TENTATIVE TRANSLATION

June2025

Report Regarding Generative AI ver. 1.0

公正取引委員会
Japan Fair Trade Commission

TENTATIVE TRANSLATION

Contents

TENTATIVE TRANSLATION

TENTATIVE TRANSLATION

[Glossary]

Graphics Processing Unit (GPU)

A type of semiconductor chip specialized for screen display and image processing. GPUs are known to be suitable for the development of generative AI due to its superior parallel processing capability that can process large amounts of data simultaneously.

Foundation Model (FM)

The core technology behind generative AI and the foundation for creating individual models. FMs are created through pre-training using large amounts of data. Fine-tuning of the foundation model allows it to be optimized for a specific task or field.

Large Language Model (LLM)

A type of language-specific generative AI model. LLMs are pre-trained using a vast amount of text data and can generate text for conversations, writing papers, etc. similar to humans.

Pre-training

The process of initial learning through large amounts of data that takes place when developing foundation models. Pre-training on text data, for example, may involve billions of tokens (words and phrases). This process allows the model to acquire extensive knowledge and lays the foundation for its application to a wide variety of tasks.

Fine-tuning

Additional training performed on a pre-trained model to optimize it for a specific task or field. The process further improves the performance of the model by specializing it to a specific business domain or application.

Open-source/Closed-source

Open-source means that the technical specifications are made available to the public so that anyone can freely improve and optimize them. In contrast, closed-source refers to when the technical specifications are not made public and can only be accessed by specific parties.

1. Introduction

1 Background of The Start of The Market Study

Generative AI offers numerous benefits to the economy and society, such as enhancing business productivity and providing a range of services. It also has the potential to drive transformations in existing businesses and foster the creation of new business models, indicating its capacity to spur new innovations and further development. For example, in the field of applications utilizing generative AI, new applications that provide individually optimized services based on user behavior and preferences are emerging one after another.

On the other hand, various risks have been pointed out with regard to generative AI, including the risk of disrupting society through infringement of intellectual property rights including copyright, the proliferation of false or misleading information, as well as security risks such as its potential use in cyberattacks. Some of these risks are pointed out as potential risks from a competition policy perspective, such as limited access to computational resources, etc., necessary for the development of generative AI. In light of this, the Japan Fair Trade Commission (JFTC) has launched a market study to understand the actual conditions of the market for generative AI from the viewpoint of maintaining a fair and free competition environment in Japan's generative AI markets, ensuring the sustainable development of the technology to generate further innovation, and integrating generative AI in a sound manner for the economy and society at large. In October 2024, the JFTC published a discussion paper titled "Generative AI and Competition" (hereinafter referred to as the "previous paper") to solicit information and opinions widely from parties concerned[1]. In light of the current state of flux in generative AI markets, the JFTC decided to proceed with this survey in a more agile, rapid, and flexible manner than previous surveys.

In the previous paper, the JFTC organized the market structure of the generative AI chain into three layers: "infrastructure," "models," and "applications." The paper also provided an overview of each layer to the best of the JFTC's knowledge at the time of publication, and presented antitrust and competition policy topics for the purpose of gathering input for future discussions.

2 Publication of The Report Based on The Results of The Information and Opinions Solicited, etc.

As a result of soliciting information and opinions through the previous paper, domestic and foreign generative AI developers and users offered their unique input, indicating a high level of interest in the generative AI markets. The majority of the 712 comments and opinions received were from individual businesses and general consumers, but various enterprises and trade associations, including domestic and foreign model developers, also submitted numerous

[1] The Japan Fair Trade Commission, "Discussion Paper, Generative AI and Competition" (October 2024), https://www.jftc.go.jp/en/pressreleases/yearly-2024/October/1002.html

comments and opinions on the current landscape and situation of generative AI markets. In addition to analyzing these inputs, the JFTC also conducted interviews with domestic and foreign enterprises, experts, relevant ministries and agencies, overseas authorities, etc. totaling roughly 50 stakeholders in parallel, collecting various opinions, including the most up to date information.

The JFTC analyzed trends in the information and opinions collected through the request for comments, hearings, etc. (hereinafter referred to as "hearings, etc."), carefully selected major themes, and compiled this "Report Regarding Generative AI ver.1.0" (hereinafter referred to as the "Report") in a more targeted form than the previous reports.

The JFTC will continue its market study on generative AI, and will update and add information to this report as it gains a better understanding of the actual conditions in the market.

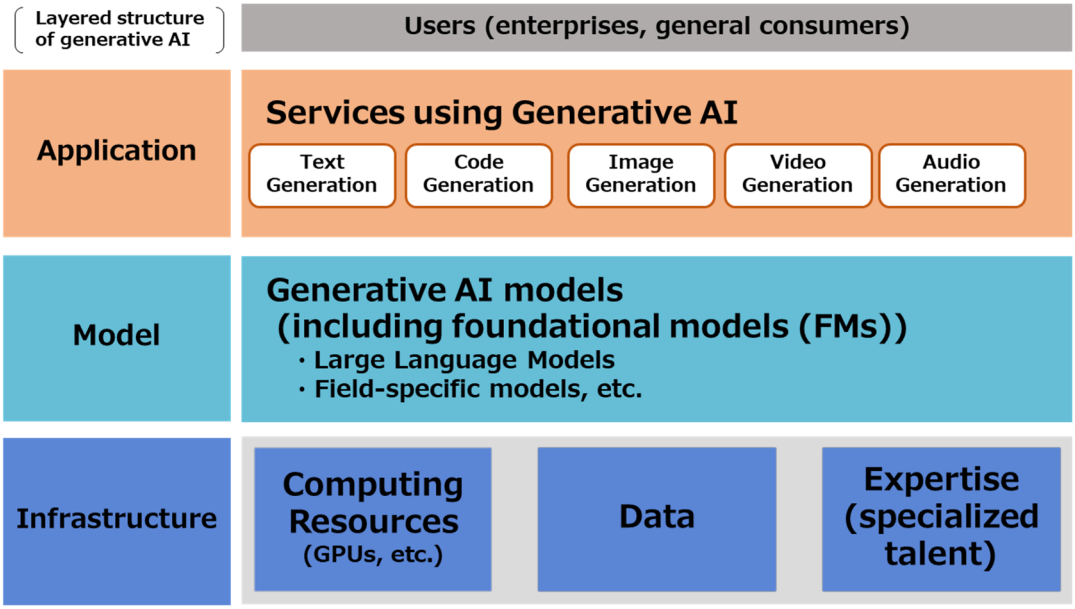
2. Structure of Generative AI Markets

Generative AI is an artificial intelligence-related technology[2] that has the ability to generate various types of content, such as text, videos, still images, illustrations, and audio, in response to prompts (questions or task instructions). Generative AI is developed using a model that has been fine-tuned for specific tasks or fields on top of a foundation model pre-trained with large amounts of data. Applications and services capable of generating text, movies, pictures, audio, and other content using this model (hereinafter referred to as “generative AI products”) are being newly deployed. The market structure of the generative AI markets is layered, consisting of multiple layers, with both domestic and foreign enterprises engaged in economic activities in each layer.

In the previous paper, the JFTC examined the structure of the generative AI sector, organizing it into three layers, as described in 1. 1 above, and in this report, we similarly organize the three layers, with some updated content, as shown in Figure 1.

[2] “Artificial intelligence-related technology” refers to technology necessary to realize functions that replace the intellectual abilities related to human cognition, reasoning, and judgment by artificial means, as well as information processing system technology for realizing functions that process input information using such technology and output the results.

Figure 1: Structure of Generative AI Markets



Source: Compiled by the Japan Fair Trade Commission

1 Infrastructure Layer

The infrastructure layer represents a market that acts as the foundation for generative AI. The three main elements of this layer are computing resources (such as GPUs), data, and expertise (specialized talent).

(1) Computing Resources (such as GPUs)

a. Summary of Previous Paper

- In order to significantly improve the versatility of the generative AI model[3], it is essential to develop the model using high-performance semiconductor chips. Among various types of semiconductor chips, GPUs (Graphics Processing Units) are said to be suitable for the development of generative AI models because they can significantly reduce training time.

- In the GPU market, NVIDIA holds approximately 80% of the global market share. The reasons for this are pointed out as (1) high performance (fast computation), (2) a good usage environment for development enterprises by providing mechanisms for efficient use of GPUs and facilitating parallel computation, and (3) continued improvement of these technologies by devoting resources to them.

[3] "Generative AI model" is a generic term for models used for generative Al products, and includes foundation models and fine-tuned foundation models.

- The supply of GPUs is in short supply worldwide relative to demand, and competition to acquire them is intensifying. As a result, many enterprises around the world are working on development for the supply of GPUs, but the gap between NVIDIA and them has not been bridged.

b. Updates After Hearings, etc.

(a) NVIDIA GPUs

As noted in the previous paper, there is a strong correlation between the number of parameters[4] and computational complexity in the performance of a generative AI model. In addition, a certain model size is required to significantly improve versatility, and in order to develop such a high-performance generative AI model, it is essential to deploy a sufficient amount of semiconductor chips suitable for developing generative AI models. Among various types of semiconductor chips, GPUs, among others, have high parallel processing capability because they were originally designed for image processing, and are considered to be the most suitable semiconductor chips for developing generative AI models, with the performance to significantly reduce computation time. In fact, some enterprises consider GPUs essential for the development of generative AI models.[5] In the GPU market, NVIDIA holds approximately 80% of the global market share, and its products continue to enjoy a high market share.

In the previous paper, we mentioned that NVIDIA GPUs have a high market share because of their high performance (fast computation) and development environment for development enterprises, etc. In particular, NVIDIA's development environment (e.g. CUDA[6]) is said to offer high usability and scalability because deep learning frameworks like PyTorch and TensorFlow[7] are highly optimized for NVIDIA GPUs by utilizing CUDA libraries (such as cuDNN, etc.). In addition, only NVIDIA GPUs are generally supported for these NVIDIA development environments.

On the other hand, regarding the supply and demand situation for NVIDIA GPUs, it

[4] Generative AI models are composed of numerous variables, known as parameters, which are essential for processing input data in foundational models. Generally, a higher number of parameters correlates with improved performance of the generative AI model. However, this also brings challenges such as longer training times, higher costs, and increased computational demands during inference.

[5] See Bruegel, "COMPETITION IN GENERATIVE ARTIFICIAL INTELLIGENCE FOUNDATION MODELS," published June 18, 2023, p. 7.

[6] CUDA is a parallel computing platform and programming model developed by NVIDIA that enables general-purpose computation utilizing GPUs.

[7] A deep learning framework is a software library for efficiently building, learning, and reasoning about neural networks.

used to be difficult to secure GPUs, but the situation has changed due to the increased supply of NVIDIA GPUs. The following comments were offered by enterprises[8] .

- The global supply of GPUs falls short of demand, the cost for access to chips is incredibly high – indeed, it is prohibitively high for many companies. [Overseas model development enterprise].
- Currently, GENIAC[9] and NEDO[10] have secured ample computing resources, but before that, it was difficult to secure GPUs and the price was so high that there was internal debate as to whether the cost was worth the development costs. The difficulty in accessing and the high cost of using GPUs may lead companies to give up on R&D. [Domestic model development enterprise]
- Although the depletion of GPU resources was pointed out from early spring to summer of 2024, supply and demand for chips are matching, and as a general matter, procurement may be possible if there are sufficient funds, although there may be some restrictions on procurement due to lack of funds. [Domestic model development enterprise]

(b) Training and Inference Phases

As described in 2. 1 (1) b. (a) above, NVIDIA GPUs have a high market share in the GPU market, but on the other hand, it has been pointed out that the supply and demand situation is beginning to change, as semiconductor chips that can compete with NVIDIA GPUs are being developed mainly by Big Tech companies. However, different situations were observed with regard to the efforts of enterprises in supplying computing resources, including GPUs, at the training phase and inference phase of the generative AI models.

a Training Phase

The training phase refers to the phase in the process where a generative AI model uses a large amount of data to find common rules and features, and then generates appropriate responses and content based on these rules and features. Since the training phase requires the ability to process a huge amount of calculations, GPUs are the mainstream chips, as they are suitable for parallel processing, and among

[8] In this report, we have faithfully reproduced the comments we received to the extent possible so that nuances are not lost (the same applies to the following comments from enterprises, etc.).
[9] GENIAC is a project by the Ministry of Economy, Trade and Industry (METI) to strengthen the development capability of generative AI in Japan by providing computational resources for the development of basic models, the core technology for generative AI, promoting collaboration among related parties, and disseminating information to the public. (https://www.meti.go.jp/policy/mono_info_service/geniac/index.html)
[10] NEDO is a national research and development corporation that creates innovation through the promotion of research and development necessary to realize a sustainable society. (https://www.nedo.go.jp/)

them, NVIDIA GPUs are market leaders due to their good usage environment and other features. The following comments were offered by enterprises.

- For the training phase, we believe NVIDIA will be the dominant player for the next 10 years. Since the training phase uses a larger number of GPUs than the inference phase, the ability to integrate and operate these chips as a whole is a major difference from the inference phase, and NVIDIA is ahead in terms of technological achievements, including this. [Domestic model development enterprise]
- Since NVIDIA GPUs are the gold standard in training and the life cycle is advanced even for NVIDIA GPUs, we believe that we will continue to use NVIDIA GPUs in our development. [Overseas model development enterprise]
- For the training phase, NVIDIA has maintained its dominance in the market, primarily due to the advanced development environments and libraries it offers, such as CUDA. However, these libraries are gradually being adapted to support devices from other manufacturers. For instance, code initially written for NVIDIA GPUs is now increasingly capable of running on GPUs from other companies, which has started to ease NVIDIA’s exclusive hold on the market. [Domestic enterprise]

b Inference Phase

The inference phase refers to the phase in the process where a model that has gone through the training phase analyzes new input data and generates optimal responses and content based on patterns and knowledge acquired during training.

In the inference phase, semiconductor chips that can process at high speed and low power consumption are required for efficiency, and generally do not need the ability to process as large a computational volume as in the training phase. In addition to semiconductor chips manufactured by NVIDIA, semiconductor chips developed by Big Tech companies, existing semiconductor manufacturers, and startups are also attracting attention in the market for semiconductor chips for inference. On the other hand, NVIDIA is also strengthening its GPUs for inference, and competition for semiconductor chips for inference is active. The following comments were offered by enterprises.

- NVIDIA's presence will remain strong in inference as well, but other companies are also developing chips, and it is possible that in two to three years, chips offered by other companies will be the chips of choice. Inference can be

completed on a relatively small scale, and there are fewer situations where severe operations are required compared to training. [Domestic model development enterprise]

- NVIDIA has a full range of GPUs used for inference, but in a number of cases, inference is processed by CPUs, and there are companies that are developing chips for inference. [Domestic enterprise]
- For the inference phase, various companies are developing their own chips, but the transition away from GPUs is sluggish. One reason for this is the high implementation cost in terms of human resources. With NVIDIA GPUs, companies can immediately run code they've developed in the past, thanks to the widespread use of CUDA. On the other hand, with non-NVIDIA chips, even if a specific model works, there's no guarantee it will work across other models, necessitating additional engineering measures. This process can take one to two months, creating significant time costs. As a result, companies might hesitate to consider switching to non-NVIDIA chips. [Domestic model development enterprise]

c. Summary

In the GPU market, NVIDIA's GPUs hold a significant share globally, primarily due to factors like development environments and CUDA. However, changes have been observed in the supply-demand dynamics as supply has increased. In the provision of computing resources, particularly during the training phase, NVIDIA's GPUs are likely to maintain their status as leaders. On the other hand, the market for inference phase computing is witnessing heightened competition compared to the training phase, driven by factors such as the emergence of new entrants. This has led to innovation across diverse players in the market.

(2) Data

a. Summary of Previous Paper

- In developing a generative AI model, it is necessary to perform pre-training using training data. Fine-tuning using additional training data is also necessary to tailor the model to specific tasks and uses. Large amounts of training data are required in the development of generative AI models.

- Problems have arisen over the use of data between data owners and generative AI model development enterprises, especially in overseas. For example, lawsuits have been filed

against foundation model development enterprises on the grounds of copyright infringement, etc., in relation to the use of various information available on the Internet for pre-training.

● Japanese language data is important for generative AI products using large language models used in Japan, but Japanese has fewer users than English, resulting in less data in Japanese than in English, and the amount of data available for training is relatively small. Under these circumstances, domestic enterprises may be able to develop a generative AI model specialized for Japanese that is superior to the language models developed by Big Tech companies and others.

b. Updates After Hearings, etc.

(a) Quantity and Quality of Data

The data is used in a variety of situations and usage scenarios, including the various stages of generating AI models and subsequent usage scenarios. For example, RLHF[11] can be used to generate more pertinent responses. It is also possible to generate more appropriate responses utilizing external knowledge by combining a mechanism called RAG.[12]

The amount of data required for the development of a generative AI model, for example, depends on the situation and usage. A large amount of data is required for pre-training in order to learn language structures and basic knowledge. In the fine-tuning stage, data is required for purposes such as improving the accuracy of responses to user instructions and questions. Data used in RAG needs to be organized so that the necessary information can be searched appropriately. On the other hand, when developing specialized models for specific industries and usage, such as finance and medicine, high-quality data for specific industries and usage is required. The following comments were offered by enterprises and others.

- There are various technical debates, but as a general rule, pre-training tends to require a large amount of data. In the fine-tuning stage, it is necessary to improve the accuracy of responses to user instructions and questions, which requires data tailored to the purpose. In mechanisms that retrieve external knowledge, such as RAG, it is important to be able to easily search and extract the information that

[11] Reinforcement Learning from Human Feedback, or “RLHF,” is reinforcement learning using human evaluation.
[12] Retrieval-Augmented Generation, or “RAG,” is, according to Patrick Lewis, defined as "a language generation model that combines parametric and nonparametric memory (i.e., retrieval-based memory).” In addition to being used by companies to improve the accuracy of generative AI responses by searching internal documents, databases, etc., it is also used to search information on the Internet in real time to enable responses based on the latest data.

should be used in business, so the structure and condition of the document data used are key points. [Relevant Japanese government body]

- The first step in training a generative AI model is to have a large amount of data. At this stage, the quality of the data is not so important, and the model will focus primarily on learning basic facts about the language. The subsequent fine-tuning phase, on the other hand, requires very high quality and specific types of data. While it is relatively easy to collect the large amount of data needed for this first step, it is much more difficult to ensure the high quality data needed for fine tuning. [Overseas model development enterprise]

(b) Data Types

There are various types of data for developing generative AI models, including data available on the web, data sets available in open form, data held independently by enterprises, and data held by government entities. A large amount of data is needed to develop a general-purpose large language model. However, much of the accessible high-quality data on the Internet is already being used to train generative AI, and the open data available for training may be depleted in the future.[13] Under these circumstances, synthetic data[14] is increasingly being used to supplement the quantitative shortage of training data, or to improve the quality of training data in order to make effective use of limited data. The following comments were offered by enterprises.

- The differences in performance among generative AI models will almost disappear as the differences in available data on the Internet disappear. In the future, a small number of large general-purpose models will survive, and the training data used for those models will be almost the same. [Domestic model development enterprise]
- We recognize that the exhaustion of training data has already begun to some extent, and negotiations between model developers and data rights holders are expanding. Negotiations with publishers of textbooks, for example, are becoming increasingly important. [Overseas model development enterprise]
- Publicly accessible data is depleted, and proprietary data will become more important to competition. Big Tech companies have proprietary data sets that can be used to train models and improve them. In the future, such proprietary data will become more valuable and will make a difference in competition. [Overseas

[13] In an article in the journal Nature dated December 11, 2024, summarizes the current state of AI development and points out that training data may be exhausted by 2028. Nature: "The AI revolution is running out of data. What can researchers do?"(December 11, 2024) https://www.nature.com/articles/d41586-024-03990-2

[14] Synthetic data is artificial data, generated by computer algorithms, that is close to real data in the real world.

model development enterprise]

- Synthetic data has both positive and negative aspects, and we use synthetic data because it has the greater positive impact. Synthetic data is an effective means of overcoming data exhaustion, and the advantages are lower data procurement costs and the ability to secure large data sets. The disadvantage is that it entails issues regarding data reliability and the risk of bias. However, this point can be adjusted for quality in the subsequent development process. Furthermore, if the model is an open source model, it is expected to be subjected to verification and improvement measures can be proposed. [Overseas model development enterprise]

(c) Possibility of Uneven Distribution of Data to Certain Enterprises

In light of the situation described in 2.1. (2) b. (b) above, the importance of the proprietary data held by each enterprise is increasing. In particular, it has been pointed out that Big Tech companies have a competitive advantage because they can purchase data and use it for training with their abundant financial resources, and they can use data obtained through the services they provide for training. The following comments were offered by enterprises.

- Big Tech companies and others have an advantage in data collection because they use their non-AI customer base to collect data, and they use those data to provide AI services. [Domestic model development enterprise]
- The high demand and scarcity of high-quality input data for training tends to make it very expensive, and Big Tech companies and others with large amounts of capital at their disposal will have an advantage. [Overseas model development enterprise]
- We believe that we cannot compete with services provided by platforms that use data collected via social networking sites, for example. Therefore, we compete in niche areas. The fact that Big Tech companies have an advantage in terms of data is common in all areas, not just images and videos. [Domestic model development enterprise]

On the other hand, some pointed out that there is limited or no competitive advantage due to the uneven distribution of training data. The following comments were offered by enterprises.

- The purpose of data collection varies from enterprise to enterprise. Since it is sufficient for each player to collect data in its own area of interest, large-scale data collection is not always required, and competitive advantages due to uneven

distribution of training data is limited. [Domestic enterprise]

- The fact that there are companies that own large amounts of data but are not leading the development of generative AI models, and startups that do not own large amounts of data but have succeeded in developing generative AI models, makes it clear that companies with large amounts of data do not have a competitive advantage. The reason why companies without data pools can succeed in developing generative AI models is because the hurdle for acquiring data is low. Data on the Internet can be obtained at low cost, and open-source data and synthetic data can also be used for model development. [Overseas model development enterprise]

While the importance of proprietary data held by each enterprise is considered to be increasing, there are some who argue that Big Tech companies have an advantage in competition due to the uneven distribution of training data, while others argue that this advantage is limited or none. At this point in time, a vast amount of data accessible on the Internet has been used, and it is difficult to determine whether or not Big Tech firms have an advantage, meaning continued close monitoring is necessary.

(d) Japanese-language Data

One path for domestic companies is to develop models that focus on the Japanese language. Since it has been pointed out that to improve the performance of Japanese in generative AI, it is necessary to specialize in training Japanese grammar and syntax, and since translation data may not adequately reproduce natural Japanese expressions and cultural context, it is important to ensure high-quality Japanese-language data in the development of models for the Japanese market. Under these circumstances, it has been pointed out that domestic businesses hold an advantage in their ability to refine the generated Japanese into accurate and polished language, as well as allocate personnel to effectively reach the Japanese market. The following comments were offered by enterprises.

- When it comes to the Japanese language performance of generative AI models, two factors come into play: linguistic capability and knowledge. For linguistic capability, achieving a level equivalent to 70–80 points can be accomplished with English-based training data. However, reaching a 90-point level likely requires training on Japanese-based data. As for knowledge, the inherent information embedded in English-language data and Japanese-language data differs fundamentally. In this context, the Japanese language performance of generative AI models from Big Tech companies has improved to an estimated 50–70 points, but has not yet reached the 90-point level.

That said, it seems unlikely that Big Tech companies see a need for their models to achieve such a high level of Japanese proficiency in the first place. [Domestic model development enterprise]

- The use of models made overseas may result in unnatural Japanese output. While it may be difficult for foreign companies to judge this, we believe that the ability to accurately correct unnatural Japanese is a strength of Japanese companies. [Domestic model development enterprise]
- Domestic companies lagged behind their foreign counterparts in model development by many rounds. Although language is not a factor in model development itself, since training data is mainly in English, it is important for domestic companies to develop models with an emphasis on Japanese. [Domestic model development enterprise]

On the other hand, it has been pointed out that it is relatively easy for foreign enterprises to collect Japanese-language data. The following comment was offered by an enterprise.

- It is not clear how much advantage Japanese companies have with Japanese language models. Since data on the Web is used for model development, data can be collected regardless of language, and global models can learn Japanese well enough. In addition, from the perspective of improving model performance, the current basic idea is that by learning a large amount of data, even global models, which have a smaller percentage of Japanese being learned than other languages such as English, show high Japanese performance, and there is not much of a language barrier. [Domestic model development enterprise]

Some point out that high-quality Japanese-language data is important for improving the Japanese-language performance of the generative AI models and that Japanese-language-based training is necessary, giving domestic firms an advantage. On the other hand, some point out that even if the percentage of Japanese in the training data is low, high Japanese performance can be achieved. At this point, a certain amount of Japanese-language data is considered necessary, but it is difficult to determine how much Japanese-language data is needed to train a high-performance Japanese-language model and how the availability of Japanese-language data will affect the market position of model development enterprises. It is necessary to continue close monitoring of the situation.

(e) Processing of Rights, etc. in The Collection and Use of Data

Regarding data, there were opinions expressing concerns about whether appropriate

rights management and handling procedures are being followed. Additionally, some highlighted the evaluation of access environments, including Japan's institutional framework. The following comments were offered by individuals and an enterprise.

- Under the current Copyright Act, data can be used for AI training if the purpose is not to enjoy the thoughts or sentiments expressed in a work, but it seems that it is often used for to enjoy the thoughts or sentiments expressed in a work. [Individual]
- If AI developers do not handle data rights appropriately and are able to offer services at lower prices than developers who do handle rights appropriately, this cannot be considered fair competition. [Individual]
- In future AI development, clean models that use training data collected after appropriate rights processing may become a new axis of competition. [Individual]
- I believe that there is good access to data in Japan. I believe this is due to the domestic legal framework, including Article 30-4 and Article 47-5[15] of the

[15] (Exploitation without the Purpose of Enjoying the Thoughts or Sentiments Expressed in a Work)
Article 30-4: It is permissible to exploit a work, in any way and to the extent considered necessary, in any of the following cases, or in any other case in which it is not a person's purpose to personally enjoy or cause another person to enjoy the thoughts or sentiments expressed in that work; provided, however, that this does not apply if the action would unreasonably prejudice the interests of the copyright owner in light of the nature or purpose of the work or the circumstances of its exploitation:
(i) if it is done for use in testing to develop or put into practical use technology that is connected with the recording of sounds or visuals of a work or other such exploitation;
(ii) if it is done for use in data analysis (meaning the extraction, comparison, classification, or other statistical analysis of the constituent language, sounds, images, or other elemental data from a large number of works or a large volume of other such data; the same applies in Article 47-5, paragraph (1), item (ii));
(iii) if it is exploited in the course of computer data processing or otherwise exploited in a way that does not involve what is expressed in the work being perceived by the human senses (for works of computer programming, such exploitation excludes the execution of the work on a computer), beyond as set forth in the preceding two items.
(Minor Exploitation Incidental to Computerized Data Processing and the Provision of the Results Thereof)
Article 47-5 (1) A person undertaking an action as set forth in one of the following items that contributes to facilitating the exploitation of a work by creating new knowledge or information through computerized data processing (this includes a person undertaking a part of such an action; limited to one doing so in accordance with the standards prescribed by Cabinet Order) may exploit a work that has been made available or presented to the public (this includes a work that has been made available for transmission; the same applies hereinafter) (hereinafter in this Article and Article 47-6, paragraph (2), item (ii) referred to as a "available or presented work") (limited to a publicized work or a work made available for transmission), in any way and to the extent considered to be necessary in light of the purpose of the action set forth in the relevant item, when exploiting it incidental to the undertaking of that action (limited to exploitation that is minor in light of the percentage it constitutes of the part of the available or presented work that has been provided for exploitation, the volume of the part of that work that has been provided for exploitation, the accuracy of indications made at the time it was provided for exploitation, and other elements; hereinafter in this Article referred to as "minor exploitation"); provided, however, that this does not apply if the person undertakes that minor exploitation knowing that the making available or presentation of the works to the public constitutes copyright infringement (for a work made available or presented to the public abroad, this means that the action would constitute copyright infringement if it took place in Japan); nor does this apply if the action would otherwise unreasonably prejudice the interests of the copyright owner in light of the nature or purpose of the available or presented work or the circumstances of its minor exploitation:
(i) using a computer to search for the title or author name of a work in which information that a person is searching for (hereinafter in this item referred to as "information being searched for") has been recorded, for the transmitter identification code (meaning the letters, numbers, symbols, or any other code by which the transmitter of an automatic public transmission is identified; the same applies in Article 113, paragraphs (2) and (4)) associated with information being searched for that has been made available for transmission, or for any other information concerning

Copyright Act. [Overseas model development enterprise]

In relation to this, for example, the relevant ministries and agencies have organized their thinking on the relationship between AI and intellectual property rights, and are working to raise awareness and understanding[16].

c. Summary

The demand for data is not only about "quantity" but also about "quality" depending on the situation and usage. However, it is difficult to determine at this point whether the uneven distribution of training data will give rise to a competitive advantage, how much Japanese-language data will be needed to train high-performance Japanese-language models, and what impact the availability of Japanese-language data will have, and these issues require continued close monitoring.

(3) Expertise (specialized talent)

a. Summary of Previous Paper

- There are only a limited number of highly skilled professionals who can develop GPUs and generative AI models, and the difficulty in acquiring such personnel has been a bottleneck in the development of generative AI models and other products.

- Because Big Tech companies are well-funded, they tend to have a high concentration of

---

the identification or location of information being searched for; and making the results of that search available;
(ii) undertaking computerized data analysis and furnishing the results of that analysis;
(iii) an action that Cabinet Order prescribes as contributing to increased convenience in the lives of the citizenry by creating new knowledge or information through computerized data processing and making the results of this available, beyond what is set forth in the preceding two items.
(2) A person that prepares to undertake an action set forth in one of the items of the preceding paragraph (limited to a person who collects, organizes, and provides information in preparation to undertake the action in accordance with the standards prescribed by Cabinet Order) may reproduce or make public transmissions of an available or presented work (or make the relevant work available for transmission, if such transmission is being made via an automatic public transmission; the same applies in this paragraph and Article 47-6, paragraph (2), item (ii)) or distribute copies thereof, to the extent considered to be necessary in order to prepare for minor exploitation under the preceding paragraph; provided, however, that this does not apply if the action would otherwise unreasonably prejudice the interests of the copyright owner in light of the nature or purpose of the available or presented work, the number of copies that would be reproduced or distributed, or the circumstances of the reproduction, public transmission, or distribution.

[16] ・Agency for Cultural Affairs, “Approach to AI and Copyright” (March 2024) (https://www.bunka.go.jp/seisaku/bunkashingikai/chosakuken/pdf/94037901_01.pdf)
・“Checklist and Guidance on AI and Copyright” (July 2024) (https://www.bunka.go.jp/seisaku/bunkashingikai/chosakuken/seisaku/r06_02/pdf/94089701_05.pdf)
・Cabinet Office Intellectual Property Strategy Promotion Office, “AI Era Intellectual Property Rights Review Committee Interim Report: A Guide for Rights Holders” (May 2024) (https://www.kantei.go.jp/jp/singi/titeki2/chitekizaisan2024/0528_ai.pdf)
• Ministry of Economy, Trade and Industry “Guidebook on the Utilization of Generative AI for Content Creation” (July 2024) (https://www.meti.go.jp/policy/mono_info_service/contents/aiguidebook.html)

highly specialized personnel, making it a hurdle for domestic enterprises to attract these individuals.

b. Updates After Hearings, etc.

(a) Competition for Human Resources, Including Big Tech Companies, etc.

The rapid development and market expansion of generative AI technology has led to a significant increase in the demand for specialized talent in the development of generative AI, and the highly specialized researchers, engineers, data scientists, and governance professionals who support the development and application of generative AI models and products are an essential driver of industry growth. In particular, Big Tech companies are noted to have an advantage in the competition for highly specialized human resources, backed by their financial strength, computational resources, and other assets. The following comments were offered by enterprises.

- Big Tech companies have an advantage in the competition for specialized human resources, backed by their financial strength and development resources. [Domestic enterprise]
- The development of generative AI models is not profitable on its own on a global scale at this time, and it is difficult for companies without sufficient capital to make investments, including in human resources, in this field. Therefore, the number of businesses that can provide an environment in which highly specialized personnel can demonstrate their abilities is extremely limited in Japan. [Domestic enterprise]
- The shortage of skilled workers may become increasingly apparent. There's a risk of losing out in the competition for talent to Big Tech companies, leading to further brain drain. Unable to match their offers, this disparity in company strength becomes evident—but from the perspective of those on the ground, the situation feels quite harsh. [Domestic model development enterprise]

However, it has been noted that Big Tech companies do not always have a monopoly on highly specialized talent, and that there is a certain degree of mobility of highly specialized talent. For example, there are examples of engineers who start their own businesses or leave Big Tech companies to take up new positions at startups or other companies. In addition, there were opinions that securing highly specialized talent depends on the brand power of the company and other factors. The following comments were offered by enterprises.

- We do not believe that highly specialized talent is concentrated within any one company. We recognize that highly specialized personnel are highly mobile and

there are many examples of engineers leaving Big Tech companies to take new positions. [Overseas model development enterprise]

- The problem of difficulty in securing human resources and data is a problem that exists globally, and since brand power and attractiveness can attract human resources, we believe it is the responsibility of each company to solve such problems. [Domestic model development enterprise]

(b) Competition for Human Resources among Domestic Enterprises

As mentioned in 2.1 (3) b (a) above, it has been pointed out that there is a certain degree of mobility of highly specialized talent, but such mobility is thought to occur mainly in Big Tech companies, and not all companies, including domestic enterprises, are benefiting from it in the same way. Challenges faced by domestic enterprises in securing professional human resources include lower compensation levels and lack of R&D resources compared to Big Tech companies. On the other hand, domestic enterprises have strengths such as the ease of training and hiring local human resources were also pointed out. The following comments were made by enterprises.

- The situation is different in Japan and the U.S. Big Tech companies have been acquiring Silicon Valley startups, but almost no Japanese companies have been acquired by Big Tech companies. Japanese people are conservative, so few will go directly to Silicon Valley to work. There are also not many people who go to work for the Japanese subsidiaries of foreign companies. Although there is often talk of job transfers from large Japanese companies to the Japanese subsidiaries of Big Tech companies, I don't think there are many other channels for the influx of human resources. [Domestic model development enterprise]
- Various research institutions in Japan have created a rich and fluid local talent pool that can develop and commercialize models and applications for the Japanese market, especially with their Japanese language skills. [Overseas model development enterprise]

c. Summary

As the demand for skilled professionals in generative AI continues to rise, conditions of competition for human resources are becoming more active, both domestically and internationally. Under such conditions, Big Tech companies are expected to have an advantage in the competition for professional human resources, backed by their financial strength and development resources. In contrast, domestic enterprises have some challenges, such as low compensation levels and lack of R&D resources. However, they

also have strengths, such as the ease of training and recruiting local human resources.

2 Model Layer

The model layer represents a market focused on developing generative AI models, which are pre-trained using vast amounts of data, often leveraging cloud technologies.

(1) Summary of Previous Papers

- Generative AI models can be classified into three main categories: (1) text generation, (2) image (video and 3D model) generation, and (3) voice generation, with text generation being the current mainstream. Generative AI models can be classified into general-purpose models that can be applied to a variety of tasks and specialized models for specific industries and usage, such as finance and medicine. Domestic companies have developed large language models that are both lightweight and highly precise, with some highlighting these features as strengths.

- With regard to large language models, there is currently active competition in development, mainly among Big Tech companies. In addition, domestic enterprises are also developing generative AI models that enable output based on Japanese language performance and business practices unique to Japan, etc. Rather than competing with general-purpose models developed by Big Tech companies, etc., there is a trend toward the development of specialized models for specific industries and usage.

- Image and voice generation technologies have developed based on existing image recognition and voice recognition technologies, and further development is expected. Multimodal technologies that integrate and process different types of data, such as text, images, and voice, are also making progress.

(2) Updates After Hearings, etc.

a. Trends in Large Language Models, etc.

As for large language models, where development is underway by various companies, competition for development is intensifying, especially among Big Tech companies, and new models are being released one after another. For example, in March 2025, Google

unveiled “Gemini 2.5 Pro[17]” followed by Meta's “Llama 4[18]” and Open AI's “o3” and “o4-mini”[19] in April of the same year. Additionally, in May 2025, Anthropic released “Claude Opus 4” and “Claude Sonnet 4”[20].

These large language models were developed from scratch by the companies themselves, and are said to be primarily for general-purpose use.

Although general-purpose large language models have been developed by domestic companies, developing such models generally requires a large investment in computational resources, a vast amount of data collection, many highly specialized human resources, and long-term development time, so Big Tech companies with abundant capital and technical capabilities are said to have an advantage. The following comments were offered by enterprises.

- The key resources required for large language model development, such as computing resources, specialized human resources, and funds, are concentrated in Big Tech companies, which further anchors the structure of competition. [Overseas model development enterprise]
- In terms of the development of generative AI models, companies that possess large-scale, general-purpose language models have the upper hand, and we believe that an oligopoly situation among several companies is inevitable in the future. [Domestic model development enterprise]

On the other hand, mainly by domestic companies and startups, there are examples of companies that use the foundation models (open source models, etc.) provided by other companies to specialize in Japanese language performance, etc., or to achieve efficiency in computing resources and adaptation to specific usage (medical, financial, legal, manufacturing, etc.) to differentiate their models from the generic models of Big Tech companies. It is also said that Big Tech companies that possess high-performance foundation models are still in an advantageous position in the development of specialized models. The following comments were offered by enterprises.

- In the future, generative AI models will not be devoted to a particular type (e.g., large language models), but rather many different types will thrive and even complement each other. As a recent trend, many companies, including startups, are developing powerful AI models that operate with fewer parameters and less computing power.

[17] The Experimental version was released on March 25, 2025, and has been continuously updated since then. https://blog.google/technology/google-deepmind/gemini-model-thinking-updates-march-2025/
[18] https://ai.meta.com/blog/llama-4-multimodal-intelligence/
[19] https://openai.com/ja-JP/index/introducing-o3-and-o4-mini/
[20] https://www.anthropic.com/news/claude-4

[Overseas model development enterprise]

- Japan is a step behind in the competition for large- language models for the general public, but as future generative AI becomes more decentralized and specialized, we believe that the source of our competitiveness will be to create highly specialized generative AI, especially in areas where Japan has strengths, such as manufacturing and materials science. [Overseas model development enterprise]
- We suppose many of the companies that have developed language models that achieve world-class Japanese language proficiency in benchmarks that measure Japanese language knowledge and reading comprehension are actually using models developed by foreign Big Tech companies as foundation models. Few companies are developing Japanese-specific models from the training stage of the foundation model, nor is this an efficient way to do so. Therefore, while the development of Japanese-specific models remains possible, it is anticipated that such models will primarily emerge as fine-tuned adaptations of the latest foundational models (which are not Japanese-specific). This situation is not expected to change significantly in the foreseeable future. [Domestic enterprise]
- The development of specialized models and applications based on data on the Web ultimately favors Big Tech companies, because the creation of specialized models and applications also requires an foundation model, and the performance of that model affects the performance of the specialized models and applications. [Domestic model development Enterprise]

Chinese-made generative AI models have also attracted attention. For example, in January 2025, Chinese startup DeepSeek released a large language model (DeepSeek-R1). This model is said to have performance comparable to the latest models of other companies, and has had a major impact on the U.S. stock market, although concerns about its safety and other issues have been reported[21]. The following comment was offered by an enterprise.

- DeepSeek has been reported to have very low development costs, resulting in NVIDIA's stock price decline. Consequently, technologies such as MoE (Mixture of Experts) and knowledge distillation[22] are also being talked about and are in active development. New models using knowledge distillation and MoE are emerging from Alibaba's Qwen, and the power of open source is stimulating competition. [Overseas model development enterprise]

[21] DeepSeek was also mentioned as one of the factors influencing the demand outlook of the companies (Nihon Keizai Shimbun, "NVIDIA 'One Strong' Turning Point, Driving the Shifting AI Market" (February 27, 2025), https://www.nikkei.com/article/DGXZQOGN260GV0W5A220C2000000/).

[22] See 2.2 (2) c. below for MoE (Mixture of Experts) and knowledge distillation.

b. Other Trends in Generative AI models

In addition to the large language models described in the 2.2(2)a above, competition in the development of generative AI models that generate images, audio, video, etc. is intensifying, and new models are being released. Although those models basically face the same issues as those in the development of large language models, some of the situations specific to each model were also observed. The following comments were offered by enterprises.

- The amount of storage required for the development of a speech model is large because of the large amount of speech data being stored. For text and voice data in the same situation, the amount of storage required is 100 to 1,000 times greater than for text. Simply storing the data is the bottleneck, and keeping the data on hardware during training is also a bottleneck. [Domestic model development enterprise]
- The development of image generation models, like LLM development, requires an enormous amount of computation. The difference is that in LLM development, a small number of global tech companies such as OpenAI, Google, and Anthropic are extremely strong, making it difficult for Japanese enterprises to win, but in the image field, there is no single leader, and there is a lot of open data, so it is not difficult to compete on a global scale. [Domestic model development enterprise]

In recent years, the range of applications for generative AI has expanded significantly, resulting in the mainstream evolution from large language models to multimodal AI that integrates and processes different types of data, such as text, images, audio, and video. The emergence of multimodal AI is expected to expand the range of applications of AI in medicine, drug discovery, education, entertainment, and other fields, as well as improve processing capabilities in general tasks such as inference and analysis.[23]

These technological advances will further expand the range of applications of generative AI, and competition for multimodal AI and products that use it, etc. will likely increase, contributing to the growth of the generative AI markets in the future. The following comments were offered by enterprises.

- The video/image modality and multimodal foundation models have been the focus of much attention as foundation models in the pipeline. Models that combine text, image, and audio generation products are bringing more dynamism to generative AI,

[23] Ministry of Internal Affairs and Communications and Ministry of Economy, Trade and Industry, "AI Enterprise Guidelines (Version 1.1) Appendix (Appendix)" (March 28, 2025), p. 13 (https://www.soumu.go.jp/main_content/001000988.pdf https://www.meti.go.jp/shingikai/mono_info_service/ai_shakai_jisso/pdf/20250328_3.pdf)

as evidenced by the development of numerous models that offer all modalities, including one that can generate high-definition video based on input text. [Overseas model development enterprise]

- Although each company has its own definition of multimodal, it is generally appropriate to interpret multimodal as "a generative AI model that can process not only text, but also a combination of multiple elements (modals) such as images, video, and audio." One notable aspect of multimodal is that it is not only capable of processing multiple elements simultaneously, but also of input and output in different formats. For example, inputting an image and describing it in text is an example of multimodal. [Overseas model development enterprise]

c. Model Development Methods

Behind the highly accurate technology of generative AI lies the challenge of requiring huge models, large numbers of parameters, and vast amounts of computational resources. In addition, in order to enable operation in resource-constrained environments such as on mobile terminals, there is a need to make models smaller and inference more efficient. To address these issues and others, methods such as "distillation" and "Mixture of Experts (MoE)," which are not necessarily new methods, are attracting attention as approaches aimed at making models smaller, lighter, and more efficient.[24]

(a) Distillation

Distillation (knowledge distillation) is generally a method of transferring the wealth of knowledge learned by a pre-trained generative AI model (teacher model) to another generative AI model (student model) that is smaller and more efficient. Typically, a teacher model requires a huge amount of parameters and computational resources, while a distilled student model is said to have a major advantage in that it can achieve the same or close performance as a teacher model in a low-cost, fast inference environment.[25] Therefore, distilled student models are widely used as optimization methods, especially for edge AI[26] for smartphones. The following comment was offered by an enterprise.

- It is now common for modern AI systems to use a combination of models of

[24] For example, DeepSeek, which is credited with developing low-cost, high-performance generative AI models, has publicly announced that it is incorporating distillation and MoE techniques to optimize computational efficiency. (https://github.com/deepseek-ai/DeepSeek-R1/blob/main/DeepSeek_R1.pdf) Such efficiency efforts are said to have the potential to reduce the relative reliance on GPUs and other computational resources.

[25] Dave Bergmann (IBM), "What is Knowledge Distillation?" (April 16, 2024) https://www.ibm.com/jp-ja/topics/knowledge-distillation

[26] Edge AI is a technology in which inference is processed on the device (terminal) side, and is said to offer advantages such as reduced communication costs, low latency processing, reduced privacy risks, and reduced network load.

various sizes. Smaller models can be derived from larger models through distillation and other techniques, and offer advantages such as faster processing and increased efficiency. On the other hand, large models can process and analyze more data and can handle more complex tasks. [Overseas model development enterprise]

(b) Mixture of Experts (MoE)

Mixture of Experts (MoE) is a method of demerging a single huge model into multiple subnetworks of "experts" and generate output by selecting the most appropriate expert for each input. This method reduces the computational costs associated with training and inference of the generative AI model, and achieves larger scale, lower power consumption, and higher speed. Through this mechanism, each expert is said to be able to specialize in a particular area or feature of the input, and the amount of computation is said to be reduced while maintaining high expressive power as a whole.[27] The following comments were offered by an enterprise and research institute.

- There is no strict definition of what exactly is a MoE, although it will be used more and more in other model development in the future. Currently, the Transformer architecture is the most commonly used, but that may change in the future, and other new methods may emerge. [Overseas model development enterprise]
- The trend that the amount of computational resources one possesses determines the outcome of the competition has eased, as model size has changed from being an absolute necessity to the emergence of high performing smaller models. In recent years, MoE, a method of bundling multiple models, has been introduced, so the amount of computational resources no longer directly influences model performance. [Domestic research institute]

(3) Summary

At present, there is active competition in the development of generative AI models, mainly among Big Tech companies, and domestic enterprises are trying to differentiate themselves from the general-purpose models of Big Tech companies by developing general-purpose models with high Japanese-language performance or specialized models for specific usage. At the same time, various development methods aimed at making models smaller, lighter, and more efficient are attracting attention, and it is possible that the conditions for competition may be renewed in the future.

[27] Dave Bergmann (IBM), "What is a Mix of Experts" (April 5, 2024), https://www.ibm.com/jp-ja/think/topics/mixture-of-experts

3 Application Layer

The application layer represents a market where generative AI products are developed and offered. Generative AI products have been developed for a variety of usage, including text, code, images, video, and voice/music, and are offered by a number of enterprises, including startups. In addition, Big Tech companies are providing services that make generative AI models available via their platforms.

(1) Summary of Previous Papers

- Generative AI models used by generative AI product development enterprises can be divided into three categories: (1) those using generative AI models provided as open source, (2) those using generative AI models provided as closed source, and (3) those using generative AI models developed for their own use.

- Generative AI products are widely deployed in industries and fields such as finance, construction, healthcare, and law. Big Tech companies and other enterprises are developing and providing generative AI products, for example, using cloud services provided by large digital platform operators.

- Generative AI products are emerging to be functionally integrated with existing digital services, such as search services and office productivity software. Big Tech companies may enhance the competitiveness of their existing digital services and further strengthen their position through such functional integration.

(2) Updates After Hearings, etc.

a. Application Layer Market Trends

The structure and status of the application layer are described in the previous paper. As shown in Figure 2 below, in the application layer, a wide variety of enterprises, from Big Tech companies to startups, are offering a variety of generative AI products, and there is active competition to meet diverse user demands. In Japan, a variety of generative AI application services, including chatbots for internal and external use, search systems, text generation, image generation, and voice generation, have been commercialized and are being used in the market. For example, according to OpenAI, the number of ChatGPT users

in Japan doubled year-on-year to 6 million in 2024.[28] On the other hand, there is also data[29] that shows that the utilization of generative AI by Japanese companies is low among major countries. The following comments were offered by enterprises.

- Companies large and small, both domestic and international, are competing to develop new and creative applications that utilize generative AI to learn languages, shop, edit and create photos, discover music, and make work more efficient. Given the endless potential for generative AI applications, it should come as no surprise that established and emerging companies from a variety of industries are entering and expanding in this space, developing new and innovative tools for developers with a variety of use cases and technological requirements. [Overseas model development enterprise]
- While many language models have emerged and there is a lot of competition in development, we have the impression that penetration of the user side within companies is slower in Japan than in other countries. This is due to the fact that domestic business users cannot tolerate mistakes by generative AI and demand 100% accuracy, and there is an impression that domestic businesses are wondering if the generative AI is really usable. Minor uses, such as taking minutes, have been progressing, but the use in actual business operations has not progressed yet. [Domestic model development enterprise]

[28] According to the Nihon Keizai Shimbun, “OpenAI 'ChatGPT' Users Reach 400 Million Worldwide, Up 30% in Two Months” (February 21, 2025, https://www.nikkei.com/article/DGXZQOGN20EDF0Q5A220C2000000/), ChatGPT's worldwide users reached 400 million in February 2025. Note that as of December 2024, the number of users was 300 million, and while it took three months for the number of users to reach 300 million from 200 million, the increase from 300 million to 400 million took just over two months, accelerating the speed at which the number of users increased.

[29] According to the Ministry of Internal Affairs and Communications' “Information and Communications White Paper for 2024” (July 2024, https://www.soumu.go.jp/johotsusintokei/whitepaper/ja/r06/html/nd151120.html#f00061), the status of policy formulation for the utilization of generative AI indicates that 42.7% of corporate respondents in Japan have established policies to utilize generative AI (including those with “a policy of active utilization” and those with “a policy of limited use in specific areas”), while approximately 80% of respondents in the United States, Germany, and China have established policies to utilize generative AI.

Figure 2: Types and Functions of Major Generative AI Services

Main Services	Main Uses
Language Generation AI	
ChatGPT (OpenAI) Gemini (Google) Copilot (Microsoft)	- Questioning, summarizing, calculating, paraphrasing, translating, knowledge discovery - Interactive sentence generation combined with search - Assistance in programming, etc.
Video generation AI	
StableDiffusion AdobeFirefly RunwayGen	- Image generation, partial image editing, automatic image coloring, line drawing extraction - Video generation, etc.
Sound Generation AI	
MusicGen Synthesizer V So-VITS-SVC	- Generation of music and sound effects - Singing voice generation - Voice conversion, voice language conversion, etc.
Other	
-	- Generation of 3D objects, generation of molecular structures, etc.

Source: Prepared by the JFTC based on various publicly available materials.

b. Integration of Functions with Existing Digital Services

As mentioned in the previous paper, at the application layer, a trend toward functional integration of generative AI products with existing digital services (search services, office productivity software, cloud services, SNS, digital advertising tools, etc.) is emerging. Big Tech companies have established strong positions in specific markets and sectors through the digital services they provide, and functional integration between generative AI products and existing digital services could increase the competitiveness of existing digital services and further strengthen their positions. Examples of functional integration are shown in Figure 3.

Figure 3: An example of functional integration with existing digital services

Company Name	Existing Services	Function Overview
Google	Google Search	The integration of AI Overviews (a feature that generates and displays answers in the form of an AI summary on the search results page) to Google Search.
	Android	Gemini on Android devices.
Microsoft	Office	Copilot to Office software.
	Windows PC	Copilot on Windows PCs.
Amazon	Shopping Apps	Rufus chatbot on Amazon's mobile shopping app.
Apple	iPhone	Apple Intelligence on iPhone devices and other devices.
X	X (SNS)	The "Grok" generative AI on the SNS platform.

Source: Prepared by the JFTC based on various publicly available materials.

c. AI Agents

In recent years, AI agents[30] have attracted attention as an emerging technology. It is now possible for AI agents to replace some or all of the routine flow tasks. Personal agents, in which AI is endowed with character or specific personality traits, are also being put to practical use. In Japan, AI agent services are being introduced[31] and efforts to improve business efficiency and create new business models are expanding. The following comments were offered by enterprises.

- There is a possibility that new markets will emerge as the current mainstream use of conversational AI chat evolves into agent-based solutions that are deeply integrated into companies' business processes. [Domestic enterprise]
- One feature of AI agents is that they can connect to and operate existing IT systems, such as human resources and accounting systems. Another feature is that they can determine what actions are necessary for a given task and operate on their own. Since compact models specialized for specific tasks are easier to use, we believe that the development of agents is a natural progression. [Overseas model development enterprise]

(3) Summary

The application layer is a particularly competitive area of the generative AI markets, with a

[30] In this paper, an AI Agent is an AI system that recognizes its environment and makes decisions autonomously to achieve a specific goal.

[31] For example, SoftBank and OpenAI have established SB OpenAI Japan, a joint venture to provide AI services for the Japanese market, and through this partnership they are developing customized AI agents that utilize data from each company (https://www.softbank.jp/corp/news/press/sbkk/2025/20250203_01/).

diverse range of players from Big Tech companies to startups. In addition, the increasing integration of functions between generative AI products and existing digital services may increase the influence of existing digital services, and the proliferation of AI agents may further intensify competition.

4 Other Generative AI Specific Issues and Considerations That Transcend Layers

(1) Summary of Previous Papers

● Cloud Services

・Most of the enterprises that develop generative AI do not have their own computational resources, but use cloud services provided by Big Tech companies. In addition, multiple domestic enterprises are engaged in cloud service business for enterprises developing generative AI models and generative AI products.

● Switching and migration of development environment, etc.

・Some pointed out that when switching to a different development environment for the generative AI model, there is hesitation in switching due to the cost of system restructuring, and that switching to cloud services may make the transition more difficult.

● Open source/closed source

・Open source technology can easily accelerate technological progress because anyone can check and improve the technical specifications, in addition to lowering the entry barriers for new enterprises. Closed source is easier for companies, etc. to control the use of the technology because technical specifications are not disclosed, and the risk of misuse is lower. From the perspective of competition policy, it is difficult to say which is preferable, but it is important to ensure a variety of options.

● Partnership

・In generative AI markets, partnerships among enterprises are active, especially between Big Tech companies and startups. While there are some indications that partnerships have the potential to increase competition, there are also indications that they may weaken competition.

(2) Updates After Hearings, etc.

a. Cloud Services

As mentioned in the previous paper, since the developers of generative AI models and generative AI products are mainly limited to Big Tech companies and other large enterprises

that have their own large computational resources, most enterprises are provided with GPUs and other high-performance chips from enterprises that offer cloud services for generative AI. Most operators receive servers equipped with high-performance chips such as GPUs and other resources from providers of cloud services for generative AI. Such service provision enables enterprises that develop generative AI models and generative AI products to train and infer generative AI models with high computational load.

There are indications that there is active competition in the cloud market, as the demand for GPU servers is increasing in line with the growing demand for generative AI, and each cloud enterprise is trying to differentiate its products. The following comments were offered by enterprises.

- With the growing demand for generative AI, more and more companies are competing to provide the necessary resources, such as computing power, human resources, and data. With respect to computing, AI development enterprises can choose from major cloud providers, specialized cloud services, and traditional hardware manufacturers. [Overseas cloud provider]
- We believe that competition will continue to be very fierce in the cloud sector. Competition is fierce from a variety of perspectives, including competition among cloud providers, competition in the area of providing apps to consumers, competition for model developers, and competition on platforms for third parties wishing to deploy models. The key element in such competition is differentiation, and each cloud provider needs to make its platform attractive enough to make model developers want to use its product. [Overseas cloud provider]

Against this backdrop, foreign cloud service enterprises, especially Big Tech companies, have been securing computational resources with their abundant financial resources and providing cloud services optimized for generative AI development, as shown in Figure 4.

As noted above, some pointed out that the generative AI cloud services business is experiencing strong competition, while others said that the barriers to entry are high because of the large investment required[32] for the data centers that operate them and the cooling systems that handle the enormous amount of heat generated[33]. The following comments were offered by enterprises.

[32]FTC, "Partnerships Between Cloud Service Providers and AI Developers," January 17, 2025, p. 10. https://www.ftc.gov/news-events/news/press-releases/2025/01/ftc-issues-staff-report-ai-partnerships-investments-study

[33] As for the Japanese market share of GPU cloud computing, it is estimated to be 55.0% for Amazon Web Services Japan, 18.7% for Microsoft Japan, 13.4% for Google Cloud Japan, and 1.0% for Sakura Internet, 12.0% for others as of 2023 (Fuji Chimera Research Institute, "2025 Generation AI/ AI Market Study that Leapfrogs with LLM" (November 2024), p. 188).

- The market for cloud services and development of generative AI models is dominated by foreign enterprises, which are ahead of the competition, and some believe that there are high barriers to entry. Continuous monitoring of the situation by the JFTC is desirable, as there are potential concerns that prices will remain high in the future. [Domestic enterprise]
- The cloud business is perceived as having no structural barriers to entry. However, due to the business model requiring significant upfront investment and prolonged periods of deficit, these factors act as entry barriers. Large corporations, with their financial resources, can enter the market if they commit to it. However, they tend to avoid entering if immediate profits are not anticipated. Many companies also shy away from businesses that require a long-term investment of time and effort. [Domestic enterprise]

Figure 4: Overview of Major Generative AI-Related Services of The Three Major Cloud Computing Companies

cloud	Major semiconductor chips	Major Generative AI-related Services
Amazon Web Services (AWS)	Trainium Inferentia NVIDIA H100, etc.	Amazon Q Amazon Bedrock Amazon SageMaker, etc.
Microsoft Azure	NVIDIA H100, etc.	Azure AI Foundry Azure Machine Learning Azure OpenAI Service, etc.
Google Cloud	TPU v5p/v5e NVIDIA H100, etc.	Vertex AI Cloud AutoML Gemini, etc.

Source: Prepared by the JFTC based on various publicly available materials.

Although there are indications that competition in the generative AI cloud services business is active, it is believed that competition is centered on leading foreign enterprises, as evidenced by the continued high market share of the three major cloud providers' GPU cloud services in Japan.

b. Switching/Migrating Development Environments

In the development of generative AI models, it has been pointed out that changes in the development environment, etc., such as changes in hardware or cloud environments, are not

easy because of the costs and man-hours involved in rebuilding the entire system, reimplementing software, migrating data, etc.

For example, if the generative AI model is designed to depend on a specific semiconductor chip, it may be technically difficult to switch to a chip made by another company. As described in 2. 1 (1) b. (a) above, especially for NVIDIA's development environment (e.g., CUDA), in principle, only NVIDIA GPUs are supported, and switching to chips made by other companies requires software redesign, etc., which is said to increase development costs and time. The following opinions were offered by enterprises.

- In the research and development of the model, it is technically possible to switch from an NVIDIA ecosystem to another company's ecosystem, but the switch will not occur if the performance of the GPU itself is comparable. There are factors such as cost, performance, stability in maintaining a leading position, and the performance of writing code to operate in the ecosystem, and while some have exceeded one of these in the short term, we do not believe that there is currently anything that exceeds them in overall performance. [Domestic model development enterprise]
- If hardware becomes more open source in the future, there is a possibility that NVIDIA-compliant hardware can be developed as open source. If specifications are standardized, hardware and development environments based on the same specifications will be available, even if hardware manufacturers and software vendors providing development environments differ. Users will then be able to freely select their hardware and development environment. However, the current situation has not progressed that far, and the development process is still dependent on NVIDIA, so we recognize that it will take time to reach the point of open source. [Domestic model development enterprise]

In addition, when switching from one cloud service to another in the development of generative AI models and products, data migration costs and other costs associated with the switchover will be incurred. Furthermore, if a certain semiconductor chip can be used only through a certain cloud service, the cost of changing or switching the specific semiconductor chip to be used will also be incurred, resulting in overall switching costs. The following comments were offered by enterprises.

- Switching cloud service providers requires significant time and resources, as all proprietary data stored in the cloud must be migrated to the new cloud service provider. [Overseas enterprise]
- While cloud switching itself is possible, the interface of the mechanism for managing billing, running machines, feeding data, and a series of other tasks is completely

different. It is not that difficult to move one server to another single server, but if multiple servers are linked together and a large number of GPUs are used, it takes a lot of time and effort to move to another service. [Domestic model development enterprise]

On the other hand, some commented that switching to other clouds is easy, and some providers pointed out that they are competing by offering services that simplify the switching process, although the switching costs are not small. The following comments were offered by enterprises.

- Our company also ensures that customers can easily export and transfer their data if they choose to switch to another cloud provider, allowing developers to move their models and data without being constrained by technical or commercial barriers. [Overseas cloud provider]
- The simple physical process of moving data always requires network equipment, electricity, fiber optic cables, and other connections, resulting in at the least a small amount of switching costs. However, cloud providers are competing by offering services that simplify IT provider switching. [Overseas cloud provider]

Although switching cloud providers is technically possible, and there will be at least small amounts of switching costs involved, it is difficult to determine at this point whether this will be a barrier or not, and this issue will continue to require close monitoring.

c. Open-source/Closed-source

As noted in the previous paper, there are two types of generative AI model offerings: open source and closed source. In the case of open source, it is expected to contribute to ensuring a competitive environment by lowering the entry barriers to the development and use of generative AI products, especially for new entrants and startup enterprises that generally have difficulty bearing high costs. In addition, since anyone can check and improve the technical specifications, it is said to facilitate verification of security issues and accelerate technological progress. However, since the technical specifications are open to the public, it is difficult for enterprises to maintain their competitive advantage, and since it is easy to use, there is a risk of being used for malicious purposes.

On the other hand, in the case of closed source, technical specifications are not disclosed, and companies and research institutions can independently manage and restrict the use of generative AI, which is said to be characterized by a lower risk of misuse by third parties. However, there are concerns that it is difficult to accept feedback and contributions from third

parties, which may stall innovation, and that the technology lacks transparency because the technical specifications are not disclosed.

In light of the above conditions, it is difficult to say whether open source or closed source is preferable from the standpoint of competition policy, but in any case, it is important to ensure a variety of options in the development of generative AI models and in the development and provision of generative AI products. The following comments were offered by enterprises and others.

- The AI technology market is still nascent and rapidly evolving, with new models and applications emerging frequently. This dynamism creates many possibilities as to how these models will integrate with applications and how business models will develop. Against this backdrop, both open-source and closed-source models are likely to interact and compete, fostering free and fair competition in the generative AI model and application markets. [Overseas model development enterprise]
- Although the performance advantage of closed source is sometimes pointed out, the performance of open source continues to follow the performance of closed source, so that, against the backdrop of the development of open source, open source has also become a price check on closed source. [Domestic enterprise]
- Open source allows smaller players to take advantage of opportunities for innovation without having to rely on the proprietary tools of the dominant player in the field of generative AI, and reduces monopolistic practices by reducing dependence on said player. [Foreign trade association]

d. Partnerships

As mentioned in the previous paper, in the generative AI markets, partnerships are actively being formed across layers, both domestically and internationally, including partnerships between semiconductor chip providers and generative AI model developers, between generative AI model developers, and between generative AI model developers and generative AI product developers. Partnerships are actively being formed across layers in Japan and overseas. In addition, looking at partnerships among development enterprises in the generative AI markets, as shown in Figure 5, there are multiple examples of partnerships through Big Tech companies that provide existing digital services and startups that develop generative AI models.

Figure 5: List of partnerships

Companies Related to Generative AI Services

Investor	OpenAI	Anthropic	Hugging Face	SakanaAI
NVIDIA	cooperation[34]		○ 2023/8	○ 2024/9
Microsoft	○ 2019/7		cooperation[35]	
Amazon		○ 2023/9	○ 2023/8	
Google		○ 2023/2	○ 2023/8	
Apple	cooperation[36]			
SoftBank Group	○ 2025/2			

A ○ indicates a partnership with equity investment.

Source: Prepared by the JFTC based on various published information as of March 31, 2025

As noted in the previous paper, such partnerships offer benefits to startups developing generative AI models, such as access to development capital through large investments, resources for development in generative AI markets, and access to existing related technologies. On the other hand, Big Tech companies may also benefit from licensing innovative generative AI models offered by startups through the partnership, thereby increasing the competitiveness of existing digital services offered by Big Tech companies. The following comments were offered by enterprises.

- Investments and partnerships are just one efficient way to bring together resources and

[34] On January 21, 2025, OpenAI announced the Stargate Project and revealed that NVIDIA would be a key initial technology partner. (https://openai.com/ja-JP/index/announcing-the-stargate-project/)

[35] On May 24, 2023, Microsoft and Hugging Face announced a partnership to launch the Hugging Face Model Catalog on Azure (https://huggingface.co/blog/hugging-face-endpoints-on-azure)

[36] On June 10, 2024, OpenAI announced a partnership with Apple to integrate ChatGPT into the Apple experience (https://openai.com/ja-JP/index/openai-and-apple-announce-partnership/).

technology to bring inventions to fruition more quickly and effectively than any one firm could accomplish alone. Partnerships can increase access and choice for customers to the foundation model, and can bring additional competition to the vibrant field of generative AI. [Overseas model development enterprise]

- We believe that partnerships in the field of generative AI need to be evaluated on the basis of their individual merits. We believe that partnerships allow companies access to essential resources for development and use, are often pro-competitive, and also promote innovation. [Overseas model development enterprise]

On the other hand, it was also pointed out that such partnerships may allow Big Tech companies and others with strong positions in the existing digital market to strengthen their own positions in the generative AI markets or weaken competition[37] [38]. The following comments were offered by a trade association and enterprise.

- While strengthening competitiveness through partnerships is desirable for the market, the convergence of a large number of AI technologies in the hands of a small number of companies requires social checks and balances to ensure that there is no excessive concentration in terms of training data, etc. [Domestic trade association]
- While partnerships among development enterprises usually help foster innovation when structured properly, partnerships that make innovative startups overly dependent on a single cloud provider and result in collaboration with a single provider can be detrimental to competition, as they reduce the cloud provider's own incentive to compete and risk preventing competing cloud providers from offering services to key customers, among other things. [Domestic enterprise]

3. Antimonopoly Act and Competition Policy Issues in Generative AI

In the previous paper, we presented possible Antimonopoly Act and competition policy issues based on the conditions of the generative AI market in order to contribute to future discussions on the maintenance and promotion of fair and free competition in the generative AI market in Japan. The following is a summary of the issues presented in the previous paper.

● Access restrictions and exclusion of competitors

- GPU and data required for developing generative AI are currently concentrated in the hands of Big Tech companies. In this situation, if access restrictions or exclusion of competitors are implemented, the potential for market entry may be lost, which may affect competition.

[37] FTC, "Partnerships Between Cloud Service Providers and AI Developers" (January 17, 2025), p. 37, https://www.ftc.gov/news-events/news/press-releases/2025/01/ftc-issues-staff-report-ai-partnerships-investments-study

[38] See Appendix for efforts by foreign competition authorities related to partnerships related to generative AI.

●Self-preferencing

・If an enterprise that provides a generative AI model develops a model in such a way that it treats its own goods and services more favorably than other goods and services in terms of inference results, this may affect competition with respect to such goods and services.

● Tying

・If an enterprise with a leading position in a certain service tying the use of its own generative AI model as a condition for providing that service, it may affect competition for the generative AI model.

●Parallel conduct using generative AI

・While price competition may increase due to price surveys by generative AI, the common use of underlying data and algorithms may lead to the same or similar pricing strategies, production targets, etc., which may affect competition.

● Acquiring specialized talent via partnerships

・Competition may be affected if the partnership is intended to enclose highly skilled professionals and the partnership creates an effect that is substantially similar to a business transfer.

Various opinions were received in response to these issues in the solicitation of information and opinions and in the hearings with enterprises, etc. Some opinions expressed concerns about competition, especially with regard to “access restrictions and exclusion of competitors” and “tying.”

Therefore, taking into account the conditions of the generative AI market mentioned in 2. above, the JFTC has summarized the antitrust and competition policy approach to "access restrictions/exclusion of competitors" and "tying" as follows.

1. Access Restrictions and Exclusion of Competitors

The development of generative AI models requires infrastructure such as computational resources (e.g., GPUs), data, and expert personnel. If certain enterprises that have established a strong position in these markets restrict access to, for example, the computational resources and data necessary for development to enterprises that do not have this infrastructure, competition may be affected, for example, by excluding other enterprises that compete with them or by denying them the potential for market entry.

As stated in 2.1(1) b. (a) above, there are opinions that it is difficult to secure semiconductor chips due to the supply and demand situation, but no opinions indicating specific actions to restrict access to semiconductor chips were received at this point in time. On the other hand, there

were opinions that there is no merit in restricting access in the semiconductor chip market, as described below by an enterprise.

- We have an incentive to offer our semiconductors to a wide range of companies. For example, if we provide our semiconductors to only a small number of companies and restrict access to our semiconductors from other companies, it would mean the end of our company if those few companies cut off business with us, so there is no advantage for us to restrict access in this way. The best scenario for us is a system that allows everyone to enjoy the benefits of access to semiconductors. [Overseas semiconductor enterprise]

There is also the possibility of access restrictions regarding mobile OSs. In February 2023, the JFTC published a report on markets for mobile OSs, etc.[39]. In the report, it is reported that a business enterprise in a leading position in the mobile OS market may restrict access to smartphone functions via its own mobile OS, which its own applications, goods, and services can access, by limiting API connections to other competing enterprises in the app market and its peripheral markets. By doing so, and thereby reducing the business opportunities of other competing enterprises or excluding these enterprises, there is a possibility of a problem under the Antimonopoly Act (private monopolization, unfair trade practice(s) (paragraph 14 of the General Designation (interference with a competitor's transactions)), etc.).[40]

Smartphones are a daily necessity for consumers and have become a major point of contact and entry point for accessing diverse digital content and services. The miniaturization of generative AI models and other factors have made it possible to use generative AI on smartphones,

---

[39] The Japan Fair Trade Commission, "Market Study Report on Mobile OS and Mobile App Distribution" (February 2023), https://www.jftc.go.jp/en/pressreleases/yearly-2023/February/230209.html

[40] The Japan Fair Trade Commission, "Market Study Report on Mobile OS and Mobile App Distribution." (February 9, 2023), pp. 135-136, which the concept is shown in the following table.
"In determining whether an act violates the Antimonopoly Act, it is necessary to comprehensively consider various factors. In evaluating these allegations, the reasonableness of the purpose and reasonableness of the means (e.g., whether there are other, less restrictive alternative means) should be taken into account.
(omitted)
If we examine the antitrust evaluation of claims from the perspective of ensuring security and protecting privacy, we can consider the following.
In the case of elimination of ineligible goods and services and ineligible enterprises from the viewpoint of safety, (omitted) the reasonableness of purpose is usually recognized if a certain offender conduct can be evaluated as having been undertaken for the purpose of ensuring consumer security and protecting the privacy of the consumer.
On the other hand, whether or not there is no other appropriate alternative to the conduct and whether or not it is a socially appropriate means to achieve the objective of ensuring competition and protecting the privacy of consumers requires careful consideration, taking into account various individual circumstances and whether or not there are other means that are less restrictive of competition. In order to determine the reasonableness of the means, which is to say whether or not it is a socially appropriate means, careful consideration is required, taking into account various individual circumstances and whether there are other means that are less restrictive of competition.
In judging the reasonableness of such measures, it may be necessary to receive detailed technical information from the parties to the act, compare and evaluate individual technologies, and determine whether technologies other than those currently in use can produce the same or better security and privacy protection effects, Quantitative and qualitative studies may be required to determine whether technologies other than those currently employed can provide the same or better security and privacy protection."

and as indicated in the section “Integration of Functions” (2.3(2)b above). Competition for generative AI models and generative AI products that function on smartphones is also intensifying. Therefore, any restrictive practices such as those described above could be problematic in the generative AI markets. The following comments were offered by an enterprise.

- The third-party developers of generative AI models need access to specific software on the mobile OS in order to run the models on the device. If third-party developers of generative AI models are unable to access such software, they will need to take individual measures, which will be very costly. There are many other concerns about the challenges of running AI models on devices. In the future, these restrictions will prevent users, device manufacturers, and AI model developers from having meaningful choices when running generative AI models on devices, and will prevent model developers from having non-discriminatory access to critical digital infrastructure, data, and other resources, which will put them at a competitive disadvantage. [Overseas enterprise]

In response to these comments, one enterprise offered the following comment

- We provide specific non-OS software to support our on-device generative AI model, and this software is distributed only to devices intended to provide our on-device generative AI model. However, this does not imply that it is difficult to run third-party on-device generative AI models. In fact, third-party on-device generative AI models can function on devices without specific software by using frameworks and other tools. This is evidenced by the fact that OEMs and app developers are releasing their own on-device models that can be integrated with terminals without depending on specific software. [Overseas model developer]

As mentioned in the 2. above, some pointed out that the generative AI markets will continue to be an oligopoly dominated by certain enterprises due to their technological and financial strengths. In general, when an enterprise that has established a strong position in the market for computing resources (GPUs, etc.), data, specialized human resources, etc., restricts access through the aforementioned acts, it becomes difficult for new entrants and existing competitors to secure alternative suppliers, which raises the costs of business activities and undermines the motivation to enter the market or develop new products, etc. When there is a risk of creating a situation in which new entrants or existing competitors are excluded or their business opportunities are reduced (market foreclosure effect), this may become an issue under the Antimonopoly Act (private monopolization, unfair trade practice(s) (paragraph 14 of the General Designation (Interference with a competitor's transactions)), etc.).

2. Tying

In general, providing services to counterparties by combining multiple functions to add new value is a method of technological innovation(s) and sales promotion, and such an act itself is not immediately problematic under the Antimonopoly Act.

However, if a particular digital service provider integrates a generative AI model into its digital service and provides it to users, it may lead to conditions of the competitors in the market for generative AI models that impede their business activities or raise barriers to entry, depending on their position in the market for such digital services, etc. The following comments were offered by enterprises.

- Company A, a competitor with a leading position in the digital services market, integrates its generative AI products into its existing digital service offerings and sells them as a bundle. Since these digital services already hold a dominant position among business customers, such tying practices are expected to give Company A a significant distribution advantage in the generative AI services market, making it difficult for competing generative AI providers to enter or expand. [Overseas model development enterprise]
- We intend to sell generative AI products in the future, but some companies that are already selling generative AI products are offering them by integrating their existing digital service products with generative AI products. This practice may make it difficult for users of such AI products to easily adopt a latecomer's AI product, partly due to the linkage with applications and security issues, and we see this as one of our future challenges. [Domestic model development enterprise]

In response to these comments, other enterprises offered the following comments.

- It is quite natural to offer one's own products on one's own services, and because of the development and maintenance costs involved in adapting products to other platforms and services, tying one's own products to those of other companies is inevitable, and it seems important that other options exist on those services and platforms. It is important to have another option on the service or platform. [Domestic model development enterprise]
- The generative AI function integrated into our company's existing in-house SaaS service is merely an enhancement to improve the functionality of that service, and is not a separate product. Therefore, we believe that it cannot be considered tying-in sales in the first place. In addition, all major commercial software and SaaS enterprises are integrating generative AI functionality into their solutions. Given the existence of many leading competing generative AI developers and SaaS solutions in general, we believe that such integration is an extremely pro-competitive offender as it meets consumer demand for cutting-edge solutions. [Overseas model development enterprise]

- The continuous enhancement of the Generative AI model is not only enabling entirely new use cases and new classes of apps to better serve customers, but is also bringing a new dimension of competition to existing digital services. [Overseas model development enterprise]

In general, if a generative AI model provider has a strong position in a specific digital service market, integrating generative AI models into that digital service and offering it to users as a new digital service may make it difficult for other generative AI model providers or new entrants seeking to offer generative AI models to secure users as customers, thereby raising the costs of business activities and discouraging new entrants and the development of new goods. When there is a risk that existing competitors or new entrants will be excluded or that their business opportunities will be reduced (market foreclosure effect), this may become an issue under the Antimonopoly Act (private monopolization, unfair trade practice(s) (paragraph 10 of the General Designation (tie-in sales, etc.))).

3. Other Issues

Regarding other issues, the following comments were received, although no comments were received at this time indicating actions such as those exemplified in the previous paper. The JFTC will continue to monitor market trends with regard to the other issues.

(1) Self-Preferencing

- Since clearly identifiable self-preferencing can be easily detected, the possibility of controlling/tuning performance in such a way as to intentionally give preferential treatment is low. [Domestic model development enterprise]
- The background is that the development of generative AI models optimized for integration with the company's own services, rather than merely as a preferential measure, and as a result, we believe that the situation is such that integration costs can be easily optimized compared to models developed by other companies. [Domestic enterprise]

(2) Parallel Conduct Using Generative AI

- In reality, it is unlikely that pricing strategies, etc. will depend solely on the responses of the generative AI, so concerns about parallel actions using the generative AI are somewhat unrealistic. [Domestic model development enterprise]
- Even if similar results are shown by AI, they will not lead to the same conclusion in many cases because of differences in the criteria for judging the results and culture of each company. In addition, by learning information about individual companies, they may act to become more strategically competitive by aiming to use services that produce

results specific to individual companies, even if they use the same model. [Domestic trade association]

(3) Acquiring Specialized Talent Via Partnerships

- Our company has not observed or recognized any such events as described, at least in Japan. [Domestic enterprise]
- We believe that this is not limited to the generative AI market. We have heard rumors of team-by-team transfers and enclosures as urban legends in the industry, but have never been able to confirm any cases. [Domestic model development enterprise]

4. The Japan Fair Trade Commission's Next Steps

As mentioned in the previous paper, the size of the generative AI market in Japan is currently 118.8 billion yen (in 2023), and the market itself can be said to be in its infancy, but it is expected to grow rapidly with an average annual growth rate of 47.2%, reaching 1.7774 trillion yen by 2030[41]. The market is expected to further expand and grow. Against this backdrop, in Japan, discussions at the AI System Study Group,[42] including hearings from researchers and enterprises, were published as an "Interim Summary" on February 4, 2025, after public comments. Based on this, the Cabinet approved the "Draft Law on Promotion of Research, Development and Application of Artificial Intelligence-related Technology" on the 28th of the same month, and submitted it to the 217th ordinary session of the Diet, leading to its enactment at the House of Councillors plenary session on the 28th of May 2025. In addition, several competition authorities in various foreign countries and regions have also published reports on the actual situation, as shown in the Appendix. Thus, with the growth of generative AI markets, efforts to actively address issues related to generative AI are being made both domestically and internationally.

In order to present the current status of generative AI markets, while based on the previous paper, this report was reorganized in more detail based on the results of the hearings, etc., to reflect changes in the situation from October 2024, and to delve deeper into the facts. In the 3. above, the JFTC has summarized the issues under the Antimonopoly Act and competition policy, and clarified the concept under the Antimonopoly Act in the generative AI markets. The JFTC hopes that these ideas will help prevent the negative effects that are of concern in the generative AI markets and further promote fair and free competition. In addition, the JFTC will continue to take rigorous and appropriate measures when it comes in contact with specific cases that raise issues under the

[41] Japan Electronics and Information Technology Industries Association, "Survey on Trends in Focus Areas 2023," 2023, p. 1.
[42] The "AI Strategy Council" was established under the "Integrated Innovation Strategy Promotion Council" chaired by the Chief Cabinet Secretary and composed of all cabinet ministers. The "AI System Study Group" was established under it.

Antimonopoly Act, including conduct that is problematic under the Antimonopoly Act as pointed out in this report, as well as conduct that is listed as an issue in 3. above but not pointed out as a specific problem or issue.

With regard to the negative effects that may be of concern in the generative AI markets, we will actively work in close coordination and cooperation with relevant ministries and agencies with the aim of ensuring a fair competition environment, as necessary. In addition, since some of the competitive entities in generative AI markets operate on a global scale, competition authorities in various countries and regions overseas are also concerned and interested in the business activities of such enterprises. The JFTC will continue to exchange views with competition authorities in various countries and regions at various levels, and will also make use of forums such as the International Competition Network (ICN) and the Organization for Economic Cooperation and Development (OECD) to continuously cooperate with relevant overseas authorities in order to improve the competition environment.

As stated in the introduction, this report is not the final report of this market study. The JFTC will maintain closely monitoring generative AI markets. Therefore, the JFTC will continue to solicit a wide range of information and opinions by opening a new form on the JFTC's website to solicit such comments.

In the generative AI markets, changes and technological innovations that could affect the market structure are occurring every day, and the market environment is likely to change drastically in the future, so it is extremely important to properly understand the market trends. The JFTC will continue its market study in the hope that the generative AI market will continue to develop under fair and free competition.

[Appendix] Foreign Competition Authorities' Work in The Generative AI Space

1 European Union

On January 9, 2024, the European Commission invited public comments on competition in the field of virtual worlds and generative AI (deadline: March 11, 2024) to identify potential competition issues that may arise in the technology sector and to conduct forward-looking analysis of technology and market trends[43]. The public comments were organized into 12 questions to identify potential competition issues that may arise in these technology sectors, and approximately 120 comments were submitted in response to this call.

On June 28 of the same year, a workshop was held on "Competition in Virtual Worlds and Generative AI" as a follow-up to the call for comments[44]. The workshop also clarified that the partnership between Microsoft and OpenAI had been reviewed under the EU Merger(s) Regulation, but that the review was now closed[45].

On September 19 of the same year, based on the comments submitted in the Call for Comments, findings from the workshop, and research conducted by the Commission, a policy brief on competition in generative AI and virtual worlds was published[46]. The policy brief notes trends in the generative AI market, including increased vertical integration and partnerships, smaller and more efficient models, and the existence of open- and closed-source models. The policy brief points to trends in the generative AI market, such as increased vertical integration and partnerships, downsizing and increased efficiency of models, and the existence of open source and closed source models, while also pointing out competition concerns, such as limited access to key resources, exclusionary conduct leveraging market dominance, and network effects. The report points out the committee will continue to address potential issues in the area of generative AI in order to achieve a fair competitive environment.

2 United Kingdom

The Competition and Markets Authority (CMA) conducted a study of AI foundation models from the perspective of competition and consumer protection policies. It published a survey report on generative AI foundation models on September 18, 2023[47], and an updated version of the report was published on April 13 of the following year[48], which indicated concerns in terms of fairness, efficiency, and openness of competition.

---

43 https://ec.europa.eu/commission/presscorner/detail/en/ip_24_85
44 https://competition-policy.ec.europa.eu/about/reaching-out/virtual-worlds-and-generative-ai_en
45 https://ec.europa.eu/commission/presscorner/detail/en/speech_24_3550
46 https://digital-strategy.ec.europa.eu/en/news/commission-publishes-policy-brief-competition-generative-ai-and-virtual-worlds
47 https://www.gov.uk/government/publications/ai-foundation-models-initial-report
48 https://www.gov.uk/government/publications/ai-foundation-models-update-paper

In addition, the CMA was reviewing business combinations, including partnerships between several Big Tech companies and generative AI startups, and it was decided not to move to the phase 2 review for all of the partnerships.

- Microsoft and Mistral AI (review completed on May 17, 2024)[49]
- Microsoft and Inflection AI (review completed on September 4 of the same year)[50]
- Amazon and Anthropic (review completed on September 27 of the same year)[51]
- Google and Anthropic (review completed on November 19 of the same year)[52]
- Microsoft and OpenAI (review completed on March 5, 2025)[53]

3 United States

Technologists from the Federal Trade Commission's (FTC) Office of Technology have been discussing competition issues related to technology on their blog[54]. On June 29, 2023, a blog post titled "Generative AI Raises Competition Concerns" explained that competition could be harmed if a critical component of generative AI delivery is controlled by a single operator[55]. The blog also shared key insights from the panels at the AI Tech Summit[56] held by the FTC on January 25, 2024[57].

In January 2024, FTC issued an order to five companies - Alphabet, Amazon, Anthropic, Microsoft, and OpenAI - to provide information on recent investments and partnerships involving generative AI companies and major cloud service providers under Section 6 of the FTC Act[58]. The order was issued to companies involved in three business partnerships (Microsoft-OpenAI, Amazon-Anthropic, and Google-Anthropic) including billions of US dollars of investments, to better understand how partnerships and investments formed between AI developers and cloud service providers impact competition.

On January 17, 2025, the FTC released a staff report on "AI Partnerships and Investments,"

---

[49] https://www.gov.uk/cma-cases/microsoft-slash-mistral-ai-partnership-merger-inquiry
[50] https://www.gov.uk/cma-cases/microsoft-slash-inflection-ai-inquiry#invitation-to-comment-closed
[51] https://www.gov.uk/cma-cases/amazon-slash-anthropic-partnership-merger-inquiry
[52] https://www.gov.uk/cma-cases/alphabet-inc-google-llc-slash-anthropic-merger-inquiry
[53] https://www.gov.uk/cma-cases/microsoft-slash-openai-partnership-merger-inquiry
[54] https://www.ftc.gov/policy/advocacy-research/tech-at-ftc
[55] https://www.ftc.gov/policy/advocacy-research/tech-at-ftc/2023/06/generative-ai-raises-competition-concerns
[56] https://www.ftc.gov/news-events/events/2024/01/ftc-tech-summit
The FTC's AI Tech Summit brought together people from various positions in academia, industry, civil society, and government agencies for three panels focusing on different layers of AI: hardware and infrastructure, data and models, and consumer applications.
[57] "Semiconductor Chips and Cloud Computing" (March 14, 2024):
https://www.ftc.gov/policy/advocacy-research/tech-at-ftc/2024/03/semiconductor-chips-cloud-computing-quote-book
"Data and Models" (April 17, 2024):
https://www.ftc.gov/policy/advocacy-research/tech-at-ftc/2024/04/data-models-quote-book-tech-summit-ai
"Applications for Consumers" (Apr. 24, 2024):
https://www.ftc.gov/policy/advocacy-research/tech-at-ftc/2024/04/consumer-facing-applications-quote-book-tech-summit-ai
[58] https://www.ftc.gov/news-events/news/press-releases/2024/01/ftc-launches-inquiry-generative-ai-investments-partnerships

based on information obtained through the above order and other sources[59]. The report notes that partnerships between large cloud service providers (Alphabet, Amazon, and Microsoft) and generative AI developers (Anthropic and OpenAI) have resulted in competition concerns regarding limited and preferential access to inputs, higher switching costs, and access to sensitive information[60].

4 European Union, United Kingdom, United States

On July 24, 2024, the European Commission, CMA, DOJ (Antitrust Division of the United States Department of Justice), and FTC jointly issued a declaration on competition for generative AI foundation models and AI products[61]. The main points are as follows:

- While law enforcement is subject to the independent decision-making of national competition authorities, AI-related risks are often cross-border. Therefore, it is important to share an understanding of these issues and, when necessary, to exercise their respective powers collaboratively.
- Threats to competition in the AI market include the risk that a few companies might exploit bottlenecks in existing or emerging AI-related fields, allowing them to exert excessive influence over the future development of specialized chips, massive computing power, and other critical resources. Additionally, there is a concern that the dominant positions established by large companies through past technological transformations could become entrenched, hindering future competition.
- To promote competition and foster innovation in the AI ecosystem, it is essential to ensure fair trade practices, interoperability, and other supportive measures.

5 France

On February 8, 2024, the French Competition Commission issued a call for comments on

---

[59] https://www.ftc.gov/news-events/news/press-releases/2025/01/ftc-issues-staff-report-ai-partnerships-investments-study

[60] Note that while Chairman Andrew N. Ferguson (who was a commissioner at the time of the report's release) and Commissioner Melissa Holyoake expressed approval of the report's release itself, they also expressed the opinion that Chapter 5 (Areas to Watch for Potential Impacts of AI Partnerships) should either be skipped or read with a skeptical eye, etc. The report should be skipped or read with a skeptical eye.
https://www.ftc.gov/legal-library/browse/cases-proceedings/public-statements/concurring-dissenting-statement-commissioner-andrew-n -ferguson-joined-commissioner-melissa-holyoak

[61] https://competition-policy.ec.europa.eu/about/news/joint-statement-competition-generative-ai-foundation-models-and-ai-products-2024-07-23_en
https://www.gov.uk/government/publications/joint-statement-on-competition-in-generative-ai-foundation-models-and-ai-products/joint-statement-on-competition-in-generative-ai-foundation-models-and-ai-products
https://www.justice.gov/opa/pr/leaders-justice-department-federal-trade-commission-european-commission-and-uk-competition
https://www.ftc.gov/news-events/news/press-releases/2024/07/ftc-doj-international-enforcers-issue-joint-statement-ai-competition-issues

generative AI, with a deadline of March 22, 2024[62]. The French authority sought information on the strategies of major digital companies in the generative AI market, focusing on practices in the cloud services market and issues related to access to cloud services, data, and expert talent necessary for developing foundation models. Following this call for comments, a report was published on June 28, 2024[63] that highlighted high barriers to entry in the generative AI market and the increasing advantages generative AI provides to businesses in other digital areas. It also outlined competition concerns upstream in the value chain and made recommendations to address these issues.

On September 27, 2023, the French Competition Commission conducted a court-authorized inspection of a company suspected of carrying out anti-competitive practices in the graphic cards sector[64].

6 Canada

On March 20, 2024, the Competition Bureau Canada launched a public consultation on competition in generative AI with a discussion paper[65], inviting comments until July 7, 2024. The goal was to better understand how competition is developing in generative AI markets, how to safeguard and promote it, and how to tackle potential issues. After receiving 28 submissions[66], the Bureau published a report on January 27, 2025[67]. This report highlighted concerns that significant investments by Big Tech and other firms could stifle competition and innovation. However, it also acknowledged that it's currently unclear whether existing antitrust laws can effectively address potential anti-competitive practices fostered by AI. The Bureau stated its commitment to ongoing engagement with stakeholders to gain a deeper understanding of emerging competition challenges related to AI.

On September 16, 2024, the Canadian Competition Summit 2024[68] titled "Market Dynamics in the AI Era" was held, and on November 7, 2024, a report summarizing the Summit was released[69].

---

62 https://www.autoritedelaconcurrence.fr/en/press-release/generative-artificial-intelligence-autorite-starts-inquiries-ex-officio-and-launches

63 https://www.autoritedelaconcurrence.fr/en/press-release/generative-artificial-intelligence-autorite-issues-its-opinion-competitive?referrer=content_seehereview

64 https://www.autoritedelaconcurrence.fr/en/press-release/general-rapporteur-autorite-de-la-concurrence-indicates-unannounced-inspection-was
The French Competition Commissioner has announced that it will conduct a review of the case, but has not disclosed the recipient of the review.

65 https://www.canada.ca/en/competition-bureau/news/2024/03/competition-bureau-seeks-feedback-on-artificial-intelligence-and-competition.html

66 https://competition-bureau.canada.ca/en/how-we-foster-competition/consultations/written-responses-consultation-artificial-intelligence-and-competition-discussion-paper

67 https://www.canada.ca/en/competition-bureau/news/2025/01/competition-bureau-issues-report-summarizing-feedback-on-artificial-intelligence-and-competition.html

68 https://competition-bureau.canada.ca/en/how-we-foster-competition/education-and-outreach/canadas-competition-summit

69 https://competition-bureau.canada.ca/en/how-we-foster-competition/education-and-outreach/report-summit-2024-competition-ai-era

7 South Korea

On August 1, 2024, the Korea Fair Trade Commission (KFTC) announced that it would launch a fact-finding survey of major domestic and foreign operators in the AI sector to proactively understand potential competition and consumer rights issues[70]. On December 17 of the same year, the report was published[71]. The report is significant in that it provides an overview of the rapidly changing conditions of the competition in the domestic generative AI market across the entire value chain, and it also provides an overview of the dynamic market for generative AI, which is not only actively competitive, but is also rapidly changing and developing through technological innovations, etc. Given concerns arising from the emergence of firms with market power in specific segments, the KFTC stated its intent to continue monitoring the competitive landscape of the generative AI market and to conduct additional market analysis if necessary.

8 G7 Competition Summit

On October 3 and 4, 2024, the G7 Competition Authorities and Policymakers' Summit was held in Rome, Italy, hosted by the Italian Competition and Market Protection Commission and attended by the heads of the G7 competition authorities and policy-making bureaus, including the JFTC.

The Summit discussed topics such as competition concerns in the AI value chain (key AI inputs and partnerships), competition concerns in AI (downstream and adjacent markets for AI and algorithmic collusion), and regulatory and policy approaches in AI markets. The "Digital Competition Communiqué" was adopted as an outcome document. The "Digital Competition Communiqué" outlines the following ideas[72].

- ○ Creation of innovation by AI
- ○ Competition concerns within the AI
- ○ Guiding principles for ensuring contestability and fostering innovation
- ○ Role of competition authorities (vigorous antitrust enforcement, enhancement of international cooperation, etc.)

End

---

70 https://www.ftc.go.kr/www/selectBbsNttView.do?pageUnit=10&pageIndex=32&searchCnd=all&key=12&bordCd=3&searchCtgry=01,02&nttSn=43530

71 https://www.ftc.go.kr/www/selectBbsNttView.do?pageUnit=10&pageIndex=14&searchCnd=all&key=12&bordCd=3&searchCtgry=01,02&nttSn=43730

72 https://www.jftc.go.jp/houdou/pressrelease/2024/oct/241007_G7_result.html