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

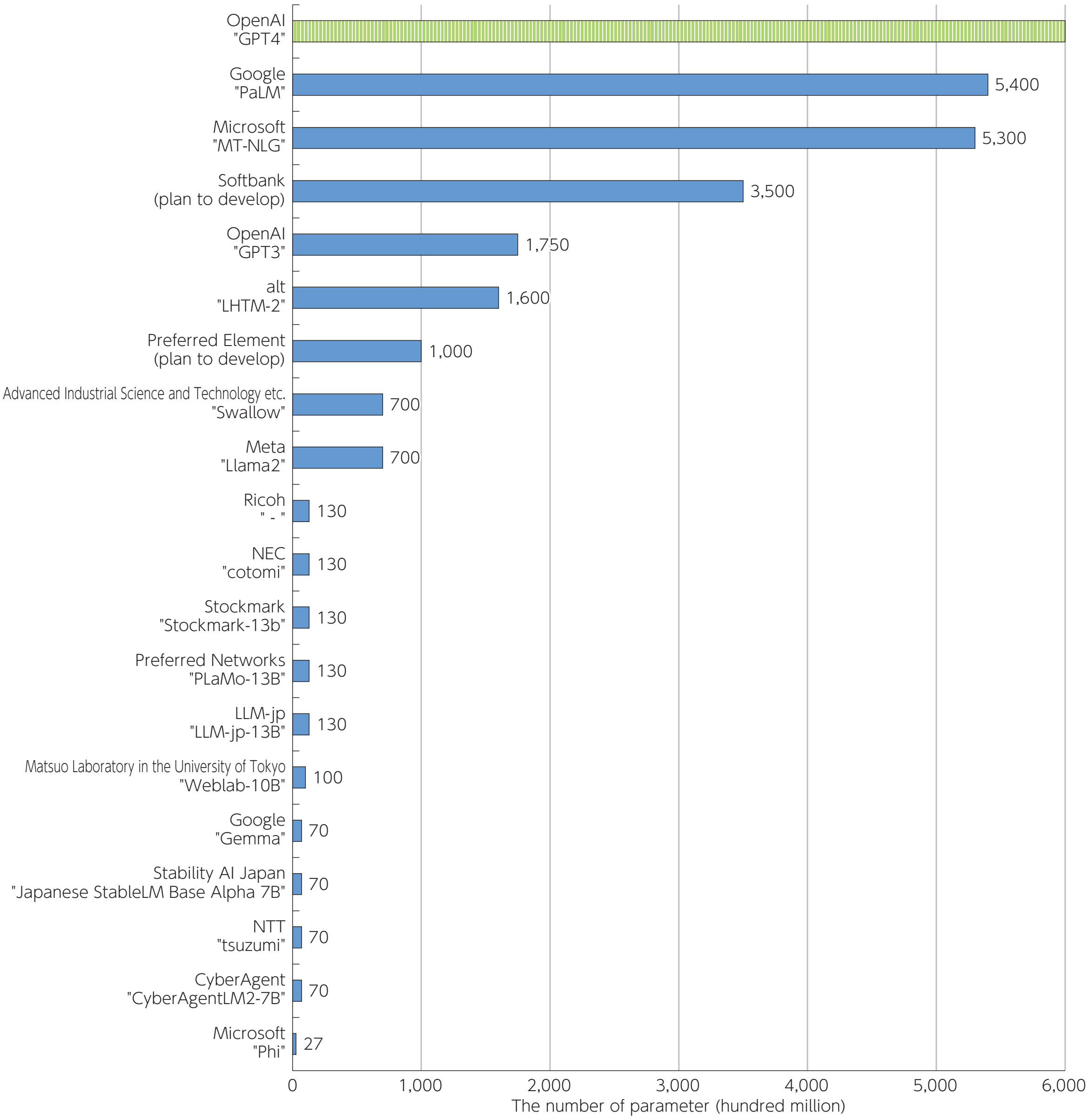
The development of Large Language Models (LLMs), which form the foundation of generative AI, is being led by major tech companies such as Microsoft and Google in the U.S.

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 nonconsensual 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

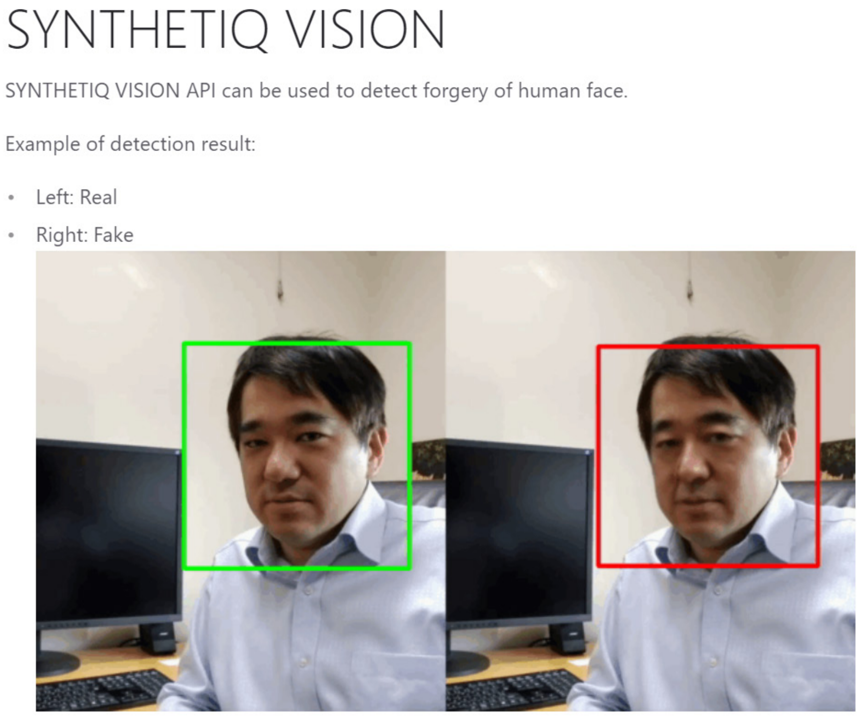
[27] Passed in December 2020. FY2021 National Defense Authorization Act and the Identifying Outputs of Generative Adversarial Networks Act (IOGAN Act) are related to the defense budget for FY2021.

[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)