a trade-off relationship where performance decreases as it is made more explainable. How is this handled by your company?

A. To ensure explainability, there are a number of companies that are using so-called white-box AI as a guarantee of explainability, where the way of making the AI is completely a white box. However, it is difficult to guarantee explainability with a white box when analyzing a large amount of data through deep learning, such as images and sounds, and creating an AI model for analysis. Our company's technology focuses to help customers understand AI models from the outside, without restricting how the model is created. Like the approach taken with Open Scale, we look at the model from the outside, rather than tampering with the model contents. In other words, our customer support approach is based on having the customers achieve both full utilization of deep learning technology and explainability.

[Commitment to fairness and trust and institutional responses]

Q. You mentioned that fairness and trust are set by the customers and your technology does the

verification. Considering the social responsibility of companies, if the customer company causes a problem, your company, which provided the system, will probably be named guilty as well. When this occurs, though you may say that it is only a tool, will your company make any commitments toward these social responsibilities, or are preparing to do so in the future? We would like to know where your company stands. Also, when this happens, there is a high possibility that the person in charge of development alone does not have clear answers regarding fairness and trust of the model. Recently, there are ideas such as including a specialist in the project or having a person in charge always review the model. Are there any such systems in place?

A. During the actual process of building customized AI models together with customers, we do not do things like creating an AI behavior model and immediately publishing it as is. We take an approach to guarantee the point that the AI model does not conduct inappropriate behavior by repeating tests on the model many times together with the customer. At present, we have not yet implemented approaches such as not making it public unless we pass a specific test.

5. NTT DATA Corporation ("NTT DATA Group AI Governance Initiatives")

(1) Interview outline[75]

a. Background to the establishing of AI guidelines

While various activities related to the AI principles are spreading, in June 2019, Japan's "AI Social Principles" was announced at the G20 Ibaraki-Tsukuba Ministerial Meeting on Trade and Digital Economy. As an IT service provider that spreads AI globally, the company felt responsible for presenting AI guidelines and announced it in May 2019.

b. AI Governance Initiatives in the company group

The "AI guidelines" currently published are about the philosophy behind AI development, and the content is very typical.

In the future, the company plans to establish more specific AI development processes based on the "AI guidelines". When developing an AI system, the company first has to coordinate the awareness of what kind of development process it will use to create the AI system with its customers, and then it must do a breakdown of what kind of details it will concretely confirm during the development process. Therefore, by the end of this fiscal year, it is aiming to create an "AI Development Process" which arranges the process and checking viewpoints of AI service realization. However, it is difficult to know what to actually do with the process alone, so the company is considering creating an "AI diagnostic method" that has more concrete methods to solve problems and check them, including the use of

[75] For interview materials (excerpt), refer to <https://www.soumu.go.jp/main_sosiki/kenkyu/ai_network/02iicp01_04000232.html>.

general tools and in-house development of insufficient tools.

c. Process for establishing AI guidelines

First, the company considered what kind of guidelines should be used to realize the group vision "Trusted Global Innovator", and then completed the AI guidelines by referring to the principles and guidelines already announced by the government and other companies, and incorporating opinions from the members of its domestic and overseas group companies.

Further, the AI Guidelines were submitted to the board of directors for approval and announced after reaching a consensus with the management level. In the future, the AI guidelines will be further refined as society changes and technology advances.

d. Business activity initiatives

For example, the company established the "AI Center of Excellence" in collaboration with overseas AI experts. The core members consist of 90 members from Spain and 50 members from Japan, making a total of 140 members. The members are from eight companies and seven countries.

At present, focusing on "AI Center of Excellence", it is developing the "AI Ethics Framework" which systematizes assessment and methodology on AI ethics for customers. This framework organizes the items to be checked in terms of governance, algorithms, organizations, and societies.

e. Examples of "AI Development Process" and viewpoint

AI development tends to lean towards agile development than traditional waterfall development commonly used for large systems. "AI Development Process" defines specific confirmation items and methods for each phase of agile AI development.

For example, in the case of data collection in the "AI Development Process," a specific development process is presented to developers by combining an assessment sheet that arranges important check items such as data volume and data quality, and a check tool such as AI bias visualization/detection tool. By using the "AI Development Process", the company believes that it is possible to confirm whether the AI services that it is actually creating together with its customers really follow the AI principles, and whether it can realize the customers' requests.

(2) Discussion

[Systematic measures for the use of guidelines in corporate activities]

Q. In implementing the guidelines and processes in corporate activities, how do you think about institutional measures, such as asking specialists to join development, or having a specific section conducting checks upon development?

A. At present, we are not considering checking or confirming being run by the organization or such

mechanism. This is because the development process has become the company standard methodology. With that said, the extent of application is determined through discussions with the customer, but if there is any problem, there is a technical team and a quality assurance team, and consulting to these teams have become common in normal development, so we plan to proceed in this way.

[About the standards of fairness]

Q. How do you set standards for fairness?

A. The basic idea is to confirm whether or not this is acceptable through discussions with the customer. When providing a service, it is not usually suddenly applied on a large scale, and a realistic solution will be to reduce the risk in small increments and closely noting reactions, so there are no standards for ensuring fairness at present.

[About data minimization]

Q. I think that the idea of data minimization, where unnecessary excess data should not be collected even when it is necessary to collect information from customers, is of concern these days. What do you think about this in terms of your company's philosophy? Also, I believe the information needed to be collected differs depending on the customer; how does this proceed?

A. Whether or not the idea of data minimization should be included in the philosophy is a subject for future consideration. We have experienced that kind of discussion in our interactions with customers, and customers say they don't want to submit it, and we don't want to receive unnecessary data. When receiving data, we provide tools for processing and partially masking the data, and masking data by tools is a prerequisite for providing data, especially for highly confidential medical and healthcare data.

[About actual cost invested for data governance]

Q. We do not know the cost of data governance each vendor is spending, but we believe that the cost of data governance as a percentage of overall costs will become a serious problem for IT vendors in the future. How do you feel about this point?

A. We have a sense that the invested cost varies depending on the type of data and the industry. There is no doubt that the medical, healthcare, and life sciences sectors that often handle highly confidential data are very sensitive to data governance. However, it is not possible to determine the specific percentage at this point. We also offer a data management platform, and although it applies its data management guidelines and checklists, in the areas of medical, healthcare, and life sciences, customers may have stricter data management guidelines, in which case we create the platform and services according to their guidelines. It's also clear that the retail sector is pretty

sensitive to data these days.

[Information sharing on data governance]

C. If there is a place in the industry where the invested cost for data governance can be shared, it may become relatively sound. If there is a vendor that does not take care of it too much, they may fall behind industrially. Your company is the largest system integrator in our country, so we hope that you can play such a role.

A. We think it is a good idea to collect and share the best practices through case studies. There are many AI principles being published, but it is hard to say a rule where everything will be absolutely OK as long as it is followed, so we have a feeling that we need to collect some examples of how things worked well within certain conditions and constraints. Although it is a part that requires consideration for customers, we would like to provide examples in a somewhat abstract form. Also, the "AI Development Process" and "AI Diagnostic Method" take the form of taking a few examples and generalizing it to some extent, so if this type of idea becomes recognized to some extent to the world, we would be most happy if we can incorporate data governance into the idea. Furthermore, for example, there is an activity called "QA 4 AI" in QA (Quality Assurance) of AI, and a wonderful idea has been proposed, so we think that we should refer to such an activity more and more and discuss how it was incorporated on a case basis.

6. Oki Electric Industry Co., Ltd. ("Improvement of Foundations for the implementation of AI - Establishment of "OKI Group AI Principles"-")

(1) Interview outline[76]

a. AI initiatives

Until the 2010s, developed AI in the lab, a bit distanced from the product level. In the 2010s, AI shifted its focus to business. In terms of products, the company has been selling products that use machine learning for image recognition since the 2000s, but those products were only a part of the market, and sales were not so high. In recent years, the company has decided to use AI for its main products.

b. AI productization/business foundation improvement project

While AI businesses and products have gradually emerged within the company, individual departments have been dealing with contracts and quality assurance when delivering products to customers based on the idea that they are the most appropriate, but when viewed as a streamline,

[76] For interview materials (excerpt), refer to <https://www.soumu.go.jp/main_sosiki/kenkyu/ai_network/02iicp01_04000232.html>.

expectations differ between departments, and the business divisions themselves were also concerned whether this was really appropriate. For this reason, the company decided there is a need to establish uniform rules within the company, so in FY2019, the company launched a company-wide AI productization/business foundation improvement project.

There are technical and non-technical aspects to business promotion using AI, but this project focuses mainly on non-technical aspects, ethics, quality assurance, and contracts.

Almost every department involved in the AI business participated in the project. The goal here was to design a realistic mechanism for implementing AI policies that can be implemented without difficulty, by having those who operate, make, and use the AI-related measures all take part in creating the guideline.

c. Process for establishing AI principles

Established in September 2019 to express the company's distinctive features while maintaining consistency with the shared values of the group by identifying all the adopted items of each country and organization that have already been submitted and comparing them with the company's Charter of Corporate Behavior and Action Guidelines.

d. Positioning of principles and purpose of establishment

Firstly, positioned for it to be a foundation for fulfilling social responsibility, and a principle for ensuring that all AI activities are based on this AI principle.

As a specific aim, clarified that it would present the company's basic philosophy and that it would further reinforce the existing action guidelines of the company as shared values.

e. Future initiatives

In the future, following the establishment of the AI principles, guidelines will be developed on how to create a system to comply with it and how to ensure compliance.

For future operations, the company will consider how to manage risks that need to be kept in mind when launching various AI products in the course of business promotion. Last year, as a trial operation, in a project that carries out a variety of checking processes upon delivering products to customers such as whether the contracts are valid, and how the quality is, the company conducted a preliminary check to the ICT business division that produces many AI products regarding items such as whether or not this is an AI-related project, and what kind of consideration should be given to it. Based on the results, the company is planning to start full-scale operations within this fiscal year.

Through the trial, the company confirms how much workload is given to the employees and how much the cost is. As exemplified in tradeoffs[77] in the Conference "AI Utilization Guidelines", people

[77] See 1. of this chapter.

in the field have an image that unnecessary processes hinder business, and they always ask why they have to do them. The purpose of the trial in FY2019 was to identify tradeoffs and identify what is necessary, as to prevent events such as implementing a system that focuses on minimizing the impact on the workplace as much as possible, so much that problems are experienced by customers in the future. If events like these occur, opinions stating that AI products are troublesome and costly may arise and also work as a disincentive, so after discussions through this trial, the company will establish a system and rules so that it can be reasonably operated from FY2020.

(2) Discussion

[Awareness of development principles and utilization principles and interest in FAT[78]]

Q. Until now, rules and principles have been created deductively, but when it comes to implementation, in the future, practices will most probably be gathered inductively by listening to the voices of customers, for example, by creating models together and discussing fairness and deploying in small increments. Most probably the company will go in this direction. Your company's main focus is on B-to-B operations, but I wonder if customers are aware of the government's development principles and utilization principles, such as if they are highly interested in the so-called FAT.

A. It depends on the customer company, and we think customer companies in the financial services industry are very concerned about fairness. Looking at customer companies in general, we think we have to convey it as a message.

[Spread of "social acceptability" in the company]

Q. By mentioning "social acceptability", which is one of the purposes of establishing the principle, "to provide AI products that can be accepted by many customers and society", is it easier to be accepted by workers in the field?

A. It depends on the person in the field. SE and salespeople who look into the future are interested in such matters and can understand them, but those who have difficulty in understanding the social situation cannot accept such matters by just that. Literacy education in this area needs to be carried out steadily into the future.

[78] Abbreviation for Fairness, Accountability, Transparency.

7. Microsoft (“Business, Responsibilities, and Challenges Surrounding AI -Ethics and the Potential of AI-”)

(1) Interview outline[79]

a. Resolving issues through collaboration

The history of AI utilization is short, and many unknown issues remain, so collaboration with organizations and groups in various fields is indispensable for safe and secure AI utilization.

b. Microsoft AI principles and the AI ethics review board

In 2017, the company was among the first to establish principles for AI. Based on this principle, an AI ethics review board has been established to control corporate activities. The two bases are transparency and responsibility. There are six principles, (1) fairness, (2) reliability, and (3) safety on the transparency base, and (4) privacy, (5) security, and (6) inclusive on the responsibility base. It is based on how to conduct business in accordance with the six principles in product R&D and actual service provisioning.

c. Ethical and responsible AI (fairness and reliability)

A very famous case published in the NY Times in 2018 where age identification quality differed between races. This is a symbolic case of how data offset, or data bias, led to raise doubts about the fairness and reliability regarding the use of AI.

Fairness varies from scene to scene. A case that made clear that in addition to AI algorithms, the distribution and bias of data used for learning must be carefully examined according to the purpose and use at the same time. AI vendors and IT vendors tend to focus on algorithms, tuning, and so on, but when providing them as actual services, it is necessary to carefully examine the data including the model and the fairness based on it.

d. Ethical and responsible AI (transparency)

It is very important to clarify the way of thinking that was used to collect data and build various AI systems. (While introducing a case study for predicting the risk of death from pneumonia, where the death rate from pneumonia in asthmatic patients is very low,) there are many cases in which the essence cannot be seen when only looking at the actual visible and comprehensible data. From a business perspective, it is important to understand the target and its meaning and essence of the data to develop systems using AI.

e. Ethical and responsible AI (privacy and security)

[79] For interview materials (excerpt), refer to <https://www.soumu.go.jp/main_sosiki/kenkyu/ai_network/02iicp01_04000232.html>.

(While introducing a case study where Target, a United States supermarket identified whether a teen girl was pregnant before her father through behavioral history such as internet searches,) one issue is how to operate an AI system on the assumption that private unrevealed information may become uncovered from the outside through data, analysis, modeling, etc.

f. Ethical and responsible AI (accountability)

Including face recognition, various things such as analysis and modeling using extremely large amounts of data, which were not possible until now, have become possible by utilizing AI. There may also be areas where the use of AI for undesirable purposes can trigger various social impacts. The company plans to conduct research and development and carefully evaluate and examine the system.

g. Considerations for responsible development and use of AI systems (concrete evaluation of the ethics review board)

- Purpose of the system. The company's vision is to show that its products and services maximize people's potential. Of course, the first evaluation point is whether or not the development and provision of AI will have a positive impact on people and society.
- Technology Readiness. A bit different from ethics, but rather than using AI for everything, to judge and evaluate whether it is technically correct or appropriate to use AI.
- Quality and reliability. Maintaining quality is a major key point. When building an AI model that involves a huge amount of data, the volume of data collected is so large that collecting data for system development becomes costly. In terms of maintaining quality and reliability, using AI can easily be costly and difficult to maintain for some targets. It is important not to blindly use AI, but to carefully assess its sustainability.
- Careful use of AI. Systems using AI needs careful evaluation and correspondence according to its use. When developing systems using AI and providing them as services, evaluation is done carefully in advance for what they are intended and how they will affect people and society.
- Service denial by AI. Evaluate from the viewpoint of whether the use will make it possible, whether it will result in fostering discrimination and inequality, and whether it will infringe upon the freedom and privacy of individuals.

h. Challenges surrounding AI and data (initiatives by the AI Data Utilization Consortium)

It is important to promote the resolution of social issues through the realization of smooth data distribution (case studies of the use of AI data for persons with disabilities and the elderly are described in Chapter 5, Initiatives for Consumer Users.).

AI research, open innovation, and creating solutions (commercializing) have many challenges to

be solved, including fragmented data providers, different licensing concepts, personal information, and integration with computing resources.

i. Various elements and issues required for data distribution infrastructure

(i) Different nature of data

Ideally, data shall be widely distributed and used, and different vendors shall be able to offer different solutions and different organizations will be able to contribute to solving social issues. However, when looking at data, it is different from conventional data analysis and utilization. In the case of machine learning, the data changes the shape of the model after the learning and becomes sustained in a sense. Value of data are completely different from conventional data distribution, and have characteristics close to technology patents. In the case of conventional technology patents, the content of the contract and the price are determined after interviewing and negotiating the commercial flow of products to which the technology is applied. On the other hand, in the case of the AI model, the value of the data cannot be determined until the commercial flow is determined, because the AI system created and constructed using the learned data itself is provided as a service or sold by being incorporated into a device. In other words, the data itself has secondary and tertiary values. As a recent trend, data holders and data acquirers with higher AI literacy tend to recognize it to be the same as technology patents. This, on the other hand, makes it difficult, including data distribution and contracts.

When promoting data distribution, it assumes that data retrievers, AI developers, and users will all be different. Since there is no such contract model that assumes this, the AI Data Utilization Consortium discussed, examined, and created contract models and templates.

(ii) Software programs and product liability

In the case of the implementation of algorithms in software development, it is relatively possible to clarify who is responsible for defects.

On the other hand, in the case of systems using AI, depending on the application target of the AI system, the use is very important. While the algorithm itself may be problematic, and there's room for improvement, there's still a long way to go before an explainable AI can be technically established.

(iii) Data Provenance at current AI quality (responsibility for the data itself)

Annotations are always used when images are used as learning data. Vendors often outsource annotations. If, for example, there was a malicious worker in the annotation vendor company, and problems were found in the AI system, it will be important to be able to trace back to the data whether the included data itself was malicious, or in the case of AI, whether the problem was the data used for learning, or whether a malicious worker was there in the labeling process for annotation, and if so, who.

Levels are broadly divided into three. For example, in the case of level 3 (applications of AI systems that have a variety of irreversible impacts on the body, etc.), it is important to be able to go back to the source of the original data used for learning and who labeled the label data, and to be able to fulfill accountability.

In terms of quality, changes in the data environment will lead to data obsolescence, and there will be a need for new data, and make impact on AI quality. It is necessary to carefully evaluate and consider whether this is due to changes in the environment or whether or not there was malicious intent.

(2) Discussion

[Malicious annotations and product liability]

Q. It is often thought that annotations in learning data are maliciously or even non-maliciously biased, which is a big problem for modern AI. In such cases, how much product liability can be questioned?

A. With regard to malicious annotations and product liability, we believe that liability arises depending on the object and content of the annotation. We believe that manufacturers and system developers must examine the learning data to a certain extent and confirm that no erroneous data are included. Naturally, there is a big difference between a case in which a problem occurs in spite of efforts to examine the fact that there is no wrong data and to confirm the quality, and a case in which a problem occurs without doing so at all. In that sense, the AI Data Utilization Consortium is also considering the establishment of a mechanism to ensure and secure data provenance and traceability, including who performed the annotation work.

[AI checks for mislabeling]

Q. With so much data being used as learning data, reviewing it with the human eye is time-consuming and extra work. Rather, if you're going to use a technology like AI, you should consider how to implement it to check for potential malicious data, or else you won't be able to defend your product liability.

A. As an example of such efforts, we do not believe that it is practical to visually check all contents of the huge amount of learning data. On the other hand, it is possible to collect a limited amount of data that has been checked but does not contain mislabeling. So, we are currently researching a recursive method of learning with these datasets and checking quality with some other label data.

[Transparency of AI data]

Q. In terms of transparency of personal data, there is a possibility that, for example, one's insurance is presented inaccurately. How is this ensured?

A. System are not made with intentions to result in a detrimental outcome from decision by wrong scoring. Because of the risks involved, we decided not to develop systems using AI in those fields. We made decisions not to develop systems or provide services in areas where accountability and transparency are difficult to achieve.

[History of consideration of product liability]

Q. What is the background to the consideration of product liability?

A. We have been dealing with the risk of the model changing rapidly as malicious learning data gets mixed in little by little without being noticed from three or four years ago, and various security threats of AI have been considered in the research field for a long time. At the same time, there was an actual problem called Tay, and we set up a team to work and research on it urgently. This was how product liability became of concern.

[About monitoring annotations]

C. When I went to a Silicon Valley company a few years ago, I was skeptical whether they were really monitoring annotations or data cleansing thoroughly, and I think it will be very important in the future. There was a comment that the more AI becomes available, the more important it becomes to understand crowdsourcing contracts, especially for companies working as brokers and what kind of responsibility and intellectual property management they will manage.

A. Upon surveying labeling and annotation vendors, we found that none of them made clear statements including withdrawing quality assurance. Some vendors provided service options for inspection, screening, and checking, but basically, all of them did not guarantee the results.

8. Anonymous (“Report on Support Tools for the use of AI by Private Enterprise Volunteers”)

(1) Interview outline[80]

a. Purpose

A research activity (Hereinafter referred to as “research activity”) conducted by volunteers from the private enterprise, which considers the establishment of a checklist for enterprises to securely incorporate AI up considering the incorporation of AI as an AI support utilization tool.

b. Background

In this research activity, with the background that many enterprises have a high interest in the incorporation of AI and are considering its incorporation, the checklist for introducing AI, which has

[80] For interview materials (excerpt), refer to <https://www.soumu.go.jp/main_sosiki/kenkyu/ai_network/02iicp01_04000232.html>.

yet to be started in Japan, was set as the subject of research by referring to the "AI Utilization Guidelines" and comparing the initiatives of Japan and EU for AI utilization.

c. AI implementation support tool as a deliverable

Composed of (1) AI Assessment List (Hereinafter referred to as "Assessment List"), (2) Instruction Manual for Use of the Assessment List, and (3) Check Items Explanation Document, from the viewpoint that anyone can easily incorporate AI, assuming users as AI service providers and business users. Further, the list can be narrowed down according to the position of the introducer and the incorporation phase.

d. Validation

In this research activity, feedback was obtained by having the Assessment List be used in the actual field (one manufacturing and one insurance company), and out of 118 check items, 84% were verified as valid items. Additionally, the improvement process was implemented based on items pointed out, such as overlapping questions. Further, by comparing the EU "Trustworthy AI Assessment List" and the Assessment List, it was confirmed that there were items only in the Assessment List (the reason is most likely items are covered by the GDPR in EU), while there were items only in the EU "Trustworthy AI Assessment List", and revisions were made to add these items to the Assessment List.

e. Future prospects

To make use of AI in business, it is necessary to have a "Guidelines for using AI" based "Self-Inspection and Self-Evaluation" system, as well as a "neutral evaluation" system.

As "Guidelines for using AI," the "AI Utilization Guidelines" and the Assessment List play the role of "Self-Inspection and Self-Evaluation".

The "neutral evaluation" can be an audit, but no specific statement has been made by any auditing firm as of this point. As with the evaluation of cloud services, we hope that the use of AI will expand in the future and that the system will be established in response to market demands.

Activities of this research group will continue and will promote the establishment of a system for self-inspection and self-evaluation of AI utilization in cooperation with the Conference based on the verification results.

(2) Discussion

[Expectations for future initiatives]

C. This is an excellent initiative.

C. We hope that this research activity will lead to the improvement of the AI Utilization Guidelines.

C. As a bottom-up approach in our country, it would be desirable to report this to OECD, etc.

C. In the EU, they put a lot of effort into the Assessment List, revise the principles through the Assessment, and try to create a solid database, and in that sense, they can contribute a lot.

9. Summary

(1) Significance of establishing AI principles

Enterprises that actively establish and utilize AI principles are sending out clear messages to society regarding concerns on AI development, such as "fairness" "accountability" and "transparency". Further, in establishing the AI principles, some characteristics reflect the corporate philosophy, etc. While the establishment of AI principles does not directly lead to corporate earnings, it does provide a basic corporate policy for AI development. This leads to dispelling concerns about AI development for corporate stakeholders and builds confidence in AI development and other initiatives. In light of this significance, the following are considered as possible future initiatives.

a. Follow up, disseminate, and publicize initiatives by enterprises to establish AI principles

As described in this chapter "1. Starting point for discussion", initiatives to establish AI principles have not been sufficiently spreading. From the viewpoint that these initiatives are beneficial to corporate activities, it is considered necessary to continue to collect case studies through interviews, etc., conducted by the Conference and to disseminate and publicize the significance of establishing and utilizing AI principles. Further, it is important to continue to disseminate the "Draft AI R&D Guidelines" and "AI Utilization Guidelines" to contribute to the establishment of the AI principles.

b. Follow up on trends in international discussions on the development of AI principles

Further, in establishing the AI principles, it is effective to make initiatives such as collaborating with overseas related organizations, because AI-related initiatives have a global nature. Therefore, it is important to follow not only domestic trends, but also trends in overseas and international discussions. However, it is not always easy for individual companies to follow such trends appropriately in a timely manner. With that said, the Conference must continue to follow and provide information on trends in overseas and international discussions. It is also necessary to actively communicate the status of our country's initiatives to the OECD and other international organizations in order to follow trends in international discussions.

(2) Use of AI principles in AI development and utilization

AI principles not only represent a company's philosophy on AI development and utilization, but can also serve as a guide for actual development and utilization of AI, thereby demonstrating its specific functions. For example, how to deal with fairness and trust in the development of AI, and how

to deal with trade-offs such as the relationship between explainability and accuracy are important issues, and principles become the premise for the judgment of these issues. In light of this significance, the following are considered as possible future initiatives.

a. Collecting, disseminating, and publicizing specific examples of the use of AI principles among enterprises

Knowing how AI principles are used in actual AI development cases, etc. will be useful information for enterprises considering the establishment of AI principles. In the future, it is important to share and refer to such information from the perspective of supporting business decisions in AI development and utilization as well. Therefore, it is considered necessary to continue to collect case studies through interviews, etc. at the Conference and promote information dissemination and information sharing.

b. Research on checklists as specific diagnostic tools

In order to utilize AI principles. in actual AI development and utilization, it is important to establish checklists, etc. based on AI principles from the viewpoint of ensuring objectivity, uniformity, and verifiability, etc. of the judgment methods for AI development and utilization. It can be said that the fact that the status of the development of such checklists was presented in this interview is a valuable reference. On the other hand, the EU has also established and tested an Assessment List and is making efforts to revise it around this summer based on the test results. In light of these trends, it would be beneficial for the Conference to continue to follow up on trends in the EU and other countries, collect case studies of the establishment of checklists in Japan, and conduct research on checklists as concrete diagnostic tools[81].

(3) Governance system required for safe, secure, and trustworthy AI development[82]

To ensure implementation of AI principles., governance (mechanism) is considered to be

[81] For example, the University of Tokyo has proposed a risk chain model that classifies risk factors for AI service provisioning into three layers: technical factors, normative factors for service providers' behavior, and factors for user understanding/behavior/usage environment. This model is being used to develop risk scenarios and visualize the relationships (risk chain) of relevant elements to consider incremental risk reduction and effective and efficient risk control, and to facilitate stakeholder dialogue and build up best practices.
(The University of Tokyo's Future Vision Research Center proposes a risk chain model to reduce the risk of AI services <https://ifi.u-tokyo.ac.jp/project-news/7079/>)

[82] For a summary of the implementation of the AI principles based on discussions on corporate governance, see Ministry of Internal Affairs and Communications, Information and Communication Policy Research Institute, Information and Communication Laws Study Group, AI Section Meeting FY2020 1st meeting, Member KOZUKA Soichiro (Professor, Faculty of Law, Gakushuin University) presentation material “Business implementation and corporate governance of AI development principles and utilization principles”. <https://www.soumu.go.jp/main_content/000689960.pdf>

necessary. As mentioned in this interview, there are some initiatives to improve the governance system, such as establishing an internal committee composed of diverse external human resources. Concerning governance, there are various possible forms of how it is secured and to what extent it is affected. Therefore, the following measures are considered as possible future initiatives.

a. Collection and publicizing of examples of self-inspection and self-evaluation initiatives

To disseminate initiatives for self-inspection and self-evaluation, the Conference must collect and disseminate examples of initiatives that serve as reference examples for the "Governance for implementation of AI principles". Further, startup companies in particular may hesitate to take such measures from the perspective of the priority of internal resources. Therefore, it is considered beneficial to hold meetings to exchange opinions with startup companies through cooperation with related organizations.

b. Consideration of external audits

No case studies of external audits have been reviewed. It is considered necessary for the Conference to continue to hold interviews with the parties concerned on the ideal form of external audits and to proceed with discussions[83].

c. Establishment of open forums to share the content of governance implementation and issues

As described above, there are a growing number of companies (and corporate groups) in Japan that develop and utilize AI principles and establish governance systems, and it would be beneficial to establish a forum for public discussion, such as holding a domestic symposium, to share the content of implementation and issues faced by each company. It is also necessary for the Conference to become one of the platforms for sharing the details of implementation and issues.

(4) Establishment of the "Best practices for utilizing AI"

While the significance of ethical initiatives and governance initiatives, including the formulation of AI principles, is gradually being understood, it has been pointed out that as a sense of the actual workers, it seems necessary to first gain the understanding of the significance and benefits of utilizing AI among end-users, including enterprises, who are business users. Since the understanding of the

[83] In December 2018, there were opinions that the draft proposal on profiling presented by the Personal Data plus α Study Group (<https://www.shojihomu-portal.jp/nbl1137pc>) would be a useful reference for examining checklists and external audits, and that an industry group (People Analytics & HR Technology Association) has published the principles of AI use in the area of human resources (HR Tech) (profiling) based on this proposal, and that it would also be a useful reference for application in other fields.

usefulness of AI utilization is a major prerequisite for advancing AI implementation in society, it is considered necessary for the Conference to actively promote AI utilization, hold interviews with people making use of it to improve management, and formulate an "AI Best Practices" that introduces the usefulness of AI utilization. In doing so, it is necessary to collect cases from a variety of industries without focusing on specific industries.

(5) Others

As described in the interview outline and the discussion above, this interview covered a wide range of issues concerning "safe, secure and trustworthy implementation of AI in society" with AI ethics as the starting point. Issues raised in this way will continue to be considered in cooperation with other relevant organizations as necessary.

Chapter 4 Initiatives by Business Users

1. Starting point for discussion

AI business users are defined as "End users who use AI systems or AI services on a business basis" in the AI Utilization Guidelines. As examples of such AI business users, in Report 2019, physicians providing services using AI services such as Medical AI Cloud Services (ex: medical services), financial institutions providing services using AI systems such as Credit Review System (ex: loan business), and manufacturers using AI services such as Abnormality Detection Service were mentioned. These are just examples, and AI, like ICT in the past, can be utilized in all aspects of industrial activities and social life. "Chapter 2: Prospects for the ecosystem formed as AI networking progresses" is also being implemented in anticipation of future developments in AI utilization in all aspects of industrial activities and social life.

As described above, AI can be used for a wide range of purposes, but on the other hand, as mentioned in the "Introduction" it is thought that not only AI developers and service providers but also end-users with relatively little expertise in AI have ethical concerns about AI. In order to promote the spread of ICT, it is important to facilitate the use and application of AI as well as to develop infrastructure. Therefore, to examine what are the challenges in advancing AI utilization and what measures are necessary to solve them, it was decided to conduct interviews among various types of business users on examples of active initiatives for AI utilization at present and examples of challenges faced upon utilization.

2. Sumitomo Mitsui Financial Group, Inc. ("The SMBC Group's Digitalization Initiatives")

(1) Interview outline[84]

a. Purposes and direction of AI utilization

There are 3 main reasons for utilizing AI. (1) Convenience. The company wants to maximize the value of the services it provides to the customers by using AI. (2) Efficiency. For streamlining and automating internal operations. SMBC would like to utilize AI for the automation of clerical work and application to illegal transactions and anti-money laundering. (3) Profitability. As a company, SMBC would like to use AI from the perspective of profitability.

b. Examples of AI utilization in financial services

As case studies, introduced (1) contact center support using AI, (2) chatbots, (3) advanced

[84] For interview materials (excerpt), refer to <https://www.soumu.go.jp/main_sosiki/kenkyu/ai_network/02iicp01_04000232.html>.

analysis, (4) detection of changes in business conditions of enterprises, (5) diagnosis of AI stock portfolio, and (6) correction technology of program failures by AI. Further, here are two other technologies that are being paid close attention.

(i) GAN (Generative Adversarial Networks)[85]

An example of using GAN to implement deep fake. As seen in the case of President Barack Obama, it has come to a level where counterfeits achieve high accuracy that even human beings cannot distinguish them as fake, and there is a growing concern about their misuse, which leads to a requirement of even higher ethical standards.

(ii) Explainable AI (Blackbox problem)

There is talk of becoming a black box, but there is a trade-off between interpretability and accuracy. There are active moves to improve the interpretability of black-box models and to improve the accuracy of highly interpretable models, and it is necessary to pay close attention to technological trends.

As an example capturing the essence of becoming a black box, introduced the case of the U.S. Army, which created an AI to distinguish enemy and friendly tanks. High accuracy was obtained in tests, but low accuracy was obtained when put into practice. Upon studying the training data, it was found that while many of the images of friendly tanks were taken on sunny days, many of the images of enemy tanks were taken on cloudy days, which resulted in the AI judging the situation by observing the weather. Since verification using the verification data set also produced good results, it was considered that there will be a reasonable degree of accuracy when new and unknown data of the actual data come. However, if the basis of judgment is not known, a situation may occur in which learning is made from elements completely different from the intention of the learner and the expected results cannot be obtained. An example that well represents the risks of using poorly interpreted models.

c. Challenges, guideline improvement status, etc.

Established guidelines for introducing AI at an early phase by summarizing risks specific to AI. Risks converge into four categories: (1) accuracy of results, (2) black box, (3) characteristics specific to each AI engine, and (4) data bias.

Clarified these risks at each stage of the AI implementation flow, (1) planning, (2) training, and (3) utilization after implementation and specified countermeasures.

[85] A framework in which 2 AI models, a generator (generate data that will be misidentified by the discriminator) and a discriminator (identify learning and generated data), are trained in competition to produce outputs with properties similar to learning data.

d. Composition of AI implementation guidelines

The first part describes the definition and risks of AI and points to note when implementing AI. In the latter portion, the AI implementation process is divided into stages from planning to release according to the AI implementation flow, condensed to important points and made into a guideline. Further, the points to be taken into consideration in order not to result in ethically inappropriate results, the relation with intellectual property rights, and the points to be noted in the contract with vendors when developing AI, etc. were organized. In compiling these documents, the "AI utilization Principles" (including draft version) issued by the Conference served as a reference and were used as the basis.

e. Future and challenges of AI utilization

(i) Spread of AI into all business systems and cooperation among AI

In the past, AI was introduced independently for each application, but in the future, AI will become commonplace in all business systems. For example, it is assumed that AI will permeate a range of business operations, including AI that predicts changes in business conditions of companies, AI that calculates credit based on those AI results, and AI that further proposes appropriate amounts of loans and financial products to be provided based on those AI results. This is also specified in the "AI utilization Guidelines" as the "principle of collaboration".

(ii) Organizational aspect

The implementation system is a very important theme for a company to actively utilize AI. The effective introduction of AI will not be discussed unless the division in charge of collecting technical information and verifying technology and the division in charge of business both promote the project integratively. To introduce not only AI but also other advanced technologies, it is essential to build a solid system in which each department of the bank promotes cooperation.

f. Future prospects

(i) Utilization policy

The company has been working on the use of AI from the early stages and has drawn up guidelines. In the future, the company will continue to actively utilize emerging technologies such as quantum computers, AR, VR, and speech recognition as well as AI.

(ii) Promotion system (As stated in e.(ii))

(iii) Human resource development

Awareness of the need to further strengthen and increase human resources who are well versed

in AI and advanced technologies. Training activities are conducted through participation in learning programs on the web and acquisition of certification by the Japan Deep Learning Association. In the future, it will continue to develop human resources in order to enhance the Group's AI knowledge and planning capabilities.

(2) Discussion

[Challenges and solutions for AI collaboration]

Q. What are the future challenges for AI collaboration?

A. When it comes to AI collaboration in which one AI receives an output of another AI as input, there is a possibility that the decision process of the final output given by the AI will be broadly black-boxed, leading to interpretability even lower than when incorporating a single AI. As measures to improve this situation, we believe that it is necessary to clearly establish a policy of countermeasures, such as introducing highly interpretable AI as much as possible or have people see what AI has provided as a flow of work, and minimizing the risk of erroneous decisions through collaboration.

[Trade-off between interpretability and accuracy, about explainable AI]

Q. It has been often reported that even a dead technology has come to be able to produce something similar to deep learning, and in some cases, it can produce data with almost the same degree of accuracy as deep learning with small data, and the black box can be somewhat alleviated. What is the future direction in total?

A. Regarding the tradeoff of interpretability, research has been conducted to increase the accuracy of conventional interpretive algorithms, which tend to be less accurate. Our chatbots are built using deep learning algorithms known as Recurrent Neutral Networks (RNN). In addition to the background that the development of this model was for internal use, this model is designed to do additional learning after being checked by bank employees, pursuing accuracy rather than interpretability. In this way, we choose AI algorithms on a case-by-case basis upon promoting the incorporation of AI.

[System's concept of text data]

Q. There are a lot of data that come in form of text, and I think how powerful the analysis of text documents is will definitely affect the performance of the system. What do you think about the analysis of text documents, and what is the course of action?

A. We have already introduced a text analysis platform and constructed a model to extract information that contributes to effective proposals to customers from, for example, negotiation records (text data) between sales representatives and customers. Upon training, correct answer

data labels are attached from several viewpoints from the enormous negotiation records until now.

[Problems with dialect speech recognition]

Q. With regard to call centers, how are you dealing with the problem of speech recognition of dialects and what kind of difficulties are you having?

A. In terms of training for the contact center, it was initially a problem that the accuracy of speech recognition for dialects and customer phone calls was not high. Therefore, as a mechanism of usage, both the voice of the customer and the voice of the operator were made into text, and the clear text that the operator has spoken was able to be selected as the text input to the AI.

[Utilization of External Data]

Q. In relation to the Code of Ethics, regarding the use of external data and trained models, you may be concerned about the content of the data and its bias. Is the data evaluated when utilizing such external data? If so, what kind of mechanisms and policies are available?

A. AI-OCR is a classic example of an AI model that's become smarter from a history of many cloud companies converting images into text and modifying them, and we know results are far more accurate than training with less data prepared by the bank. If possible, we would like to use the model as is, but we cannot see what kind of data it has learned, so we are considering how to capture the risks. However, we believe that OCR is only used to convert images into text, and we are planning a flow in which employees will always make corrections, and therefore believe that there will be no significant business risk. Conversely, if, as a different example, you bring in an AI model from outside to use in credit decisions, the risks are different. It is assumed that we will use external resources depending on the application.

[AI output and transparency issues]

C. What used to be standalone AI engine power could be integrated into a standard system in the form of embedded machine learning. We'll soon see a world in which we cannot identify at which stages' AI data output was used to generate the systems' generated data. If those systems further collaborate with other systems, it will become more difficult to see at what stage the AI output comes out and how the output is processed in a normal system. We are focusing on how to manage those risks and move on to the next step.

3. Tokyo (“Tokyo Metropolitan ICT-related measures”)

(1) Interview outline[86]

a. AI chatbots

- Bureau of Social Welfare and Public Health <Chatbot for inquiries about measures to prevent passive smoking>

Introduced AI chatbots to answer questions about Tokyo’s passive smoking prevention measures, such as the facilities subject to the Tokyo Metropolitan Passive Smoking Prevention Ordinance and the meanings of the terms in the ordinance (service launched January 2019)

- Bureau of Waterworks <”Suiteki-kun” advice room>

Introduced an AI chatbot to support access to information such as procedures for setting up water and construction information to improve services related to inquiries from customers (service launched July 2018)

b. AI chatbot general counter service (one-stop operation) (Planned release FY2020)

Have set up an integrated AI chatbot platform, which serves as a general contact point for all chatbots, and users can make inquiries to the general contact point and ultimately reach their desired answers.

c. Administrative reform

Regarding internal administrative work (salary and travel expenses, human resources, mutual aid, welfare, contract, accounting, goods, documents), which is common to all organizations regardless of the business they are in charge of, promoted the centralization of work and the improvement of efficiency through the use of ICT technology.

- AI utilization

Promote operational efficiency and productivity improvement by utilizing AI in clerical work and operations common to all organizations. For the time being, will promote initiatives targeting inquiries, preparation of minutes, and checking and proofreading of documents and materials.

d. Data conversion using RPA, AI-OCR, and ETL[87]

To promote the administrative use of digital data, RPA and other technologies will be utilized to make available the enormous amount of paper-based and other forms already owned by the metropolitan government, and will effectively utilize existing data assets.

---

[86] For interview materials (excerpt), refer to <https://www.soumu.go.jp/main_sosiki/kenkyu/ai_network/02iicp01_04000232.html>.

[87] Abbreviation for Extract, Transform, Load.

e. Support for preparation of AI minutes

The accuracy of the minutes, the frequency of use, and the cost of implementing the system were carefully examined on a trial basis within the metropolitan government, and efforts were made toward full-scale implementation.

f. AI support for operator operations

Implemented AI to the customer center of the Bureau of Waterworks to support operators by recognizing, documenting, and analyzing conversations and providing information such as possible responses. (Implemented starting February 2020) In addition, currently considering using big data obtained through AI for business improvement, etc.

(2) Discussion

[Dissemination of the use of AI to citizens of Tokyo]

Q. How are you explaining that the Tokyo Metropolitan Government uses AI to respond to the needs of its citizens?

A. In terms of how we have incorporated AI, the current state of AI being used by the Tokyo Metropolitan Government Office is in such a way that the person using it can recognize that it is AI (the user can see it is a chatbot or a machine). However, even if the AI application phase becomes a broader and more general function, the involvement of people at the end of the line is an important point, and the metropolitan government believes that face to face interaction is very important. In that sense, as the use of AI becomes more widespread and general in the future, we have recognized that we have a need to explain how we are using AI.

[The way of using AI]

C. If you expect AI to be 100% accurate, I think it will slow down the use of AI.

C. Even if it is not 100%, we would like it to be incorporated more and more if the administrative system will be improved. It is not easy to go to the city hall during daytime on weekdays. If it is a system where the user can use the system if they recognize the system is run by an AI and understands the risks that there may be mistakes, we believe there will be a lot of people who would like to use it. As a user, it would be great if you can start from anywhere wherever possible.

A. We don’t think AI will replace everything, but we believe it can be used as one type of multi-channel tool and be utilized as one of many tools and methods. At this stage, we think it is important not to burden everything with AI, including chatbots.

4. Yamaha Corporation (“Yamaha’s AI Singing Synthesis -Initiatives to Revive Hibari Misora-”)

(1) Interview outline[88]

a. Background and premises

Has been working on synthesizing singing voices for more than 20 years, and developed a greatly advanced technique by introducing a completely new AI technique last year. Using this, cooperated in the project to restore MISORA Hibari’s singing voice (MISORA Hibari: a Japanese singer, actress and cultural icon, deceased 24 June 1989). The company has started working on VOCALOID and AI technologies, but has not come to the point where enough discussion has been made on how to use AI and how to think about AI ethics on a company-wide basis. In this presentation, the point will be on the company’s thoughts while engaging in singing voice synthesis.

b. Differences between the current VOCALOID and VOCALOID:AI

In contrast to the current VOCALOID, which combines pre-recorded sounds, VOCALOID:AI uses a deep neural network to learn the correspondence between a large number of scores and singing voices in advance, so that when a new piece of music arrives, its corresponding singing voice can be estimated by the deep neural network. The point superior to the former is that people can request musical intent to the VOCALOID:AI.

c. Requesting musical intent

Mr. AKIMOTO Yasushi requested the company to create the atmosphere of Ms. MISORA Hibari who is making a comeback after 30 years, or to synthesize a song that feels it is sung to each person. If you submit these requests directly to VOCALOID:AI, it will not work as well as expected. There are two places where you can request human intentions in this system. One way is from the many songs of MISORA Hibari used for training, to request to sing in a style close to a specific song. The other is a more basic approach, in which the training data are limited to songs with certain desired characteristics.

(i) Music style requests

Based on Mr. AKIMOTO’s request, interpreted that the singing voice in the later years like that of a song “Ai Sansan” shall be appropriate, and the singing voice shall have characteristics like that of kindness and richness. This song was synthesized with somewhat musical intentions such as Mr. AKIMOTO and our way of thinking.

[88] For interview materials (excerpt), refer to <https://www.soumu.go.jp/main_sosiki/kenkyu/ai_network/02iicp01_04000232.html>.

(ii) Atmosphere reproduction by data selection

In order to synthesize the narrative part of the song, we borrowed Ms. MISORA Hibari's precious voice data pre-recorded and kept for her son and finally synthesized the narrative part of the song based on this data. Even if the text is the same, the meaning of what is transmitted changes depending on the atmosphere of the way of speaking. Instead of just generating Ms. Hibari's voice after inputting some data, choices were made based on the musical intentions of Mr. AKIMOTO and us.

d. About VOCALOID:AI

They sing in the context of the musical score, but on the other hand, can only see the context of the musical score. For example, it is impossible to change the way of singing to match the atmosphere of the accompaniment like a real singer does as of now. It is impossible for the AI to understand by itself the historical context, the intention of the project, and the intention to stand in front of the fans whom she cared about for the first time in 30 years and delivering a message. So, these parts were synthesized by reflecting Mr. AKIMOTO's intentions and producing intentions by requesting musical intentions, selecting music style and using selected data.

e. How AI singing voice synthesis technology was conveyed and interpreted

There was an impression that there was a difference in reaction on SNS right after this project was broadcasted in NHK Special (TV program) and right after it appeared in Kohaku Utagassen (annual New Year's Eve TV marathon song festival, also referred to as "Kohaku")

Right after the NHK Special, very positive reactions such as great, moving, funny, wonderful, and good were dominant, but right after the Kohaku, quite negative reactions such as fear, disgusting, punching, feel sick, and "blasphemy" were common.

f. To what is the "blasphemy" being felt?

Analyzed what people who use the word "blasphemy" is feeling blasphemy towards. Possibilities are (i) blasphemy to the dignity of life, (ii) blasphemy to the dignity of the personality, or (iii) blasphemy to the dignity of art or creative activity. What we think about this, and why were these reactions not presented after the airing of the NHK Special are to follow.

(i) Dignity of life

Perhaps there was an impression that AI is something very mysterious that is imitating human beings or life. In the NHK Special, the mechanism of the AI was clearly delivered, so most probably this impression was corrected. As with the term artificial intelligence, we tend to use words that are likened to the activities of life, and we have to be careful that we may get the impression that it is a kind of imitation of life that we cannot fully understand.

(ii) Dignity of the personality

One big difference between NHK Special and Kohaku is whether the listener/receiver were mentally prepared. This mental preparation is for listening to a reproduction made by AI technology. The difference may be whether or not there was a mental preparation that what is coming up is not Ms. Hibari herself, but that we shall enjoy the mood of it.

(iii) Dignity of art or creative activity

This is a question of the attitude of the creator, or whether the creator and the listener share deep respect. This project was carried out with that intention. If this was a project where Ms. Hibari's data was thrown into a black box called AI and something that resembles her is outputted with a conclusion that AI is amazing and impressive, it may have appeared that the project looked down on artistic and creative activities, and I might have felt that way myself. In this project, everyone wanted to further understand Ms. Hibari's music, get closer to her music, and reproduce it, and we believe that at least there is nothing blasphemous about this attitude.

g. Where we are going

The project is one example of how we want to move into a future where people can create music with AI.

There was difficulty in conveying in terms of both technology and ethical demarcation, but we would like to continue research and development with hopes that it will be possible for these technologies to be passed on to people and used in a truly good way in the future.

It seems that there is a tendency that AI is perceived as something undecipherable, and that this tendency is preventing the use of the good aspects of AI. Firstly, we want everyone to understand AI correctly, including the negative aspects, what we can and cannot do. Then, we would like to establish a system to actively promote its good use. We would be very happy if we could discuss this topic.

(2) Discussion

[VOCALOID:AI PR method]

Q. It was mentioned that there was little antipathy after the NHK Special because they were informed about the development process in the documentary, but on the other hand, there was no such thing at Kohaku which led to antipathy, so I wonder what kind of PR you are thinking about when putting VOCALOID:AI out into the world in the future.

A. If it is an AI synthesis of a living singer, we believe it is relatively easy to clear ethical issues as long as the singer itself acknowledges it. In such a case, it can generate songs sung in the singer's style depending on the context of the score, and it is also possible to enter the creator's intentions

to make it into a work of their own. We hope that this will move in the direction where more people, such as people who can't hire a real singer, are enabled to send out better music to the world, of course where all parties are agreeing.

[Support as the responsibility of the manufacturer/developer]

Q. (If the customer) has the same affection for robots and VOCALOID as you do for real people and pets, I think there is a possibility that you will not be able to stop product production and service provisioning. What was introduced today was mentioned that it will not go into product development, but what do you think about the possibility of it?

A. Certainly some people may be sad that we stop software support and people can no longer use it to synthesize. It's good to hear opinions that there could be some developer responsibility.

[Dignity of the personality]

Q. I think that the newest thing this time is that a topic about the dignity of the personality was raised. As your company aims, to impress people means to move people's minds, and in other words, I think it can induce people's values. In particular, in the case of the deceased, the person themselves cannot appear and deny it. Therefore, I think that the origin of discomfort is that something that has too much influence on society can be created and used by someone, and that is the important point. I'd like to hear your thoughts on that.

A. We believe you are right. There are many examples where voices are generated in President Obama's voice saying inappropriate things, so we think that's a problem. Now, it seems the responsibility is held on the ethics of developers. On the other hand, it cannot be stopped technology-wise, and we would like to continue the development of technology in the sense that it has positive aspects as well. Perhaps we have no choice but to think about how to use it and discuss what we can and cannot do. Now that these new technologies are born, perhaps the formation of ethics based on them should be done at the educational level.

[Relationship with accountability]

C. In relation to the "Draft AI R&D Guidelines" and the "AI Utilization Guidelines", accountability and other issues are discussed with multiple stakeholders.

[About videos and sound]

Q. Do you think there was a difference in the effect (reaction) between the video and the voice on the reaction to the AI MISORA Hibari?

A. If you say that the audience was impressed only by the sound, it is probably not like that, and there is an aspect that the audience was impressed because it was as if Ms. Hibari was revived on

the stage of NHK Special, and that is why it was moving even though the audience knew it was in an entertainment space, or other words, knew that it was only an AI. Most probably the sound stood out better thanks to the video. On the other hand, if the listener feels that AI is creepy, we wonder whether they are thinking that the video is AI or not. People who can identify what is AI-made and what is not were having discussions in a calm matter commenting how the audio was, and how the video was. But those that cannot understand may have thought that the very image of that incomprehensible something has created a mysterious AI. We have to admit that the impact of video is greater than that of sound alone, so we may have to be more careful.

[Prevention of voice abuse]

C. For example, if the technology develops in the future and you can enjoy using your voice as a song, it is fine, but for example, it is possible that your voice may be used in an unexpected way and used for something outside your intentions, so I would appreciate it if you could develop the technology including preventive measures and security.

A. Even in cases where you are not thinking to leave your voice for AI use, it could be used by someone in some unexpected way. It is really becoming possible to extract the characteristics of a person's voice from a small voice sample on a video-sharing site, so it's strange to say to have people be aware of the sound they produce, but it may be a necessary awareness. For example, many people are conscious of photos, and some people say that they should be careful of uploading photos with their faces because you don't know how they can be used after posting on the Internet. On the other hand, there has been a lot of technical work done to identify whether a sound is synthetic or real. Maybe this is what we should be developing at the same time, but this is really a game of cat and mouse, and there are ways to train the synthesized voice using systems that detect synthesized voices (by using GAN, etc.), so the development is really in a cat and mouse situation.

[Problem of distinguishing between real and fake]

C. Perhaps the real fear is that humans cannot identify whether it is real or synthesized, so together with the development of AI, it's best to make it so that it can be recognized that it is synthesized, because it is most likely impossible to identify.

A. Fundamentally, we think it is the responsibility of the creator to make clear that this is made by AI, and that they should enjoy it while knowing the fact. On the other hand, in the aspect of entertainment, there is a feeling that we don't want to do that. For example, we think some people may feel that the fun is spoiled if we make a regulation to announce that "this is AI-generated, please keep this in mind when listening" before airing the work. In the case of TV dramas and

movies, it mentions that the work is fiction after you are done watching, and we hope we can aim for something similar to this.

[The rights of the voice]

C. Is the singing voice a personal right or a related right? If it is a related right, it can be an object of inheritance, and it would hold monetary value. On the other hand, if this is a personal (right) of MISORA Hibari, it cannot be anything other than MISORA Hibari. Perhaps the first task is to make a distinction in this regard. Further, the created AI may become a personality itself, learning and becoming smarter. Whether to allow such changes is an important issue in the scope of rights. To what extent can the bereaved family control it? Or, I think it would be better to organize whether it can be used to make money in the entertainment business. This is a very good time to create a map because complex relationships of rights can become entangled.

5. Anonymous

(1) Interview outline

a. Evolution of security systems and AI

Security systems have evolved to a high level that handles drone use and image analysis by incorporating technology evolution such as high-capacity transmission and sensors/cameras. The background is that high-definition (4K) cameras are getting cheaper, AI that can be used in practice is born, and the launch of 5G. Security models will be changed with this opportunity.

The business model so far was all post-processing, either by watching the monitor all the time by a guard or by watching the camera image stored on the hard disk afterwards. With this approach, it cannot cope with unexperienced incidents such as soft targets and truck terrorism, so we need a business model that considers how to predict incidents in advance. We are also making initiatives to incorporate AI because of the worker shortage. Every industry is our customer, and we are working hard to develop AI based on the various needs of our customers.

b. Case (1) Automation of true/false alarm judgment

AI determines the presence or absence of a person from the alarm image and assists the guard in determining true alarms. In the future, we aim to develop the “support” of identifying true alarms into “automated judgment.”

We are promoting this system as a means of increasing operational efficiency and introducing it voluntarily.

c. Case (2) X-ray inspection assisting AI (Efficiency and laborsaving with X-ray

inspection work AI)

Research that aims to improve the efficiency of X-ray inspection which is now done by experienced inspectors by using AI.

d. Case (3) Suspicious behavior (shoplifting) detection AI (development of a suspicious person (shoplifting) detection system using a security camera)

(i) Model

Conducted in response to customer needs to find shoplifters. In the model, a camera is set up, and when the AI detects an image from the camera as a suspicious person, the staff, employee, or security guard is notified via smartphone that a suspicious person is seen in the number X camera, and they will approach and speak to the customer to prevent the incident.

(ii) Challenges for social implementation

Sharing information about a suspicious person among multiple stores in the same industry is technically possible by face authentication and there are needs for it. There is also a need to share that information between other industries and creating a shared blacklist of suspicious persons. However, there is a problem that at the present time, this can be done only in the same store, and there are multiple requests from customers to share it among multiple stores in the same industry or across industries.

e. Case (4) New hospitality service

(i) Model

Conducted a demonstration experiment of a “hospitality service” in the Shin-Marunouchi Building, where an AI analysis of camera images detects the movement of “people who want help” and notify security guards.

(ii) Challenges for social implementation

In this demonstration experiment, since it was conducted in an open space, there were problems such as what to do with the people in the camera images and whether to perform face authentication. Procedures were made based on the protection of personal information in this case, but how to implement this on a more open scale is an upcoming challenge.

f. Case (5) Ministry of Internal Affairs and Communications 5G comprehensive demonstration experiment 2017

(i) Model

This system enables wide-area monitoring with a high-altitude camera, detection of the

occurrence of fires and associated traffic jams, and detection of their locations, thereby enabling optimal fire-fighting activities by firefighters, and appropriate notice of evacuation guidance by municipalities, etc.

(ii) Challenges for social implementation

While the initiatives of E. are made from a micro point of view and from the perspective of personal information, this is in a macro scale of open space, for example, whether a building in the picture is really allowed to be on camera, or the model of the passing car can be identified, so it is technically possible, but challenges are how to implement this in society.

(2) Discussion

[How to address the challenges]

C. Facial recognition is a very difficult issue, and there are various ways of thinking about how to deal with it. We think it would be better for your company, which is facing problems on the spot, to raise the issue because it would be a certain fact when the government discusses it. Also, there are hearings on the revision of the Act on the Protection of Personal Information, etc., and we think the easiest route for the opinion to be reached is by your company to raise the issue. We would like your company to voice opinions in various ways.

[Human intervention and course of action of business in AI utilization]

Q. This is never concluded with just AI, and there is always human intervention. Can you tell us whether introducing AI systems will eventually save the company's overall labor or increase the new variety of cases to handle?

A. Both labor-saving aspects and new handling cases exist depending on the work. Rather than slowly decreasing the number of people positioned in 2400 locations, it is more like having those people conduct multitasking like monitoring social infrastructure. The idea is to decrease the time spent on current basic work so that the multitasking portion can expand.

6. Mr. Hiroyuki Sanbe, a member of the Conference ("Rikunabi Case Reviewed from the Perspective of AI Utilization")

* Translation Note: Rikunabi Case is a famous case where Recruit Career Co., Ltd. and companies which received certain information from Recruit Career were subject to administrative despositions. Recruit Career operated a website called "Rikunabi" where job seekers, typically students of universities and colleges, may obtain information of companies and may apply for recruiting process of the companies. Recruit Career illegally provided services called the "Rikunabi DMP Follow",

where, among other things, Recruit Career calculated score of each student's declining informal job offers with algorhythms and sold the score to Recruit Career's client companies. These client companies purchased this "Rikunabi DMP Follow" because they would like to prevent the situation where they offer jobs to students who will later decline the offers.

(1) Interview outline[89]

(Premise) At the request of the Secretariat of the Conference, the presentation will be done under the title of "Rikunabi Case Reviewed from the Perspective of AI utilization". We would like to ask for your understanding that we have no intention of criticizing the companies involved and would simply like to share the points of discussion in light of the activities of the Conference.

a. Four fundamental differences between AI and traditional businesses

First, risks occur from the time when data is collected.

Second, AI is inductive and hard to predict results of AI's reasoning in advance, so you have to cope with the results.

Third, AI may conduct things which have been traditionally difficult, and may substitute human judgment.

Fourth, many points are unexpected and not covered with existing laws, including those from the Meiji period, and AI may do things outside the scope of the law.

These differences make those who are working in traditional business methods get into unfamiliar territory, and make it difficult to be aware of the laws and stakeholders involved. Also, laws and ordinances are not yet established to cope with these differences. There are many issues that need to be coordinated with the competent authorities, but there are also situations where it is difficult for a company to deal with the issues because the issues do not come to the company's mind at first.

b. Key points of the Rikunabi case reviewed from the perspective of AI utilization

Based on the above, the following four key points are important when looking at the Rikunabi case.

The first key point is that "We must consider laws, legal issues, and regulatory authorities". The point here is that even if AI businesses and traditional ones are based on the same laws, AI will have different problems. Therefore, coordination with the supervisory authorities is necessary. In other words, we need to be aware that "issues arise because it is AI."

Secondly, "We need to consider ethical and reputational risks".

---

[89] For interview materials (excerpt), refer to <https://www.soumu.go.jp/main_sosiki/kenkyu/ai_network/02iicp01_04000232.html>.

Third, "We need to be aware of the breadth of stakeholders involved", which is unique to AI businesses.

Based on these points, the fourth key point is that the challenge will be "efforts and activities within each company". Since the efforts and activities are related to "governance", it is very important for each company to consider their internal corporate systems and organizations.

Hereinafter, the Rikunabi case will be examined from these four viewpoints (see below sections C. to F. The fourth point will be discussed before the third point.).

c. On the point that "legal, regulatory issues, and regulatory authorities must be considered"

- In the case of Rikunabi, what the Personal Information Protection Commission is most concerned about is not the lack of consent from job-hunting students. Of the three facts that led to their administrative disposition in August 2019, two were based on the fact that safety control actions, which have been required under the Act on the Protection of Personal Information, have not been taken, and these points were discussed earlier than the fact that consent had not been obtained. As a result, the most important issue was the governance system such as organizational structure and company-wide awareness. This is an important point here.
- Further, the Tokyo Labor Bureau (the Ministry of Health, Labour and Welfare) issued their administrative disposition pursuant to the Employment Security Act. Under the Employment Security Act, certain companies, such as employment placement business providers (in other words, employment agencies), shall, in collecting, retaining and using the personal information of job seekers with respect to their businesses, collect the personal information of job seekers within the scope necessary to achieve the purpose of their businesses and retain and use the same within the scope of the purpose of said collection. Therefore, it is not just a matter of the Act on the Protection of Personal Information. This is another important point here.
- Also, the Japan Fair Trade Commission released new guidelines at the end of 2019. Unfair acquisition or use of personal information may constitute an abuse of a dominant bargaining position and constitute a violation of the Act on Prohibition of Private Monopolization and Maintenance of Fair Trade (*i.e.*, Japanese antitrust law). At least from now on, these points must also be considered.
- For the foregoing reasons, it is wrong to recognize that "the issue in the Rikunabi case is that Recruit Career have not obtained the job-hunting students' consent which was required under the Act on the Protection of Personal Information," and that is merely one of the issues. The points are that governance is also an issue under the Act on the Protection of Personal Information, and that other laws are also involved.
- When it comes to other laws, there is also an issue related to the Constitution, and in theory, there

is a possibility that Civil Code violation may be admitted.

- The Companies Act, the main pillar of corporate governance, will also pose a problem. Under the Companies Act, it is possible to raise the concern that Recruit Holdings, the holding company of Recruit Career, may have had problems in managing its subsidiary. The Financial Instruments and Exchange Act also becomes related. Major shareholders of Recruit Holdings sold shares of the company by way of the Secondary Distribution as defined in the Financial Instruments and Exchange Act. It is important to understand that governance issues and market laws like the above also have an impact.
- The EU's GDPR (General Data Protection Regulation) also affects discussions in Japan. In January 2019, Japan received the so-called adequacy decision from the EU. To maintain this decision, it is expected that Japan will take a severe attitude toward business operators regarding the protection of personal information.
- AI businesses involve almost all of the laws. In the case of Rikunabi, the point here is that not only the Act on the Protection of Personal Information is in question, but also legal problems not found in the past business occur or are likely to occur, even if the same statutory law is applied.

d. On the key point that there is a "We need to consider ethical and reputational risks"

- The Rikunabi scandal concerns the Constitution's principle of respect for individuals. In August 2019, the Personal Information Protection Commission issued administrative dispositions stating that Recruit Career "has a grave responsibility for the proper treatment of the information, because it can affect the future lives of the students." The Act on the Protection of Personal Information and its guidelines do not explicitly state that "this is how you should treat information that can affect people's lives," but when a company improperly handles important matters relating to the principle of respect for individuals, supervisory authorities will pursue the company's responsibility. Therefore, it is important to bear in mind that what is said in the context of AI ethics, including constitutional principles, can be incorporated into the legal interpretation and used to determine whether a business is appropriate or not.

e. About "efforts and activities within each company" and "governance"

According to administrative dispositions issued by the Personal Information Protection Commission in December 2019, Recruit Career and other companies are required to "review the organizational structure and bring about a drastic change in awareness throughout the company, starting with the management". The 35 companies that have become clients of Recruit Career were required to take actions including "systematic legal review and necessary actions". Further, Recruit Career was required that "when it acquires personal information, it shall specify the content of goods and the like as much as possible, and shall appropriately inform a data subject,

or disclose to the public, the utilization purpose of the personal data." It is not necessarily enough to simply publish a privacy policy or have users click "accept."

For this reason, companies that will be clients of AI products and services must also consider internal control systems concerning protection of personal information which are adopted by companies developing or utilizing AI. Now it became important for clients of AI products and services to incorporate the "process of considering internal control systems concerning protection of personal information which are adopted by companies developing or utilizing AI" into their business processes.

Further, even if the company is not a large company and the like and is not required to establish internal control systems under the Companies Act, we have entered an era in which mishandling of personal information could violate other laws such as the Act on Prohibition of Private Monopolization and Maintenance of Fair Trade (again, Japanese antitrust law) which I have just mentioned earlier. Companies that develop and/or utilize AI, as well as those that receive it, must consider governance issues in relation to various laws when dealing with data and AI. This is also an important point here.

f. On the key point that there is a "We need to be aware of the breadth of stakeholders involved"

When I look at the case of Rikunabi, many people tend to focus on the fact that Recruit Career provided their client companies with information about job seekers without the job seekers' permission, but that's not all. Universities, shareholders (of Recruit Holdings that is a listed company), supervisory authorities, and employees and other companies of the Recruit Group unrelated to the case are also stakeholders. Furthermore, the client companies themselves have stakeholders. Moreover, the mass media reported it day after day.

Looking at it in this way, the range of stakeholders tends to be very broad in AI businesses. When you start an AI business, it is essential that you go through a process where you consider what legal, ethical, and social issues are associated with each stakeholder, at least in the early stages of the business.

The client companies have also received administrative dispositions from the Tokyo Labor Bureau and the Personal Information Protection Commission, and this fact has been publicized and the companies were criticized by the public. Therefore, a negative impact may have been possibly generated in the relationship between Recruit Career and the client companies that Recruit Career wanted to value.

In this way, the breadth of stakeholders unique to AI businesses is also an important point.

g. Summary

There is a wide range of laws, ethics, and stakeholders related to AI businesses. The risks are likely to materialize if you proceed with AI businesses in the same way as traditional businesses. To avoid such a situation, it is necessary to establish a governance system to consider stakeholders and legal and ethical issues. Because each company has different businesses and governance structure, how they deal with issues related to AI also varies greatly. Based on these points, I would like to contribute more and more by providing feedback on practical ideas and issues with the Conference when the Conference considers future initiatives.

(2) Discussion

[Comments on implementation of AI guidelines and governance]

C. I completely agree with this explanation, as I have been thinking from the start that implementing "AI Guidelines" would lead to corporate governance.

[Q & A on the concept of the information use itself in this case]

Q. What do you think about whether negative evaluation of the use of the information itself was conducted in this case? As a job seeker, I believe there is a desire to be free to decline an unofficial job offer. On the other hand, there was a need for companies to predict such a situation. Has there been recognition in this process that the use of such information itself is a problem?

A. Recently, a "job offer declination set", which consists of a series of goods which may be useful when a student declines a job offer of a recruiting company from a point of view of courtesy in Japan, has been put on the market. Both the Rikunabi case and the job offer declination set suggest that the products and services related to declining job offers are quite a large market. It is clear, however, that the market is based on the premise that people are allowed to decline job offers. That is why it was a real problem for the big companies that purchased the Rikunabi DMP Follow.

In light of this situation, what the Personal Information Protection Commission raised as problems were the lack of appropriate process which led to non- explanations to job seekers and the lack of proper consent from them, and the lack of effective governance for the appropriate process. In other words, I recognize that the Personal Information Protection Commission has not made a clear statement as to whether it is definitely unacceptable to provide the third parties with the data of the score showing each student's declining rate or to utilize it.

With that said, as a general statement, it is hard to say what companies utilizing or developing AI should have done in the Rikunabi case. I would like to refrain from expressing my personal opinion on whether or not the utilizing the data of the score showing a student's declining rate should be admitted.

However, it is important to consider how to approach cases where AI and data may be utilized

and at the same time no social consensus has been reached for such utilization.

In the Rikunabi case, the interests and needs of those who obtain data and those who are taken data are completely different, and the interests and needs of those who hire and those who seek employment are also completely different, and therefore, no social consensus has been reached. I believe that it was necessary to deepen awareness of whether or not society would accept the business that the company wanted to realize by considering the interests of opposing stakeholders or by fostering awareness of how to think legally and ethically within the company or in society. If we go ahead with AI businesses without this kind of process, problems like the Rikunabi case are likely to occur.

[Q&A on understanding personal information]

Q. What we are very concerned about is that personal information is tied to individuals. Even if it is estimated information or false information, it becomes personal information. I am very concerned whether everyone understands such a realistic scheme. Rather, I am worried that there may have been a feeling of the people in the field that they treated it as if it was not personal information because they were not aware of it. I think this will be an important point for other companies as well.

A. I believe you are right. The Personal Information Protection Commission, in its December 2019 administrative dispositions, said that it was problematic that Recruit Career provided information to its clients without the consent of the job seekers, knowing that, although data held by Recruit Career alone would not fall under personal information, if it were provided to the clients, it would be possible to identify a specific individual by being combined with the data owned by the client, whereby the provided data would be transformed into personal information and personal data as defined in the Act on the Protection of Personal Information. This has been a point at issue when interpreting the Act in the past, and legal practitioners have recognized that this was such a point. The Personal Information Protection Commission said that, in the foregoing situation, Recruit Career's providing data to clients without consent of job seekers is an "evasion of the law".

Recruit Career also seems to have judged that hashed information may not fall under personal information under the Act on the Protection of Personal Information. Because hashing is not even encryption and it is still possible to identify a specific individual, the company's view is not considered valid under the Act on the Protection of Personal Information. In fact, the Personal Information Protection Commission has also pointed out that Recruit Career's views are wrong.

Following these discussions, for example, in the "Act on the Protection of Personal Information 'The Every-Three-Year Review', Outline of the System Reform" published in December 2019, it was clarified that restrictions on the provision of personal data to third parties would be applied to cases where, although relevant information does not fall under personal data at a data provider, it

is clear that the information to be provided to the other party will become personal data at the other party[90]. Likewise, in the future, there will be more problems that will be highlighted especially because what is utilized is AI. As new problems related to AI and data will appear more and more in the future, we need to make efforts to foster awareness of these problems.

[Q&A on the use of data such as of education and learning]

Q. In the future, universities will analyze student data and create portfolios like at MIT and the National University of Singapore to see how they can contribute to raising the market value of their students. For example, they may become able to advise based on the student's portfolio that they have not completed certain courses yet, that it is not good and they should study more to raise their education level. I think companies will ask to submit a certificate of completion upon hiring. Please tell me anything that comes to mind in this area.

A. Putting it into the context of Japan would require a lot of analysis, so I refrain from giving an immediate answer, but I often hear people saying that it is difficult to utilize data related to education and learning. I have been appointed by the Cabinet Office and the Ministry of Education, Culture, Sports, Science and Technology as a member of the committees on learning support technology and university initiatives in relation to Society5.0, and I have also been teaching at universities, so I have met many people in the educational field and university, and I believe that there are many people who are interested in the utilization of data. Some municipalities are also concerned about the legal system concerning data, including the issue of so-called 2000 personal information protection ordinances. If there is any useful discussion that can be made public on how to promote utilization, I would like to provide information at the Conference.

It is also necessary to consider how to incorporate information that has not been made into digital data. It is not a matter of law, but there is a debate over whether it is hard to make good use of information of students unless we make each and every student use a tablet. I would like to

[90] The Act on the Protection of Personal Information was amended through The Amendment Act of the Act on the Protection of Personal Information, etc. (Act No. 44 of 2020), which was published after the interview of our member Mr. Sanbe, and was approved in the 201st Session of the National Diet. After the enforcement of the said amendment, in cases where a business operator intends to provide personally referable information (*i.e.*, information relating to a living individual which does not fall under any of the categories of personal information, pseudonymously processed information, or anonymously processed information) to a third party and it is assumed that the third party will acquire said personally referable information as personal data, unless one of the exceptions apply, the business operator shall not provide said personally referable information to said third party without confirming, in advance, that the data subject's consent (to the effect that the data subject approves that the third party acquires the personally referable information as personal data which can identify the data subject) has been obtained (and also confirming certain more facts depending on the situation).

keep my interests high and give feedback in the future.

* Translation Note: The issue of so-called 2000 personal information personal information protection ordinances refers to the issue where transfer and utilization of personal information held by municipal governments, hospitals, schools and any other municipal agencies have been difficult because municipalities have ordinances with different contents and municipalities have managed the ordinances differently. The issue was recently addressed by the act amending relevant statutory laws. This act was approved by the National Diet in 2021, and it is expected that this act will be enforced against municipalities in 2023.

7. Summary

(1) Supporting AI business users in establishing AI Utilization Principles and Guidelines

Only a portion of AI business users has developed AI Utilization Principles and Guidelines[91]. There are various purposes for enterprises, etc. to utilize AI in their business activities, but the formulation of the principles for AI utilization has the advantage that, concerning AI developers and service providers, it will become easier for enterprises, etc. to promote the development of AI systems and services in cooperation by clarifying the basic concept of AI utilization. In the case of in-house use, it will show that the company is utilizing AI safely and securely concerning their employees, etc., and most importantly, in the case of providing products and services as a business by utilizing AI, it will be an important message to gain the trust of customers that they are receiving products and services safely and securely. It is considered necessary for the Conference to support AI business users in establishing the AI principles and guidelines by following up on their ambitious initiatives to establish the AI principles, introducing such advanced cases, and continuing to disseminate the "AI Utilization Guidelines".

(2) Compilation and accumulation of ideas on each principle of AI utilization through specific case studies

In this interview, the views on each principle of the AI Utilization Principles were also discussed through actual initiations for AI utilization. Although this interview was not an exhaustive discussion due to the limited scope of the subject fields of the hearing, we consider that organizing and accumulating the ideas on the principles in such specific cases will serve as a reference for companies when they make decisions on AI utilization in the future. From this perspective, it is considered

[91] Aforementioned, see footnote 70 Ministry of Internal Affairs and Communications, Information and Communication Policy Research Institute, Information and Communication Laws Study Group, AI Section Meeting FY2019 1st meeting, Member SHIMPO Fumio (Professor, Faculty of Policy Management, Keio University) presentation material "Do AI principles work?" <https://www.soumu.go.jp/main_content/000660996.pdf> See p6 and beyond.

necessary to continue holding interviews on ambitious initiatives, including areas that were not included in this interview, and to sort out and accumulate actual ideas on each principle in specific case studies.

(3) About the establishment of "Best practices for utilizing AI" (Same subject as Chapter 3, 9. (4))

In Chapter 3, 9. (4), In Chapter 3, paragraph 9. (4), the subject is written from the perspective of AI developers and service providers, but it can be said the relationship between the establishment of AI Utilization Principles and Guidelines by AI business users and the progression of AI utilization initiatives by enterprises itself are a chicken-and-egg relationship. From this perspective, it is considered necessary to work on the formulation of a "Best practices for utilizing AI" that will serve as a reference for AI utilization initiatives.

(4) Follow-up of institutional challenges necessary for AI utilization

In providing products and services using AI, it may be necessary to reorganize and review the relationship with existing systems, such as voice copyrights. It is considered necessary to continue to follow up on these institutional challenges at the Conference by cooperating with the AI Section Meeting of the Information and Communication Laws Study Group and other relevant organizations.

(5) Importance of governance in businesses utilizing AI

There are many laws, ethics, and stakeholders involved in AI businesses and the risks are likely to materialize if companies proceed with AI businesses in the same way as traditional businesses. To avoid such a situation, it will be necessary to establish a governance system to consider legal and ethical issues and stakeholders. Further, when considering governance systems, because each company has different businesses and governance structure, how they deal with issues related to AI may also varies greatly[92]. With this in mind, it is necessary for the Conference to continue to follow up on and study governance systems for businesses utilizing AI.

(6) Other

In this chapter also, as described in the interview outline and the discussion above, a wide range of issues concerning "safe, secure and trustworthy implementation of AI in society" were raised, like Chapter 3. Issues raised in this way will continue to be considered as necessary in cooperation with other relevant organizations.

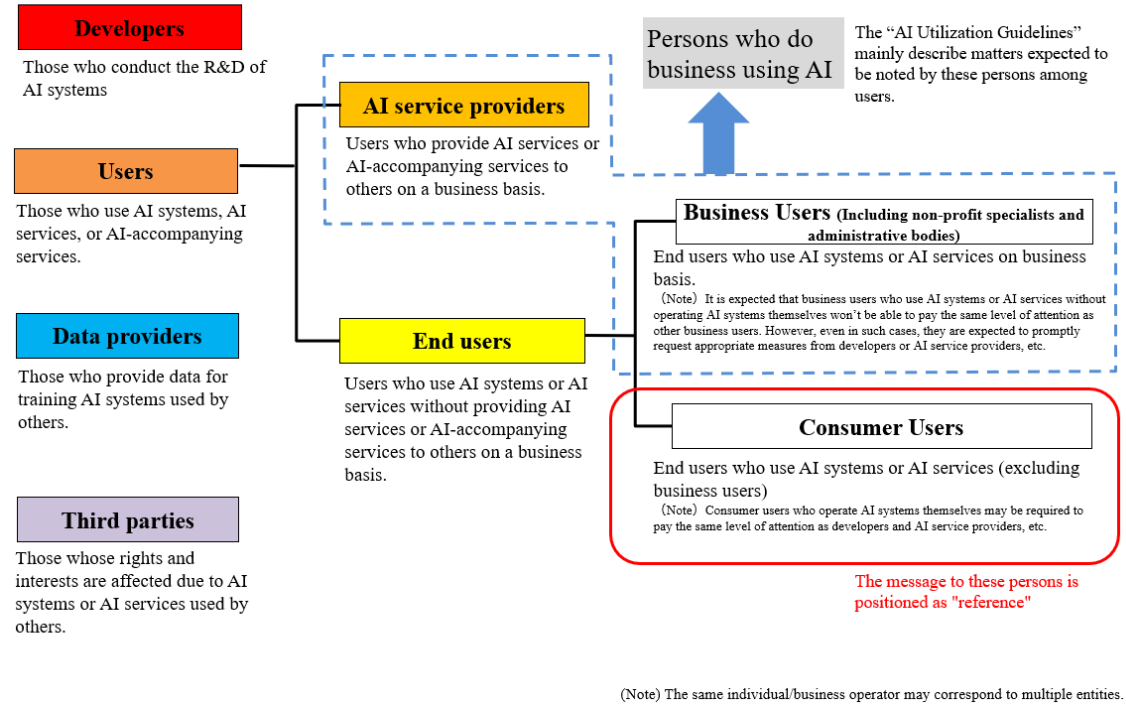
[92] See 6. (1) G. of this chapter.

Chapter 5 Initiatives for Consumer Users

1. Starting point for discussion

In the AI Utilization Guidelines, the points to be noted by AI service providers, business users, and data providers are explained concerning the 10 principles organized as the AI Utilization Principles, which are the pillars of the Guidelines. As for consumer users, points desirable to be noted are described together as a reference.

Organization of related subjects



(Note) The same individual/business operator may correspond to multiple entities.

From the perspective of actively promoting the "safe, secure and trustworthy implementation of AI in society" in the future, it will be necessary to initiate to enable consumer users to safely utilize AI and enjoy its benefits. Regarding this point, it is also described in Report 2019 **"Chapter 3: Future Issues" "1. Promotion of AI development and utilization and the sound development of AI networking"** in **"(1) Dissemination and development of the Draft AI R&D Guidelines and the AI Utilization Guidelines"** that "It is important to send easy-to-understand messages to consumer users as well. Based on the descriptions in the AI Utilization Guidelines, it is desirable to consider creating literacy materials such as "Handbooks" and "Documentation" (user guides) and holding workshops based on them."

Further, among consumer users, it is considered to be one of the most important initiatives to

realize a human-centered AI society to eliminate the inconvenience associated with aging and disability by making use of AI for the elderly and disabled so that everyone can equally achieve self-realization. From this perspective, it was decided to hold interviews with relevant parties on initiatives to implement safe, reliable AI for the elderly and disabled in society.

2. Initiatives for Consumer Users

(1) Government and Consumer Affairs Agency discussions on consumer involvement in AI

The Government of Japan, in its "Basic Plan for Consumer Policies,"[93] states the following as one of the "Plan of action of consumer policy in the Basic Plan for Consumer Policies in this fiscal year" regarding the relationship with AI and other innovative technologies to realize a society in which consumers play a leading role:

"With the advancement of ICT, innovative technologies such as AI, IoT, big data, and robotics are being developed, and our country, both public and private, is aiming to accelerate the realization of Society 5.0 by making the most of these technologies. Society 5.0 is a people-centered society that achieves both economic development and solutions to social issues through a system that highly integrates cyberspace (virtual space) and physical space (real space), and it is expected that consumer lifestyles will change significantly in this context. Given the rapid pace of change brought about by technological innovation, the impact on consumer life is expected to be significant. Therefore, it is necessary to respond intensively and promptly to (1) rapid changes in the fields of transactions and settlement accompanying the digitalization of the economy, and (2) rapid changes in the environment for utilizing big data."

In line with this description, the "Study Group on Responding to Consumer Digitization"[94] is being held at the Consumer Affairs Agency. One of the items to be studied is "Current situation of AI surrounding consumers and how to face it," and the "AI Working Group" (Chairperson: Member KOZUKA Soichi of the Conference)[95] has been launched and studies have begun.

(2) Cooperation as the Conference to the Consumer Affairs Agency's discussion

The Conference has been responding and cooperating to the efforts of the Consumer Affairs

---

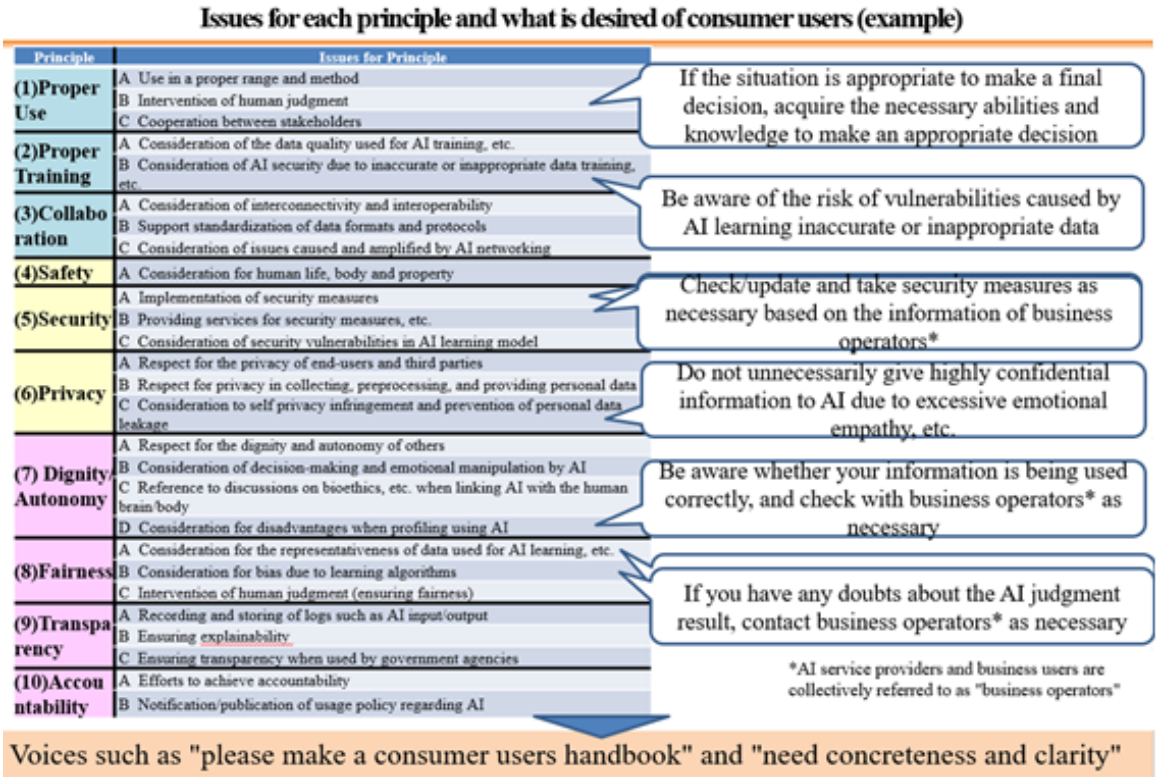
[93] Basic Plan for Consumer Policies (2020 Cabinet Decision on March 31)
<https://www.caa.go.jp/policies/policy/consumer_policy/basic_plan/>

[94] Study Group on Responding to Consumer Digitization
<https://www.caa.go.jp/policies/policy/consumer_policy/meeting_materials/review_meeting_003/>

[95] AI Working Group of Study Group on Responding to Consumer Digitization
<https://www.caa.go.jp/policies/policy/consumer_policy/meeting_materials/review_meeting_004/>>
TAKAGI Koichi, Senior Researcher (Senior Analyst, Co-Creation Strategy Group 2, Future Design Division 2, KDDI R&D Laboratories, Inc.) from the Research Department of the Information and Communication Policy Research Institute, Ministry of Internal Affairs and Communications, who was a Secretariat of the Conference, participated in this working group as a member.

Agency. Through such cooperation, efforts will lead to promotion to enable consumers to use and enjoy the benefits of AI with peace of mind.

Specifically, based on the discussions held at the meeting, in addition to the examples of the contents of the "AI Utilization Guidelines" as a reference regarding matters that should be kept in mind by consumer users, it mentions the necessity of specifying and clarifying such matters and creating handbooks for consumer users, as mentioned above.



(3) Overview of discussions at the Consumer Affairs Agency "AI Working Group"[96]

The following missions were shared by the working group during the discussion:

"With the ultimate goal to be providing new benefits and enriching consumer life by using AI wisely and appropriately, the mission is to consider what kind of information should be disclosed to businesses, what kind of knowledge is expected to be acquired by consumers for this purpose, and from the viewpoint of encouraging the exchange of such information, what position the government and other agencies should take"

Based on this content, after organizing information and sharing awareness about AI, the relationship between AI and consumers, and challenges, etc. of AI, discussions are held on what

[96] "AI Working Group of Study Group on Responding to Consumer Digitization Meeting Material" < https://www.caa.go.jp/policies/policy/consumer_policy/meeting_materials/review_meeting_004/>

consumers are expected to learn (focus in particular on AI functions and data collected by AI), the course of action for compilation, and points to be noted upon action. Further, there are plans to prepare a handbook for consumers based on the above discussion, and the targets, composition, editorial policy, and contents of the handbook are discussed as well.

3. Initiatives for the elderly and persons with disabilities

(1) Interviews with Member KONDO Noriko and Ms. WAKAMIYA Masako[97] ("The Future of the Elderly and AI Speakers")

a. Interview outline of Member KONDO Noriko ("Interim Report on Smart Speaker Survey")[98] [99]

It is the serious disability of the elderly that makes nursing care lonely and harsh, and from 30 years ago, the Rou-Tech Research Society (Aging and Technology Research Society) has been carrying out research and activities to promote nursing care based on the awareness that if you can enjoy using a computer, a smartphone and an AI speaker, it will improve nursing care.

The penetration rate of AI speakers by country was 22% for China and 20% for the U.S., while it was about 3% in Japan at the end of FY 2018. As for smart speakers, there is no experience of network trouble as reported in Europe and America, and no one is showing concern. We have been holding smartphone classes for seniors, but in the future, we plan to introduce precedent cases in Europe and the United States in scenes such as smart speaker use support classes.

b. Interview outline of Ms. WAKAMIYA Masako ("Report on the Status of AI Speaker Usage at Home and the Results of the AI Speaker Survey (interim report)")[100]

Since AI speakers have the convenience of being operated orally, the advantage of using the AI speaker in homes is that the AI speaker can be selectively used according to the situation between other electric appliances operated by contact.

With the introduction of monitors on AI speakers, when the response could not be seen or heard, information can be complemented via the speaker or monitor, overcoming the difficulties of poor hearing and viewing of the elderly.

Of the results of the survey on AI speakers, no one answered "Yes" when asked about concerns about information leakage, etc. There is more concern for hidden human operation error than internet

---

[97] Director, NPO Broadband School Association

[98] For interview materials (excerpt), refer to <https://www.soumu.go.jp/main_sosiki/kenkyu/ai_network/02iicp01_04000232.html>.

[99] The report on the "Smart Speaker Survey" was presented at the "Smart Aging Forum/Cyber Hina (Doll) Festival 2020" held in March 2020. The forum can be viewed at the following URL: < https://www.youtube.com/watch?v=Vk0czALp21U >

[100] For interview materials (excerpt), refer to <https://www.soumu.go.jp/main_sosiki/kenkyu/ai_network/02iicp01_04000232.html>.

security.

There is request to develop AI devices that are useful in supporting the independence of elderly people, because elderly people tend to be the target of nursing care and supervision, but elderly people do not want to be considered targets of nursing care and supervision, and they tend to think more positively when mentioned to support independence.

To popularize AI speakers for the elderly, the network environment is important, and it is expected that there will be support measures to prepare the network environment when constructing serviced residences for the elderly. In addition, it is expected that supporters or so-called help-man are registered in the community comprehensive center for setting up the network environment, and they will be able to do needed configurations upon setting it up. Further, improvement of the operability of home appliances and household equipment is expected.

c. Discussion

[Information leakage from AI speakers]

C. Given that there was no response to concern about information leaks from AI speakers, it is not clear how much information is actually at risk, nor is it often heard that there was economic damage, so it will be nice to understand how far it can be used through considering risk assessment.

C. Given the case in which information leakage is likely to occur in the immediate surroundings such as family members, if the intelligence of AI speakers is relied on in the future, there is a concern that not only privacy but also various information such as financial assets may be leaked.

[Concerns about dependence on AI speakers]

C. Given the current dependence on smartphones, there may be concerns that we will depend on AI speakers in the future.

A. Given the current state of AI speaker use, we have yet to become acquainted with them, so it is important to get acquainted with them first.

[Risk of AI speakers being used in a variety of applications]

C. The advantage of AI speakers is that they have a good interface that can be easily operated by voice. From this perspective, it is expected that various operations will be possible in the future without using hands. However, for example, in the case of operating an autonomous driving car by AI speaker, it is expected to cause great damages upon misoperation, so it is necessary to consider the risks when AI speakers are used in various applications in the future.

(2) Interview of Member TAMARU Kenzaburo[101]

a. Outline of initiatives for the elderly

(i) Background

The recognition accuracy of standard Japanese of people in their twenties, thirties, and forties has significantly improved, but the speech recognition of elderly people and people in rural areas is still very poor. In particular, in the aging society, when we observe the conversation of a person with dementia in nursing care, although it depends on the stage, it is naturally quite different from the conversation of a person without dementia.

(ii) Examples of initiatives by the AI Data Utilization Consortium (Promotion of conversation between elderly people and robots by AI using voice data of elderly people)

Because there is a shortage of nursing care personnel, research and development with ATR and professors of Osaka University were promoted to suppress the progression of dementia as much as possible by making conversation robots, androids, etc., for elderly people suffering from dementia. The accuracy of voice recognition for the elderly tends to be low because of the scarce voice data. Therefore, voice data of the elderly was specifically collected to utilize for robots that talk with the elderly.

b. Outline of initiatives for persons with disabilities

(i) Background (Comparison between Japan and the United States)

In the United States, at universities where there are students with hearing impairments, it is common practice to use speech recognition, display subtitles, and conduct classes. Speech recognition is very accurate, even if there is a dialect. Even for technical terms, it is now possible to upload power points and material in advance, and even if the pronunciation is the same, it can be recognized as the words used in the subject class, and display appropriately as subtitles.

On the other hand, Japanese language recognition, including dialects, is still difficult. This is largely due to the difference in data volumes in Japanese. The AI Data Utilization Consortium is initiating how to improve domestic service and quality in Japan when looking domestically.

(ii) Examples of initiatives by the AI Data Utilization Consortium (Data utilization project for people with difficulty in computer operation)

A technology where people with physical disabilities can enter text based on their gaze when using IT devices to communicate. Algorithm-based implementations have been slow to improve (the

[101] Conducted an interview on initiatives done by Microsoft and the AI Data Utilization Consortium. For interview materials (excerpt), refer to <https://www.soumu.go.jp/main_sosiki/kenkyu/ai_network/02iicp01_04000232.html>.

problem was that one operation takes a long time). Aiming to realize this by utilizing deep learning and AI, where input is performed by gaze prediction based on the training of past input. The Japan Assistive Technology Committee, in which companies that provide support for the disabled participate, and the AI Data Utilization Consortium are working together on this project, and Microsoft Japan is also collaborating. Specifically, data collection, analysis, modeling, and systematization are conducted together.

(iii) Microsoft Initiatives (1) (PuCom, a communication system for the severely disabled using pupil responses)

An initiative for those who cannot move their bodies at all, to support their decision making by analyzing images of pupils and creating a model.

(iv) Microsoft Initiatives (2) (AI Mimi, an AI-human hybrid information security system)

An information security system that enables people with hearing impairments to participate together in conferences and lectures.

As mentioned above, the quality of speech recognition has improved considerably overseas; this is an initiative to collect Japanese speech data and improve the speech recognition quality so that students with hearing impairments can actively participate in learning opportunities.

c. Challenges

It will also become very important in the future what kind of social services and support Japan can provide to people especially in remote areas, in an aging society, using AI. Looking at the data, although it is obvious in the field of natural language, there is no data in other fields that is inversely proportional to the areas where it would be better for elderly people and people in remote areas to use AI. Therefore, it is also important to seriously consider the distribution and acquisition of such data.

4. Summary

(1) Initiatives for consumer users

a. Cooperation with the Consumer Affairs Agency

The Consumer Affairs Agency is considering the creation of a handbook for consumers, so it is thought to be effective to follow up and cooperate such as by promoting through publicizing, etc.

(2) Initiatives for the elderly and persons with disabilities

b. Cooperation in activities to disseminate information on how to use AI speakers

While products and services such as AI speakers are already in practical use due to their

convenience, their penetration rate in our country is not necessarily high. According to the survey results, there is no concern about network troubles, but on the other hand, the possibility for human error is raised. In light of this, it is considered necessary to disseminate information on how to use AI speakers, etc., although it is an individual choice whether to use it or not. Therefore, it is considered beneficial for the Conference to cooperate with organizations engaged in such publicity activities.

c. Promotion of dissemination of examples of advanced initiatives

The cases introduced in this interview are useful as advanced initiatives for supporting the lives of the elderly and disabled through the use of AI. It is considered necessary to continue to collect such advanced cases through interviews, etc., and disseminate information on them.

Chapter 6 Initiatives for Security

1. Starting point for discussion

As part of the development of an environment for safe, secure, and trustworthy AI, it is necessary to consider technical measures. For example, various initiatives can be considered, such as ensuring quality[102], improving explainability, certification, and security.

Of these, the AI Development and Utilization Guidelines explicitly provide for "security principles" with a focus on ensuring security, based on the recognition that for the promotion of AI development and utilization and the sound development of AI networking, there is a particular need for security that protects AI users.

On the other hand, if we reconsider the relationship between AI and security (hereinafter referred to as "AI x Security") from the viewpoint of "environmental development for safe, secure, and trustworthy AI," the following 4 viewpoints will be raised[103].

(A) Attack using AI

(B) Attack by AI

(C) Attack to AI

(D) Measure using AI

Of these, as the "Security Principles" in the AI Development and Utilization Guidelines focus on point (C) as described above, it is considered important for the Conference to dig deeper into point (C) individually. However, there is an opinion that "It is important to deepen each of them, but combining them will deepen research."[104], so it is considered important to consider other points as well.

Based on the above, it was decided to hold interviews with the JNSA (Japan Network Security Association), a trade association that examines measures and conducts awareness-raising activities regarding AI x Security, to organize the points at issue regarding AI x Security.

---

[102] There is a QA4AI consortium that considers guidelines to ensure the quality of AI products such as machine learning. In addition to the common framework for quality assurance, the Consortium's guidelines also refer to specific case studies such as smart speakers, industrial processes, and automated driving, and there are opinions that these examples will be useful for the discussion in Chapter 4 (Initiatives by business users).
< http://www.qa4ai.jp/ >

[103] SASAKI Ryoichi (Visiting Professor, Center for Research Collaboration, Tokyo Denki University) "AI and Security" <https://digitalforensic.jp/2018/09/18/column531/>

[104] Cyber Security Task Force, Ministry of Internal Affairs and Communications (20th Session) Material 20-3 "Research and Development to be Focused in the Future"
<https://www.soumu.go.jp/main_content/000666230.pdf>

2. Specified non-profit organization Japan Network Security Association (JNSA)

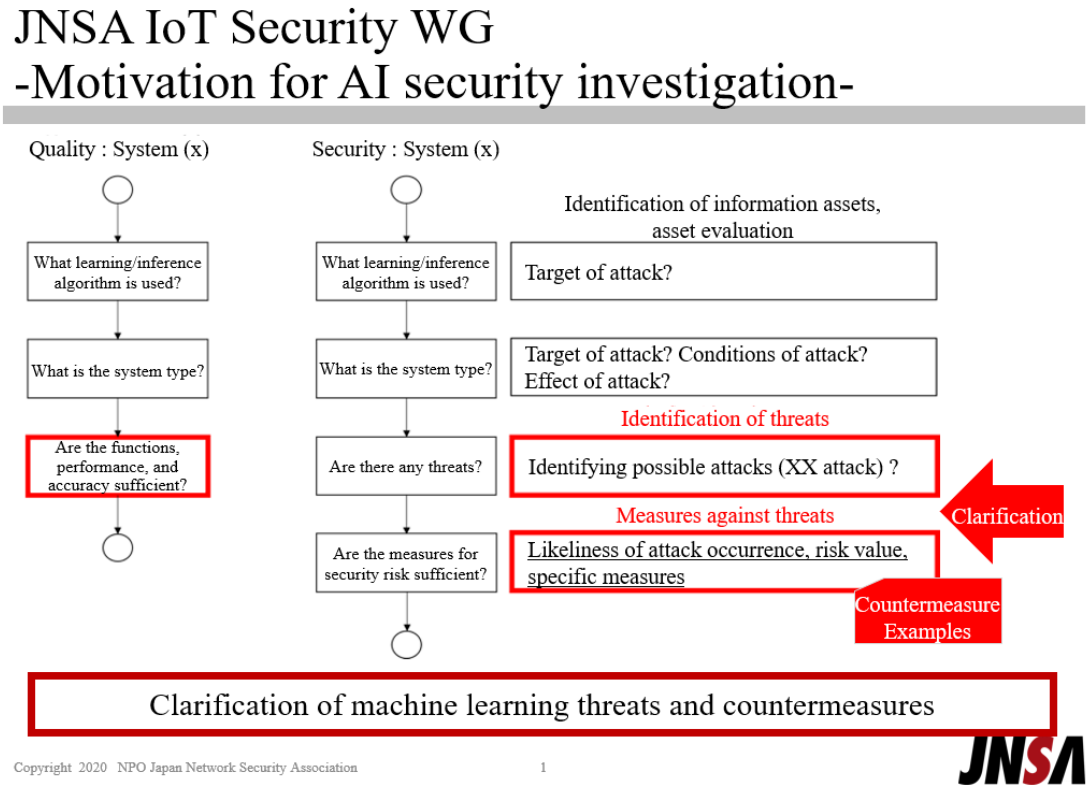
(1) Interview outline[105]

a. JNSA initiatives

The JNSA was founded in April 2000 as a security business group, and now, with the participation of security officers from various companies, is an organization with 239 companies as members (as of 25 October 2019). To realize a secure network society, information sharing among vendors, and enlightenment and information sharing for users are conducted through various subcommittees.

b. Overall structure of "AI x Security" and the following discussion

As described in this chapter "1. Starting point for discussion", there are various viewpoints regarding AI x Security, but not many people are studying viewpoint (A), and the other 3 viewpoints are often discussed. The following describes (C) Attack to AI and (A) Attack using AI. Concerning the former, since there are no specific threats or countermeasures against AI systems, the IoT Security Workgroup of the JNSA will explain the content investigated to clarify these issues (refer to the figure below). On the other hand, regarding the latter, research based on the content of presentations at overseas security conferences, etc. will be described.



[105] For interview materials (excerpt), refer to <https://www.soumu.go.jp/main_sosiki/kenkyu/ai_network/02iicp01_04000232.html>.

c. About (C) Attack to AI (classification, countermeasures, and policies)

Attacks (threats) to AI, especially machine learning (models, systems, etc.,) fall into three broad categories:

- Evasion Attacks: An attack that causes contradicting recognition between humans and machine learning inference engines by inputting data unrecognizable by humans
- Poisoning Attacks: An attack that shifts the boundaries of the learning model in some way by inputting malicious data into the learning data; and
- Inversion Attacks: An attack that extracts sensitive information such as original data by data input/output or response to the inference engine of machine learning

AI-specific threats

Attacks to machine learning

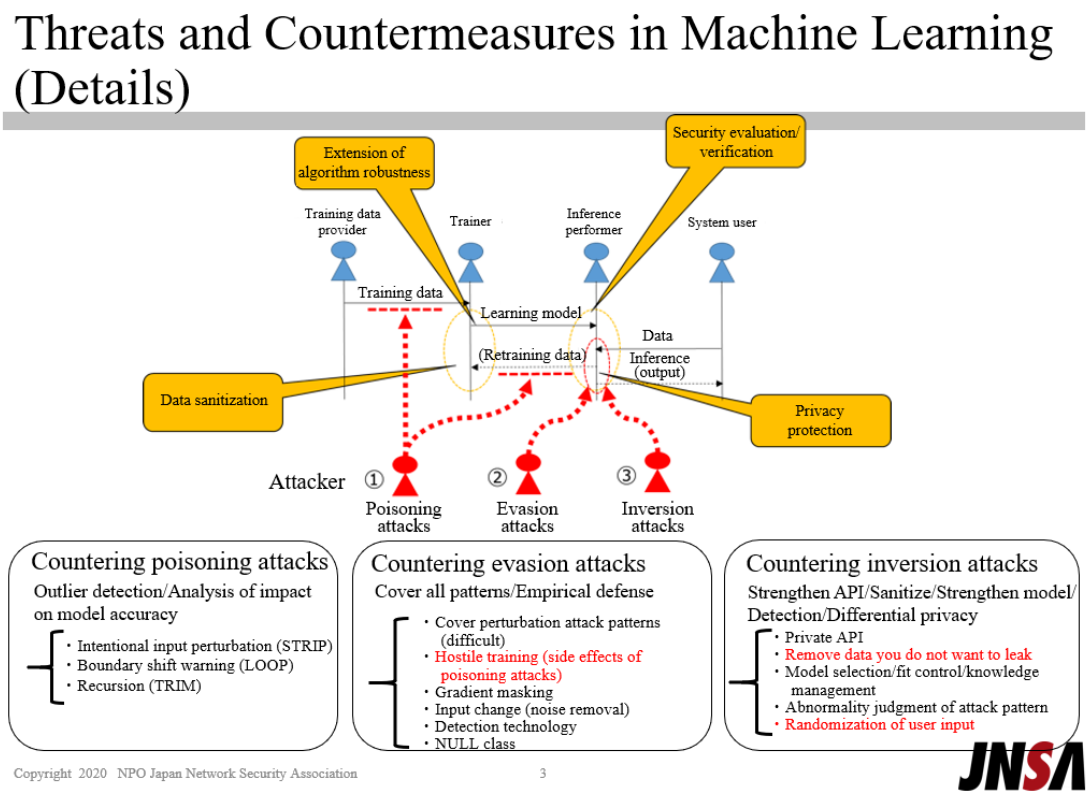
Attack (Threat)	Subcategory	Content
Evasion Attacks	・Stealth ・ Impersonate	Attacks that cause different recognition between humans and machine learning inference engines (images, sounds, characters, etc.) due to "data input including perturbations" that cannot be recognized by humans
Poisoning Attacks	・Availability ・Backdoor	Attacks that shift the boundaries of the training model in some way by "inputting (injecting) malicious data" into the training data (There are availability attacks that render machine learning model boundaries unusable by injecting large amounts of bad data, and backdoor attacks that generate backdoors by injecting small amounts of sophisticated data)
Inversion Attacks	・Privacy ・Membership	Attacks that extract confidential information (or the model itself) such as original data through the "data input/output or reaction to the inference engine" of machine learning (With membership inference attacks, the adversary searches whether the data at hand is included in the other party's dataset)

* There are different names and classifications, but here we have summarized three attacks.

JNSA

As for countermeasures against these attacks, effective measures are introduced in reports. In summary, for poisoning and evasion attacks, it is desirable to fix the AI-based technology, that is, to correct the algorithmization of AI. On the other hand, it has been found that for types of data theft such as inversion attacks, it is better to use anonymization technology, which is a security technology (refer to the diagram below for details).

Further, there was an explanation that key points are to limit attacks according to the situation as much as possible, and to identify who the attacker was and take appropriate measures.



d. About (A) Attack using AI (types of attacks and measures taken)

There are several AI (especially machine learning) attacks under consideration, but here, DeepLocker and DeepFake and measures to be taken for each will be explained.

DeepLocker is a targeted attack that embeds malware into the deep learning model introduced in black hat[106] USA 2018. In normal times, the app runs as a benign app with encrypted malware built into it, but it uses a deep learning model to determine the presence or absence of a target while collecting information about the target, and when it recognizes the target, it attacks the app. There are no specific measures for this.

DeepFake, on the other hand, is a technology for creating so-called fake videos. Various measures have been taken for this, including efforts in the United States to detect fake videos in anticipation of a presidential election, and there is a state law in California that allows citizens to sue fake video creators[107].

[106] <https://www.blackhat.com/>
[107] Assembly Bill No.602:
<https://leginfo.legislature.ca.gov/faces/billTextClient.xhtml?bill_id=201920200AB602>

Attack using AI

Attack name	AI Use	Content
DeepExploit	Determining the best attack method	A PoC using deep reinforcement learning for penetration testing. Based on the information collected from the intrusion target system (OS, product name/version, etc.), the attack method with the highest probability of successful system intrusion is determined and the intrusion action is executed. After successful intrusion, the intruded system is used as a stepping stone to repeatedly invade the internal system.
DeepLocker	・Target identification ・Malware concealment	A PoC that uses deep learning for targeted malware attacks. Encloses encrypted malware and collects information on the target person via webcam/microphone while acting as a face recognition application or video application in normal times. When the target person is identified, the built-in malware is decrypted and an attack is performed. It is possible to enter the target PC/smartphone without being detected by anti-virus software.
tAIchi	Automatic malware generation	A PoC that uses GAN and reinforcement learning to generate malware. Transforms known malware with GAN and reinforcement learning to automatically create a malware (variant) that evades detection by anti-virus software.
Deepfake	・Face replacement ・Reproduction of facial expressions	A group of technologies that use autoencoder/decoder to create fake videos. Creates a fake video of the target person by replacing the face part of the original video with the face of the target person.

JNSA

(2) Discussion

[Treatment of viewpoint (B)]

Q. It was mentioned that few people are studying (B) Attack by AI (Attack by AI itself), but what kind of discussions have been made?

A. Although there is no particular discussion, human risk sensitivity is not high and assumptions may not be sufficient. Perhaps Professor SASAKI, who conducted the classification of this point of view, may have included this risk in the classification based on his expertise in IT risk studies that risks related to this matter should also be assumed.

[Viewpoint (C): Attacks to machine learning and presence or absence of intent]

Q. Regarding poisoning and evasion attacks, are intentional attacks (abuse) and attacks that are not known to be intentional (misuse) considered from the security perspective? (For example, whether the multiple inputs of biased training data are caused by a particular person or is a result of collective misuse.)

A. From a security point of view, it is impossible to treat objects that do not contain intent as a target, and the premise is based that those will be problems of ensuring quality. However, it is very difficult to distinguish between them (for example, in the section of privacy protection shown in the figure “Threats and Countermeasures in Machine Learning” above, if an inversion attack is

attempted, and if the attack continues many times, there is a possibility of network trouble (quality problem), but if it can be judged to be intentional based on the number of attacks and the type of site it is attacking from, security measures should be taken).

Q. Also, in the case of poisoning attacks, it states "input of malicious data," but how do you judge whether or not it is malicious?

A. There may be methods such as detecting an abnormal value by setting a threshold value (determine whether it is fraudulent based on the value), but distinguishing is very difficult.

[Viewpoint (A): Fake image creation and defining good and bad]

Q. About DeepFake, it is hard to define whether a fake is bad or good in the first place, but how do you separate the bad fakes?

A. If there are victims in some way or another based on the fake video, it shall be classified as malicious, but it may be hard to draw a line for those in the gray area.

3. Supplementary theory

- About viewpoint (D), both domestic and international research is progressing in terms of technology, and as with viewpoints (A) and (C), information is constantly updated at security conferences such as black hat. In our country, the "AI Strategy 2019"[108] has a goal to "establish highly efficient and sophisticated AI technologies for each phase of 'prevention' 'detection' and 'response' to counter increasingly complex and sophisticated cyber-attacks" and various research institutes such as the National Institute of Information and Communications Technology are studying this[109] [110].
- In general, there are opinions that it is important "to conduct discussions not only with security engineers but also with those involved in AI" for examining this field.[111] There are also opinions that "research in the humanities and social sciences is becoming a relatively useful element in cybersecurity. ... Systems such as social engineering and eDiscovery are entwined with organizations and psychology"[112].

---

[108] <https://www.kantei.go.jp/jp/singi/ai_senryaku/pdf/aistratagy2019.pdf>

[109] Conference 3rd Committee on AI Governance, National Institute of Information and Communications Technology, Takahashi Research Manager Lecture Material

[110] See Ministry of Internal Affairs and Communications "IoT/5G Security Comprehensive Measures Progress Report 2020" Chapter III, Progress of Comprehensive Measures and Future Initiatives (cross-cutting approach), (1) Promotion of Research and Development (6) Research and Development of Cyber Attack Detection and Analysis Technology Using AI
< https://www.soumu.go.jp/main_content/000688845.pdf >

[111] Cyber Security Task Force, Ministry of Internal Affairs and Communications (20th Session) Material 20-3 "Research and Development to be Focused in the Future" (aforementioned)

[112] Cyber Security Task Force, Ministry of Internal Affairs and Communications (20th Session) Minutes of Proceedings

4. Summary

(1) Deepening measures against attacks on AI and addressing other issues

- To make the principles of security and other items practical, based on the classifications such as (C) attacks on AI, it is considered technically important to limit attacks as much as possible and to identify who the attacker is and strengthen countermeasures.
- It is also important to proceed discussions with considering not only attacks on AI, but also other issues related to AI. In particular, (B) attacks by AI have not been studied in detail, but it is necessary to consider them as one of the risks, just as AI and AGI, which operate autonomously, were considered when considering AI Development and Utilization Guidelines.

(2) Need for determining intent in an attack

- Since it is difficult to determine what constitutes malicious or evil in both (C) attacks on AI and (A) attacks using AI (to decide what attacking is), it is important to consider how to identify the intent of such attacks.
- For reference, the Platform Services Study Group of the Ministry of Internal Affairs and Communications presented in its final report[113] a number of issues, including the promotion of fact-checking[114] and the promotion of ICT literacy, on the grounds that there are various levels of responses to false information (information that is intentionally false). Further, as AI may control and delete the distribution of information, it is expected to refer to the "Draft AI R&D Guidelines" and the "AI Utilization Guidelines" of the Conference for ensuring transparency and accountability.

(3) Need for multistakeholder interdisciplinary discussion

- Since this field contains problems that cannot be solved by technical aspects alone, it is important to continue interdisciplinary discussions, not only by (security) engineers but also by incorporating knowledge from psychology, sociology, etc.

---

[113] The report is posted on the website of the following URL.
< https://www.soumu.go.jp/main_content/000668595.pdf >
A summary of measures against fake news and false information on the internet, including this report, is available on the website of the following URL.
< https://www.soumu.go.jp/main_sosiki/joho_tsusin/d_syohi/ihoyugai_05.html >

[114] For example, it is reported that for reasons of responding to posts containing slanderous injuries to individuals, on top of strengthening countermeasures, Yahoo! Japan has set up a technical provision and study group for a natural language processing model (AI) using deep learning. Specifically, there is a case where 20,000 comments posted on Yahoo! News were deleted per day as posting of unrelated information.
(From Yahoo! Japan press release: <https://about.yahoo.co.jp/pr/release/2020/06/01a/>)

Chapter 7 Initiatives Related to Insurance

1. Starting point for discussion

As initiatives to create a safety net upon promoting the "safe, secure and trustworthy implementation of AI in society," it is considered important to establish an insurance mechanism to compensate for damage caused by the development and utilization of AI, taking into account the characteristics of AI. It is also described in Report 2019 "Chapter 3: Future Issues" "1. Promotion of AI development and utilization and sound development of AI networking" "(3) Issues related to environmental improvement addressed by relevant stakeholders," it states the "consideration of how to provide relief (insurance, etc.) to victims of AI accidents, etc. and prevent the occurrence of damage." For "(6) Protection of the interests of users," it states the "consideration of measures to prevent damage to users, etc. due to AI system risks, allocation of responsibility if risks become apparent (at the time of occurrence of an accident, etc.), and consideration of ideal mechanisms to protect users, etc. (insurance, etc.)" are listed. Based on these issues, interviews were conducted with companies that are taking advanced initiatives to AI insurance.

2. Tokio Marine & Nichido Fire Insurance Co., Ltd. (About "Insurance to help spread AI")

(1) Interview outline[115]

a. Analytical technique

The players associated with AI were classified into "AI providers" "AI users" and "indirect AI users" and examples of assumed risks associated with AI were organized by player. Based on the above, an analysis was conducted of parties responsible for accidents involving AI-related devices.

b. Awareness of challenges of existing regulatory frameworks

Regarding the responsibility of AI operators and AI users, for AI operators, it is determined that "when AI is incorporated into a product, it is possible to organize the responsibility within the framework of conventional product liability, but due to the inherent characteristics of AI (learning ability and unpredictability of their behavior), it will not necessarily work well in all accidents. However, as AI operators, it is required to control or reduce risks caused by AI devices at an allowable level, so it is likely for them to bear some responsibility." On the other hand, for AI users, it is determined that "if an AI user is found to be at fault, tort liability exists, but it is difficult to uniformly discuss the predictability of events caused by autonomous AI devices, and it is impossible to determine

[115] For interview materials (excerpt), refer to <https://www.soumu.go.jp/main_sosiki/kenkyu/ai_network/02iicp01_04000232.html>.

whether or not there is a fault." For these reasons, awareness is expressed regarding the problem that it is difficult to determine who will take the responsibility and the distribution of responsibility based on the existing legal and regulatory framework alone.

c. Case (1) (Insurance products for accidents of autonomous driving)

Based on this awareness, the "special agreement on compensation for victims' relief expenses" as an insurance product for accidents caused by autonomous driving was developed. The system allows insurance companies to pay insurance money regardless of liability for accidents during autonomous driving operated by the driver, and to pay insurance money for automobile accidents with unclear causes for prompt relief to victims. If it is later decided that the manufacturer is liable, the insurance company will seek compensation from the manufacturer.

d. Case (2) (Insurance products for defects in new houses)

The Housing Quality Assurance Act stipulates that major structural parts of new houses must be covered by liability for defects for 10 years in the sale and purchase of new houses. With this as a basis, this is an example of a product for home sellers purchasing residential defect liability insurance.

e. Proposed insurance approaches for the risks associated with AI (1) Victim relief cost compensation insurance

Based on the cases described in C and D above, two approaches are proposed as insurance approaches to the risks associated with AI. One example is, as seen in the example of autonomous driving, if an indirect AI user suffers injury or other damage from a product or service provided by the AI user, as a "victim relief cost compensation insurance" the indirect AI user will be compensated for the damage as a special contract of liability insurance arranged by the AI user, regardless of whether the AI user is responsible or not. This approach is one way to achieve social acceptance of AI in the sense of direct victim relief. However, this will not be the direct reason for AI to become widespread. In order for AI to become widespread, measures must be taken to prevent accidents in the first place.

f. Proposed insurance approaches for the risks associated with AI (2) AI quality assurance liability insurance

Using examples of home quality assurance as a reference, the idea is that AI providers will guarantee the quality of AI to AI users, and insurance will be provided for those who are responsible for guaranteeing the quality of AI. However, the most important issue is what kind of quality should be guaranteed, and the system will need a quality evaluation standard to be based on. An effective approach for ensuring quality only when there are quality assessment standards for AI and an auditing body that evaluates quality, where the insurance will support that quality assurance.

g. Summary

"Insurance" provides compensation that reflects the inherent characteristics of AI for the risks faced by each AI operator, user, and indirect user, in order to realize the "safety" and "social acceptance" necessary for the social implementation and spreading of AI. Its functions will be more effective when combined with legal and technical approaches to clarify the responsibilities associated with AI and the development of governance and quality assessment guidelines by auditing bodies.

(2) Discussion

[Framework of "Special agreement on compensation for victims' relief expenses" for autonomous driving accidents]

Q. There were comments that it was a desirable approach from the viewpoint of victim relief. However, because autonomous driving cars have a very complex structure and the cause of accidents may not be easily identified, and there is a large asymmetry of information between insurance companies and automobile manufacturers, so insurance companies will have difficulties from the viewpoint of ex post facto reimbursement. There were questions about what kind of institutional collateral should be used to solve the problem. Further, it is expected that there will be difficulty in determining the fault of the victim, such as in cases where the driver, who is the victim of the accident, forgot to update the software of the autonomous driving car; what is the future course of action in this regard.

A. There are currently no standards or methods to clearly identify the liability for accidents caused by autonomous driving. On the other hand, there are some very difficult aspects of offsetting negligence in the normal world of existing vehicles, but there are certain criteria based on historical precedents, and insurance payments are made on the basis of these criteria. Therefore, even in the field of autonomous driving, it is difficult to decide in advance before the accumulation of accident cases, and in consideration of the asymmetry of information with automobile manufacturers, it is necessary for insurance companies to acquire the necessary specialized knowledge by the time when autonomous driving is fully realized.

[Significance of quality assurance]

Q. By providing quality assurance insurance, the cost will come back to AI providers later in the form of insurance premiums, so providers must engage in it properly. Insurance companies don't have to pay insurance money for those properly engaging in quality assurance. For those who don't, insurance money will be paid to the victims, and of course, it will reflect the insurance premiums of the AI providers in the following year. Is it correct to understand that this is a proposal that provides incentives to create good quality products?

A. That is right. The insurance that covers AI quality assurance is exactly to secure the incentive for

the business to improve quality. This proposal is based on the idea that if companies do not ensure quality, it will affect the spread of AI itself.

[Benefits of quality assurance liability insurance]

Q. If it is covered by quality assurance liability insurance, is it correct to assume that if it is marked as having a quality that is properly inspected, guaranteed to a certain degree, and covered by insurance companies, consumers will preferably buy such products, leading to a virtuous cycle?

A. That is right. In addition, in the case of the special agreement on compensation for relief expenses, for example, if accident data are accumulated and the accident rate of a certain manufacturer's autonomous driving car is high, it is conceivable that the premium of the automobile insurance of the owner of that manufacturer's car will rise. Currently, auto insurance premiums vary by manufacturer and type of vehicle, so there is a concept of differentiating premiums for autonomous driving vehicles to provide incentives for accident prevention. However, the more direct incentive for quality improvement will be to guarantee quality assurance responsibility. Quality assurance is a more direct approach for proving quality than directly reflecting it in insurance premiums. However, this will be a challenge because there is a precondition that mechanisms such as an organization evaluating quality exists.

[System design for quality assurance liability insurance]

Q. The level of quality assurance can vary among AI providers, so is insurance design possible in this case? Normally, accident rates are calculated by collecting cases under as uniform a condition as possible. However, if the SLA (Service Level Agreements) of products and services are different, whether insurancing is really possible will rise as a question.

A. Since it is insurance, we believe that a stable system cannot be created without collecting the same kind of risks. In that sense, even if we say AI, it has various uses, so it is desirable to create a certain parent set by use and type. Since it is difficult for insurance companies if SLAs are set freely per company, we think it is necessary to make efforts to establish quality evaluation standards and guidelines and ensure that the guarantees are in line with those guidelines.

[Guarantee of quality assurance standards]

Q. In the case of housing, I think there is a minimum quality level that must be guaranteed required by law, but does this mean that a legal minimum, such as a standard law, may be necessary for AI quality assurance liability insurance as well?

A. We believe it doesn't have to be a law, but we may need some kind of industry standard.

[About product design]

Q. In terms of product design, for relief cost compensation insurance, in the case of automobiles, the third-party and the first-party can be the insured, since the end-user has liability insurance. But for AI in general, the third-party is the insured person, and the indirect AI user is insured for cost compensation, so I cannot imagine that being in a package, but what are some ideas for other than automobiles? And as for quality assurance liability insurance, there are probably several ways to do this, one of which is to rate and change insurance premiums by rating, as is done in the world of ships. The second way is to include a disclaimer, and if the product does not meet the quality evaluation standards, it will be disclaimed and insurance will not be paid. In this case, the quality evaluation standards must be very low or at a minimum level, or the insurance money will be zero with immunity. Third, in the area of AI development, for example, there may be a way to make the payment of insurance payments conditional on compliance with governance and user principles. What do you think about this?

A. The image is that third party insurance, such as product liability insurance, that the company subscribes to will have an expenses contract attached to it. Rather than including users as insured, the idea is that if the company is held responsible, the company will pay the cost in the form of liability insurance, but the cost will be covered by insurance, assuming that the company will bear the cost once in the form of a response cost to provide moral relief to the victim.

[Problems regarding causal relation]

C. There has been an explanation that insurance design is difficult unless who holds responsibility is clear, but in the case of AI, problems regarding causal relation may be more of an issue. In particular, I believe that the Insurance Act in Japan is unexpectedly unclear, including whether the causality of civil liability, the causality of insurance, and the causality of insurance accidents are the same. How is this seen from the perspective of insurance practitioners? In addition, I think it is quite difficult to determine whether the relationship is causality, predictability, or proximate causality.

A. Regarding the liability insurance that we handle, we are considering applying insurance per the liability stipulated by civil law, and if a considerable causal relationship is not established concerning AI, we will naturally not be liable for compensation. As to whether it is a matter of predictability or a matter of proximate causality, I think it is both in the end, but when it comes to AI, there is no standard to judge whether it is the manufacturer's responsibility or the company's responsibility, so this quality assurance is a matter of simply making it a contractual responsibility. The concept is to conclude a contract under which action will be taken if a certain standard is not met. By doing so, the responsibility will be recognized without considering the issue of proximate causality, etc., and the insurance will cover it.

3. Sompo Japan Insurance Inc. (“Utilization of Insurance in Smart Factories”)

(1) Interview outline[116]

a. IoT/AI solutions and insurance

We recognize that the penetration of IoT/AI solutions is still insufficient and is in a transitional period. Examples of factors that hinder the implementation, consideration of countermeasures (types of countermeasures according to the type of risk), and implementation of those considered countermeasures, this model combines technology and finance to realize “IoT/AI solution + insurance cost < existing cost” and expects the creation of new business opportunities through the accumulated data and know-how through (1) the use of insurance (insurance products corresponding to each risk), and (2) minimizing risks through the accumulation of know-how of IoT/AI implementation.

b. Insurance products corresponding to each risk

Assuming that each product is to be designed custom-made, there will be 5 types of insurance: (1) performance assurance insurance that responds to performance risk and (2) predictive failure cost insurance that responds to failure (predictive) risk both based on business collaboration (small damage high frequency), and (3) predictive failure profit insurance that responds to shutdown risk, (4) cyber insurance that responds to cyber risk, and (5) IoT/AI liability insurance that responds to compensation (contract) risk, all of which are based on the risk scale (large damage low frequency).

(i) Performance assurance insurance

An agreement concluded with a user company by an IoT/AI vendor to guarantee the cost for implementing an IoT/AI solution if the vendor cannot meet a certain level of performance (SLAs — Service Level Agreements) at the time of a production implementation, and by fulfilling the agreement, the vendor company is compensated for damages. The content of the SLA and the amount of compensation must be set within a range that is appropriate in terms of social norms. The concept is that vendors are usually at risk of AI performance failures and malfunctions, and many companies are not moving from PoC (Proof of Concept) to production, so risk sharing (assurance) has the effect of encouraging implementation.

(ii) Predictive failure cost insurance

To back up costs borne by a user company by providing insurance for additional costs borne by the user company upon pre-detection of failure (costs for investigation of causes, field engineer dispatch, emergency measures, etc.). The concept is positioned as the utilization of insurance against the negative scenarios assumed before actual implementation.

---

[116] For interview materials (excerpt), refer to <https://www.soumu.go.jp/main_sosiki/kenkyu/ai_network/02iicp01_04000232.html>.

(iii) Predictive failure profit insurance

Covers the loss of profits (items that are not covered by the predictive failure cost insurance) incurred by a user company due to the shutdown of a factory, facility line, etc., upon pre-detection of failure, insurance is provided to back up the loss incurred by the user company. The concept is to prevent failures from occurring through pre-detection, and to compensate for the loss of profit in the case of shutdown and restoration work by insurance, thereby reducing the total loss of user companies compared to prior to implementation.

(iv) Cyber insurance

Compensates for compensation risks held by insured persons, emergency expenses required upon accidents, etc., and loss of profits/operating expenses of insured persons, as well as compensation for outside agency investigation consignment costs that were expended to determine the presence or absence of unauthorized access, etc., and costs that were expended in the case where measures to shut down networks were outsourced when a risk of unauthorized access, etc. was discovered based on a notice from a security management company, etc. The concept is to respond to unexpected compensation risks, especially when providing cloud services or remote maintenance services on the premise of Connected when the vendor is the insured, and to compensate for profit damages and operating costs due to cyber-attacks when the user is the insured.

(v) IoT/AI liability insurance

Compensation is provided to compensate for damage to user companies' equipment, machines, etc., caused by malfunction of IoT sensors provided by IoT/AI vendors or false AI (ex: failure prediction) detection, etc. if the vendor is liable for compensation in accordance with the contract. The concept is that, for example, failure prediction cannot be said to be 100% effective, and the vendor holds the risk of the damage of the user company caused by overlooking predictive failure, which leads to the expansion of the service for the vendor company or budget fixation of the user company.

Note that risks other than the five products described above are also possible, and this is not limited to these five types of insurance but is positioned as initiatives to create new types of insurance.

c. Insurance design considerations

There are two important points in considering insurance. The first is to sort out the responsibilities of the parties involved in the IoT/AI business model, and it is necessary to clearly sort out who bears what risks to whom. The second point is whether or not the risk is a risk that should be passed on to the insurance company. We believe that a design suitable for the risk management of each company is necessary.

(2) Discussion

[Estimating the probability of risk occurrence]

Q. How will the probability of risk occurrence be estimated?

A. As it is most likely not a traditional risk, we believe data will be a key point. Assuming that the data collected through the PoC can be obtained, we consider that the occurrence frequency of predictive failures can be calculated.

[Cyber insurance]

Q. In the case of cyber-attacks, there are cases where incidents are caused by attacks such as layered attacks, in which attacks are made in situations that are difficult for the users of the actual system to protect. In such cases, what are the criteria for determining how much of an attack should be made before the insured business becomes liable?

A. Not limited to cyber-attacks, liability insurance is determined by whether the customer is liable or not. Therefore, whether the vendor is liable for the liability or not depends on each case, and in some cases, it may be determined in court.

[Reasons for insurance exemption]

Q. About the reasons for insurance exemption, although intent may be included, is it correct to understand that cyber risk is basically the only reason for exemption? Does this mean that the overall picture is that cyber insurance will cover such cases?

A. In general, intent is exempt from liability. As for the performance guarantee and predictive failure, the design is performed based on the PoC data received from the customer. Therefore, if the determined conditions or operation environment are different, the preset thresholds and conditions are changed, so that the conditions and the immunity will be determined individually.

[About product design system]

Q. I think PoC is especially important in the overall process of cooperation, from the confirmation of needs to a meeting, and then to the conclusion of insurance contracts, but I think that a considerable number of people must be involved. In particular, how are the contracts dealt with for performance assurance insurance? What kind of team and how do they start and create the whole picture?

A. In designing this product, we asked opinions of those who are actually operating in vendor companies and user companies, and asked opinions of those who are facing troubles in the field, what needs exist, and confirmed what insurance can be applied to them. In the past, as an insurance company, we have been selling packaged products, but we are currently working with a team based on experts on such new risks.

4. Summary

(1) Collection, dissemination, and sharing of insurance products and their case studies related to AI

Considering the risk characteristics of AI, the development of various insurance products for the purpose of compensating for losses, etc. is considered important as an initiative to establish a safety net for utilizing AI. Therefore, it is considered necessary for the Conference to continue to follow the development of insurance products suited to the risk characteristics of AI, and to collect, disseminate, and share case studies of insurance utilization.

(2) AI quality assurance and insurance

It is important to disseminate AI by ensuring the quality of AI, and from the perspective of ensuring AI quality, an approach of quality assurance liability insurance combined with AI quality evaluation standards and governance by audit organizations that evaluate quality is considered an effective means. In this case, the quality assessment of AI will differ depending on the application and business type of AI, and the methods and feasibility of quality assessment audits will vary accordingly. Therefore, it is considered necessary to continue to consider the quality assessment of AI and the consideration of insurance from an expert point of view while following up on the consideration.

(3) Legal issues regarding AI insurance

Upon considering AI insurance, it may be difficult to determine who will take the responsibility and the distribution of responsibility based on the existing legal and regulatory framework alone. With regard to these issues related to civil law, such as the location and distribution of responsibilities, as well as the legal issues arising from the development of new insurance products, it is necessary to continue to consider them while following up on the consideration of the legal issues of insurance related to AI from an expert perspective.

In Place of Conclusion

The concept of compiling this report is as described in the “Introduction” but we would like to emphasize again the fact that it was compiled based on interviews and free-spirited discussions on individual concrete and ambitious initiations. Until now, Report 2017 has compiled the draft AI R&D Guidelines, Report 2018 has compiled the Draft AI Utilization Principles, and Report 2019 has compiled the AI Utilization Guidelines based on the Draft AI Utilization Principles as the core of each report. Compared with previous reports, this report has a different feature. This is because discussions on the development and utilization of AI have shifted from the “formulation of principles” phase to the “social implementation of AI” phase based on the “formulation of principles”. As can be seen from the start of operation of the OECD observatory (a platform for sharing AI information) introduced in Chapter 1, 2., moves to advance the “social implementation of AI” have been active in international and overseas discussions. Not enough gratitude can be expressed to those who made reports of ambitious and valuable initiations for the interviews with the Chairman of the Conference, and those who cooperated in free and vigorous discussions with the members of the Conference.

Also, unlike previous reports which aimed to develop guidelines, submitting this report will not mark punctuation to the activities. In order to promote social implementation of AI in our country, it is necessary to continue to listen to motivated efforts related to “safe, secure and trustworthy implementation of AI in society”, have free and vigorous discussions, and organize them into “shared knowledge”. We hope that interviews will be held in the future for those who did not make it this time. We would also like to have the opportunity to report on and discuss the progress of future initiatives with those who cooperated in interviews for this report. The current trend of AI initiatives in relation to COVID-19 measures have been surveyed and outlined in Chapter 1, 1., but specific initiatives shall be discussed through interviews in the future[117].

Finally, in advancing the “safe, secure and trustworthy implementation of AI in society,” various “collaboration” once again seems to have become important perspectives. For example, collaboration with organizations and groups in various fields to resolve social issues, the collaboration between developers and users in AI development, the collaboration between organizations to promote AI development and utilization within the company, collaboration with

[117] As the social situation changes over time, changes that should be emphasized, such as the COVID-19, are abrupt and impactful, far beyond the previously anticipated scope, and are covered in this report. It is expected that such a sudden change will occur in the future and that society must respond quickly. In such a case, it would be necessary to take emergency measures to avoid serious damage even if some of the principles are relaxed. There was an opinion that in the future, it may be necessary to consider the priority and importance of each principle by generalizing the lessons learned from the COVID-19 response. For related descriptions, see 'Report 2019' page 45.

various stakeholders in governance. The Conference is also a meeting body formed by the collaboration of various members[118], and the interviews were conducted in collaboration with the people who cooperated. With "Collaboration" as an important perspective, the Conference will continue to promote initiations toward "safe, secure and trustworthy implementation of AI in society".

[118] At this Conference, there were opinions from the members on the necessity to consider the following points:
- As the range of common sense that AI can handle expands in the future, society will correspondingly expand its reliance on AI autonomous judgment. How about discussing the impact on society based on the possibility that AI will change from a tool of humanity to a partner? Also, how about interviewing AI researchers for that purpose?
- In response to the COVID-19 measures, the use of ICT technologies, including AI, are being promoted in the short term, but issues such as privacy and decision making will likely be reconsidered. How about conducting discussions foreseeing those reconsiderations at this Conference?
- In response to the formulation of the AI Principles, how about considering measures to form a trust that holds the key to the proper and smooth formation and use of AI networks?

(Reference) “Future Issues” as listed in Report 2019

1. Matters concerning the sound development of AI networking

(1) Dissemination and development of the Draft AI R&D Guidelines and the AI Utilization Guidelines:
Holding symposiums to disseminate the AI R&D Guidelines and the AI Utilization Guidelines, and disseminating detailed explanations to realize principles in international frameworks

(2) Follow-up of discussions on principles and guidelines for AI development and utilization
Follow-up and ongoing review of international discussions on principles and guidelines for AI development and utilization

(3) Issues related to environmental improvement addressed by relevant stakeholders:
Cooperation among stakeholders, sharing of best practices, consideration of desirable legal systems, etc.

(4) Ensuring smooth coordination between AI systems or services:
Consideration of the scope of relevant information expected to be shared among related stakeholders

(5) Securing a competitive ecosystem:
Continued monitoring of relevant market trends

(6) Protection of the interests of users:
Consideration of ideal mechanisms for the voluntary provision of information from developers to users, and consideration of ideal mechanisms to protect users, etc. (insurance, etc.)

2. Matters related to the evaluation of social and economic impacts of AI networking

(1) Scenario analysis of social and economic impacts of AI networking:
Continuous implementation and international sharing of scenario analysis

(2) Establishment of indicators to assess the impact of the progress of AI networking and indicators to assess richness and happiness:
Study for the establishment of indicators

(3) Fostering social acceptance regarding the utilization of AI systems:
Continued monitoring of the degree of acceptance of AI utilization in society

3. Issues related to humans in a society where AI networking is progressing

(1) A study on the desirable relationship between humans and AI:
Consideration of the desirable division of roles between professionals (doctors, lawyers, accountants, etc.) and AI systems

(2) A study on the desirable relationship among stakeholders:

Consideration of the desired distribution of responsibilities in the event of actualization of AI risks

(3) Establishment of safety nets:
Continued monitoring of trends in the labor market and consideration of desirable ways to prevent disparities such as income redistribution in line with the progress of AI networking