Organisation for Economic Co-operation and Development

DAF/COMP/WD(2025)53

Unclassified

English - Or. English

14 November 2025

DIRECTORATE FOR FINANCIAL AND ENTERPRISE AFFAIRS
COMPETITION COMMITTEE

Competition in Artificial Intelligence Infrastructure – Note by Japan

5 December 2025

This document reproduces a written contribution from Japan submitted for Item 7 of the 147th OECD Competition Committee meeting on 4-5 December 2025.

Antonio CAPOBIANCO
Antonio.Capobianco@oecd.org, +(33-1) 45 24 98 08

JT03576762

Japan

1. Overview

1. Generative AI, which is rapidly evolving, is expected to offer numerous benefits to the economy and society, such as enhancing the productivity of enterprises and providing a range of services. It is also anticipated to develop further, as it has the potential to spur new innovations.

2. On the other hand, various risks associated with generative AI have been highlighted, including the risk of disrupting society through the infringement of intellectual property rights (such as copyright) and the proliferation of false or misleading information, as well as security risks such as its potential use in cyberattacks. From a competition policy perspective, there are also potential risks, such as limited access to computing resources (like GPUs) necessary for generative AI development. In light of this, the Japan Fair Trade Commission (JFTC) launched a market study on generative AI to understand the market conditions in Japan, from the viewpoint of maintaining fair and free competition, ensuring the sustainable development of the technology for further innovation, and implementing generative AI in a sound manner for the social economy.

3. In October 2024, the JFTC published a discussion paper titled "Generative AI and Competition (Discussion Paper)" to provide an overview of generative AI and its market structure. At the same time, the JFTC invited stakeholders to provide information and comments to contribute to the maintenance and promotion of fair and free competition in Japan’s generative AI sector.

4. As a result, the JFTC received 712 comments and opinions. The JFTC also conducted interviews with roughly 50 stakeholders, including domestic and foreign enterprises, to gather a wide array of views and up to date information. The JFTC analyzed trends in those information and opinions, carefully selected major themes, and compiled "Report Regarding Generative AI ver.1.0 (Report)" in June 2025, in a more targeted form than the Discussion Paper.

2. Structure of Generative AI Market

5. A 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 are being provided. Similar to other digital sectors, the structure of the generative AI market is multi layered with both domestic and foreign enterprises engaged in economic activities in each layer.

6. The JFTC examined the structure of the generative AI sector, and organized it into the following three layers.

1. Infrastructure Layer: This market acts as the foundation of generative AI. The three main elements of this layer are computing resources, data, and expertise (specialized talent).
2. Model Layer: This market focuses on developing generative AI models, which are pre-trained using vast amounts of data, often leveraging cloud technologies.

3. Application Layer: This market focuses on developing and offering generative AI products.

7. This contribution paper focuses on the infrastructure layer and provides its overview based on the Discussion Paper and the Report.

2.1. Computing Resources[1]

2.1.1. Summary of the Discussion Paper

8. To develop high-performance generative AI models, it is crucial to invest in a sufficient quantity of semiconductor chips. There are various types of semiconductor chips, with GPUs being particularly suitable for developing generative AI models. Originally designed for image processing, GPUs have high parallel processing capabilities, significantly reducing the training time for the development of these models.

9. A certain manufacturer holds approximately 80% of the global GPU market share.[2] This dominance is attributed to several factors: the high performance of the manufacturer's GPUs (fast computation), the developer-friendly environment that facilitates parallel computation, and the availability of supporting software. The manufacturer continues to enhance these technologies by investing substantial resources into their development.

10. However, the global supply of GPUs falls short of demand, leading to intense competition to acquire them among companies, including Big Tech firms. In response, many companies worldwide are developing alternative semiconductor chips. Despite these efforts, the gap between the manufacturer and other companies remains significant, with some experts suggesting it is extremely challenging to bridge this gap. The situation in Japan mirrors the global landscape, and Japanese domestic companies are striving to innovate in terms of power efficiency and pricing as they develop their own semiconductor chips.

2.1.2. Updates in the Report

GPUs

11. In the GPU market, the manufacturer holds approximately 80% of the global market share, and its products continue to enjoy a high market share. One reason why the manufacturer holds high market share is that its development environment is said to offer high usability and scalability because deep learning frameworks[3] are highly optimized for its GPUs. In addition, only the manufacturer's GPUs are generally supported for the manufacturer's development environment.

12. On the other hand, regarding the supply and demand situation for the manufacturer's GPUs, it used to be difficult to secure, but the situation has changed due to the increased supply.

[1] Computing resources are essential for the development and utilization of generative AI models, with semiconductor chips like GPUs playing a particularly crucial role. Therefore, this contribution paper focuses on semiconductor chips.

[2] Bruegel, "Competition in Generative Artificial Intelligence Foundation Models", June 18, 2023, p. 7.

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

Training Phase[4] *and Inference Phase*[5]

13. 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.

14. In the training phase, 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 them, the manufacturer's GPUs are market leaders due to their good development environment and other features.

15. On the other hand, 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 the manufacturer's semiconductor chips, semiconductor chips developed by Big Tech companies, existing semiconductor manufacturers, and startups are also attracting attention in the inference phase. Therefore, the market for inference phase is witnessing heightened competition compared to the training phase.

2.2. Data

2.2.1. Summary of the Discussion Paper

16. In developing a generative AI model, it is essential to pre-train the model using training data to understand the structure and meaning of language. This is followed by fine-tuning with additional data to tailor the model to specific tasks, industries, or applications. A large amount of training data is required in the development of generative AI models. However, it has been noted that Japanese domestic enterprises, in particular, are cautious about acquiring training data to avoid potential copyright and other legal issues.

17. Among generative AI products utilizing large language models, Japanese data is crucial for those used in Japan. However, as mentioned in above, a large amount of task-specific data is required. It has been noted that the number of Japanese users is smaller than English users, resulting in less Japanese data available on the Internet. Consequently, the amount of data available for training in Japanese is relatively limited. In this context, Japanese domestic enterprises have the potential to develop superior language models specialized for Japanese, compared to those developed by Big Tech companies and leading startups.

2.2.2. Updates in the Report

18. 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. However, much of the accessible high quality data on the Internet is already being used to train generative AIs, and the open data

[4] The training phase refers to the phase 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.

[5] The inference phase refers to the phase where a trained model analyzes new input data, and generates optimal responses and content based on patterns and knowledge acquired during the training phase.

available for training may be depleted in the future.[6] Under these circumstances, synthetic data[7] 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.

19. 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.

20. Regarding the improvement of Japanese language performance of the generative AI models, a certain amount of Japanese language data is considered necessary, but it is difficult to determine how much it is needed to train a high-performance Japanese-language model and how the availability of the data will affect the market position of the enterprise developing the model.

2.3. Expertise (specialized talent)

2.3.1. Summary of the Discussion Paper

21. The development of generative AI models requires highly skilled researchers and engineers. However, it has been noted that there is a limited pool of highly specialized professionals capable of developing GPUs and generative AI models. Consequently, hiring these experts is challenging, and their scarcity is becoming a bottleneck in the development of generative AI models. In this context, some have pointed out that Japanese domestic enterprises face significant difficulties in hiring these limited specialized professionals. This is primarily because Big Tech companies, with their abundant financial resources, are more capable of attracting and retaining such talent, leading to a concentration of these resources within Big Tech companies.

2.3.2. Updates in the Report

22. As stated in item A above, the demand for specialized talent in the development of generative AI is significantly increasing, and Big Tech companies are noted to have an advantage in the competition for highly specialized human resources. 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.

23. Challenges faced by Japanese 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, Japanese domestic enterprises have strengths such as the ease of training and hiring local human resources were also pointed out.

[6] *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

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

2.4. Other Issues and Considerations that Transcend Layers

2.4.1. Cloud Services

Summary of the Discussion Paper

24. Among developers of generative AI models and products, only a few, such as Big Tech companies and leading startups, possess their own computing resources. Most developers lack these resources and therefore rely on specialized cloud services from cloud service providers.

25. The use of cloud services in developing generative AI models and products is expanding, and the spread of AI-related cloud services is driving the growth of the cloud service market. It is anticipated that competition in this market will remain intense, primarily among Big Tech companies.

Updates in the Report

26. 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. Against this backdrop, foreign cloud service enterprises, especially Big Tech companies, have been securing computing resources with their abundant financial resources and providing cloud services optimized for generative AI development. 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.

2.4.2. Switching/Migrating Development Environments

Summary of the Discussion Paper

27. It has been noted that switching a development environment built with specific semiconductor chips and dedicated software to another environment can incur significant costs, such as system reconstruction. This may cause developers to hesitate to switch. Furthermore, when using cloud services for the development and deployment of generative AI models and products, it can be challenging to switch to other cloud services or migrate to an on-premise[8] environment. Moreover, it has been noted that when users, such as generative AI product developers, customize generative AI models, they face higher switching costs to transfer trained components. This tendency for switching costs to arise at each layer makes the entire generative AI market structurally prone to lock-in.

Updates in the Report

28. For example, if a generative AI model is designed to depend on a specific semiconductor chip model, it may be technically difficult to switch to a competitor’s chip model. In particular, because the manufacturer's development environment primarily supports only its GPUs, switching to a competitor’s chips would require software redesign, etc., which is said to increase development costs and time.

29. 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

[8] When the company owns and operates its own infrastructure (servers, network equipment, software, etc.) necessary to operate its IT systems.

the switchover will be incurred. Furthermore, if a certain semiconductor chip model can be used only through a certain cloud service, the cost of changing or switching the specific semiconductor chip model to be used will also be incurred, resulting in overall switching costs. 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.

3. Issues on the Antimonopoly Act and Competition Policy in Generative AI

30. The Discussion Paper outlines, as summarized below, the potential issues under the Antimonopoly Act and competition policy, taking into account the current state of the generative AI market.

- Access restrictions and exclusion of competitors

GPUs 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.

- Self-preferencing

If an enterprise that provides a generative AI model develops a model, treating its own goods and services more favorably than other's goods and services in terms of inference results, it may affect competition with respect to the goods and services.

- Tying

If an enterprise with a leading position in a certain service ties the use of its own generative AI model as a condition for providing the service, it may affect competition with respect to 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

If a partnership is intended to enclose highly skilled professionals and the partnership creates an effect that is substantially similar to a business transfer, it may affect competition.

31. Various opinions were received in response to the hearings with enterprises, etc., conducted after the Discussion Paper was published. Some competitive concerns were expressed, especially with regard to "access restrictions and exclusion of competitors" and "tying." Therefore, the JFTC has compiled its views on such conduct under the Antimonopoly Act and competition policy as follows.

3.1. Access Restrictions and Exclusion of Competitors

32. There are opinions that it is difficult to secure semiconductor chips due to the supply and demand situation, but the JFTC received no opinions indicating specific actions to restrict access to semiconductor chips at this point. On the other hand, there were opinions that there is no merit in restricting access in the semiconductor chip market.

33. Some pointed out that the generative AI market 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, data, specialized human resources, etc., restricts access through the aforementioned acts, it may become 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. Through such actions, 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 be problematic under the Antimonopoly Act (Private Monopolization, Interference with a Competitor's Transactions (Unfair Trade Practice), etc.)

3.2. Tying

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

35. However, if a particular digital service provider integrates a generative AI model into its digital service and provides it to users, depending on the position in the market for such digital services, it may lead to a situation which impedes business activities of competitors in the market of generative AI models or raises barriers to entry.

36. In general, such act may make it difficult for other generative AI model providers or new entrants seeking to offer generative AI models to secure 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 be problematic under the Antimonopoly Act (Private Monopolization, Tie-in Sales (Unfair Trade Practice), etc.)

4. The JFTC's Next Steps

37. 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 1.7 trillion yen by 2030. [9]

38. The JFTC hopes that the views presented in the Report will help prevent the negative effects concerned and further promote fair and free competition in the generative AI market. In addition, the JFTC will continue to take rigorous and appropriate measures against problematic acts under the Antimonopoly Act, when it encounters specific cases, including those involving conduct not explicitly pointed out in the Report.

39. With regard to the negative effects concerned in the generative AI market, the JFTC will actively work for ensuring a fair competition environment, in close coordination and cooperation with relevant ministries and agencies as necessary. In addition, since some of the players in the generative AI market operate their business globally, the JFTC will continuously cooperate with overseas competition authorities to improve the competition

[9] Japan Electronics and Information Technology Industries Association, "Survey on Trends in Focus Areas 2023", 2023, p. 1.

environment. This cooperation will involve exchanging views at various levels and participating in international forums such as the ICN and the OECD.

40. The generative AI market has been continuously facing changes and technological innovations that could affect the market structure, and the market environment is likely to change drastically in the future. Therefore, it is extremely important to properly understand the market trends. The JFTC will continue its study on the generative AI market in the hope that the market develops under fair and free competition.