Title

Japan's Approach to AI: Promoting Innovation with Responsible Risk Management

Key sentence

The FSA actively supports the sound use of AI in the financial sector.

Q. How is the use of AI developing in the Japanese financial sector, and what are the main drivers and obstacles to scaling its use further?

In Japan's financial services sector, the use of conventional AI technologies—such as fraud detection, market analysis and forecasting, and marketing—is already well established. With the growing adoption of generative AI, new use cases are emerging that offer further improvements in operational efficiency and customer experience. Many financial institutions are now aiming for more advanced applications, including the potential development of entirely new financial services.

At the same time, there is increasing awareness of critical challenges associated with complex AI systems like generative AI. These include ensuring transparency, explainability, and fairness, as well as managing the risks of misuse in financial crimes and the potential impact on financial system stability. In light of risks such as hallucination—where AI generates plausible yet entirely inaccurate responses—, most use cases involving generative AI do not present outputs directly to customers. Instead, they incorporate human judgment in the process—a practice commonly referred to as "human in the loop."

Q. How are the Japanese authorities approaching the regulation and supervision of AI use in finance (e.g. in terms of hard and soft laws, sector specific regulation...)?

As Japan's integrated regulatory and supervisory authority responsible for financial stability, consumer protection, and market integrity, the Financial Services Agency (FSA) is committed to fostering an environment in which financial institutions can confidently pursue AI-driven innovation. When challenges arise in the course of innovation, the FSA

seeks to resolve them through dialogue and collaborative engagement, ensuring that regulatory responses do not unduly discourage financial institutions from taking initiative.

To support this, the FSA is working to clarify the applicability of existing regulations and provide safe harbors where appropriate. Its fundamental stance is technology-neutral: existing laws and regulations apply regardless of whether specific technologies such as AI are used. However, where the unique characteristics of AI warrant special consideration, the FSA is prepared to review and revise relevant laws, guidelines, and supervisory frameworks.

Given the rapid pace of AI development, it is essential to maintain flexible and adaptive policymaking through ongoing dialogue with financial institutions. The FSA actively supports the sound and responsible use of AI in the financial sector. In line with this approach, the FSA launched a public-private forum in June 2025 and held its first meeting in the same month. The forum serves as a platform for incorporating the views of market participants and academia into the development of necessary policies, including the formulation of regulations and guidance.

Q. How do you see the downside of new developments such as generative AI?

A recent survey by FSA identified hallucination as a prominent risk associated with generative AI, alongside its potential misuse in financial crimes. Hallucination is particularly problematic from the standpoint of financial consumer protection. Unlike fraud or scams, it arises from the model's inherent design, making it difficult to regulate through traditional frameworks. As AI-based services begin to be offered directly to ordinary retail investors, the need to ensure their protection will become increasingly important.

This risk is especially concerning in financial services like investment advisory (e.g., robo-advisors), where misleading outputs could result in real financial losses for retail consumers. Similar concerns apply to bank loan approvals and insurance pricing, particularly when consumers have limited alternatives.

In my understanding, three core challenges stand out: First, hallucinated outputs are often highly convincing and confidently presented, making them difficult to detect. Second, financial institutions may lack incentives to implement human-in-the-loop oversight, as

expert review across disciplines can offset the cost benefits of AI. Third, some AI developers include liability waivers in their terms of service, placing the burden of responsibility on consumers.

To address these issues, I would like to propose a three-tiered defense strategy: The first tier is developing new architectures that prevent hallucination or further enhancing safeguards such as retrieval-augmented generation (RAG) which mitigate it. The second tier is proper intervention where service providers ensure sufficient human oversight before outputs reach consumers, with industry self-regulation playing a complementary role. The third tier is compensation mechanisms such as introducing liability insurance for AI-related losses, and creating compensation funds to provide a safety net.