NEWSLETTER

Legal Risk Management and Governance Framework for the Use of AI
– Based on the FSA's AI Discussion Paper –

Finance Law & Robotics / Artificial Intelligence Newsletter August 18, 2025

Authors:
Shinnosuke Fukuoka
s.fukuoka@nishimura.com
Toshiyuki Yamamoto
to.yamamoto@nishimura.com

As AI technology advances, financial institutions are adopting AI on an increasing basis, and the establishment of risk management and governance frameworks for the use of AI have become a critical issue. In March 2025, the Financial Services Agency of Japan (FSA) published the "AI Discussion Paper (Version 1.0)" ("Discussion Paper"), which presents an initial set of discussion points to promote the sound use of AI in the financial sector. The Discussion Paper not only highlights potential negative impacts of AI but also points to the "risk of not taking on challenges"—the risk that financial institutions may fall behind in technological innovation and struggle to provide high-quality financial services over the medium to long term. The paper thus encourages proactive use of AI by financial institutions, while calling for the establishment of corresponding governance and risk management systems.

I Barriers to Adoption of AI and Countermeasures

The Discussion Paper identifies three common challenges that financial institutions face in connection with the use of conventional and generative AI:

1. Data preparation and quality control
2. Third-party risk management (management of external vendors)
3. Uncertainty in return on investment (ROI)

Proposed countermeasures include:

- For issue 1: establishing systems for data collection and management
- For issue 2: implementing contractual safeguards and adopting management frameworks
- For issue 3: budgeting within digital departments and establishing/monitoring KPIs.

II Challenges Significantly Compounded by Generative AI (Part 1)

The Discussion Paper also identifies the following six challenges, which have become particularly difficult due to generative AI:

(1) Ensuring Explainability

Generative AI produces output through processes that involve massive numbers of parameters, making them inherently difficult for humans to understand or explain. However, financial institutions are expected to provide certain levels of accountability, particularly with regard to decisions that impact customers. The use of generative AI creates a risk of failure to comply with these accountability obligations.

(2) Risks of Bias and Unfairness

If training data or algorithms contain biases, there is a risk that specific groups of customers may be unfairly

disadvantaged. For example, there is a risk that loan approvals may show bias with regard to gender or race, despite otherwise identical conditions.

(3) Development and Operation of AI Systems and Model Risk Management

Like other financial models, AI models can lead to adverse outcomes in the event of errors or inappropriate use. A risk-based approach, such as that described in the FSA's "Principles for Model Risk Management" (published November 2021), can be applied here. A key countermeasure is the establishment of a "model inventory" that identifies and tracks the key AI models being used.

(4) Protection of Personal Information and Information/Cybersecurity

Many generative AI models are hosted on large-scale cloud platforms provided by IT vendors, and carry risks of personal information leaks when data is input, as well as risks of personal information being inferred from output. When using AI services provided by third parties, there are risks of cyberattacks, manipulation, or security vulnerabilities at points of integration with internal systems. However, risks described in the Act on the Protection of Personal Information can be mitigated by complying with the law; the key issue is how to ensure compliance when using generative AI.

(5) Hallucinations

Generative AI may produce seemingly plausible but factually incorrect information, a phenomenon known as "hallucinations." For example, reports or answers generated by AI may contain misinformation that misleads customer decision-making. One countermeasure is human intervention and review. However, humans being involved in every step of the process may undermine the benefits of AI. Uses of generative AI can be categorized as: (i) internal use, (ii) indirect use in customer-facing services, and (iii) direct use in customer-facing services. Employee education can be a significant countermeasure in connection with uses (i) and (ii). To address the risks inherent in use (iii), institutions may employ human oversight or clearly disclose the use of generative AI and warn of the risk of hallucinations.

(6) Misuse for Financial Crime

It is easy for generative AI to create convincing phishing emails or deepfake voices and images that impersonate real individuals, which creates risks of fraud and cybercrime. Financial institutions must build systems to prevent deception, train employees properly, and develop mechanisms to detect unusual activities initiated by customers who may have been deceived. In addition, misinformation spread via social media or AI-driven algorithmic trading may cause unintended chain reactions and heightened market volatility. Risk management departments also must respond from the perspectives of market and reputational risks.

III Challenges Significantly Compounded by Generative AI (Part 2)

In addition to the items in the Discussion Paper, the following issues have been raised as specific challenges associated with the use of generative AI:

(7) Leaks of Confidential Information/Breaches of Confidentiality Obligations

As with personal data, when confidential information is input into generative AI systems, there is a risk of information leaks, breaches of a financial institution's confidentiality obligations, or violations of nondisclosure agreements. Leaks of trade secrets also could constitute violations of the Unfair Competition Prevention Act. While inputting confidential information into generative AI is not categorically prohibited, the implementation of

internal rules specifying the conditions under which such input is permitted—as well as system structures that prevent the disclosure of information externally—can serve as effective countermeasures.

(8) Copyright Infringement

Copying and using data to train AI systems at the development phase may infringe copyrights. The generation and use of AI-generated output also may present a risk of copyright infringement. Article 30-4 of the Japanese Copyright Act generally permits the use of copyrighted materials for machine learning without authorization from the copyright holder. However, if AI is trained with the purpose of reproducing the creative expressions of copyrighted works, in whole or in part, the applicability of Article 30-4 becomes questionable, and remains a debated issue in legal theory. Particular caution is needed when using RAG (Retrieval-Augmented Generation), as certain configurations may result in the reproduction of third-party copyrighted material found online. Unless a proper license for use of copyrighted materials has been obtained, there is a high risk of copyright infringement in the generation and use phase. Therefore, one mitigation strategy is to avoid using third-party copyrighted work in AI prompts. Output that resembles copyrighted works also should be avoided.

IV Key Considerations for Building AI Governance Frameworks for Financial Institutions

1. What is AI Governance?

The adoption of new technologies inevitably introduces new risks, and it is not feasible to eliminate risks entirely. These technologies also offer valuable benefits and returns. Therefore, it generally is believed that AI should be used within a governance framework that manages risks appropriately.

One key concept highlighted in the “AI Guidelines for Business (Version 1.1)” (March 2025) issued by the Ministry of Internal Affairs and Communications and the Ministry of Economy, Trade and Industry, is the “risk-based approach.” This involves identifying the potential magnitude and likelihood of harm associated with specific AI uses in advance and adjusting the level of safeguards accordingly. This concept aligns well with financial-sector-specific guidelines, such as the FSA’s “Principles for Model Risk Management” mentioned in II.

While ensuring compliance with relevant laws and regulations, financial institutions should consider building AI governance frameworks based on a risk-based approach. Because there are no express legal requirements that define AI governance, each financial institution must construct its framework in accordance with its own organizational circumstances. Given the fast pace of innovation in AI, agile governance is a key principle also.

2. Specific Issues in the Financial Sector

While AI governance is relevant to any enterprise that uses AI, issues unique to financial institutions include:

(i) The regulated nature of the industry, public interest responsibilities, and the importance of maintaining public trust and credibility;
(ii) Application of the three lines of defense model in risk management;
(iii) Determining which department or individual should lead governance efforts, given the existence of many relevant stakeholders;
(iv) Designing governance mechanisms that do not stifle the motivation of business and IT departments.

The Discussion Paper addresses issue (i) by emphasizing the need to recognize both the risks of AI use and the “risk of not taking on challenges.” This means that financial institutions must take a proactive stance toward the adoption of AI. However, as part of a regulated industry with high public expectations, financial institutions

must strike a careful balance between risk-taking and risk aversion.
Financial institutions typically address issue (ii) by adopting a three lines of defense model, involving frontline business departments (first line), risk management and compliance departments (second line), and internal audits (third line). This framework can be leveraged for AI governance, and attention should be paid to how AI-specific governance mechanisms can be embedded within the existing structure.
Issue (iii) arises in part because the adoption and use of AI often involves multiple departments, beyond the risk and compliance functions, such as IT, development, and planning. Without unified leadership, these projects are unlikely to succeed. Therefore, a central department with authority to coordinate these functions is vital. At many institutions, a DX (digital transformation) unit takes the lead in AI issues also. However, approaches vary depending on the institution's structure and culture. Moreover, because expertise in AI remains limited, the leadership of knowledgeable individuals may be just as important as departmental coordination.
While risk-awareness is important, overly burdensome rules—such as excessive manuals or rigid checklists—can become counterproductive, as recognized by issue (iv). The focus may shift toward procedural formality rather than substantive risk mitigation. A balanced and prioritized approach, via a risk-based model, is necessary to avoid these pitfalls.
It is important to note that implementation of AI governance does not guarantee legal compliance. Therefore, it remains essential to involve the legal department in the governance process.

3. Key Elements to Consider When Building an AI Governance Framework

While not exhaustive, the following table outlines key elements that financial institutions should consider when establishing an AI governance framework. Chapter 4 of the Financial Data Utilizing Association (FDUA)'s Guidelines for the Use of Generative AI in Financial Institutions (Version 1.1) also serves as a useful reference.

[Key Considerations for Building an AI Governance Framework]

Item	Key Considerations
1. Management Involvement	(1) Senior management should clearly define the purpose and strategy for AI use and lead the development of an institution-wide governance framework. (2) Clearly articulate the risk appetite (i.e., the acceptable level of risk) to align AI initiatives across the organization. (3) Implement regular reporting, monitoring, and reviews through the board of directors and executive committees.
2. AI Policy and Internal Rules	(1) Establish and publish a basic policy for AI use (AI Policy). - Define core principles, scope of use, accountability, fairness, ethics, etc. (2) Develop detailed internal rules, guidelines, and manuals. - Departments in charge. - Clarify roles and responsibilities of departments involved in AI. - Establish procedures for reporting to the board of directors and executive committees. - Establish methods for identifying and evaluating risks. - Approval processes for AI use. - Legal aspects: establish rules for handling personal information/personal data, customer data, confidential/secret information, and intellectual property such as copyrights.
3. Organizational Structure	(1) Designate a lead department and assign responsible officers. (2) Set up cross-functional committees or working groups involving

Item	Key Considerations
	business, planning/DX, risk management, legal/compliance, and IT/development departments. - Maintain awareness of the three lines of defense model. (3) If necessary, invite external advisors to incorporate up-to-date technologies and governance practices (e.g., advisory boards or external committees).
4. Risk Identification, Evaluation, and Approval Processes	(1) Define "risk." (2) Identify and evaluate risks based on various perspectives: - Internal use vs. customer-facing use - Handling of personal, customer, and confidential/secret data - Degree of human involvement - Model complexity and explainability - Fairness - Whether the model is developed internally or externally (third-party risk) (3) Handling of high-risk AI. (4) Third-party risk when using outside vendors. (5) Clarify risk assessment and approval procedures before development or deployment. (6) Use checklists to assess and approve model performance, accuracy, bias, and data quality.
5. Monitoring	(1) Perform regular monitoring of working AI models (e.g., prediction accuracy, anomaly detection, error rates). (2) Consider a combination of real-time monitoring and human-in-the-loop oversight for high-risk uses. (3) Share monitoring results with management and relevant departments and take prompt corrective actions. (4) Revise AI policies and internal rules as necessary. (5) Continue performance reviews and evaluations post-deployment (e.g., change management, version control).
6. Employee Training and Culture Building	(1) Provide AI literacy training for all employees and risk-response training based on role. (2) Use case studies and real-world scenarios in educational programs. (3) Clearly communicate usage guidelines and prohibitions. (4) Foster a culture that emphasizes explainability and transparency, and avoids overreliance on AI decisions. (5) Encourage reporting and sharing of incidents and near-misses to promote organization-wide learning and prevent recurrence. (6) Build a culture of shared risk awareness and accountability from the executive level to the front line.

Under "1. Management Involvement," while all items are important, particular attention should be paid to "4. Risk Identification, Evaluation, and Approval Processes," as building a robust framework—from developing evaluation criteria to actual implementation—may require considerable time.

For instance, even defining "risk" presents challenges. While guidelines such as the "AI Guidelines for Business" define risk as the magnitude of harm and its likelihood, this differs from the quantitative models commonly used in financial institutions, which assume a certain probability distribution and use historical data to calculate numerical outcomes. In contrast, risks associated with AI are not necessarily easy to quantify, and qualitative evaluation may be unavoidable. Therefore, even after risks are identified (see II and III), quantifying the "magnitude of harm" can be difficult, and estimating "likelihood" based on past data is challenging in emerging fields like AI. As such, it may be necessary to respond by applying qualitative assessments and categorizing

risks into certain types. In this regard, as noted in "(2) Identify and evaluate risks based on various perspectives" in the table above, it is important to consider a wide range of evaluation perspectives. Moreover, evaluating risks across various AI systems used in different business and operational departments using a unified standard—essentially applying a "horizontal approach"—is also difficult. An agile mindset of "thinking while running" will likely be essential.

For third-party risks when using external vendors, in addition to complying with existing laws and regulations on outsourcing, in managing third-party risks, it is necessary to consider not only oversight and monitoring but also contingency plans for service suspension, termination, or contract cancellation.

Each financial institution needs to independently assess risks and build its own framework for AI use. However, this also presents opportunities for creativity and innovation.

V Conclusion

Since the use of AI also entails the "risk of not taking on challenges," it is important to adopt a forward-looking attitude toward technological innovation while addressing the associated risks. The FSA has expressed its intent to support the development of flexible and effective governance frameworks in dialogue with the industry, through platforms such as the "FSA AI Public-Private Forum."

Financial institutions must not only develop governance structures and rules that suit their operational realities, but also continually adapt to the evolving technological landscape. It is imperative to establish a framework that enables the safe and responsible use of AI.

Public Relations Section, Nishimura & Asahi newsletter@nishimura.com