FSB FINANCIAL STABILITY BOARD

Sound Practices for Responsible Adoption of Artificial Intelligence (AI): Consultation report

Response to Consultation

The Finvocates

1. **Do you agree with the benefits and risks of AI adoption by financial institutions described in this report? Are there any substantive benefits or risks not covered?**

   The benefits and risks identified in the report are broadly agreed with.

   However, wider access to Artificial Intelligence (hereinafter "AI") is likely to accelerate both desirable and undesirable outcomes with comparable ease, since it lowers the effort required to act on any given intent. This creates a genuine risk of increased employee-induced fraud and underscores the need for auditable logs covering the use of all AI tools. Banking Frontiers reported a 73% rise in employee fraud over eight years (Source: Banking Frontiers, https://bankingfrontiers.com/employee-fraud-rises-73-in-8-years/).

   Third-party vendors that outsource essential or critical parts of their AI services should carry liability for the explainability of those outsourced parts. Such vendors should not be permitted to plead ignorance where the outsourced components fall outside their own visibility, as this would allow non-explainability to be used as a shield by bad actors.

   AI models typically retain a copy of the data fed into them. Given that financial institutions will necessarily process personally identifiable information (PII) when working with AI, data safety requires particular care. Anonymisation should be considered wherever possible, and the unnecessary creation of copies of sensitive data on the cloud purely to increase productivity should be avoided, consistent with the principle of data minimisation. Proper data-purging measures should also be in place to ensure that copies of sensitive data do not linger in the cloud environment: every replication of sensitive data proportionally increases the risk associated with it.

   It is common practice among financial-sector employees to access work email on personal mobile devices. Access to AI-based environments via private devices introduces multiple non-organisational cyber environments into the picture, creating a risk of unintentional data exchange.

   A further use case where AI may be successfully deployed is consumer grievance handling. Similar grievances could be resolved through a set of pre-approved remedial actions, and reporting of repetitive data requests within such grievances could be automated via AI.

AI-based kiosks could be deployed in rural areas where compliance knowledge is scarce among staff or the population, providing an efficient channel for employee training and query resolution. This would also help expedite the organisation-wide rollout of policy amendments and bring uniformity to services across jurisdictions.

In certain cases, such as fraud or suspicious-transaction reporting, human interpretation remains necessary to determine the correct course of action. For example, a customer who sets up an electronic mandate for daily payments to a firm they indirectly own, and another customer who uses a similar mandate for a daily bus commute, may both appear innocuous on the surface if each transacts through a single dedicated account for the purpose. An AI-based monitor may struggle to distinguish which of the two is layering a money-laundering scheme, since neither pattern is likely to trigger an alert at initial screening.

Permissioned distributed ledger technology (DLT) may be considered for storing logs of sensitive data or actions with real impact on stakeholders, since this would flag any attempt to tamper with or manipulate official records immediately.

Another way to validate GenAI-based output may be to deploy validators built using GenAI itself. A deterministic, prompt-frozen agent could help flag and cross-verify the factual accuracy and logical coherence of the primary AI's output. Such frozen-prompt validator models could themselves be improved by users on an ongoing basis.

Where an organisation develops its own AI model, clear boundaries must be enforced on what data goes into training, since creating any copy of sensitive data carries a corresponding, specific liability.

AI could be used to provide staff at the ground level with step-by-step, live compliance guidance. Embedding automated checkboxes for important disclaimers and disclosures within such a process would help prevent staff from later pleading ignorance, ensuring that liability for the legal consequences of a specific action transfers immediately to the employee responsible for it.

Redressal of grievances arising from erroneous AI action should be governed by a clear, categorical allocation of liability. Financial institutions may wish to devise appropriate schemes that preserve their ability to claim non-liability in instances that fall categorically outside their remit to have prevented.

AI could help reduce corrupt practices at customer touchpoints by curbing undue discretion through AI-based legal warnings and communications.

Joint Liability Groups and Self-Help Groups could be extended additional benefits, such as AI-based record management with reduced scope for error.

AI models could be trained to remain grounded in verifiable facts and measurable data points, maintaining clean citability of their output.

2. Are the sound practices sufficiently comprehensive and clear to enable financial institutions' responsible AI adoption?

The suggested practices are broadly sufficient; the following may also be considered.

Financial institutions should reserve, for themselves, the direct ability to undo the actions of any AI-based process or procedure.

Cross-institutional sharing of audit findings should be kept confidential and withheld from extra-organisational release unless the underlying issue has been resolved or is confidently and internally controlled or mitigated; exposing such sensitive industry information to bad actors risks unnecessary exacerbation of the underlying risk.

Care should be taken to ensure that cross-organisational AI access does not worsen the material risk associated with related-party transactions (RPTs).

Sensitive information (corporate secrets) arising at high-level meetings, such as board meetings, must be protected with appropriate security and internal classification.

3. **Do the sound practices strike an appropriate balance between managing risks relating to all forms of AI, and addressing some of the risks relating to emerging and new complex forms of AI, such as GenAI and agentic AI?**

   The suggested approach to managing risks across different forms of AI is consistent.

   The ability to undo the actions of both traditional and agentic AI must remain reserved to organisations.

   Clear, step-by-step logs of GenAI and agentic AI processing should be maintained and periodically reviewed by human auditors on a random-sampling basis to support continuous improvement.

   Software should be built with the ability to produce clean, traceable, logical decision trees.

4. **Are the sound practices sufficiently flexible to accommodate and address newer types of AI and responsible adoption over time?**

   The flexibility of the proposed framework is agreed with.

5. **Do the case studies in this report sufficiently highlight how different types of financial institutions can benefit from responsible AI adoption? Are there additional case studies for inclusion in the report? If so, please provide such case studies, particularly for nonbanks.**

   Addressed above, in the responses to Questions 1 and 2.

6. **Do the case studies in this report provide actionable insights for financial institutions in their responsible AI adoption?**

   The suggested framework accommodates sufficient space for organisational interests, leaving it to individual organisations to design their executive machinery around it. This makes the framework more than sufficiently actionable.

7. **Are the definitions in the glossary clear and aligned with industry sound practices, including recent developments in AI?**

While the glossary broadly covers the requisite terms, the following may be considered for inclusion, given the rise of cloud-based and localised automation in the industry using applications such as Microsoft Power Automate, n8n, and Activepieces.

Composite AI – Composite AI is a methodology that brings together multiple artificial intelligence techniques—such as machine learning, natural language processing (NLP), knowledge graphs, and rule-based reasoning—to tackle tough, real-world business problems. By blending the strengths of techniques, such as statistical, symbolic, and knowledge-based, into a single solution, businesses can take on challenges that are too complicated or nuanced for a lone AI model to manage alone. (Source: Oracle, https://www.oracle.com/in/artificial-intelligence/composite-ai/)

Agentic Orchestration – AI agent orchestration functions like a digital symphony. Each agent has a unique role and the system is guided by an orchestrator—either a central AI agent or framework—that manages and coordinates their interactions. The orchestrator helps synchronise these specialised agents, ensuring that the right agent is activated at the right time for each task. This coordination is crucial for handling multifaceted workflows that involve various tasks, helping ensure that processes run seamlessly and efficiently. (Source: IBM, https://www.ibm.com/think/topics/ai-agent-orchestration)

8. Are there any comments you would like to make on this report? Please fill in.

On page number 21, in 2.2 Governance and accountability, Establish cross-functional coordination mechanisms paragraph, the second sentence has an error. The intended word "facilitates" is written as "facilities", which must be corrected as follows:

This [wrong - facilities] [right - facilitates] knowledge sharing so that stakeholders across the financial institution understand the full range of benefits and risks of AI, optimise implementation plans, and clearly delineate responsibilities and accountabilities.