FSB FINANCIAL STABILITY BOARD

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

Response to Consultation

Infocomm Media Development Authority Singapore

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?**

   Please see feedback to question 8 on specific inputs related to AI agents

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

   Please see feedback to question 8 on specific inputs related to AI agents

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?**

   Please see feedback to question 8 on specific inputs related to AI agents

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

   Please see feedback to question 8 on specific inputs related to AI agents

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.**

   Please see feedback to question 8 on specific inputs related to AI agents

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

   Please see feedback to question 8 on specific inputs related to AI agents

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

Please see feedback to question 8 on specific inputs related to AI agents

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

On "Agentic AI Risks" (page 16) - Other key agentic risks to highlight:

1) Biased or unfair actions: While the agent's actions may not be erroneous (i.e. it follows the prescribed steps) nor unauthorized (i.e. it acts within its authorization boundaries), it's actions may still lead to outcomes that are not ethically / socially acceptable e.g. hiring decisions that are unfairly skewed towards a certain demographics

2) Agent sprawl: This is a significant emergent risk; as organizations deploy more and more agents, this can lead to the uncontrolled proliferation of AI agents without centralized oversight. This can then lead to issues with provenance, incompatibility between old and new agents etc.

On "Developmental testing of Gen AI and agentic AI models" (page 39) - For developmental testing of AI agents, some key additional challenges / considerations are:

1) Testing entire agent workflows: Agents can take multiple steps in sequence without human involvement. Thus, beyond testing an agent’s final output, agents should be tested across their entire workflow, including reasoning and tool calling.

2) Testing agents individually and together: Beyond individual agents, testing should be carried out at the multi-agent system level, to understand any emergent risks and behaviours when agents collaborate, such as competitive behaviours or the impact on other agents when one agent has been compromised.

3) Testing in real or realistic environments: As agents may be expected to navigate real-world situations, testing should occur in a properly configured execution environment that mirrors production as closely as possible, such as using tool integrations, external APIs, and sandboxes that behave as they would in deployment. However, organisations should calibrate the need for realism against the risk of prematurely allowing agents to access tools that affect the real world.

On "Additional considerations for Gen AI and agentic AI" (page 43) - Additional considerations for human oversight of agentic AI:

1) Gradual / phased roll out as an oversight mechanism: Progressively expanding an agent's deployment — by user group, by tools/systems accessible, or by risk tier — allows human overseers to build calibrated trust and catch issues before wide-scale exposure, rather than deploying at full autonomy from day one.

2) Designing approval requests to be digestible and contextual: As agents operate at machine speed and continuously, the logs produced may bve voluminous and reasoning chains may be very long. Hence to facilitate ease of review by humans, the info can be formatted to present approval requests concisely with the key risk factors and a confidence score, so humans can make an informed decision quickly without being overwhelmed.

On "reducing the potential attack surface by implementing strong network security" (page 45) - On a related note (specifically for AI agents), another measure to consider is MCP / protocol-layer governance i.e. MCP is currently the de facto protocol for agent-to-agent comms, hence the MCP layer itself can be treated as a security control point i.e. whitelisting trusted MCP servers and only allowing the agent to interact with these servers on the whitelist. Organizations can also explore additional controls on the MCP layer, such as filtering sensitive data that passes through the servers, logging all agent-to-system interactions etc.

On "threat modeling" (page 45) - Beyond generic threat modelling, for AI agents a method called "taint tracing" could be helpful in mapping out all the workflows and interactions to track how untrusted data can move through the system, since agentic systems can become very complex.