Chapter 4

Issues and Current Responses to Digital Technologies

Section 1 Issues and current initiatives along with the advancement of AI

The development of AI has brought convenience to our lives, but it also comes with risks and challenges that need to be considered. In the past, using inappropriate or biased data for training AI models has led to increased bias and errors, resulting in decreased reliability of predictions. Many traditional machine learning models have also been criticized for being black boxes (lack of transparency), making it difficult to understand their internal workings and potentially causing issues in critical decision-making scenarios. Additionally, as generative AI rapidly develops and becomes more widespread, specific challenges and risks have become apparent. Below is an overview of the risks and challenges associated with generative AI from both a technical and social/economic perspective.

1. Issues of generative AI

The “AI Business Guidelines (Version 1.0)” formulated by the MIC and the Ministry of Economy, Trade and Industry (hereinafter referred as to METI) in April 2024 provide examples of risks that have become apparent due to the use of generative AI, in addition to the risks associated with conventional AI **(Figure 1-4-1-1)**. For instance, the risks associated with conventional AI include biased or discriminatory outputs, the occurrence of filter bubbles and echo chamber phenomena[1], and the risk of data pollution attacks (such as the degradation of AI performance and misclassification due to the mixing of learning data). Furthermore, the expansion of AI usage leading to increased computational resources resulting in higher energy consumption and environmental impact[2] is also highlighted. As for the risks that have become apparent due to generative AI, the guidelines mention the potential for hallucinations. Generative AI may convincingly produce disinformation not based on facts, which is referred to as “Hallucination.” While technical measures are being considered, it is not entirely suppressible. Therefore, when utilizing generative AI, it is desirable for users to keep in mind the possibility of hallucination and verify the accuracy of the output by cross-referencing or using other means. Additionally, in the use of generated AI, there are concerns about the risk of personal and confidential information being input as prompts and then leaked through the output from the AI. There is also the risk of uncritically accepting false or misleading information, such as fake images and videos created by deepfakes, which could be used for information manipulation and propaganda. Additionally, there is a risk of perpetuating biases and amplifying prejudices present in existing information if AI-generated responses based on such information are uncritically accepted, leading to the continuation or exacerbation of unfair or discriminatory outputs (re-generating bias).

The guidelines emphasize that “the existence of these risks should not immediately hinder the development, provision, or use of AI”. Instead, they “encourage the recognition of risks, the consideration of risk tolerance and the balance with benefits, and the proactive development, provision, and use of AI to enhance competitiveness, create value, and ultimately drive innovation”.

[1] A “Filter Bubble” refers to an information environment in which an algorithm analyzes and learns from an individual internet user’s search history and click history, and the information that the individual user wants to see is displayed first, whether they want it or not, and they are isolated from information that does not match their perspective, and are isolated in a “Bubble” of their own way of thinking and values. An “Echo Chamber” refers to a phenomenon in which people with the same opinions gather together and reinforce each other’s opinions, leading them to believe that their own opinions are correct and to become unable to be exposed to diverse perspectives. For measures against these, refer to 2 in Section 1, Chapter 6.

[2] The guidelines also point out that introducing AI into energy management can also contribute to the environment, such as making electricity use more efficient.

Figure 1-4-1-1 Issues of generative AI

	Risks	Examples
Risks from traditional AI	Output of result that includes bias or discrimination	● AI human resources recruitment system developed by an IT company had a defect in machine learning that discriminated against women.
	Filter bubble and echo chamber phenomena	● The social division is caused by recommendations given by SNS, etc.
	Loss of diversity	● If the whole society uses the same model in the same way, the derivedopinions and replies might converge through LLM, losing diversity.
	Inappropriate use of personal data	● The nontransparent use of personal data and the political use of personal data are problematic.
	Infringement on lives, bodies, and properties	● During AI training, there is a risk of intrusion of invalid data into learningdata, causing performance degradation and misclassification. ● In medical settings, if AI has an ethical bias for determining prioritization,fairness might be lost.
	Data poisoning attack	● During AI training and service operation, there is a risk of intrusion of invalid data into learning data and cyberattacks aimed at the applicationitself.
	Black-box AI, and requirements for explanation about judgment	● Black-box AI's judgments caused a problem as well. ● There is also a rising demand for transparency regarding AI's judgments.
	Energy consumption and environmental load	● As the use of AI spreads, the demands for calculation resources also increase. As a result, data centers are enhanced, and some people are concerned about the increase in energy consumption.
Risks that have become apparent with generative AI	Misuse	● The use of AI for fraud is also problematic.
	Leak of confidential information	● In using AI, there is a risk that personal data or confidential information isentered as a prompt becomes leaked through output.
	Factual errors	● Response represented by generative AI as facts containeddis/misinformation, and a lawsuit was filed against an AI developer and AI provider
	Blindly trusting disinformation and misinformation	● Blindly trusting misinformation produced by generative AI can be a risk. ● Misuse of deepfakes has occurred in various countries.
	Relationship with copyright	● The handling of intellectual property rights is an issue that needs discussed.
	Relationship with qualifications, etc.	● There might be risks of infringement of legally prescribed licenses andqualifications caused by using generative AI.
	Reproduction of bias	● Because generative AI creates answers based on existing information,biases contained in existing information might be amplified, continuingand enhancing unjust output containing discrimination.

(Source) Outline of "AI Guidelines for Business Appendix Ver1.0"

(1) Summary of major LLMs

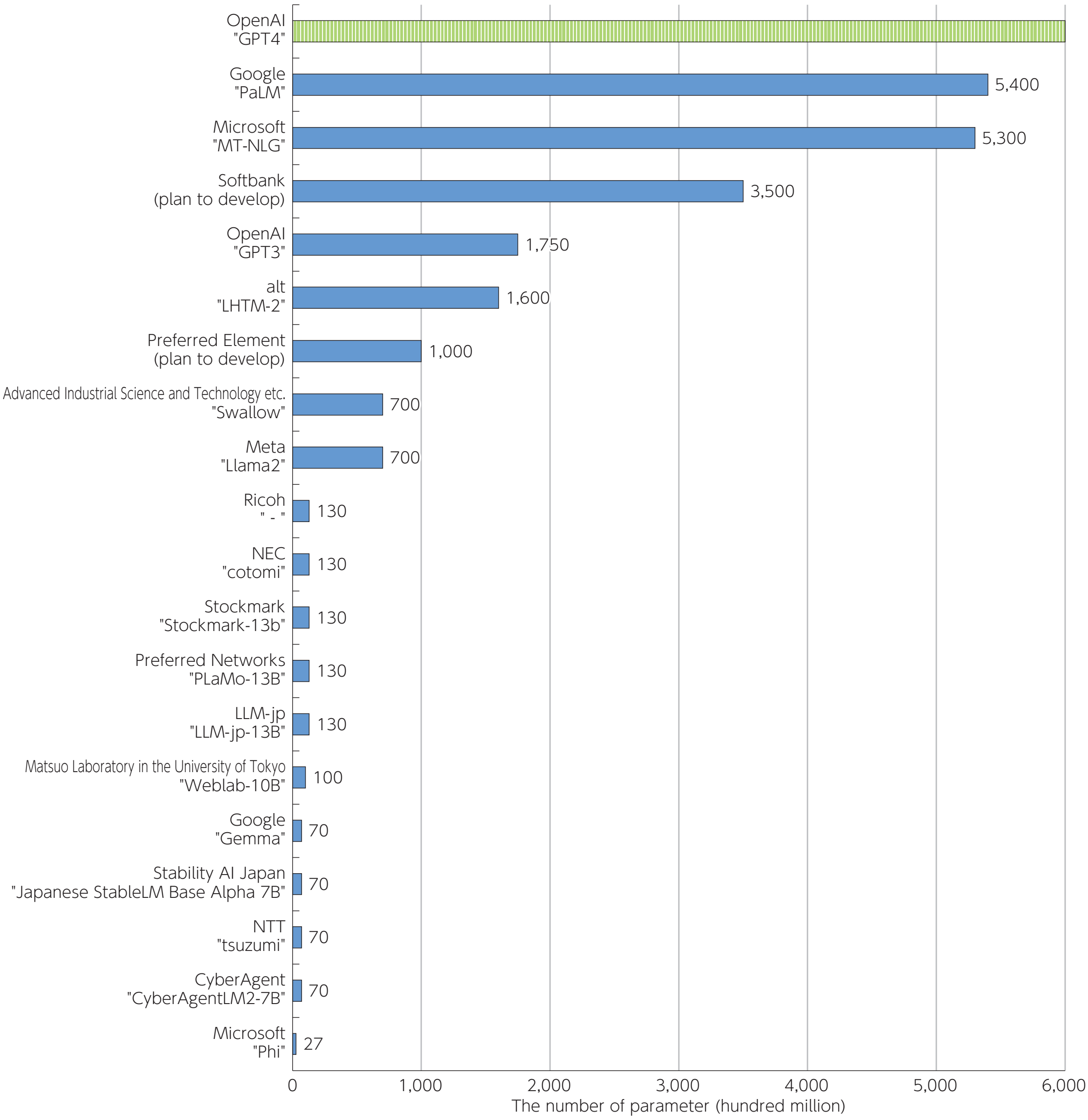
The development of Large Language Models (LLMs), which form the foundation of generative AI, is being led by major tech companies such as Microsoft and Google in the U.S.

However, simply utilizing LLMs developed through closed research and development by non-Japanese entities other than Japan may lead to the black-boxing of the LLMs construction process, raising concerns about rights infringement and information leakage when utilizing LLMs. To ensure the effective utilization of LLMs with a strong focus on the Japanese language, it is essential to have domestically developed LLMs with high transparency, where the construction process and the data used are clearly visible, providing a sense of security[3]. Some Japanese companies are already independently working on LLMs development, and here we will introduce the trends in this area.

In contrast to the LLMs developed by Big Tech companies, there is a tendency in Japan to develop medium-sized LLMs **(Figure 1-4-1-2)**.

[3] National Institute of Advanced Industrial Science and Technology press release "Development of world-class generative AI begins using AIST's computational resource ABCI - AIST, Tokyo Institute of Technology, and LLM-jp (hosted by the National Institute of Informatics) cooperate –" (October 17, 2023), <https://www.aist.go.jp/aist_j/news/pr20231017.html> (accessed on March 22, 2024)

Figure 1-4-1-2 Number of parameters of each model



(Source) Prepared based on companies' websites and news articles etc.[4]

(2) Domestically developed LLMs

A Domestically developed LLMs by the NICT[5]

In July 2023, the NICT announced the development of a large-scale language model with 40 billion parameters using 350GB of high-quality Japanese web text with minimal noise. The LLM developed by the NICT has not undergone fine-tuning or reinforcement learning, and while its performance level is not comparable to ChatGPT, it has reached a level where it can facilitate Japanese language interactions. The NICT plans to further expand the scale of learning texts, focusing on Japanese, and is also working on pre-training a model with 179 billion parameters similar to GPT-3. Additionally, the NICT is aiming to improve both positive and negative aspects in the construction of larger pre-training data and language models, as well as enhancing existing applications and systems such as WISDOM X and MICSUS. (As of May 2024, the NICT is continuing its development efforts, including the development of multiple LLMs with up to 311 billion parameters, and researching the impact of parameter and learning data differences on performance).

[4] The number of parameters for OpenAI's "GPT4" is undisclosed.

[5] NICT, "Prototype of a large-scale language model (generative AI) specialized for Japanese ~ Developing a 40 billion parameter generative large-scale language model trained only on Japanese web data ~" July 4, 2023 <https://www.nict.go.jp/press/2023/07/04-1.html> (accessed on March 22, 2024)

B “CyberAgentLM” LLMs in Japanese developed by CyberAgent[6,7]

In May 2023, CyberAgent announced the development of LLMs in Japanese with a maximum of 6.8 billion parameters. In November 2023, they released a higher-performance model with 7 billion parameters and 32,000 token support, named “CyberAgentLM2-7B,” along with a chat-tuned version called “CyberAgentLM2-7B-Chat.” These models are capable of processing approximately 50,000 characters equivalent to Japanese text. They are provided under the Apache License 2.0 for commercial use.

C “tsuzumi” LLMs in Japanese developed by Nippon Telegraph and Telephone Corporation (NTT)

In November 2023, NTT announced the development of “tsuzumi,” a lightweight Japanese language model with world-class processing capabilities ranging from 6 to 7 billion parameters. “tsuzumi” addresses the challenge of reducing costs for learning and tuning in cloud-based LLMs. It supports both English and Japanese and is capable of modalities such as visual and auditory processing, allowing for specialized tuning for specific industries or corporate organizations. Commercial services for “tsuzumi” began in March 2024, and future plans include enhancing tuning capabilities and gradually implementing multimodal features[8].

2. Issues caused by generative AI

In addition to the constraints faced by generative itself, there are many social and economic challenges associated with the advancement and proliferation of generative AI. Various tech companies, platform operators, industry organizations, and governments both domestically and internationally are working on measures to address these issues.

(1) Challenges and countermeasures for the circulation and spread of dis-/mis-information

The term “Deepfake” is a combination of “Deep Learning” and “Fake,” and it refers to audio, images, or video content that is synthesized using AI technology to falsely represent as genuine or truthful, depicting speech or actions that individuals have not actually made. In recent years, the use of deepfakes for information manipulation and criminal activities has been increasing worldwide, and efforts to address this issue are being made from various quarters. However, the situation presents a cat-and-mouse game, with ongoing challenges in effectively combating deepfakes.

A Challenges posed by deepfakes

(A) Circulation and spread of AI-generated dis-/mis-information

With advancements in generative AI, it has become possible to create highly realistic text, images, audio, and video, making it feasible to produce convincing dis-/mis-information. Using deepfake technology, it is easy to create videos that make it appear as though real people are saying things they never actually said. In Japan, for instance, a fake video of Prime Minister KISHIDA created by using generative AI was spread on social media[9]. Additionally, related to the Noto Peninsula Earthquake on January 1, 2024, numerous posts on social media linked footage from the 2011 Great East Japan Earthquake’s Tsunami and the 2021 Atami Landslide to the Noto Peninsula Earthquake, leading to widespread viewing and dissemination[10]. In 2020, disinformation claiming a connection between COVID-19 and 5G signals led to the destruction of mobile phone base stations[11], demonstrating the societal impact of such disinformation.

The proliferation of various digital services like social media has enabled anyone to become an information disseminator, resulting in a vast amount of information and data circulating on the Internet. In this information-overloaded society, the attention and time we can devote to consuming information are scarce compared to the volume of information available. This scarcity gives rise to what is known as the attention economy, where information that can easily capture the recipient’s attention is prioritized, often driven by economic incentives such as advertising revenue. This structure can lead to the spread of dis-/mis-information and exacerbate online outrage.

The spread of dis-/mis-information is a global issue. In January 2024, the World Economic Forum identified

[6] CyberAgent, “CyberAgent releases Japanese LLM (large-scale language model) with up to 6.8 billion parameters to the public - Providing a commercially available model trained with open data –” May 17, 2023, <https://www.cyberagent.co.jp/news/detail/id=28817> (accessed on March 22, 2024)

[7] CyberAgent, “Version 2 of our unique Japanese LLM (large-scale language model) released to the public - Providing a commercially available chat model with 32,000 tokens -” November 2, 2023, <https://www.cyberagent.co.jp/news/detail/id=29479> (accessed on March 22, 2024)

[8] NTT, “NTT’s commercial service using its unique large-scale language model “tsuzumi” will begin in March 2024” November 1, 2023, <https://group.ntt/jp/newsrelease/2023/11/01/231101a.html> (accessed on March 22, 2024)

[9] The video featured a voice that sounded just like the Prime Minister making obscene remarks, and the logo of a commercial news channel was displayed, giving the impression as if Prime Minister Kishida was being broadcast live as an emergency report. Yomiuri Shimbun Online, “Fake video of Prime Minister KISHIDA spread on social media using generative AI...NTV’s logo misused: “We cannot forgive this,”” November 4, 2023, <https://www.yomiuri.co.jp/national/20231103-OYT1T50260/>

[10] Nikkei Online Edition, “Fake video of Noto Peninsula Earthquake spread on social media, also soliciting remittances,” January 2, 2024, <https://www.nikkei.com/article/DGXZQOCA020JZ0S4A100C2000000/> (accessed on March 22, 2024)

[11] Nikkei Online Edition, “European 5G base station destruction, the shadow culprit is the hoax of “spreading coronavirus”” April 25, 2020, <https://www.nikkei.com/article/DGXMZO58443970U0A420C2XR1000/>

“Disinformation” as one of the most severe risks expected over the next two years, warning that it could exacerbate social and political divisions[12]. Notably, 2024 will see national elections in over 50 countries, including the U.S., Bangladesh, Indonesia, Pakistan, and India. Already, there have been instances of deepfake videos related to the Indonesian presidential election and fake audio impersonating the U.S. President Biden before the U.S. presidential primaries, highlighting the use of generative AI for information manipulation **(Figure 1-4-1-3)**.

Figure 1-4-1-3 Examples of information manipulation by deepfakes made by generative AI

Date	Country and Region	Content
February, 2021	Japan	•When a strong earthquake with a seismic intensity of 6+ struck Miyagi and Fukushima prefectures, a doctored image of then-Chief Cabinet Secretary KATO Katsunobu, making it appear as if he was smiling during a press conference, circulated.
March, 2022	Ukraine	•After the Russian invasion of Ukraine, a fake video was circulated on social media, showing President Zelensky calling for the Ukrainian army to surrender.
September, 2022	Japan	•When Typhoon No. 15 made landfall, fake images claiming that many houses in Shizuoka Prefecture were submerged were spread on Twitter (now X).
March, 2023	The U.S.	•Using image-generating AI, a fake image of former President Trump being arrested was created and circulated on Twitter (now X).
May, 2023	The U.S.	•A fake image depicting an explosion near the Pentagon spread on social media (SNS), causing the Dow Jones Industrial Average to temporarily drop by more than 100 points.
November, 2023	Japan	•A fake video depicting Prime Minister KISHIDA Fumio making sexually suggestive remarks spread on social media (SNS).
November, 2023	Argentina	•During the Argentine presidential election, fake videos allegedly created using AI circulated on social media (SNS).
January, 2024	Taiwan	•During the Taiwan presidential election, a fake video was created and posted, making false claims about President Tsai Ing-wen's personal life.
January, 2024	The U.S.	•A spoof call imitating President Biden's voice urged voters to refrain from voting in the upcoming presidential primary in New Hampshire over the weekend.

(Source) Prepared based on BBC News Japan(2024)[13] etc.

(B) Other use of AI for criminal activities

The use of AI for criminal activities is on the rise, extending beyond information manipulation. The same AI used in the ChatGPT, an automated conversational program developed by the US-based OpenAI, has been exploited to create “BadGPT” or “FraudGPT” - illicit chatbots that mass-produce phishing scam emails. These hacking tools began to surface on dark web sites a few months after OpenAI released ChatGPT in November 2022. It’s estimated that within 12 months of ChatGPT’s release, phishing scam emails increased by 1,265%, resulting in an average of around 31,000 phishing attacks per day[14].

Furthermore, AI’s image generation capabilities have been misused for extortion. Criminals are using AI to transform commonly shared images on social media into inappropriate content, which they then use to blackmail victims. The Federal Bureau of Investigation (FBI) has issued warnings, noting that victims, including minors, have been targeted by such activities[15].

[12] World Economic Forum “How to navigate an era of disruption, disinformation, and division” January 15, 2024, <https://jp.weforum.org/agenda/2024/01/no-wo-ri-rutameni-fo-ramu-sa-dhia-zahidhi/>
NHK NEWS WEB “”Disinformation” becomes the most serious risk. Report before the Davos Conference” January 11, 2024, <https://www3.nhk.or.jp/news/html/20240111/k10014317071000.html> (accessed on 22 March, 2024)

[13] BBC NEWS Japan, “[U.S. presidential election 2024] Automated voice call impersonating Biden disrupts primary election in New Hampshire,” January 23, 2024 <https://www.bbc.com/japanese/68065455> (accessed on February 28, 2024)

[14] “[Focus] Welcome to the era of generative AI “bad GPT””, “Dow Jones US Corporate News”, March 1, 2024 issue

[15] Federal Bureau of Investigation, “Malicious Actors Manipulating Photos and Videos to Create Explicit Content and Sextortion Schemes”, <https://www.ic3.gov/Media/Y2023/PSA230605> (accessed on February 28, 2024)

B Measures against information manipulation and criminal use of deepfakes

(A) European Union (EU)

The European Union (hereinafter referred as to EU) is at the forefront of legal regulations concerning disinformation. The "Digital Services Act"[16] (hereinafter referred as to DSA), which came into effect in November 2022[17], mandates very large online platforms (VLOPs[18]) to conduct risk assessments (including those related to disinformation) and implement risk mitigation measures. Companies that violate these regulations can face penalties of up to 6% of their global annual revenue. The European Commission (hereinafter referred as to EC), the EU's executive body, initiated a formal investigation in December 2023 into X (formerly Twitter) for potentially not complying with the DSA, particularly in relation to the spread of illegal content and the effectiveness of countermeasures against information manipulation on the platform, in light of the spread of illegal content related to terrorist attacks by Hamas and others against Israel[19]. The EC is focusing on the effectiveness of features like "Community Notes," which allow third parties to add annotations to posts anonymously. In March 2024, the European Parliament passed the final draft of the "AI Act,"[20] a comprehensive legal framework for AI, which includes some regulations on deepfakes. The AI Act was formally approved by the EU Council in May 2024 and is expected to be fully applicable by around 2026.

(B) The UK

In the UK, the "Online Safety Act 2023,"[21] which came into effect in October 2023, includes provisions for a six-month prison sentence for those who knowingly transmit disinformation online with the intent to cause psychological or physical harm to the recipient. If it is proven that the perpetrator intended to cause distress, anxiety, humiliation, or sought sexual gratification, the maximum sentence can be up to two years in prison.

(C) The U.S.

In the U.S., the Biden administration announced in July 2023 that it had secured voluntary commitments from seven leading AI companies, including Google, Meta Platforms, and OpenAI[22], to improve AI safety and transparency[23]. In September 2023, an additional eight companies, including IBM, Adobe, and NVIDIA[24], joined this commitment[25]. These 15 companies are promoting the development of technologies to identify AI-generated content, such as "Digital Watermarks" that can indicate authenticity[26]. Some states in the U.S. have specific regulations concerning the use of deepfakes for purposes like pornography and election activities. For example, nine states, including California, Texas, Illinois, and New York, have criminalized the distribution of non-consensual deepfake pornography. Texas and California also have laws regulating the use of deepfakes in political campaigns. At the federal level, laws have been enacted requiring federal agencies like the Department of Defense and the National Science Foundation to strengthen research on disinformation, including deepfakes[27]. However, under Section 230 of the "Communications Decency Act" of 1996, providers are generally not held responsible for third-party content, although the Biden administration is considering legislative changes to hold platform operators accountable for dis-/mis-information.

(D) Japan

In Japan, the MIC has been holding discussions since November 2023 on ensuring the healthiness of information circulation in the digital space in the "Study Group on Ensuring the Healthiness of Information Circulation in the Digital Space", with plans to publish a summary by the summer of 2024[28].

Technological measures include the development of the Originator Profile (OP) technology, which links in-

[16] The law began to apply to VLOPs, etc. from August 2023, and to all regulated businesses from February 2024.

[17] European Commission, "The Digital Services Act package", <https://digital-strategy.ec.europa.eu/en/policies/digital-services-actpackage> (accessed on February 28, 2024)

[18] Abbreviation for Very large online platform. Among online platform services, there are 45 million users within the EU (10% of the EU population) refers to the above services.

[19] European Commission, "PRESS RELEASE18 December, Commission opens formal proceedings against X under the Digital Services Act", <https://ec.europa.eu/commission/presscorner/detail/en/ip_23_6709>(accessed on February 28, 2024)

[20] European Commission, "AI Act", <https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai>(accessed on February 28, 2024)

[21] Legislation.gov.uk," Online Safety Act 2023", <https://www.legislation.gov.uk/ukpga/2023/50/enacted>(accessed on March 2, 2024)

[22] Amazon, Anthropic, Google, Inflection, Meta Platforms, Microsoft, OpenAI

[23] The White House, "FACT SHEET: Biden-Harris Administration Secures Voluntary Commitments from Leading Artificial Intelligence Companies to Manage the Risks Posed by AI", <https://www.whitehouse.gov/briefing-room/statements-releases/2023/07/21/factsheet-biden-harris-administration-secures-voluntary-commitments-from-leading-artificial-intelligence-companies-to-manage-the-risks-posedby-ai/>(accessed on March 8, 2024)

[24] Adobe, Cohere, IBM, NVIDIA, Palantir, Salesforce, Scale AI, Stability

[25] The White House, "FACT SHEET: Biden-Harris Administration Secures Voluntary Commitments from Eight Additional Artificial Intelligence Companies to Manage the Risks Posed by AI", <https://www.whitehouse.gov/briefing-room/statementsreleases/2023/09/12/fact-sheet-biden-harris-administration-secures-voluntary-commitments-from-eight-additional-artificial-intelligencecompanies-to-manage-the-risks-posed-by-ai/>(accessed on March 8, 2024)

[26] "US companies agree to develop AI video identification system; President Biden announces, 'Measures to be taken'", NHK News, July 22, 2023

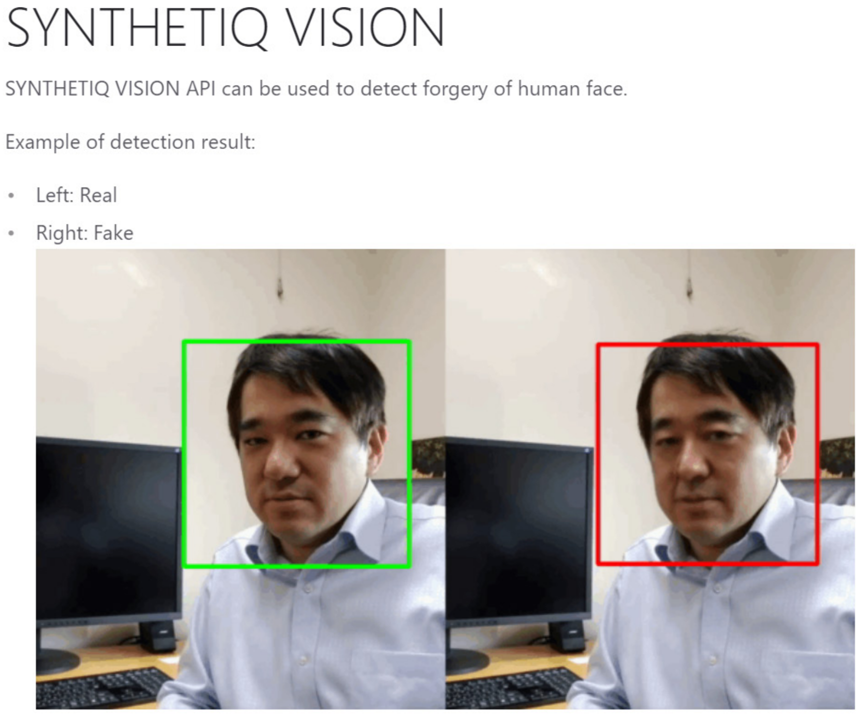
[27] Passed in December 2020. FY2021 National Defense Authorization Act and the Identifying Outputs of Generative Adversarial Networks Act (IOGAN Act) are related to the defense budget for FY2021.

[28] MIC "Study Group on Ensuring the Healthiness of Information Circulation in the Digital Space", <https://www.soumu.go.jp/main_sosiki/kenkyu/digital_space/index.html>

formation content such as news articles and advertisements to the originator's information. This technology is expected to have several effects: it will make impersonation and alterations visible, allowing web users to view highly transparent content; it will make it more difficult to generate advertising revenue from fake news or easy attention-grabbing content; it will reduce the infringement of rights and interests of legitimate web media and content distributors; and by clarifying the identity of web content publishers where ad spaces are placed, advertisers will be able to place ads with confidence.[29].

The National Institute of Informatics (hereinafter referred as to NII) has been engaged in research on countermeasures against fake technologies from an early stage. In September 2021, they developed a tool called "SYNTHETIQ VISION: Synthetic video detector" that automatically determines whether a face image generated by AI is fake **(Figure 1-4-1-4)**. This tool allows users to upload an image they want to verify to a server, and the tool determines whether it is fake or not. The NII is also developing more advanced deepfake countermeasure technologies, such as "Cyber Vaccine," which is expected to provide not only authenticity judgments but also information on where alterations have been made[30,31].

Figure 1-4-1-4 SYNTHETIQ VISION



SYNTHETIQ VISION

SYNTHETIQ VISION API can be used to detect forgery of human face.

Example of detection result:

- Left: Real
- Right: Fake

(Source) Global Research Center for Synthetic Media, National Institute of Informatics[32]

(2) Discussion on intellectual property rights including copyright

The outputs of generative AI primarily include text, images, and music/audio. These are developed using "Machine Learning" techniques that learn features from large amounts of data and generate appropriate results based on prompts (inputs). During this process, there are issues related to the development and learning stages, such as whether collecting and duplicating data to create training datasets and using these datasets for training AI (trained models) infringe on the rights of the original data creators. Additionally, when generating images or other content using generative AI, or when uploading and publishing generated images or selling reproductions (such as illustration collections), there is a risk of infringing on the rights of existing content creators if the generated content is similar to existing works (issues related to the generation and usage stages).

[29] https://originator-profile.org/ja-JP/

[30] "Breakthrough Special Feature 1 - Unmanned Defense 2 - [Part 4: Deepfake Countermeasures] - Tools to Detect Deepfakes, Vaccines to Automatically Repair Tampering," Nikkei Electronics, January 20, 2024 issue

[31] However, these measures also have the issue of the accuracy of the authenticity determination tool. According to OpenAI, the probability that the company's independently developed determination tool correctly determines that documents created by generative AI (mainly ChatGPT) are created by AI is 26%, and conversely, there is a 9% probability of a "False Positive" where documents written by humans are mistakenly determined to be created by generative AI. Therefore, this level of accuracy is not actually an effective judgment tool, and the company has stopped offering the tool. In the future, it is highly likely that the AI for generating text, images, voice, etc. and the judgment tools for these will compete with each other and both technologies will improve, so even if such technology is used, it is considered difficult to accurately distinguish fake information.

[32] https://www.synthetiq.org/

A Issues related to intellectual property rights including copyright with the advancement and spread of generative AI

The issues of copyright and portrait rights infringement related to generative AI are gaining international attention, leading to numerous lawsuits. In the U.S., in November 2022, a class-action lawsuit was filed against Microsoft, GitHub, and OpenAI, alleging that the open-source code used for training GitHub Copilot might infringe on programmers' copyrights[33]. Additionally, in July 2023, three American authors filed a lawsuit against OpenAI and Meta Platforms, claiming damages for the unauthorized use of their works in ChatGPT's machine learning. As a result of this lawsuit, OpenAI announced that instead of removing copyrighted works from its training data, it would cover the legal costs if sued for copyright infringement[34].

Media organizations such as newspapers and news agencies are cautious about using AI. In July 2023, the Associated Press (AP) announced a partnership with OpenAI to explore ways to use generative AI in news reporting. However, by August, they decided not to use AI for creating distributable content. On the other hand, the New York Times filed a lawsuit against OpenAI and Microsoft for the unauthorized use of articles by AI, marking the first lawsuit by a news organization[35]. In Japan, newspapers and news agencies have also expressed concerns about the unauthorized use of articles by generative AI and have called for fundamental legal reforms.

In Japan, in response to concerns raised by rights holders and AI developers about infringement of intellectual property rights including copyright due to the rapid development and spread of generative AI technology, the Legal System Session, Copyright Subcommittee, Cultural Affairs Council compiled a report on "AI and Copyright" in March 2024[36]. Additionally, in May 2024, the "Interim Report on Intellectual Property Rights Review Committee for the AI Era" was published by the Intellectual Property Rights Review Committee for the AI Era[37].

B Measures against the risk of infringement of intellectual property rights including copyright

To address the issue of copyright infringement when using generative AI, it is conceivable for both data/content rights holders and AI businesses to address the issue through mutual contracts. Technically, there are measures such as the practical implementation of electronic watermarks to indicate that the content is generated by AI, and OpenAI providing specifications to suppress the input and output of data/content that may infringe on intellectual property rights. Meanwhile, media organizations such as the New York Times, CNN, Bloomberg, Reuters, and the Nikkei have taken self-protective measures by blocking GPT bots from OpenAI and other AI businesses[38].

There are also initiatives to commit to legal risks of copyright infringement while utilizing technology. In September 2023, Microsoft announced the "Copilot Copyright Commitment," taking responsibility for legal risks associated with its productivity tool "Microsoft Copilot," which incorporates large language models (LLMs). If a copyright claim is made against the output generated by Microsoft Copilot, Microsoft will take responsibility[39]. Another way to avoid the risk of copyright infringement is to use non-copyrighted or licensed works. For example, Adobe's "Adobe Firefly" uses images with open licenses or other non-copyrighted images during the training stage, allowing commercial use of the generated images without concerns about copyright infringement.

[33] The three companies claim that GitHub Copilot uses knowledge gained from open source code and does not infringe copyright, and have asked the court to dismiss the lawsuit. Reuters, "OpenAI, Microsoft want court to toss lawsuit accusing them of abusing open-source code," <https://www.reuters.com/legal/litigation/openai-microsoft-want-court-toss-lawsuit-accusing-them-abusing-open-source-code-2023-01-27/> (accessed on February 27, 2024)

[34] Generative AI Utilization Promotion Association, "What will happen to AI copyright? A thorough explanation of copyright and legality of images and illustrations generated by generative AI, and points to be aware of" December 28, 2023, <https://guga.or.jp/columns/ai-copyright/> (accessed on March 2, 2024)

[35] Reuters, "OpenAI, Microsoft want court to toss lawsuit accusing them of abusing open-source code," <https://www.reuters.com/legal/litigation/openai-microsoft-want-court-toss-lawsuit-accusing-them-abusing-open-source-code-2023-01-27/> (accessed on February 27, 2024)

[36] "About the Concept of AI and Copyright," the Legal System Session, Copyright Subcommittee, Cultural Affairs Council (March 15, 2024), <https://www.bunka.go.jp/seisaku/bunkashingikai/chosakuken/pdf/94037901_01.pdf

[37] Intellectual Property Rights Review Committee for the AI Era "Interim Report of the Intellectual Property Rights Review Committee for the AI Era" (May 2024), <https://www.kantei.go.jp/jp/singi/titeki2/chitekizaisan2024/0528_ai.pdf>

[38] Intellectual Property Rights Review Committee for the AI Era "Interim report of the Intellectual Property Rights Review Committee for the AI Era" (May 2024), <https://www.kantei.go.jp/jp/singi/titeki2/chitekizaisan2024/0528_ai.pdf>

[39] Do AI characters have copyright? What happens if you violate the law? We asked a lawyer, <https://webtan.impress.co.jp/e/2023/12/19/46093> (accessed on March 2, 2024)

Section 2 Responses to AI by country

In the midst of the rapid proliferation of AI, including generative AI, addressing the ethical and societal issues that have arisen requires collaborative efforts not only domestically but also internationally.

1. Trends in international discussion

(1) Hiroshima AI Process

Discussions on the ethical and societal issues of AI have been intensifying since around 2015, and our country has been at the forefront of discussions in G7/G20 and the Organisation for Economic Co-operation and Development (hereinafter referred as to OECD), playing a significant role in formulating AI principles. In April 2016, at the G7 ICT Ministers' Meeting held in Takamatsu, Japan proposed discussions on the development principles of AI, leading to the agreement on AI principles at the OECD in May 2019, followed by the agreement on "G20 AI Principles" at the G20 Summit in June of the same year[1]. From 2019 to 2020, there has been an international consensus forming around AI principles, and discussions have been transitioning to the formulation of specific institutional and regulatory frameworks to implement these principles in society. Furthermore, the rapid proliferation of generative AI in 2022 has led to an intensification of discussions on AI governance in international cooperation forums such as the G7 and within individual countries.

In April 2023, G7 Digital and Tech Ministers' Meeting in Takasaki, Gunma was held in Takasaki City, Gunma, where discussions were held on "Responsible AI and Global AI Governance" in light of the rapid proliferation and advancement of generative AI. At this meeting, the importance of interoperability between different AI governance frameworks among G7 members was confirmed, and a ministerial declaration consisting of six themes, including "Responsible AI and Global AI Governance," "Secure and Resilient Digital Infrastructure," and "Internet Governance," was compiled. This declaration was subsequently reflected in the discussions at the G7 Hiroshima Summit held in May, and the leaders' communiqué at the summit instructed the establishment of the Hiroshima AI Process for discussions on generative AI. Specifically, it was decided to collaborate with relevant organizations such as the OECD and the Global Partnership on AI (GPAI) and to advance investigations and deliberations in G7 working groups.

In September 2023, a ministerial-level meeting was held to discuss the development of advanced AI systems, including generative AI, based on reports drafted by the OECD in July and August. It was confirmed that transparency, disinformation, intellectual property rights, privacy, and personal information protection are priority issues. Subsequently, on October 30, the "G7 Leaders' Statement on the Hiroshima AI Process"[2] was issued, and International Guiding Principles and Code of Conduct for Organizations Developing Advanced AI Systems were first published. Furthermore, in December of the same year, a Comprehensive Policy Framework for the Hiroshima AI Process, including Project-Based Cooperation on AI, and Work Plan to advance Hiroshima AI Process were announced.

(2) Movements of the OECD/GPAI/UNESCO

A OECD

Many international organizations, including the OECD, the GPAI, and the UNESCO, are advancing the consideration of AI governance systems from a global perspective. Since the publication of the OECD AI Principles in May 2019, various OECD reports have been released, and projects have been promoted in collaboration with the G7, actively engaging in these activities. Additionally, in September 2023, the three organizations—the OECD, the GPAI, and the UNESCO—announced the "Global Challenge to Build Trust in the Age of Generative AI,"[3] a global collaborative project aimed at advancing innovative solutions to social risks posed by disinformation and deepfakes, based on the comprehensive framework of the G7.

At the OECD Ministerial Council Meeting held in May 2024, a side event on generative AI titled "Towards Safe, Secure, and Trustworthy AI: Promoting Inclusive Global AI Governance" was held, where Prime Minister KISHIDA announced the establishment of the "Hiroshima AI Process Friends Group,"[4] a voluntary framework of countries and regions that support the spirit of the Hiroshima AI Process, with participation from 49 countries and regions.

B GPAI

The "Global Partnership on AI" (hereinafter referred as to GPAI) was established in 2020 through a joint state-

[1] METI Study Group on Implementation of AI Principles, "AI Governance in Japan ver1.1", <https://www.meti.go.jp/shingikai/mono_info_service/ai_shakai_jisso/pdf/20210709_1.pdf> (accessed on March 4, 2024)

[2] Ministry of Foreign Affairs, "G7 Leaders' Statement on the Hiroshima AI Process" <https://www.mofa.go.jp/mofaj/ecm/ec/page5_000483.html> (accessed on March 4, 2024)

[3] Global Challenge partners, "Global Challenge to Build Trust in the Age of Generative AI", <https://globalchallenge.ai/>(accessed on March 21, 2024)

[4] https://www.kantei.go.jp/jp/101_kishida/statement/2024/0502speech2.html

ment by the OECD and the G7, based on a human-centered approach to realize the development and use of "Responsible AI." The organization, with the OECD serving as its secretariat, is an international public-private partnership consisting of governments, international organizations, industries, and experts who share common values, with 29 countries currently participating. The GPAI has four working groups: "Responsible AI," "Data Governance," "Future of Work," and "Innovation and Commercialization," where experts engage in discussions and practical research.

At the "GPAI Summit 2023," the establishment of the GPAI Tokyo Expert Support Center, which is a new support center for the GPAI experts, was approved. This center is set to prioritize projects related to the investigation and analysis of generative AI.

C UNESCO

The United Nations Educational, Scientific and Cultural Organization (UNESCO) adopted the "UNESCO Recommendation on the Ethics of Artificial Intelligence"[5] in 2021, supporting initiatives in various countries. In September 2023, the UNESCO published the "Guidance for Generative AI in Education and Research,"[6] the first global guidance on generative AI in the fields of education and research. This document provides definitions and explanations of generative AI, ethical and policy issues, implications for the education sector, necessary steps for regulatory considerations, curriculum design, and learning. Given that most generative AI is primarily designed for adults, it suggests restricting its use in educational settings to those aged 13 and above. It also calls on governments to implement appropriate regulations, including data privacy protection, and to provide teacher training.

(3) AI Safety Summit

In May 2023, OpenAI announced the possibility of AI systems surpassing human expert skill levels within the next decade, naming this "Frontier AI." Considering existential risks such as nuclear energy and synthetic biology, the company emphasized the need for international regulations rather than reactive measures. In response, the UK Prime Minister Sunak hosted the "AI Safety Summit"[7] in Bletchley, the UK, on November 1 and 2, 2023. This summit was notable for its focus on "AI Safety," aiming to prevent "Severe and Catastrophic Harm" caused by AI, beyond the traditional "AI Ethics" concerns of human rights and fairness.

The summit concluded with the adoption of the "Bletchley Declaration."[8] The UK also decided to establish the AI Safety Institute.

From May 21 and 22, 2024, the "AI Seoul Summit" was co-hosted by the Republic of Korea and the UK (with the leaders' session held online on the 21st and the ministerial session held in person in Seoul on the 22nd). The summit deepened discussions on AI safety, promoted innovation in AI development, and addressed the equitable enjoyment of AI benefits. The summit resulted in the adoption of the "Seoul Declaration for Safe, Innovative, and Inclusive AI" and its appendix, the "Seoul Statement of Intent toward International Cooperation on AI Safety Science," as leaders' outcome documents. The ministerial outcome document, the "Seoul Ministerial Statement for Advancing AI Safety, Innovation and Inclusivity," was also adopted. The next meeting is scheduled to be held in France in February 2025.

(4) Developments in the United Nations

In light of the growing interest in international governance frameworks for Frontier AI, the UK led discussions on AI at the United Nations Security Council in July 2023. In October of the same year, the UN Secretary-General António Guterres established a High-Level Advisory Body on AI, which includes Japanese members. On March 21, 2024, the UN General Assembly adopted by consensus the "Resolution Seizing the opportunities of safe, secure, and trustworthy artificial intelligence systems for sustainable development,"[9] co-sponsored by Japan. This resolution is the first UN General Assembly resolution on safe, secure, and trustworthy AI. It promotes safe, secure, and trustworthy AI to accelerate progress towards the "2030 Agenda for Sustainable Development" and to bridge the digital divide. The resolution encourages member states to develop and support regulatory and governance approaches related to safe, secure, and trustworthy AI. It also recommends that member states and stakeholders promote innovation for identifying, assessing, and mitigating risks during AI design and development, and establish, implement, and disclose risk management mechanisms for data preservation to ensure AI systems can address global challenges. Furthermore, it emphasizes that human rights and fundamental freedoms should be respected, protected, and promoted throughout the AI

[5] UNESCO, "Recommendation on the Ethics of Artificial Intelligence", <https://unesdoc.unesco.org/ark:/48223/pf0000381137>(accessed on March 13, 2024)

[6] UNESCO, "Guidance for generative AI in education and research", <https://www.unesco.org/en/articles/guidance-generative-aieducation-and-research>(accessed on March 13, 2024)

[7] GOV.UK, "About the AI Safety Summit 2023", <https://www.gov.uk/government/topical-events/ai-safety-summit-2023/about> (accessed on March 12, 2024)

[8] GOV.UK, "The Bletchley Declaration by Countries Attending the AI Safety Summit, 1-2 November 2023", <https://www.gov.uk/government/publications/ai-safety-summit-2023-the-bletchley-declaration/the-bletchley-declaration-by-countries-attending-the-ai-safetysummit-1-2-november-2023> (accessed on March 12, 2024)

[9] United Nations General Assembly, A/78/L.4 <https://documents.un.org/doc/undoc/ltd/n24/065/92/pdf/n2406592.pdf?token=0e5FKl9eh5r1MmYPD3&fe=true> (accessed on March 22, 2024)

system lifecycle.

This resolution reflects discussions from the Hiroshima AI Process, G7, G20, the OECD, and other forums, and although it does not have binding force under international law, its adoption by consensus signifies its political weight as the collective will of the international community.

2. Trends in creation of legal rules and guidelines by country

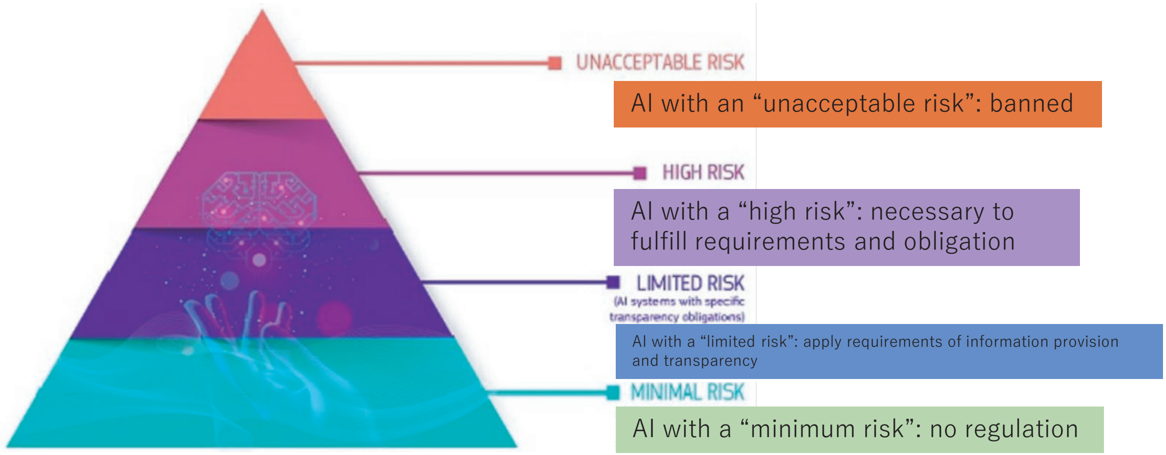
Currently, discussions on legal frameworks and international standards related to AI are actively taking place in various countries around the world. The year 2023 has become a significant milestone for AI policy, marked by the adoption of the EU AI Act by the European Parliament, the issuance of an executive order on AI safety in the US, and the publication of draft guidelines for AI-related businesses in Japan. Observing the regulatory movements concerning AI in each country and region, the rapid rise in interest in generative AI has necessitated a review of the governance systems that have been under consideration. In the establishment of regulations for rapidly evolving technologies, it is essential for governments to take the lead while also requiring voluntary efforts from AI businesses. This dual approach of public and private sector collaboration is currently being advanced.

(1) European Union (EU)

The EU, which lacks major Big Tech companies originating within its borders, has aimed to implement the strictest regulations ahead of other regions and has been discussing AI regulations since 2020. On May 21, 2024, the AI Act[10], which is positioned as the world's first comprehensive AI regulation with legal binding force targeting businesses that develop, provide, and use AI systems in the European market, was established. This AI Act marks the first comprehensive AI regulation law to be established in major countries and regions, and it is expected to be gradually applied, with full implementation anticipated around 2026.

The AI Act is based on a "Risk-based Approach," which changes the regulatory content according to the level of risk[11]. It classifies regulatory targets into four risk levels: (1) unacceptable risk; (2) high risk; (3) limited risk; and (4) minimal risk AI applications and systems, and imposes different regulations for each level. Businesses that violate these regulations may face fines of up to 35 million euros (approximately 5.6 billion yen) or 7% of their annual turnover for the most severe violations[12] **(Figure 1-4-2-1)**.

Figure 1-4-2-1 Risk-based approach in the AI Act


UNACCEPTABLE RISK
AI with an "unacceptable risk": banned
HIGH RISK
AI with a "high risk": necessary to fulfill requirements and obligation
LIMITED RISK
(AI systems with specific transparency obligations)
AI with a "limited risk": apply requirements of information provision and transparency
MINIMAL RISK
AI with a "minimum risk": no regulation

(Source) Prepared based on the European Commission (2024)[13]

(2) The U.S.

The U.S., home to many Big Tech companies, has focused on protecting its own companies, prioritizing voluntary measures by the private sector over government regulations. The government steps in with regulations only when necessary[14].

In July 2023, seven leading AI development companies (including Google, Meta Platforms, and OpenAI)[15] committed to voluntary measures for safe AI develop-

[10] European Commission, "AI Act", <https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai>(accessed on March 2, 2024)
[11] European Parliament, "Artificial intelligence act", <https://www.europarl.europa.eu/RegData/etudes/BRIE/2021/698792/EPRS_BRI%282021%29698792_EN.pdf>(accessed on March 12, 2024)
[12] "EU regulates AI development and operation by law...Copyright protection of learning data, fines of 5.6 billion yen for violators", "Yomiuri News" March 13, 2024 issue
[13] European Commission, "AI Act", <https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai>(accessed on March 15, 2024)
[14] "AI legislation: Industry, government, and academia debate with the world" Ask an expert, "Nihon Keizai Shimbun Electronic Edition" January 1, 2024 issue
[15] Amazon, Anthropic, Google, Inflection, Meta Platforms, Microsoft, OpenAI

ment. In September, an additional eight companies (including IBM, Adobe, and NVIDIA)[16] agreed to these measures, as announced by the U.S. government[17]. These companies have established principles from the perspectives of safety, security, and reliability as part of their voluntary commitments[18].

While the White House indicated that these companies would continue their efforts until mandatory regulations were introduced, President Biden announced the "Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence"[19] on October 30, 2023. This executive order expands the scope of AI issues from ethical considerations to national security concerns. It includes not only Big Tech companies but also biotechnology firms and other businesses that could impact national security and the economy. The order mandates new safety assessments for AI, guidance on fairness and civil rights, and studies on AI's impact on the labor market[20]. It aims to establish new standards for AI safety and security, protect American privacy, and promote fairness and civil rights[21].

Following the publication of the executive order, Vice President Harris announced the "New U.S. Initiatives to Advance the Safe and Responsible Use of Artificial Intelligence"[22] at the UK AI Safety Summit in November 2023. This initiative includes the establishment of the U.S. AI Safety Institute (hereafter referred as to US AISI) within the National Institute of Standards and Technology (NIST). The US AISI, established within the National Institute of Standards and Technology (NIST), develops guidelines, tools, benchmarks, and best practices to evaluate and mitigate harmful functionalities, conducts evaluations to identify and mitigate AI risks, including red team assessments. It also plans to develop technical guidance related to the authentication of human-generated content, electronic watermarking for AI-generated content, identification and mitigation of discrimination by harmful algorithms, ensuring transparency, and introducing privacy protection. This includes the collaboration with international counterparts such as the UK's AI Safety Institute for information sharing and research cooperation, as well as potential partnerships with external experts from civil society, academia, and industry.

Meanwhile, the U.S. Congress is also discussing federal-level AI regulation bills. In June 2023, the Senate proposed the "SAFE Innovation Framework," a comprehensive framework to address the rapid advancement of AI, and held nine thematic forums with industry representatives and experts by December 2023[23]. The House of Representatives announced the establishment of a bipartisan task force on AI in February 2024, which will prepare a comprehensive report with principles and policy recommendations for AI policy[24]. Although several bills regulating AI use in specific areas, such as elections, have been introduced in both chambers, none have yet passed. With the U.S. presidential election approaching in the fall of 2024, discussions on AI regulation are expected to intensify, particularly concerning issues like deepfake-driven information manipulation.

(3) The UK

The UK is considered one of the leading countries in AI research, following the U.S. and China. Although it fell to fourth place for the first time in 2023 due to the rise of Singapore in terms of private investment in the AI sector, it has maintained its position as the third in the world since 2019, following the U.S. and China[25]. The current Sunak administration is reluctant to implement legally binding AI regulations. Instead, it aims to promote the development of AI systems with safety considerations, thereby leading to economic growth. Consequently, it has expressed its intention not to establish strict new regulations like the EU's AI Act for the time

[16] Adobe, Cohere, IBM, NVIDIA, Palantir, Salesforce, Scale AI, Stability

[17] The White House, "FACT SHEET: Biden-Harris Administration Secures Voluntary Commitments from Eight Additional Artificial Intelligence Companies to Manage the Risks Posed by AI", <https://www.whitehouse.gov/briefing-room/statementsreleases/2023/09/12/fact-sheet-biden-harris-administration-secures-voluntary-commitments-from-eight-additional-artificial-intelligencecompanies-to-manage-the-risks-posed-by-ai/>(accessed on March 8, 2024)

[18] (1) Ensuring Safety Before System Release: The companies commit to internal and external security testing of their AI systems before their release. The companies commit to sharing information across the industry and with governments, civil society, and academia on managing AI risks. (2) Building Systems that Put Security First: The companies commit to investing in cybersecurity and insider threat safeguards to protect proprietary and unreleased model weights. The companies commit to facilitating third-party discovery and reporting of vulnerabilities in their AI systems. (3) Earning the Public's Trust: The companies commit to developing robust technical mechanisms to ensure that users know when content is AI generated, such as a watermarking system. The companies commit to publicly reporting their AI systems' capabilities, limitations, and areas of appropriate and inappropriate use.

The White House, "FACT SHEET: Biden-Harris Administration Secures Voluntary Commitments from Leading Artificial Intelligence Companies to Manage the Risks Posed by AI", <https://www.whitehouse.gov/briefing-room/statements-releases/2023/07/21/factsheet-biden-harris-administration-secures-voluntary-commitments-from-leading-artificial-intelligence-companies-to-manage-the-risks-posedby-ai/>(accessed on March 8, 2024)

[19] The White House, "Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence", <https://www.whitehouse.gov/briefing-room/presidential-actions/2023/10/30/executive-order-on-the-safe-secure-and-trustworthy-developmentand-use-of-artificial-intelligence/>(accessed on March 4, 2024)

[20] "Thinking about AI governance (5) Different responses depending on social and cultural backgrounds", "Nihon Keizai Shimbun" morning edition, February 8, 2024

[21] The White House, "FACT SHEET: President Biden Issues Executive Order on Safe, Secure, and Trustworthy Artificial Intelligence", <https://www.whitehouse.gov/briefing-room/statements-releases/2023/10/30/fact-sheet-president-biden-issues-executive-order-on-safesecure-and-trustworthy-artificial-intelligence/>(accessed on March 10, 2024)

[22] The White House, "FACT SHEET: Vice President Harris Announces New U.S. Initiatives to Advance the Safe and Responsible Use of Artificial Intelligence", <https://www.whitehouse.gov/briefing-room/statements-releases/2023/11/01/fact-sheet-vice-president-harrisannounces-new-u-s-initiatives-to-advance-the-safe-and-responsible-use-of-artificial-intelligence/> (accessed on March 10, 2024)

[23] "U.S. Senate Leader Schumer Announces Action Framework for Formulation of AI Bill", "JETRO Business Bulletin" June 22, 2023 issue

[24] "U.S. House of Representatives establishes bipartisan task force on AI," "JETRO Business Bulletin" February 28, 2024 issue

[25] Tortoise media," The Global AI Index", <https://www.tortoisemedia.com/intelligence/global-ai/#rankings> (accessed on March 21, 2024)

being, opting to handle matters flexibly within the existing framework. In line with this policy, the UK government published a policy document in March 2023 titled "A pro-innovation approach to AI regulation,"[26] which outlines the basic framework for AI regulation in the country. This document sets forth five principles from the perspectives of security, transparency, fairness, accountability, and contestability[27]. When addressing AI governance, the approach is described as "pro-innovation, flexible, non-statutory, proportionate, trustworthy, adaptable, clear, and collaborative." For the time being, the government plans to encourage the implementation of these principles within the industry under existing regulations through the collaboration of various government agencies. In the future, there may be an effort to make these principles mandatory.

Additionally, on November 27, 2023, the UK's National Cyber Security Centre (NCSC) and the U.S.' Cybersecurity and Infrastructure Security Agency (CISA) led a joint effort with 18 countries, including Japan, to publish the "Guidelines for secure AI system development."[28] These guidelines compile the necessary actions to be taken at each stage of AI design, development, deployment, operation, and maintenance.

(4) Japan

While Japan shares the same stance as Western countries regarding democracy and fundamental human rights, cultural and social norms differ, leading to a unique societal perception of AI. Consequently, in terms of AI governance, Japan currently favors a soft law approach that emphasizes voluntary efforts by private businesses, rather than a cross-cutting legal regulatory approach. This contrasts with Europe, which aims for legally binding hard laws. The MIC and the METI have been at the forefront of these efforts. The "AI Development Guidelines"[29] by the MIC's AI Network Society Promotion Council were published in 2017, followed by the "AI Utilization Guidelines"[30] in 2019. Additionally, in March of the same year, guidelines based on the "Human-Centric AI Social Principles"[31] decided by the Cabinet Office's Integrated Innovation Strategy Promotion Council were formulated. In July 2021, the METI published the "Governance Guidelines for the Implementation of AI Principles" (revised in January 2022)[32], which outlines action goals for AI businesses along with practical examples. These guidelines are organized by items such as environmental and risk analysis, system design, and operation, to serve as a reference for businesses developing and operating AI.

In May 2023, the government established the "AI Strategic Council" to discuss various themes such as addressing AI risks, optimal AI utilization, and measures to strengthen AI development capabilities. The council published the "Tentative Summary of AI Issues"[33] and began work on integrating guidelines from various ministries. In September of the same year, the council presented a "New AI Business Operator Guidelines Skeleton (Draft)" that included governance for generative AI. In December, the government published the "AI Business Operator Guidelines Draft," which outlines ten principles, including considerations for human rights and countermeasures against disinformation, and prohibits the development of AI that unjustly manipulates human decision-making, cognition, or emotions. However, unlike in the West, these guidelines do not have certain legal binding force. After a public comment period, the "AI Guidelines for Business Ver 1.0" were published on April 19, 2024.

Additionally, at the AI Strategic Council meeting in December 2023, Prime Minister Kishida announced the establishment of the "AI Safety Institute" (hereinafter referred as to AISI)[34] in Japan, similar to institutions in the US and UK, in response to the growing international concern over AI safety. On February 14, 2024, the AISI was established under the Information-technology Promotion Agency (IPA), which is under the jurisdiction of the METI. The AISI will collaborate with similar institutions in the UK, the US, and other countries to develop standards and guidance to improve the safety of AI development, provision, and utilization, conduct research on AI safety evaluation methods, and investigate technologies and case studies related to AI safety.

[26] GOV.UK, "AI regulation: a pro-innovation approach", <https://www.gov.uk/government/publications/ai-regulation-a-pro-innovationapproach> (accessed on March 19, 2024)

[27] (1) Safety, Security, and Robustness: AI systems must be robust, secure, and safe throughout their lifecycle, and risks must always be identified, assessed, and managed. (2) Appropriate Transparency and Explainability: Developers and implementers of AI systems must provide sufficient information to stakeholders about when, how, and for what purpose the AI system is being used, and must offer adequate explanations of the AI system's decision-making processes to stakeholders. (3) Fairness: AI systems must not infringe on the legal rights of individuals or entities throughout their lifecycle, and must not be used to unfairly discriminate against individuals or produce unfair commercial outcomes. (4) Accountability and Governance: An effective governance framework must be established to ensure the monitoring of the supply and use of AI systems, and clear accountability must be maintained throughout the AI system's lifecycle. (5) Disputability and Redress: In cases where AI decisions or outcomes are harmful or involve significant risks, those affected must be provided with opportunities to appeal and seek redress.

[28] National Cyber Security Centre, "Guidelines for secure AI system development", <https://www.ncsc.gov.uk/collection/guidelinessecure-ai-system-development>(accessed on March 12, 2024)

[29] MIC, "Publication of AI Network Society Promotion Council Report 2017", <https://www.soumu.go.jp/menu_news/s-news/01iicp01_02000067.html>

[30] MIC, "Publication of AI Network Society Promotion Council Report 2019", <https://www.soumu.go.jp/menu_news/s-news/01iicp01_02000081.html>

[31] Cabinet Office, Integrated Innovation Strategy Promotion Council Decision, "Social Principles of Human-Centric AI", <https://www8.cao.go.jp/cstp/aigensoku.pdf> (accessed on March 12, 2024)

[32] METI, "Governance Guidelines for the Implementation of AI Principles ver. 1.1", <https://www.meti.go.jp/shingikai/mono_info_service/ai_shakai_jisso/20220128_report.html> (accessed on March 12, 2024)

[33] Cabinet Office AI Strategic Council "Tentative Summary of AI Issues", <https://www8.cao.go.jp/cstp/ai/ronten_honbun.pdf> (accessed on March 12, 2024)

[34] AI Safety Institute, <https://aisi.go.jp/> (accessed on March 12, 2024)

Section 3 Trends in discussion on other digital technologies

1. Trends in discussion on the metaverse, robotics and automated driving

(1) Metaverse

In the report compiled in July 2023 by the MIC's Study Group on the Utilization of Metaverse toward the Web3 Era, issues related to the metaverse are broadly categorized into "issues within the metaverse space" and "issues related to the outside of the metaverse space."

For issues within the metaverse space, the report identifies (1) challenges related to avatars, (2) interoperability between platforms, (3) issues during the construction and utilization of the metaverse, and (4) issues related to data acquisition and use. For issues related to the outside of the metaverse space, the report identifies (5) challenges related to user interfaces (UI) and user experiences (UX), and (6) the trends and social impacts of the metaverse. The study group has examined these issues related to from (1) to (4) and outlined directions for addressing them, including forming an international consensus on the principles of the metaverse, efforts to ensure interoperability (such as standardization), and the development of guidelines (provisional) for metaverse-related service providers. For issues related to (5) and (6), the group has outlined directions for continuous follow-up on market, technology, and user trends, and research on the relationship between the metaverse and UI/UX[1]. From October 2023, the MIC has been holding a "Study Group on Realizing the Safe and Secure Metaverse" to examine issues identified in the previous report that require continuous follow-up. The group aims to realize safer and more secure metaverse for users, based on the democratic values confirmed at the G7 Digital and Technology Ministers' Meeting in Gunma-Takasaki in April 2023 and the G7 Hiroshima Summit in May 2023. The group is considering the "Principles of the Metaverse (First Draft)" , which consists of (1) principles for voluntary and autonomous development of the metaverse (openness and innovation, diversity and inclusiveness, literacy, community) and (2) principles for improving the trustworthiness of the metaverse (transparency and explanation, accountability, privacy, security)[2], with plans to compile a report by the summer of this year.

International organizations are also examining immersive technologies such as the metaverse. For example, the OECD announced the establishment of the Global Forum on Technology (GFTech)[3] in December 2022 and has set up focus groups to discuss immersive technologies. Discussions in the focus group on immersive technologies began in December 2023, with plans to compile a report by the fall of 2024. The MIC is also contributing to international discussions, such as co-hosting a session with the OECD on "Pursuing a metaverse based on democratic values" at the Internet Governance Forum Kyoto, organized by the United Nation in October 2023.

(2) Robotics

Robotics has traditionally been a strong technology for our country, and particularly in the case of industrial robots, it holds a 46% share of the global market. In a country where the labor force continues to decline, there are high expectations for the use of robotics to improve productivity, address the shortage of labor, and create new industries. In 2015, our country formulated the "New Robot Strategy" and has since implemented over 30 technology development projects through public-private partnerships. While the robots themselves and the individual technologies supporting them have evolved, there is a reality that social implementation has not progressed due to the gap between the needs of robot introduction sites. In response to this situation, the New Energy and Industrial Technology Development Organization (NEDO) published the "Comprehensive Action Plan for Research and Development and Social Implementation in the Field of Robotics" in April 2023, which outlines the direction for promoting the use of robots that contribute to solving social issues and the early start of projects to develop robot technology strategies[4]. The action plan addresses eight fields where the use of robots is expected (manufacturing, food production, facility management, retail/food service, logistics warehouses, agriculture, infrastructure maintenance management, and construction) and compiles short-term measures for accelerating social implementation by around 2030 as the "Action Plan for Accelerating Social Implementation" and medium- to long-term measures for creating impact toward the next-generation technology infrastructure by around 2035 as the "Action Plan for Building Next-Generation Technology Infrastructure."

In the future, based on the actions extracted from the robot action plan for technology development and environmental improvement, efforts will be made to advance discussions on future national projects and social implementation.

[1] MIC, "Study Group Report on Utilization of Metaverse toward the Web3 Era", <https://www.soumu.go.jp/main_content/000892205.pdf>
[2] MIC, "Study Group on Realizing a Safe and Secure Metaverse," <https://www.soumu.go.jp/main_sosiki/kenkyu/metaverse2/index.html>
[3] OECD, Global Forum on Technology, <https://www.oecd.org/digital/global-forum-on-technology/>
[4] NEDO releases "Overall action plan for research and development and social implementation in the field of robots" –Promoting the resolution of social issues through both social implementation and next-generation technology development–" <https://www.nedo.go.jp/news/press/AA5_101639.html>

(3) Autonomous driving technology

The utilization of autonomous driving technology is expected to contribute to the maintenance of public transportation and logistics in regions facing population decline and aging. Efforts to expand its societal use are being sought. In the "Comprehensive Strategy for the Vision for a Digital Garden City Nation (Revised in 2023)," the government has set a goal to promote regional transportation through autonomous driving. Various relevant government ministries and agencies are collaborating to achieve the target of implementing unmanned autonomous transportation services in approximately 50 locations by FY2025, and in over 100 locations by FY2027. Additionally, in the "National Comprehensive Development Plan for Digital Lifelines" (METI), the establishment of autonomous driving service support roads is listed as one of the Early Harvest Projects. It aims to set up priority lanes for autonomous driving vehicles of over 100 km on certain sections of the Shin-Tomei Expressway by FY2024, with the goal of realizing the operation of autonomous driving trucks. Furthermore, it aims to enable the provision of autonomous driving vehicle-based mobility services in 50 locations nationwide by FY2025, and in 100 locations nationwide by FY2027. To achieve this plan, collaborative efforts are being undertaken by the National Police Agency, the MIC, the Ministry of Land, Infrastructure, Transport and Tourism, and other relevant ministries and agencies.

2. Trends in discussion to ensure cybersecurity

To ensure that each citizen can utilize digital technology with peace of mind, it is important to ensure cybersecurity. In recent years, due to the increasing complexity of international situations, cyberattacks targeting government agencies and others have been occurring frequently in various countries, including our own. Additionally, with the emergence of technologies such as generative AI, while convenience has increased, there are also concerns about the expansion of risks through their misuse.

Traditionally, cybersecurity has mainly focused on ensuring the availability and confidentiality of systems, in other words, ensuring that systems do not stop and preventing the theft or leakage of data, in order to maintain business continuity and convenience. Alongside this, in recent years, various risks related to the integrity and reliability of information have become apparent, such as the spread of disinformation, deepfakes, and the tampering or leakage of information. The spread of disinformation and deepfakes and the tampering or leakage of information not only undermines societal trust and stability, and national security, but also has the potential to seriously impact political processes and decision-making, posing a significant threat to the health of democracy.

As highlighted in the National Security Strategy (December 2022), "cyberattacks across national borders on private critical infrastructure and the spread of disinformation are occurring constantly, blurring the line between peacetime and wartime." The threats surrounding cyberspace are becoming increasingly serious, and it could be said that the situation has become one of "Constant Crisis."

In light of this situation, efforts are being made to further ensure the safety and reliability of information and communication networks, enhance autonomous capabilities to address cyberattacks, respond to disinformation, promote international cooperation, and raise awareness and promote dissemination[5].

[5] National Security Strategy (December 2022), <https://www.cas.go.jp/jp/siryou/221216anzenhoshou/nss-j.pdf>