Section 2 Trends in Explosive AI Progress

AI continues to evolve explosively, with progress being made in the development of large-scale general-purpose Large Language Models, while new technologies are emerging every day, and it is pointed out that this is a field with great potential for technological change. Furthermore, the AI field, which requires huge investments, tends to be led by overseas big tech companies (giant digital companies with global influence) and overseas startups with AI talent and advanced technology that receive huge investments.

Amidst this trend, while Japan's presence in the AI field is not necessarily high on a global scale, domestic companies and organizations are actively engaged in technological development, including model development.

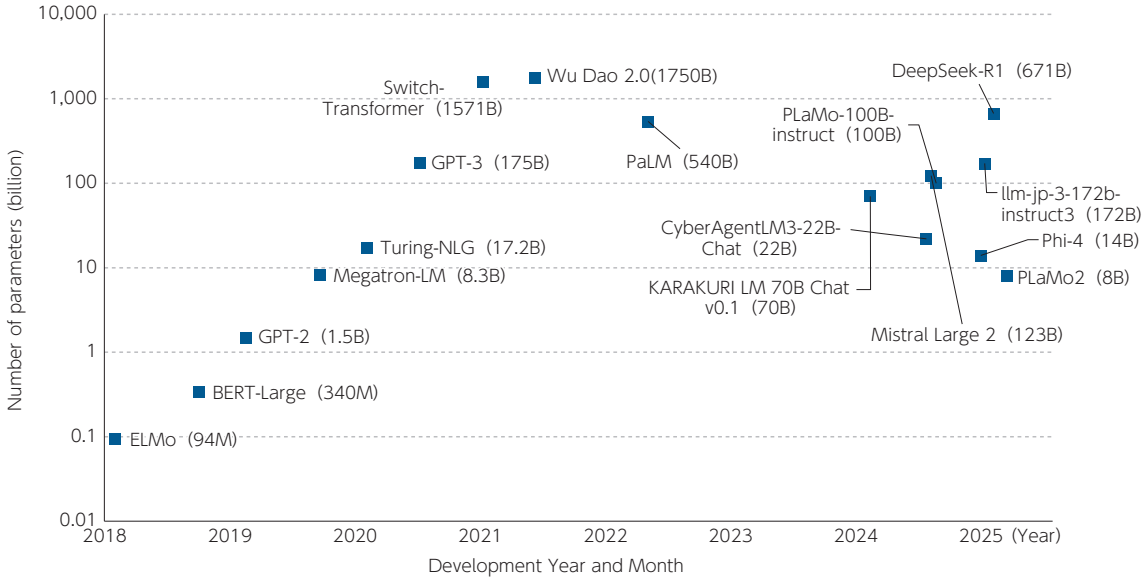
As AI continues to evolve, penetrate and integrate across all digital fields, it is increasingly likely to become a fundamental element underpinning the digital society.

1. Current Status and Trends in AI Technology Development

(1) Intensifying global AI development race

AI comes in many forms, but one of the major trends in recent AI technology development and applications is "generative AI," which generates text, images, videos, etc., and one of the technologies within this is the Large Language Model (LLM), which applies deep learning technology. In 2020, OpenAI proposed the Scaling Law, stating that the performance of LLM improves as the size of the data used for learning, the amount of calculations used for learning, and the number of model parameters increase. For example, while the GPT-2 model announced by OpenAI in 2019 had 1.5 billion parameters, the parameters in GPT-3, announced by the company in 2020, were scaled up to approximately 175 billion, about 120 times larger. The trend toward larger scale has continued since then, with the number of parameters in PaLM announced by Google in April 2022 reaching 540 billion (**Figure 1-1-2-1**).

Figure 1-1-2-1 Changes in the number of Large Language Model (LLM) parameters[1]



(Source) Prepared from the University of Tokyo Matsuo Laboratory "AI Evolution and Japan's Strategy" (2023) and presentation materials by development organizations

Following the Scaling Law, investment in computing resources such as GPU and data centers has intensified, particularly among big tech companies. Furthermore, startups such as OpenAI, Anthropic, and Mistral AI, which have received massive investments from big tech companies and venture capital firms, have also become major players in LLM development and are participating in the development race.

Figure (related data) Investment status of major AI developers
Source: Prepared from related reports and materials
URL: https://www.soumu.go.jp/johotsusintokei/whitepaper/ja/r07/html/datashu.html#f00028
(Data collection)

[1] This figure does not include models for which the number of parameters has not been published. Note that the number of parameters for recent LLMs is often not disclosed.

(2) Recent trends in AI research and development

The AI field is one where new technologies and models are announced daily, and technological innovation is advancing at a remarkable pace. While it is not possible to describe all of these technological innovations, the major technological trends that have recently emerged will be discussed, focusing on the LLM field.

A LLM research and development trends

(A) Emergence of reasoning models

In September 2024, OpenAI announced the development of the "OpenAI o1" series as a reasoning model for solving difficult problems. o1 outperformed OpenAI's GPT-4o model on numerous evaluation metrics in the fields of science, code generation, and mathematics, areas where conventional generative AI has traditionally struggled. For example, in the U.S. Mathematical Olympiad Preliminary Round, o1 was able to solve approximately 83% of the problems. This is considered to be an academic achievement comparable to the top 500 students in the U.S. o1 also outperformed experts with doctorates in the performance evaluation index of expertise in chemistry, physics, and biology[2]. It was also reported that when the o1 model was used to solve the 2025 University of Tokyo entrance exam questions,[3] it achieved a higher score than the minimum passing score[4].

(B) An open model developed by a Chinese AI startup and its impact on the market

In January 2025, Chinese AI startup DeepSeek announced the development of "DeepSeek-R1." This model is said to achieve performance equivalent to OpenAI's reasoning model "o1" through various technical innovations[5]. DeepSeek-R1 has garnered particular attention for being made openly available to anyone, being developed by an emerging Chinese startup, and its low development cost[6].

Immediately after the announcement of DeepSeek-R1, the stock prices of semiconductor companies and companies that provide AI infrastructure through cloud services, etc. fell, affecting all AI-related stocks. The emergence of DeepSeek-R1, which is said to have achieved both low cost and high performance, may have caused the fall of such stock prices as it cast doubt on the long-held belief that developing high-performance models requires massive investment.

However, developing and operating advanced AI requires more data processing, and if costs further decrease, the number of developers and users is expected to increase, therefore, it is suggested that demand for AI semiconductors will actually increase[7,8].

(C) Development of relatively small language models

While the competition to develop LLMs is intensifying, attention is also being paid to the development of models that are composed of LLMs with relatively fewer parameters. Generally, the more parameters an LLM has, the higher its learning ability. However, due to their large scale, their services are often provided only via cloud services or APIs, which can result in long response times due to the massive amount of computational processing, and the input data may be used as training data by the LLM provider. On the other hand, small scale models are relatively lightweight and capable of fast processing, making them advantageous in environments without network connectivity (local environments) or for specific applications. As a result, their development is being actively pursued.

For example, Microsoft has developed the "Phi" series as a small scale model, and released the 14 billion parameter "Phi-4" in December 2024. Phi-4 is a model capable of complex reasoning, and is said to have outperformed other companies' models of the same size in mathematics competition problems[9].

---

[2] OpenAI "Learning to reason with LLMs" <https://openai.com/index/learning-to-reason-with-llms/> (Reference March 11, 2025)

[3] The Common Test for University Admissions and the University of Tokyo Second Stage Examination

[4] Kyodo News (April 5, 2025) "AI Meets the 'Passing Standard' for the University of Tokyo, College of Arts and Sciences, Science III Exams: Exceeds the Minimum Passing Score for 2025 Entrance Exams" <https://nordot.app/1281148575208423464> (Reference April 14, 2025)

[5] DeepSeek-R1 and its predecessor, DeepSeek-V3, are said to have achieved high performance at low cost through various technological innovations. For example, DeepSeek-V3 employs a technique called MoE (Mixture of Experts), which incorporates multiple specialized units (Experts) called upon for each input. To optimize the parallel computation required for this, various innovations were implemented (based on an interview with Okanohara Daisuke, CEO of Preferred Networks).

[6] DeepSeek explains that it costs approximately $5.6 million (approximately 860 million yen) to develop one model. According to development cost calculations by Stanford University, the U.S. Google's "Gemini Ultra" cost $191 million and OpenAI's "GPT-4" cost $78 million, which means an even simple comparison with these costs shows the DeepSeek model was developed at less than one-tenth the cost of the major American companies. It should be noted that the figures cited in the paper may not represent the total cost of the entire development process. While there is debate regarding the development costs, many within the U.S. tech industry view the DeepSeek model as having been developed at a significantly lower cost compared to previous costs. Nikkei "DeepSeek Shock: AI Development Costs "Less Than One-Tenth" Using Open-Source Technology <https://www.nikkei.com/article/DGXZQOGN27CE40X20C25A1000000/> (Reference March 12, 2025)

[7] Based on expert interviews (including Kinoshita Masafumi, Executive Officer, ABEJA, Inc.).

[8] Meta announced plans to invest up to $65 billion in AI-related projects by 2025, while Alphabet, the parent company of the U.S. Google, expects to spend $75 billion on capital expenditures for data centers and AI infrastructure development in 2025. (Bloomberg "Meta Shares Rise, Longest Streak of 12 Consecutive Gains -Market Cap Gains $240 Billion" (February 5, 2025) < https://www.bloomberg.co.jp/news/articles/2025-02-04/SR63J7DWLU6800> (Reference March 18, 2025), Bloomberg "Alphabet shares plummet revenue falls short of expectations -cloud growth slows" (February 5, 2025) < https://www.bloomberg.co.jp/news/articles/2025-02-04/SR6G6DT0AFB400> (From reference March 18, 2025)

[9] Microsoft, "Introducing Phi-4:Microsoft's Newest Small Language Model Specializing in Complex Reasoning" <https://techcommunity.microsoft.com/blog/aiplatformblog/introducing-phi-4-microsoft%E2%80%99s-newest-small-language-model-specializing-in-comple/4357090> (Reference March 16, 2025)

B AI agents

As generative AI advances, services called "AI agents" are becoming more widespread. The definition of an AI agent varies depending on the development company, but recently, it is thought that the term AI agent tends to refer to what has the ability to automatically determine tasks (and subdivide tasks as necessary) and carry out processing in response to set goals or instructions given in natural language.

Since the second half of 2024, many services claiming to be "AI agents" have been developed and launched (**Figure 1-1-2-2**) . Some point out that AI agents themselves are not particularly technologically advanced. [10]However, it is true that advances in LLMs are enabling AI to handle more complex tasks, such as multimodal AI that can process multiple data formats such as text, images, audio, and video in an integrated manner, and reasoning models that are capable of logical thinking. It is expected that more user-friendly services leveraging LLMs and other technologies will emerge in the future.

Figure 1-1-2-2 Service overview and definition of so-called "AI agents"

Company name	Service overview
Microsoft	・Three types of AI agents, including built-in agents integrated into Microsoft 365 Copilot are offered. The built-in agents are embedded in various Microsoft 365 apps and can play the roles of interpreter, facilitator, self-service, project manager, etc.
Salesforce	・In October 2024, the company announced that it would offer the autonomous AI agent "Agentforce" in Japan. Responding to customers in natural language, arranging appointments with prospective customers, coaching sales representatives through role-playing, and assisting with narrowing down customer segments and creating content for campaign development.
Amazon Web Services	・"Agents for Amazon Bedrock," a service for developing AI agents, is offered. The acquired information can be used to respond to user instructions and tasks. The system also includes enhanced functionality to track the process the AI agent executed to arrive at the final answer after the user entered instructions, as well as enhanced functionality to improve response accuracy.
OpenAI	・In January 2025, the company announced the launch of "Operator," an AI agent service in which the AI performs tasks on behalf of users. When the user inputs the task they want to perform, the AI reads the screen like a screenshot and executes it.
Anthropic	・In October 2024, the company announced the development of "computer use," an AI that operates computers on behalf of people. The screen the user is viewing on their computer is recognized as a screenshot converted into image data, and operations such as moving the cursor, clicking, and entering text can be performed.

Company name	Definition of an AI agent
Microsoft	A system designed to perform specific tasks, answer questions, and automate a series of processes on behalf of a user[11].
Salesforce	An intelligent system that can understand and respond to user inquiries without human intervention[12].
Amazon Web Services	A software program that interacts with its environment, collects data, and uses that data to perform self-determined tasks to achieve predetermined goals. Humans set the goals, but the AI agent independently chooses the optimal actions it needs to take to achieve them[13].

(Note) As this information is based on materials from each company at the time of announcement, it may differ from the current service content and definitions.

(Source) Prepared from reference materials from each company and related reports[14]

[10] Based on an interview with Professor Satoshi Kurihara of Keio University.
[11] Microsoft "What is an AI Agent?" <https://learn.microsoft.com/ja-jp/azure/cosmos-db/ai-agents> (Reference March 13, 2025)
[12] Salesforce "What is an AI Agent – The Complete Guide" <https://www.salesforce.com/jp/agentforce/what-are-ai-agents/> (Reference March 13, 2025)
[13] Amazon "What is an AI Agent?" <https://aws.amazon.com/jp/what-is/ai-agents/> (Reference March 13, 2025)
[14] Microsoft, "Operational Efficiency and Innovation Enabled by AI Agents: Latest Case Studies from Japan" (December 18, 2024 <https://news.microsoft.com/ja-jp/2024/12/18/241218-operational-efficiency-and-innovation-enabled-by-ai-agents-latest-case-studies-from-japan/> (Reference March 18, 2025), NIKKEI X Tech, "MS Announces Genuine AI Agent, CEO Nadella Aims for an 'Agentic World'" (November 20, 2024) https://xtech.nikkei.com/atcl/nxt/column/18/03012/112000001/ (Reference March 18, 2025), Salesforce "Agentforce" <https://www.salesforce.com/jp/agentforce/> (Reference March 17, 2025), NTT EAST "Explanation of AI Agents and Amazon Bedrock Agents" (March 12, 2025) <https://business.ntt-east.co.jp/content/cloudsolution/column-578.html> (Reference March 18, 2025), NIKKEI X Tech "AWS, Google, and MS's AI Agent Development Services Enable Building AI Agents without Coding" (October 10, 2024) <https://xtech.nikkei.com/atcl/nxt/column/18/02968/100100003/> (Reference March 18, 2025), NIKKEI X Tech "Open AI has Announces its AI agent "Operator," which Executes User-Directed Tasks Using its Own Custom Browser" <https://xtech.nikkei.com/atcl/nxt/news/24/02098/> (Reference March 17, 2025), NIKKEI "AI Developed by Anthropic in the U.S. to Operate Computers on Behalf of Humans" (October 23, 2024) https://www.nikkei.com/nkd/company/us/CRM/news/?DisplayType=1&ng=DGXZQOGN22DYL022102024000000 (Reference March 17, 2025), Bloomberg "Anthropic Reveals New AI Tool, Making the Experience More Intuitive" (October 23, 2024) https://www.bloomberg.co.jp/news/articles/2024-10-23/SLS7T9T0AFB400 (Reference March 17, 2025)

C AI Robotics

Competition for development and investment in the field of AI robotics, which applies AI technology to robotics, is heating up. The background to this is that dramatic innovations in image recognition and natural language processing in the field of AI have made flexible and complex processing possible, and robots are expected to replace labor in developed countries overseas where there are concerns about labor shortages due to declining birthrates and aging populations.

For example, as one area of such trend, the development of humanoid robots has recently become active. There is no clear definition of a humanoid robot, but it is generally modeled after the shape and capabilities of a human, and is designed to handle a wide range of tasks, from use in factories, such as grasping objects and arranging parts, to tasks intended for use in the home, such as folding laundry. Because humanoid robots have a human-like shape, they are thought to be easily introduced into human-centered social infrastructure, and it is expected that in the future they will be able to assist or replace human tasks[15].

Against this backdrop, the race to develop humanoid robots for commercialization is heating up, particularly in the U.S. and China. Currently, humanoid robots are being developed for pilot deployment in manufacturing sites and factories and for industrial use, but some companies are aiming to create humanoid robots that can be used in everyday life, such as for household chores and entertainment in the long run. It is expected that research and development into highly versatile robots will continue in the future **(Figure 1-1-2-3)**.

Figure 1-1-2-3 Humanoid robot development examples

Company name/Model name	Characteristics/Direction
Tesla (U.S.) Optimus	·In October 2024, the company unveiled the humanoid robot "Optimus." ·In addition to industrial use, the robot can also be used for everyday purposes such as household chores and entertainment. ·The company plans to produce small quantities for its own factories in 2025, and mass-produce for other companies in 2026.
Figure AI (U.S.) Figure 02	·In August 2024, the company announced its new humanoid robot, "Figure 02." ·In partnership with Open AI, the humanoid robot incorporates a voice recognition system that can have natural conversations with humans. ·A pilot introduction of the humanoid robot was conducted at a BMW plant. ·The company aims to alleviate labor shortages and improve industrial efficiency, while also considering developing a consumer version.
Agility Robotics (U.S.) Digit	·In March 2023, the company announced the next generation model of the bipedal multipurpose robot "Digit." ·Designed to work with humans in logistics facilities such as warehouses.
Boston Dynamics (U.S.) Atlas	·In April 2024, the company announced the fully electric robot "Atlas" as the successor to its previous hydraulic model "HD Atlas." Compared with conventional hydraulic models, the new robot has a wider range of motion. ·In 2025, a pilot introduction of the new robot will be conducted at an automobile manufacturing plant, with the aim of using it in the industrial sector.
Unitree Robotics (China) G1	·In May 2024, the company announced the bipedal humanoid robot "G1." It features a balancing system that allows it to continue walking without falling over even when pushed by a person, and a joint mechanism with a wide range of motion. ·The price starts at $16,000 (approximately 2.5 million yen), making it a low-cost option that will encourage social implementation.
Pudu Robotics (China) PUDU D9	·In December 2024, the company announced its first bipedal humanoid robot, "PUDU D9." It walks on two legs at the same speed as an adult, 2m/s, and can handle a variety of terrain, including stairs, slopes, and uneven ground.

(Source) Prepared from reference materials[16]

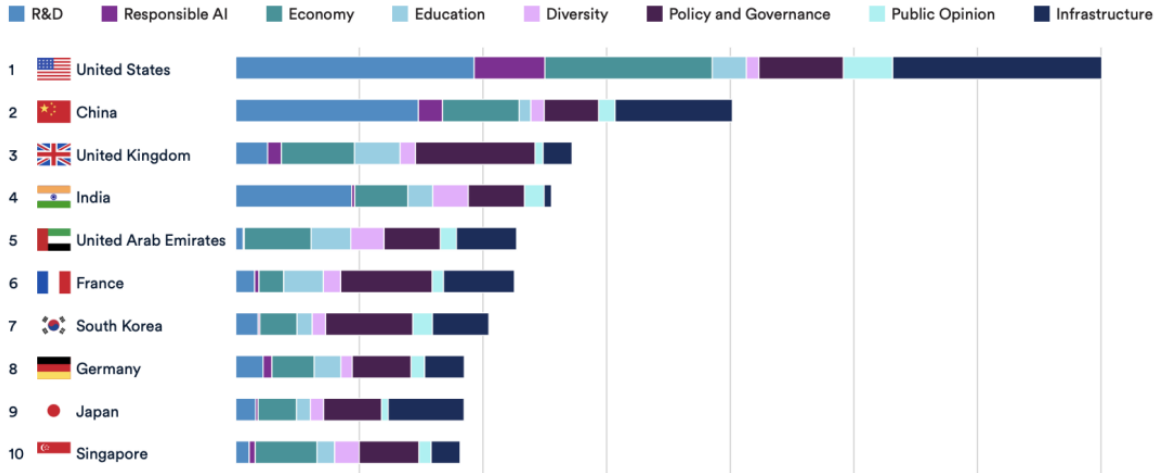
[15] Mitsui & Co. Global Strategic Studies Institute "Humanoid Robots - Technological Advancement Driven by Generative AI and the Beginning of Pilot Introduction" <https://www.mitsui.com/mgssi/ja/report/detail/__icsFiles/afieldfile/2025/02/07/2501btf_tsuji_matsuura.pdf> (Reference March 19, 2025)

[16] NIKKEI "Tesla Unveils Humanoid Robot Performing Household Chores and Other Everyday Tasks" (October 11, 2024) https://www.nikkei.com/article/DGXZQOGN114AB0R11C24A0000000/ (Reference March 19, 2025), Ledge.ai "Figure, in Partnership with Open AI, Unveils New Humanoid Robot "Figure 02" and Successfully Tests with BMW for Practical Use in Automotive Manufacturing" (August 11, 2024) <https://ledge.ai/articles/figure02>, CNET Japan "Revamps Ability Robotics, Bipedal Robot "Digit" (April 4, 2023) <https://japan.cnet.com/article/35202097/> (Reference March 19, 2025), NIKKEI X Tech "Interview with Boston Dynamics, Humanoid Robots, Starting as Automobile Plant Workers" (March 19, 2025) https://xtech.nikkei.com/atcl/nxt/column/18/03118/00004/ (reference March 19, 2025), NIKKEI X Tech "Where are Bipedal Humanoid Robots Headed? AI Advancement are Reaccelerating the Development Race among Overseas Companies" (July 31, 2024) <https://xtech.nikkei.com/atcl/nxt/mag/nmc/18/00011/00263/?P=3> (Reference March 19, 2025), NIKKEI X Tech "Lower Prices Drive Social Implementation, Developing a Wide Range of Applications from Factories to Homes" (January 31, 2025) <https://xtech.nikkei.com/atcl/nxt/mag/nmc/18/00163/00004/?P=3> (Reference March 19, 2025), Unitree <https://www.unitree.com/g1> (Reference March 19, 2025), PRTIMES "PUDU Unveils its First Full-sized Humanoid Robot "PUDU D9" (December 20, 2024) <https://prtimes.jp/main/html/rd/p/000000051.000087027.html>

(3) Trends in AI development and business expansion in Japan

Looking at various evaluation reports on AI, it cannot be said that Japan is highly rated compared with other world-leading countries in terms of its AI research and development capabilities and utilization of AI. For example, according to the 2023 Global AI vibrancy ranking released by Stanford University's HAI (Human-Centered Artificial Intelligence) in November 2024, Japan is ranked 9th overall, lagging behind countries such as the U.S., China, and the UK **(Figure 1-1-2-4)**. In addition, in AIRankings, which ranks AI research capabilities based on factors such as the number of papers on AI, the top countries in recent years have been the U.S., China, the UK, and Germany, in that order, with Japan hovering around 11th or 12th place. However, Japanese companies and organizations are also taking various steps toward AI development.

Figure 1-1-2-4 Global AI vibrancy ranking: top ten countries (2023)



(Source) Stanford University Human-Centered Artificial Intelligence(2024) "Global AI Vibrancy Tool" [17]

Figure (related data) Changes of AI rankings by country
Source: Prepared based on AIRankings (data obtained on February 25, 2025)
URL: https://www.soumu.go.jp/johotsusintokei/whitepaper/ja/r07/html/datashu.html#f00032
(Data collection)

A LLM research and development trends

While overseas big tech companies and AI startups are globally leading the development of LLMs, Japanese organizations are also working on developing LLMs. Compared with the world's most cutting-edge models, Japanese models tend to be relatively small-scale. In addition, in recent years, development has also progressed on relatively small-scale yet high-performance models **(Figure 1-1-2-5)**.

[17] Stanford University "Global AI Vibrancy Tool" <https://hai.stanford.edu/ai-index/global- vibrancy-tool> (Reference March 14, 2025)

Figure 1-1-2-5 Examples of LLM development by Japanese organizations

Development organization: Model name (number of parameters)	Announcement date	Overview
Tokyo Institute of Technology (currently Institute of Science Tokyo),, Tohoku University, Fujitsu Limited, RIKEN, Nagoya University, CyberAgent Inc., Kotoba Technologies Inc.: Fugaku-LLM (about 13 billion)	May 2024	Learning is done using the supercomputer "Fugaku." Improved Japanese performance by having Fugaku learn original data from scratch. Demonstrated high performance in an evaluation index for Japanese, particularly well in humanities and social science tasks. Excellent in terms of transparency and security as the entire learning process can be monitored[18].
CyberAgent: CyberAgentLM3-22B-Chat (about 22.5 billion)	July 2024	The model was developed without being based on any existing models. Its Japanese proficiency is on par with Meta's model "Meta-Llama-3-70B Instruct[19]."
Fujitsu and Cohere Inc: Takane (unknown number of parameters)	September 2024	This model was developed based on the LLM from Canadian AI startup Cohere and can be used for highly specialized tasks. This LLM achieved the world's highest record in terms of the Japanese language performance evaluation index (at that time). This model can be used in a secure private environment[20].
National Institute of Advanced Industrial Science and Technology and Institute of Science Tokyo (A research team led by Professors Okazaki Naoaki and Yokota Rio): Llama 3.1 Swallow (about 8 billion/ about 70 billion)	October 2024	While maintaining the capabilities of a large language model (Meta's Llama 3.1) with high English proficiency, the model succeeded in improving Japanese language understanding, generation, and dialogue capabilities[21].
Preferred Networks Group[22]: PLaMo-100B (about 100 billion)	October 2024[23]	It was developed from scratch with proprietary architecture and training data. It achieved high Japanese language performance by surpassing OpenAI's GPT-4o and others in a Japanese language performance benchmark (Later, "PLaMo2 1B" and "PLaMo2 8B" were developed with the number of parameters reduced)[24].
National Institute of Informatics llm-jp-3-172b-instruct3 (about 172 billion)	December 2024	Trained a large language model with about 172 billion parameters (the same scale as GPT-3) from scratch at the institute's Research and Development Center for Large Language Models and made it available to the public[25].

* Model descriptions are those at the time the press releases were published and models were released by each organization.

(Source) Prepared from published materials of each development organization

The government support measures are supporting such LLM development originating in Japan.

For example, the "GENIAC" project, launched by the METI and the New Energy and Industrial Technology Development Organization (NEDO), supports the provision of computing resources necessary for the development of foundation models. In the first phase (development period for providing computing resources: February to August 2024), the project worked on the development of foundation models for a total of 10 development themes. The second phase began in October 2024, with development of foundation models underway for a total of 20 development themes.

Additionally, in order to strengthen AI development capabilities, the Ministry of Internal Affairs and Communications (MIC) is leading efforts, through the National Institute of Information and Communications Technology (NICT), to develop and expand the large amount of high-quality training data, primarily in Japanese, required for LLM development, and provide it to Japanese LLM developers[26].

Furthermore, the Ministry of Education, Culture, Sports, Science and Technology is leading efforts, through the Inter-University Research Institute Corporation/Research Organization of Information and Systems/National Institute of Informatics (NII), to conduct research and development contributing to the transparency and reliability of generative AI. It is also working to share the knowledge and experience gained through a study group (LLM-jp) that brings together AI researchers and engineers from industry and academia, centered around the NII, in an effort to contribute to fostering research and development capabilities related to generative AI in Japan.

[18] Fujitsu "Release of 'Fugaku-LLM' – a large language model trained on the supercomputer 'Fugaku'"<https://pr.fujitsu.com/jp/news/2024/05/10.html> (Reference March 17, 2025)
[19] CyberAgent "publicly releases version 3 of its proprietary Japanese LLM (Large Language Model) - Providing a commercially available model with 22.5 billion parameters" <https://www.cyberagent.co.jp/news/detail/id=30463> (Reference March 17, 2025)
[20] Fujitsu "launches 'Takane' - A large language model for enterprises offering the highest Japanese language proficiency in the world" <https://pr.fujitsu.com/jp/news/2024/09/30.html> (Reference March 17, 2025)
[21] National Institute of Advanced Industrial Science and Technology "releases 'Llama 3.1 Swallow' by improving the Japanese language capabilities of the open source LLM" <https://www.aist.go.jp/aist_j/press_release/pr2024/pr20241008_2/pr20241008_2.html> (Reference March 17, 2025)
[22] Developed by Preferred Elements, a 100%-owned subsidiary of Preferred Networks
[23] Release date of PLaMo-100B-Pretrained<https://www.preferred.jp/ja/news/pr20241015/>. Note that PLaMo-100B-Instruct was announced in August 2024<https://tech.preferred.jp/ja/blog/plamo-100b-post-training/>
[24] Preferred Networks "releases PlaMo-100B-Pretrained, a Large Language Model as a result of developing the first cycle of GENIAC" (October 15, 2024) <https://www.preferred.jp/ja/news/pr20241015/ > (Reference March 17, 2025)
[25] National Institute of Informatics "publicly releases 'llm-jp-172b-instruct3,' a fully open large language model with about 172 billion parameters (GPT-3 level)" <https://www.nii.ac.jp/news/release/2024/1224.html> (Reference March 17, 2025)
[26] NICT is using the language data it has built for learning to prototype an LLM with up to 311 billion parameters.

B Trends in AI robotics research and development and social implementation

The U.S. and China are particularly ahead in the research, development, and social implementation of humanoid robots, but in Japan, research and development in the field of AI robotics (including collaboration between AI and robots) is also being conducted by industry and academia, and efforts are also underway toward social implementation. Here are some examples.

(A) Development of a robot foundation model by the AI Robot Association[27]

In December 2024, the AI Robot Association (AIRoA), a general incorporated association, was established with the aim of building a robot data ecosystem through the fusion of AI and robots. As AI technology evolves, there is a growing need for foundation models capable of integrating large amounts of data and learning efficiently, even in the field of robotics. However, it has been pointed out that the current market does not have a fully developed framework for sharing and utilizing large amounts of data, and that development efficiency remains difficult to improve as each company and research institution handles data separately. AIRoA promotes open, large-scale data collection across industry boundaries and the development and publication of foundation models, with the aim of building a scalable robot data ecosystem to realize advanced general-purpose robots.

(B) Social experiment using the guide robot "ugo" and generative AI developed by NTT

A social experiment using the guide robot 'ugo' developed by ugo, Inc. that provides autonomous mobile business robots, and several generative AI models, including the NTT-developed LLM "tsuzumi," began in January 2025 by Osaka Metro, NTT WEST Group, NTT Communications[28], and ugo, Inc., in anticipation of an increase in tourists from Japan and abroad visiting EXPO 2025 Osaka, Kansai. The experiment explores ways to improve the ability to respond quickly to questions in multiple languages **(Figure 1-1-2-6)**.

Figure 1-1-2-6 Social experiment using multilingual guide robot "ugo" and generative AI "tsuzumi," etc.

(Source) NTT WEST, Inc.[29]

C Trends in AI agent development and business expansion

Japanese companies are also making moves regarding AI agents.

For example, in February 2025, SoftBank Group Corp. and OpenAI announced a partnership to develop and sell "Cristal intelligence," a cutting-edge AI for businesses. The partnership will see SoftBank and OpenAI exclusively sell "Cristal intelligence," which has AI agent functions, to major Japanese companies[30]. SoftBank Group will jointly develop AI agents with OpenAI, and plans to operate its AI data center in Japan.[31,32]

Fujitsu has also developed and begun offering globally an AI service called "Fujitsu Kozuchi AI Agent" that enables AI to autonomously and collaboratively handle highly challenging tasks.[33]

[27] NIKKEI X Tech, "Toyota and other major companies join forces to start building a 'robot foundation model' originating in Japan" (March 12, 2025) <https://xtech.nikkei.com/atcl/nxt/column/18/03118/00001/> (Reference March 21, 2025)
[28] NTT Communications changed its company name to "NTTDOCOMO BUSINESS, Inc." as of July 1, 2025
[29] Osaka Metro, NTT WEST Group, NTT Communications, ugo inc. "A social experiment will be conducted at Umeda Station on the Osaka Metro Midosuji Line, using the multilingual guide robot "ugo" and the generative AI "tsuzumi." <https://www.ntt-west.co.jp/news/2501/250108a.html>
[30] SoftBank News: "OpenAI and SoftBank Group go into partnership. "Crystal intelligence," a cutting-edge AI for businesses, will be offered in Japan, ahead of the rest of the world" (February 5, 2025) <https://www.softbank.jp/sbnews/entry/20250204_02> (Reference March 26, 2025)
[31] Nikkei "SoftBank, OpenAI to run autonomous AI agents at former Sharp Sakai plant" (March 14, 2025) <https://www.nikkei.com/article/DGXZQOUC12A7X0S5A310C2000000/> (Reference March 10, 2025)
[32] OpenAI announced the "Stargate Project," which plans to invest $500 billion (approximately 78 trillion yen) over four years from 2025 to build a new AI infrastructure for OpenAI in the U.S. SoftBank is among the initial investors. (Open AI "Announcing the Stargate Project" <https://openai.com/ja-JP/index/announcing-the-stargate-project/> (Reference March 15, 2025)
[33] Fujitsu "Began offering "Fujitsu Kozuchi AI Agent" in which AI autonomously collaborates with humans to promote advanced business processes" (October 23, 2024)
<https://www.nikkei.com/article/DGXZRSP680645_T21C24A0000000/?msockid=32429ef1e23f674838988bb8e35b6633> (Reference March 10, 2025)

2. Current status of AI use

(1) Current status of AI use among individuals

In order to understand the current state of AI use among individuals, a survey was conducted targeting the general public (hereinafter referred to as the "FY2024 survey").

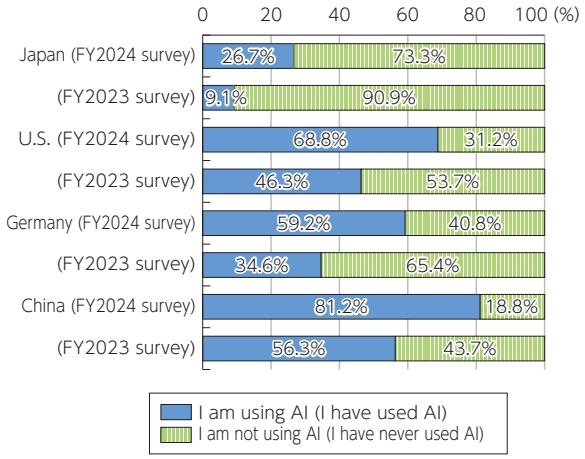
In Japan, the percentage of respondents who answered that they "use (or have used in the past)" some kind of generative AI service was 26.7% in the FY2024 survey. Considering that the percentage of those who answered "I use generative AI (or have used it in the past)" was 9.1% in a survey conducted in FY2023 (hereinafter referred to as the "FY2023 survey"), usage has been expanding (**Figure 1-1-2-7**) . In addition, looking at the age groups, 44.7% of people in their 20s responded that they had used AI services in the FY2024 survey (**Figure 1-1-2-8**) .

Similar surveys were also conducted in the U.S., Germany, and China, and AI use is expanding in each of these three countries (the percentage of people who have experience using generative AI increased from 46.3% in the FY2023 survey to 68.8% in the FY2024 survey in the U.S., from 34.6% to 59.2% in Germany, and from 56.3% to 81.2% in China).

In Japan, when people who answered that they "do not use (have never used)" text generation AI services were asked about the reasons for not using them, the most common response was "not necessary for my life or work," followed by "I don't know how to use it," suggesting that the barriers to use are still high. Many also responded that there were "no attractive services" (**Figure 1-1-2-9**) .

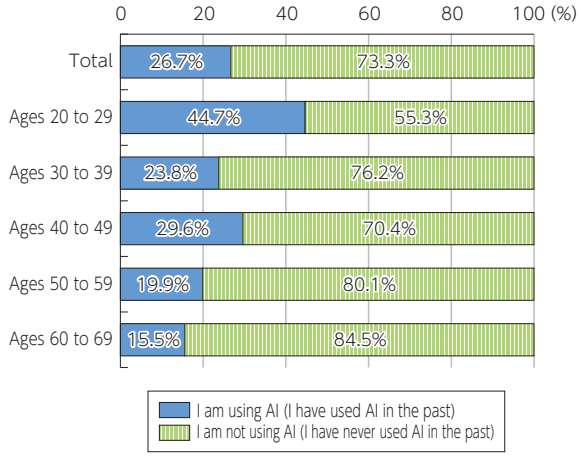
Additionally, when asked about their intention to use AI (including generative AI), there was a relatively high intention to use it for research and content summarization and translation, a trend similar to that seen in the FY2023 survey (**Figure 1-1-2-10**) .

Figure 1-1-2-7 Experience using generative AI services (by country)



(Source) MIC (2025) "Survey on the latest trends in ICT, R&D, and digital utilization in Japan and other countries"

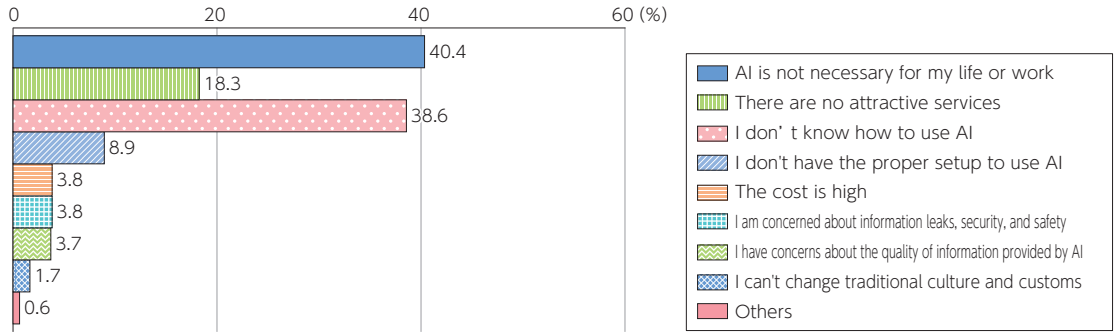
Figure 1-1-2-8 Experience using generative AI services (by age group, Japan)



(Source) MIC (2025) "Survey on the latest trends in ICT, R&D, and digital utilization in Japan and other countries"

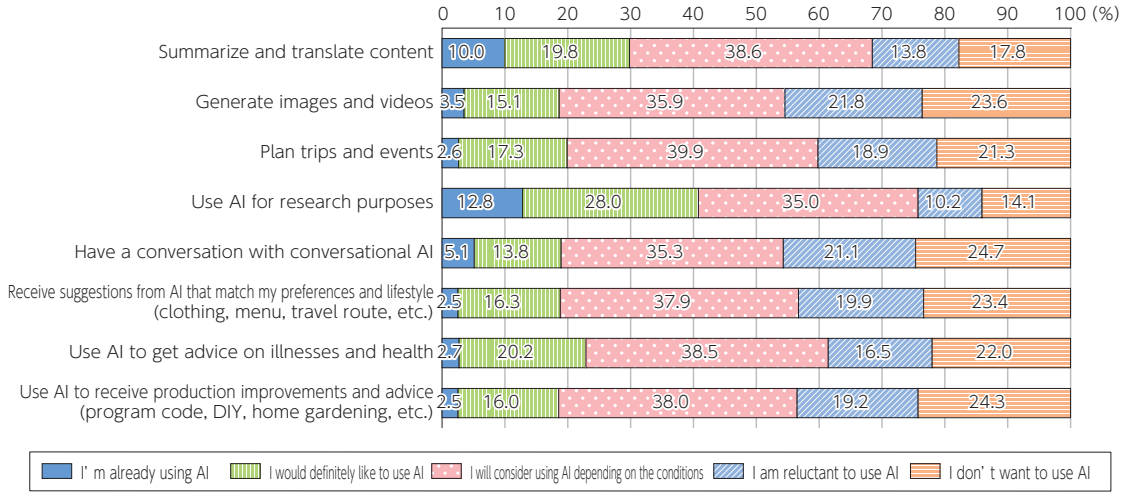
Figure (related data) Experience using each generative AI service (Japan, by age group)
Source: MIC (2025) "Survey on the latest trends in ICT, R&D, and digital utilization in Japan and other countries"
URL: https://www.soumu.go.jp/johotsusintokei/whitepaper/ja/r07/html/datashu.html#f00037
(Data collection)

Figure 1-1-2-9 Reasons for not using text generation AI services



(Source) MIC (2025) "Survey on the latest trends in ICT, R&D, and digital utilization in Japan and other countries"

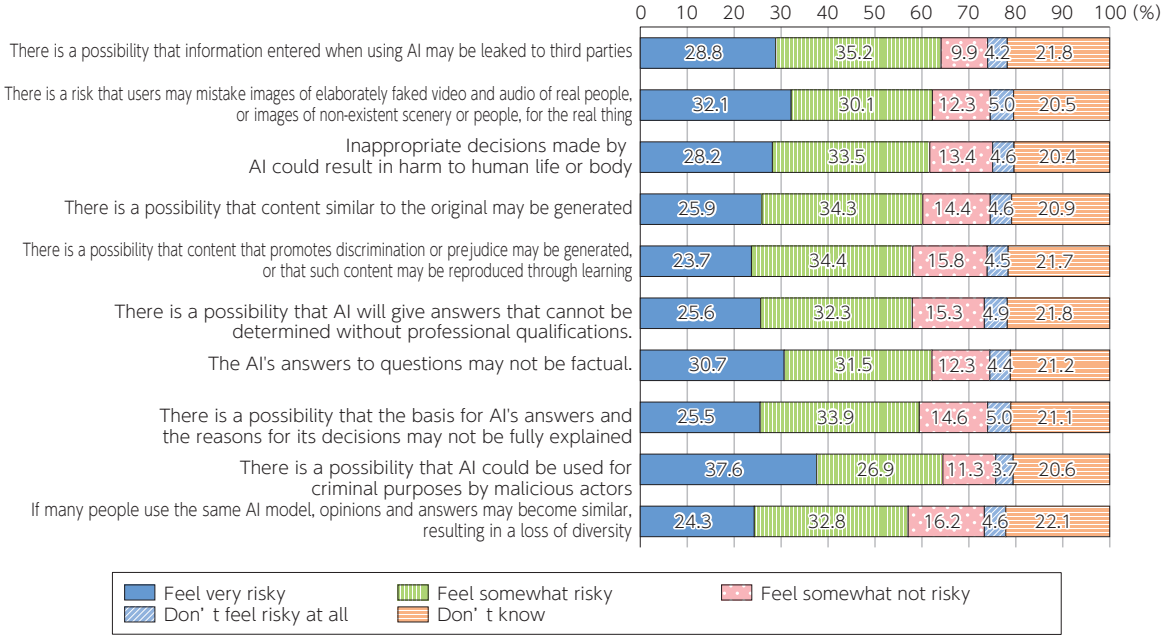
Figure 1-1-2-10 Intention to use generative AI and AI



(Source) MIC (2025) "Survey on the latest trends in ICT, R&D, and digital utilization in Japan and other countries"

In a survey of people's awareness of the risks of using AI, the number of respondents who said that "I feel it is very risky" was relatively high for such reasons as criminal use by malicious actors, being deceived by elaborate fakes, and the possibility that AI's answers to questions may not be factual (**Figure 1-1-2-11**) .

Figure 1-1-2-11 Perspective on the risks of AI use



(Source) MIC (2025) "Survey on the latest trends in ICT, R&D, and digital utilization in Japan and other countries"

Figure (related data) Intention to use generative AI and AI (by country)
Source: MIC (2025) "Survey on the latest trends in ICT, R&D, and digital utilization in Japan and other countries"
URL: https://www.soumu.go.jp/johotsusintokei/whitepaper/ja/r07/html/datashu.html#f00041
(Data collection)

Figure (related data) Perspective on the risks of AI use (by country)
Source: MIC (2025) "Survey on the latest trends in ICT, R&D, and digital utilization in Japan and other countries"
URL: https://www.soumu.go.jp/johotsusintokei/whitepaper/ja/r07/html/datashu.html#f00042
(Data collection)

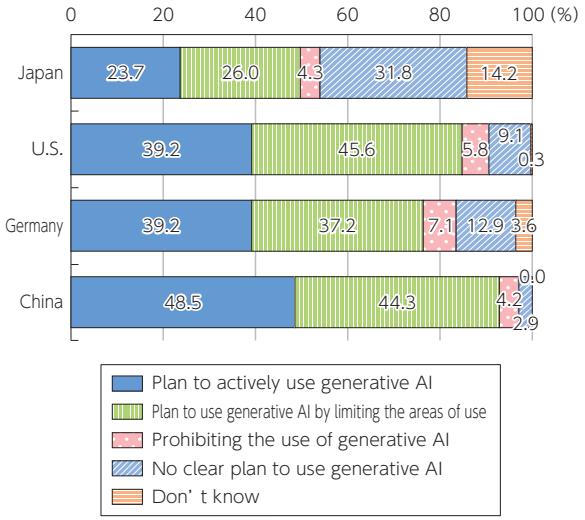
(2) Current status of AI use among companies

Based on a survey conducted in four countries, namely Japan, the U.S., Germany, and China, the current status of AI use in companies is summarized.

Respondents were asked about the policy for using generative AI at their companies, the percentage of companies in Japan that formulate a policy of "positively using generative AI" or "limiting domains for using generative AI" was 49.7% in the FY 2024 survey, an increase from the FY2023 survey (42.7%). On the other hand, compared with other countries surveyed this time, Japan continues to show a lower tendency to use generative AI than other countries (**Figure 1-1-2-12**) .

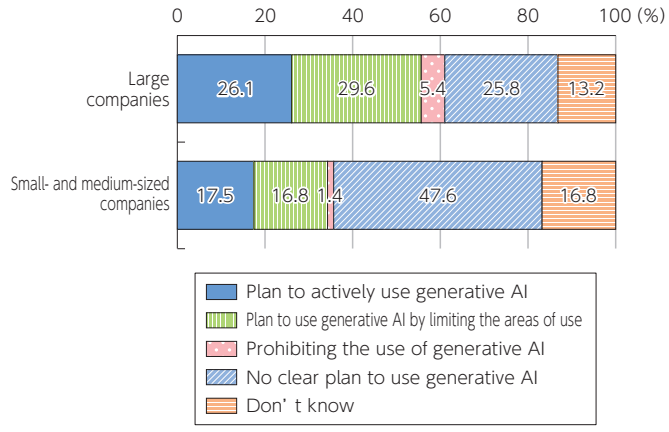
In addition, when looking at the situation in Japan by company size, the majority of small and medium-sized companies responded that they have "not clearly formulated a policy for using generative AI," accounting for about half of the total. The survey indicates that Japanese small and medium-sized companies are lagging behind large companies in deciding on a policy of using generative AI (**Figure 1-1-2-13**) .

Figure 1-1-2-12 Status of policy formulation for using generative AI (by country)



(Source) MIC (2025) "Survey on the latest trends in ICT, R&D, and digital utilization in Japan and other countries"

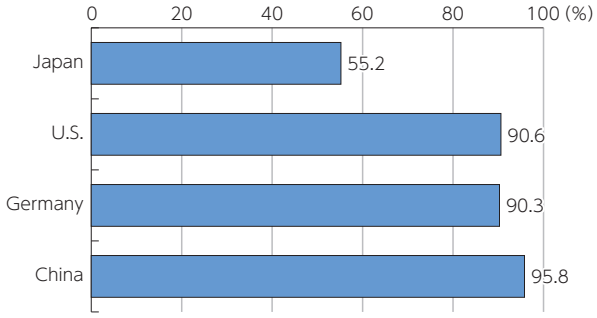
Figure 1-1-2-13 Status of policy formulation for using generative AI (by company size (Japan))



(Source) MIC (2025) "Survey on the latest trends in ICT, R&D, and digital utilization in Japan and other countries"

Furthermore, when asked about the current status of using generative AI for business areas where it is expected to be used, the percentage of respondents in Japan who answered that they are using generative AI for some kind of business was 55.2% (the percentage who answered "currently using generative AI for business") **(Figure 1-1-2-14)**. For example, regarding individual tasks, 47.3% of respondents in Japan (those who answered "currently using generative AI for business") said they use generative AI to "assist with emails, minutes, document preparation, etc.[34]" Both percentages are low compared with other countries.

Figure 1-1-2-14 Percentage of using generative AI for business in companies (by country)[35]



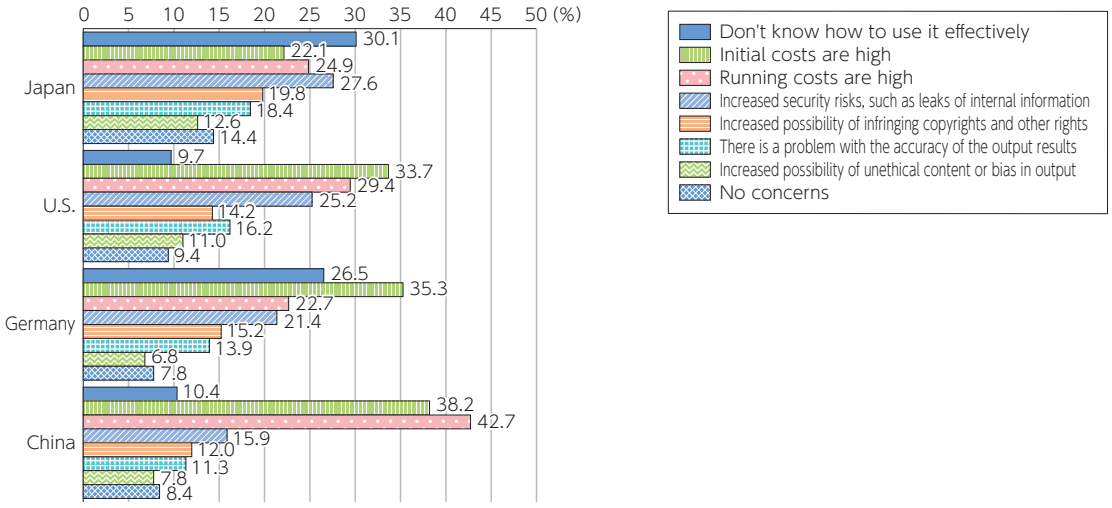
(Source) MIC (2025) "Survey on the latest trends in ICT, R&D, and digital utilization in Japan and other countries"

Figure (related data) Percentage of using generative AI for business in companies (by business, by country)
Source: MIC (2025) "Survey on the latest trends in ICT, R&D, and digital utilization in Japan and other countries"
URL: https://www.soumu.go.jp/johotsusintokei/whitepaper/ja/r07/html/datashu.html#f00046
(Data collection)

When respondents were asked about concerns about introducing generative AI, the most common answer in Japan was "I don't know how to use it effectively," followed by "security risks such as leaks of internal information," "high running costs," and "high initial costs" **(Figure 1-1-2-15)** .

When respondents were asked about their thoughts on the impact that promoting the use of generative AI will have on their company, the most common answer in Japan was that it will "improve business efficiency and alleviate labor shortage." In the other three countries, respondents tend to cite business expansion, acquiring new customers, and new innovations. Overall, all four countries appear to be focusing on the positive aspects of generative AI, such as increased operational efficiency and business expansion, rather than the negative aspects, such as increased security risks **(Figure 1-1-2-16)**.

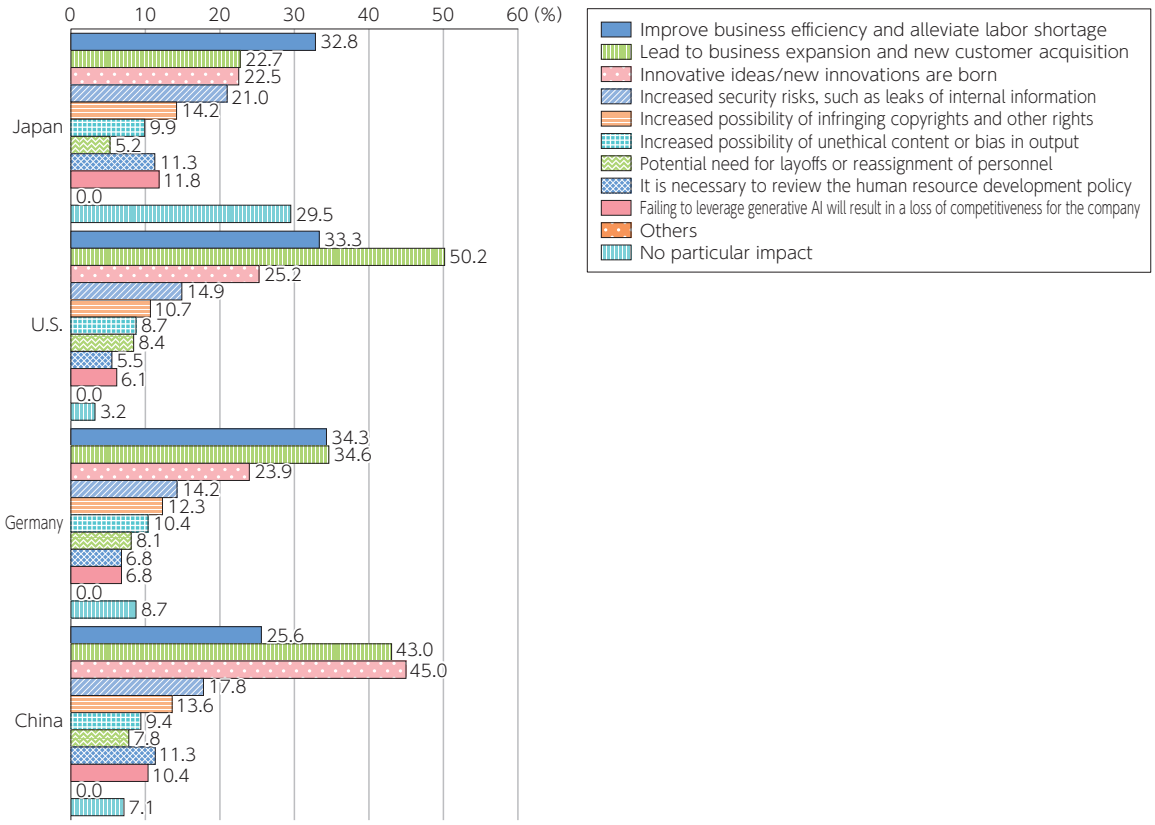
Figure 1-1-2-15 Concerns about introducing generative AI (by country)



(Source) MIC (2025) "Survey on the latest trends in ICT, R&D, and digital utilization in Japan and other countries"

[34] Regarding the percentage of using generative AI for each individual task within a company, as the estimation method used for the FY2024 survey was changed from that of the FY2023 survey to improve estimation accuracy, a simple comparison between the FY2024 survey and the FY2023 survey is not possible.
[35] Estimated based on the percentage of respondents who know their company's policy of using AI and answered that they use generative AI for some kind of business.

Figure 1-1-2-16 Effects and impacts of using generative AI (by country)



(Source) MIC (2025) "Survey on the latest trends in ICT, R&D, and digital utilization in Japan and other countries"