AI Data Consortium　Support | Microsoft Japan

2023

# AI Data Symposium

## What will happen to the intellectual property of data with the advent of generative AI

While generative AI is attracting attention, on December 22, 2023, the AI Data Consortium held a symposium on AI by prominent researchers and experts in Japan.

In addition to the basics of generative AI and the current state of system development, a panel discussion was held with speakers after explaining policies related to AI and data, regulatory trends and legal issues surrounding AI.

Director of the AI Data Consortium / Professor, The University of Tokyo
Toshiya Watanabe

Director of AI Data Consortium / Microsoft Japan, National Technology Officer
Kenzaburo Tamaru

Professor, Faculty of Law, Waseda University
Tatsuhiro Ueno

Digital Agency Planning Officer, International Data Strategy
Maiko Meguro

Attorney at TMI Associates
Yuto Noro

Director of AI Data Consortium / Partner, PwC Consulting LLC
Takuya Fujikawa

*Honorifics omitted

Director of AI Data Consortium / Professor Toshiya Watanabe, The University of Tokyo

## Purpose of the AI Data Symposium

The AI Data Consortium was established in 2019 and focuses on the utilization of AI data. The organization is developing a learning data infrastructure for AI and is exploring new ways to deliver data through a platform called AIDC Data Cloud. We are also utilizing limited data provided based on the Unfair Competition Prevention Act and introducing a simple contract preparation system.

The consortium is also actively working on issues from the perspectives of AI data legislation, governance, and intellectual property rights, and will host a symposium on AI data utilization to address these issues. Here, we will hear the opinions of experts on a wide range of topics, including generative AI, copyright, international trends, and personal information protection.

Under the leadership of Chairman Ken Sakamura, the consortium aims to harness the full potential of AI and solve related challenges.

## Generative AI and Changing system development

Speakers

General Incorporated Association
Director of AI Data Consortium

Microsoft Japan,
National Technology Office

Kenzaburo Tamaru

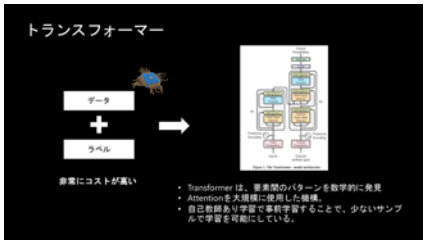
Microsoft is mainly in charge of data infrastructure in the AI Data Consortium. They introduce in detail the background to generative AI and how system development is about to change. AI's capabilities, such as object recognition, speech recognition, machine reading comprehension, and machine translation, have been achieved as equal to human quality on benchmark from 2016 to 2019. However, the quality of Japanese in natural language processing is not sufficient compared to English and other Western languages.

Generative AI has been widely featured in the media in recent years.

However, it is by no means a new area of research. It is well known that SCIgen, an automatic paper generation program published in 2005, passed peer review in academic journals. In addition, Microsoft Research has conducted a research project in 2016 that studies Rembrandt's paintings and generates images in the same painting style. AI-generated technology has improved significantly today, resulting in a quality that makes it very difficult to distinguish between real subjects and generated images.

The difference between conventional AI and AI using large-scale modeling that has been attracting attention in recent years is that conventional AI focuses on a specific domain and learns to improve accuracy, while generative AI can be used in a wider variety of fields. With the advent of data transformers, automated research on labeling and data organization is accelerating. GPT has evolved and improved its performance due to the invention of Transformers and other inventions, as well as the increase in the number of parameters.

In the past, a specific model was used to deal with a specific problem, but now it is possible to respond to multiple problems using a large-scale model as a base.




Data transformers enable self-supervised learning with less sample.

The reduction in the cost of extremely large data storage and computational resources required to train the large-scale model that forms the basis of this large-scale model has greatly contributed to the progress of research and development of large-scale models.

Microsoft's Azure OpenAI provides a service that allows users to use their data in a completely closed environment, allowing them to use AI with a focus on transparency, fairness, and privacy protection. When dealing with intellectual property, it is common for data to be stored and processed across multiple countries and regions. These variations in where computational resources are hosted and international regulations are influencing these challenges.

As an example of change made by generative AI, Microsoft Security using Sentinel and log monitoring. With the advent of generative AI, operations can now be directed in natural language, and the user interface has changed significantly. For example,based on the natural language instruction "Summarize it in PowerPoint format", it is now possible to visualize the flow of the attack and create a summary. In the past, report creation was done manually based on research and analysis, but generative AI is about to change the way user interfaces are conducted.

Microsoft Sentinel now uses generative AI to visualize monitoring and create summaries.

## Generative AI and Copyright

Speakers

Professor, Faculty of Law,
Waseda University

Tatsuhiro Ueno

The debate on AI and copyright can be divided into two main issues: the possibility of copyright infringement through information analysis, and the presence or absence of copyright protection for AI products.

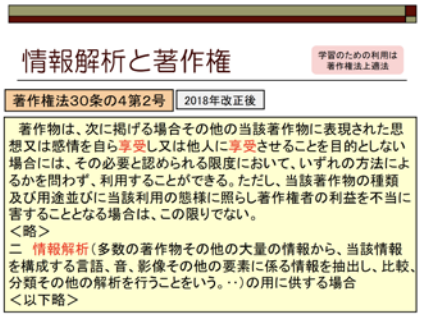
First, there is the issue of copyright infringement in information analysis. The question is whether the act of collecting content from the Internet without permission is at risk of copyright infringement in both the learning and output stages. In particular, if the AI uses text, images, etc. on the Internet to learn, the question arises whether it may infringe the copyright of these contents. Of course, in the case of private use, permission is not required to use the copyrighted work of others, but in the case of organizations and companies, it is a problem because it does not fall under private use.

Next, there is the issue whether the AI product is copyrighted.This is because it is generally believed that content created entirely autonomously by AI is not copyrightable. However, there is also the idea that if a human makes a creative contribution, such as trial and error to input prompts, in the process of generating content by AI, the AI product should be recognized as a copyrighted work. In some countries, such as the United Kingdom, there are also legislative precedents that grant copyright to content generated entirely by AI, even without human creative involvement.

Regarding the use of copyrighted works for AI learning,in Japan, Article 30-4, Item 2 of the Copyright Act permits the use of copyrighted works for information analysis under certain conditions. This provision was introduced in 2009 as Article 47-7 of the Copyright Act and amended in 2018. Currently, this provision allows information analysis as long as it does not unduly harm the interests of the copyright holder.

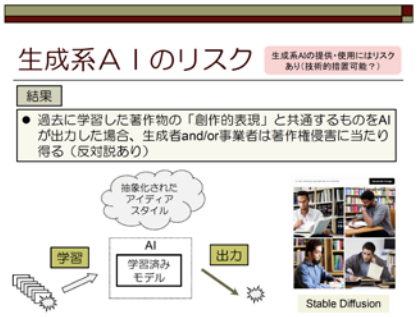
Japan was the first country in the world to introduce copyright restrictions for information analysis. Subsequently, many

countries, including the United Kingdom, Germany, Switzerland, and Singapore, introduced similar provisions. The EU has made it mandatory for member states to introduce information analysis provisions by 2021. While it is necessary to be careful about strict comparisons, internationally, Japan's regulations are considered to be relatively extensive.
In particular, Japan's regulations has the feature that the subject is not limited to the research institutes, and that it is not uniformly prohibited to use it for commercial purposes. In the EU, there is a provision that the right holder can refuse analysis only for commercial use, but there is no such provision in Japan. In addition, in Japan, there is no restriction that information can be analyzed only for legally accessed content. As a result, this provision may also apply to the analysis of copyrighted works accessed in breach of contract or bypassing technical measures, as well as information obtained from illegal sources.




In Japan, the Copyright Act permits the use of copyrighted works for information analysis under certain conditions.

In addition, under Japan regulations, creating and providing datasets to others for information analysis and selling analyzed datasets may also be subject to this provision. It can be said that this makes it possible to take advantage of the benefits of information analysis.
However, there are criticisms of such a broad provision, and there is an opinion that the copyright law needs to be revised. However, this provision is only for the learning stage and does not allow up to the output stage, so this point should not be misunderstood. In other words, if generative AI outputs something similar to the creative representation of the content it learns from, such output may infringe copyright.
Of course, in copyright law, there is a principle called the "dichotomy of ideas and expressions." This protects only concrete expressions, not the ideas themselves. For example, data from newspaper articles, the style of illustrations, and the expressive style of music are not subject to copyright protection. Ideas and methods of expression are important for creators, but copyright protection is a strong monopoly for a long period of 70 years after the death of the author, and if copyright infringement is recognized, there is a risk of criminal penalties, including imprisonment, as well as injunctions and damages, so abstract styles and ideas are not protected by copyright law. This is an undisputed principle around the world.
Regarding AI, even if the AI that comprehensively analyzes works created by specific authors generates a new work in the style of Ghibli, Banksy, or Beatles, it is common to think that it does not constitute copyright infringement unless the creative essence of the original work does not exist. On the other hand, if an AI product outputs a specific representation of the learning source work as it is, such as Ever filter, an app that processes images in the style of anime directed by Makoto Shinkai, which was once a problem, it is clearly copyright infringement.
Article 30-4, Item 2 of the Copyright Act regulates the use of copyrighted works for learning, but if learning is carried out with the intention of outputting content that constitutes copyright infringement, this provision does not apply because it does not constitute non-enjoyment use as defined in the main paragraph of the same article. However, in addition to Article 30-4, Item 2 of the Copyright Act, Article 47-5, Paragraph 1, Item 2 was also stipulated in the revision of the law in Heisei 30. Article 47-5, Paragraph 1 also targets search engines, for example, and permits minor output such as displaying a snippet of text or a thumbnail of an image as a search result under certain conditions, so I think there is room for this provision to be applied to generative AI that has a purpose of enjoyment.
In this way, for the debate about AI and copyright, careful consideration is required, distinguishing between both learning phases and output phases. Personally, I believe that it is important to prevent infringement at the output stage, but that freedom at the learning stage should be maintained, but we must continue to pay attention to international trends and find the right balance.
As for AI and copyright, the problem is that whenever AI learns someone else's work and outputs something that co-exists with that "creative expression", it is a copyright infringement. If the AI is said to be relying on a work that it has learned in the past, the output is considered to be a copyright infringement.
However, if this is the case, this is a risk in the use of AI, so there is a lot of discussion.




If AI outputs "creative expressions" of works learned in the past, it may constitute copyright infringement.

This issue is not new, and has been discussed at the Intellectual Property Headquarters in 2016 and through the revision of the law in Heisei 30. And if there is a risk of copyright infringement by AI, it can extend to users who use AI, AI service providers, and AI development companies. For example, recently, AI services such as DALL-E3 have taken measures not to be able to input the name of a specific character, such as Pikachu, if someone tries to output it.
In the United States, there is no clear provision for information analysis, and in the event of a lawsuit, the court will decide. In fact, litigation has begun, so it is expected that discussions on the copyright of AI learning will proceed in the United States in the future.
There are various opinions regarding the interpretation of the information analysis provisions in Japan's copyright law. In particular, the proviso that it does not unduly harm the interests of the copyright holder is the focus of the discussion.
However, regardless of the interpretation of the information analysis provisions, that is, regardless of whether or not copyright is extended, it is possible and useful to have a data provision contract for the purpose of information analysis. Already, we are seeing contracts with publishers and newspapers to provide digital data suitable for analysis. As such, it is important for data owners to facilitate data delivery agreements. On top of that, it is necessary to take measures at the output stage, which can be said not only for infringement of rights such as copyright infringement, portrait infringement, and publicity infringement, but also for various illegal and harmful information. With regard to the freedom of the learning stage, I think we should maintain this to the last, and then concentrate our wisdom and technological ingenuity to prevent illegal and harmful information at the output stage.

## AI Development and Utilization and Data Governance

Speakers

Digital Agency
Planning Officer,
International Data Strategy

Maiko Meguro

Data governance for artificial intelligence (AI) can be broadly divided into two categories: "data issues in the formulation of policies and rules related to the development and use of AI" and "data governance related to all stages of AI development and use."

The former includes specific examples as AI strategy in 2022, the summary of issues (provisional) issued at the AI Strategy Meeting, and economic measures in Japan. These address specific data issues and provide draft guidelines for businesses. The latter includes strategic documents and relevant laws and regulations (e.g., intellectual property laws, limited availability data, privacy laws, etc.) that are relevant to the overall data strategy.

Internationally, the Government of Japan has taken the lead in Data Free Flow with Trust (DFFT), which is international initiatives. This is an initiative to accept the reality that data crosses borders, regardless of differences in legal systems such as privacy, security, and intellectual property, and to constructively discuss principles for building trust between countries, specific legal systems, specific international guidelines, and specific policies.

In terms of specific domestic policies, the AI strategy in 2022 shows an awareness of the problem that although data is accumulated in each field, it is not being used as effectively as in other countries. In addition, at the AI Strategy Meeting (May 26, 2023), the emphasis was similarly on enhancing the available data and building a data linkage platform, which is a prerequisite for AI.

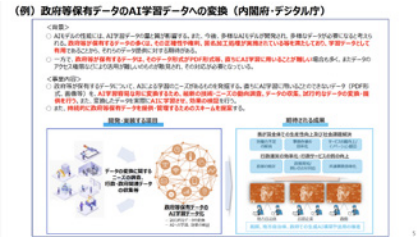
In addition, the Comprehensive Economic Stimulus Plan (2023) focuses on AI, and specific initiatives related to AI and data governance are being promoted in many areas, such as "responding to risks," "promoting the use of A(I, mainly generative AI)," and "strengthening AI development capabilities." The Digital Agency is making concrete efforts to verify the technology and improve the usage environment for the business use of generative AI.

As mentioned above, various discussions and initiatives are taking place in AI data governance both in Japan and overseas, and each of them is collaborating to improve the environment for better data utilization. It is believed that discussions on data governance should proceed from various perspectives, such as policies and rules related to data, and AI development and use in the future.

Next, I will explain the "Action Plan for the Development and Collaboration of Public-Private Data in the Age of AI" announced by the Digital Agency of Japan on December 20, 2023. This action plan focuses on the theory of data governance at all stages involved in the development and use of AI.

The development of data governance has gone through three stages. The first focus will be on core data for the realization of a digital society, and "Comprehensive Data Strategy(June 2021)." Next, after the establishment of the Digital Agency (September 2021), based on this strategy, Data Free Flow with Trust (t DFFT). Through the "Priority Plan for the Realization of a Digital Society (June 2023)," we are proceeding with specific implementations and initiatives in priority areas. However, these efforts are still in the middle stage, and in response to the rapid development of generative AI technology, the G7 Gunma Takasaki Digital and Technology Ministers' Meeting agreed to establish an Institutional Arrangement for Partnership (IAP) on DFFT.

In the "Action Plan Background" section, it describes how to respond to the rapid development of generative AI technology. Converting the massive amount of data held by the government into a format that can be used by AI is a time-consuming and labor-intensive task. The Digital Agency plans to open up the data held by the government and promote the use of past administrative data.




The Digital Agency is planning to open up the data held by the government and others and convert it into AI training data.

The Data Strategy Action Plan has two main goals: "Develop and open high-quality and easy-to-use data" and "Develop tools and mechanisms that enable the use and linkage of the data with confidence." In addition, it is important not only to develop and collaborate with domestic data infrastructure, but also to collaborate internationally.

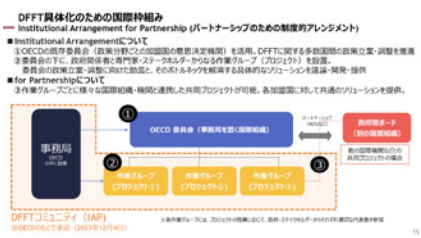
In 2023, the G7 Hiroshima Summit was held, where discussions were held on issues related to international data governance and data utilization. DFFT is defined as "reliable and free flow of data" and relates to the cross-border flow of international data. This concept was proposed by Prime Minister Abe at the Davos Summit in January 2019 and agreed upon at the G20 Osaka Summit in June 2019. The cross-border data problem is complicated by differences in privacy, security, and intellectual property legislation from country to country. While data localization legislation is on the rise, the need for country-specific placement of data centers and country-specific management of data collection and processing can overturn the fundamental assumptions of cloud technologies.

At the 2019 Davos Forum, Prime Minister Abe said that personal data, intellectual property, and data related to national security should be protected, but impersonal and anonymous data such as healthcare, industry, and transportation should be freely circulated. The G20 Osaka Leaders' Declaration in 2019 highlighted DFFT and declared that "continuing to address issues related to privacy, data protection, intellectual property rights and security can promote the free flow of data and strengthen consumer and business trust."

Expertise in different areas, such as privacy, security, data protection, and intellectual property, influences the flow of data, and how to coordinate and facilitate them can be challenging. In particular, AI-related developments and advances in cloud technology are rapidly advancing and are widely incorporated into our daily lives, but discussions on international governance among governments and governments have not kept up.

AI developers and the research community are calling for steps to ensure the data quality of AI training. Contains disinformation or information that infringes intellectual property rights can lead to legal consequences later on. Against this background, the introduction of "rating tags" to indicate

the safety of data has been proposed, but the setting of standards requires international discussion. Even if it is technically possible, if appropriate evaluation criteria are not determined, there is a risk that the international flow of data will stagnate.
In addition, there are various stages of AI development and service provision, with different regulatory trigger points. However, if regulations are concentrated at a specific stage, it will be inefficient when viewed from the perspective of the entire life cycle. In addition, it is difficult to talk about AI and applications in data governance discussions, and discussions on individual regulations such as privacy, security, and intellectual property are not shared from a holistic perspective.
Under these circumstances, the Japan government is looking for ways to connect it to international governance discussions. At the 2023 G7 Gunma Takasaki Digital and Technology Ministers' Meeting, it was acknowledged that it is difficult for experts from different fields to discuss with each other, and that there are gaps in international governance. Therefore, it is proposed to gather evidence on specific issues and barriers to the reality of cross-border data access, and to look for the way to use the concrete way to eliminate gaps, not abstract rules. However, since many projects tend to be temporary, an approach to set up a permanent secretariat for governance and to solve the individual problem is important. Also, the establishment of an international system for the realization of Data Free Flow with Trust (DFFT) at the summit level was also approved, and it was decided to place it in the OECD. It is believed that this will make it easier to advance discussions with developed countries that have similar values.
The Japan government aims to back up these movements and build large projects and organizations to come up with a wide range of solutions with a wide range of experts.



The International Framework for the Implementation of the DFFT (IAP) was approved by the OECD in December 2023

## AI and Issues in the Personal Information Protection Law

Speakers

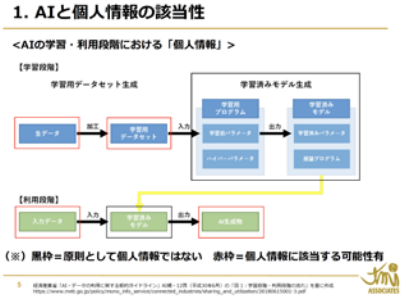
Attorney at TMI Associates

Yuto Noro

I thought it would be useful to check the relationship between AI and the Personal Information Protection Act in order to consider how to establish regulations for AI in the future, so I would like to talk about the theme of "AI and the Personal Information Protection Act" today.
Today, I would like to first talk about the extent to which Japan's Personal Information Protection Law can affect AI in regulation. I will explain that this law does not directly regulate AI, but indirectly regulates it through AI input and output data. Next, I would like to talk about the extent to which Japan's Personal Information Protection Law regulates AI at this time. This law regulates AI at each stage of acquisition, use, and provision of personal information, but there are practical countermeasures for all of them, and I plan to explain that the regulations are not so strict as to impede the use of AI at this time.

### 1.Applicability of AI and personal information

First, I would like to discuss the issue of the applicability of AI and personal information in order to confirm the extent to which Japan's Personal Information Protection Law can affect AI. Here, I would like to explain two issues, the first of which is the applicability of AI to personal information. I think it raises an uncomfortable question about whether AI itself personal information is, but in other words, it can be explained as a debate about whether AI algorithms and parameters contain personal information. The second is the "applicability of personal information in input and output data," which is the issue of whether the input data and output data are personal information.
There are two elements to the definition of personal information: (1)Information about a living individual" and "(2) Information that can identify a specific individual by the description contained in the information, or information that includes an individual identification code."
The first point of contention is the applicability of personal information in AI. In the debate on whether this AI is personal information, the factor (1) is important, and it is said that it may not be personal information because it is not related to the individual in the first place. Here, if AI is defined as an algorithm and a parameter, the Personal Information Protection Commission explains that it does not fall under personal information because it does not fall under (1) as far as the correspondence between the specific individual is excluded, regarding the latter "parameter".
The second point of contention is the applicability of input and output data to personal information. A variety of things can be input and output data, so it is a case-by-case decision. On top of that, we believe that input and output data that includes (1) and (2) fall under the category of personal information.
To illustrate these issues, let's take AI as an example of disease prediction AI in the medical field, where it inputs patient information and outputs problematic symptoms as images. In this case, while the AI itself is considered not to fall under the category of personal information, these inputs and outputs meet the requirements of (1) and (2) and may fall under the category of personal information.
Next, we will explain what is "personal information" and what is not evaluated as "personal information" at the learning and use stage of AI in a diagram in accordance with the process of AI learning and use.



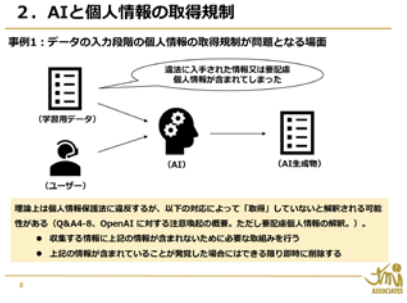
[Figure] "Personal information" at the stage of learning and using AI.

First, in the [Learning Stage], the "raw data" is processed to create a "training dataset". This The "training dataset" is input to the "training program" to generate the trained

model, and the "trained model" is output.
Next, in the [Usage Stage], the user inputs "Input Data" to the "Trained Model" output in the [Learning Stage], and the "AI Product" is output.
The above is the process of learning and using AI that is generally assumed but based on the discussion of the applicability of personal information earlier, it can be organized as follows, first, the items in the red frame in the figure, "raw data", "learning dataset", "input data", and "AI product" are the input and output data I mentioned earlier. It may fall under the category of personal information. On the other hand, the item in the black frame is "Learning Program" and the "trained model" is the AI itself that I mentioned earlier, and in principle it is not personal information. As shown in these figures, the Act on the Protection of Personal Information does not cover the AI itself, which is enclosed in a black frame, but the peripheral part of the AI surrounded by a red frame, that is, the data input to the AI and the data output.
Based on the discussion so far, if we summarize the relationship between AI and the Personal Information Protection Act, it can be explained that this law does not directly regulate AI, but indirectly regulates AI through input and output data.
This is due to the EU's data protection law, the GDPR (General Data Protection Regulation) also regulates personal data, so like the Personal Information Protection Law, it does not directly regulate the AI itself, but inputs it to the AI. It is considered to be restricting the data to be output. And recently, the EU's AI bill, which has been politically agreed, allows for direct regulation of the AI itself, which is enclosed in a black frame in the diagram that was not covered by the GDPR. I think the background to this is that the existing GDPR did not directly regulate the dangers posed by AI as long as it covered personal data.
As a result of that the EU's AI bill regulates AI itself, we have been able to apply regulations to various areas that could not be covered by the existing GDPR. For example, the existing GDPR is a law that targets personal data, and its rules are built around the protection of individuals in the background, but if AI adversely affects society as a whole or a part of it, it is difficult to regulate it because it is not linked to the protection of individuals. On the other hand, the EU's AI bill directly regulates AI itself, so it is possible to ban algorithms that adversely affect society as a whole. We believe that the EU's AI Bill will serve as a reference when considering regulations on AI in Japan.

**2.Regulation of AI and the acquisition of personal information**

Next, in order to discuss the extent to which Japan's Personal Information Protection Law regulates AI at this time, I will explain from the regulation of AI and the acquisition of personal information. Here, the regulation of the acquisition of personal information at each of the data input and output stages is an issue. There are two regulations on the acquisition of personal information that can be problematic in any of these situations. The first is the prohibition of unauthorized acquisition of personal information as stipulated in Article 20, Paragraph 1 of the Personal Information Protection Law. This prohibition stipulates that personal information must not be obtained by unauthorized means. The another is a regulation on the acquisition of special care-required personal information stipulated in Article 20, Paragraph 2 of the Personal Information Protection Law. As a general rule, this regulation stipulates that special care-required personal information must not be acquired without the consent of the individual.
As a situation where these regulations are problematic for the acquisition of personal information, we will first introduce "Case 1: Situations where the regulation of the acquisition of personal information at the data entry stage is a problem". In this case, if the AI training data contains illegally obtained information or special care-required personal information, the problem is that these two regulations will be violated.
This problem has not arisen recently as a problem with generative AI, but has existed as a problem that is difficult to solve since the 27th revision of the Personal Information Protection Law established new regulations for the acquisition of special care-required personal information. For example, in conversational AI, there is a possibility that a third party's special care-required personal information will inevitably be included in the input data of the AI, but in that case, it is practically difficult to obtain consent for the special care-required personal information from a third party, and it does not necessarily fall under the exception of the regulation, so there was a concern that conversational AI would violate the regulations for obtaining special care-required personal information.
However, during the debate on the personal information protection law surrounding generative AI in recent times, this point was raised again, and the discussion deepened. What is important in considering this issue is the "Summary of Alerts to OpenAI" published by the Personal Information Committee on June 2, Reiwa 5. In the outline of this warning, based on the premise that OpenAI physically acquires special care-required personal information from users and third parties without the consent of the user, a certain amount of prior. If you take action after the fact, you can understand that it is a legal arrangement that prescriptively evaluates that you have not acquired personal information in the first place. In the outline of this alert, it is required to take measures such as taking necessary measures to prevent the above information from being included in the information to be collected as a preemptive response, and as a follow-up response, if it is discovered that the above information is included, it is required to delete it as soon as possible. If the Personal Information Protection Commission is able to take these preventive measures and take follow-up measures, it can be understood that it has not acquired personal information in the first place and is not subject to the acquisition regulations related to special care-required personal information. Such an interpretation is theoretically questionable, but at least as an administrative interpretation, it can be understood that such a legal arrangement is made so as not to interfere with practice. Although this interpretation is an interpretation of the regulations regarding the acquisition of special care-required personal information, I think the same can be said about the interpretation regarding the prohibition of unauthorized acquisition.

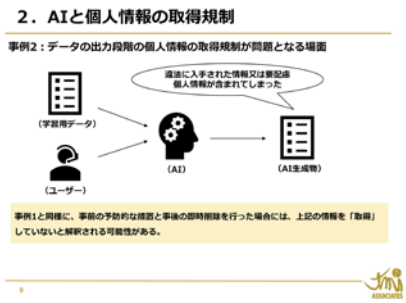


Case 1: Situations where restrictions on the acquisition of personal information at the data entry stage are problematic.

Next, we will introduce "Case 2: A situation where the regulation of the acquisition of personal information at the data output stage becomes a problem". In this case, it is assumed that illegally obtained information or special care-required personal information is included in the AI product instead of the training data. For example, when using ChatGPT, a third party's personal information has been output for some reason.

In this case, there is an argument that the user who has acquired the AI product will be subject to the regulation on the acquisition of special care-required personal information and the prohibition of unauthorized acquisition. However, as in Case 1, if precautionary measures are taken in advance and immediate deletion is taken after the fact, it may be interpreted that the above information was not acquired in the first place.

If this is the case, it will be possible to comply with the rules of the Personal Information Protection Law in both Cases 1 and 2, so I think it can be said that the Personal Information Protection Law is not a regulation that hinders the use of AI in the context of AI and the acquisition of personal information.




Case 2: Situations where restrictions on the acquisition of personal information at the data output stage are problematic.

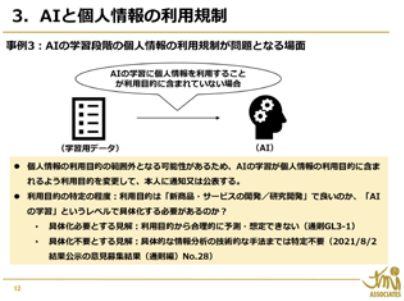
### 3.Regulation of the use of AI and personal information

Next, I will explain the regulations on the use of AI and personal information. One of the problematic usage regulations is the regulation of the specification of the purpose of use and the handling within that scope. First of all, Article 17, Paragraph 1 of the Personal Information Protection Act stipulates that the purpose of use must be specified as much as possible. In particular, when analyzing information such as behavior and interests related to the individual, it is understood that the purpose of use must be specified to the extent that the person can predict and assume (Guidelines on the Act on the Protection of Personal Information (General Rules) 3-1-1 (※1). This interpretation was added to the guidelines based on consideration at the time of the revision of the Personal Information Protection Law in Reiwa 2. Prior to that, there was an interpretation that in order to specify the purpose of use, it was necessary to specify the final purpose of use of personal information, so it was not necessary to specify in detail the process leading up to the final purpose of use. However, in the case of analysis of individual behavior monitoring, it is necessary to explain not only the final purpose of use but also the analysis of the process, which has been added in the interpretation based on the consideration at the time of the revision of the Personal Information Protection Law in Reiwa 2. In addition, Article 18, Paragraph 1 of the Personal Information Protection Act stipulates that, in principle, personal information must be handled within the scope necessary to achieve the specified purpose of use, and these regulations are established for the identification of personal information and the handling within that scope.

Another usage regulation is the prohibition of improper use. Article 19 of the Personal Information Protection Law stipulates that personal information must not be used in a manner that may encourage or induce illegal or unjust acts. This provision is a prohibition added by the amendment to the Personal Information Protection Law in Reiwa 2, and broadly prohibits the improper use of personal information, but it may be applied in situations where AI is used.

One example of the use of personal information where these regulations can be problematic is "Case 3: A situation where the regulation of the use of personal information at the AI learning stage is a problem". In this scenario, it is assumed that the use of personal information for AI training is not included in the existing purpose of use. In this case, the use of personal information for AI learning will be handled outside the scope of the existing purpose of use of personal information, so it is necessary to change the purpose of use in order not to violate the Personal Information Protection Law. Therefore, in practice, I think it is necessary the purpose of use is changed so that AI learning is included in the purpose of use of personal information, and the person is notified, or it is publicized.

There is a debate about whether the purpose of use should be for the general purpose of "development/research and development of new products and services" or whether it is necessary to have a specific purpose of use at the level of "AI learning" to the extent of specifying the purpose of use in this case. The view that needs to be concretized is that AI learning cannot be reasonably predicted or assumed for general purposes, and that AI learning should be specified until the learning of AI. The basis for this view is a specific interpretation of the purpose of use when analyzing information such as behavior and interests about the person mentioned earlier, but there are doubts as to whether this interpretation will be applied to AI learning in the first place. The view that it is not necessary to specify the purpose of use is understood that it does not need to specify the specific technical method of information analysis when specifying the purpose of use, so it is not mandatory to mention AI. The rationale for this view is No. 28 of the Guidelines for the General Rules in the Result of the Call for Opinions published by the Personal Information Protection Commission on August 2, 2021. In this regard, if you read the privacy policies of some companies, you will find that some AI development startups specify "AI learning", but there are not many that specify to that extent as an overall trend. I have the impression that there are many descriptions of the purpose of use with a granularity.



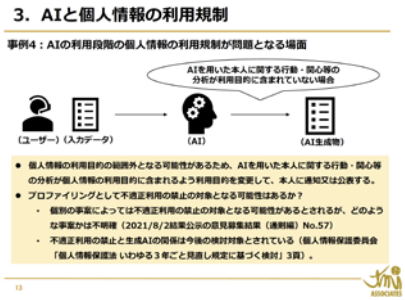

Case 3: Situations where the regulation of the use of personal information at the AI learning stage becomes a problem.

Another example is "Case 4: A situation where the regulation of the use of personal information at the stage of using AI becomes a problem." In this situation, it is assumed that the purpose of use does not include analysis of behavior and interests related to the person using AI. In this case,

since there is a possibility of violating the Personal Information Protection Law as handling personal information outside the scope of the purpose of use, it is required to change the purpose of use so that the analysis of the behavior and interests of the person using AI is included in the purpose of use of the individual, and to notify or announce it to the person. As in Case 3, it is possible to comply with the rules of the Personal Information Protection Law by notifying or announcing to the person, so I think it can be said that the Personal Information Protection Law does not impede the use of AI even in the case of the use of AI and personal information.

In this case, the so-called profiling is being performed and there is also an interesting question as to whether it is subject to the prohibition of improper use depending on the manner in which it is done. Profiling generally refers to the automated analysis of private aspects of an individual. Whether to establish profiling regulations in Japan's Personal Information Protection Law has been discussed since around the 27th revision, but at this time there are no profiling regulations themselves. Then, as for whether profiling falls under the prohibition of inappropriate use, No. 57 of the Guidelines for the General Rules in the Results of the Call for Opinions, which was announced by the Personal Information Protection Commission on August 2, 2021, may be subject to the prohibition of inappropriate use depending on the individual case. In general, I think it is reasonable to interpret that it may apply depending on the individual case, but at this point it is not clear in what specific cases it falls under the prohibition of inappropriate use.

It has been suggested that the relationship between the prohibition of inappropriate use and generative AI may be the subject of future consideration in the examination based on the three-year review regulations of the Personal Information Protection Commission (Personal Information Protection Committee, "Examination based on the so-called three-year review provisions of the Personal Information Protection Law" (November 15, Reiwa 5), p. 3) ). If, as a result of this study, the use of generative AI that falls under the category of inappropriate use is clarified in guidelines, Q&A, etc., it may be possible to create a new regulation on generative AI. For example, the EU's AI Bill sets out a list of cases where the use of AI is prohibited for unacceptable risks, and it is conceivable that it will be something like that list.

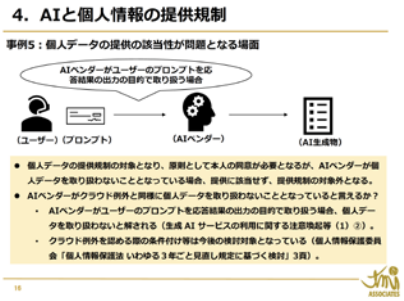



Case 4: Situations where the regulation of the use of personal information at the stage of using AI becomes a problem.

**4.Regulation of the provision of AI and personal data**

Finally, we will explain the regulations on the provision of AI and personal information. There are two regulations on the provision of personal information that can be problematic here. The first is the general regulation of personal data on third parties stipulated in Article 27, Paragraph 1 of the Personal Information Protection Law. When providing personal data to a third party, in principle, the consent of the person is required. However, if the AI vendor does not handle personal data, it can be considered as a so-called cloud exception, and there is an interpretation that it does not mean that the personal information is “provided” to the AI vendor in the first place. (Q&A: 7-53 regarding the "Guidelines on the Act on the Protection of Personal Information", Caution regarding the use of generated AI services, etc. (1) (2). Even if it is understood that personal data is being provided, there is an exception in the case of outsourcing the handling of personal data (Article 27, Paragraph 5, Item 1 of the Personal Information Protection Act). The other is the regulation of the provision of personal information to third parties in foreign countries stipulated in Article 28 of the Personal Information Protection Law. When providing personal data to a third party in a foreign country, in principle, it is necessary to obtain consent after providing certain information to the person. In the case of provision to a third party in a foreign country, the entrustment of the collection of personal data is not excepted. Of course, even if personal data is provided to a third party in a foreign country, if the recipient meets the standards for system development equivalent to Japan law, as an exception, this regulation does not apply.

Here are two examples. The first is "Case 5: A situation where the applicability of the provision of personal data is questionable." In this scenario, the user inputs a prompt into the AI and the AI product outputs it, but it is assumed that the AI vendor handles the user's prompt only for the purpose of outputting the response result. In principle, the consent of the user is required to provide personal data to the AI vendor, but if the AI vendor is not supposed to handle personal data, the cloud exception will apply. In this regard, according to the "Warning on the Use of Generative AI Services" published by the Personal Information Protection Commission, if an AI vendor treats the user's prompt only for the purpose of outputting response results, it can be understood that it does not fall under the category of "provision" of personal data, as with the cloud exception, and is not subject to regulations regarding the provision of personal data. However, in other cases, it is unclear under what circumstances the use of AI vendors' services is exempt from the provision regulations. For example, if the AI is fine-tuned on the user side, it may not be applicable if the AI vendor handles the user's prompt only for the purpose of outputting the response result, and it may not be interpreted in the same way as a cloud exception. It has been suggested that the cloud exception may be considered in the future in the review based on the three-year review provision by the Personal Information Committee, which I mentioned earlier.




Case 5: Situations where the applicability of the provision of personal data becomes an issue.

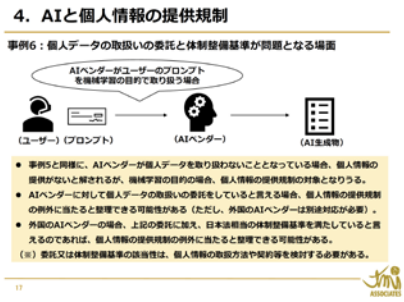
The second is "Case 6: Situations where the outsourcing of the handling of personal data and the standards for the establishment of systems are problematic." This is also a scene where the user inputs a prompt into the AI and the AI product outputs it, but it is assumed that the AI

vendor will use not only the user's prompt response, but also the prompt itself as machine learning data. In this case, according to the "Warning on the Use of generative AI Services", it cannot be interpreted in the same way as the so-called cloud exception, and it must be understood that personal data is provided to AI vendors.

There is a possibility that it can be interpreted as falling under the consignment of the handling of personal data, and it may be possible to organize it as not requiring the consent of the person. However, in the case of foreign AI vendors, regulations regarding the provision of personal data to third parties in foreign countries apply, so in principle, consent after providing information to the person is required. However, even if this regulation is applied, there is an exception if the provider meets the standards for system development equivalent to Japan law. This point will be judged on an individual basis, and it will be necessary to consider contracts with AI vendors. If this contract meets the standards for the establishment of the system, it will be possible to comply with the rules of the Personal Information Protection Law, so I think it can be said that the Personal Information Protection Law is not a regulation that hinders the use of AI even in the case of the provision of AI and personal information.

Thank you for your time.

Thank you very much for listening.




Case 6: Situations where the outsourcing of the handling of personal data and the standards for the establishment of systems become problematic.

## Japan's Generative AI Utilization Status and AI Regulatory Trends in Each Country

Speakers

General Incorporated Association Director of AI Data Consortium

Partner, PwC Consulting LLC

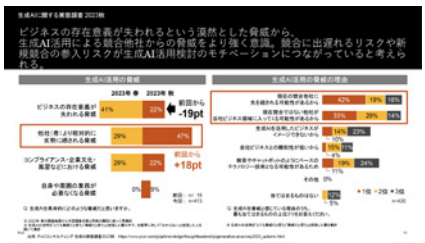
Takuya Fujikawa

PwC's "Fact-finding Survey on Generative AI Autumn 2023" is available on the web, and if you are interested, you can search for "PwC Generative AI Fact-finding Survey 2023 Autumn" and download the PDF file. This data is a valuable source of information for understanding how Japan companies perceive generative AI and what strategies they are taking.

According to the survey, awareness of generative AI among managers of large companies in Japan has risen to 96%, 73% have used it, and 87% are considering introducing it. The survey, which was conducted in October 2023 and released in December, surveyed 1,000 people in section managers and above who belong to companies and organizations in Japan with sales of 50 billion yen or more. Expectations are high for generative AI that aims to improve efficiency, and 12% of companies are already providing it as an external service.

On the other hand, about half of the companies feel that we are relatively inferior to other competing companies about generative AI, and their motivation to aim for the first time in the market has been confirmed. The amount of investment varies from several million yen to billions of yen or more, with 24% investing hundreds of millions of yen or more. Also, with 43% of companies planning to implement generative AI by March 2024 and 58% by September, companies with large investment budgets are more risk-conscious and using tools like conversational generative AI is becoming increasingly important.

While each company has its own departments leading AI governance, it typically involves departments such as information systems, security, legal, and business. The key is for various departments within a company to work together to build governance.




About half of companies feel that generative AI is inferior to competitors and is more aware of the threat

The survey results are classified into five clusters: 12% are "indifferent," 34% are "interested but not taking action," 22% are "promoting projects but postponing governance," 18% are "using it for internal operations and focusing on governance," and 14% are "company-wide support inside and outside the company." In particular, the technology and telecommunications industries are leading the way, and the automotive industry is also making positive moves. The retail industry tends to be interested but not take actions.

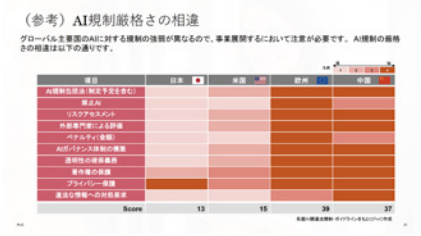
Comparing the Spring and Fall Surveys, the Technology Industry, while the industry and telecommunications continue to be pioneers, the healthcare and automotive industries are gaining momentum. The financial industry is slowing down.

In terms of use cases, the technology industry is developing applications and automated programming, the telecommunications industry is providing services for call centers, the healthcare industry is improving the efficiency of doctors' office work, and the automotive industry is supporting vehicle design. PwC is looking at emerging trends such as automated CAD generation, using generative AI for design defect checking in the construction industry, landscape design generation in the real estate industry, and product packaging and advertising production in the retail industry.

Regarding generative AI legislation in major countries, the EU and China have adopted a strict hard-law model, and the EU regulations are particularly strict. The EU aims to provide robust protection for civil rights under the GDPR, and fines for violations can reach up to 7% of sales. This is also important for non-EU companies, as it can affect all companies that provide services to EU citizens.

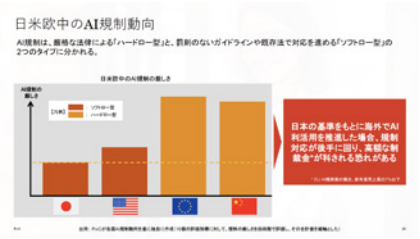
In China, the National Internet Information Office (CAC) has enforced the AI Algorithm Regulation. This has a notification system, and severe fines can be imposed in the event of a violation. Thus, state-led digital strategies for AI technologies are being

strengthened.
The U.S. has previously adopted a soft-law type of regulation, but with a new executive order, there are signs of a shift to a hard-law model. This could make AI developers obliged to test and report.
Japan has relatively loose regulations and regulations based on guidelines have been the mainstream, but with the sharing of the new "Draft Guidelines for AI Business Operators", regulations may be strengthened in the future. Japan companies need to understand overseas regulations before developing their business in the global market, so they are required to consider measures in advance.
The EU's AI Regulation sets strict regulations on prohibited and high-risk AI systems, and it includes many areas such as biometrics services, critical infrastructure, and personnel evaluation.
In the EU, providers and users are obliged to perform quality control, technical documentation, conformity assessment, and post-market monitoring.




Since the strength and weakness of regulations on AI in major global countries are different, it is necessary to be careful in business development.

In line with these developments, Japan companies operating globally need to grasp international regulatory trends and establish an appropriate governance system. In addition, it is expected that more detailed industry-specific guidelines will be required in the future. This will enable companies to drive innovation while effectively managing risk.
The EU has the strictest regulations, followed by China. While these countries have adopted a hard-law type of regulation, Japan has so far promoted a soft-law type of regulation, that is, regulation with loose guidelines. However, in Japan, new "Draft Guidelines for AI Operators" has been submitted, suggesting that this could increase the severity of regulations.
In the U.S., there have been signs of a shift to a hard-law model due to a recent executive order, although the country has adopted a loose regulation through guidelines. This means that AI developers may be obliged to test and report.




AI regulation is divided into two types: "hard law" with strict laws and "soft law" with no penalties.

Japan companies need to take into account the international business development, and to understand and adapt the regulations of each country. This is especially true in the EU, where there are strict regulations, such as specific obligations imposed on providers and users regarding AI systems.
The G7 Leaders' Statement on the Hiroshima AI Process calls for organizations developing AI systems to apply an international code of conduct, which includes risk checks and strengthening security measures.
"Improving user literacy" is also taken up due to Japan's strong demand, and it is said that users are also responsible.
In PwC's view, companies need to build their own rules with international rules in mind in addition to Japan's guidelines. In the future, it is expected that individual guidelines for each industry will be required, and it is speculated that EU regulations will move in a more harmonized direction. China is also moving toward international cooperation while continuing to operate under the leadership of the state.
In closing, it is stated that it is important to immediately develop a governance system in order to promote innovation in generative AI. Preparations for stringent regulations need to start now, and consideration is required to establish global standards.


It is expected that each country will develop a governance system and consider the establishment of global standards.

Discussion

## Panel discussion by speakers

The panel covered specialized topics related to data governance, personal data protection, intellectual property rights, cloud technologies, and international data flows. The main focus is on cross-border data transfers, the relationship between machine learning and copyright, data security and privacy issues, and international data governance approaches.

Mr. Meguro explained that through the OECD project, countries will cross borders in anticipation of the cloud, and will work with engineers and the field to solve problems. In addition, the OECD has nearly 40 member countries, and while policy coordination is required, the data topics are wide-ranging, he mentioned that some topics are appropriate to use IAP, while others can be narrowed down by frameworks such as bilateral and trilateral. We believe that challenging topics also require a rethinking of the scope of the project. Furthermore, he indicated that the issue of cross-border data transfer will lead to multilateral and multilateral discussions.

In addition, in response to a question from Mr. Watanabe about security of database, Ms. Meguro acknowledged that information related to security is basically unshakeable and that it is difficult to separate it, but he believes that the government needs to separate the frame that should deal with data discussions, the frame that should be discussed by isolating only security, and the issue that should be discussed in terms of technical protection and should use multiple tracks to advance the discussion. And he emphasized that it is necessary to create a form that satisfies everyone while he wants to make the domestic regulatory environment the standard overseas.

Mr. Ueno pointed out that if the output of machine learning is common to the creative expression of the original work, it may be a copyright infringement. Therefore, he said that the liberalization of machine learning should be maintained, but consideration should be given to ensuring that the output result is not the same as the creative expression of the learning source work. Mr. Watanabe pointed out that there are many opinions that "I don't want to be learned," and explained that many of them are not copyrighted works, and that measures are required to prevent them from becoming technically copyright infringement.

He Mr. Ueno pointed out that the copyright holder cannot refuse to study for the non-profit research purpose in the EU and the UK, but rather that the EU and the UK have more freedom to study for non-profit research purposes. He said that even if there is a strong concern about his work being AI-learned at the moment, he hopes that as time goes by, generative AI will become commonplace, and there will be no more concerns about AI learning itself.

Mr. Ueno introduced Japan's first provision for information analysis and said that with the growing awareness of the new knowledge that big data analysis can provide, it is unclear how the understanding of this will ultimately converge. However, he pointed out that Japan's theoretical framework that copyright does not need to extend to uses that no one enjoys in the first place is attracting attention from other countries, and that it may converge in that direction in the future.

Mr. Noro said that if we are complying with the law, we should not be overly afraid of flaming because of the vague fear, about a business that involves new technologies. In areas where new technologies such as AI are involved, even if the law is complied with, there is a potential possibility that it will be flamed at some point, but the question is whether the business should always be shut down due to concerns about that possibility. He said that the supervisory authorities are also paying attention to whether or not new businesses that are born every day can gain the understanding of the public, but as long as they comply with the law, they do not always take measures to stop the business even if a fire breaks out.

On the other hand, as a practical matter, in companies

Mr. Noro points out that when promoting AI-related businesses, it is also important to consider reducing the risk of flaming as much as possible in practice. As a countermeasure against flaming, he emphasized the user's point of view and said that it is necessary to consider both the law and social acceptability, and that it is important to proceed with business

with consideration not only of the law but also of social acceptability, especially in areas where new technologies such as AI are involved.

Mr. Noro also pointed out that Japan's Personal Information Protection Law does not currently pose a major obstacle to AI-related businesses. For example, when handling personal information in an AI-related business, while it may be possible to request deletion of data input and output data to AI, it is difficult to request deletion of the AI itself on the premise that AI itself is not a personal information.

He also explained that the provision of personal data to AI providers can rely on so-called cloud exceptions, which may not be subject to their regulations. However, Mr. Noro explained that if AI user companies are making adjustments in the form of finetuning, they may not be able to rely on so-called cloud exceptions, so it is necessary to be careful.

Tamaru pointed out that there is a big difference in the way companies in Japan and the United States handle data. In Tamaru's experience, U.S. companies provide data unless it's a trade secret, while Japan companies rarely provide data. He said that this is affecting AI development in Japan, and expressed a sense of crisis that there is less data specific to Japan than in other countries and regions.

Mr. Tamaru also pointed out that the perception of the value of data is changing, and there is a shift from disposable to utilized. He cited the establishment of the AI Data Consortium and explained that efforts are underway to make data trading more open and fairer. On the other hand, he said that one of the reasons why companies do not want to provide data is that they see their data as monetary value, and that it is necessary to build an ecosystem that returns the value of data.

Mr. Watanabe emphasized the importance of creating a data ecosystem and utilizing AI. He said that in order to increase the delivery of valuable data, technical options are needed. He also pointed out that data providers have trouble setting prices and expressed the need for value-based pricing.

Mr. Tamaru and Mr. Watanabe explained about the data exchange platform "Data Cloud" developed by the AI Data Consortium. This allows the data provider to set the terms of the transaction and transact while protecting the limited offer data. However, they shared the issue that the existence of this trading platform is not sufficiently recognized.

Mr. Fujikawa stated that his mission is to bring together companies to solve common problems and promote the distribution and sharing of data. He also aims to bring together companies in the same industry to solve major industry issues and social problems.

Regarding the use of AI data, he explained that although it would be ideal to form a community that crosses industries, it should start small and within the same industry. Regarding AI governance, he pointed out that the framework of the industry may not be necessary, but also said that it is necessary to create guidelines within the industry because people in the same industry are using AI for similar use cases, and that such efforts should be carried out in a manner that is not bound by the industry.

Finally, Mr. Tamaru said that the development team is transforming through the use of generative AI, within Microsoft. He showed the idea that the method of creating system services using generative AI is not an extension of the past, but a completely new line. He also said that young engineers and R&D are bringing in new ideas, and that he feels that it is fresh and good when it becomes a product. On the other hand, he said that he was surprised by the high usage of ChatGPT in the senior community, and when he saw how it was used, he felt that he had to come up with new ideas when creating things.

The discussion highlighted the complexity of data governance and its impact on society and business, as well as the need for international collaboration and standardization. The discussion addressed the new challenges associated with technological advances and explored approaches to solving these problems within an international framework. Recognizing the importance of ensuring the free flow and reliability of data (DFFT), we are committed to building new mechanisms and partnerships for international data governance. It also discusses in depth the interpretation of copyright law and the handling of personal information in relation to AI, and it explores how these issues should be incorporated into international data governance frameworks.



For inquiries, please use the following information.
■Internet Homepage: https://aidata.or.jp/
■Secretariat: info@aidata.or.jp

AI Data Consortium 〒150-0034 Daikanyama Front 5F, 16-2 Daikanyama-cho, Shibuya-ku, Tokyo