Monetary Authority of Singapore

Information Paper

December 2024

ARTIFICIAL INTELLIGENCE MODEL RISK MANAGEMENT

OBSERVATIONS FROM A THEMATIC REVIEW

Contents

1. Overview

1.1 This information paper sets out good practices relating to Artificial Intelligence (AI) (including Generative AI)[1] model risk management (MRM)[2] that were observed during a recent thematic review of selected banks. The information paper focuses on the following key areas[3]: AI governance and oversight; AI identification, inventorisation and risk materiality assessment; as well as AI development, validation, deployment, monitoring and change management.

1.2 While the thematic review focused on selected banks, the good practices highlighted in this information paper should generally apply to other financial institutions (FIs), which should take reference from these when developing and deploying AI.

2. Background

Industry use of AI and Generative AI and associated risks

2.1 The launch of ChatGPT in November 2022 and recent advancements in AI, particularly Generative AI, has led to an increased interest in leveraging AI and Generative AI in the banking and broader financial sector. Prior to these developments, FIs have used AI in a wide range of areas and use cases. Key areas where we observed significant use of AI by banks during the thematic review include risk management, customer engagement and servicing, as well as to

[1] Generative AI is a subset of AI, and an AI or Generative AI *system* can comprise one or more AI or Generative AI *models* and other machine-based components. For the purposes of this paper, AI generally refers to both AI and Generative AI models and systems. Where a point pertains specifically to an AI model or an AI system, or to Generative AI, we will use the respective terms explicitly in the paper. We define the terms AI and Generative AI, as well as AI model, system and use case in greater detail in Annex A.

[2] In line with the footnote above and recognising that the AI MRM is intrinsically linked to the risk management of systems in which AI models are used, when we refer to AI MRM or AI risk management in this paper, it generally refers to the risk management of AI models and systems.

[3] The aim of this information paper is not to cover all aspects of model risk management, but to focus on good practices in areas that are more relevant to AI MRM.

support internal operational processes. For example, we have seen banks use AI, particularly decision tree-based machine learning (ML) models such as XGBoost, LightGBM and CatBoost[4], in financial risk management to detect abnormal financial market movements, or to estimate loan prepayment rates. They are also commonly used in anti-money laundering (AML) systems to detect suspicious transactions, and in fraud detection systems. In customer engagement and servicing, banks use AI to predict customer preferences, personalise financial product recommendations and manage customer feedback. AI is also widely used to support internal operational processes across a wide range of business functions, for example, to automate checking and verification processes (e.g., for customer information), prioritise incident management (e.g., triaging IT incidents for attention), or forecast demand for services (e.g., ATM cash withdrawals).

2.2 While the use of AI in these areas can enhance operational efficiency, facilitate risk management and enhance financial services, they can also increase risk exposure if not developed or deployed responsibly. Potential risks include:

- **Financial risks**, e.g., poor accuracy of AI used for risk management could lead to poor risk assessments and consequent financial losses.

- **Operational risks**, e.g., unexpected behaviour of AI used to automate financial operations could lead to operational disruptions or errors in critical processes.

- **Regulatory risks**, e.g., poor performance of AI used to support AML efforts could lead to non-compliance with regulations.

- **Reputational risks**, e.g., wrong or inappropriate information from AI-based customer-facing systems, such as chatbots, could lead to customer complaints and negative media attention, and consequent reputational damage.

[4] Decision tree-based ML models make predictions based on a tree-like structure learnt from data. Models such as XGBoost, LightGBM and CatBoost utilise a series of decision trees together with a boosting technique. Each decision tree in the series focuses on the errors made by a prior decision tree to improve predictions. Such models are also explainable as the relative importance of different features to model predictions can be extracted.

2.3 While natural language processing (NLP)[5] and computer vision (CV)[6] techniques were already in use in the financial sector prior to the emergence of Generative AI[7] for text or image-related tasks, recent Generative AI models such as OpenAI's GPT[8] large language models (LLMs) and DALL-E[9] image generation models, or Anthropic's Claude LLMs[10] offer better performance in tasks such as clustering documents. They have also enabled new use cases, e.g., to generate text content and images for marketing, or to process multimodal data[11] for financial analysis.

2.4 Based on the thematic review, use of Generative AI in banks appears to still be at an early stage. The current focus is on the use of Generative AI to assist or augment humans for productivity enhancements, and not in applying Generative AI to direct customer facing applications. Use cases being explored by banks include risk management (e.g., detecting emerging risks in text information); customer engagement and service (e.g., summarising customer interactions or generating marketing content); and research and reporting (e.g., investment analyses). Banks are also exploring the use of Generative AI in copilots[12] to support staff, for example, in coding, or for general text-related tasks such as summarisation and answering queries based on information in internal knowledge repositories.

2.5 With Generative AI, existing risks associated with AI may be amplified[13]. For example, Generative AI's potential for hallucinations and unpredictable

[5] Natural language processing (NLP) is commonly used to refer to techniques that process, analyse, make predictions or generate outputs relating to human language, both in its written and spoken forms.

[6] Computer vision (CV) is commonly used to refer to techniques that enable machines to process and generate outputs based on visual information from the world.

[7] For example, for news sentiment analysis, information extraction, clustering documents based on underlying topics, or digitising physical documents.

[8] Generative Pre-trained Transformers (GPT) are a family of Generative AI models developed by OpenAI, that includes models such GPT 4 and GPT-4o.

[9] DALL-E is a Generative AI model that generates images from text prompts or descriptions.

[10] Claude models are a family of Generative AI models developed by Anthropic and include models such as Claude 3.5 Haiku and Sonnet.

[11] Multimodal data refers to datasets that comprise multiple types of data, e.g., text, images, audio or video.

[12] In the context of Generative AI, the term copilot is typically used to refer to Generative AI being used to assist or augment humans on specific tasks.

[13] More details on risks associated with Generative AI have already been covered extensively in Project MindForge's white paper on "Emerging Risks and Opportunities of Generative AI for Banks" and will not be repeated in this information paper. The whitepaper can be accessed at https://www.mas.gov.sg/schemes-and-initiatives/project-mindforge.

behaviours may pose significant risks if Generative AI is used in mission-critical areas. The complexity of Generative AI models and lack of established explainability techniques also creates challenges for understanding and explaining decisions, while the diverse and often opaque data sources used in Generative AI training, coupled with difficulties in evaluating bias of Generative AI outputs, could lead to unfair decisions.

MAS' Efforts on Responsible AI for the Financial Sector

2.6 Alongside the growing use of AI in the financial sector and such associated risks, MAS had established key principles to guide financial institutions in their responsible use of AI.

2.7 In 2018, MAS co-created the principles of Fairness, Ethics, Accountability and Transparency (FEAT) with the financial industry to promote the deployment of AI and data analytics in a responsible manner. To provide guidance to FIs in implementing FEAT, MAS started working with an industry consortium on the Veritas Initiative[14] in November 2019. The Veritas Initiative aimed to support FIs in incorporating the FEAT Principles into their AI and data analytics solutions, and has released assessment methodologies, a toolkit, and accompanying case studies.

2.8 With the emergence of Generative AI, Project MindForge[15], which is also driven by the Veritas Initiative, was established to examine the risks and opportunities of Generative AI. The first phase of Project MindForge was supported by a consortium of banks and released a risk framework for Generative AI in November 2023.

2.9 More recently, MAS released an information paper relating to Generative AI risks in July 2024[16]. The paper provides an overview of key cyber threats arising from Generative AI, the risk implications, and mitigation measures that FIs could take

[14] See https://www.mas.gov.sg/schemes-and-initiatives/veritas
[15] See https://www.mas.gov.sg/schemes-and-initiatives/project-mindforge
[16] See https://www.mas.gov.sg/regulation/circulars/cyber-risks-associated-with-generative-artificial-intelligence

to address such risks. The paper 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.

3. Objectives and Key Focus Areas

3.1 This information paper, which focuses on AI MRM, is part of MAS' incremental efforts to ensure responsible use of AI in the financial sector. A key difference between an AI-based system and other systems is the use of one or more AI models within the system, which potentially increases uncertainties in outcomes. Robust MRM of such AI models is important to support the responsible use of AI.

3.2 As the maturity of AI MRM may vary significantly across different FIs, MAS conducted a thematic review of selected banks' AI MRM practices in mid-2024. The objective was to gather good practices for sharing across the industry.

3.3 Based on information gathered during the review, MAS observed good practices by banks in these key focus areas[17]:

- **Section 4: Oversight and Governance of AI**
    - Updating of existing policies and procedures of relevant risk management functions to strengthen AI governance;
    - Establishing cross-functional oversight forums to ensure that evolving AI risks are appropriately managed across the bank;
    - Articulating clear statements and principles to govern areas such as the fair, ethical, accountable and transparent use of AI; and

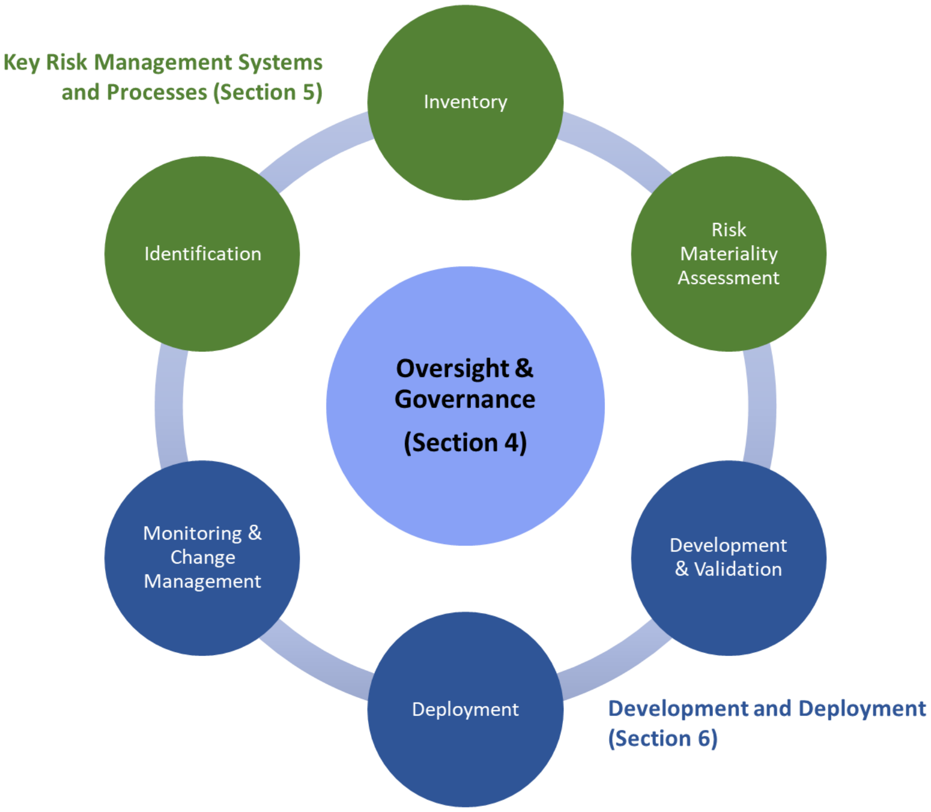
[17] For the purposes of the subsequent parts of this information paper, the good practices relating to AI would also apply to Generative AI as practicable. Specific considerations relating to Generative AI will be covered in Section 7.1.

- Building capabilities in AI across the bank to support both innovation and risk management.

- **Section 5: Key Risk Management Systems and Processes**
  - Identifying AI usage and risks across the bank so that commensurate risk management can be applied;
  - Utilising AI inventories, which provide a central view of AI usage across the bank to support oversight; and
  - Assessing the materiality of risks that AI poses using key risk dimensions so that relevant controls can be applied proportionately.

- **Section 6: Development and Deployment of AI**
  - Establishing standards and processes for key areas that are important for the development of AI, such as data management, robustness and stability, explainability and fairness, reproducibility and auditability;
  - Conducting independent validation or peer reviews[18] of AI before deployment based on risk materialities; and
  - Instituting pre-deployment checks and monitoring of deployed AI to ensure that it behaves as intended, and application of appropriate change management standards and processes where necessary.

[18] The terms validations and reviews are usually used interchangeably by banks to refer to assessments or checks of the AI model development process, whether by an independent party, or another peer developer. More details on such validations and reviews are provided in Section 6.4.


Key Risk Management Systems and Processes (Section 5)
Inventory
Risk Materiality Assessment
Identification
Oversight & Governance (Section 4)
Monitoring & Change Management
Development & Validation
Deployment
Development and Deployment (Section 6)

Overview of Key Thematic Focus Areas

3.4 These key focus areas are generally also applicable to Generative AI, as well as AI (including Generative AI) from third-party providers. Nonetheless, there may be additional considerations for Generative AI, as well as AI from third-party providers. Hence, additional observations on good practices relating to Generative AI and third-party AI are also outlined in Sections 7.1 and 7.2 of this information paper respectively.

3.5 The risks posed by AI and Generative AI extend beyond MRM and relate to non-AI specific areas such as general data governance and management, technology and cyber risk management, as well as third party risk management. These are not covered in this information paper, and existing regulatory requirements and supervisory expectations, including but not limited to notices, guidelines or information papers on data governance, technology and outsourcing risk management would apply, where relevant[19].

[19] Links to relevant publications are provided in Annex B.

4. Governance and Oversight

Overview

While existing control functions continue to play key roles in AI risk management, most banks have updated governance structures, roles and responsibilities, as well as policies and processes to address AI risks and keep pace with AI developments. Good practices include:

- establishing cross-functional oversight forums to avoid gaps in AI risk management;
- updating control standards, policies and procedures, and clearly setting out roles and responsibilities to address AI risks;
- developing clear statements and guidelines to govern areas such as fair, ethical, accountable and transparent use of AI across the bank; and
- building capabilities in AI across the bank to support both innovation and risk management.

Existing governance structures and such good practices are important to help support Board and Senior Management in exercising oversight over the bank's use of AI, and ensure that the bank's risk management is robust and commensurate with its state of use of AI.

4.1 While existing risk governance frameworks and structures[20] continue to be relevant for AI governance and risk management, a number of banks have established cross-functional AI oversight forums. Such forums serve as key platforms for coordinating governance and oversight of AI usage across various functions. They also play an important role in addressing emerging challenges and potential gaps in risk management as the AI landscape evolves, and ensuring that standards and processes, such as relevant AI development and deployment standards, are aligned across the bank.

[20] Aside from MRM, risk governance frameworks and structures from other areas that are usually relevant to AI risk management include (but are not limited to) data, technology and cyber, third-party risk management, as well as legal and compliance.

4.2 The mandates of these forums often include establishing a consistent and comprehensive framework for managing AI risks, evaluating use cases that require broader cross-functional inputs, and reviewing AI governance requirements to ensure they keep pace with the state of AI usage in the bank. Data and analytics, risk management, legal and compliance, technology, audit, as well as other relevant business and corporate functions, are typically represented at such cross-functional oversight forums.

4.3 A number of banks have also found value in compiling policies and procedures that are relevant to AI into a central guide to ensure that consistent standards for AI are applied across the bank. As more AI use cases are rolled out in banks, and the state of AI technology evolves, the use of AI may accentuate existing risks or introduce new risks. Hence, most banks have reviewed and, where necessary, updated existing policies and procedures to keep pace with the increasing use of AI across the bank, or new AI developments, e.g., updating policies and procedures relating to performance testing of AI for new use cases, or establishing new policies and procedures for AI models that are dynamically updated based on new data.

4.4 Given the broad range of use cases for AI, and the potential for inappropriate use, most banks have set out central statements and principles on how they intend to use AI responsibly, including developing guidelines to govern areas such as fair, ethical, accountable, and transparent use of AI[21]. Such efforts are important in setting the tone and establishing clear guidance on how AI should be used appropriately across the bank, and to prevent potential harms to consumers and other stakeholders arising from the use of AI. In addition to central statements and principles, some banks have also taken steps to operationalise such central statements and principles by mapping them to key

[21] More details on these areas can be found in MAS' publications relating to the FEAT principles under the Veritas Initiative. Similar principles covering areas relating to the responsible or ethical use of AI in the financial sector have also been published in other jurisdictions , e.g., the Hong Kong Monetary Authority (HKMA) issued guiding principles for the use of big data analytics and AI covering governance and accountability, fairness, transparency and disclosure, and data privacy and protection in 2019; De Nederlandsche Bank (DNB) issued the SAFEST principles on soundness, accountability, fairness, ethics, skills, and transparency in 2019.

controls, which are in turn mapped to the relevant functions responsible for these controls.

4.5 Given the growing interest in AI, banks also recognised the need to develop AI capabilities and have established plans to upskill both their staff and senior executives. Aside from building awareness, banks have developed AI training that facilitate staff in leveraging and using AI in an effective and responsible manner. Some banks have also set up AI Centres of Excellence to drive innovation, promote best practices and build AI capabilities across the bank.

5. Key Risk Management Systems and Processes

Overview

Most banks have recognised the need to establish or update key risk management systems and processes for AI, particularly in the following areas:

- policies and procedures for identifying AI usage and risks across the bank, so that commensurate risk management can be applied;
- systems and processes to ensure the completeness of AI inventories, which capture the approved scope of use and provide a central view of AI usage to support oversight; and
- assessment of the risk materiality of AI that covers key risk dimensions, such as AI's impact on the bank and stakeholders, the complexity of AI used, and the bank's reliance on AI, so that relevant controls can be applied proportionately.

5.1 Identification

5.1.1 Identifying where AI is used is important so that the relevant governance and risk management controls can be applied. Even when using widely accepted definitions, such as the Organisation for Economic Co-operation and

Development's definition of AI[22], considerable ambiguity remains around the definition of AI due to its broad and evolving scope.

5.1.2 Most banks leveraged definitions in existing MRM policies and procedures as a foundation for identifying AI models[23], and extended or adapted these definitions to account for AI-specific characteristics. Some banks shared that the uncertainty of model outputs is a common source of risk for both AI and conventional models[24], and that the presence of such uncertainties was a key feature that was usually considered when identifying AI. MRM control functions also typically play a key role in AI identification, often serving as the key control function responsible for AI identification systems and processes, e.g., setting up attestation processes, or acting as the final arbiter in determining whether AI is being used. Some banks have also developed tools or portals to facilitate the process of identifying and classifying AI across the bank in a consistent manner.

5.2 Inventory

5.2.1 Banks mostly maintain a formal AI inventory[25] with a comprehensive record of where AI is used in the bank. A key area that an AI inventory supports, alongside the relevant policies, procedures and systems, is to ensure that AI are only used within the scope in which they have been approved for use, e.g., the purpose, jurisdiction, use case, application, system, and other conditions for which they have been developed, validated and deployed. This is critical because unapproved usage of AI, particularly in higher-risk use cases, can lead to

[22] The OECD's definition of AI: An AI system is a machine-based system that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments. Different AI systems vary in their levels of autonomy and adaptiveness after deployment.

[23] This could entail step-by-step guide to facilitate the identification of techniques that meet the bank's definition of AI.

[24] Models usually refer to quantitative algorithms, methods or techniques that process input data into quantitative estimates which may be used for analysis or decision making. Apart from AI models, which typically refer to machine or deep learning models, banks also routinely utilise conventional models, such as economic, financial, or statistical models. Some quantitative algorithms, methods or techniques, such as logistic regressions, are commonly regarded as both AI and statistical models. A more detailed definition of models can be found in Annex A.

[25] Most banks have established software systems for their AI inventories that not only record where AI is used in the bank, but may also include additional features outlined above, such as automated tracking of approvals and issues, and identification of inter-dependences between AI. A small number of banks still rely on spreadsheets for their AI inventories, but this approach is more prone to operational issues, e.g., outdated records, and would not allow for the additional features outlined above.

unintended consequences. For example, AI approved for use in one jurisdiction should not automatically be treated as approved for use in others as the data, assumptions and considerations may not be similar, and the AI may not perform as expected in a different context.

5.2.2 A few banks also utilised their AI inventory system to track the use of AI through their lifecycle, and to establish checkpoints for different risk management processes at the various stages of the AI lifecycle. A few banks also used the AI inventory to support the identification and monitoring of aggregate AI risks and interdependencies across different AI models and systems. The AI inventory may also serve as a central repository for AI artifacts needed for model maintenance, validation and incident or issue management.

5.2.3 Most banks have established clear policies on the scope of AI assets to be inventoried, the roles responsible for maintaining the inventory, and the processes for updating it. AI models are typically included within regular model inventories but specific tags or fields added to identify AI and capture AI-relevant attributes. One bank built an AI use case inventory that aggregated information from the AI model inventory and other inventories or repositories relating to assets and controls in areas such as data, technology and operational management. This provided the bank with a comprehensive and clear view of the linkages between AI models and other relevant assets and controls.

5.2.4 Across banks, AI inventories generally capture key attributes such as the AI's purpose and description, scope of use, jurisdiction, model type, model output[26], upstream and downstream dependencies, model status, risk materiality rating, approvals obtained for validation and deployment, responsible AI requirements, waiver or dispensation details[27], use of personally identifiable information (PII)[28], personnel responsible such as owners, sponsors, users, developers, and validators. For third-party AI, additional attributes such as the AI provider, model

[26] Model output refers to the type of output generated by the AI model. For example, the model output attribute could be the likelihood of customer attrition, or the credit score of a customer.

[27] Waiver or dispensation details refer to information about exceptions/special permissions granted, regarding the development or deployment of AI, that deviate from the bank's standard policies and procedures.

[28] For example, full name, national identification number, personal mobile number.

version, endpoints utilised, as well as other details from the AI developers[29] may also be included.

5.3 Risk Materiality Assessment

5.3.1. Risk materiality assessments are critical for banks to calibrate their approach to risk management of AI across the diverse areas in which AI can be used (e.g., to map the risk materiality of AI to the depth and scope of validation and monitoring required). In assessing risk materiality, most banks considered both quantitative and qualitative risk dimensions that could generally be grouped into three broad categories:

a. **Impact** on the bank, its customers or other stakeholders, including but not limited to financial, operational, regulatory and reputational impact. A few banks developed granular, function-specific definitions of impact to provide greater clarity.

b. **Complexity** due to the nature of the AI model or system, or the novelty of the area or use case in which AI is being applied.

c. **Reliance** on AI, which takes into account the autonomy granted to the AI, or the involvement of humans in the loop as risk mitigants.

5.3.2 Most banks have also established processes to review that risk materialities assigned to AI remain appropriate over time. Similarly, quantitative and qualitative measures and methods used to assign risk materialities were also reviewed, e.g., measures used to quantify financial impact would be updated if the nature of the business in which AI was used had evolved.

[29] These may be provided in AI or AI model cards, which are documents or information usually released alongside open-source AI models that facilitate transparency and accountability by providing essential information on key areas such as the AI model's purpose, performance, limitations, ethical considerations. More information on details that may be included in such cards are available in papers such as https://link.springer.com/chapter/10.1007/978-3-031-68024-3_3.

6 Development and Deployment

Overview

Most banks have established standards and processes for development, validation, and deployment of AI to address key risks.

- For development of AI, key areas that banks paid greater attention to include data management, model selection, robustness and stability, explainability and fairness, as well as reproducibility and auditability.
- For validation, banks required independent validations or reviews of AI of higher risk materiality prior to deployment, to ensure that development and deployment standards have been adhered to. For AI of lower risk materiality, most banks conducted peer reviews that are calibrated to the risks posed by the use of AI prior to deployment.
- To ensure that AI would behave as intended when deployed and that any data and model drifts are detected and addressed, banks performed pre-deployment checks, closely monitored deployed AI based on appropriate metrics, and applied appropriate change management standards and processes.

6.1 Standards and Processes

6.1.1. To support robust risk management of AI across its lifecycle, banks have established standards and processes in the key areas of development, validation, deployment, monitoring and change management. Most banks built upon existing MRM standards and processes for development, validation, deployment, monitoring and change management, but updated these standards and processes to address risks posed by AI.

6.1.2. Key standards and processes relating to conventional model development, validation, deployment, monitoring and change management that banks generally regard as relevant to AI are listed below[30]. Observations on key areas of focus for AI, and how banks have adapted or updated these standards and processes in these areas to address AI risks will be outlined in the subsequent sections.

a. **Data management** - Determining suitability of data, such as the representativeness of data for the intended objective, assessment of completeness, reliability, quality, and relevance of data, and approaches for determining training and testing datasets.

b. **Model selection** - Defining the intended objective of the model and justifying how the selection and design of the model is relevant and appropriate for achieving the desired objective, including the selection of architectures[31] and techniques[32] that are appropriate for the use case and objective.

c. **Performance evaluation** - Setting appropriate evaluation approaches and thresholds, and assessing the model's ability to perform under a range of conditions in accordance with its intended usage and objective.

d. **Documentation** - Providing sufficient detail to facilitate reproducibility by an independent party, including details on data sources, lineage, and processing steps; model architecture and techniques; evaluation and testing approaches and results.

[30] As highlighted previously, even prior to the use of AI models, banks already utilised conventional models, such as economic, financial, or statistical models, and would have instituted model risk management standards and processes for such models. While these standards and processes may have preceded the use of AI models in the bank, their general principles and considerations may also be applicable to AI models.

[31] Model architecture, in the context of AI, relates to the underlying structure and design of the model. It could involve choosing between decision tree-based models such as XGBoost, which were previously described in Section 2, or neural network-based models such as recurrent neural network or transformer models, based on various considerations. For example, decision tree-based models may be more suitable for structured data, such as tabular data, while recurrent neural network or transformer models may be more suitable for text or time-series data as they are designed for sequential data.

[32] Techniques may include methods that are used to train a model from the data. In the context of AI, these may include supervised learning techniques that use labelled data during training to learn how to generate predictions, or unsupervised learning techniques which learn general patterns from unlabelled data. For more details on supervised and unsupervised learning, please refer to Annex A.

e. **Validation** - Setting out the depth of review expected of validators across the areas above; frameworks for determining the prioritisation and frequency of validation (including any revalidation conducted on deployed models).

f. **Mitigating model limitations** - Frameworks and processes for testing key assumptions, identifying limitations and their expected impact, and establishing appropriate mitigants which are commensurate with the impact of the limitations.

g. **Monitoring and change management** - Setting appropriate tests and thresholds to evaluate the ongoing performance of a deployed model, including the frequency of monitoring; as well as the processes to be followed (e.g., additional validations and approvals) for changes made to a deployed model.

6.1.3. When implementing standards and processes for risk management of AI, most banks established baseline standards and processes that applied to all AI across the bank, regardless of risk materiality. For AI that were of greater risk materiality, or where there were requirements specific to the use case, baseline standards and processes would be supplemented by enhanced standards and processes. For example, additional evaluation or enhanced validation standards and processes could apply to AI used for risk and regulatory use cases where there may be heightened requirements on performance evaluation or thresholds. The alignment of baseline standards and processes across the bank helped ensure that key model risks were addressed consistently for AI with similar characteristics and risks regardless of where they were used in the bank.

6.2 Data Management

6.2.1 Robust data management is essential to support the development and deployment of AI. General bank-wide data governance and management standards and processes[33] would apply to data used for AI. For example, whether data was used for reporting purposes or for AI systems, the same data governance committees generally oversee approvals and management of data issues. Similarly, standards and processes for key data management controls such as basic data quality checks would also apply. However, to address AI-specific requirements, all banks had established additional data management standards and processes to ensure that data used for AI development and deployment are fit for purpose. An overview of key data management areas for AI development and deployment that most banks generally focused on are listed below. Standards or processes relating to data management that are specific to AI development, validation, deployment, monitoring or change management are covered in the subsequent sections.

a. **Appropriateness of data for AI use cases** - Ensuring data used for development and deployment of AI are suitable for the context in which the AI is used, including assessing the use of such data against fairness and ethical considerations.

b. **Representativeness of data for development** - Ensuring data selected for training and testing AI models are representative of the real-world conditions, including stressed conditions, under which the AI would be used.

c. **Robust data engineering during development** - Ensuring data processing steps, [34] which may include additional data quality checks [35] , feature

[33] Please see MAS' information paper on Data Governance and Management Practices for more details on general data governance and management standards and processes. The paper covered governance and oversight, data management function, data quality and data issues management, which would also apply to data used for AI. Other relevant regulations and publications include the Personal Data Protection Act (PDPA), which comprises various requirements on data privacy governing the collection, use, disclosure and care of personal data, and provides a baseline standard of protection for personal data in Singapore; and Advisory Guidelines on Use of Personal Data in AI Recommendation and Decision Systems issued by the Personal Data Protection Commission (PDPC) in March 2024. Please refer to Annex B for the relevant links.

[34] Examples of data processing steps include missing value imputation, replacement of outlier values and standardisation or normalisation of data values.

[35] To ensure data quality, key areas such as data relevance, accuracy, completeness and recency may be assessed.

engineering[36], augmentation and labelling[37] of datasets, are robust and free of bias, and that the integrity and lineage of data are checked and tracked across these data engineering steps.

d. **Robust data pipelines for deployment** - Establishing robust controls around data pipelines for deployment, including continuous monitoring of the quality of data passed to deployed AI, as well as checks for anomalies, drifts, and potential bias that may have an impact on performance or fairness.

e. **Documentation of data-related aspects for reproducibility and auditability** - Ensuring key data management steps, such as data sourcing, data selection, data lineage, data processing, approvals and remediation actions taken for data issues are documented to enable reproducibility and auditability.

6.2.2 Some banks have also established additional data management standards and processes in the areas below:

a. To ensure that data is being used appropriately when developing or deploying AI, a few banks have required approvals to be obtained for high-risk data use cases, such as data use where a third party may have access to the bank's internal data, use of employee data for monitoring, or the collection of biometric data to identify individuals.

[36] 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.

[37] When training AI models for a specific task, such as predicting a credit default or recommending a suitable financial product to a customer, we need data that includes the input variables (e.g., data relating to a past loan, or customer history), as well as a target variable (e.g., whether there was a credit default for the loan, or a recommendation that the customer accepted). Data labelling refers to the process of assigning such target variables, typically based on past historical data or via human annotation.

b. To support data reusability and reduce the time needed for feature engineering across the bank, as well as enhance consistency and accuracy in model development, a few banks have also built feature marts[38].

c. To account for the greater use of unstructured data[39], there were also ongoing efforts to more effectively manage such unstructured data, such as improving metadata management and tagging for unstructured data to enable better data governance[40]. Most of the data management areas outlined in paragraph 6.2.1 are also generally applicable to unstructured data, where relevant.

6.3 Development

Model Selection

6.3.1 Given the trade-offs of adopting more complex AI models (e.g., higher uncertainties, limited explainability), most banks required developers to justify their selection of a more complex AI model over a conventional model or a simpler AI model[41], (e.g., balancing the need for performance against explainability for a specific use-case). Some banks required developers to go beyond qualitative justifications, and develop challenger models (which could be either conventional or simpler AI models) to explicitly demonstrate the performance uplift of the AI model over the challenger model as part of this justification.

[38] A feature mart is a centralised repository or database that stores curated, pre-processed and reusable features (variables or attributes) that can be used for training models. Aside from supporting data reusability, feature marts may also help improve data governance by maintaining metadata on each feature, including details on its sources, transformations, lineage and quality. Feature marts may also allow for version control, ensuring that any updates to features are tracked.

[39] Unstructured data refers to information that does not follow a predefined format or organised structure, making it more difficult to store and analyse using traditional databases or methods for structured data. Unstructured data typically includes data types such as text, images, videos, and audio. While the use of unstructured data is not new to banks, e.g., using surveillance videos from cameras at ATMs, the use of such data is growing due to Generative AI.

[40] These may also include updating and adapting other areas such as data discovery and classification, access rights, data lifecycle management, data sanitisation and validation, and security controls for unstructured data.

[41] For example, a developer who wishes to use a more complex neural network-based deep learning model may be required to justify the need for such an AI model over a simpler tree-based machine learning model or a logistic regression model, and consider the trade-offs based on the use case requirements.

Robustness and Stability

6.3.2 In assessing the overall suitability of AI models, banks placed heavy focus on ensuring that AI models were both robust and stable[42], and accordingly paid significant attention to i) the selection and processing of datasets used for training and testing AI models; ii) determining appropriate approaches, measures and thresholds for evaluating AI models; and iii) mitigating overfitting risks[43] that often arise due to the complexity of AI models. We outline some of the practices in these key areas below.

Selection and Processing of Datasets for Training and Testing

6.3.3 Datasets chosen for training and testing or evaluation[44] of AI models were expected to be representative of the full range of input values and environments under which the AI model was intended to be used. Training and testing datasets were also checked to ensure that their distributions or characteristics are similar[45].

6.3.4 Most banks also invested efforts in collecting testing datasets that allowed predictions or outputs from AI models to be tested or evaluated in the bank's context as far as possible. For example, curating datasets that allowed for AI model generated answers to queries from customers to be compared against answers from in-house human experts, or getting actual feedback from the bank's customers on the quality of these AI model generated answers.

Evaluation Approaches, Measures and Thresholds

[42] The concepts of robustness and stability in AI systems often overlap and what these terms cover can vary. For the purpose of this information paper, robustness refers to AI's ability to achieve its desired level of performance under real-world conditions, while stability refers to the consistent performance of AI across a representative range of real-world scenarios. These concepts are also related to the reliability of the AI system or model.

[43] Overfitting is when an AI model learns the training data overly well, to the point where it performs extremely well on training data but very poorly on new data that it has not seen in the training dataset. Intuitively, this may mean that the model has memorised the training examples rather than learning general patterns, resulting in poor performance in real-world conditions.

[44] The terms "testing" and "evaluation" of AI models are commonly used interchangeably to refer to the assessment of the performance of AI models on datasets that it had not been trained on.

[45] This issue is also commonly referred to as training-testing skew, which are discrepancies between the distribution of data used to train an AI model and the distribution of data it encounters during testing.

6.3.5 Given that AI is developed to meet specific business needs or objectives, banks' standards and processes on the robustness and stability of AI models generally required testing or evaluation approaches to be aligned with the intended outcomes that the AI models were meant to support. The exact approaches selected could differ depending on the nature of the AI models, as well as the needs of the use case. For example, assessing a fraud detection model's ability to flag out known fraud cases by comparing against ground truth in historical data, or the usefulness of a financial product recommendation model through human feedback.

6.3.6 Correspondingly, while there are many established performance measures for AI models[46], banks paid significant attention to aligning the choice of performance measures with the intended outcomes that the AI models were meant to support. In some cases, this could involve trade-offs between different performance measures. For example, if the intended outcome was to detect as many instances of fraud as possible, performance measurement would need to focus more on the proportion of false negatives (i.e. fraudulent instances that were not detected), even though this may come at the expense of a higher proportion of false positives (i.e. instances falsely flagged by the model as being fraudulent).

6.3.7 Other tests that banks may include to ensure robustness and stability[47] include the following:

a. **Sensitivity analysis** to understand how predictions or outputs of AI models change under different permutations of data inputs. This also helps to identify important features that significantly influence predictions or outputs, and facilitate explanations of the behaviour of AI models.

[46] There are a wide range of performance measures for AI models, and these are often specific to the task or use-case; for example, recall, precision, or F1 for classification tasks, mean absolute error or root mean squared error for regression tasks, mean average precision or mean reciprocal rank for recommendation tasks.

[47] The objectives of some of these tests overlap, and may also relate to data management aspects that we outlined earlier. Nonetheless, we list all the tests that we observed across the banks for completeness.

b. **Stability analysis** to compare the stability of data distributions and predictions or outputs, e.g., assessing whether the distribution of a training dataset from an earlier period matches the distribution of testing datasets from more recent periods, and how differences affect the performance of AI models.

c. **Sub-population analysis**, which are evaluations of how AI models perform across different sub-populations or subsets within the datasets (e.g., to identify any significant differences in performance between different customer segments). Such analysis of sub-populations or subsets within the datasets help to identify potential issues that might not be obvious in the aggregated testing dataset, as well as potential sources of bias, which could support fairness assessments of AI models where necessary (e.g., where sub-populations relate to protected features or attributes such as race or gender).

d. **Error analysis** to identify potential patterns in prediction errors (e.g., misclassified instances), which helps to understand the limitations of AI models.

e. **Stress testing** the response of AI models to edge cases or inputs outside the typical range of values used in training. This allowed banks to better determine performance boundaries and identify limitations of AI models. Some banks also tested the behaviour of AI models in the context of unexpected inputs or conditions. Examples included adversarial testing or red teaming types of exercises. Such testing is especially important in the context of AI models used in high risk or customer-facing applications, as it allowed the bank to establish conditions under which AI models would not perform as expected or could introduce potential security or ethical concerns.

6.3.8 Most banks would establish criteria or thresholds for performance measures, to define what was considered acceptable performance. Such thresholds need to be clearly defined and documented, as well as mutually agreed upon by developers and validators. Such thresholds were usually use case specific, and could also be used subsequently to facilitate validation, pre-deployment checks, as well as monitoring and change management.

Mitigating Overfitting Risks

6.3.9 The large number of parameters and inherent complexity of AI models increases the risks of them overfitting on training data (in-sample data) and hence performing poorly when deployed on out-of-sample data. Banks employed a variety of mitigants to address this risk:

a. **Model selection** – Generally favouring AI models of lower complexity unless there are clear justifications to do otherwise; or adopting approaches that constrained the complexity of AI models[48].

b. **Feature selection** - Applying explainability methods to identify the key input features or attributes that are important for the AI model predictions or outputs[49] and assessing that they are intuitive from a business and/or user perspective[50].

c. **Model evaluation** - Additional performance testing requirements to test the performance of AI models on unseen data where possible, such as cross-validation techniques[51] and testing against more out-of-sample/out-of-time[52] datasets.

Explainability

6.3.10 All banks identified explainability as a key area of focus for AI, particularly for use cases where end-users or customers need to understand key features or attributes in the data influencing predictions of AI models. For example, explainability would be more important in higher risk materiality use cases where

[48] Examples include regularisation techniques or limiting the number and depth of trees for gradient boosting trees. Such techniques generally try to limit the number of parameters used so that the trained model is less complex. For example, some regularisation techniques force less important parameters to values of zero.

[49] As discussed in the next section on explainability methods.

[50] Additional justification would typically be required to retain features or attributes that were not intuitive, or which did not meaningfully contribute to the overall performance of the models. Such data may introduce more noise, and cause the AI model to overfit on the noise, leading to poor performance in real world conditions.

[51] Cross-validation generally refers to techniques to evaluate models by resampling the dataset for training and testing. An example would be K-fold cross-validation (which involves splitting the dataset into K parts for K training and testing rounds).

[52] An out-of-sample testing dataset is a subset of data not used in model training, whereas an out-of-time testing dataset is a subset of data obtained from a time period distinct from the time period of the subset of data used in training the model.

bank staff making decisions based on predictions of AI models need to understand the key features or attributes[53] contributing to the prediction; or in use cases where a customer may ask for reasons for being denied a financial service. Hence, development standards for AI across all banks had been expanded to include a section on explainability.

6.3.11 Explainability requirements in banks' standards and processes generally required developers to apply global and/or local[54] explainability methods to identify the key features or attributes used as inputs to AI models and their relative importance; assess whether these features or attributes were intuitive from a business and/or user perspective; and provide additional justification for retaining features or attributes which were not intuitive. Such methods could also help identify the usage of potentially sensitive features as part of fairness assessments. Some banks had set out a list of global and local explainability methods that could be applied to explain the outputs[55] of AI models. Such methods could be directly applied during development as part of the feature selection process, or used within explainability tools developed as part of the AI system so that either global and/or local explanations can be provided alongside predictions or outputs generated by AI models post-deployment.

6.3.12 In terms of the level of explainability required for different use cases, some banks established standards and processes to clearly define the minimum level of global and/or local explainability required for different use cases. For these banks,

[53] An example of a feature or attribute in this context could be the income of the customer.

[54] Global explainability is the ability to understand the overall functioning of the model by identifying how input features drive model outputs at an overall model level. Local explainability is the ability to identify how input features drive the model output for a specific observation or instance. Taking a fraud detection model as an example, global explainability methods allow for identification of the most important features, such as the high values of transactions, used to detect fraudulent transactions for the model in general. However, the key features that are important for a specific transaction (i.e. the local instance) may not necessarily be the same, e.g., the value of the transaction may be small for a specific instance but the transaction is still detected as a fraudulent transaction due to specific characteristics of the parties involved in the transaction, such as an unfamiliar geographic location of one of the parties. Local explainability methods help to identify such features for the local instance.

[55] Common examples of explainability methods include SHAP (for global and local explainability) and LIME (for local explainability). SHAP generates Shapley values for each feature based on its contribution to a given model output. A global-level explanation can be generated by generating a summary plot of the Shapley values of the key features, across the entire set of model outputs. LIME is based on training a separate model for the local instance that needs to be explained. The explanation that is generated is based on the separately trained model.

factors considered when applying a higher standard of global and/or local explainability included risk materiality or the extent to which AI-driven decisions were likely to require explanations (e.g., to the bank's customers) for the use case. For example, AI models used for credit decisioning could require the most exacting standards for global and local explainability, requiring developers to carefully consider all features used as inputs and provide justifications for their use, as well as the ability for users to easily identify key features influencing any given prediction post-deployment. Other banks required global and/or local explainability to be explored across all AI, but allowed users and owners to decide on the acceptable level of explainability, and justify their decision based on the use case.

Fairness

6.3.13 The outputs of AI models are inherently influenced by the patterns learnt from its training data. If the training data contained biases that unfairly represent or disadvantage specific groups of individuals, AI models may perpetuate these unfair biases in its predictions or outputs. This could lead to decisions or recommendations that disproportionately and unfairly impact certain demographic groups.

6.3.14 The earlier section on data management had outlined the need for fairness to be considered during development, and for checks and monitoring of potential biases during deployment. More specifically, during AI development, for use cases that could have a significant impact on individuals, most banks would undertake a formal assessment on whether specific groups of individuals could be systematically disadvantaged by AI-driven decisions. The scope of such assessments could vary between banks depending on the relevant rules, regulations or expectations applicable to the bank[56], and between use cases depending on the risk materiality of the AI.

[56] Examples of such expectations on fairness for AI used by banks across jurisdictions include the Principles to Promote Fairness, Ethics, Accountability and Transparency (FEAT) in the use of Artificial Intelligence and Data Analytics in Singapore's Financial Sector, published by MAS in 2018; General Principles for the use of Artificial Intelligence in the Financial Sector, published by

6.3.15 Generally, the approach for assessing fairness used by banks involved the following steps:

a. Defining a list of protected features or attributes, for which use of such features or attributes in AI models would require additional analysis and justification. Common examples of such protected features or attributes include gender, race or age.

b. Determining whether such features or attributes[57] were used in training AI models. Based on this assessment, to define groups of individuals at risk of being systematically disadvantaged by the AI-driven decisions (at-risk groups).

c. Where necessary, determining the extent to which AI-driven decisions systematically disadvantaged against at-risk groups. The was usually assessed via fairness measures (e.g., fairness measures that are available in the toolkit released by the Veritas Initiative).

d. Where necessary, providing adequate justifications on the use of protected features or attributes in AI models (e.g., trade-offs against the intended objectives of the AI model[58]).

Reproducibility and Auditability

6.3.16 Reproducibility and auditability[59] of AI development are essential for ensuring accountability and building trust in AI systems. To facilitate reproducibility and auditability of AI, most banks expanded existing documentation requirements to incorporate the relevant AI development processes and considerations. A list of

De Nederlandsche Bank in 2019; and the High-level Principles on Artificial Intelligence, published by the Hong Kong Monetary Authority in 2019.

[57] These could include proxy attributes that are heavily correlated with such protected attributes.

[58] This could be supported by, for example, analysis on the difference in performance between an AI model which included these protected features or attributes, and an AI model which did not. An informed assessment could then be made on whether this difference in performance was necessary to achieving the model's intended objective, taking into consideration the level of potential harm done to at-risk groups arising from the use of the AI model.

[59] Reproducibility refers to “the ability of an independent verification team to produce the same results using the same AI method based on the documentation made by the organisation”, while audibility refers to “the readiness of an AI system to undergo an assessment of its algorithms, data and design processes” (Model AI Governance Framework, IMDA Singapore.)

key documentation requirements for AI commonly seen across banks are as follows:

a. **Data** - Documentation of key data management steps is important to facilitate reproducibility and auditability. During development, key information that would usually be documented include datasets and data sources used for model development and evaluation, details of how these datasets were assessed as fit-for-purpose, processed ahead of model training, and split into relevant training, testing and/or validation[60] datasets.

b. **Model training** - Details of how the AI model was trained or fit to the training dataset. Such details could include codes (along with software packages/environment used and their relevant versions), key settings (e.g., hyperparameters[61] used and the approach for selecting hyperparameters[62]), random seed values[63] and any other configurations required for a third party to reproduce the training process.

c. **Model selection** - Details of how the performance of the AI model was evaluated and how the final model was selected. Such details could include the evaluation approaches, thresholds and datasets applied [64] and the corresponding results, comparisons of performance across multiple AI models and justifications for selecting the final model.

d. **Explainability** - Global and/or local explainability methods used, feature selection process, analysis of results, as well as description of key features selected and additional justifications for inclusion of certain key features (e.g., features that may not have appeared to be important to a human expert).

[60] Testing and validation datasets refer to datasets used to evaluate the performance of the model outside of the dataset used to train the model. This should be distinguished from independent validation, which is the process of independently assessing the overall suitability of the model.

[61] E.g., number of trees and maximum tree depth for gradient boosted trees.

[62] E.g., grid search, random search of hyperparameters.

[63] AI models usually need to be initialised with a random set of numbers (e.g., for the model parameters) before training, and documenting the random seed value that is used to initialise the AI models is necessary to reproduce the AI model's behaviour and results.

[64] As detailed in the earlier sub-section on Robustness & Stability.

e. **Fairness** - Metrics and associated thresholds, results of fairness assessments and justifications for the use of any protected features or attributes.

6.3.17 Alongside documentation requirements in the relevant standards and processes, most banks also set up documentation templates that developers were required to follow for consistency. Such templates were typically designed by the bank's MRM function. Templates could differ between business domains (as different performance tests or metrics could apply) or between AI of different risk materialities (as documentation requirements could be higher for AI of higher risk materiality).

6.4 Validation

6.4.1 Independent validation provides an objective and unbiased assessment of the suitability, performance and limitations of AI. It acts as an important challenge to developers, and ensures that the relevant standards and processes have been adhered to when developing AI.

6.4.2 The validation process typically involves an independent unit[65] reviewing the AI development process and documentation, assessing that AI performs and behaves as intended, and undertaking pre-deployment checks. Actions to address issues identified during validation, such as the application of suitable adjustments or other mitigating or compensatory controls, would typically be proposed by developers and agreed to by validators before deploying AI.

6.4.3 Building on their conventional MRM processes, banks have equipped independent validation functions with the skills and incentives needed to conduct independent review of AI used in the bank, which include investments in efforts to ensure that independent validation staff have the relevant technical expertise for AI.

[65] For example, the Federal Reserve/Office of the Comptroller of the Currency's SR Letter 11-7 on Supervisory Guidance on Model Risk Management states that validation should generally be done by individuals not responsible for development or use and do not have a stake in whether a model is determined to be valid.

6.4.4 Banks adopted a range of approaches in establishing independent validation requirements across different AI. One bank required all AI to be subject to independent validation, with the depth and rigour of validation varying based on the AI's risk materiality rating. Most other banks required independent validation only for AI of higher risk materiality, with other AI subject only to peer review[66]. Even for AI of lower risk materiality, the involvement of either an independent validator or peer reviewer allowed for some degree of challenge that helped to better manage the added uncertainties and risks posed by AI, and check that such AI was developed in accordance with the bank's standards and processes.

6.5 Deployment, Monitoring and Change Management

Pre-Deployment Checks

6.5.1. Aside from checks during the validation process, pre-deployment checks and tests are important to ensure that the AI has been correctly implemented and produces the intended results before being deployed for use. Banks placed significant focus on implementing controls for the deployment of AI to ensure that the AI functions as intended in the production environment[67]. These controls were usually based on existing technology risk management guidelines. For example, banks would apply standard software development lifecycle (SDLC) processes to ensure that the AI application or system was secure, free from error and performed as intended before deployment[68]. Some banks also conducted additional checks to ensure that the deployed AI's scope, output and performance, and associated controls align with that of the validated AI:

a. **Additional tests**, such as:

[66] As compared to independent validation, peer reviews were usually conducted by a non-independent function (e.g., a different development team in the same unit/reporting line as the original model developers).

[67] A production environment is a live operational setting where deployed systems, such as deployed AI models, are run under real world conditions to deliver services or perform tasks for end-users.

[68] Please see MAS' Technology Risk Management Guidelines for further details on the adoption of sound and robust practices for the management of technology risk in these areas: https://www.mas.gov.sg/regulation/guidelines/technology-risk-management-guidelines

i. **forward testing**, which are experimental runs using a limited set of production data or with a limited set of users, for selected high materiality use cases to assess the behaviour of AI in an environment similar to when the AI is fully deployed; and

ii. **live edge case testing** to assess how AI handles edge cases in the production environment, which helps to verify that AI can handle a variety of improbable but plausible scenarios when deployed.

b. **Automated pipelines**, such as setting up automated deployment and continuous integration/continuous deployment (CI/CD) pipelines [69] to minimise human error and maintaining a consistent process for how AI is deployed, monitored, and maintained, which is important for AI given the need for regular data and model updates.

c. **Process management,** which includes checks to ensure that key processes important for the deployed AI, such as human oversight, backup models, and other appropriate controls and contingencies, are in place; and business process change management, such as training users to understand AI limitations and to use AI appropriately.

6.5.2. Non-AI specific pre-deployment checks [70] remain relevant, hence key control functions, such as those in the areas of technology, data, legal and compliance, third-party and outsourcing, would also confirm that the checks have been undertaken and sign off before AI is deployed into production.

[69] Continuous integration/continuous deployment (CI/CD) pipelines automate the process of building, testing, and deploying code changes, and reduce the potential of errors arising from manual interventions. Approvals and checks are also usually integrated into the CI/CD process to ensure that new code pushed into production are checked for errors. More details on CI/CD, as well as other related terms such as MLOps and AIOps are provided in Annex A.

[70] For example, checks relating to cyber-security, or compliance with outsourcing policies.

Monitoring Metrics and Thresholds

6.5.3. Monitoring is particularly critical for AI given their dynamic nature and the potential for AI model staleness due to drifts[71] in either data or the model behaviour over time. All banks paid significant focus to the ongoing monitoring of their AI to ensure that they continue to operate as intended post-deployment. Key measures that were monitored generally follow those that were covered during development and validation, and include robustness, stability, data quality, and fairness measures.

6.5.4. Measures used for monitoring were tracked against predefined thresholds, usually determined at the development and validation stages, to ensure models perform within acceptable boundaries. Some banks have also implemented tiered thresholds, for example, additional early warning thresholds to pre-empt model deterioration, and different thresholds to determine when retraining or a full redevelopment of the AI may be necessary.

6.5.5. Most banks also have a process or system for reporting, tracking and resolving issues or incidents if breaches or anomalies arise from the monitoring process. Banks generally track issues or incidents from discovery to resolution, and incorporate a relevant escalation process based on the materiality of the issue or incident. The resolution process may include AI model retraining, redevelopment, or decommissioning as possible outcomes. Where a major redevelopment was undertaken, revalidation and approval would be needed before the updated model could be redeployed.

[71] AI models can perform poorly when they become stale due to factors such as data drift, concept drift or model drift, which are essentially due to changes in the data distributions, relationships between input data and predictions/outputs, or the general environment in which the AI model is being used. More details on data, concept and model drifts are provided in Annex A.

Contingency Plans

6.5.6. All banks would generally have standards and processes relating to contingency plans for AI, particularly those supporting high-risk or critical functions[72]. These plans, which may not be specific to AI, typically outline fallback options, such as alternative systems or manual processes, and would be subject to regular reviews and testing to ensure readiness for rapid activation when necessary. For mission-critical AI applications[73], a few banks may also have kill switches in place. Kill switches are used to deactivate AI if they exceed risk tolerances, and require clear contingency plans to be quickly rolled out.

Review and Revalidations

6.5.7. Aside from ongoing monitoring, banks also conducted periodic reviews of their portfolio of AI. Key aspects that that were usually reviewed include changes in the models' materiality, risks, scope and usage, performance, assumptions and limitations, and identification and remediation of issues.

6.5.8. Banks also have standards and processes for ongoing revalidations of AI in production, with the intensity and frequency based on the materiality of the AI. In general, AI deemed critical to risk management, regulatory compliance, business operations, or customer outcomes are revalidated more frequently and intensely.

Change Management

6.5.9. Standards and processes relating to AI change management are needed to ensure that what constitutes a change is clearly defined, and that the appropriate development and validation requirements are applied. Most banks required

[72] Such contingency plans may not apply specifically to AI, but to technology systems in general. Nonetheless, they may require additional considerations in the case of AI, e.g., AI-specific performance monitoring thresholds to determine when to trigger the contingency plan, or a backup plan that involves another AI system or model.

[73] For example, for AI that are used for trading.

significant or material changes[74] to AI in production to be reviewed and approved by the control functions prior to implementation, so as to ensure that any modifications made to the model do not negatively impact its performance. To manage changes to AI, banks have also established systems and processes for version control of both internal and third-party AI (which do not only cover code relating to AI, but also data and other artifacts such as hyperparameters and the trained model parameters or weights). Version control enables banks to track changes across different aspects of AI and roll-back to previous versions of AI where necessary[75]. Most banks have also set up processes for third-party AI providers to provide notifications of version updates[76].

6.5.10. AI for certain use cases, such as fraud detection, may need to be changed or updated more frequently[77], due to drifts in the data or the behaviour of the AI model over time. To deal with such frequent changes, some banks have established systems and processes for the automatic updating of such AI. Such AI, which some banks refer to as "dynamic AI", need to be subject to enhanced requirements and controls to ensure that change management is well governed. Key additional requirements and controls include justifications for enabling automatic updating of AI, clearly defining what can be updated automatically, for example, restricting changes to the retraining of AI model with more recent datasets, but not allowing for changes to AI model architectures or hyperparameters. Such dynamic AI would also be subject to enhanced risk

[74] Examples of significant or material changes include fundamental changes to AI model architectures or training techniques. Such changes may necessitate an in-depth revalidation, compared to less significant changes, such as retraining the AI model with more recent data, which may only require checks on AI performance to ensure the AI is still behaving as expected.

[75] While we cover version control here under change management where the AI is already deployed, it is important to note that version control for AI also plays a key role during the development and validation stages. For example, version controls are needed to support iterative improvements and collaboration during development, and also help to ensure reproducibility and auditability during validation.

[76] While banks generally try to require third-party providers to notify them of any changes to the AI model or service, there may be circumstances where such notifications may not happen, e.g., the third-party provider may not notify end-users on changes that they view as immaterial. We have observed banks trying to address this by setting out clearer terms in their legal agreements, for example, adding a clause that requires the third-party provider to notify banks on any upcoming changes to the AI model or system.

[77] For example, if we compare a fraud detection use case with an NLP use case such as summarisation of customer call transcripts, data relating to the behaviour of scammers would usually change much more frequently than data relating to customer calls due to the active efforts of scammers to evade detection.

management requirements, such as enhanced data management standards, e.g., additional checks on data quality and drifts, as well as enhanced performance monitoring requirements, e.g., more stringent monitoring notification thresholds.

7 Other Key Areas

7.1 Generative AI

Overview

While the use of Generative AI in banks is still in the early stages, banks generally try to apply existing governance and risk management structures and processes where relevant and practicable, and balance innovation and risk management by adopting:

- Strategies and approaches, where they leverage on the general-purpose nature of Generative AI by focusing on the development of key enabling modules or services; limit the current scope of Generative AI to use cases for assisting/augmenting humans or improving internal operational efficiencies that are not direct customer facing; and building capacity and capabilities by establishing pilot and experimentation frameworks;
- Process controls, such as setting up cross-functional risk control checks at key stages of the Generative AI lifecycle; establishing more detailed development and validation guidelines for different Generative AI task archetypes; requiring human oversight for Generative AI decisions; and paying close attention to user education and training on the limitations of Generative AI tools; and
- Technical controls, such as selection, testing and evaluation of Generative AI models in the context of the bank’s use cases; developing reusable modules to facilitate testing and evaluation; assessing different aspects of Generative AI model performance and risks; establishing input and output filters as guardrails to address toxicity, bias and privacy issues; and mitigating data security risks via measures such as the use of private clouds or on-premise servers, data loss prevention tools, and limiting the access of Generative AI to more sensitive information.

7.1.1. In addition to the key areas highlighted in the prior sections, there are some aspects relating to Generative AI (compared to conventional AI) that require further consideration:

a. **Higher uncertainties associated with Generative AI** – The risks of hallucinations and unexpected behaviours by Generative AI given its greater complexity may lead to less robust and stable performance, and was a key concern highlighted by banks. This concern was particularly pronounced for use cases of higher risk materiality or those that are directly customer-facing, where greater reliability was required.

b. **Difficulties in evaluating/testing Generative AI and mitigating its limitations** – Compared to conventional AI, which were typically used by banks for specific use cases that the AI models had been trained for, Generative AI are more general-purpose in nature and can be used in a wider range of use cases in the bank. However, there may be a lack of easily available ground truths[78] in some of these newer use cases to evaluate and test Generative AI. Use cases involving Generative AI also typically involve unstructured data, such as text data, for which there are significantly more possible permutations, compared to structured data usually used for conventional AI. This makes it challenging to foresee all potential scenarios and perform comprehensive testing and evaluations[79].

c. **Lack of transparency from Generative AI providers** - Unlike conventional AI models, which are often developed and trained internally by the bank's developers, Generative AI used by banks were pre-dominantly based on pre-trained models from external providers. As disclosure standards relating to such AI are still evolving globally, banks may lack full access to essential risk

[78] Ground truth refers to reliable or factual information that serves as a standard against which the outputs or predictions of AI models, including Generative AI models, can be evaluated.

[79] For example, it is significantly harder to evaluate the quality of a summary or of an image generated by Generative AI, compared to evaluating the accuracy of a simple yes/no prediction from conventional AI. It is also harder to foresee all possible permutations of text or images that may be used as inputs to Generative AI, as well as all possible permutations of text or images that may be generated by Generative AI.

management information, such as details about the underlying data used in model training and testing, as well as the extent of evaluation or testing applied to these models.

d. **Challenges in explainability and fairness with Generative AI** – The lack of transparency from external providers may also contribute to challenges in understanding and explaining the outputs and behaviour of Generative AI, and ensuring that the outputs generated by Generative AI are fair. There is also a general lack of established methods currently for explaining Generative AI outputs and assessing their fairness.

7.1.2. Most banks are in the process of reviewing and updating parts of their AI model risk management framework for Generative AI to balance the benefits and risks of its use.

7.1.3. The subsequent paragraphs outline observations from the thematic on key approaches and controls that banks have adopted to balance innovation and risks based on the current state of use of Generative AI. It should be noted that these approaches and controls will need to be updated as Generative AI technology evolves, and that risk management efforts will need to be scaled accordingly based on the state of Generative AI use across the institution.

Strategies and Approaches

7.1.4. Some banks have invested significant effort in identifying and building key enabling services and modules for Generative AI that can be utilised across multiple use cases, e.g., vector databases[80], retrieval systems[81], evaluation and

[80] Data, particularly unstructured data, such as text and images, need to be encoded into numerical representations before they can be used for AI or Generative AI. Such numerical representations are commonly referred to as vectors. Vector databases are specialised database systems designed to store, index, and efficiently query such data.

[81] Retrieval systems help to search information repositories and retrieve the most relevant information for a specific task. For example, to help answer a query relating to information in a corporate information repository, the retrieval system will help to search for the most relevant pieces of information in the corporate information repository. The retrieved information is then usually used as context for the Generative AI model to generate an answer from.

testing modules[82]. Such an approach enables scalability, reduces time and costs for implementation, and facilitates the development of more robust and stable Generative AI.

7.1.5. To manage the potential impact of Generative AI risks, such as hallucinations, most banks have started with a more limited scope of use, focusing on the use of Generative AI for assisting or augmenting humans, or improving internal operational efficiencies, rather than deploying Generative AI in direct customer-facing applications without a human-in-the-loop. Banks felt that such an approach would allow them to learn how to utilise Generative AI effectively and understand its limitations, while managing the potential impact of risks posed by Generative AI.

7.1.6. Similarly, to gain greater comfort with the use of Generative AI, most banks have established clear policies and procedures for Generative AI pilots and experimentation frameworks. Aside from helping the bank to build capacity and capabilities while managing risks associated with Generative AI, such pilots and experimentation frameworks are needed to evaluate and test Generative AI in real-world scenarios and understand how Generative AI would behave when deployed. Such pilots are typically bound by time and user limits[83].

Process Controls

7.1.7. To address the cross-cutting nature of Generative AI use cases and risks, as well as the fast-evolving landscape, some banks have instituted cross-functional risk control checks at key stages of the Generative AI lifecycle.

7.1.8. As most Generative AI use cases usually fall within a few task archetypes, e.g., summarisation, information extraction, conversational agents, question answering, one bank established detailed development and validation guidelines

[82] An example of such a module could be a separately trained AI model that estimates the probability of an answer generated by an LLM being a hallucination.

[83] Aside from setting time and user limits, other requirements that may apply to such pilots or experiments include setting clear criteria for success at the end of the pilot, conditions on the terms of use for owners and end-users, and close monitoring of usage patterns and outputs for anomalies and to ensure compliance with the limited scope of usage.

specific to different Generative AI task archetypes to support development and validation processes.

7.1.9. Due to the uncertainties associated with Generative AI, banks continue to require human oversight or have a human-in-the-loop when using Generative AI to aid in decision-making. Extensive user education and training on the limitations of Generative AI tools was another key area of focus.

Technical Controls

7.1.10. As most Generative AI models used by banks, whether closed or open-source, originate from third parties, selection of the appropriate model continues to be an important step for most banks. To assess suitability, some banks would typically start by conducting significant research on the capabilities of these models for their needs, including utilising public benchmarks and the latest research papers to guide decisions. Testing and evaluation of Generative AI models in the context of the bank's use cases was also an important area of focus.

7.1.11. More advanced banks would undertake a range of assessments, from standalone, functional to end-to-end assessments. Standalone assessments involve the evaluation of the Generative AI model itself. This is usually based on publicly available data or resources, such as evaluation results in research articles, model leaderboards, or using open-source evaluation datasets. Functional assessments involve evaluations of Generative AI model performance on tasks and contexts specific to the bank, e.g., evaluating the performance of a Generative AI model when used for retrieval of information from the bank's repository. Finally, end-to-end assessments would evaluate the performance of the entire Generative AI system, which may involve multiple Generative AI or AI models.

7.1.12. Such banks also paid significant attention to establishing methods for assessing different aspects of Generative AI model performance such as accuracy, relevance, and bias[84], as well as creating reusable modules to facilitate testing and evaluation.

7.1.13. The more advanced banks also paid significant attention to curating testing datasets that were specific to the use cases and tasks that Generative AI models were being used for in the bank. Such testing datasets were critical to ensuring that Generative AI models and systems were fit-for-purpose in the bank's context. For example, if Generative AI was used for summarising complaints from the bank's customers, the performance of Generative AI on general summarisation tasks may not be indicative of its performance in the bank's context as it may not have been trained on such complaints that are not in the public domain, and the complaints may also contain information specific to the bank, e.g., the bank's services. To ensure the proper evaluation of Generative AI in the bank's context, the bank will need to curate bank-specific testing datasets from the bank's internal historical data, or use expert human annotators to generate good quality summaries for a set of customer complaints to evaluate against. Such testing datasets are also important for monitoring the ongoing performance of Generative AI models, and for evaluating newer Generative AI models as part of the onboarding process. Other key tests that banks adopted included model vulnerability testing to assess cyber security risks[85], as well as stability and sensitivity testing to ensure consistent performance. Human feedback also played a key role in testing, evaluating and monitoring Generative AI performance.

7.1.14. Most banks have established input and output guardrails that utilise filters to manage risks relating to areas such as toxicity, biasness, or leakage of sensitive information. Such filters may use rules or AI to detect such undesired or inappropriate information. For example, input filters may be used to reject requests with toxic language, or replace PII information in requests with generic

[84] In this context, accuracy refers to whether the generated text aligns with factual information; relevance refers to how pertinent the generated text is to the specific query; and bias refers to scenarios where the generated text may be biased to specific groups of people, e.g., the generated content may favour one gender over another.

[85] These were discussed at length in MAS' information paper on Cyber Risks Associated with Generative Artificial Intelligence and will not be repeated here. See Annex B for link to the paper.

placeholders. Output filters may be used to detect biasness or toxic language in the outputs of Generative AI and trigger a review by a human or another Generative AI model, or redact PII information in the outputs of Generative AI before they are presented to the user. Similarly, some banks also focused efforts on developing guardrails that were reusable.

7.1.15. Banks mitigated data security risks when using Generative AI by either using private cloud solutions for Generative AI models, or open-source models on-premise, which keep sensitive data within controlled environments (either dedicated cloud resources not shared with other organisations, or on-premise servers) which can reduce the risks of exposure of data to external parties. Legal agreements with solution providers, data loss prevention tools, as well as limits on the classification of data that could be used in Generative AI were also important to mitigate data security risks.

7.1.16. Another common area that banks were exploring to address Generative AI risks were grounding methods[86] such as retrieval augmented generation (RAG)[87] where the outputs of Generative AI models are constrained based on internal knowledge bases, and source citations are provided to allow end-users to check for the accuracy of Generative AI outputs.

[86] Grounding methods help to ground or anchor the Generative AI outputs to factual, verifiable information, which can help reduce hallucinations and improve robustness.

[87] Retrieval-Augmented Generation (RAG) methods typically retrieve relevant information from a pre-defined knowledge base, and provide the retrieved information as context to the Generative AI model for the generation of outputs. For example, to generate an answer to a question, information relevant to the question would be first retrieved, and the retrieved information would then be provided as context to an LLM. The LLM would usually be instructed to answer the question based on the retrieved information. Links to the retrieved information could also be provided as source citations in the answer. There is however still the possibility of hallucinations occurring even with such approaches.

7.2 Third-Party AI

Overview

Existing third-party risk management standards and processes[88] continue to play an important role in banks' efforts to mitigate risks associated with third-party AI. As far as practicable, most banks also extended controls for internally developed AI to third-party AI. When considering the use of third-party AI, banks would weigh the potential benefits against the risks of using third-party AI. To address the additional risks arising from third-party AI, banks were exploring areas such as:

- conducting compensatory testing;
- enhancing contingency planning;
- updating legal agreements; and
- investing in training and other awareness efforts.

7.2.1 The use of third-party AI is increasingly common among banks, particularly in the context of Generative AI where most banks utilise Generative AI models that were pre-trained by an external party. However, the use of such third-party AI and Generative AI presents additional risks, such as unknown biases from pre-training data, data protection concerns, as well as concentration risks due to increased interdependencies, e.g., from multiple FIs or even third-party providers relying on common underlying Generative AI models. The lack of transparency is often cited as a key challenge in managing such third-party risks. Third-party AI providers may be reluctant to disclose proprietary information about their training data or algorithms, hindering banks' efforts in risk assessment and ongoing monitoring.

7.2.2 To mitigate these additional risks, banks were exploring various approaches, such as:

[88] This includes processes required to comply with MAS' Notice and Guidelines on Outsourcing (refer to https://www.mas.gov.sg/regulation/third-party-risk-management).

a. **Compensatory testing** - conducting rigorous testing of third-party AI models using various datasets and scenarios to verify the model's robustness and stability in the bank's context, and to detect potential biases.

b. **Contingency planning** - developing robust contingency plans to address potential failures, unexpected behaviour of third-party AI, or discontinuing of support by vendors. This can include having backup systems or manual processes in place to ensure business continuity.

c. **Legal agreements** - updating contracts with third-party AI providers to include clauses such as those pertaining to performance guarantees, data protection, the right to audit, and notification when AI is introduced (or not incorporating AI without the bank's agreement) in existing third-party providers' solutions. Such clauses could facilitate clearer expectations and responsibilities.

d. **Awareness efforts** – investing in training of staff on AI literacy and risk awareness to improve understanding and mitigation of risks; conducting surveys with third-party providers to gather more information about whether AI is being used in their products or services, and third-party providers' practices, including their AI development and risk management processes.

8 Conclusion

8.1. Robust oversight and governance of AI, supported by comprehensive identification, inventorisation of AI and appropriate risk materiality assessment, as well as rigorous development, validation and deployment standards and processes are important areas that FIs need to focus on when using AI. As the AI landscape continues to evolve, AI MRM frameworks will need to be regularly reviewed and updated, and risk management efforts scaled up based on the state of AI use. Aside from AI MRM, controls in non AI-specific areas such as general data governance and management, technology, cyber and third party risk management, and legal and compliance will also need to be reviewed to take AI developments into account.

8.2. As the AI landscape continues to evolve, MAS will continue to work with the industry to help facilitate and uplift AI and Generative AI governance and risk management efforts across the financial industry, through information sharing efforts such as this paper to promulgate industry best practices, and industry collaborations such as Project MindForge. MAS is also considering supervisory guidance for all FIs next year, building upon the focus areas covered in this information paper.

Annex A - Definitions

- **Model** – *A model is a method, system or approach which converts assumptions and input data into quantitative estimates, decisions, or decision recommendations (based on the Global Associate of Risk Professionals' definition of a model)*. Apart from **AI models**, which typically refer to machine or deep learning models which we define below, banks also routinely utilise **conventional models**, such as economic, financial, or statistical models. Some models, such as logistic regression models, are commonly used in both statistical and AI fields and may be regarded as both AI and conventional models.

- **Artificial Intelligence (AI)** – *An AI system is a machine-based system that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments. Different AI systems vary in their levels of autonomy and adaptiveness after deployment (based on the Organisation for Economic Co-operation and Development's definition of AI).* Such a definition would include Generative AI. An **AI or Generative AI *system*** can be based on one or multiple AI or Generative AI *models* and may also involve other machine-based components.

- **AI Use Case –** An AI or Generative AI use case usually refers to a specific real-world context that the AI or Generative AI model or system is applied to. For example, an AI recommendation model or system that is applied to a financial product recommendation use case.

- **Machine learning** – Machine learning is a subset of AI where the AI directly learns from data. The machine learning model learns model parameters (or model weights) to transform inputs into estimates or outputs from the data by updating these parameters iteratively based on an objective. For example, the machine learning model may be provided with historical data that consists of the information on customers, e.g., income and existing value of debt (which we refer to as input data), and whether the customer had defaulted on a loan obligation (which we refer to as the target variable or label). The machine learning model can then be trained by learning model parameters that allow it to transform input data to target variables or labels with maximum accuracy (or minimum error).

- **Deep learning** – Deep learning is a subset of machine learning, usually based on neural networks (that were inspired by how neurons in the brain recognise complex patterns in data) that comprise multiple layers of neurons. Deep learning models are able to learn more complex patterns due to the many layers of neurons in the model.

- **Discriminative versus Generative AI models** – AI models that generate predictions, e.g., predicting a credit default based on customer information, or recommending a financial product based on customer information, are usually referred to as discriminative AI models. This is in contrast to Generative AI models that are usually used to generate content such as text, images, audio or videos.

- **CI/CD, DevOps, MLOps, AIOps, LLMOps** – Continuous integration/continuous deployment (CI/CD) or DevOps pipelines automate the process of building, testing, and deploying code changes. These terms are closely related to the term MLOps, which is used to describe tools and systems that help to automate the process of building, testing, deploying and monitoring the performance of machine learning systems. More recent terms such as AIOps and LLMOps have also been used to describe such tools and systems for AI in general or for Large Language Models (LLM).

- **Data Drift** - This occurs when the statistical properties of the distribution of the data changes. For example, the underlying distribution of customer data may have drifted or changed over time due to changes in the lifestyles of customers. Hence, an AI model that was trained on data from a more distant time period may not perform as well on data from a more recent time period due to data drift. A common measure of how much a population distribution has changed over time is the Population Stability Index (PSI).

- **Concept Drift** - This occurs when the underlying relationships between the features in input data and what the AI model is being used to predict or generate changes. For example, customer preferences for financial products may have shifted due to broad industry changes (e.g., a shift in the relationships between customer information and their preferences for financial products), and an AI model used to generate financial product recommendations may no longer perform as well due to such concept drifts. A common measure of concept drift is the Characteristic Stability Index (CSI).

- **Model Drift** - Model drift is a broader term that usually encompasses both data drift and concept drift, as well as other factors that can cause a model's performance to degrade over time. Aside from measures such as PSI and CSI, monitoring the statistical characteristics of AI predictions can also be used to detect drifts in general.

- **Supervised learning** – Supervised learning is a machine learning approach where a model is trained on a labelled dataset. In this process, each data point includes input features paired with the corresponding output (label). The model learns to map inputs to outputs by comparing its predictions with the actual labels and updating the model parameters iteratively. Classification, which involves the prediction of classes or categories, and regression, which involves the prediction of continuous values, are common examples of supervised learning.

- **Unsupervised learning** – Unsupervised learning is a machine learning approach where a model discovers patterns in data without the use of labels. An example of unsupervised learning is clustering, where data points are grouped together based on their inherent similarities or dissimilarities.

Annex B - Useful References

Publications for the Financial Sector issued by MAS

- MAS FEAT Principles: *https://www.mas.gov.sg/publications/monographs-or-information-paper/2018/feat*

- Veritas Initiative: *https://www.mas.gov.sg/schemes-and-initiatives/veritas*

- Project MindForge: *https://www.mas.gov.sg/schemes-and-initiatives/project-mindforge*

- Information Paper on Implementation of Fairness Principles in Financial Institutions' use of Artificial Intelligence/Machine Learning: *https://www.mas.gov.sg/publications/monographs-or-information-paper/2022/implementation-of-fairness-principles-in-financial-institutions-use-of-artificial-intelligence-and-machine-learning*

- Information Paper on Cyber Risks Associated with Generative Artificial Intelligence: *https://www.mas.gov.sg/regulation/circulars/cyber-risks-associated-with-generative-artificial-intelligence*

- Information Paper on Data Governance and Management Practices: *https://www.mas.gov.sg/publications/monographs-or-information-paper/2024/data-governance-and-management-practices*

- Technology Risk Management Guidelines: *https://www.mas.gov.sg/regulation/guidelines/technology-risk-management-guidelines*

- Business Continuity Management Guidelines*:* *https://www.mas.gov.sg/regulation/guidelines/guidelines-on-business-continuity-management*

- Notice and Guidelines on Third-Party Risk Management: *https://www.mas.gov.sg/regulation/third-party-risk-management*

- Information Paper on Operational Risk Management - Management of Third Party Arrangements: *https://www.mas.gov.sg/publications/monographs-or-information-paper/2022/operational-risk-management---management-of-third-party-arrangements*

Non-Financial Sector Specific Publications

- AI Verify: AI governance testing framework and software toolkit: *https://www.aiverifyfoundation.sg/what-is-ai-verify/*
- Project Moonshot: *https://www.aiverifyfoundation.sg/project-moonshot/*
- Model Governance Framework for Generative AI: *https://www.aiverifyfoundation.sg/resources/mgf-gen-ai/*
- Trusted Data Sharing Framework: *https://www.imda.gov.sg/how-we-can-help/data-innovation/trusted-data-sharing-framework*
- Personal Data Protection Act (PDPA): *https://www.pdpc.gov.sg/overview-of-pdpa/the-legislation/personal-data-protection-act*
- 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*
- Guidelines and Companion Guide on Securing AI Systems: *https://www.csa.gov.sg/Tips-Resource/publications/2024/guidelines-on-securing-ai*