Monetary Authority of Singapore

Consultation Paper

P017-2025 – November 2025

Consultation Paper on Guidelines on Artificial Intelligence Risk Management

Contents

1. Preface

1.1. The Monetary Authority of Singapore (MAS) is proposing to introduce Guidelines on Artificial Intelligence (AI) Risk Management (the "Guidelines")[1] to enhance management of AI risks in financial institutions (FIs), and set out MAS' supervisory expectations relating to AI risk management in the financial sector. The Guidelines focus on oversight of AI risk management in FIs, key AI risk management systems, policies and procedures, key AI life cycle controls, as well as capabilities and capacity needed for the use of AI.

1.2. The Guidelines aim to establish a set of expectations that are generally applicable across the financial sector, and may be applied in a proportionate manner across FIs of different sizes and risk profiles. The Guidelines should be generally applicable to different AI applications and technologies, including Generative AI, as well as newer developments such as AI agents. Nonetheless, MAS recognises the evolving nature of AI and will update or augment these Guidelines when necessary.

1.3. MAS invites comments from FIs and other interested parties on the Guidelines.

1.4. Please note that all submissions received will be published and attributed to the respective respondent unless they expressly request MAS not to do so. As such, if respondents would like:

(a) their whole submission or part of it (but not their identity), or

(b) their identity along with their whole submission,

to be kept confidential, please expressly state so in the submission to MAS. MAS will only publish non-anonymous submissions. In addition, MAS reserves the right not to publish any submission received where MAS considers it not in the public interest to do so, such as where the submission appears to be libellous or offensive.

1.5. Please submit your written comments to the consultation paper by **31 January 2026** via this link: *https://form.gov.sg/690b2a3b024ee5eebbfcf7f1*

[1]The scope of AI in this consultation paper includes AI based on machine learning, deep learning, reinforcement learning techniques, as well as Generative AI, AI agents and any newer AI technologies that fall within the proposed scope of AI set out in paragraph 1.2 of the Guidelines.

2. MAS' Supervisory Approach to AI Risk Management

2.1. The use of AI in the financial sector is not new. However, recent advancements in AI have led to an increased interest in leveraging AI in the financial sector. While AI can deliver significant benefits across business and functional areas, the use of AI, in particular newer and more complex AI technologies such as Generative AI and AI agents, introduces new challenges. As AI adoption becomes more pervasive across business and functional areas within FIs, it may accentuate existing risks or introduce new risks. For example, poor performance of AI models used for risk assessments could lead to substantial financial losses, unexpected behaviours in AI systems could disrupt critical operations, and inappropriate outputs from customer-facing AI systems could result in harm or financial loss to customers.

2.2. Generative AI also introduces risks that are less well understood and harder to mitigate, such as hallucinations that produce convincing but false information, unpredictable behaviours that emerge from using more complex methods, and fundamental challenges in explaining decision-making processes. Other risks include security vulnerabilities such as prompt injection attacks, privacy risks from potential data leakage when using third-party services, intellectual property violations from training on copyrighted content or generating outputs that may infringe existing copyrights, concentration risks from over-reliance on a few dominant Generative AI providers, and human-factor risks arising from over-reliance on Generative AI.

2.3. The use of even newer technologies, such as AI agents that leverage Generative AI, but with greater autonomy and the ability to access tools, could introduce even more significant risks. An AI agent granted access to an FI's internal systems might autonomously execute actions misaligned with business objectives or customer interests, while compromised AI agents could exfiltrate sensitive data or execute malicious commands.

2.4. Alongside the growing use of AI in the financial sector and such associated risks, MAS had established key principles to guide FIs in their responsible use of AI. In 2018, MAS co-created the principles of Fairness, Ethics, Accountability and Transparency (FEAT)[2] with the financial industry to promote the deployment of AI and data analytics in a responsible manner. To support FIs' efforts in implementing FEAT, MAS started working with an industry consortium on the Veritas Initiative[3] in November 2019. The Veritas Initiative aimed to support FIs in incorporating the FEAT principles into their AI and data analytics solutions. The Veritas Initiative has released assessment methodologies, a toolkit, and accompanying case studies.

2.5. With the emergence of Generative AI, Project MindForge[4] was established to examine the risks and opportunities of Generative AI. The first phase of Project MindForge was led by a consortium of banks and released a risk framework for Generative AI in November 2023. In the second phase of Project MindForge, the consortium was expanded to include FIs from other parts of the financial sector, such as capital markets and

[2] https://www.mas.gov.sg/publications/monographs-or-information-paper/2018/feat
[3] https://www.mas.gov.sg/schemes-and-initiatives/veritas
[4] https://www.mas.gov.sg/schemes-and-initiatives/project-mindforge

insurance. The expanded consortium is working on an industry-led AI Risk Management Handbook that will serve as a companion guide for FIs implementing the Guidelines[5].

2.6. MAS has also recently released information papers relating to AI and Generative AI from the perspective of an FI's use of AI, as well as the use of AI by third parties against the FI:

(a) Cyber Risks Associated with Generative Artificial Intelligence (July 2024)[6]. The paper provided an overview of key cyber threats arising from Generative AI, the risk implications, and mitigation measures that FIs could take to address such risks. The paper also covered areas enabled by Generative AI, such as deepfakes, phishing and malware, as well as threats to deployed Generative AI, such as data leakage and model manipulation.

(b) AI Model Risk Management (MRM) in Banks (December 2024)[7]. The paper highlighted good AI MRM practices in banks relating to robust oversight and governance of AI, key risk management systems and processes for comprehensive AI identification, risk materiality assessments and inventories, as well as standards and controls for rigorous development, validation and deployment of AI.

(c) Cyber Risks Associated with Deepfakes (September 2025)[8]. The paper provided an overview of the emerging threats and risks posed by deepfakes, their potential impact on the financial sector, and mitigation measures that FIs could implement to address such risks. The paper covered three key areas where deepfakes could impact the financial sector. This included the use of deepfakes for defeating biometric authentication, carrying out social engineering for impersonation and scams, and facilitating the dissemination of misinformation and disinformation.

2.7. The work by the financial sector is also supported by key initiatives at the national level, such as the Infocomm Media Development Authority's Model AI Governance Framework[9], as well as the various initiatives under the AI Verify Foundation[10].

2.8. These Guidelines build on the FEAT principles, experiences from collaborating with industry, and MAS' information papers relating to AI risks. While FEAT sets out principles in the areas of fairness, ethics, accountability and transparency when using AI, these Guidelines focus on articulating high-level supervisory expectations relating to risk management when AI is used in the financial sector. The Guidelines focus on

[5]The Project Mindforge AI Risk Management Handbook will be released by January 2026.
[6]https://www.mas.gov.sg/regulation/circulars/cyber-risks-associated-with-generative-artificial-intelligence
[7]https://www.mas.gov.sg/publications/monographs-or-information-paper/2024/artificial-intelligence-model-risk-management
[8]https://www.mas.gov.sg/regulation/circulars/cyber-risks-associated-with-deepfakes
[9]https://www.pdpc.gov.sg/-/media/Files/PDPC/PDF-Files/Resource-for-Organisation/AI/SGModelAIGovFramework2.pdf
[10]https://aiverifyfoundation.sg/

oversight of AI risk management in FIs, key AI risk management systems, policies and procedures, key AI life cycle controls, and capabilities and capacity needed for the use of AI[11].

2.9. The Guidelines should be generally applicable to different AI applications and technologies, including Generative AI, as well as newer developments such as AI agents. Nonetheless, MAS recognises the evolving nature of AI and will update or augment these Guidelines when necessary.

3. Applicability of the Guidelines

3.1. MAS proposes to apply the Guidelines to all FIs[12]. FIs in Singapore that are branches or subsidiaries with parent entities in other jurisdictions may leverage the AI risk management frameworks of their parent entities, as long as such frameworks meet the expectations set out in these Guidelines.

3.2. MAS recognises that AI can be applied to a wide range of use cases, and that the risks associated with different usage of AI may vary based on the scale, scope and business models of FIs. FIs may implement these Guidelines in a proportionate manner[13], i.e., commensurate with the size and nature of their activities, use of AI, and their risk profiles, as well as the relevance of these Guidelines to the specific AI model, system or use case.

3.3. For the purpose of the Guidelines, the proposed scope of AI includes an AI model, system or use case, defined as follows:

(a) A model is a method or approach which converts assumptions and input data into outputs such as estimates, decisions, or recommendations;

(b) A system can comprise one or more models and other machine-based components;

(c) A use case refers to a specific real-world context that the model or system is applied to; and

(d) AI includes use cases involving models or systems that learn and/or infer from inputs to generate outputs such as estimates, predictions, content, summaries, recommendations, or decisions that

[11]There are other risks to FIs that may arise from the use of AI by external actors, such as for AI-powered cyber-attacks or scams. These are not within the scope of these Guidelines, and are covered by other MAS publications such as those mentioned in paragraphs 2.6(a) and 2.6(c).

[12] As defined in Section 2 of the Financial Services and Markets Act 2022.

[13]All FIs should institute basic policies for the use of AI commensurate with the FI's level of AI adoption. These policies should address who is responsible for overseeing AI use, guidelines on allowed and disallowed uses of AI, as well as the communication, checks and reviews of such guidelines. Expectations relating to AI oversight and key AI risk management systems, policies and procedures would only apply to FIs using AI as an integrated part of their business processes. The expectations relating to AI life cycle controls, as well as capabilities and capacity for the use of AI may be applied based on the relevance and risk materiality of the AI use cases, systems and models in the FI. More details on the implementation of these Guidelines in a proportionate manner are set out in paragraph 1.5 and the Annex of the Guidelines.

may influence physical or virtual environments, and vary in their levels of autonomy and adaptiveness after deployment.[14] Calculators or tools whose outputs are solely based on predefined programming logic or rules would not be regarded as AI for the purpose of these Guidelines.

Question 1. MAS seeks comments on the application of the Guidelines to all FIs in a proportionate manner, and the guidance on the proportionate application set out in paragraph 1.5 and the Annex of the Guidelines.

Question 2. MAS seeks comments on the proposed scope of AI use cases, systems and models for the application of the Guidelines.

4. Proposed AIRG

AI Oversight

4.1. The Guidelines set out MAS' expectations on the Board and senior management of FIs to govern and oversee AI-related risks across a range of areas. These include the establishment and robust implementation of frameworks, structures, as well as policies and processes in FIs to identify and inventorise the use of AI, assess their risk materiality, put in place the necessary governance and risk management frameworks, policies and processes, as well as manage risks of AI throughout its lifecycle. Board and senior management should also foster the appropriate risk culture for the use of AI, and ensure that existing organisation-wide risk management is updated to address risks introduced by the use of AI. Where the overall AI risk exposure of an FI is deemed material[15], MAS proposes that the FI establish a dedicated cross-functional committee to ensure adequate oversight and to proactively address potential gaps in risk management.

Question 3. MAS seeks comments on the proposed responsibilities of the Board and senior management in overseeing AI risk management.

[14]For the purposes of these Guidelines, this would generally include AI based on machine learning, deep learning, reinforcement learning techniques, as well as Generative AI, AI agents and any newer AI technologies.

[15]FIs should assess the materiality of their AI use cases, and the overall AI risk exposure based on the guidance on risk materiality assessment set out in paragraphs 3.8 to 3.11 of the Guidelines. The materiality of the overall risk exposure of the FI to AI should take into consideration the dimensions set out in paragraph 3.10 of the Guidelines, as well as the impact on FI's business strategies and overall risk profile. In general, if there are one or more AI use cases in the FI that expose the FI to significant risks (e.g., due to the deployment of high risk AI in critical business lines or functional areas) that could adversely impact the FI or its customers, the FI should assess if such heightened management oversight, i.e., establishment of a dedicated cross-functional committee, is necessary.

Question 4. MAS seeks comments on the proposal for FIs to establish a dedicated cross-functional committee to oversee AI risk if the overall AI risk exposure of an FI is deemed material; and how such overall AI risk exposure should be assessed at the organisational level.

Key AI Risk Management Systems, Policies and Procedures

AI Identification

4.2. Identifying where AI is used within FIs is a critical prerequisite for applying the appropriate governance, risk management standards, and controls effectively to such usage of AI. Hence MAS expects FIs to establish clear definitions, criteria and processes, supported by robust systems, to facilitate this identification process and ensure the consistent identification of AI usage across all relevant business and functional areas.

AI Inventory

4.3. Unapproved usage of AI, particularly in higher-risk use cases, can lead to unintended consequences and an FI being exposed to AI risks beyond its risk appetite. To mitigate this risk, MAS proposes that FIs establish and maintain an accurate and up-to-date inventory of AI use cases, systems or models to support governance and oversight, as well as risk management throughout the AI lifecycle. Such an inventory can be established specifically for AI or by enhancing existing inventories. In either case, there should be clear linkages between the AI inventory and other relevant inventories in the FI.

Risk Materiality Assessment

4.4. MAS recognises that AI is used across a wide range of business and functional areas with varying levels of risks. MAS expects FIs to implement an appropriate assessment methodology to evaluate the risk materiality of AI use cases, systems, or models. MAS proposes that the risk materiality assessment minimally cover the key dimensions of impact, complexity and reliance.

Question 5. MAS seeks comments on the proposal for FIs to establish clear definitions, criteria and processes, supported by robust systems, to facilitate the consistent identification of AI usage across all relevant business and functional areas.

Question 6. MAS seeks comments on the proposal for FIs to establish and maintain an accurate and up-to-date inventory of all AI usage.

Question 7. MAS seeks comments on the proposed risk dimensions of impact, complexity and reliance that should be captured by FIs in AI risk materiality assessments, and whether there are any other risk dimensions that should be included.

AI Life Cycle Controls

4.5. MAS expects FIs to plan for and implement robust controls covering the entire AI life cycle, based on their relevance to the AI model, system or use case and proportionate to the assessed risk materiality of the specific AI model, system or use case. Key areas that FIs should assess for relevance to the AI model, system or use case, and apply in a proportionate manner include data management, fairness, transparency and explainability, human oversight, management of third-party AI risks, selection of AI, evaluation and testing, technology and cybersecurity, reproducibility and auditability, reviews, monitoring and change management.

Question 8. MAS seeks comments on the proposed standards, processes and controls that should be applied across the entire AI life cycle, and the key areas that FIs should assess for relevance to the AI model, system or use case, and apply in a proportionate manner.

General

4.6. In addition to the areas highlighted above, MAS welcomes comments on other aspects of the Guidelines.

Question 9. MAS seeks comments on any aspects of the Guidelines that have not been covered in earlier questions, as well as aspects of AI risk management that have not been covered in the proposed Guidelines.

Implementation

4.7. MAS recognises that the maturity of AI risk management practices vary among FIs. Hence, MAS proposes to provide a transition period of 12 months after the Guidelines are issued, for FIs to assess and implement the Guidelines as appropriate.

Question 10. MAS seeks comments on the proposed transition period of 12 months.

5. List of Questions

	the key areas that FIs should assess for relevance to the AI model, system or use case, and apply in a proportionate manner.	
Question 9	MAS seeks comments on any aspects of the Guidelines that have not been covered in earlier questions, as well as aspects of AI risk management that have not been covered in the proposed Guidelines.	9
Question 10	MAS seeks comments on the proposed transition period of 12 months.	9

6. Proposed Guidelines on AI Risk Management

1 INTRODUCTION

1.1 The Guidelines on Artificial Intelligence (AI) Risk Management (AIRG) set out MAS' supervisory expectations relating to AI risk management in financial institutions (FIs)[1]. Existing principles on Fairness, Ethics, Accountability and Transparency (FEAT)[2] continue to apply for guiding the use of AI in the financial sector[3]. These Guidelines complement the FEAT principles by setting out MAS' supervisory expectations in the areas of oversight, key AI risk management systems, policies, procedures, AI life cycle controls, and capabilities and capacity for the use of AI. An overview of the key sections covered in these Guidelines is set out below.

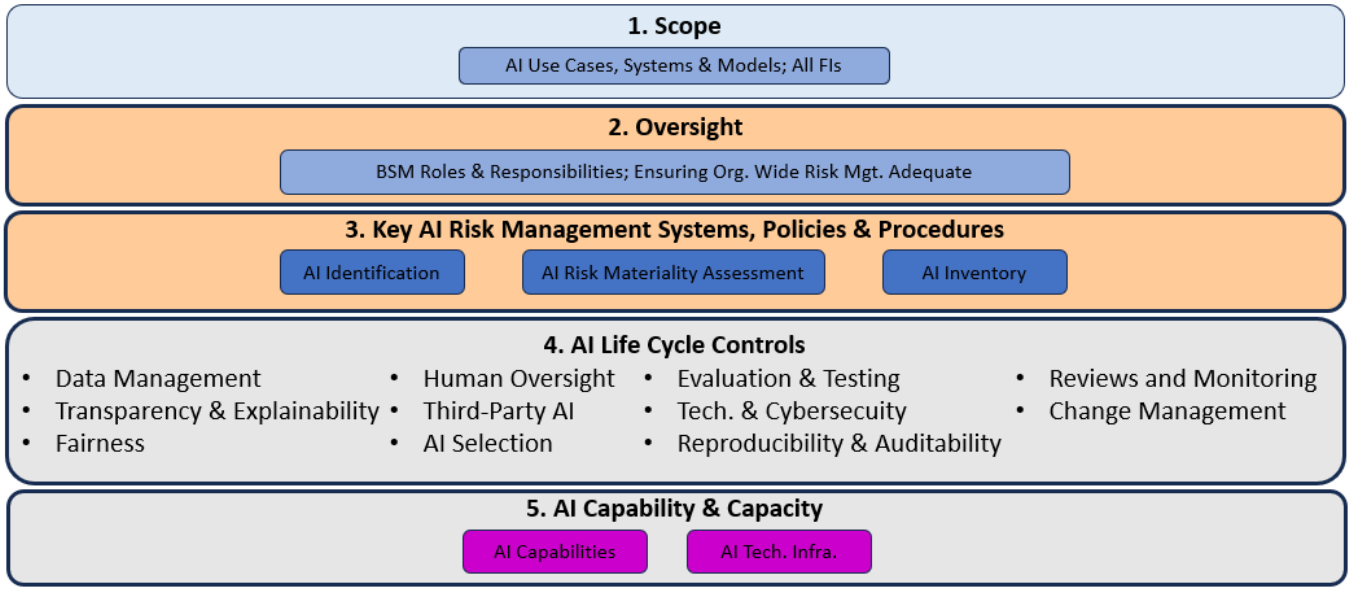

Figure 1: Overview of AIRG

1.2 For the purpose of these Guidelines, AI may refer to an AI model, system or use case[4], defined as follows:

a. A model is a method or approach which converts assumptions and input data into outputs such as estimates, decisions, or recommendations.
b. A system can comprise one or more models and other machine-based components.
c. A use case refers to a specific real-world context that the model or system is applied to.
d. AI includes use cases involving models or systems that learn and/or infer from inputs to generate outputs such as estimates, predictions, content, summaries, recommendations, or decisions that may influence physical or virtual environments, and vary in their levels of autonomy and adaptiveness after

[1]As defined in Section 2 of the Financial Services and Markets Act 2022.
[2]https://www.mas.gov.sg/publications/monographs-or-information-paper/2018/feat
[3]FIs may also refer to key initiatives at the national level, such as the Infocomm Media Development Authority's Model AI Governance Framework, as well as the various initiatives under the AI Verify Foundation to inform their use of AI.
[4]Where a point pertains specifically to an AI use case, AI system or AI model, we will use the respective terms explicitly in the paper.

deployment[5]. Calculators or tools whose outputs are solely based on predefined programming logic or rules would not be regarded as AI for the purpose of these Guidelines.

1.3 The Guidelines establish a set of high-level expectations that all FIs should adhere to as they deploy AI and AI technologies, including Generative AI and newer developments such as AI agents. FIs in Singapore that are branches or subsidiaries with parent entities in other jurisdictions may leverage the AI risk management frameworks of their parent entities to meet the expectations set out in these Guidelines.

1.4 FIs should implement these Guidelines in a manner commensurate with the size and nature of their activities, and the extent to which their use of AI could lead to material risks to them. AI can be applied to a wide range of use cases, and the risks associated with different usage of AI may vary based on the scale, scope and business models of FIs.

1.5 All FIs should minimally institute basic policies for the use of AI commensurate with the FI's level of AI adoption. These basic policies should address who is responsible for overseeing AI use, guidelines on allowed and disallowed uses of AI, as well as the communication, checks and reviews of such guidelines. FIs using AI as an integrated part of their business processes should minimally establish frameworks, policies and procedures to oversee their use of AI; apply clear identification and robust risk materiality assessments of AI use cases, systems or models; and have an adequate AI inventory in place. The extent of such oversight and AI risk management systems, policies and procedures may be proportionate to the size and nature of the FI's activities, use of AI, and risk profile. The expectations for FIs that use AI as an integrated part of their business processes are set out in Section 2 and 3 of these Guidelines. More details and examples on what constitute the use of AI as an integrated part of business processes in the FI, and the scope of basic policies are provided in the Annex.

1.6 The application of AI life cycle standards and controls, as well as establishment of capabilities and technology infrastructure for the use of AI (as set out in Section 4 and 5 of these Guidelines) may be calibrated based on their relevance to the AI use cases, systems or models in the FI. Where such aspects are relevant to the AI use case, system or model, the FI may implement them based on risk materiality. For example, MAS would expect an FI using AI in a manner that has a material impact on customer or risk management outcomes[6] to have more robust AI life cycle standards and controls, as well as stronger capabilities and technology infrastructure. Certain AI life cycle standards and controls, such as those relating to transparency, explainability, and fairness may also be more relevant for AI used in areas with customer impact such as credit scoring, insurance underwriting, provision of financial advisory or fund management services. For AI used in low risk materiality areas to assist humans and not for decision making, such as copilots to assist in writing, AI life cycle standards and controls relating to selection of the AI algorithm or certain evaluation and testing methods for AI, e.g., stress testing, may be less relevant. However, AI life cycle standards and controls relating to data management, safety and cybersecurity would be relevant and should be applied in a proportionate manner.

[5]For the purposes of these Guidelines, this would generally include AI based on machine learning, deep learning, reinforcement learning techniques, as well as Generative AI, AI agents and any newer AI technologies.

[6]This could include computation of regulatory capital, regulatory reporting, management of key financial, operational, or anti-money laundering and fraud risks, or conduct of regulated activities with high impact on customer outcomes (e.g., provision of financial advisory services or fund management). MAS recognises that the degree of impact depends on how AI is used in these areas and expects FIs to apply proportionate standards and controls based on the assessed risk materiality.

1.7 Given the pace of AI developments, MAS also expects FIs to regularly review the adequacy of its AI risk management efforts against AI developments and address new or accentuated AI risks that may arise due to such developments.

Risks

1.8 The use of AI can improve performance across business and functional areas but its complexity and probabilistic nature can lead to greater uncertainty, as well as unexpected or more biased behaviour that is harder to identify compared to the use of simpler methods[7]. The greater complexity of AI also leads to challenges in understanding and explaining its outputs that may be inaccurate or biased. Hence, general risks that could arise from the use of AI include:

a. Financial risks, e.g., when used for risk management, the greater uncertainty and unexpected behaviour of AI could lead to poor risk assessments and consequent financial losses.
b. Operational risks, e.g., unexpected behaviour of AI used to automate an FI's operations could lead to operational disruptions or errors in critical processes.
c. Conduct risk, e.g., use of AI can result in biased outputs, leading to unfair treatment of certain customer groups, a bias for certain investment types over others, or conflicts of interests between an FI and its customers.
d. Financial crime risks, e.g., when used to support anti-money laundering efforts, the greater uncertainty and unexpected behaviour of AI could mean that suspicious transactions go undetected.
e. Reputational risks, e.g., customer-facing systems, such as chatbots, could provide incorrect information or make offensive remarks that lead to negative media attention and reputational damage.

1.9 With Generative AI, such existing risks associated with AI may be amplified. The greater complexity of Generative AI gives rise to even greater uncertainty and unexpected behaviour compared to AI. The unstructured nature of Generative AI inputs and outputs, and a lack of established techniques in this area also make it harder to evaluate and test, as well as understand and explain Generative AI's behaviour and outputs. The diverse and often opaque data sources used in Generative AI training, coupled with difficulties in evaluating bias of Generative AI outputs, could also result in decisions that lead to unfair customer outcomes. Generative AI may also lead to a range of other inter-related risks, such as:

a. Security risks, e.g., arising from adversarial attacks on Generative AI systems via prompt injection or data poisoning[8].
b. Privacy risks, e.g., arising from the leakage of confidential or customer data due to the use of third-party Generative AI products or services, or the inappropriate use of customer data in Generative AI without proper consent.
c. Legal and intellectual property risks, e.g., arising from the use of Generative AI trained on data that infringes existing copyrights or patents.
d. Third-party risks, e.g., arising from over-reliance on a few dominant Generative AI providers for mission critical areas, or using open-source models with poor security controls.
e. Operational risks, e.g., potential service disruptions due to Generative AI outages and problems arising from inadequate pre-deployment testing.

[7]Such uncertainty or behaviour may arise from the noise that is inherent in training data, training data that may not be representative, or when the model encounters scenarios that are not present in its training data. It may also arise due to data or model drifts over time.

[8]More details on such risks are covered extensively in MAS' Information Paper on Cyber Risks Associated with Generative Artificial Intelligence. The paper can be accessed at https://www.mas.gov.sg/regulation/circulars/cyber-risks-associated-with-generative-artificial-intelligence.

f. Human-factor risks, e.g., arising from poor human oversight over Generative AI used in mission critical areas, or skill degradation due to over-reliance on Generative AI.

1.10 Similarly, the use of newer technologies such as AI agents, which may be granted greater autonomy and access to tools, could further amplify these risks. For example, an AI agent with access to tools could autonomously execute actions that are not aligned with an FI's business objectives or a customer's best interests due to a divergence between human goals and how the AI agent translated such goals into actions. Compromised AI agents with access to internal systems and external tools could be used to exfiltrate sensitive data or execute malicious commands at scale, amplifying security risks.

1.11 MAS recognises that AI applications continue to grow and will update or augment these Guidelines when necessary.

2 AI OVERSIGHT

2.1 The Board of Directors ("Board") and senior management of an FI play critical roles in establishing and overseeing robust frameworks, policies and procedures to support AI risk management across the FI.

2.2 **The Board and senior management should maintain effective oversight of AI-related risks, foster the appropriate risk culture for the use of AI, and ensure that its use of AI would not conflict with its ability to meet other supervisory expectations**[9]**.** This includes establishing and implementing robust frameworks, structures, policies, and processes to:

a. Identify AI use cases, systems or models (both internally developed and third-party AI[10]);
b. Assess the materiality of AI-related risks;
c. Maintain an inventory of AI use cases, systems or models and govern their use according to a predefined risk appetite;
d. Manage an AI use case, system or model throughout its entire lifecycle; and
e. Develop the capabilities and capacity needed for the development and deployment of AI use cases, systems, or models in the FI.

2.3 **The Board and senior management should ensure that existing risk management frameworks, policies, and practices across the organisation[11] adequately identify, assess, and address risks posed by AI**. Across existing risk management areas across the organisation, the FI should:

a. Identify and assess all relevant AI risks;
b. Update relevant policies and procedures to address such risks;
c. Institute appropriate strategies and controls to mitigate such risks;
d. Define the FI's risk appetite in relation to such risks;
e. Establish relevant indicators and appropriate risk appetite thresholds for such risks;

[9]For example, FIs should continue to adhere to MAS' Guidelines on Fair Dealing even if it leverages AI to deliver products and services.

[10]For the purpose of these Guidelines, third-party AI includes all providers of third-party AI products and services, which can include third-party systems, models, as well as data used for AI. Existing third-party products and services where AI has been introduced would also be included.

[11]Key existing risk management areas where AI risks are relevant include but are not limited to model, operational, reputational, data, technology and cybersecurity, third-party, legal and compliance, financial, conduct, and environmental risks.

f. Monitor such indicators and the adherence to risk appetite thresholds;
g. Articulate clear roles and responsibilities for managing AI risks across different business lines and functions;
h. Set out clear policies and procedures for updating the Board and senior management on breaches of such risk appetite thresholds and AI related incidents; and
i. Perform regular reviews to take into account newer AI developments, changes in the FI's risk profile and business strategies, and AI regulatory developments.

2.4 **The Board and senior management should ensure consistent standards, clear accountability, and robust coordination across the FI to manage AI risks.** While the FI may adopt varying approaches to AI risk management, such as creating a new centralised function to manage all AI-related risks or adopting a less centralised approach where incremental risks arising from AI are assigned to and managed by existing risk management functions, the Board and senior management should still ensure consistent and coordinated management of AI risks. Where overall AI risk exposure of the FI is deemed material[12], the FI should establish a dedicated cross-functional committee to ensure adequate oversight and to proactively address potential gaps in risk management coverage. Board and senior management must also ensure that the FI continues to comply with all existing regulatory requirements relevant to these risk areas, even when AI is adopted.

2.5 The Board, or a committee delegated by it, is responsible for:
a. approving the overall governance approach for AI risk management, including key frameworks, structures, policies and procedures designed to assess and manage the FI's AI risk on an ongoing basis;
b. ensuring that AI risks, where material, are explicitly addressed within the FI's risk appetite framework, including the setting of appropriate qualitative statements and quantitative measures or limits;
c. setting clear roles and responsibilities for the Board and senior management concerning AI risk management oversight;
d. ensuring it has an adequate understanding of AI to provide effective oversight and challenge; and
e. ensuring that the FI's approach, risk appetite framework, roles and responsibilities, capabilities and culture for risk management of AI use are regularly reviewed to keep pace with newer AI developments, as well as changes in the FI's risk profile and business strategies.

2.6 Senior management is responsible for:
a. ensuring the effective implementation of AI-related risk management policies and procedures across the FI, consistent with the FI's risk appetite;
b. regularly reviewing the effectiveness of the AI-related risk management policies and procedures, making appropriate revisions to keep pace with newer AI developments and changes in the FI's risk profile and business strategies, as well as relevant regulatory requirements;

[12]FIs should assess the materiality of their AI use cases, and the overall AI risk exposure based on the guidance on risk materiality assessment set out in paragraphs 3.8 to 3.11. The materiality of the overall risk exposure of the FI to AI should take into consideration the dimensions set out in paragraph 3.10 as well as the impact on FI's business strategies and overall risk profile. In general, if there are one or more AI use cases in the FI that expose the FI to significant risks (e.g., due to the deployment of high risk AI in critical business lines or functional areas, or the use of AI in relation to the conduct of regulated activities) that could adversely impact the FI or its customers, the FI should assess if such heightened management oversight, i.e., establishment of a dedicated cross-functional committee, is necessary.

c. ensuring robust mechanisms for coordination and accountability for AI-related risk management are established and maintained across the FI;
d. establishing an internal escalation process for managing material AI risks and exceptions, such as incidents or breaches of risk thresholds, and ensuring appropriate and timely actions are taken;
e. updating the Board on material AI risk issues in a timely manner; and
f. ensuring the necessary competence of personnel and allocating adequate resources (such as human, technological, financial resources) for effective AI risk management, including appropriate training and capacity building.

3 KEY AI RISK MANAGEMENT SYSTEMS, POLICIES AND PROCEDURES

3.1 **An FI should ensure that its AI risk management framework encompasses key systems, policies and procedures for the identification, inventorisation, and risk materiality assessment of AI.** The FI should apply consistent and robust approaches to identify, inventorise and determine the risk materiality of AI, and apply controls proportionate to the assessed risk materiality.

AI Identification

3.2 **An FI should establish systems, policies and procedures to ensure the consistent identification of AI usage across all relevant business and functional areas.** This identification process is a critical prerequisite for applying the appropriate governance, risk management standards, and controls effectively to AI throughout their lifecycle. Clear definitions, criteria and processes, supported by robust systems, should be implemented to facilitate this identification process.

3.3 **An FI should assign clear roles and responsibilities for AI identification, including the designation of a control function to be responsible for AI identification systems and processes**, in areas such as ensuring the consistent application of the identification process across the FI, setting up attestation processes, or acting as the final arbiter in determining whether AI is being used**.** The designated control function should also ensure that clear documentation of the identification process and outcomes is maintained, and that identification systems and processes are regularly reviewed and updated to take into account newer AI technologies.

AI Inventory

3.4 **An FI should establish and maintain an accurate and up-to-date inventory of AI use cases, systems or models across the FI to support governance and oversight, as well as risk management, throughout the AI lifecycle.** There should be clear policies and procedures on the maintenance of the inventory, with new, updated or decommissioned AI use cases, systems or models reflected accurately. The FI can enhance existing inventories to include AI use cases, systems or models or establish a dedicated inventory for AI use cases. In either case, there should be clear linkages between the AI inventory and other relevant inventories in the FI.

3.5 **The AI inventory should capture key attributes to enable effective governance and oversight, as well as risk management.** Specific attributes may vary based on the FI's context, but they may include the following for each AI use case, system or model: purpose and description, approved scope of use (e.g., jurisdiction), model type, data used, dependencies, lifecycle status, assigned risk materiality rating, validation status, key roles and

responsibilities (e.g., owners, developers), and links to essential documentation. Maintaining this information facilitates risk aggregation, monitoring of scope adherence, and consistent application of controls.

3.6 **The design of the inventory should be regularly reviewed to ensure that the attributes captured take into account newer AI technologies** where there may be additional relevant attributes or guardrails needed, or additional information relating to third-party AI.

3.7 **The FI should assign clear roles and responsibilities for the inventorisation of AI, including the designation of a control function to be responsible for the AI inventory**, in areas such as policies and procedures relating to the inventory, maintenance and update of the inventory, attestation process, and regular reviews of the scope of the inventory.

AI Risk Materiality Assessment

3.8 **An FI should establish an assessment methodology to evaluate the risk materiality of an AI use case, system or model based on the nature of its business.** The assessment methodology should be applied consistently to perform risk materiality assessments for each AI use case, system or model used by the FI. Such assessments are critical for calibrating the AI risk management approach, ensuring that controls are applied proportionate to the risks posed by AI. This ensures that the AI use case, system or model assessed as having higher risk materiality receives more stringent scrutiny and robust controls, supporting effective oversight and confirming that AI usage remains within approved boundaries.

3.9 **The assessment should take into account both the inherent risk materiality of an AI use case, system or model before the appropriate risk management controls are applied, as well as the residual risk materiality after risk management controls are applied.** The FI should ensure that the residual risk materiality of the AI use case, system or model meets the FI's risk appetite before deployment. The risk materiality assessment methodology and risk materiality assessments for each AI use case, system or model should also be regularly reviewed to ensure their continued relevance and appropriateness.

3.10 **The risk materiality assessments should consider various risk dimensions relevant to the FI's context**, minimally covering:

a. **Impact**: The potential consequences of a failure, malfunction or poor performance of the AI system or model on the FI (e.g., financial, operational, regulatory, reputational) and its customers or other stakeholders (e.g., fairness, ethical breaches, consumer protection). The nature and sensitivity of the data processed by the AI system or model should also be considered.
b. **Complexity**: Arising from the nature of the AI technology used, the novelty of its application, or the data it uses. As the understanding of AI technologies evolves across time, this risk dimension could also evolve, e.g., with more research and greater familiarity, the complexity of a new AI technology that was not as well understood initially may change.
c. **Reliance**: Considering the level of autonomy granted to the AI system or model, the degree of human involvement or oversight in the process it supports; as well as the availability of alternatives.

3.11 **An FI should assign clear roles and responsibilities for AI risk materiality assessment.** A control function should be assigned to ensure the consistent application of the assessment process across the FI, setting up

attestation processes, ensuring clear documentation is maintained, and acting as the final arbiter in determining the risk materiality of an AI use case, system or model.

4 AI LIFE CYCLE[13] CONTROLS

4.1 **An FI should plan for and implement robust controls covering the entire life cycle of an AI use case, system or model and assign clear roles and responsibilities for such controls.** AI life cycle controls should be regularly reviewed to take into account the use of newer AI technologies.

4.2 **For each AI use case, system or model, an FI should clearly define its use case and assign clear roles and responsibilities for the entire life cycle across different business and functional units.** The FI should also conduct risk materiality assessments and capture available information in its AI inventory. Such assessments and inventory information should be reviewed and updated across the AI life cycle when new information is available.

4.3 **An FI may implement the relevant AI lifecycle controls in a proportionate manner, ensuring that they are appropriate to the assessed risk materiality of AI[14].** Where there are practical constraints, e.g., where there may not be adequate disclosure of information by the third-party AI provider, the FI should identify risks that arise from such constraints, and put in place the necessary mitigants, such as limiting the usage of AI. After putting in place such mitigants, the FI should ensure that residual risks remain within the risk appetite of the FI.

4.4 **For an AI use case, system or model assessed as high risk, an FI should develop and implement contingency plans.** These plans should outline fallback options, such as alternative systems or manual processes, to ensure business continuity in case of AI failure or unexpected behaviour. Plans should be regularly reviewed and tested for effectiveness. For AI with "kill switches"[15], clear contingency activation protocols should be in place and tested regularly.

Data Management

4.5 **An FI should put in place data management controls to ensure data used across the AI life cycle is fit for purpose and representative, of high quality, and subject to robust data governance.** General data governance and management standards, in areas such as data ownership, access controls, and intellectual property rights, should be applied to data used for AI, and, where necessary, uplifted to address AI-specific requirements. Key areas that the FI should consider include:

---

[13]The AI life cycle refers to its evolution from inception to retirement or decommissioning (adapted from ISO/IEC 22989 – AI Concepts and Terminology).

[14]In certain cases, AI may not be fully deployed from the onset and could be partially deployed, e.g., as pilots or phased roll-outs. Such partial deployments of AI may require adjustments to lifecycle standards, processes and controls designed for full deployments, e.g., certain testing may only be conducted after information from the partial deployments are available, or pre-deployment reviews may not have been fully completed. For such partial deployments, the FI should establish clear policies and procedures to govern any deviations from standard lifecycle standards, processes and controls. Such policies and procedures to govern any deviations from standard lifecycle standards, processes and controls may include time and user limits, clear criteria for success, conditions on the terms of use for owners and end-users, close monitoring of usage patterns and outputs for anomalies and ensuring compliance with the limited scope of usage.

[15]"Kill switches" are usually used to refer to mechanisms that can be used to deactivate AI expeditiously if they exceed risk tolerances.

a. **Fit for Purpose:** Suitability of data used in an AI use case, system or model based on the intended objective and context, including assessing the use of data against other considerations, such as fairness.
b. **Representativeness of Data:** Representativeness of data used for training and testing the AI use case, system or model across the full range of real-world conditions, including stressed conditions, under which the AI will be used.
c. **Data Quality:** Adequacy of the quality of data used in an AI use case, system or model, including assessing data relevance, accuracy, completeness, and recency; as well as regular monitoring of data quality and checks for anomalies, drifts, and potential bias.
d. **Data Classification:** Appropriate data classification processes to guide appropriate use of data in an AI use case, system or model, taking into account criticality and sensitivity of data used.
e. **Data Security:** Secure data lifecycle management practices, including timely destruction or sanitisation of training data, model artefacts and outputs once decommissioned or no longer required; appropriate data protection measures, such as data encryption in transit and at rest, secure data handling, including securing the inputs and outputs of the AI use case, system or model.
f. **Data Privacy:** Appropriate data privacy measures based on relevant regulatory requirements and guidance relating to data privacy[16]; and obtaining permission when using sensitive personal data of customers or employees to train the AI models, or allowing AI to access such data in real-time.
g. **Auditability and Lineage of Data:** Appropriate documentation of key data management aspects, such as data sourcing, selection, processing and lineage; how data was assessed as fit-for-purpose and representative; and recording approvals and remedial actions related to data.

Transparency and Explainability

4.6 **An FI should determine the extent of transparency and explainability[17] required of an AI use case, system or model according to its assessed risk materiality, and establish the relevant controls accordingly[18].** Transparency is important to support accountability and trust in the use of AI, and enables customers, internal users and other stakeholders to make informed assessments of the risks, reliability and limitations of an AI use case, system or model. It also allows FIs to demonstrate that their use of AI is in line with their stated objectives and risk management standards, and helps to strengthen trust among customers and other stakeholders.

4.7 **Key considerations on the degree of transparency and explainability required may include reliance on AI for the final decision (i.e., the degree of AI autonomy), level of impact on customer or risk management outcomes.** For example, AI that is heavily relied upon for credit decisioning, insurance underwriting or other regulated activities with high impact on customer outcomes (e.g., provision of financial advisory services or fund management), will require more exacting standards for explainability. For such AI use cases, the FI should pay greater attention to the different features or attributes in data used as inputs and justifications for their use, the

[16]Such as the Personal Data Protection Commission's Advisory Guidelines on use of Personal Data in AI Recommendation and Decision Systems https://www.pdpc.gov.sg/guidelines-and-consultation/2024/02/advisory-guidelines-on-use-of-personal-data-in-ai-recommendation-and-decision-systems

[17]Transparency refers to disclosures made to individuals or groups of individuals related to the use of AI that affects them, as well as the provision of explanations where relevant and requested. Explainability relates to methods used for facilitating an understanding of AI generated outputs or decisions. The level of explainability required may vary for different AI use cases, and the standard of explainability required may take into account risk materialities as well as the extent to which AI-driven decisions are likely to require explanations (e.g., to facilitate decision making by end-users, or to account to the FI's customers) for the AI use case.

[18]For example, based on areas relating to transparency and explainability covered in the NIST AI Risk Management Framework at https://doi.org/10.6028/NIST.AI.100-1.

ability for users to identify key drivers of the output, the need to inform customers of the use of AI, the consequences of AI-driven decisions, and channels for redress.

Fairness

4.8 **An FI should define what it considers "fair" outcomes and have appropriate controls to identify and mitigate harmful biases and discriminatory outcomes across the AI life cycle, calibrated to its assessed risk materiality**. For example, greater attention should be paid to AI used in areas such as credit decisioning or insurance underwriting that may lead to unfair access or denial of financial services or products offered to individuals.

4.9 Where fairness considerations are relevant, the FI should conduct fairness assessments, which may involve defining relevant protected attributes[19], assessing whether the use of AI leads to systematic disadvantages for specific groups using appropriate fairness metrics, and documenting the results and any mitigation steps taken.

Human Oversight

4.10 **An FI should put in place and regularly review controls to ensure appropriate human oversight over an AI use case, system or model[20] across its life cycle.** The need for and degree of human oversight should take into account the objective of using AI, and be proportionate to the risk materiality of AI used. It should also take into account automation bias and decision fatigue as the use of AI increases in speed and scale. Key areas that the FI should consider include:

a. **Roles and Responsibilities:** Clear assignment of roles and responsibilities for human oversight, including escalation and decision-making processes relating to human oversight.
b. **Capabilities:** Equip competent personnel assigned to monitor AI use with the necessary capabilities, including the necessary authority and ability to intervene.
c. **Design**: Designing and developing AI systems or models from the outset to enable and facilitate appropriate human oversight.
d. **Documentation and Review:** Establishing processes to document and regularly review human oversight decisions and interventions (including incidents as well as near misses) to assess the effectiveness of human oversight.

Third-Party AI Management

4.11 **An FI should ensure that onboarding, development and deployment controls for third-party AI are adequate for the risk materiality of the use case, system or model which uses or depends on third-party AI[21].** This would include testing third-party AI products and services in the context of the FI's use cases (including using the FI's own data), and performing compensatory testing to address informational gaps arising from inadequate disclosures by third-party AI providers. The FI should also ensure that AI developed by third parties are subject to appropriate reviews, as well as establish processes to receive notifications of updates or changes to third-party AI and manage and assess the impact of such updates or changes. Key areas that the FI should consider:

[19]Attributes that are substitutes for or highly correlated to the protected attributes should also be considered.
[20]This would include all instances where the FI has assessed that human involvement is necessary, regardless of how human oversight is implemented – e.g., human in the loop or over the loop.
[21]MAS expectations and requirements on managing the risks from outsourcing and use of third-party services would apply to the use of third-party AI.

a. **Transparency**: Assessing the level of transparency from third-party AI providers on how key risks, such as those relating to data, model, technology and cybersecurity risks, are addressed during the development and deployment of such third-party AI; and setting out clear and consistent expectations on the level of transparency needed from third-party AI providers across the FI. Where transparency and explainability is required but not available for third-party AI, the FI should consider employing compensatory measures, such as additional testing to understand the behaviour of third-party AI, applying greater human oversight, or making the appropriate disclosures to users.
b. **Fairness**: Exercise due diligence regarding the fairness practices of third-party AI providers, as the FI remains accountable for the fairness outcomes of third-party AI used in the FI.
c. **Supply Chain Assessments**: Checking whether key third-party and open-source AI models, datasets, and dependencies have undergone supply chain risk assessments and validation, including reviews of model provenance, training data integrity, and other known vulnerabilities.
d. **Concentration Risks**: Assessing potential concentration risks arising from over-reliance on key third-party AI providers (whether direct or indirect).
e. **Contingency Plans**: Developing robust contingency plans to address potential failures, unexpected behaviour of third-party AI, or discontinuing of support by vendors, particularly for third-party AI used in use cases, systems or models where the risk materiality is high.
f. **Legal Agreements**: Updating legal agreements to facilitate clearer expectations and responsibilities, such as clauses pertaining to performance guarantees, data protection, the right to audit, notification when AI is introduced, or seeking the FI's agreement before incorporating AI.
g. **Capabilities**: Building awareness and developing capabilities of staff involved in procurement, development, deployment and use of third-party AI.
h. **Complexity**: Conducting more detailed assessments when using more complex third-party AI products and services that the FI may have less experience with[22].

Selection

4.12 **When selecting AI algorithms[23] or features[24] in data to use, an FI should consider the objectives and risks of the AI use case, system or model.** The FI should require developers to justify and document their selection process, particularly when selecting more complex algorithms or less understood features over simpler or conventional alternatives. This may involve balancing factors like the needs of the use case and performance requirements against complexity, the need for fairness, or transparency and explainability. The selection should be supported by theory, research, or accepted industry practice where possible. The FI should also consider incorporating reviews by domain experts or users (e.g., by subject matter experts or users in the relevant business lines or functional units), to ensure that the selection of algorithms or features aligns with the context of the AI use.

[22]For example, the FI may need to do more detailed evaluation and testing, as well as technology and cybersecurity checks if they have not had any experience with developing and deploying AI agents provided by third-party providers.
[23]This may include specific AI models (e.g., specific machine or deep learning models) or techniques (e.g., optimisation or finetuning techniques).
[24]Features refer to the attributes of data points in a dataset, e.g., for data relating to a loan, the income of the obligor and outstanding value of the loan are two possible attributes or features. Feature engineering refers to the process of selecting, modifying or creating new features from the original attributes of a dataset to improve an AI model's performance, e.g., normalising income of the obligor and outstanding value of the loan to a common scale ranging from 0 to 1; or creating new derived features, such as a debt-to-income ratio, from existing attributes.

4.13 Where newer and more complex AI algorithms which are less understood are selected, an FI should carefully weigh the benefits of deploying such AI algorithms against new or heightened risks to the FI, such as hallucination, opaqueness, security risks, and the FI's capabilities to mitigate such risks.

Evaluation and Testing

4.14 **An FI should conduct relevant evaluation and testing that is proportionate to the assessed risk materiality of the AI use case, system or model.** Each AI use case, system or model should be evaluated and tested to meet an appropriate level of reliability and safety based on the assessed risk materiality before deployment. An FI should identify key AI risks and set clear, measurable thresholds. Identification of AI risks and reliability and safety assessments should take into account the context of AI use, and assessments should take reference from best practices[25]. Evaluation and testing should assess performance of the AI use case, system or model under a range of plausible conditions, from real-world scenarios to edge cases. Key areas that the FI should consider include:

a. **Evaluation Measures**: Define appropriate evaluation measures aligned with the AI's objectives and establish acceptable performance thresholds for these evaluation measures. Performance thresholds should be clearly defined, documented, and mutually agreed upon by business owners, developers and reviewers.
b. **Testing Approaches**: Employ relevant testing methods, such as out-of-sample or out-of-time testing, sensitivity analysis, stability analysis across different data distributions or time periods, sub-population analysis, stress testing (including edge cases and adversarial testing where appropriate), error analysis, and benchmarking against alternatives. Datasets used for testing should be representative of the context of use in the FI.
c. **Overfitting Mitigation**: Where possible, implement techniques to prevent overfitting, especially for more complex AI. This may involve favouring simpler models unless higher complexity is justified by a clear performance uplift, constraining complexity (e.g., via regularisation of AI models[26]), appropriate feature selection, and using robust validation techniques like cross-validation.

4.15 **Where there are risks and limitations associated with AI identified during development, an FI should put in place appropriate controls and guardrails to mitigate these risks and limitations before deployment.** Evaluation and testing of newer AI developments such as Generative AI and AI agents should cover their key failure modes[27].

[25]Such as IMDA's Starter Kit for Safety Testing of LLM-Based Applications which can be accessed at https://www.imda.gov.sg/-/media/imda/files/about/emerging-tech-and-research/artificial-intelligence/large-language-model-starter-kit.pdf. The Starter Kit is a set of voluntary guidelines that coalesces emerging best practices and methodologies for the testing of LLM-based applications.

[26]Such techniques generally try to limit the number of parameters used so that the trained model is less complex, e.g., some regularisation techniques force less important parameters to values of zero.

[27]For example, for Generative AI, evaluation and testing should pay greater attention to hallucinations, generation of undesirable content such as toxic, harmful or biased content, bias, data leakages or disclosures, and vulnerability to adversarial attacks. For more information on the testing of such risks, please refer to IMDA's Starter Kit for Safety Testing of LLM-Based Applications.

Technology & Cybersecurity Risks

4.16 **An FI should ensure that the AI system is secure, well-governed, and supported by appropriate controls to manage technology and cybersecurity risks.** Relevant regulatory requirements and guidance relating to technology risk management would apply[28]. Key areas that the FI should consider include:

a. **Security:** Deployment should occur in secured IT environments, including hardened configurations, network segmentation, and technical controls such as input validation, API authentication, encryption, and data loss prevention.
b. **Access Control:** Access to AI components and infrastructure during deployment should be tightly controlled, with role-based access and multi-factor authentication; and especially in privileged account access management. Separation of duties, such as assigning different teams to train and test the AI systems, should also be considered.
c. **Third-party:** Where third-party components or services are used, such as plugins or APIs, controls should be implemented to govern usage, restrict data exposure, and monitor for security, compliance or operational risks, such as cybersecurity issues or service disruptions at third-party providers.

Reproducibility & Auditability

4.17 **An FI should document the AI development process to enable reproducibility and auditability.** Documentation should be sufficiently detailed for an independent party, such as a reviewer or auditor, to understand and potentially replicate the implementation of the AI system or model and its results. Documentation standards should cover the entire development process and may include information such as:

a. data sources, processing, and quality checks;
b. selection rationale;
c. training procedures including code versions, environments, and hyperparameters;
d. evaluation measures and performance thresholds, testing approaches, and results;
e. explainability analysis, fairness assessments; and
f. key assumptions, limitations and mitigants.

Pre-Deployment Reviews

4.18 **Prior to deployment, an FI should subject the AI use case, system or model to reviews by parties not involved in its development to ensure that the relevant controls, such as evaluation and testing, have been adhered to**. Such reviews should be supported by clear policies and procedures, and assess the suitability, robustness, and performance of AI for its intended purpose within defined boundaries. Roles and responsibilities relating to such reviews should be clearly assigned, and the scope and independence of these pre-deployment reviews should be proportionate to the AI's assessed risk materiality.

4.19 **The scope of these pre-deployment reviews, as well as the degree of independence of the parties undertaking the assessment and reviews, should be proportionate to the assessed risk materiality of the AI use case, system or model.** AI use cases, systems or models assessed as having high risk materiality should undergo

[28]Such as MAS' Guidelines on Risk Management Practices for Technology Risk which can be accessed at https://www.mas.gov.sg/regulation/guidelines/technology-risk-management-guidelines.

formal independent validation before deployment. Such validations should be conducted by competent personnel or functions possessing the necessary expertise and objectivity, and who are independent from the development and deployment teams. The validation process should provide effective challenge to developers and should cover areas such as:

a. conceptual soundness of the design;
b. suitability and quality of data inputs;
c. integrity of the implementation;
d. evaluation measures, performance thresholds, testing approaches and results;
e. explainability analysis and fairness assessments; and
f. assumptions, limitations and mitigants.

4.20 For AI use cases, systems or models not assessed as having high risk materiality, other forms of documented review (such as peer reviews by qualified individuals that were not involved in development or deployment[29]) on key aspects of the development and deployment processes may be undertaken.

4.21 **Review findings, including any identified limitations, necessary remediation actions, or conditions for use, should be reported by reviewers to the relevant approval body.** The relevant approval body should ensure review recommendations are actioned appropriately.

4.22 **An FI should conduct technology and cybersecurity reviews to ensure that AI can be deployed into the production environment in a controlled and secure manner.** Such reviews should cover areas such as technical implementation, system and network security, system resilience, recoverability and operational readiness of AI for deployment. Key areas that the FI should consider include:

a. **Secure Design and Access Control:** Review the AI system architecture, including APIs, plugins and third-party services, for secure design, environment segregation, and access controls based on the principle of least privilege. Prior to go-live, complete deployment checklists to confirm that critical controls, including encryption, data loss prevention, firewalls, access restrictions and logging, are operational and correctly configured.
b. **Testing and Threat Mitigation:** Conduct vulnerability assessments, penetration testing, and red teaming across AI systems to identify and remediate security weaknesses. Test AI systems against adversarial threats (e.g., manipulation, evasion, poisoning) and implement safeguards, such as input validation, throttling and anomaly detection.

Post Deployment Monitoring & Review

4.23 **An FI should develop and implement comprehensive and robust controls for the ongoing monitoring of all deployed AI (including third-party AI used in the FI).** Given the uncertainties associated with AI, their dynamic nature and the potential for model staleness and performance degradation due to data or model drifts over time, ongoing monitoring is critical to ensure that deployed AI operates as intended and remains fit for purpose over time. The frequency and intensity of monitoring activities should be proportionate to the assessed risk materiality of the deployed AI. Key areas that the FI should consider include:

a. **Monitoring Measures**: Based on the risks of the AI use case, system or model, define key metrics to be monitored and acceptable performance thresholds for each metric. Metrics may include relevant

[29]This includes both individuals involved in initial deployment of AI, as well as ongoing deployment of AI, e.g., operating the AI system.

dimensions like robustness, stability, data quality, and fairness. Tiered thresholds (e.g., early warning threshold levels) should be considered to pre-empt deterioration and trigger early actions. Appropriate checks for data drifts (changes in input data distributions), concept drifts (changes in relationships between inputs and outputs), and overall model drifts should be implemented. Where relevant, information flow and decision-making paths across workflows that use AI, such as reasoning processes, actions taken, tools used should also be monitored.

b. **Incident and Issue Management**: Establish robust processes for reporting, tracking, escalating, and resolving issues or incidents if breaches or anomalies arise from the monitoring process. Resolution approaches may include retraining, adjustments, redevelopment, or decommissioning of the AI model or system, depending on the severity and nature of the issue. For high risk materiality AI, consider implementing "kill switches" or override mechanisms to rapidly deactivate the AI system if it exceeds risk tolerances. Provide users of deployed AI with appropriate mechanisms to provide feedback and report issues to support continuous improvement, redress and resolution of issues.
c. **Roles and Responsibilities**: Assign clear roles and responsibility for ongoing monitoring and managing of incidents of deployed AI, including the designation of an appropriate accountable person.
d. **Documentation**: Maintain clear records of monitoring activities, results, identified issues or incidents, and subsequent remediation actions taken for auditability and ongoing risk management. Access to AI models or systems, training data, pipelines, and configuration files should be strictly controlled and logged.
e. **Training and Awareness**: Equip persons responsible for monitoring with appropriate training to support effective monitoring and incident management; and equip users of deployed AI with the necessary training and awareness to report unintended behaviour of deployed AI.

4.24 **An FI should periodically review aggregate risks across all AI use cases, systems and models, and, where necessary, conduct more detailed reviews or re-validations of specific AI in the portfolio**. **For AI use cases, systems or models assessed as having high risk materiality, regular re-validations by independent parties should be conducted.** Such reviews or validations ensure that AI continue to be used appropriately and that risks are still managed appropriately. The need for detailed re-validations or reviews of specific AI should be informed by the AI's assessed risk materiality. Such reviews may also be triggered by findings from the periodic reviews of aggregate risks across all deployed AI, alerts or breaches from ongoing monitoring, significant changes to the AI or its operating environment, or the identification of new risks in the external environment, such as regulatory or technological developments.

Change Management

4.25 **An FI should develop and implement comprehensive and robust controls for managing changes to deployed AI.** Such controls are essential to ensure that any changes do not lead to unintended behaviour, performance degradation or misalignment with the intended use of deployed AI. Key areas that the FI should consider include:

a. **Scope of Change**: Clearly define what constitutes a significant or material change (e.g., changes to model or system architectures, key assumptions, or intended use) versus minor changes (e.g., retraining with updated data). Material changes should trigger appropriate reviews and re-approval processes before implementation.
b. **Change Control**: Implement change control mechanisms (e.g., human-in-the-loop) to prevent unauthorised alterations to AI system or models and ensure changes to AI (such as underlying code, configurations, or training data) do not result in unintended consequences such as service outages,

operational disruptions, or degradation in model performance. This may be supported by version control systems that track changes not only to the AI model code but also to associated data, parameters, hyperparameters, and other key artifacts, and support traceability, auditability, and the ability to roll back to previous versions.

c. **Dynamic Updates**: Implement enhanced controls for AI that are designed to be updated automatically, including strict justifications for enabling automatic updates, clear definitions of what can be updated automatically (such as allowing re-training or changes to hyperparameters but not changes to core architecture), enhanced data quality checks, and more stringent performance monitoring.

4.26 **An FI should develop and implement clear controls for the eventual retirement or decommissioning of AI when they are no longer needed or exceed risk tolerances.** These controls should consider dependencies, data retention policies, secure removal from the production environment, and appropriate stakeholder notifications.

5 AI CAPABILITY & CAPACITY

AI Risk Management Capabilities

5.1 **An FI should determine and ensure the necessary competence and proper conduct of personnel involved in developing and deploying an AI use case, system or model.** This includes recruiting the necessary talent and providing appropriate training and support to equip employees with the skills, knowledge, and culture for the appropriate use of AI, as well as effective AI risk management. Adequate resources, including human, technological, and financial, should be allocated proportionate to the risk profile of the AI use case, system or model.

5.2 **Regular reviews should be conducted to ensure that the relevant personnel involved in developing, deploying and maintaining an AI use case, system or model are equipped with adequate capabilities and capacity for effective AI risk management.** This includes regularly reviewing programmes for building capabilities and capacity for effective AI risk management to ensure that they are up to date and incorporate training to address risks associated with newer AI technologies.

Technology Infrastructure for AI

5.3 **An FI should ensure that its technology infrastructure is adequate for an AI use case, system or model.** Like other information technology systems, AI systems and models used in such systems are also subject to technology risks, including system availability, resilience, safety and cybersecurity risks. To address these risks, the FI should ensure that the underlying hardware and software resources, such as graphics processing units, network infrastructure, system memory, secure data pipelines are sufficient to meet performance, scalability, and resilience needs for the AI use case, system or model. In determining the appropriate technology infrastructure for a specific AI use case, system or model, the FI should take into account relevant technology risk management guidelines and notices[30], as well as relevant industry frameworks[31].

[30]https://www.mas.gov.sg/regulation/guidelines/technology-risk-management-guidelines
[31]For example, technology infrastructure related areas covered in the NIST AI Risk Management Framework at https://doi.org/10.6028/NIST.AI.100-1.

ANNEX

Proportionate Application of the Guidelines on AI Risk Management

1. The Guidelines are designed to be applied in a proportionate manner across FIs of different sizes and risk profiles. FIs may refer to the following flowchart when assessing the applicability of the Guidelines. As the use of AI in the FI evolves, the FIs should reassess the applicability of the Guidelines and adjust its framework, policies and processes to support their growing use of AI and responsible innovation.

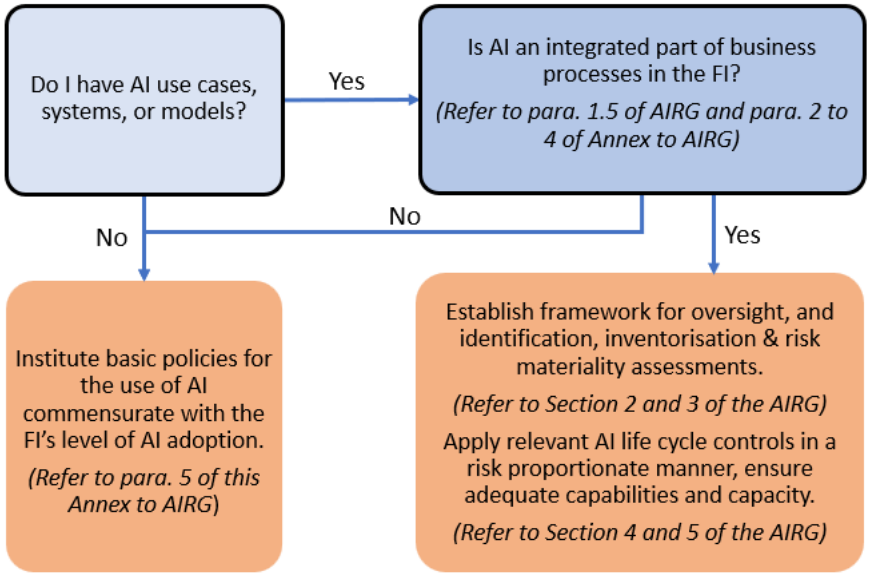

Figure 2: Assessing the Applicability of the AIRG

2. If the answer is yes to either of the guiding questions below, then AI would be regarded as being used as an integrated part of business processes in the FI.

   a. Would the lack of access to AI services or tools disrupt workflows that the FI is materially dependent on for its business activities?
   b. Is AI integrated with the FI's systems which it is materially dependent on for its business activities?

3. FIs with AI use cases similar to any of the following examples would not be regarded as using AI as an integrated part of their business processes[32], but should still apply the basic policies set out in paragraph 5 of this Annex:

   a. Individual analysts using a third-party large language model (LLM) to assist in drafting email responses to customers;
   b. Customer relationship managers using AI-enabled tools for grammar, spell checking or rephrasing of emails to customers;

[32]Humans involved should be using AI in these use cases in an assistive manner and will be reviewing and checking the outputs before using them.

c. Investment research team members using AI to summarise investment research papers for understanding;
d. Finance team members using AI to assist in generating formulas or charts in Excel or visualisation tools;
e. Marketing staff using AI image generators to assist in designing marketing materials; or
f. Insurance claims assessors using AI on an ad-hoc basis to assist in an initial review of claims documents.

4. FIs with AI use cases similar to any of the following examples would be regarded as using AI as an integrated part of their business processes [33]:

a. In-house legal team using AI legal contract review tool for standard agreements - used systematically by entire legal team; removal would significantly slow contract review processes;
b. Internal IT helpdesk chatbot powered by Generative AI used by a substantial proportion of employees - first point of contact for IT support; without it, helpdesk would be overwhelmed with ticket volumes;
c. AI-enabled financial data extraction from PDFs used by a substantial proportion of analysts - process high volumes of documents; manual extraction would be time-prohibitive;
d. AI-enabled research aggregation tool used by a substantial proportion of investment analysts - embedded in research workflow; significantly delay to research process if unavailable;
e. AI-enabled investment idea screener used by a substantial proportion of researchers - integrated into idea generation process; removal would require analysts to manually screen much larger universes;
f. Incorporation of AI into insurance claims assessment process – AI core part of workflow that processes supporting documents; unavailability of AI would delay claims processing; or
g. Incorporation of AI into financial advisory process – AI-enabled financial advisory tool used as a core part of the workflow for providing financial advice or product recommendations to customers; unavailability would significantly slow or reduce the ability to deliver advice or recommendations, and reverting to manual processes would be time-consuming.

5. Basic policies for the use of AI should be commensurate with the FI's level of AI adoption. Such policies should address the following aspects:

a. Designating a senior management member as the responsible person for AI oversight;
b. Guidelines covering areas such as allowed and disallowed uses of AI (e.g., prohibitions on inputting confidential, proprietary, or client information into public AI tools), requirement for human review and validation of all AI outputs before use, maintaining a list of approved AI tools, process for requesting approval of new AI tools;
c. Communication of guidelines to all staff;
d. Regular checks for compliance with guidelines; and
e. Annual review and update of guidelines.

6. FIs may seek clarifications from MAS on any areas relating to the proportionate application of the Guidelines if they are not addressed by this Annex.

[33] For avoidance of doubt, all AI use cases that introduce significant risks, e.g., deployment of AI in critical business lines or functional areas, or systematic use of AI in relation to the conduct of regulated activities, would be regarded as using AI as an integrated part of business processes.

Monetary Authority of Singapore