AI Discussion Paper (Version 1.0)

Preliminary Discussion Points for Promoting the Sound Utilization of AI in the Financial Sector

March 2025

金融庁
Financial Services Agency

I. Introduction

AI has historically cycled through numerous boom and bust periods. However, with the dramatic leap in performance, generative AI has finally at the stage where it is beginning to be widely implemented in society. Going forward, AI is expected to greatly improve efficiency and convenience in various fields, including industries such as finance. AI also has the potential to significantly enhance the quality of life for citizens and contribute to the development of the national economy. In the financial sector, alongside the utilization of conventional AI for applications like fraud detection, market analysis, forecasting, and marketing, the proliferation of generative AI is leading to the emergence of use cases that further enhances operational efficiency and enhanced customer experiences.

When the Internet was commercialized about 30 years ago, the use of the Internet for financial transactions was almost unthinkable due to the problems of slow communication speed, security, and fraudulent use. However, with the development of information and communications technology, it is now one of the core technologies of finance. While AI also needs to address many issues, it has the potential to become one of the core technologies supporting financial services in the medium to long term, similar to the Internet and cloud services. If AI is a technology that has the potential to fundamentally change the way financial services are provided and the business models of FIs in the future, it could also completely change the competitive environment for the FIs. For instance, pre-trained generative AI can be implemented with relative ease. This overcomes a disadvantage of conventional AI, which requires the development of custom models with self-prepared training data—a process which can be a significant hurdle for smaller FIs with limited resources. By utilizing general-purpose functions for various tasks such as document preparation and system development, AI is expected to overcome issues such as labor shortages and to achieve significant improvement in productivity. FIs need to have management-level discussions in regard to their future business models. These discussions must consider the accelerating changes happening now and project a vision for the future.

In Japan, there is a strong societal awareness of the risks associated with generative AI, such as its potential for misuse in criminal activities and the spread of false and misleading information. The ratio of Japanese people who believe that AI can be safely used under current regulations and laws is significantly lower

compared to other countries s[1]. This low level of confidence hinders the proactive utilization of AI. Under these circumstances, challenges have been identified in the financial sector, including how to ensure the transparency, accountability, and fairness of complex AI systems such as generative AI, as well as how to control the risk of misuse into financial crime and the potential impact on the stability of the financial system. While the importance of addressing risks cannot be overstated, it is also noted that AI is likely to bring benefits that greatly outweigh these risks. Additionally, it is important to fully recognize the "risk of not taking actions," which includes the risk of being left behind in technological innovation and the subsequent difficulty in providing high-quality financial services in the medium to long term. Under a risk-based approach, we hope to see steady progress in AI initiatives that boost customer convenience and efficiency, with the right controls for each use, and with strong leadership from management. We expect that FIs will identify and assess the risks associated with AI utilization and take appropriate measures to actively create new financial services and improve operational efficiency.

From this perspective, the FSA will strive to develop an environment in which FIs can confidently engage in AI-driven initiatives. Beyond AI, in the course of promoting innovation initiatives, the FSA will ensure that administrative actions do not unduly discourage FIs. We will address any issues through dialogue and other means to facilitate problem-solving. We will also make efforts to provide safe harbors by clarifying the application of regulations. It should be noted that the FSA's basic stance is technology-neutral and that existing laws and regulations apply regardless of whether or not specific technologies such as AI are used. However, if necessary, the FSA will review relevant laws, regulations, and guidelines, taking into account the characteristics of AI. Given the rapid technological innovation in AI, we believe it is important to flexibly assess policy approaches through dialogue with FIs. In line with this policy, we prepared this Discussion Paper (hereafter, "the Paper"). The FSA is determined to take the lead in controlling risks while dedicating our full efforts to creating an environment where the Japanese financial sector can develop to lead the world.

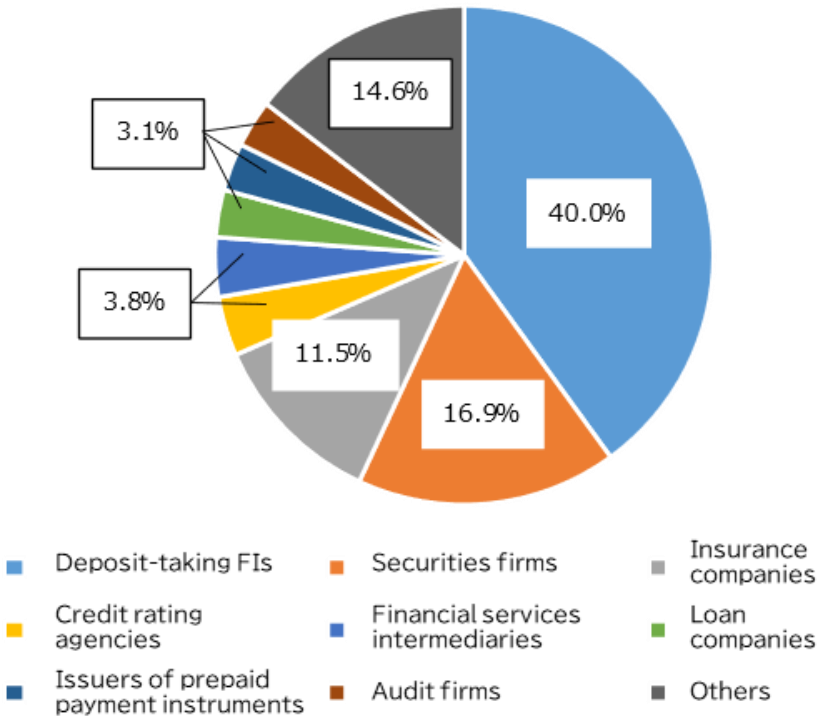
---

[1] Cabinet Office, AI Strategic Council / AI Institutional Study Group Interim Report (February, 2025)

II. Purpose and Positioning of the Paper

This Paper does not express the FSA's monitoring viewpoints or the specific actions expected of FIs. Rather, it summarizes use cases, challenges and governance examples for traditional and generative AI, based on the results of the FSA's "Survey on Use and Risk Management of Generative AI by Japanese Financial Institutions" (October 3 to November 15, 2024) and interviews, as well as progress in international discussions. This report outlines initial discussion points and FSA policy regarding AI in FIs. The survey covered a wide range of financial sectors and received responses from a total of 130 firms. Approximately 40% were deposit-taking Fis. This was followed by Financial Instruments Business Operators ("FIBOs") and insurance companies, each accounting for a little over 10%. The above three categories make up for about 70% of the total.[2]

Figure 1: Distribution of respondent FIs by business type



Due to the voluntary nature of the survey, the variability in response rates across different business sectors, and the absence of significant differences in responses across sectors as the questions were not tailored to specific sectors, a detailed analysis by business type or size was not conducted. On the other hand, we have included numerous responses obtained through open-ended comments and interviews in a manner that does not allow individual companies to be identified. To

[2] FSA "Survey on Use and Risk Management of Generative AI by Japanese Financial Institutions" (October, 2024)

understand the AI utilization landscape across diverse FIs, including small and medium-sized FIs, we analyzed various use cases and challenges, distinguishing between conventional and generative AI. We hope that this information will serve as a useful reference for the future initiatives of FIs in a way that is tailored to their specific circumstances.

Conventional AI in this Paper refers to AI that learns characteristics and trends by being provided with data in advance (for example, machine learning), and obtains answers to input data (including even rule-based models and chatbots that create and operate complex rules from data). Generative AI refers to models with large parameters, such as LLM[3], that have the function of generating new products such as documents, images, audio and video by using data and content (unstructured data such as text and images) on the Internet for training.

While Chapter IV highlights various challenges, it is important to note that these are based on preliminary analysis. As the degree of risk varies depending on the use case and implementation method, the chapter should not be interpreted as requiring all mentioned challenges to be addressed before AI can be adopted. FIs are expected to take on challenges proactively without being unduly intimidated by challenges and risks. The Paper is positioned as a foundation for examining the FSA's policies and the state of AI governance in FIs over the medium to long term. Specifically, it provides an overview of current AI use cases and risk management mechanisms within Japanese FIs. The FSA intends to further explore these issues through ongoing dialogue with businesses and will take necessary actions such as updating this Paper and clarifying the application of regulations.

[3] Large Language Model, in general, refers to a language model that is developed by large data set using deep learning technology.

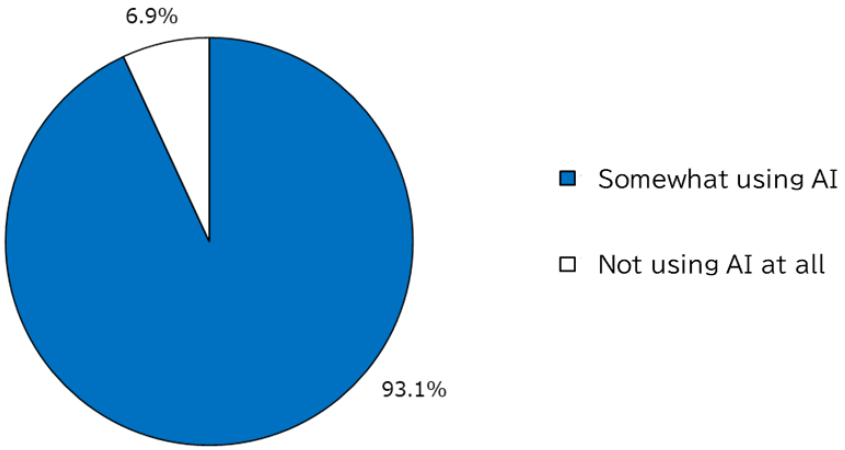
III. Potential Applications and Use Cases of AI in Finance

AI has great potential to be utilized by FIs that handle a large volume of data including documents and images. For example, as customer needs and preferences become increasingly diverse, FIs can leverage AI to move beyond the provision of uniform financial products and services. By analyzing customer-related data such as transaction history, FIs can offer personalized financial products, thereby enhancing the provision of customer-centric service. Furthermore, using technologies such as RPA to automate routine tasks like document creation can significantly reduce costs and improve operational efficiency. Additionally, there is considerable scope for AI to be utilized in asset management; namely, using alternative data to enhance asset management and refine market forecasts. In response to increasingly sophisticated fraud techniques, it is also possible to strengthen risk management and compliance through early detection of abnormal patterns that are difficult to identify manually or based on rules. Whether or not AI is introduced will affect the competitiveness of the FIs in a society where digitalization is progressing. Therefore, FIs should proactively consider the use of AI with the proactive involvement of senior management.

The survey results revealed that the use of AI is already progressing at majority of the FIs. In this chapter, we classified AI into conventional AI and generative AI and arranged the status of their implementation for each use case. First, we discovered that more than 90% of the FIs surveyed are using either conventional or generative AI in some way. Although these findings are based on participants to this survey, they suggest that various FIs and Fintechs are already incorporating the use of AI in their operations[4].

[4] Survey conducted by Bank of Japan during April to May 2024 shows results of around 60% of FI already using conventional AI and around 30% of FI using generative AI. The results show that use of AI is rapidly spreading; around 60% of FI are using/trial phase of using AI and around 80% of FIs are using/trial phase of using AI/considering using AI. Source: Bank of Japan 金融システムレポート別冊「金融機関における生成 AI の利用状況とリスク管理」：日本銀行 Bank of Japan (October, 2024)

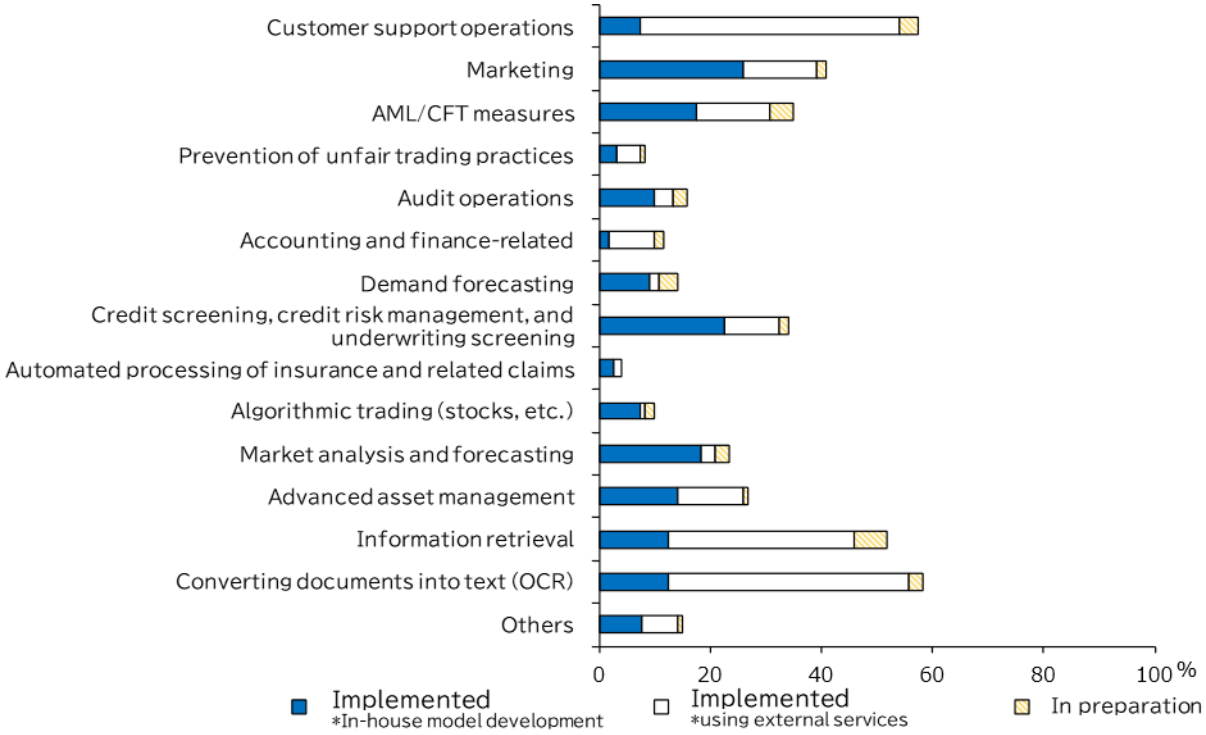
Figure 2: FIs using conventional AI or generative AI



1. Key Use Cases for Conventional AI

According to the results of a survey on status of conventional AI by use case, the majority of usage is for purposes such as optical character recognition (OCR), customer service, information search, and marketing. Focusing on these use cases, this chapter will introduce specific examples of application, referring also to the results of interviews.

Figure 3: Status of conventional AI by use case



① Streamlining Operations

It's clear that more than half of the respondents have already introduced text conversion of documents (OCR) and information search, indicating that the use of these tools to contribute to operational efficiency is expanding. In the case of OCR, use cases for text extraction from document files include the digitization of application forms and identification documents received from customers during account opening, as well as the text conversion of financial statements and PDFs received from business partners via hand delivery or fax. Information search included the introduction of an internal chatbot to efficiently search documents related to internal procedures. Some FIs are developing their own models for these use cases, but the majority of FIs use external services. This is probably due to the relatively easy integration of external services compared to other areas. On such example is cloud accounting providers offering OCR functionality to their customers. This ease of implementation contributes to the higher adoption rate observed compared to other items. Some FIs are also considering combining external systems with generative AI due to performance challenges of external systems.

② Use in Customer Service

AI is increasingly used in customer-related operations, including service and marketing. In the context of customer support operations, the use of chatbots for handling inquiries has been particularly noted. They analyze input questions and provide responses based on the closest match to pre-defined questions. Next, the following usage methods were confirmed for marketing: use as a supplementary tool for sales staff (for example, machine learning of customers' ages and transaction status to create efficient sales lists, identifying busy periods for customers and refraining from sales activities during those periods, etc.), provision of reliable services to customers through optimization of personnel allocation (for example, AI demand forecasting in call center operations to realize optimal allocation of people at the necessary time, etc.), and transmission of personalized messages to customers through applications.

③ Advancement of Risk Management

AI is also used to a considerable extent in initiatives to enhance compliance and risk controls, such as AML/CFT and credit scoring, risk management, and underwriting. Regarding the use of AI in AML/CFT operations, some FIs are exploring sophistication of detection rules in transaction monitoring operations by

having AI learn historical transaction data. Some FIs responded that they have decided not to use AI in actual operations because they could not obtain sufficient accuracy with their own in-house development. However, some FIs are considering using AI together with rules-based fraud detection system, some FIs are jointly developing models with other FIs in the same sector, and some companies have started or are preparing to provide services that contribute to the enhancement of transaction monitoring and name screening operations using AML/CFT data from multiple FIs. Observed AI applications in credit scoring, risk management and underwriting include the following: loan repayment history-based scoring models, guarantee assessment models based on past data, purchase data-driven credit scoring, news/social media based early warning systems, and delinquency/default estimation. While model accuracy hinges on data quality and quantity of training data, many FIs found that compared to fraud detection, their abundant internal historical data enables relatively high accuracy with AI-only judgments.

Furthermore, AI applications for advanced compliance and risk management include the following: detection of fraudulent insurance claims and use in audit operations (e.g., detection of inappropriate accounting treatment, scoring of internal e-mails with AI to help detect fraud), and the detection of compliance violations in sales activities (e.g., extraction of suspected compliance violations from records of meetings held during sales of financial products, etc.)

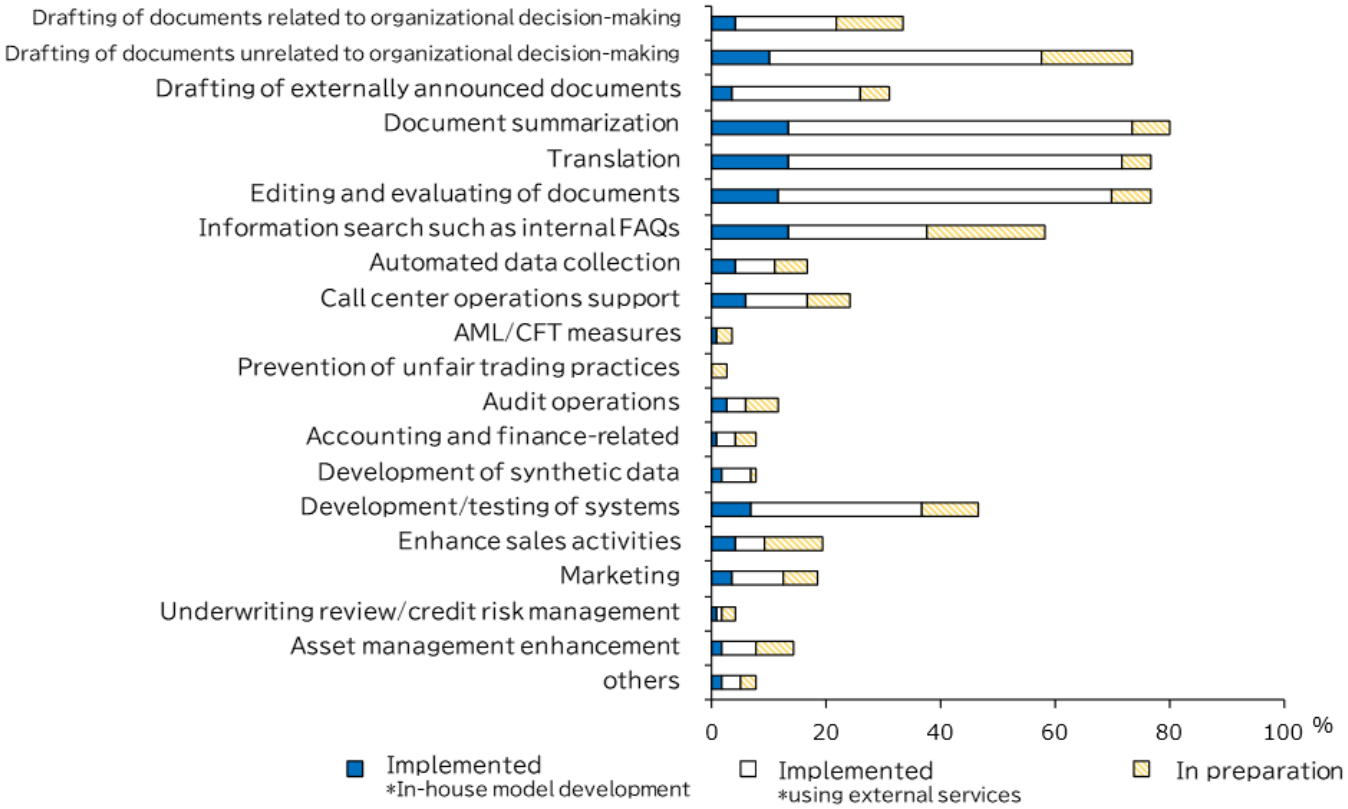
④ Market Forecasts and Others

AI is widely used in asset management, securities, and forex. Examples include exchange rate and interest rate forecasting, portfolio optimization, enhancing investment strategies, and understanding real-time market sentiment through NLP analysis of news sites and social media posts. It was confirmed that the use of alternative data, such as satellite imagery, geolocation, traffic, and supply chains, is increasing to capture leading indicators that are difficult to grasp through financial information alone. These data types have historically been underutilized, and their application is expanding.

2. Status of Generative AI Adoption and Its Main Use Cases

Next, we will analyze in more detail the status of generative AI by FIs and major use cases. First, we will explain the general status and then introduce specific use cases in three categories.

Figure 4: Status of generative AI adoption by use case

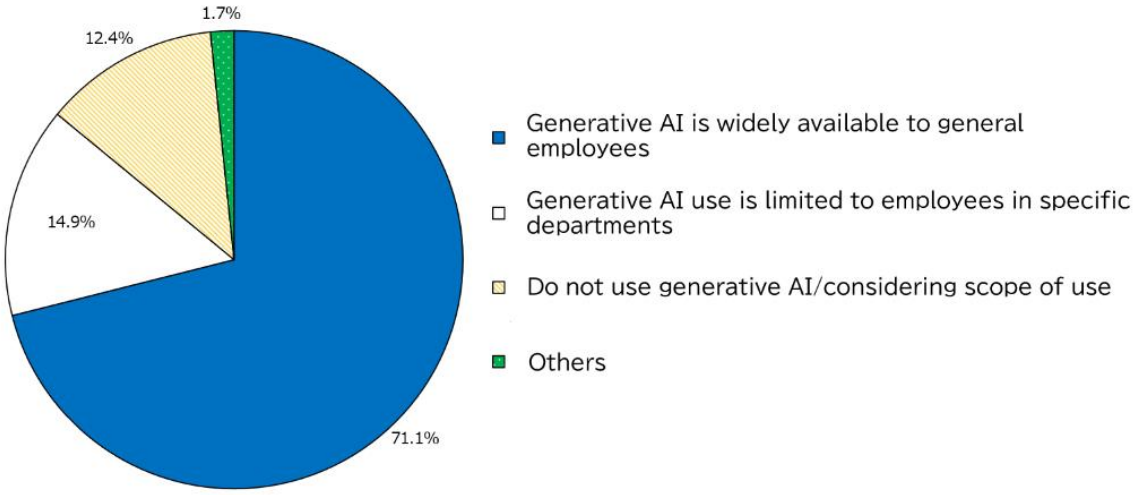


① Scope of Generative AI Usage

Generative AI has different characteristics from conventional AI, and the form of its use by FIs is inevitably different.

One of the major differences is the high degree of versatility in generative AI. Conventional AI models have been developed and operated according to specific applications. This often involves staff with a certain level of expertise, such as data scientists and engineers from the digital strategy division, the IT system division, and the risk management division. On the other hand, generative AI has potential to be used as a tool to improve the efficiency of daily operations of staff, such as drafting, translation, and summarization of documents. The majority of FIs that have already introduced generative AI have approved its widespread use.

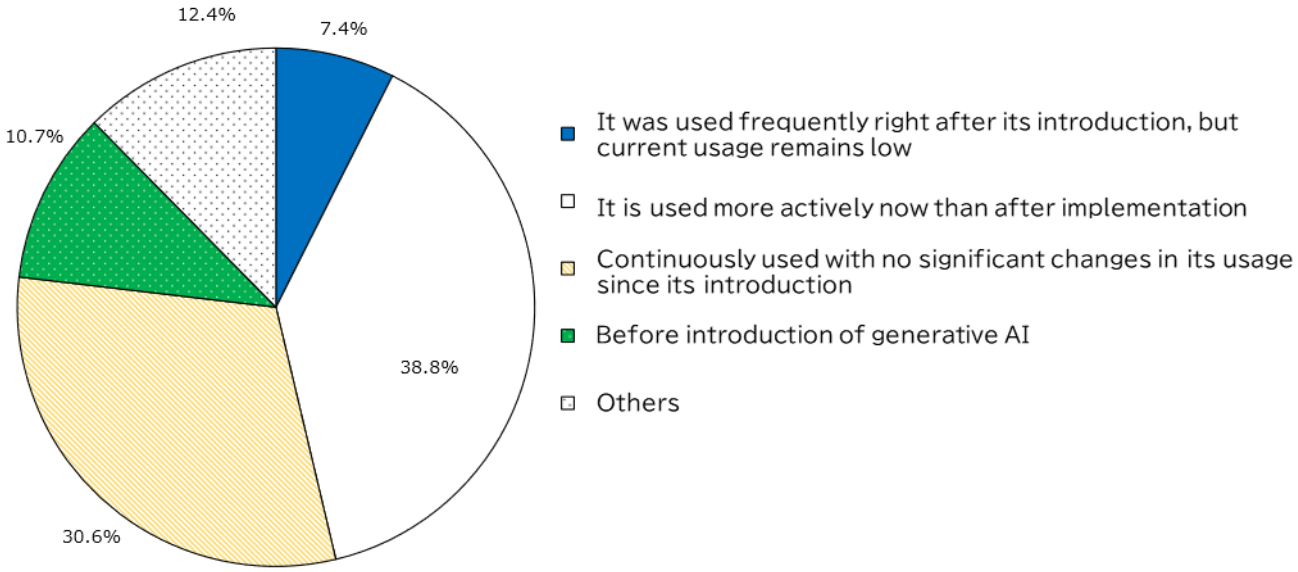
Figure 5: Status of generative AI usage



When we inquired about the use of generative AI to companies that had responded that they use AI in some way, we found that more than 70% of those companies widely allow the use of generative AI by general employees. A closer look at the responses to the survey revealed several different approaches from the viewpoint of cost effectiveness, such as companies that allow the use of generative AI by all employees, companies that adopt an application system, companies that allow the use of generative AI only by the headquarters, and companies that set different scopes of use depending on the type of generative AI.

In regard to the use of generative AI after its introduction, the results show that a large number of FIs either continued to use generative AI or are now using it more actively than immediately after its introduction. Numerous FIs have only introduced general-purpose generative AI without any customization (see the next section), but the fact that they are still actively using it supports the high possibility that generative AI will be used to improve the operational efficiency of FIs. Some FIs responded that the use of generative AI has not progressed as much as expected due to challenges such as prompt creation skills. Nevertheless, a lot of FIs responded that the use of generative AI is spreading through internal study sessions, idea contests, and other initiatives.

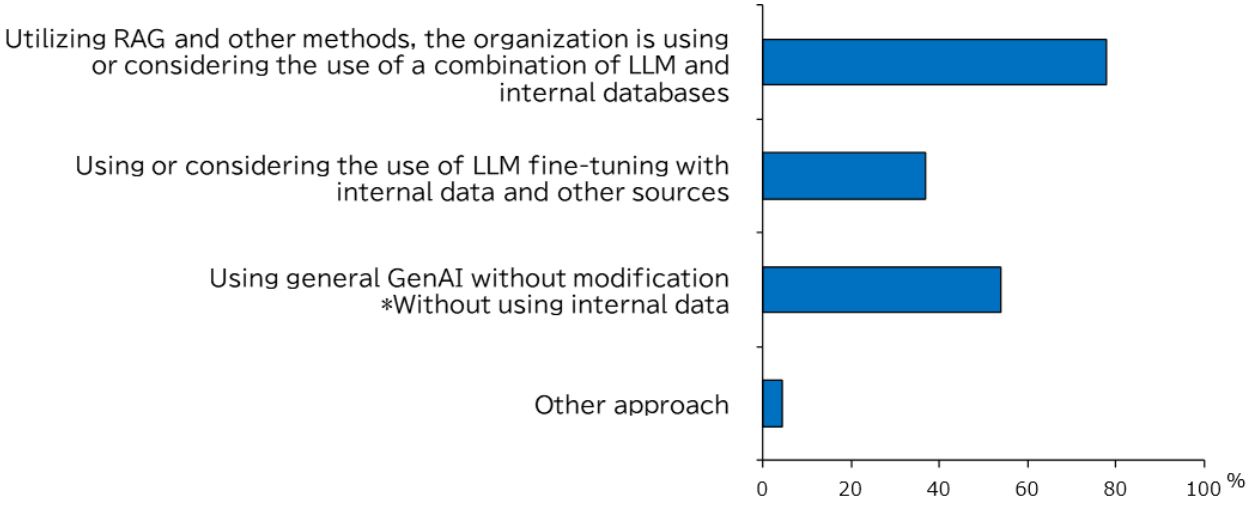
Figure 6: Status after the introduction of generative AI



② Presence of Customization and Implementation methods

LLM is a pre-trained model. This means that FIs do not need to build a model from scratch like conventional AI and can immediately utilize general-purpose functions with the minimum settings. Consequently, the barriers to AI adoption, including initial investments and securing specialized staff, have been reduced. The survey results indicate that approximately half of the surveyed financial institutions and others are using generic generative AI as-is. Reflecting these characteristics, the overall adoption rate of generative AI is higher than that of conventional AI. (For example, less than 60% of FIs have introduced generative AI for customer service and OCR but more than 70% of FIs have already introduced general-purpose generative AI, such as text summarization and translation.)

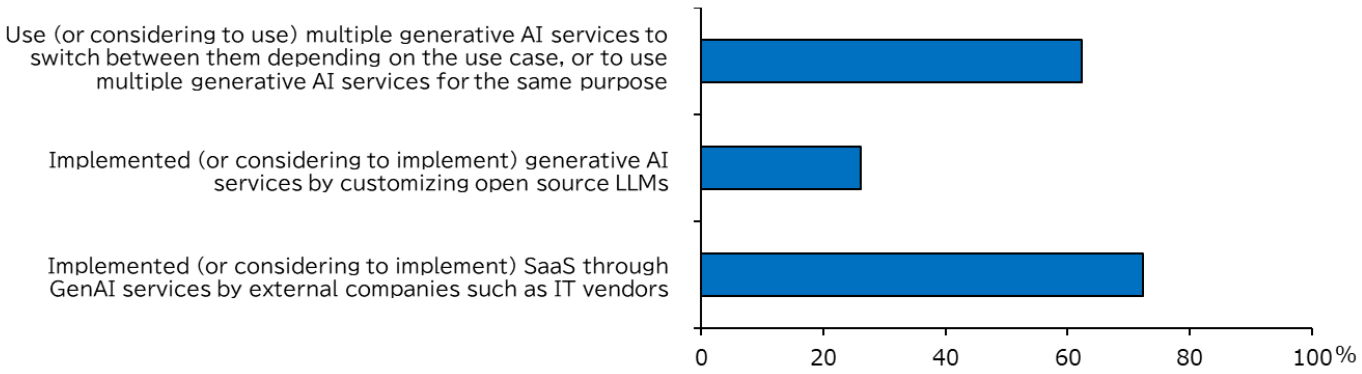
Figure 7: Forms of introducing generative AI (with or without customization)



On the other hand, many organizations are using or considering using LLMs provided by vendors in combination with internal databases through RAG (Retrieval-Augmented Generation)[5] and fine-tuning[6]. Compared to conventional AI, generative AI, which is increasingly multimodal[7], has more room to use unstructured data such as text and images owned by FIs and alternative data provided by external vendors. Various FIs are exploring the possibility of using unstructured data in collaboration with vendors, Fintechs, and consulting firms.

There are also many companies that use multiple generative AI depending on the use case. Examples include using a general generative AI provided by a major cloud provider for all employees, while employing another provider's generative AI for specialized RAG that is specifically tailored to certain tasks and used as the main environment for in-house development. Furthermore, while LLM has both open source and proprietary models, there are companies that use LLM as open source in on-premises servers, although it takes time to build their own environment. On the other hand, there are companies that use a proprietary generative AI service in a closed environment via a dedicated line. It was also found that the adoption of generation AI is not uniform, with some companies adopting generative AI tools provided by external vendors as SaaS and others developing their own user interfaces.

Figure 8: Forms of introducing generative AI (including use in SaaS and others)



5 RAG generally refers to a technology to improve accuracy of the output by combining text generation by LLM with retrieval of external information.

6 Fine-tuning generally refers to retraining trained LLMs with specific data sets and delicately modifying the parameters in the model.

7 In this context, multimodal refers to generative AI analyzing and developing by combining different data formats such as text, image, audio, and video.

With the emergence of various fundamental models and generative AI services and tools, as well as frequent model updates, we get the impression that FIs are currently experimenting with various approaches rather than sticking to one implementation approach.

Box 1. Various approaches toward building a task-specific model

While generative AI is highly efficient in answering general-purpose questions and preparing documents, general-purpose models alone do not provide sufficient accuracy in cases where responses should be made based on the rules and product knowledge specific to FIs and the past transaction histories of customers. Therefore, there is a move to construct a business-specific generative AI model by utilizing technical methods such as RAG and fine-tuning.

■ Examples of RAG application

RAG is a method of generating responses through a combination of LLM and information searches. RAG aims to improve the accuracy and reliability of responses by acquiring the latest information from external databases and incorporating it into model responses. To implement RAG, it is necessary to build a vector database and optimize the search algorithm, and the advantage is that relearning of the model itself is not required.

- One financial institution is testing a system that centrally manages internal rules, product brochures, and FAQ for customers on the cloud, and uses RAG to capture relevant information and generate answers to inquiries.

■ Examples of fine-tuning application

This refers to fine-tuning the parameters of a model over a specific data set, with a particular emphasis on consistency of output shape. It contributes to stabilization of response quality because it is fine-tuned by supervised learning. Fine-tuning involves training data of sufficient quantity and quality and a certain computational cost.

- One financial institution is taking steps to improve the accuracy of chatbot responses to customer inquiries through the use of data on past customer interactions and fraudulent claims for financial services.

There is also a method of encouraging business-specific responses in the model by devising prompts without customizing the LLM. For example, by including internal rules and a summary of the financial products handled in the prompt, it is possible to respond with a certain level of expertise without customization. Another approach, known as in-context learning, involves strategically including examples and key points in the prompt so that the model can refer to them while generating responses. Since these methods do not necessarily involve relearning or linkage with external databases, they have the advantage of reducing implementation costs and data maintenance burdens. On the other hand, the information included in the prompt tends to be lengthy, and the challenge is to organize and optimize it so that the model can understand it accurately.

③ Three Types of Use Cases

A) As described above, the introduction of generative AI is still in the trial-and-error stage, but it is possible to classify use cases from the following perspectives. Purpose of use: Internal use such as business efficiency improvement, or use in customer service

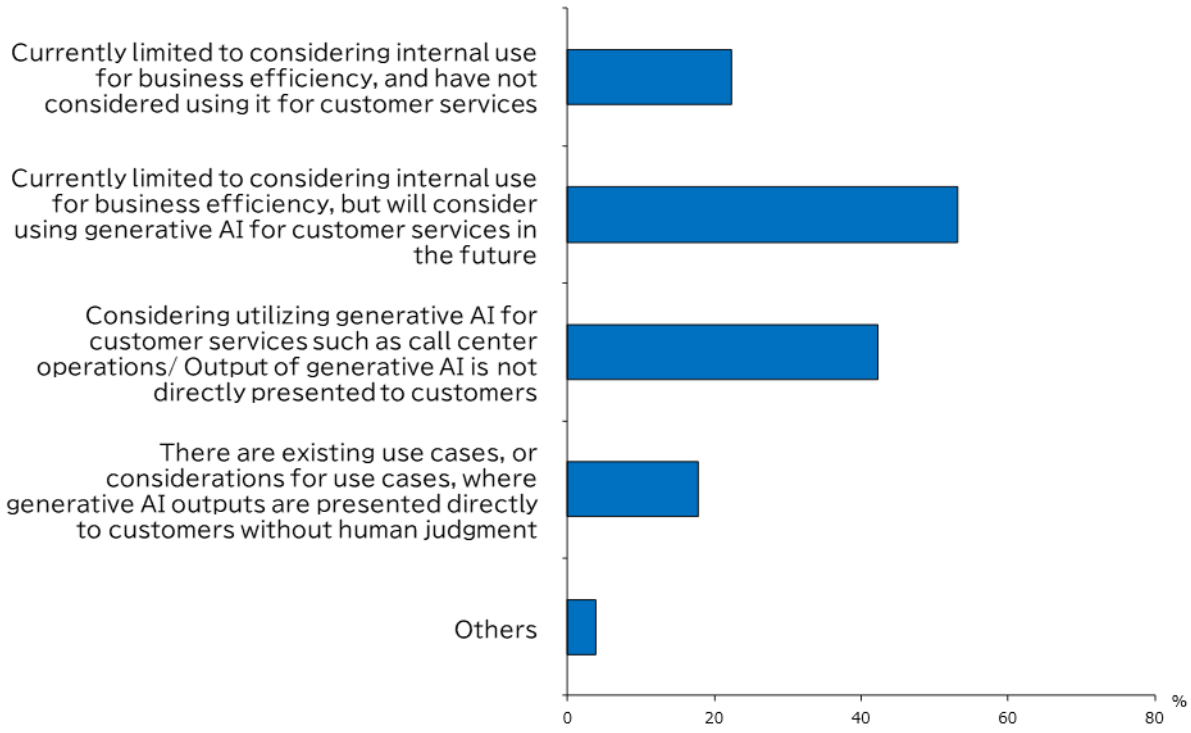
B) General or special-purpose: Use of the general-purpose generative AI or special-purpose generative AI using RAG

C) Product: natural language, such as text, or something other than natural language, such as program code or images

In the Paper, we classify the main use cases of generative AI at this stage into the following three categories in line with the perspective A).

- Internal use such as business efficiency improvement
- Indirect use in customer service
- Direct use in customer service

Figure 9: Utilization of generative AI in customer service



According to the results of the survey, most of the respondents answered that at present, the use of generative AI is limited to internal use for operational efficiency.

In many cases, off-the-shelf AI solutions, which are relatively easy to introduce, are first introduced and used for use cases such as document drafting, summarizing, translation, and correction. More than half of the respondents answered that they are only using such AI internally at present but will consider using it for customer services in the future. Several FIs are working for more developmental use with the aim of developing new financial services. There are also a considerable number of firms that have introduced generative AI for customer services such as call center operations, but in consideration of risks such as hallucination, the output of generative AI is not directly presented to customers, and the majority of use cases are through human judgment. However, there are also firms that believe that advances in technology will alleviate hallucination and similar challenges in the future, and that use cases such as directly presenting generative AI outputs to customers could expand in the medium to long term.[8]

i. Internal Use Such as Business Efficiency Improvement

The results of the survey revealed that more than 70% of FIs have already introduced the three use cases of “document summary,” “translation,” and “proofreading, correction, and evaluation.” Given the large volume of documents prepared and held by FIs, document use cases are the most common for internal use. Many organizations also utilize generative AI for the creation of routine documents. Specific examples include summarizing publicly released information, preparing minutes (transcription of online meetings), translating foreign languages (including those other than English), and proofreading for typographical errors and omissions. In the preparation of text and translation, it is possible to adjust the tone of text according to the expected reader. For example, when preparing text for regulators, it is possible to instruct the preparation of text in a formal style. The versatility is already realized at a high level. Additionally, many respondents indicated that they utilize generative AI for brainstorming during the consideration of new business ventures or document preparation.

Another typical use case is internal FAQs. Around 40% of FIs have already introduced an information search function. Similar use cases exist for conventional AI, but there are many instances where generic generative AI has been implemented in internal chatbots. In some cases, specifications have been

[8]The "Handbook for Practical Application of Generative AI in FIs" published in May 2024 by the Financial Data Utilizing Association (FDUA), a general incorporated association, classifies the level of generative AI utilization into three levels. In Level 1, individuals utilize ChatGPT’s generative AI internally. Level 2 consists of building applications in specific fields by incorporating internal information using the RAG mechanism. Level 3 is the provision of services using generative AI to external customers.

developed to allow the generation of answers by utilizing RAG to reference unique internal data such as internal rules and regulations. In regard to such customization, we have observed its application not only for data already shared among all employees, such as internal regulations, but also for the purpose of visualizing the tacit knowledge and expertise of highly experienced staff who are well-versed in company operations. Some respondents emphasized the significance of actively working to pass on personal know-how and tacit knowledge to junior employees, as each company is aware of the shortage of personnel.

Furthermore, the use of AI in fields other than natural language, such as system development/testing, has already expanded to a certain extent. Specifically, the use of AI to assist coding in programming is often heard. Some interviewees point out that the use of generative AI has achieved significant results in this field, and AI is being recognized as an effective utilization method.

Box 2. Specific examples of initiatives towards widespread internal use and application of generative AI

- To provide an environment where employees can use generative AI daily, a personable generative AI assistant was added to Slack, the main internal communication tool. Equipped with practical capabilities to support tasks, it is also used as a casual advisor.
- Instead of web page that needs to be opened every time, the generative AI functionality was embedded in the collaboration tool to improve usability. It is used for research, translation, transcription, etc.
- Build various tools utilizing generative AI in the cloud environment. The FI utilizes not only general-purpose AI functionalities but also deploys a document-referencing GPT that generates draft answers based on internal documents. It has also introduced an add-in that enables generative AI access from document creation applications, across the entire organization.
- A generative AI chatbot was introduced on a vendor-provided platform, allowing users to interactively ask questions about administrative procedures. AI summarizes the content of the bank's administrative manual and responds along with the source of the information, thereby reducing the need to call a dedicated call center for inquiries.

ii. Indirect Use in Customer Service

A typical use case in this category is call center operations support. There is a belief that generative AI can enable accurate, automated customer service, thereby addressing past issues of understaffed call centers and missed customer needs. However, at present, it is difficult to completely resolve the risk of hallucination as described below, and most companies have determined that immediately using generative AI directly provided to customers is unfeasible. However, there are already widespread examples of generative AI being used as an aid by call center staff during customer interactions, and the use and application of generative AI in customer service is being explored. Such usage offers limited additional benefits to staff who are already proficient in the job, but it is a great help to staff who are newly engaged in the job, and some say it has contributed significantly to a decrease in the turnover rate of the entire workplace.

It also became clear that some progress is being made in document development operations that can have an impact on customers, such as drafting documents related to organizational decision-making and preparation of external publication documents. For example, several FIs have cited use cases such as drafting request for approval with reference to past request for approval and other related materials. The effects of shortening the preparation time and improving the quality of request for approval are expected. Generative AI is also used for review of proposal document. A proposal document can have an important impact on customers, such as loan decisions, and most operations require human involvement (human in the loop).

iii. Direct Use in Customer Service

Given the risk of hallucinations and the fact that FIs are still working on streamlining their governance systems regarding generative AI, most FIs are taking a more conservative approach to the use of generative AI in fields where services are directly provided to customers. However, some Fintech providers have already started providing services that directly present the output of generative AI to customers, such as life-plan advice.

Box 3. Use of AI in audits

- The development and implementation of AI tools is also progressing on a global level in the area of audit, which plays a key role in ensuring the reliability of corporate disclosure information in capital markets. While auditors are required to search, understand, and analyze vast amounts of financial and non-financial information, the development in the use of AI is expected to improve auditors' operational efficiency and data analysis. It enables auditors to focus more on tasks that require evaluation and judgment, and thus, contribute to the improvement of audit quality.
- In performing audit engagement, conventional AI has been used as a tool to detect anomalies in transaction and journal data, to identify and assess fraud risks, to search for audit firm's internal information such as audit related standards and operation manuals, and to convert documents into text (OCR). To further improve efficiency and enhance the audit engagement, some audit firms are introducing or testing tools that incorporate generative AI. Regarding generative AI, a trend is observed for audit firms to implement or consider implementing tools that assist auditors in summarizing, translating, and proofreading documents.
- On the other hand, audit firms also noted a variety of challenges regarding the use of AI. One common issue is the difficulty in ensuring the quantity and qualitative training data due to insufficient data standardization and accumulation of fraudulent accounting data. Furthermore, it has been pointed out that generative AI presents challenges that are unique to generative AI or that are more difficult than that of conventional AI. Each audit firm is working on to establish control systems, such as introducing additional rules and restrictions on the use of generative AI. Also, as common with financial institutions, audit firms noted hallucination (see IV-③-i) and the difficulty of ensuring explainability in the output process as major challenges regarding the generative AI. Audit firms are taking initiatives to enhance accuracy and explainability of AI tools by implementing or considering to implement tools to address such challenges and by verifying the results through back testing.
- The impact of advancements in AI on audit quality is also discussed at IFIAR, an international organization comprised of audit regulatory authorities around the world. Especially regarding the global networks of large-sized audit firm, they are basically promoting the development and implementation of AI tools and establishing related guidelines, at a global level. Therefore, it is crucial for the regulatory authorities across jurisdictions to monitor audit firms' use of AI in a cooperative manner. Taking into account the trends observed in international discussions, the FSA will engage in dialogue with audit firms and the Japanese Institute of Certified Public Accountants, as well as collaborative research with academia, to understand how the AI tools are used by audit firms and what their challenges are to take necessary actions.

IV. Initial Considerations for Promoting the Use and Application of AI by FIs

In this chapter, we provide an initial overview of the major challenges facing AI utilization. The overview is based on surveys and interviews with FIs, as well as discussions on AI in Japan and overseas, including at international organizations such as the FSB. In addition, as some FIs and related entities have begun initiatives to overcome these challenges, we also provide examples of specific initiatives. Furthermore, the direction of the FSA's future actions for promoting the sound use and application of AI through the establishment of AI governance is summarized. These descriptions are based on preliminary analysis at the time of writing, and do not immediately require FIs to take any specific action.

1. Major Challenges Related to the Use of AI by FIs and Examples of Initiatives to Overcome These Challenges

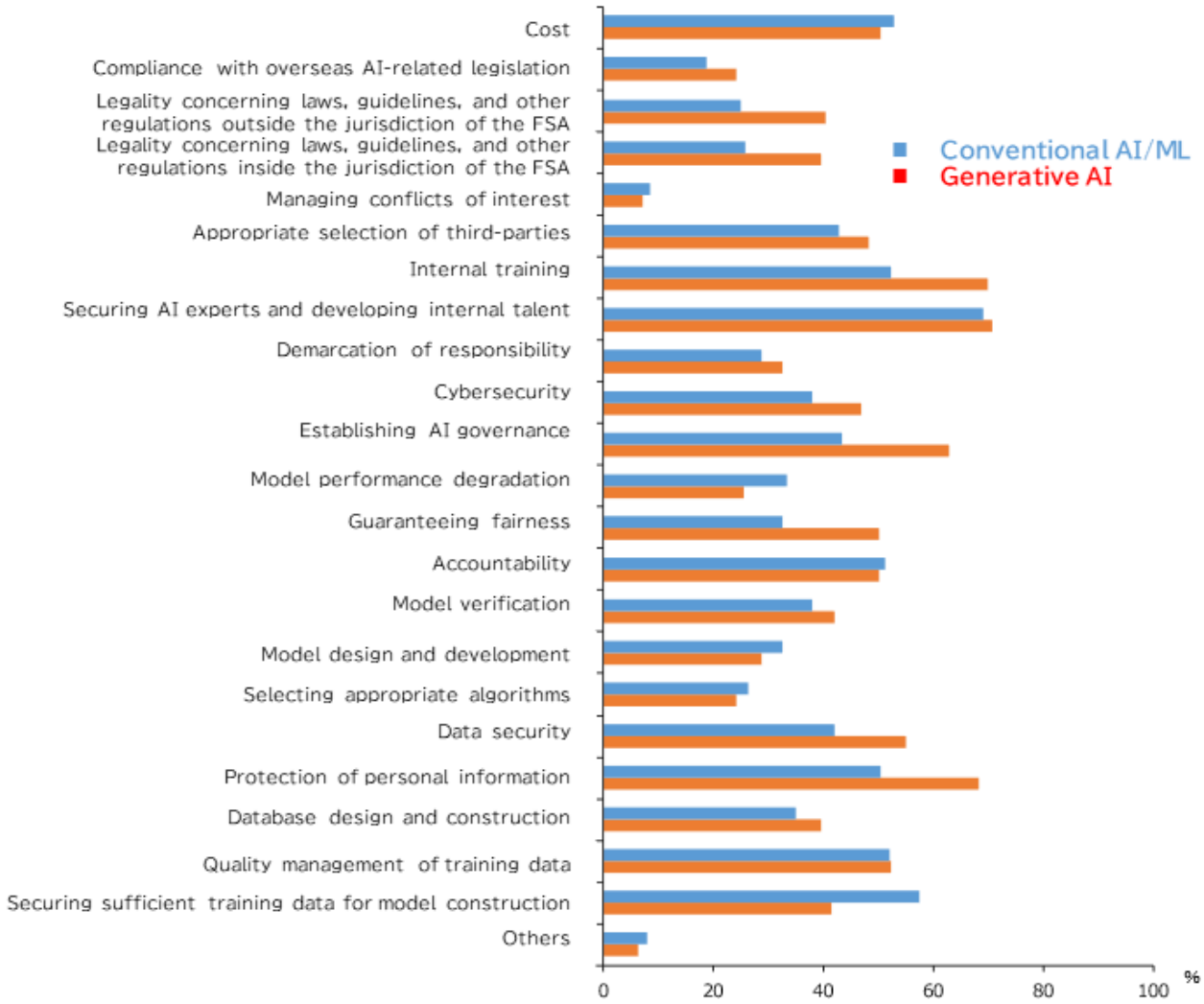
When we conducted a survey asking about the challenges in using conventional AI and generative AI, FIs cited data maintenance and other issues as common challenges. On the other hand, some issues, such as ensuring the explainability of the output results of AI models, were more frequently raised with generative AI. There are also unique challenges to generative AI, such as hallucinations. Therefore, in this section, we organized the challenges by dividing them into the following three categories.

① Common challenges for conventional AI and generative AI

② Issues made difficult by generative AI

③ New challenges created by generative AI

Figure 10: Issues in AI consideration, introduction, and use



① Common issues for conventional AI and generative AI, and examples of initiatives to resolve these issues

i. Data Preparation

In response to the digitalization of society, the increase in non-face-to-face transactions, and the diversification of customer needs, FIs have continued to make efforts to utilize data in order to improve business efficiency and create new financial services. Numerous FIs are developing data infrastructures using cloud services and APIs, and are promoting open innovation through collaboration with external businesses such as Fintech companies. Furthermore, data utilization is crucial for enhancing risk management and internal audits in FIs. The FSA has also emphasized the importance of data utilization by FIs through legal revisions related to the

promotion of open innovation, as well as the publication of discussion Papers[9] and monitoring reports[10].

Against this backdrop, the rapid development and widespread use of AI, including generative AI, is further increasing the importance of data in FIs. As described in some of the use cases in the previous chapter, using internal and other data as training and inference verification data has the potential to improve the competitiveness of FIs.

In the case of generative AI, internal data is not necessarily required when pre-trained LLMs provided by external vendors are used directly. However, many believe that generic generative AI models that are not optimized for specific business processes have limitations in terms of business efficiency improvements. Therefore, the importance of customizing foundational models using internal data in line with specific business processes, customer needs, and business models is recognized. As seen in the previous chapter, many organizations are attempting customization through RAG, fine-tuning, and in-context learning.

On the flip side, a great number of FIs face challenges in internal data management. Training and customization of AI require the development of high-quality data infrastructure, including internal data. In particular, the utilization of complex AI, including generative AI, requires ensuring the quality and quantity of data, appropriate training of models, and advanced governance related to information security and personal information protection. However, around half of the surveyed FIs responded that they face challenges in securing sufficient training data for model construction, and quality control of training data. This indicates that data management is ongoing and requires further work.

---

9 For example, Discussion Paper for Dialogues on Practices of IT Governance at Financial Institutions (June, 2023)

10 For example, Monitoring Report for Improving Internal Audits of Financial Institutions (September, 2024)

Main challenges related to data preparation raised by FIs

- The database was not built on the premise that internal data, such as data referenced by RAG, will be utilized by AI. Therefore, it will take time and cost to respond to internal consensus developing and coordination with vendors.
- Seamless connectivity between the integrated cloud data infrastructure and other databases is a challenge.
- Due to data quality issues (e.g. legacy company policy) we tried using RAG but were unable to improve LLM accuracy.
- The training data needs to be updated when laws, regulations, or internal rules are changed, and reviewing the content takes time.
- Even if we want to collect training data on fraud cases, the actual number of fraud cases is small, so our own data is not enough.
- First, we need to move away from Excel-based business processes.

To overcome these issues, various FIs are moving forward with initiatives such as developing databases suitable for AI utilization and securing sufficient training data. For example, some FIs and related entities have built a platform on the cloud to centrally collect and manage internal data, which is connected to the customer management system via an API to enable data sharing across departments and groups, allowing them to extract and analyze data according to specific needs. Other FIs have implemented solutions for efficient data access control and version management. Additionally, some are analyzing unstructured data, such as text and speech, using OCR and speech recognition.

FIs need to recognize that there are growing opportunities for AI-driven improvements in operational efficiency and the sophistication of financial services. Recognizing the potential of the vast amount of data, including unstructured data owned by FIs, are likely to become important management resources for improving profitability and transforming the business model. Therefore, FIs need to consider the development of data utilization platforms and API integration as important management issues while ensuring security.

ii. Collaboration with External Vendors and Risk Management

Expertise is required not only for data management as described above but also for the development, operation, and management of AI models and the customization

of generative AI models. Therefore, many FIs are utilizing or considering utilizing third party solutions and platforms. For example, there is a growing trend to build efficient and safe AI utilization environments through collaboration between AI implementation support vendors, consulting companies, data vendors that provide training data, and cloud service providers. According to the survey, more than 40% of FIs responded that "selecting the appropriate third party (external vendors)" is an issue for both conventional AI and generative AI. For complex AI such as generative AI, it is not practical for majority of FIs to develop a model by themselves, and they inevitably use a fundamental model provided by an external operator.

At the same time, some FIs point out the overdependence on external providers as potential drawbacks. To introduce appropriate AI systems that take into account the business environment, business processes, and customer needs of the FI, it is necessary for staff at the FI to have a certain level of knowledge about AI even in collaboration with external vendors.

Appropriate risk management of external vendors is also important. An FSB report[11] published last November identified AI-related vulnerabilities that could increase systemic risk, including reliance on external vendors and concentration of certain service providers.

Major issues raised by FIs regarding collaboration with external business operators and risk management

- Conventional AI technologies are increasingly being integrated into external services. This makes it difficult to understand and manage them effectively.
- Even when limited to the system aspect, there are multiple parties concerned such as basic designers, programmers, and system maintenance managers. It is necessary to define the scope of shared responsibility for each product and manage it correctly; for example, whether each element should be produced in-house or outsourced (there is a possibility of outsourcing to multiple business operators).
- Since it is impossible to make LLM completely in-house, it is necessary to connect to some external service. However, it is a challenge to establish a secure communication means with external services. Connecting with other companies creates a risk of data leaks and falsification. Therefore, a robust security mechanism is essential to protect confidential internal and customer information.

[11] The Financial Stability Implications of Artificial Intelligence (November, 2024)

To address these issues, FIs are conducting security checks by applying existing third-party risk management frameworks. Some organizations use open-source platforms to avoid vendor lock-in. Others contractually ensure that AI model IP remains with the FIs when outsourcing development using internal data to external vendors. Moreover, there were cases in which FIs, mainly large FIs, are internalizing the development, operation, and management of AI models.

Collaboration with external operators is essential when using complex AI, including generative AI. It is necessary to select appropriate collaboration partners based on the needs of the entity, while addressing risks associated with collaboration with external operators. These include risks related to new products and services, system failure risks, and information security risks (including cybersecurity risks).

iii. ROI (Return On Investment)

FIs noted challenges in demonstrating ROI for both conventional and generative AI, citing difficulties in estimating benefits and the time required to reach an internal consensus. The importance of establishing IT investment management processes is a matter common to not only AI but also IT systems in general. Some FIs commented that (i) AI systems, including generative AI, are likely to become obsolete in a short period of time due to the rapid progress of technology, (ii) the effects of AI systems, including generative AI, may expand for the medium to long term even if the effects are limited in the short term due to the importance of continuous model training, and (iii) the scope and impact of AI may change depending on the degree of social acceptance of AI technology and changes in the level of AI proficiency by the employees. The fact that the effects of AI systems can fluctuate is thought to complicate the challenges further. Furthermore, as generative AI provided by external operators is often charged on a usage basis, it is difficult to estimate the ROI because the cost will increase without restrictions depending on the frequency of use.

The responses to these issues are similar to those of digital transformation efforts. When ROI (Return On Investment) is difficult to measure and it is difficult for each department to secure a budget, the IT department separately allocates a budget. In other cases, projects are managed across the group to prevent multiple departments from making similar investments. While projects and businesses with potential future revenue can be evaluated using ROI indicators, and a PDCA cycle can be implemented

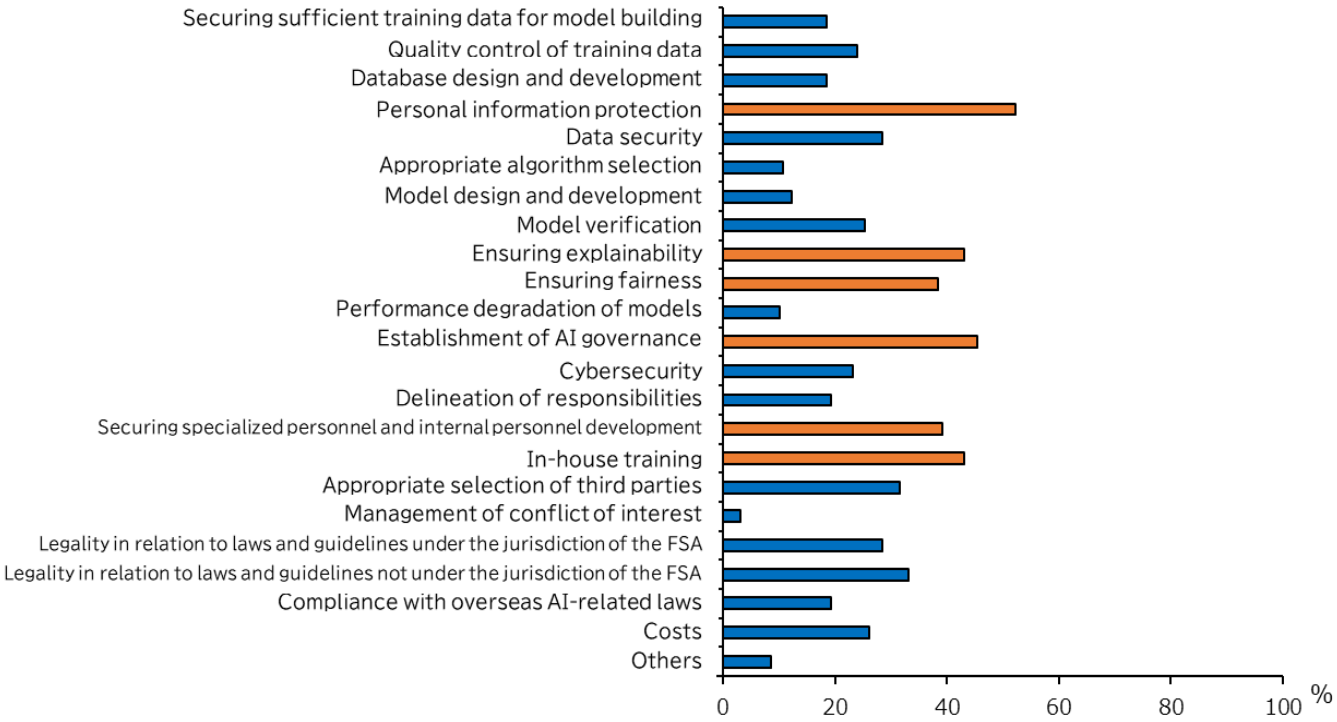
to determine resource increases or service reductions/withdrawals based on progress, strategic projects requiring a proof-of-concept stage or a medium-term perspective, such as AI-related investments, may need to set KPIs tailored to user and usage characteristics and monitor performance against the plan until revenue prospects are established.

Some FIs also expressed the view that it is important for management-level discussions and decision-making to include not only numerical values but also the visualization of long-term investment objectives as much as possible. When introducing rapidly evolving AI-related technologies and services, it is important to make strategic investment decisions with the appropriate understanding and proactive involvement of management.

② Challenges Made Particularly Difficult by Generative AI

Next, although conventional AI and generative AI are broadly grouped into the same issues, it has become clear from the results of surveys and interviews that there are multiple issues that FIs believe have been made more difficult by generative AI. Therefore, this section will cover these main issues.

Figure 11: Challenges made difficult by generative AI



i. Ensuring Explainability

Even in the case of conventional AI such as deep learning, the problem of "black boxing," in which it is difficult to explain the basis for inference (output) due to the complexity of algorithms, has been pointed out. Especially in consumer financial services, when actions based on AI outputs result in adverse effects for consumers, it is necessary to provide adequate explanations. For example, the U.S. Consumer Financial Protection Bureau (CFPB) published guidance on consumer loans in September 2023[12]. The guidance stated that lenders must provide accurate and specific reasons for decisions that are disadvantageous to consumers (for example, refusal to extend credit), even when complex algorithms or black box credit models are used.

Many argue that ensuring explainability is even more challenging with generative AI compared to conventional AI. Generative Ai requires a complex training process that uses numerous parameters. Furthermore, it dynamically generates outputs in response to multi-modal inputs (prompts) such as text, voice, and images. It is extremely difficult to clearly indicate why each inference was answered. For this reason, numerous FIs that are exploring the use of generative AI are faced with the challenge of addressing the lack of explainability.

Major challenges on explainability raised by FIs

- When models such as deep training are used, the process of outputting results creates a black box. This can make it difficult to respond in the internal audit when abnormal data is detected.
- If AI-recommended prospect lists are presented to sales branches without explaining why those prospects were recommended, there's a risk that the lists will not be prioritized or even used.
- Generative AI has a more opaque generation process compared to conventional AI. Consequently, this raises concerns about unintentional copyright infringement and the risk of employees receiving incorrect information due to hallucinations.

It is not easy to ensure the explainability of complex AI models such as generative AI. However, we confirmed that some initiatives are being taken to resolve these issues, including responses to conventional AI models. For example, we found entities using AI model evaluation tools from external vendors to

12 CFPB Issues Guidance on Credit Denials by Lenders Using Artificial Intelligence | Consumer Financial Protection Bureau (September, 2023)

visualize which parameters contribute significantly to their credit assessment models and verify the validity, entities developing technology to enhance the explainability of their fraud risk detection models, and entities adding rationale columns to sales prospect lists to ensure credibility before their usage.

Given the extreme difficulty of ensuring complete explainability with generative AI, especially those with vast parameter counts, it's crucial to focus on developing trust with relevant stakeholders, such as customers and employees, depending on the use case. Since the degree of social acceptance of AI also seems to affect the sense of trust towards AI, it is important for FIs to provide information to the extent necessary and technically possible. When providing this information, FIs must take into account use cases, risks, and the social context of AI.

ii. Fairness/Bias

Bias in AI training data, algorithms, and even model inference results increase the risk of unfair treatment of customers with certain attributes. These problems can occur not only in consumer finances such as credit scoring mentioned in the previous section, but also in internal operations such as personnel evaluation and recruitment. Therefore, they are issues that may affect the creditworthiness and social responsibility of FIs.

Fairness and bias are also major issues in conventional AI. But methods that use vast numbers of parameters, such as LLM, tend to make it difficult to examine why bias occurs in inference. Plus, there is a risk that models may strengthen bias at the training stage when training is performed without sufficient training data, or when the training data itself contains a lot of information that is biased toward specific regions or attributes. Concerns have been raised that these structural problems could inadvertently result in discriminatory results.

To address these issues, some point out that it is important to establish a framework to detect and address biases inherent in models at an early stage by utilizing model evaluation tools provided by external operators. However, while it is not always clear what specific evaluation indicators and audit processes should be established, there are various bias detection tools, and it is difficult to determine their accuracy and operating costs. Therefore, the establishment of a consistent framework to detect and address biases in AI models at an early stage is a challenge. If AI models trained on historical data are found to have fairness or bias issues, it's

important to recognize that similar problems may have existed in past decision-making processes, and this should be seen as an opportunity to review existing business processes.

iii. Development and Operation of AI Systems, and Model and Risk Management

As with other models, it is important to appropriately manage risks associated with AI models[13]. In doing so, considering the "risk-based approach" in the Principles for Model Risk Management[14], we believe it is appropriate to assess the risks inherent in models, including risks specific to AI, and manage risks according to their magnitude. At present, the level of model management varies across FIs, but it may be advisable to establish a certain level of risk management framework, such as managing a list of important AI models used internally. One example is the model inventory management described in the Principles for Model Risk Management.

However, many FIs face difficulties in testing AI models. Some FIs that have used conventional AI reported that they have developed a model and risk management framework for AI to a certain extent. However, since systematic performance evaluation indicators have not been established, even such FIs are still in the process of developing framework for development, operation and management of generative AI. Generative AI has characteristics that are significantly different from conventional AI. The former features very high versatility for use, can generate information that is not included in training data, can output different results for the same input, and its fundamental model is provided by an external provider and frequently updated. Some FIs believe that they cannot grasp and manage all the risks with the same method as before. Some FIs commented that they have no choice for the time being but to sort out potential risks for each FI and manage them through trial and error, while expecting the authorities or private entities to develop and update guidelines in the future.

---

[13] FSA Progress report on enhancing risk management and models of the FIs (December, 2024)

[14] FSA Principles for Model Risk Management (November, 2021)

As mentioned earlier, in light of issues such as the biases in AI model inference and the difficulty of ensuring explainability some FIs consider it essential to develop a framework to identify potential risks of AI models and reflect them in release management and continuous monitoring.

Major issues raised by Fis regarding development, operation, and risk management

- Generative AI's characteristic of producing varying outputs even with identical prompts necessitates more rigorous testing, ongoing monitoring, and an incident response plan.
- Unlike conventional AI, generative AI does not have an established method or axis for evaluating performance and its high versatility makes it difficult to objectively assess performance. As a result, it is currently difficult to select an appropriate third party based on systematic evaluation.
- The difficulty of verifying the generative AI model leads to the difficulty of investigating the cause of the incident and clarifying the boundaries of responsibility. In the event of a large-scale incident, it may cause serious confusion and make it difficult to resolve the problem.

Several FIs have provided responses regarding their initiatives to establish model risk management frameworks that address complex AI models, including generative AI. For example, (i) FIs confirmed that they have developed a model risk management framework applicable to AI models and are continuously strengthening existing policies, procedures, and controls for generative AI, (ii) others also confirmed that they are conducting PoC tests of tools provided by multiple specialist vendors to verify the accuracy of AI models, (iii) FIs confirmed that they have documented AI model methodologies, including assumptions, limits, and uncertainties, (iv) FIs also implemented a system where model developers and independent verifiers conduct verification before production deployment and continuous performance monitoring of models, (v) FIs also confirmed that they are adding features based on business knowledge and striving to maintain model performance.

As mentioned above, FIs are at various stages in their efforts to establish AI model management systems, and it was observed that they are exploring and experimenting with difference approaches. The FSA will continue dialogue to ensure proper risk management is being conducted, taking into account the evolving

discussions on AI both domestically and internationally, as well as the actual usage of AI.

iv. Protection of Personal Information

Personal data protection was the top challenge cited by FIs due to emergence of generative AI, and it remains a key issue in AI development, operation, and management.

For example, there are concerns that it is difficult to understand the application of rules in the case of handling personal data, including customer information, as training data, and in the case of inputting prompts that include customer information when using generative AI. Furthermore, there are concerns about the clarity of rule application when outsourcing the development and training of AI models to external vendors or when using generative AI platforms with overseas servers.

In this way, many FIs see the unclear application of regulations as a challenge.

The following are initiatives to address these issues were confirmed through questionnaires and interviews. First, we observed that some FIs have issued internal administrative notices to all employees, including a notice from the Personal Information Protection Commission, to ensure that personal data is not used as input into generative AI. We also confirmed that some FIs are considering automatically limiting the input of customer data. We also found that some FIs have established an environment in which data is not used for relearning by setting up a dedicated section to prevent input data from being used for relearning. Some FIs have clearly stated the scope of data handling by concluding appropriate contracts and memoranda when linking with external vendors and systems. Other FIs responded that they are thoroughly managing logs so that they can be traced quickly in the event of an incident, and that they are considering making it possible to view the content of inputs efficiently and accurately.

FIs are required to take both appropriate measures to protect personal data and effectively utilize data, including personal information. The FSA will continue to understand the efforts of financial institutions through ongoing dialogue. Additionally, under the AI Strategic Council, AI Institutional Study Group has been established. In an Interim Report published February this year[15], various ministries

[15] Cabinet Office, AI Strategic Council / AI Institutional Study Group Interim Report

should continue to handle existing laws and guidelines regarding the infrastructure and services (basic services) that form the basis of national life and economic activities, including some financial services. It also stated that if new risks become evident and cannot be addressed within the existing framework, it should be clarified how to interpret the related framework and consider revising or establishing new systems. To add, in February of this year, the Personal Information Protection Commission announced that it will continue discussions with stakeholders to organize the appropriate form of consent for individuals in the case of AI development deemed to be for statistical creation[16]. Looking ahead, the FSA will consider necessary actions based on these discussions and frameworks.

v. Information Security and Cybersecurity

Various issues have emerged regarding the utilization of AI by FIs in the areas of information security and cybersecurity. Some FIs have expressed their views that "the existing IT governance framework is sufficient for security-related matters" and "the effects of the diversification of AI services' algorithms and differences in release cycles are minor and can be addressed by examining existing business processes." Conversely, there is a deep-rooted recognition that new risks may arise due to the introduction of generative AI. The FSB report notes that AI, including generative AI, can enhance attackers' capabilities and increase the likelihood and impact of cyberattacks on the financial sector. A report released by the U.S. Treasury Department in March of last year[17] noted that existing risk management principles provide a framework for the safe operation of AI by FIs but emphasized that advances in AI technology are increasing cybersecurity threats and require FIs to be more vigilant than ever. Therefore, this section focuses on issues specific to generative AI.

First, there is the risk of information leakage. As discussed in the previous section of personal information protection, when FIs use generative AI, there are risks of customer information and important business information being leaked unintentionally. Additionally, when connecting to cloud-based AI services, there is a potential risk that confidential and customer information will be transmitted to the service provider. Consequently, there are concerns about the risk that prompts

---

16 Personal Information Protection Commission, "Regarding the interpretation on institutional challenges of the Personal Information Protection Act." (October, 2025)

17 U.S. Department of the Treasury "Managing Artificial Intelligence-Specific Cybersecurity Risks in the Financial Services Sector" (March, 2024)

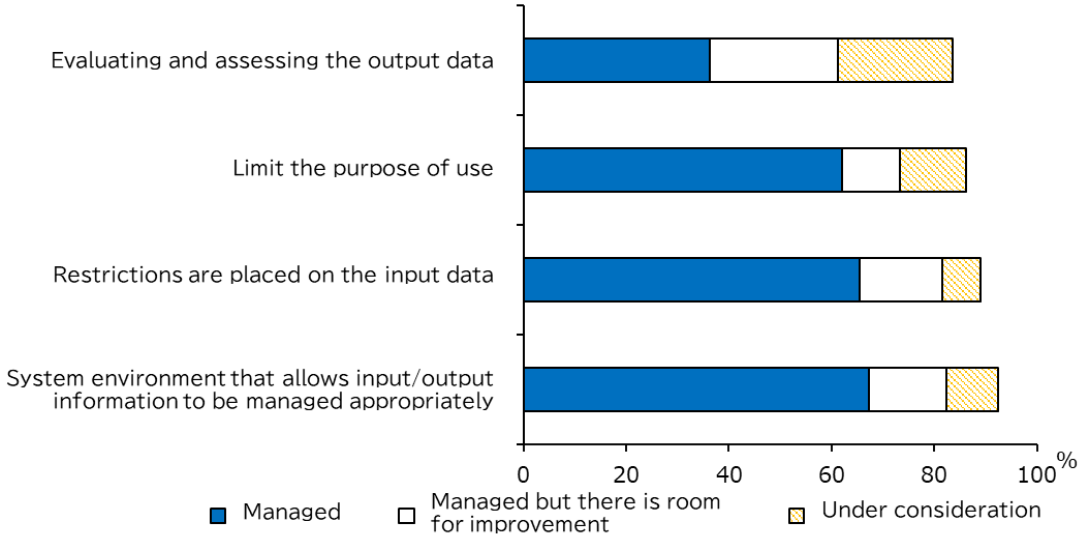
and output content may be used as the training data by the external vendor. When using platforms with overseas servers, managing international data transfers becomes a critical issue. In response to these risks, some FIs are considering mechanisms to automatically restrict the input of confidential and personal information on their systems by masking the data. However, it is not possible to completely block input. Ultimately, FIs have no choice but to rely on employee education and thorough implementation of operational rules.

Furthermore, attacks such as prompt injection can cause AI systems to malfunction or leak confidential information. Not only can sophisticated phishing e-mails and identity theft be easily created by generative AI, but AI models themselves can also be targeted for attack. Specifically, there is a need for vigilance against “data poisoning” in which training data and parameters of AI models are falsified, and for detection and defense measures against attacks in which a large number of queries are thrown to infer the internal structure of models and extract confidential information. Moreover, as pointed out in the FSB report, there are concerns that generative AI could support malware development, impersonation, social engineering, and the like, which could undermine financial stability[18].

Nevertheless, some progress has been made in addressing these issues. For example, as mentioned in the personal information protection section, it was found that most organizations impose certain restrictions on prompt input data, such as prohibiting the use of confidential information. We also confirmed that many FIs have built system environments that can manage input/output information by introducing generative AI services in their dedicated environments. Furthermore, some organizations responded that the first and second lines of defense are already collaborating to establish frameworks, methods, and rules for the verification and evaluation of output data, and that they conduct verification at the planning and development stages. During the operational phase, they are also considering establishing a system to constantly monitor output data with both human oversight and computer systems.

[18] A document published by the Center for Financial Industry Information Systems Center (FISC), last September, also pointed out three issues related to information security: information leakage, specifications/settings, and monitoring.

Figure 12: Controls over the use of generative AI



As described above, many FIs seem to have introduced guardrails to some extent. Even so, some FIs shared the view that "implementation of guardrails does not completely prevent hostile input into generative AI chatbots." Also, the widespread adoption of AI could lead to increased use of "shadow IT," which refers to the use of external AI services by employees without organizational permission. This could amplify security risks such as information leaks and unauthorized access.

Overall, the information security challenges that FIs face when implementing and operating AI are varied. As such, organizational measures, such as the monitoring of AI model-specific vulnerabilities, are essential. Furthermore, with the continued proliferation of generative AI, there will be an increasing need for the development of more advanced security techniques. It is also necessary to monitor whether the output of generative AI falls under the category of hallucination or copyright infringement. To address such multifaceted risks, it is necessary to take integrated measures, such as conducting periodic vulnerability assessment and penetration tests in cooperation with not only the security division but also the legal, compliance and the model-risk management divisions. Additionally, internal audits should be carried out by the internal control division based on the level of risks.

As pointed out in reports by the FSB and the U.S. Treasury Department, AI also contributes to the enhancement of cybersecurity. For example, by analyzing large volumes of data in real time and detecting unusual patterns and fraudulent activities, it is possible to identify attacks, malware, and fraud at an early stage. On top of that, if AI can quickly analyze incident information and take appropriate

measures, it can minimize the impact of incidents and support early recovery. Yet, the utilization of AI requires a deep understanding and expertise of AI, and FIs need to formulate and implement strategies that balance the development of AI technology with the enhancement of security.

vi. Securing and Developing Specialized Personnel and Internal Education

A great number of FIs are aware of the need to recruit and train experts in the development, operation, and management of complex AI models, as well as the need for internal education as the number of employees who use AI systems increases. It has been pointed out that internal education is becoming increasingly important for generative AI because users are diverse. Furthermore, due to variations in individual understanding of AI and anxieties about their roles being replaced by AI, many FIs express difficulty in generating use case ideas and establishing cross-departmental collaboration.

There is a deep-rooted concern that when development and operation are advanced using external vendors, know-how and intellectual property are unlikely to be accumulated within the organization, and internal human resource development will not proceed. A prominent challenge is the increasing shortage of AI experts such as data scientists and engineers, as well as personnel capable of selecting optimal solutions and continuously operating and managing those solutions, as well as personnel who can bridge the gap between frontline and the IT department. There is a growing sentiment that traditional educational content is no longer sufficient, as the rise of generative AI has introduced new risks, such as AI-specific hallucinations and copyright infringement, which demand careful consideration.

Major issues related to experts and internal training raised by FIs

- There are few AI experts in the company, and they often face difficulty with theme setting and analysis design.
- Individuals vary in their level of understanding of AI, and ideas for use cases do not advance. Plus, it is difficult to obtain active cooperation due to concerns that AI will deprive employees of their jobs.
- The scope of users of generative AI is wider than that of conventional AI, so we recognize that internal education will be more important, but there are challenges in incorporating it at the practical level.
- When utilizing external resources for development, a foremost issue is that knowledge and intellectual property do not remain within the organization, and internal human resource development does not proceed as expected.

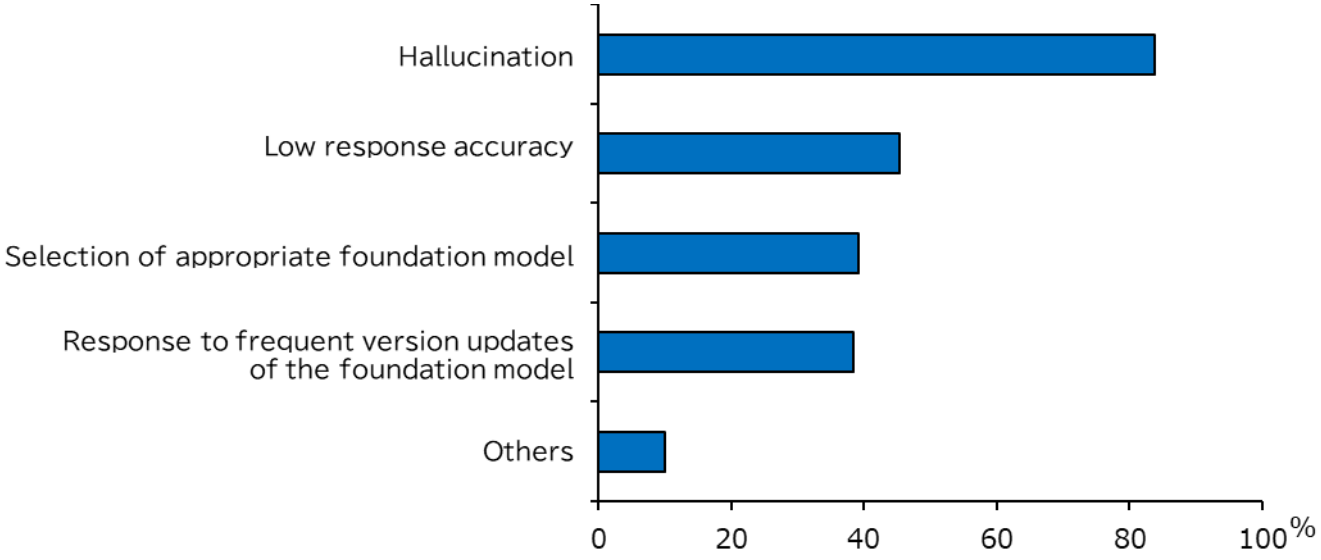
While various FIs perceived this issue as a major challenge, the surveys and interviews confirmed a variety of initiatives aimed at resolving this issue. Regarding the recruitment and development of professional human resources, the following were found: companies that are aiming for internal human resources development through OJT while seeking support for AI model development from part-time professional human resources (professional human resources); companies that are working for reskilling of highly skilled internal human resources such as SE graduates; companies that are expanding the recruitment of new graduate professional human resources; and companies that plan to actively recruit professional human resources by establishing a dedicated team to promote the use of AI. Regarding internal education for general employees, the following were found: companies that are implementing educational measures such as supporting the acquisition of AI-related qualifications; companies that are sharing practical examples of AI use with call center staff; and companies that have introduced educational programs on the risks of generative AI for all employees. As a more developmental example, there are companies that have established a joint venture with a startup company to develop and use AI in-house and build a development framework to ensure that knowledge and intellectual property remain within the group.

In addition to these individual initiatives, initiatives to develop human resources are also being undertaken by organizations in which multiple FIs are participating. In particular, since there are limitations to improving the expertise of a single small financial institution, this is an area where industry-wide initiatives are expected, such as sharing expertise from institutions that are taking advanced initiatives. As AI adoption progresses, there is a possibility that the type of personnel required by FIs and others will change significantly. Recognizing this as one of the critical management challenges, it is expected that efforts will be made in human resource development and internal education.

③ New Challenges Created by Generative AI

There are also issues unique to generative AI that have not been seen in conventional AI, with hallucination being a typical example. Another important issue is how to deal with the misuse of generative AI in financial crimes.

Figure 13: New challenges related to generative AI



i. Hallucination

The phenomenon of “hallucination,” which is a phenomenon of output that is not based on actual data, is recognized as a major challenge. The credibility of FIs could be undermined if incorrect information is presented or if responses that are directly linked to credit risk or legal risk are incorrectly generated. As mentioned in the previous chapter, numerous FIs are seeking to improve the accuracy of responses through RAG and fine tuning in order to suppress hallucination. However, if the development of reference data and the updating of model training data are insufficient, it will be difficult to suppress hallucination. In particular, even in the case of using RAG, if the business design is not appropriate, the selection of reference sources will be insufficient and the accuracy will not be improved, and there is a risk that operations will be conducted without eliminating false outputs.

In addition to internal use, the challenge of mitigating the risk of employees spreading misinformation externally when using generative AI has also emerged. At

present, there are calls for not only controls on systems but also the introduction of operating rules, including user education and how to create prompts.

Major challenges raised by FIs and others

- Since the impact of hallucination is not zero, it is important to introduce a process of human checking and to make staff understand the uncertainty of generative AI.
- In RAG or fine tuning, if a large amount of old information is mixed in the training information, hallucination is likely to occur because the number of appearances of the old information is larger.

When seeking to leverage generative AI more progressively, such as in customer-facing services, hallucinations present an unavoidable challenge, and numerous FIs are actively exploring mitigation strategies. Typically, business processes are developed with human involvement as a premise, and many organizations set up a review process by humans when introducing generative AI, under the assumption that hallucinations will not be entirely eliminated. For example, our survey has found that when employees respond to customers or agents based on the results of generative AI, they always make a judgment on whether the content is right or wrong before using it. Furthermore, to reduce hallucinations, some FIs make efforts to include supporting documents in their responses by using RAGs and prompts so that the employees can check the basis of the responses (if the search yields no results, provide a response indicating that no information for the basis was found).

As described above, at present, even if hallucination countermeasures are implemented by combining RAG and other measures, many FIs are aware that human judgment is necessary, and the utilization of AI is restrained. On the other hand, just as humans sometimes make wrong judgments, it may not be appropriate to expect an extremely high level of precision from AI; that is, to assume that AI must not make mistakes. Therefore, FIs are expected to explore the utilization of generative AI without excessive restraint, and to consider the degree of social acceptance of AI (magnitude of reputation risk), technological progress, and use cases. The FSA will continue dialogues with FIs to determine whether they are taking appropriate measures. These dialogues will consider trends in technological progress and discussions in Japan and overseas.

ii. Misuse of Generative AI for Financial Crime

The development of generative AI has made it possible to generate natural Japanese texts, sounds, and images. Thus, criminal techniques are becoming more sophisticated, and the risks surrounding FIs and their customers are also expanding[19]. For example, AI-generated emails and phishing site scams are becoming more sophisticated. Deepfake technology makes it easier to carry out scams through fake videos and audio that impersonate specific persons, and the impact is likely to be huge. Also, fraudulent activities such as presenting forged identification documents or false transaction information as if they were genuine are anticipated. From the perspective of countermeasures at FIs, conventional KYC procedures and authentication systems are increasingly challenged in distinguishing between genuine individuals and fraud. These raises concerns about the potential for bypassing identity verification.

As described above, the sophistication of generative AI has the potential to create a number of new criminal techniques that are not envisioned by existing financial crime countermeasures and may pose a threat of a scale that cannot be addressed by simply introducing individual technological measures. There are concerns that the systems and human resources of the audit and compliance departments within FIs will not be able to keep pace with repeated fraudulent acts conducted with unexpected speed and ingenuity. Hence, the risk of damaging the credibility of FIs and the stability of the financial system cannot be ruled out. Therefore, it is important for both the FSA and FIs to remain attentive to these risks.

iii. Other Issues Concerning Financial System Stability

The impact of complex AI systems such as generative AI on financial stability has also been discussed internationally, as the FSB report points out specific AI-related financial stability risks. In addition to the issues mentioned previously, like specific third-party dependencies, cybersecurity, and model risk management, the report discusses risks that could lead to financial market instability, including a herding effect. The widespread adoption of AI in financial markets could lead to herd behavior in which market participants make similar decisions based on AI-generated signals,

---

19 National Police Agency "Regarding the situation of threats in cyberspace for the first half of Reiwa 6." (September, 2024)

which has the potential to amplify market volatility and systemic risk. Also, generative AI could generate disinformation, which could lead to a sudden bank run or a sharp rise in stock market volatility through dissemination on social media, which in turn could heighten financial market unease. The FSB and SSBs continues to discuss these AI-related financial issues, and the FSA considers necessary approaches while engaging in international discussions.

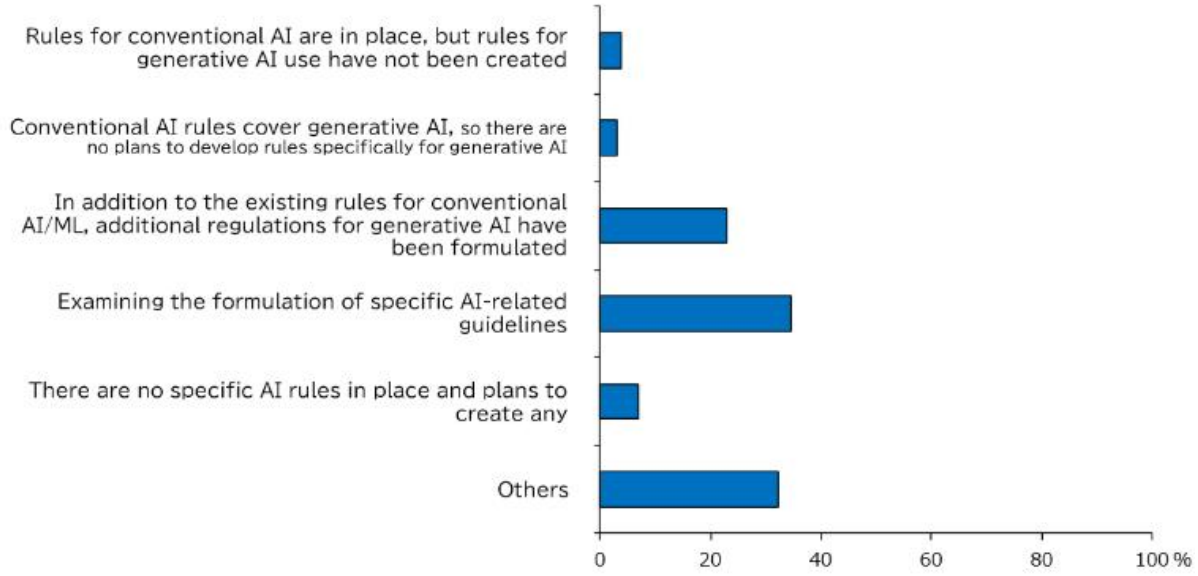
2. Initiatives by FIs toward Developing AI Governance.

This section presents initiatives by FIs and others aimed at developing governance, including the formulation of internal rules, based on the responses to the survey and the results of interviews.

① Formulation of Internal Policies and Rules on AI

In the past, numerous FIs have covered the risks related to AI to some extent through across-the-board application of existing internal rules, such as information management rules and model/risk management rules. However, with the advancement and spread of generative AI, many are calling for the urgent establishment of more specific and flexible rules specific to AI. Several FIs have reported that they had not established rules for conventional AI because there was little experience with its use, but that they have established new rules in line with the introduction of generative AI. Another FI has said that it has established rules specific to generative AI, but that company-wide rules for conventional AI have not yet been established, and it plans to integrate them in the future.

Figure 14: Current status of developing regulations and other guidelines



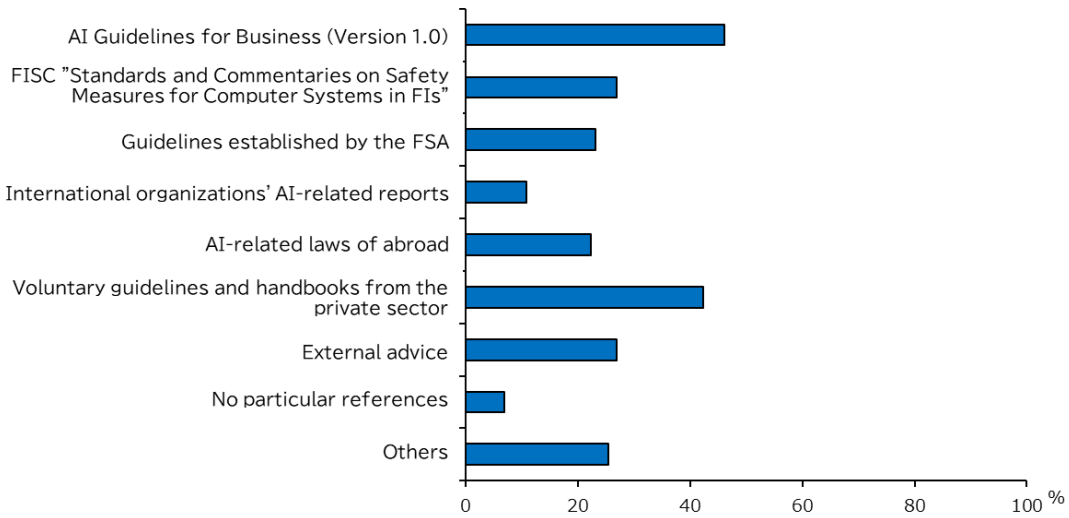
A major global financial group with overseas subsidiaries, is taking the opportunity presented by the emergence and rapid penetration of generative AI to newly establish policies for AI governance and implementation standards on a global basis and to localize implementation procedures in a way that responds to the laws and guidelines of each country. Another financial group has formulated and announced AI

governance guidelines and is working to ensure that all group companies are fully aware of the guidelines. When offering new AI services, this FI has introduced an operation which uses a checklist created in-house to confirm consistency with AI governance guidelines. The aim is to strike a balance between a “try it first” attitude and ethical and legal compliance. Another financial group has a “corporate code of conduct” and “model risk management rules” but is considering revising security check standards and rules in cooperation with overseas entities and experts because it may not be able to cope with the risks specific to generative AI.

Conversely, there are not a few voices saying that the issues are generally manageable with the existing framework. For example, some companies say that “AI is handled in accordance with the content established in the information management regulations, although no provisions dedicated to AI have been established.” Other companies say that it is sufficient to limit the scope of use on a per-business basis and obtain approval from the security and compliance departments when necessary. However, multiple FIs are starting to consider comprehensive rules and governance systems, including conventional AI, in anticipation of the future expansion of the use of generative AI. They are accelerating the launch of pilot risk management by conducting a comprehensive survey of models across their organizations.

Although not included in the selection items of the survey, rather than developing additional generative AI provisions in addition to the conventional AI provisions, there was a noticeable approach to first developing internal rules focusing on the introduction and use of generative AI, and then expanding the application of these rules to conventional AI.

Figure 15: Documents and materials referenced in the development of internal rules

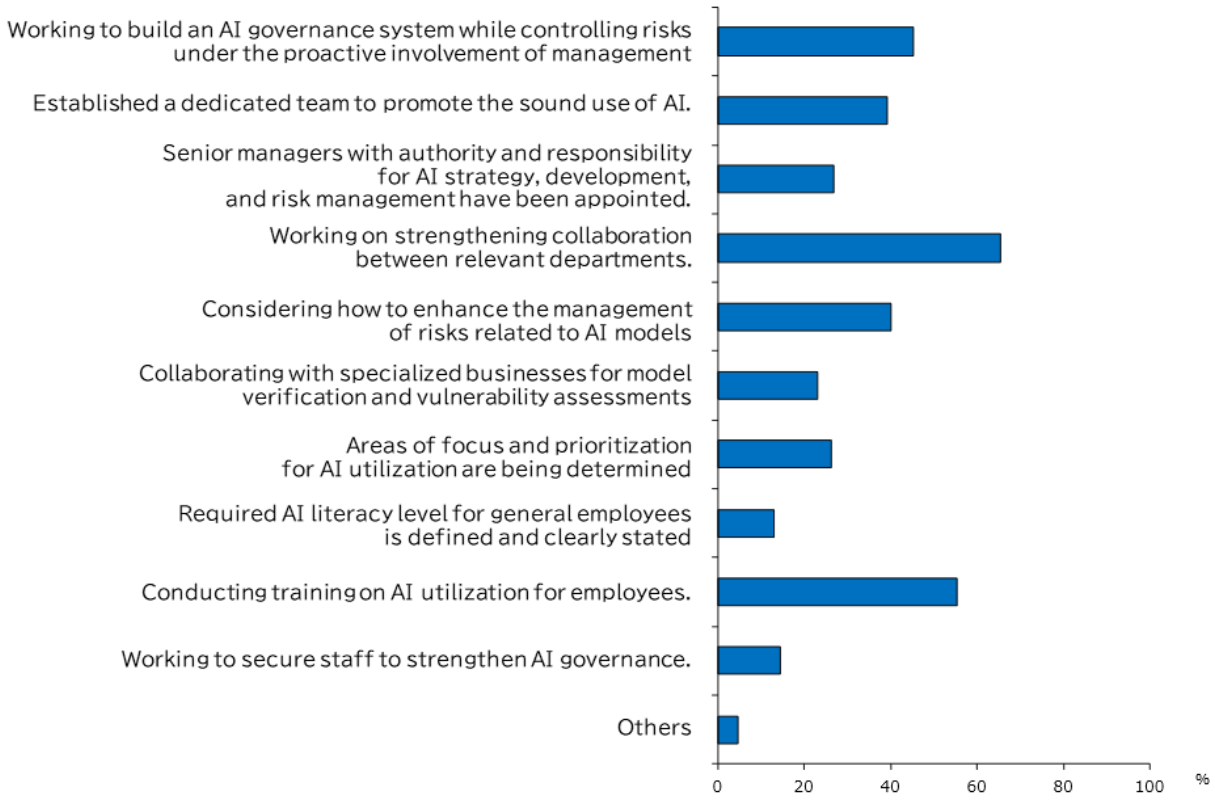


In reviewing the documents used as references during the development of internal rules, many FIs responded that they referred to the 'AI Guidelines for Business Ver 1.0' published by the Ministry of Economy, Trade and Industry and the Ministry of Internal Affairs and Communications in April 2024[20]. Additionally, guidelines concerning generative AI from private organizations, FISC publications, and various AI-related international laws from the EU and others were widely referenced. FIs are taking various approaches to the formulation of regulations, including advice from external experts.

② Other Initiatives for Developing Governance

Although there are some overlaps with the efforts to address the major issues mentioned in the first half of this chapter, we would like to introduce some examples of efforts to build governance, other than internal rule making, focusing mainly on company-wide organizational structure development.

Figure 16: Initiatives towards developing AI governance



[20] AI Guidelines for Business Ver1.01 is published November 2024.

First, a substantial number of the companies that responded to the survey are working to promote AI and build management systems through strengthening collaboration among related departments. Specific examples include (i) companies that have established a governance forum focused on promoting AI and related risks, and (ii) companies that are considering the establishment of a management system across the IT department, information security and compliance department, and model risk management department by formulating regulations on AI risk management and clarifying the roles of managers and users, the definition of AI-specific risks, and policies to control the risks.

Since AI-related risks are diverse, we confirmed that some FIs are making efforts to grasp the status of risk measures through meetings with frontline departments in collaboration with multiple departments such as the risk management department. Other FIs are making efforts to clarify AI risk evaluation standards from various perspectives in anticipation of the future development of AI technology and its further utilization within the company. We also confirmed that FIs discuss AI governance while receiving regular reports from external experts on the latest domestic and international cybersecurity trends.

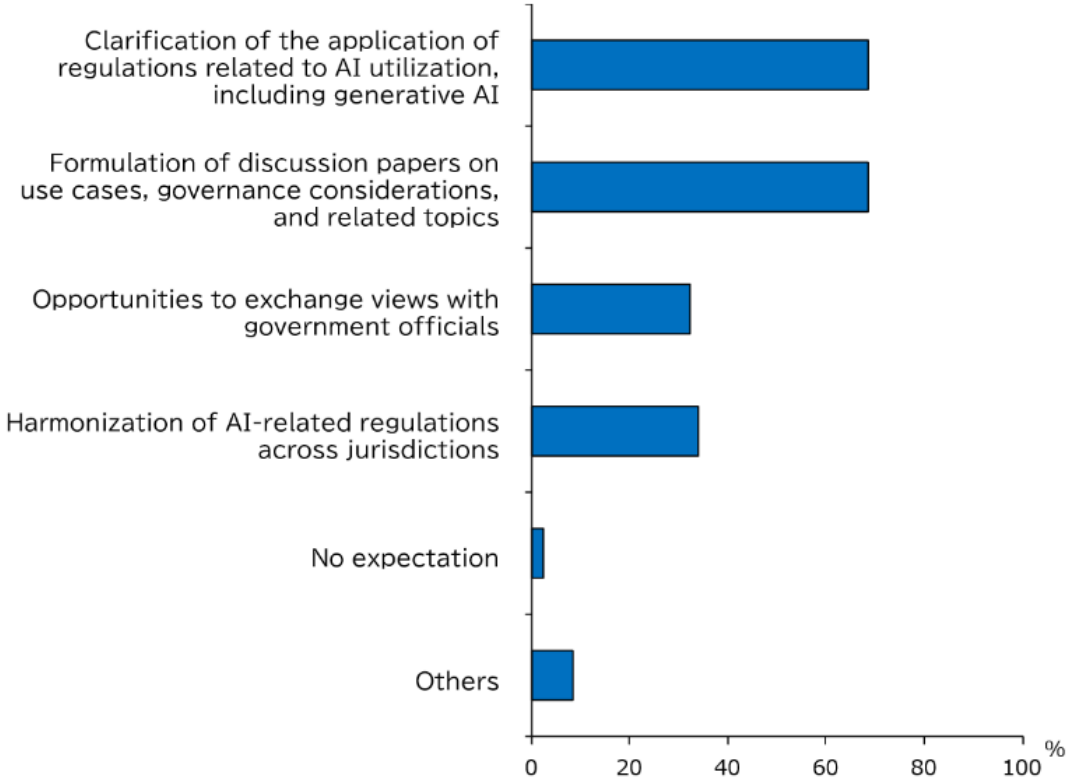
Regarding AI governance, FIs frequently pointed out that balancing offense and defense is difficult. AI-specific challenges, such as the difficulty of control due to the high versatility of generative AI and the non-deterministic nature of its services were also pointed out. Therefore, using the existing governance and risk management framework as a basis, many FIs seem to be considering sound AI utilization through trial and error, such as reviewing the framework as appropriate in line with the expansion of AI use. Given that the goals of AI governance are likely to evolve in response to technological advancements and changes in societal acceptance of AI, it is crucial to adopt an "agile governance"[21] approach that continuously and rapidly cycles through environmental and risk analysis, rule setting, operation, and evaluation, rather than establishing fixed rules and procedures in advance.

[21] Refer to AI Guidelines for Business, Chapter 2 E. Establishing AI governance by Ministry of Internal Affairs and Communications Ministry of Economy, Trade and Industry.

3. Direction of Future Initiatives

This section outlines the FSA's direction of future initiatives and immediate expectations for FIs, incorporating the issues discussed previously, feedback from business to the FSA from the survey, the Japanese government's broader AI strategy, and trends in international discourse.

Figure 17: Major requests to the FSA



① Response by the FSA

i. Recent Actions by the Japanese Government and Related Entities

In Japan, the Social Principles of Human-Centric AI were formulated in March 2019 to realize the concept of a human-centered society that balances economic development with the resolution of social issues through a system that fuses cyberspace and physical space to a high degree (Cyber-Physical System: CPS). Subsequently, the AI Business Guidelines were formulated in April last year for business operators (AI developers, AI providers, and AI users), including FIs, as unified guidelines for AI governance to promote the safe and secure use of AI amid the diversification and increased risks posed by AI.

The AI Institution Research Group was established under the AI Strategic Council in August last year, and an interim report was published in February of this year. The interim report presents the direction with a focus on “balancing innovation promotion and risks management” and “international cooperation.” The report calls for strengthening the control tower function of the Japanese government and formulating strategies with the aim of making Japan “the world’s easiest country to develop and utilize AI,” while also emphasizing the importance of initiatives aimed at enhancing safety. From the perspective of balancing innovation promotion and risk management, the report proposes that measures should be taken by appropriately combining laws and regulations with guidelines and other soft laws, that the autonomy of operators should be respected in principle, and that regulations should be limited to those that cannot be expected to be addressed through independent efforts by operators. It also states that infrastructure and foundational services underpinning national life and economic activities, including some financial services, should continue to be addressed by respective ministries and agencies under the existing legal or regulatory frameworks. Moreover, if new risks materialize and cannot be addressed by the current framework, the ministries and agencies should clarify the interpretation of the related frameworks and consider reviewing existing frameworks or developing new ones.

Additionally, key efforts within the financial sector include the Bank of Japan’s publication last October of “Utilization of Generative AI by FIs and Risk Management: Results of the Survey” (ibid.) Based on the results of the survey of FIs and the exchange of opinions with IT vendors and FIs, the BOJ summarized the status and challenges of the utilization of generative AI in the financial industry and issues related to risk management. Moreover, the Center for Financial Industry Information Systems (FISC) published “Considerations from the Perspective of Safety Measures for Utilization of AI by FIs in Their Business” in September last year. Based on the considerations, the FISC plans to revise the "Standards and Commentaries on Safety Measures for Computer Systems in FIs" by adding standards for AI within this fiscal year.

ii. Trends in International Discussions

Many respondents to the survey indicated that they expect “harmonization of AI-related regulations among jurisdictions.” Accordingly, this section provides an

overview of AI-related regulatory trends in international organizations and major jurisdictions.

The impact of AI on the financial sector has garnered significant attention at the G20. At G20 Finance Ministers and Central Bank Governors Meeting held in February 2024, the importance of understanding the benefits and vulnerabilities coming from digital innovation, including AI, was emphasized. In response to a request from the G20, the FSB published a report on the financial stability implications of AI in November 2024 and submitted it to G20 Leaders. The report analyzes the potential impact of AI on financial stability by updating the FSB's report published in November 2017, given the rapid evolution of AI, including generative AI, and the growing use of AI in the financial sector. The report encourages the Standard Setting Bodies (SSBs) and national authorities to strengthen monitoring, address data gaps, assess the effectiveness of existing regulatory frameworks, and collaborate among authorities. The International Organization of Securities Commissions (IOSCO), one of the SSBs, plans to publish a report by the first quarter of 2025 on market participants' use cases of emerging AI technologies and the issues, risks, and challenges for considering potential policy responses from market integrity and other perspectives. Other SSBs (BCBS, IAIS, IFIAR) have also been making progress on AI-related discussions.

This section provides an overview of AI-related initiatives in major jurisdictions. First, in 2024, the EU passed the AI Act (EU Artificial Intelligence Act), which is a comprehensive AI regulation law. In principle, application of the law is scheduled to begin in August 2026. Under the AI Act, AI systems are classified according to their risks, and stricter requirements are applied to AI systems with high risks. FIs are also subject to the AI Act, and it is expected that financial supervisory authorities in the EU and member states will ensure the application and enforcement of regulations consistent with the EU's existing financial services laws. From June to September 2024, the European Commission invited public comments on the use and impact of AI in financial services, with the aim of effectively implementing existing financial regulations and the AI Act.

In the UK, the government published a White Paper entitled "A pro-innovation approach to AI regulation" in March 2023, and a response to the White Paper was published in February 2024. The UK Government indicated a direction to address

the risks posed by AI on a principles-based approach, without resorting to legislation at this time. Instead, sector-specific regulators will act within their existing legal and regulatory powers, focusing on five cross-sectoral principles, including safety. In light of this, in April 2024, the Bank of England and the Prudential Regulation Authority (PRA), as well as the Financial Conduct Authority (FCA), published a Strategic Approach to AI Regulation, endorsing the direction taken by the government and stating that their current regulatory frameworks and approaches are consistent with the five principles set out in the White Paper, that they will continue to review their regulatory frameworks and approaches in light of the adoption of AI in financial markets and technological advances, and that they will continue to participate in international discussions by financial regulators.

In the United States, financial regulators oversee regulating the use of AI in financial services, and it is pointed out that existing regulations on risk management and governance can be applied to the use of AI. In March 2024, the United States Department of the Treasury released a report outlining AI use and risk management best practices in the financial services sector, primarily from the perspective of AI-specific cybersecurity risks. Furthermore, in December 2024, the Treasury released a report summarizing the results of a call for comments on AI use, opportunities, and risks in the financial services sector. The report recommended continuing international and national cooperation among governments, regulators, and financial services sectors to promote consistent and robust standards for the use of AI in financial services; encouraging stakeholders to identify and address potential gaps in the use of AI in financial services; and proposing financial regulators to continue their efforts to strengthen existing risk management frameworks.

In Singapore, the Monetary Authority of Singapore (MAS) cooperated with the financial industry to develop the Principles to Promote Fairness, Ethics, Accountability and Transparency (FEAT) in 2018 to promote the responsible deployment of AI and data analytics. In November 2019, MAS also launched an initiative (the Veritas Initiative) as part of its National AI Initiative to work with an industry consortium to help FIs implement FEAT in their AI usage and data analytics. Developing on this work, MAS and a consortium of banks and tech companies launched a project (Project MindForge) to explore the risks and opportunities of generative AI and published a comprehensive risk management framework for FIs'

use of generative AI in November 2023. Also, MAS published for FIs an Information Paper on Cyber Risk related to generative AI in July 2024 and an Information Paper on AI Model Risk Management in December 2024. MAS is considering publishing supervisory guidance for all FIs in 2025, developing on the focus areas addressed in the AI Model Risk Management Paper.

The Hong Kong Monetary Authority (HKMA) issued a circular to banks setting out high-level principles for managing risks from the use of AI, stating that it would regularly review the principles and provide additional guidance to banks as necessary considering rapidly evolving global regulatory standards and industry trends related to the use of AI in November 2019. In light of the subsequent emergence of generative AI, in August 2024, the HKMA launched a generative AI sandbox in collaboration with Cyberport. In the same month, the HKMA also issued a circular to banks on consumer protection related to the use of generative AI. The Securities and Futures Commission (SFC) also issued a circular in November 2024 to licensed entities setting out their supervisory expectations for the use of generative AI. Meanwhile, in October 2024, the Financial Services and the Treasury Bureau (FSTB) released a policy Paper on the responsible introduction of AI in the financial services sector. The paper stated that going forward, the Hong Kong government will work closely with the HKMA, the SFC and other financial regulators to provide a clear supervisory framework, and that financial regulators will continue to review existing regulations and guidelines and revise them as appropriate to keep pace with advances in AI and international practice.

iii. Future Direction of Response by the FSA

The Paper presents initial issues for future dialogues with business operators toward sound utilization of AI in the financial sector. Therefore, it does not present detailed policies for reviewing laws, regulations, or guidelines, or for monitoring AI utilization. However, because various respondents to the survey requested clarification of the application of regulations related to AI utilization, the Paper summarizes the initial directions for AI utilization.

First, regarding the clarification of regulatory application, there is a strong demand for clarification in the order of personal information protection, IT governance, model and risk management, and cybersecurity. These points were analyzed in Section 1 of

this chapter, concerning specific issues and examples of initiatives to resolve those issues. The FSA will encourage FIs and other entities to take action in accordance with existing laws, regulations, supervisory guidelines, principles, and guidelines that apply regardless of the use of AI, addressing all of these issues.

Alternatively, there are new issues arising from the characteristics of generative AI and challenges that have become more difficult to address due to generative AI. Therefore, through dialogue with FIs, we will conduct future verifications from the perspective of whether regulatory requirements for AI use are sufficiently clear and whether existing regulatory and supervisory frameworks can adequately address risks. While legal responses are not excluded in cases where significant regulatory gaps are identified, based on the policy that legal regulation should be limited to areas where self-regulation by businesses cannot be expected, we will first consider revisions to principles and guidelines.

Since it is important to understand international technology and business trends, such as the development of groundbreaking AI models by overseas businesses, the FSA will proactively participate in foreign Fintech conferences and the like, identifying the possibility of utilization and potential risks in a forward-looking manner through dialogues with AI model developers, major FIs, Fintech businesses, and others.. Furthermore, since international discussions on AI are expected to advance in the future at the FSB, SSBs, and others, the FSA will actively participate in international rulemaking on AI and consider domestic measures by referring to the documents published by these bodies.

Figure 18: Issues for which clarification of the relationship of application of regulations, and other related issues, is desired

We would like to introduce the Fintech Support Desk and the Fintech Proof-of-Concept Hub as frameworks to support the clarification of the application of regulations. The Fintech Support Desk is a centralized contact point for consultations on laws and regulations from businesses that are involved in various innovations or businesses that are considering new businesses. Since its establishment in 2015 until December 2024, it has received a total of 2,380 consultations from FIs and fintech businesses. Past consultations have included inquiries about AI applications in areas like investment advisory services, chatbots, insurance process optimization. The Desk serves as a resource for those seeking clarity on legal interpretations, particularly concerning the use of generative AI. (Note that the Desk places emphasis on prompt responses and does not provide written responses. To request written responses, one option is to use the gray zone elimination program based on the Industrial Competitiveness Enhancement Act.)

The Fintech Proof-of-Concept Hub is a framework to support Fintech businesses and FIs conducting unprecedented proof-of-concept experiments by forming a team within the FSA and collaborating with relevant ministries and agencies as necessary to conduct the proof-of-concept experiments for innovation. The results of the proof-

of-concept experiments, including the results of the legal verification, will be published on the FSA website. Among the AI-related adopted projects was the “Proof-of-Concept Test for Improving Efficiency of Compliance Operations at FIs Using Artificial Intelligence,” the results of which were published in August 2018. Verification was performed to determine if AI could improve the efficiency and sophistication of checking staff explanations for compliance breaches during financial product sales and extracting customer complaints from related response records.

To reiterate, technological innovation in AI is progressing very rapidly, and it is not appropriate to stick to a specific response at this point. The FSA will conduct policy measures flexibly in light of changes in circumstances while conducting reviews based on the above direction.

As part of this initiative, we will establish a "Public-Private Stakeholder Study Group" within the current fiscal year to delve deeper into the issues mentioned in this document, in collaboration with various stakeholders from both the public and private sectors. Utilizing the insights gained from these study groups, we will work on improving the environment and update this document as necessary.

Figure 19: Road Map

② Initiatives Expected of Business Operators

i. Review of Business Processes

According to the surveys and interviews which included vendors as interviewees, a strong consensus emerged that the impact of AI adoption is often limited if it’s merely added to existing business flows. A fundamental redesign of business processes

centered on AI utilization is essential. Examples of cases where the expected benefits could not be realized include cases where poor user experience in accessing AI service applications and the inability to integrate with internal data hindered business utilization. Additionally, there are instances where delay in the automation of overall business processes have prevented the full exploitation of AI capabilities. As is generally the case with DX initiatives, business process reviews impact a wide range of employees engaged in the relevant business. To truly optimize for AI and make significant changes, it's important to think about building a AI promotional team led by management and creating a structure that fosters synergy between different departments.

ii. Supporting Proactive Efforts to Develop Use Cases

To expand AI utilization, it is necessary to develop specific use cases that are actually used in each department. In this regard, the following initiatives were identified: implementation of an AI idea contest; use of internal newsletters to disseminate use cases that created business impacts; providing support for the development of generative AI utilization environments by the DX strategy department and accompaniment support for the review in business departments; and establishment of a person in charge of promoting the use of generative AI in each department. If successful PoC cases are consolidated and human resources development progresses through such initiatives, use cases that lead to the creation of new services and operational efficiency will increase, leading to the enhancement of services at FIs.

iii. Proactive Involvement of Management

According to dialogues with FIs to date, the more senior executives are aware of AI utilization issues, the more likely they are to adopt AI. Examples of promotion structures led by management include (i) CIO-driven systems development, internal education, and risk management utilizing AI; (ii) cross-group task forces, including international executives, established to formulate AI utilization policies and strategies; and (iii) advisory councils with external experts, where top management personally attends to listen to specialists and actively engage in promoting AI utilization.

V. Use of AI by the FSA

1. The Importance of AI Use as a Financial Supervisory Authority

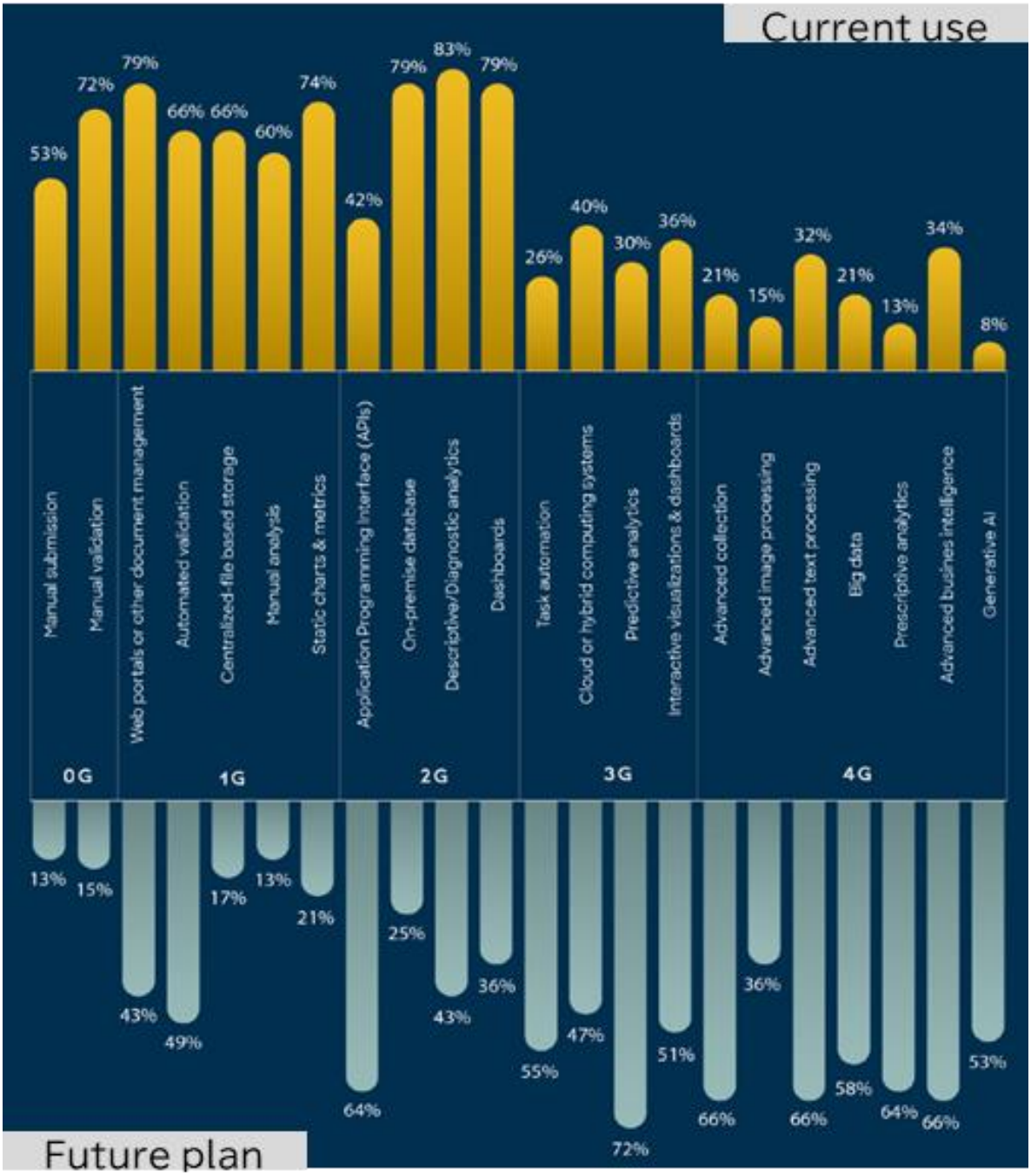
The use of AI and other technologies to enhance monitoring and operational efficiency is an extremely important challenge not only for FIs but also for financial supervisory authorities. The FSB report cited above also notes that the adoption of AI by authorities is progressing to carry out supervisory responsibilities more efficiently, and this trend is likely to intensify further in order to keep up with FIs. The use of technology in regulatory and supervisory activities (SupTech) is advancing globally. According to a survey of sixty-four financial supervisory authorities conducted by the Cambridge Centre for Alternative Finance[22], 59% of respondents have implemented one or more SupTech applications, indicating that the use of technology is advancing particularly in developed jurisdictions.

Among the technologies used in SupTech, the use of AI-related technologies (e.g., text processing and big data analytics) is currently limited, but numerous authorities have expressed a strong desire to use them in the future. They are exploring the use of AI in a wide range of monitoring operations, including the use of big data to predict financial crises, the use of unstructured data (e.g., corporate information and social media), and the detection of suspicious and fraudulent transactions.

In contrast, the following are listed as major challenges associated with the introduction of these technologies: AI model training and validation, data/privacy protection, governance and accountability, data quality, and human resource recruitment. These challenges have countless similarities to those faced by FIs in the previous chapter.

22 Cambridge Centre for Alternative Finance "Cambridge SupTech Lab: State of SupTech Report 2023" (February, 2024)

Figure 20: Status of SupTech adoption by financial supervisory authorities and outlook for future use



2. AI Utilization Efforts to Date and Further Utilization in the Future

① Enhancement of Data Analysis

The FSA is also working to enhance data utilization, including the use of AI-related technologies such as big data (granular data) analysis and text processing. In particular, the use of granular data, such as loan-by-loan level data and stock trading data, allows for a more detailed understanding of FIs' business conditions, the vulnerabilities and resilience of the entire financial system, and financial markets. In this way, the FSA is continuously working to enhance its analytical methods in a more efficient and effective manner while improving the accuracy and expanding the coverage of such granular data.

Some case examples of data analyses undertaken by the FSA are published as "FSA Analytical Notes"[23] in a timely manner. In the "FSA Analytical Notes (2024.7) vol.1," machine learning techniques are used to construct a prediction model for credit rating of a borrower using granular loan data of regional banks, and obtained suggestions on factors that affect the credit risk of real estate loans, which accounts for a large share of bank loans. In the "FSA Analytical Notes (2024.7) vol.2," a neural network is applied to the order and transaction detailed data for stock index futures to analyze the impact of high-speed trading (HFT) strategies on market fluctuations.

In addition to utilizing granular "structured" data including loan-by-loan data, the FSA is also making efforts to incorporate "unstructured" data, specifically through the introduction of text processing. For example, applying text processing technology to documents which contain large amount of information, such as FIs' annual reports and various meeting minutes with FIs, should contribute to efficiently extracting the characteristics of the entire sector, individual FI's business strategy and risk management. As a result, such information along with the expertise of staff engaging in monitoring will enable more efficient dialogues with FIs.

Enhancing data utilization using AI techniques is a key factor for more sophisticated monitoring; however, there are still challenges to fully introduce AI technology into monitoring work. One of the challenges is the accountability of AI models. Depending on the attributes of data and methods used for training of an AI model, the model may be biased, which makes it difficult to verify the reliability of the results. Other major challenges relate to information security. As information and data related to monitoring operations are highly confidential, AI models are currently being

[23] FSA Publication of "FSA Analytical Notes"

constructed in a completely closed environment without relying on publicly available advanced AI. In addition, human resources development is also a big challenge. The FSA has appointed academic experts in the field as advisors to support continuous capacity building. All of these challenges have a lot in common with those faced by FIs.

② Streamlining Operations

To effectively achieve administrative objectives with limited resources, the FSA is continuing and strengthening its efforts to utilize AI to contribute to its operational efficiency. For example, the aforementioned utilization of text processing for large volumes of documents such as annual reports not only leads to the development of monitoring operations but also directly contributes to the efficiency of operations. The FSA is working to build an environment in which many staff members can utilize such efficiency tools including text processing, while balancing information management with operational efficiency.

Furthermore, in 2018, the FSA launched a project titled “Use of AI in examining annual reports” as part of the “Open Policy Lab,” the agency's framework for voluntary policy proposals. This initiative included a verification process through empirical experiments conducted with the cooperation of a total of 20 companies, including AI startups and audit firms, to assess the feasibility of automating the review of the contents of annual securities reports using AI. The verification provided insights such as the potential for improved analytical accuracy with sufficient development time and analysis data. It also indicated the importance of human-AI collaboration, where humans interpret and provide feedback on AI-generated analyses, highlighting the need for both parties to understand their respective strengths and work together.

③ Other Initiatives

In March 2022, the FSA, aiming to enhance the dissemination and reception of information in English across the entire financial industry, not just within the agency itself, collaborated with the National Institute of Information and Communications Technology (NICT) to develop an AI translation system that can translate financial documents between Japanese and English with high accuracy in both directions. There are many technical terms in the financial sector, and it is difficult to obtain

highly accurate results with general-purpose AI translation systems. On the other hand, since it is difficult for a single financial institution to secure sufficient training data, it is also difficult to independently construct an AI translation system specialized in finance.

Therefore, the FSA collected bilingual texts within the agency, industry associations, and private financial institutions and provided them to the NICT. This helped to ensure the necessary amount of information for AI learning and to improve the accuracy of the translation. As a result, they were able to raise the level of accuracy to the point where nearly 50% of all translations were of the highest quality comparable to professional translators specialized in finance. This AI translation system has been made accessible to a wide range of entities through technology transfer from the NICT, which is the main developer, to the private sector.

④ Considering Further Utilization

The FSA will promote the utilization of AI, including generative AI, to further enhance monitoring capacities and improve operational efficiency. In line with the government-wide agreement on the operational use of generative AI, we will consider the utilization of generative AI in areas such as responding to inquiries about regulations, enhancing monitoring of market misconduct, and preparing documents related to monitoring operations. Particularly, with regard to enhancing data analysis, the FSA will use AI and other appropriate analytical methods fit for the issues and objectives to accurately understand the characteristics and trends of the FIs and financial system, and will establish a process to utilize such information in monitoring. In addition, the FSA will also promote the effective collection of information from text data by utilizing LLM. Furthermore, the FSA will engage in human resource development, including the identification of junior staff with knowledge and interest in AI, utilizing frameworks such as the Open Policy Lab (i.e. group for organization-wide tech-forming), which aims to streamline and enhance business operations at the FSA.

VI. Conclusion

1. Importance of Collaboration with Public and Private Sector Stakeholders

As we have seen in the previous chapters, the use of AI by both FIs and the authorities is accelerating. To overcome the challenges and reap the maximum benefits from AI, strengthening collaboration across the financial industry and society is critical. With LLMs being updated and new LLMs released at an almost monthly pace, it is crucial to have a system in place that accumulates knowledge and experience at both the management and staff levels. To develop, operate, and manage AI optimally while considering the company's management strategies, resources, and customer needs, it's essential to keep up with trends.

At the same time, at small FIs where development and human resources are limited, measures at the individual company level are limited, and the pace of technological renewal is likely to be delayed. To address these issues, collaborating among stakeholders in the public and private sectors to create use cases and build AI governance is considered effective. There is considerable room for collaboration in areas such as security and compliance. For example, various data that is difficult to obtain by a single FI can be aggregated through appropriate information coordination between multiple companies. This will improve the quality and quantity of training data, and there is great room for AI to further enhance its effectiveness for the risk management. Additionally, since some of the issues faced by FIs involve legal issues that are not limited to financial regulations such as copyright law and competition law, cooperation with other ministries and agencies is also necessary. Furthermore, as AI regulations are advancing globally, it is important to participate in international rule making.

From these perspectives, the FSA, as a stakeholder, intends to support sound AI utilization in the financial sector. As is the case with not only AI but also fintech in general, dialogue with businesses is essential for implementing appropriate environment and developing policies for innovation, especially with how quickly technology and business are advancing. Therefore, the FSA is determined to continue

providing opportunities for dialogues between domestic and foreign public and private stakeholders, such as Japan Fintech Week, which has been hosted by the FSA since 2024, and to make efforts to realize open innovation while improving the predictability of regulatory supervision.

2. Request for Comments on This DP

The Paper summarized the status and challenges of AI utilization by FIs, and exhaustively examined a wide range of issues. Again, the analysis in the Paper is only at an early stage, and the issues presented are likely to change significantly because of technological innovations and changes in the business landscape. The FSA will continue to strengthen its dialogue with stakeholders based on the perspectives presented here and will continue to consider specific measures in a flexible manner. If you have any comments or suggestions, please contact Fintech and Innovation Office, Risk Analysis Division, Strategy Development and Management Bureau [ai.survey★fsa.go.jp] Please replace the “★” with “@”.]