Consultation Fact Sheet

Sound Practices for Responsible Adoption of Artificial Intelligence

Body	
Financial Stability Board	
Published 2026-06-10	**Consultation deadline** 2026-07-22
Scope of application All financial institutions globally, covering organisation-wide governance and the full lifecycle of traditional, generative and agentic AI	**Impacted / targeted stakeholders** Financial institutions and market infrastructure providers including their boards and senior management, business and control functions; AI and technology vendors; financial authorities; customers and investors

Synthesis

The FSB is consulting on 12 non-binding sound practices for responsible AI adoption across all types of financial institutions. Applied proportionately across traditional AI, GenAI and agentic AI, they cover board oversight, accountability and AI inventories; materiality and risk assessment; model selection, data governance, explainability, performance testing and human oversight; and AI-specific cyber, operational and third-party concentration risks. The practices are technology-neutral, complement existing rules and standards, and are not intended to create an international standard or prescribe particular AI technologies.

01

Purpose & rationale

AI adoption is accelerating across financial services, supporting efficiency, customer service, decision-making, fraud detection and risk management. GenAI and autonomous agents can also amplify governance, model, data-quality, bias, explainability, conduct, cyber, operational and third-party dependency risks. Common models, data and infrastructure may create correlated behaviour and concentration with financial-stability implications. The FSB therefore proposes a proportional, technology-neutral menu of practices to help institutions innovate responsibly.

Specific policy objectives:

- Enable institutions to capture AI benefits while containing risks to customers, firms and financial stability.
- Embed board accountability, clear governance, risk appetite, documentation and adaptable AI capabilities.
- Apply proportionate lifecycle controls to materiality, model selection, data, explainability, performance and human oversight.
- Strengthen AI-specific cyber, operational, third-party and concentration-risk management, coordination and information sharing.

02

Key proposals

The FSB proposes 12 complementary sound practices spanning organisation-wide governance and the AI lifecycle. The practices are technology-neutral, non-prescriptive and intended as a proportional menu:

Practice	Description
Organisation-wide governance	
Set strategic direction and oversight	▪ Align AI adoption, expected outcomes and risk tolerance with business strategy and legal obligations. ▪ Define prohibited or restricted uses based on materiality, autonomy, and potential customer and market impacts ▪ Maintain effective oversight through appropriate management information, skills, technology and resources
Establish governance and accountability	▪ Define responsibilities across business functions, risk and compliance functions, and independent assurance ▪ Adopt a centralised, decentralised or hybrid operating model that supports consistent oversight and effective risk management ▪ Coordinate business, legal, compliance, risk, audit and technology functions while preserving independent challenge
Integrate AI into risk management frameworks	▪ Identify, assess, approve, document, monitor and manage AI risks through existing or enhanced frameworks ▪ Apply guardrails, controlled experimentation and staged approvals according to each use case's materiality and risk ▪ Maintain lifecycle records covering purpose, ownership, data, dependencies, limitations, approvals, testing and current status
Build organisational adaptability	▪ Develop foundational AI literacy across the organisation and role-specific expertise for boards, developers and control functions ▪ Monitor emerging technologies, risk typologies, industry practices and supervisory expectations ▪ Use incidents, near misses, testing results, validation findings and drift events to improve governance, training and controls
AI lifecycle management	
Assess materiality and risk	▪ Assess the use case's importance to critical operations, customer outcomes and legal or regulatory obligations ▪ Evaluate autonomy, complexity, system and data access, explainability, cyber exposure, human oversight and third-party dependencies ▪ Reassess materiality and risk following significant changes, incidents, performance deterioration or expansion into critical activities
Select appropriate AI solutions	▪ Decide whether to develop the solution internally, procure it from a provider, use open-source technology or combine these approaches ▪ Select traditional, generative or agentic AI based on intended use, customer impact, materiality, risk, data and institutional readiness ▪ Consider performance, explainability, cost, legal obligations, third-party dependencies, continuity and change-management requirements

Practice	Description
Strengthen data governance	▪ Establish clear accountability and controls for data access, privacy, security, classification and quality ▪ Document data provenance, preparation, labelling, limitations and lineage across internal, public and third-party sources ▪ Monitor post-deployment data quality and address drift, inappropriate use and constraints affecting third-party data
Address explainability and transparency	▪ Assess how the model's complexity, autonomy, materiality and risk affect the need for explainability ▪ Prefer more explainable solutions where feasible or compensate through stronger testing, benchmarking, data controls and human oversight ▪ Tailor disclosures, explanations, contestability and redress to boards, authorities, auditors, customers and users
Manage AI performance	▪ Define use-case-specific thresholds for accuracy, robustness, stability, bias, output ranges and other relevant outcomes ▪ Conduct developmental testing, benchmarking, sensitivity testing, stress testing, red teaming and subject-matter review ▪ Monitor performance and drift over time, then remediate, restrict, redevelop or retire systems when necessary
Implement meaningful human oversight	▪ Select human-in-the-loop, human-on-the-loop, AI-in-the-loop or human-in-command arrangements according to materiality, autonomy and risk ▪ Assign trained and appropriately independent overseers with clear escalation, approval, override, contestability and kill-switch mechanisms ▪ Apply tighter boundaries and approval checkpoints to generative and agentic AI, particularly where sensitive data or financial transactions are involved
Manage cyber and ICT risks	▪ Apply secure-by-design development, least privilege, identity and access controls, network segmentation, layered defence, patching, logging and monitoring ▪ Test threats such as poisoning, prompt injection, deepfakes, shadow AI and compromised agents through exercises, red teaming and controlled environments ▪ Share incidents, near misses and threat intelligence, and consider AI tools for detection, incident response and vulnerability management
Manage third-party AI risks	▪ Conduct AI-specific due diligence and ongoing oversight of providers, embedded AI functionality and supply-chain dependencies ▪ Secure contractual rights covering information, audit, data, security, change notification, continuity and termination ▪ Use compensating controls where visibility is limited and maintain effective portability, substitution and exit arrangements for material services

Appendix

Key concepts and definitions

Term	Definition
AI lifecycle	The stages through which an AI use case is managed: inception; design and development; verification and validation; deployment; operation and monitoring; re-evaluation; and retirement.
Agentic AI	Systems that autonomously perform complex, extended tasks, making decisions and taking actions with limited human oversight; they may incorporate generative capabilities.
Foundation model	A pre-trained model, usually built with deep learning on very large datasets, that can be adapted or refined for many downstream uses.
Generative AI (GenAI)	AI that creates new content - such as text, images, audio or video - typically in response to prompts and often using foundation models.
Large language model (LLM)	A foundation model trained on natural language for tasks such as text generation, classification, summarisation, question answering and sentiment analysis.
Retrieval-Augmented Generation (RAG)	A technique that improves LLM outputs by retrieving information from a selected source corpus, such as internal databases or live data feeds.
Explainability	The extent to which humans can understand how an AI model or system turns inputs into outputs, including its logic, influential factors and feature importance.
Data drift / model drift	Data drift occurs when data no longer reflect current conditions; model drift occurs when relationships between inputs and outputs change, weakening assumptions or performance.
LLM hallucination	A seemingly confident but inaccurate, fabricated or nonsensical output generated by a large language model.
Data/model poisoning	An attack that manipulates training data or model parameters to introduce errors or bias and compromise an AI system's integrity, reliability or security.