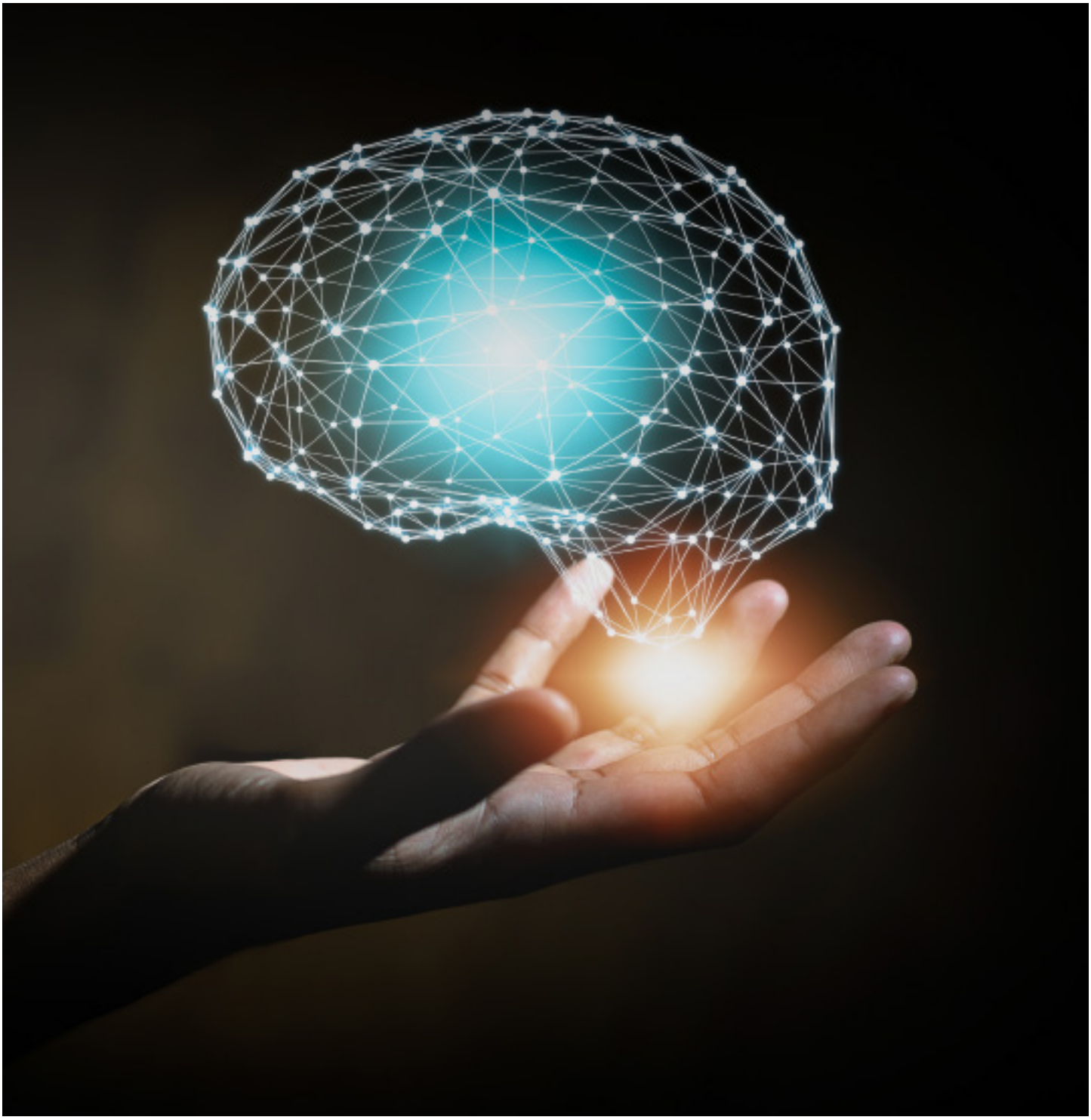
Enabling an Efficient Regulatory Environment for AI - Practical Considerations for Generative AI



JANUARY 2024

asifma
Growing Asia's Markets

Disclaimer

The information and opinion commentary in this Enabling an Efficient Regulatory Environment for AI - Practical Considerations for Generative AI (Paper) was prepared by the Asia Securities Industry and Financial Markets Association (ASIFMA) to reflect the views of our members. ASIFMA believes that the information in the Paper, which has been obtained from multiple sources believed to be reliable, is reliable as of the date of publication. As estimates by individual sources may differ from one another, estimates for similar types of data could vary within the Paper. In no event, however, does ASIFMA make any representation as to the accuracy or completeness of such information. ASIFMA has no obligation to update, modify or amend the information in this Paper or to otherwise notify readers if any information in the Paper becomes outdated or inaccurate. ASIFMA will make every effort to include updated information as it becomes available and in subsequent Papers.

asifma
Growing Asia's Markets

ASIFMA is an independent, regional trade association with over 160 member firms comprising a diverse range of leading financial institutions from both the buy and sell side including banks, asset managers, law firms and market infrastructure service providers.

Together, we harness the shared interests of the financial industry to promote the development of liquid, deep and broad capital markets in Asia. ASIFMA advocates stable, innovative and competitive Asian capital markets that are necessary to support the region's economic growth. We drive consensus, advocate solutions and effect change around key issues through the collective strength and clarity of one industry voice. Our many initiatives include consultations with regulators and exchanges, development of uniform industry standards, advocacy for enhanced markets through policy papers, and lowering the cost of doing business in the region. Through the GFMA alliance with SIFMA in the US and AFME in Europe, ASIFMA also provides insights on global best practices and standards to benefit the region.

www.asifma.org

If you have any comments or questions, please reach out to Laurence Van der Loo, Managing Director - Head of Technology & Operations (lvanderloo@asifma.org), Celina Leung, Senior Manager - Technology & Operations (cleung@asifma.org), or Leonardo Troeman, Analyst - Technology & Operations (ltroeman@asifma.org).

1
INTRODUCTION

ASIFMA's paper on "Enabling an Efficient Regulatory Environment for AI"[1] ("ASIFMA 2021 AI Paper") was launched in 2021 as a guiding set of regulatory principles for Artificial Intelligence (AI) which would help form the basis for building an efficient regulatory environment whilst at the same time supporting customer and investor protection, market integrity and financial and systemic stability. These principles recommended that regulators follow technology-agnostic, risk-based, and principle-based approaches to regulating AI in order to help increase cross-border regulatory standardisation and harmonisation. Since the launch of ChatGPT, excitement as well as anxiety have been heating up globally around the opportunities and challenges of generative AI. Governments and regulators across the world are starting to consider regulation and are at various stages of the drafting process regarding generative AI regulation.

This 2023 addendum to the 2021 paper focuses on generative AI and was developed with the support of ASIFMA's Generative AI Taskforce and the Fintech Working Group which consist of banks, brokers, asset managers, professional firms, market infrastructure providers and technology companies. The aim of this paper is to initiate further proactive engagement with regulators on generative AI, to help advance the public-private dialogue and to push forward greater collaboration so as to ensure that generative AI is used in a secure and responsible manner in the capital markets industry. We hope this paper will help to guide stakeholders in staying ahead of the curve and preparing for future regulatory developments. This paper will also serve as a useful tool in helping to educate and inform the market and regulators on the various opportunities, gaps, and challenges prevalent in the use of generative AI in capital markets and outline the industry's views on how to best seize those opportunities and address those gaps and challenges.

We wish to thank all members of the Fintech Working Group and Generative AI Taskforce for their valuable input and contributions to the paper.

[1] ASIFMA, 'Enabling an Efficient Regulatory Environment for AI', 2021, https://www.asifma.org/wp-content/uploads/2021/06/enabling-an-efficient-regulatory-environment-for-ai-report_june-2021.pdf

2

EXECUTIVE SUMMARY

Whilst the upside and benefits of the use of generative AI in capital markets are undeniable, generative AI also introduces new challenges. In this paper, we will assess the new and incremental challenges of generative AI, explore how these are already addressed by existing regulations, tools and governance frameworks, explore mitigants, identify any gaps and make suggestions on how to address the gaps with the aim to ensure the safe and responsible adoption of generative AI in the capital markets industry so as to realise the full benefits of generative AI.

As the world embarks on the journey of formulating governance frameworks for generative AI, with some further ahead than others, it is important to recognise that everyone needs to come together to have a common consensus on making it a safe environment and yet harness the advantages that generative AI can bring.

The paper continues to reiterate the importance of the 7 principles as mentioned in our 2021 AI Paper that regulators should consider. We have also put forward an additional set of principles which we believe will allow for a balanced regulatory approach for generative AI. Specifically, we recommend that regulators should:

ASIFMA PRINCIPLES FOR GENERATIVE AI

- Principle 1:
Leverage existing regulations
- Principle 2:
Maintain governance and accountability
- **Principle 3:**
Provide the appropriate level of transparency
- Principle 4:
Continue to adopt a risk-based approach
-Principle 5:
Continue to adopt a technology-agnostic regulatory approach
-Principle 6:
Address any IP Protection challenges
-Principle 7:
Strive for regulatory certainty and a harmonised framework

We look forward to engaging with regulators and other stakeholders on our suggested principles and key findings to support an enabling regulatory framework for generative AI in APAC and beyond.

3

WHAT IS GENERATIVE AI

Generative AI is a type of AI that can generate various types of content, including text, images, audio, programming code and video, and more, based on the data it has been trained on and the algorithms used. While there is no formal definition of "generative AI," it is generally understood to be a subset of traditional machine learning (ML) that generates new and useful outputs from simple inputs. In layman's terms, generative AI allows a user to provide a 'prompt' and receive a newly generated output in the desired medium. This 'generated output' is a key differentiating factor between generative AI and prior models. Generative AI has progressed ML beyond models that analyse a set of data and find resulting patterns, to machines capable of creating entirely new content.

ChatGPT, one of the most well-known generative AI tools, defines generative AI as "a category of AI techniques and models that are designed to generate new content, data, or information. These models are trained on large datasets and learn to produce outputs that are similar to the examples they were trained on. They work by learning the underlying patterns and structures present in the training data and then use generalised patterns to generate new, and potentially previously unseen content."

Generative AI is powered by ML models — but specifically for generative AI, very large models that are pre-trained on vast amounts of data and commonly referred to as Foundation Models (FM). FM is a conceptual term that is used to describe incredibly powerful, general-purpose models that can be customised for specific use cases without having to build a model from scratch each time. Generative AI tools are usually based on FMs like Generative pre-trained transformer (GPT) - a large language model (LLM) created by OpenAI, which are trained on a broad set of data that is adaptable to a wide range of tasks and can be fine-tuned for specific tasks.

A class of FMs, such as the GPT models, commonly referred to as LLMs are specifically focused on language-based tasks such as such as summarisation, text generation, synthetic text generation, classification, open-ended questions and answers, and information extraction.

What makes LLMs special is that they can perform so many more tasks because they are capable of learning advanced concepts, potentially resulting in greater scale of adoption across Financial Institutions (FI), either as standalone AI systems or embedded as an AI component in previously non-AI related software. And through their pre-training exposure to internet-scale data in all its various forms and myriad of patterns, LLMs learn to apply their knowledge to a wide range of contexts.

While the capabilities and resulting possibilities of pre-trained FMs are impressive, their adaptability through customisation to perform domain-specific functions makes it even more exciting to businesses. As a result, businesses can build highly efficient and productive applications with FMs using only a small fraction of the substantial data and compute required to train a FM from scratch.

4

LEVERAGING THE BENEFITS OF GENERATIVE AI IN CAPITAL MARKETS

Generative AI capabilities have advanced rapidly in recent years as a result of new breakthroughs in deep learning architectures (e.g., "transformer" models for text and "diffusion" models for images). These developments have made this technology a powerful tool for businesses to innovate, enhance customer experiences, boost productivity and creativity, and optimise business processes. Examples of specific use cases include:

- **Writing assistance:** Generative AI can be a useful tool to help prepare initial drafts of text (e.g., emails or speeches) or to review existing texts to make them more polished or professional.

- **Summarisation:** Generative LLMs can be used to summarise and extract key information from long articles, documents, or webpages. Summarisation can be helpful for people who want to quickly understand the main idea of a text or for researchers who need to read multiple papers and want to understand the key findings.

- **Software development:** Generate code snippets, comments, and documentation based on natural language inputs. This can improve the efficiency and accuracy of software development tasks.

- **Conversational AI:** Create natural-language-based conversational interfaces such as chatbots and virtual assistants. Conversational AI can include speech-to-text and translation capabilities.

- **Brainstorming:** Generative AI can serve as a brainstorming tool that enables a user to input. By giving generative models a specific prompt or theme, they can quickly generate unique and interesting pieces of content.

Generative AI has the potential to create significant value for FIs, driving productivity, innovation, and economic growth. Generative AI capabilities can transform processes such as portfolio management, financial documentation, intelligent advisory, fraud detection, compliance assistance, and so on. As a result, FIs can improve their internal and external customer experience, increase the efficiency of knowledge workers, automate processes, and innovate new products. For example, an investment management firm can build an investment analyst assistant using generative AI tools. Rather than employees spending significant time writing code or using software to create charts and tables, generative AI can produce charts and tables based on natural language query posed by the analysts. To understand challenges and trends at a global level, investment analysts need to sift through massive amounts of data from disparate sources. Generative AI can also power investment research, providing fast and effective Q&A capabilities with capital markets data. Capital markets organisations can also transform client engagement through the use of chatbots and virtual assistants, and they can rely on generative AI tools to ease and automate their regulatory compliance journey. Through such use cases, capital markets firms can develop unprecedented capabilities to translate data into insights and impact, develop new revenue streams, and provide better products and experiences to their customers.

Adoption of generative AI in the financial industry has been varied at this time. FIs are prudently evaluating safe and effective ways of leveraging generative AI, including by building enterprise-grade applications of FMs in controlled environments.

5

REVIEW AND APPLICATION OF 'PRINCIPLES FOR AI REGULATION' FROM JUNE 2021 PAPER 'ENABLING AN EFFICIENT REGULATORY ENVIRONMENT FOR AI'

In our 2021 AI paper, we proposed a set of regulatory principles for AI which we believe will form the basis for an efficient regulatory environment whilst at the same time supporting customer and investor protection, market integrity and financial and systemic stability:

ASIFMA 2021 PRINCIPLES FOR AI REGULATION

- Principle 1:
Support public-private collaboration
- **Principle 2:**
Allow FIs to take a risk-based approach to AI, taking materiality of the use case and stakeholders into account
- Principle 3:
Take a technology-agnostic approach to regulation
- Principle 4:
Leverage existing regulatory frameworks
- Principle 5:
Strive for regional and international harmonisation
- Principle 6:
Promote and facilitate cross-border data flow
- Principle 7:
Engage with the industry on areas that need further discussion

We continue to support these principles and they are even more important as we consider the future regulatory approaches to generative AI. At the same time, as we will outline in section 7 below, our challenges and controls assessment shows that there are some incremental or new aspects in relation to generative AI. In section 8, we put forward some additional considerations and suggestions pertaining to generative AI for regulators and industry to consider.

6
STAKEHOLDERS AND DEPLOYMENT MODELS

There are two common roles in an AI system's value chain that FIs might play, "provider" and "deployer/user". Here are their example definitions:[2]

Provider: A natural or legal person, public authority, agency, or other body that develops an AI system or that has an AI system developed with a view to placing it on the market or putting it into service under its own name or trademark, whether for a fee or free of charge.

Deployer/user: Any natural or legal person, public authority, agency, or other body using an AI system under its authority except where the AI system is used in the course of a personal non-professional activity.

Deployment options for generative AI applications impact the scope of FIs' responsibilities and approaches for risk management over the AI system lifecycle. They are classified into two categories – buyers of generative AI applications and builders of generative AI applications.

Buyers of generative AI applications

When FIs purchase generative AI applications, they act as a deployer/user. One option is that they purchase the consumer application, a public third-party generative AI service, either at no-cost or paid. Within this scope, FIs do not own or see the training data or the model, and FIs cannot modify or augment it. FIs, however, can implement an application programming interface (API) or directly use the application according to the terms of service of the provider. (e.g., an FI employee interacts with a generative AI chat application to generate ideas for an upcoming marketing campaign). The other option is that FIs purchase an enterprise application, a third-party enterprise application that has generative AI features embedded within, and a business relationship is established between FIs and the provider. (e.g., FIs use a third-party enterprise scheduling application that has a generative AI capability embedded within to help draft meeting agendas).

Builders of generative AI applications

When FIs build generative AI applications themselves, they take on the role of "provider" and "deployer/user". Firstly, FIs can build their own applications using an existing third-party pre-trained generative AI FM and directly integrate it with their workload through an API (e.g., an FI builds an application to create a customer support chatbot that uses the Anthropic Claude FM through Amazon Bedrock APIs). Secondly, FIs could build with applications using fine-tuned models. An FI refines an existing third-party generative AI FM by fine-tuning it with data specific to their business, generating a new and, enhanced model that is specialised to their workload (e.g., using an API to access a FM, an FI builds an application for their marketing teams that enables them to build marketing materials that are specific to their products and services). Thirdly, FIs could build with self-trained models. An FI builds and trains a generative AI model from scratch using data that they own or acquire in which they own every aspect of the model (e.g., an FI creates a model trained exclusively on deep, industry-specific data to license to companies in that industry, creating a completely novel LLM).

"Providers" and "deployer/users" share the responsibility for risk management associated with the use of generative AI. This establishes accountability and ensures responsibilities of the generative AI service provider and its consumers with regard to responsible use of generative AI. When understanding the risks associated with the use of generative AI, it is critical to note that the allocation of risk responsibilities is contingent upon the particular AI application scenario that could originate from the generative AI service provider or the user of the generative AI service and such allocation should be contractually agreed between the involved parties, where it is not specified by the regulators.

[2] European Parliament, EU Artificial Intelligence Act, June 2023, https://www.europarl.europa.eu/doceo/document/TA-9-2023-0236_EN.html

7

ASSESSMENT OF NEW AND INCREMENTAL CHALLENGES, EXISTING CONTROLS AND GAP ANALYSIS

In what follows, we evaluate whether there are any new or incremental challenges pertaining to generative AI. We note that whilst some challenges as highlighted in our 2021 AI paper are not new or incremental to generative AI, regulators should be aware of the amplified impact as the use of generative AI accelerates. Regulators should encourage the adoption of this new technology while encouraging their regulated entities to develop appropriate risk-based frameworks for their generative AI adoption. Such frameworks need to recognise that FIs could play different roles in the value chain of generative AI applications, and guide them to identify relevant challenges,and implement mitigants in relevant stages of the AI system lifecycle[3], given the chosen deployment option. For example, when assessing a challenge, FIs need to determine whether they should directly implement mitigants and/or via the AI model or service providers' compliance that is governed under FI's third party risk management framework along the lifecycle.

FIs should clearly establish the context, their roles, and the deployment option to determine the appropriate risk management approach.

A. Data-Related Considerations

A.1 Data Quality/Integrity

Incremental challenges of generative AI

In our 2021 paper, we covered a range of data quality issues that may affect the accuracy of AI models, including for example data poisoning. These challenges also apply to generative AI, likely to a greater extent.

- Training data disclosure

For open-source/freely available LLMs, training data is rarely disclosed, which makes it more challenging to evaluate data quality, privacy, and security in such cases.

- Data quality

Recent research has shown that data quality (as opposed to quantity) is a far more important factor in LLMs than one might have expected. Ongoing research found that fine-tuning a strong pretrained language model with a small set of carefully curated examples can produce remarkable results on a wide range of prompts compared to other state-of-the-art LLMs that are fine-tuned with large-scale instruction tuning and reinforcement learning.[4]

- Model collapse

A new challenge of generative AI is called model collapse. Model collapse is defined as a degenerative process affecting generations of learned generative models, where generated data end up polluting the training set of the next generation of models. As the subsequent models are then trained on polluted data, it would result in them mis-perceiving reality.[5] In short, if LLMs are trained on data, i.e., synthetic data that is generated by other LLMs and not properly labelled, over time, these LLMs would drift away from accuracy and precision and possibly into falsehoods that will inadvertently converge to a point that is far from reality.

[3] OECD, 'Recommendation of the Council on Artificial Intelligence', November 2023, https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449 OECD defines the AI system lifecycle as 'design, data and models', 'verification and validation', 'deployment', and 'operation and monitoring

[4] Chunting Zhou et al., 'LIMA: Less Is More for Alignment', May 2023, https://arxiv.org/pdf/2305.11206.pdf

[5] Ilia Shumailov et al., 'The Curse of Recursion: Training on Generated Data Makes Models Forget', May 2023, https://arxiv.org/abs/2305.17493

Mitigants

Below are some of the suggested mitigants of Data Quality/ Integrity challenges:

- Allocation of resources

Regarding the heightened importance of data quality, both developers and deployers of LLMs may need to invest more time, effort, and resources to ensure data quality. For example, in the Meta paper, the researchers manually curated 1,000 high-quality prompts and responses as training examples.[4] Such efforts are highly laborious, but likely necessary to uphold the quality of the data, and hence the AI models.

- Transparency

Developers will need to provide transparency and documentation on the source of the data which has been used to train a model as well as the reasons for selection of the data in order to meet the deployers' internal oversight and governance obligations.

- Quality standards for external data might be explored

To build trust in data coming from external sources, it would be positive for the industry and regulators to promote initiatives that certify the quality of external data (e.g., the origin and integrity of their data is well documented; continuity plans to ensure the provision of the same data in the future, etc). In this respect, Legal Entity Identifier data is an example of an open-source, free of charge data that is used as a public good.

- Watermarking

Regarding the challenges of model collapse, since it may be inevitable that the Internet would be inundated with AI-generated content, developers of LLMs may need to ensure that they are able to train the models as intended, for example using only genuine human-generated data as required. To do so, they will need solutions to accurately distinguish between AI-generated and human-generated data, also known as "watermarking". Traditional watermarks (e.g., stamp imprinted on an image) are unsatisfactory solutions as they present aesthetic challenges, especially for creative purposes, and can be easily edited out with technology. DeepMind recently released a beta version of SynthID, a tool for watermarking and identifying AI-generated images by embedding a digital watermark directly into the pixels of an image, making it imperceptible to the human eye, but detectable for identification. Similar technology could potentially evolve and be applied to text, audio, and video, subject to technical feasibility.

A.2 Data Privacy/Protection/ Security

Incremental challenges of generative AI

While the fundamental challenges associated with generative AI are the same as those highlighted in our 2021 AI paper, there are certain areas that require particular attention and consideration with respect to generative AI:

- Data rectification/deletion

In many jurisdictions, individuals are afforded legal rights in respect of accuracy, rectification, and deletion of their personal data (e.g. 'the right to be forgotten'). These obligations could be technically challenging for organisations to comply with in the context of generative AI as it is challenging for generative AI models to 'detrain' or 'forget' the data it was trained upon.

[4] Chunting Zhou et al., 'LIMA: Less Is More for Alignment', May 2023, https://arxiv.org/pdf/2305.11206.pdf

- **Output data**
  A generative AI model may falsely generate incorrect or flawed personal data, which is an example of hallucination, but in the context of personal data. (See Section '7B.1 Hallucination' below)

- **Disclosure of personal data**
  If generative AI models "memorise" specific data records, there is a possibility that the models will then exactly replicate those records (as opposed to "generalising" from the broad data set) when responding to a user query, which could lead to the inadvertent disclosure of personal data, as well as the revelation of sensitive and/or confidential information.[6]

- **Basis for data processing**
  One of the challenges, for example, is where the use of an individual's personal data for the purposes of training or inputting into a generative AI system is beyond the limits of the original collection consent obtained from that individual.

- **Balancing fairness and data minimisation**
  Privacy laws often limit the collection of personal data to what is strictly necessary (data minimisation). However, there is trade-off and balance to be struck between data minimisation and fairness, as generative AI developers and deployers may need data, including sensitive personal data such as race and ethnicity, to conduct fairness assessments.

Mitigants

Below are some of the suggested mitigants of Data Privacy/ Protection/Security challenges[7]:

- **Consent**
  Where practical and necessary, FIs should take into account which jurisdictions' privacy laws apply to the relevant data flow, obtain clear and unambiguous consent from users, or otherwise ensure that there is a clear, valid, legal basis for the collection of data. FIs should also provide clear and easily understandable information about how their data will be used and the purpose of the generative AI applications. This would apply to Intellectual Property (IP) Protection as well. (See Section '7A.3 Intellectual Property Protection').

- **Security measures**
  Implement robust security measures to protect user data from unauthorised access, loss, or breaches and ensure that regular assessments of security protocols take place. Where a third party is hosting the generative AI system and processing personal data, ensure technical due diligence is performed to understand the third party's security measures, and that contractual protections are in place in the event of a security breach.

- **Frozen foundational models**
  Most providers of enterprise applications of generative AI offer FIs a "frozen foundation model", i.e. a FM that is not further modified. FIs can use their own confidential data, including prompts, as input to this frozen FM and even fine-tune the model using an "adapter layer". The data and adapter layer are unique to the FI, secured within the FI's own cloud project, and only accessible by the FI. Most importantly, when the FI makes an inference, the FM receives the adapter layer, runs through the request, and returns the results, without modifying the FM or storing the request, which is why the FM is "frozen". Such enterprise-grade security measures help to ensure that FIs' data can remain confidential when using generative AI.

[6] IMDA, 'Generative AI: Implications For Trust And Governance', Jun 2023, https://aiverifyfoundation.sg/downloads/Discussion_Paper.pdf
[7] Justin Antonipillai, 'The Intersection of Generative AI, Data Privacy, and GDPR: Unlocking Marketing Opportunities Responsibly', July 2023, https://wirewheel.io/blog/the-intersection-of-generative-ai-data-privacy-and-gdpr-unlocking-marketing-opportunities-responsibly/

A.3 Intellectual Property Protection

Incremental challenges of generative AI

Below are some of the incremental challenges of IP Protection:

- Training data challenges

Generative AI models train on large amounts of data. The models, particularly FMs, can scrape the web and other sources for data which can be copyrighted and therefore, the output of generative AI models could breach copyrights.[8] It is unclear whether fair use exceptions would apply. However, if the regulators prohibit a FM from being deployed in a jurisdiction, it could put critical business services (including customer facing ones) at risk.

- IP leakage and trade secret challenges

Some generative AI models may memorise and output significant parts of the data that they are trained on, as well as the data provided in the prompts. This means that subsequent outputs, generated by the model for other users, contain or are derived from those input data and prompts. One recent example of this is when Samsung employees unintentionally leaked copyrighted source code when they were leveraging ChatGPT to optimise the code.[9]

- Software licensing challenges

Software code written by generative AI may incorporate third-party code used to train the model, including open-source code. In this event, the generated code will have to comply with any licences applicable to the original code. In particular, certain licences may require new (and proprietary) software to be made open-source.

- IP ownership challenges

There is uncertainty surrounding IP rights of outputs generated by generative AI. However, a United States (US) federal district court had ruled that AI-generated artwork is not eligible for copyright protection under the US law, explaining that human authorship is a "bedrock requirement of copyright".[10] At the same time, jurisdictions such as Japan apply copyright protection to works that are produced using a computer as a tool if there is a human's creative intention and creative contribution.

Mitigants

Below are some suggested mitigants of IP Protection challenges:

- Data source

Organisations need to understand the source of data and have a mechanism to distinguish the originality of content produced.

- Transparency

Controls and transparency requirements will have to be embedded throughout the generative AI lifecycle, by developers and users alike. However, there also may be an imbalance of power between large market vendors and FIs, resulting in an inability of FIs to obtain adequate transparency.

- Contractual requirements

FIs should embed protections into the contracts with third-party AI developers to obtain appropriate contractual protections.

- Consent

See Section '7A.2 Data Privacy/Protection/Security'. above.

- Frozen foundational models

See Section '7A.2 Data Privacy/Protection/Security' above.

[8] Jonathan Stempel, 'NY Times sues Open AI, Microsoft for infringing copyright works', December 2023, https://www.reuters.com/legal/transactional/ny-times-sues-openai-microsoft-infringing-copyrighted-work-2023-12-27/
[9] Siladitya Ray, 'Samsung Bans ChatGPT Among Employees After Sensitive Code Leak', May 2023, https://www.forbes.com/sites/siladityaray/2023/05/02/samsung-bans-chatgpt-and-other-chatbots-for-employees-after-sensitive-code-leak/?sh=3d00ab9a6078
[10] Josh Gold-Quiros et al., 'Court Finds AI-Generated Work Not Copyrightable for Failure to Meet "Human Authorship' Requirement—But Questions Remain', August 2023, https://www.jdsupra.com/legalnews/court-finds-ai-generated-work-not-2083236/

B. Model-Related Considerations

B.1 Hallucination

Incremental challenges of generative AI

AI hallucination is a phenomenon wherein generative AI generates results with falsehoods, a lack of informativeness and a lack of relevance to the context.

Text-to-text generative AI models are systems that can produce natural language texts based on some input texts (or prompts). These models are evaluated on how well they can avoid hallucination – the phenomenon of generating texts that are not grounded in reality, factual, or relevant to the input – as a key criterion. To measure the degree of hallucination, three criteria can be used:[11]

- Truthfulness

The generated text does not contain any false or misleading information. This is core to the concept of hallucination.

- Informativeness

The generated text provides useful and pertinent information that answers the user's query or satisfies the user's goal. This is the completeness and relevance of information provided and there can be other reasons leading to a lack of informativeness, for example, erroneous, inconsistent and/or outdated data. By itself, it is not a necessary indicator of hallucination. However, a lack of truthfulness, together with a lack of informativeness, become sufficient conditions for hallucination.

- Completeness

The generated text provides partial, yet correct, information. This is a common issue due to FMs' inability to retain context beyond a certain size (also known as context window). While it is noted that this is not exactly hallucination, it can be an issue when the outcomes of generative AI applications are expected to be complete and used as a basis for decision making.

In the instances when hallucination or a lack of truthfulness happens, it is not a necessarily completely black-box phenomenon but rather a combined result of various factors that have affected the model's behaviour. At the time of writing, there is no clear explanation to the cause of the hallucination or established techniques that can prevent the hallucination from happening with certainty. Some of these factors depend on the choices and actions of humans and organisations, which may either reduce or amplify the challenges of hallucination. Below are some common factors that can influence the challenges of hallucination.

- Data quality

One of the factors that exacerbates hallucination is ineffective or lack of quality training data – which means that the model is not trained on enough or relevant data to produce more accurate outputs for the domain that it is intended for use. For example, there could have been limited quality data, irrelevant data, intentional negative biases in the training data, and "noisy" – missing, contradictory or ambiguous – input data used in its training, and any of these would lead to incorrect results that can be considered as hallucinated results. Such inaccuracies can be compounded by historical data that the FM has been built on, and itself could contain errors and inconsistencies. Depending again on the context of application, if the model is not effectively combined with updated live data for it to draw from to form its responses to user queries, the challenges of hallucination would be higher. Moreover, the deployment domain of the model should match its training domain, as a model trained on wholesale financial industry data and terms may perform poorly and exhibit higher personal bias when applied to retail banking. (See Section '7A.1 Data Quality' above)

[11] Stephanie Lin et al., 'TruthfulQA: Measuring How Models Mimic Human Falsehoods', September 2021, 2109.07958.pdf (arxiv.org)

- **Model training**
  A related factor is ineffective model training, for example, the effects of inadequate "overfitting" where a model learns from the training data too well that it would fail to generalise to new data – meaning that its results are consistent with the training data but not with reality. Another factor that can influence the level of hallucination in text-to-text generative AI models is the phenomenon of "forgetfulness," which refers to the loss of some abilities or knowledge by the model due to its data manipulation skills being distorted (as a consequence of being "overfitted") during the training process. This means that the model becomes too specialised or adapted to a specific task or dataset and fails to generalise well to new or different inputs or scenarios.

- **Development process and feedback**
  The development process can also contribute to a model's tendency to hallucinate. As a significant departure from traditional technology software development lifecycle (SDLC), where domain users are primarily involved in specific requirements definition and test stages, generative AI development process requires such domain expertise to be involved – and stay involved – right from the start and in every stage of its development. Domain experts are required to define expected training results, evaluate them, and provide feedback; acting together and in parallel with technical model adjustments by data engineers and scientists.

Mitigants

Below are some suggested mitigants of Hallucination challenges:

- **Upskilling users**
  An important first step is to educate users about how generative AI actually works, so there is no expectation that the citations or news-like stories produced are always genuine or factually correct. Users should also be educated on the particular generative AI system's capabilities and limitations to prevent any misuse.

- **Verification**
  The specific case of hallucinated citations could be mitigated by augmenting LLMs with independent, verified citation databases and similar sources, using approaches such as retrieval-augmented generation.[12] Other generative AI applications are linking to search engines. For instance, Google's Bard chatbot recently updated its "Google it" feature, which allows users to easily double-check Bard's responses by reading the response and evaluating whether there is content across the web to substantiate it. When a statement can be evaluated, the user can click the highlighted phrases and learn more about supporting or contradicting information found by Google Search.

- **Data attribution**
  Another nascent but intriguing approach is to develop methods for attributing generated outputs to particular pieces of training data, allowing users to assess the validity of those sources. This could help with explainability as well. (See Section '7B.3 Transparency, Explainability, Traceability' below)

- **Implementing boundaries**
  Some users choose to intentionally limit the scope of a FM to a specific domain by implementing boundaries within their models to prevent models from hallucination.

- **Adopt an AI-centric SDLC**
  Especially for LLMs, allow for continuous user feedback and model tuning including after the initial launch.

- **Data quality and training model**
  Ensuring the trained model is deployed into the context it is meant for; and ground it with updated high-quality data that is relevant to the use case as well as regular model review and maintenance to reduce overfitting.

[12] Amog Kamsetty, 'Retrieval Augmented Generation with Huggingface Transformers and Ray', February 2021, https://huggingface.co/blog/ray-rag

- Expertise

Ensuring relevant expertise in both data science and domain fields are available for model training and feedback.

- Model evaluation

A measurement of an LLM model can provide valuable insights into the reliability of its responses, facilitating risk assessment and error, and reducing hallucinations in natural language generation tasks.[13] Industry standard measurements of LLM confidence will need to be further developed. (See Section '7B.2 Foundational Model Selection' below)

B.2 Foundational Model Selection

Incremental challenges of generative AI

When using generative AI, selecting a FM is one of the first and most critical steps. Recall that a FM is a large ML model pre-trained on a vast quantity of data at scale resulting in a model that can be adapted to a wide range of downstream tasks. FM selection has strategic implications for how a use case gets built. Models vary across several factors, including level of customisation, model size, inference options[14], licensing agreements, context windows[15], and latency[16].

- Open-source vs proprietary

There are many FMs in the market, ranging from proprietary to open-source models, each having their own trade-offs. The proliferation of open-source FMs is giving rise to thriving ecosystems, such as Hugging Face. For Open-source FMs – such as OpenAI - the API is released, but not the model or training data, are also emerging.

- Hosting

Another key consideration in model selection is how the model can be served. Open-source models, as well as self-managed proprietary models, grant the flexibility to customise how and where the models are hosted. Directly controlling a model's infrastructure can help companies ensure reliability of their applications with best practices like autoscaling and redundancy. Managing the hosting infrastructure also helps to ensure that all data generated and consumed by a model is contained to dedicated environments which can adhere to security requirements set by the company.

- Lifespan

FIs should also consider the lifespan, including support, of the FM.

Given that FMs are the foundational architecture on which many generative AI applications are built, selecting the right FM for a particular use case is key.

Mitigants

Below are some suggested mitigants of Foundation Model Selection challenges:

- Existing regulations

The existence of well-established regulations in the areas of outsourcing, third-party risk management, technology risk management, cybersecurity and operational resilience should be considered and adhered to in the selection of a FM and its underlying provider.

- Use benchmarks to assist model selection

Generalised benchmarks (such as Stanford's Holistic Evaluation of Language Models)[17] are a great starting point because they help prioritise which FMs to start experimenting with. Custom benchmarks are useful for scenarios where generalised benchmarks are insufficient, such as use cases that are focused on building for a specific customer base. Custom benchmarking may include techniques such as calculating BiLingual Evaluation Understudy and Recall-Oriented Understudy for Gisting Evaluation scores[18]. These are two metrics that help quantify the number of corrections that are necessary to AI-generated text before giving it final approval for human-in-the-loop applications.

[13] Yijun Xiao and Wiliam Yang Wang, 'On Hallucination and Predictive Uncertainty in Conditional Language Generation', March 2021, https://arxiv.org/pdf/2103.15025.pdf
[14] Inference options - Different methods of generating output including speculation, next word predictions and stopping criteria.
[15] Context windows - The textual range around a token that can be processed at generation.
[16] Latency - The overall time it takes to generate a full response.
[17] Stanford University, 'A holistic framework for evaluating foundation models', https://crfm.stanford.edu/helm/
[18] Sthanikam Santhosh, 'Understanding BLEU and ROUGE score for NLP evaluation', April 2023, https://medium.com/@sthanikamsanthosh1994/understanding-bleu-and-rouge-score-for-nlp-evaluation-1ab334ecadcb

- Consider narrow models for specific tasks

Companies can also experiment with using narrow models meant for specific tasks, like following instructions or summarisation. These purpose-built models can significantly reduce a model's parameter count while maintaining its ability to perform domain-specific tasks.

- Model customisation

Open-source FMs empower companies to further customise and fine-tune their systems with their own datasets. For example, Parameter-Efficient Fine-Tuning solutions from Hugging Face have shown how adjusting a small number of model parameters, while freezing most other parameters of the pre-trained LLMs, can greatly decrease the computational and storage costs.[19]

- Vendor

The importance of third-party risk management remains critical, especially as a firm considers the challenges that the FM may not perform as intended, particularly in edge scenarios. Typical third-party risk management assessments and governance will help to manage these model challenges. One such example is outlining responsibilities associated with the model provider/ developer embedded in a contract.

B.3 Transparency, Explainability, and Traceability

Incremental challenges of generative AI

Generative AI systems often leverage LLMs, which are more challenging to explain or provide transparency on how the model works, as compared to for example a regression model or a decision-tree. This potentially leads to operational issues (such as not being able to explain how a business decision was made) and regulatory compliance issues (such as not being transparent regarding the usage or not being able to explain how decisions affecting individuals were made, as required under data protection laws).

It is also important to adapt the explanation to the audience, which – in the case of generative AI – can be broad and therefore adaptation to existing explanation techniques might be required.

Mitigants

Below are some suggested mitigants of Transparency, Explainability, and Traceability challenges:

- References

Providing further information on the training data set(s) that have been used to train the model, with a particular attention to additional information used to supplement pre-trained models.

- Citations

Whilst further research and investment will be required, there are benefits to augmented LLMs that are able to cite their data sources, particularly when they are referencing a smaller company specific sub-set of documents, e.g. policies.

- Relevance

Consideration should be given to whether generative AI and the chosen model are appropriate after taking into account the materiality/critical nature of a given use case. (e.g., analysing log files of a system and summarising the issues vs. providing investment advice to a customer).

More importantly, generative AI is not a silver bullet and should not be viewed as the solution to every problem. If intuitive explainability is very important, generative AI might not be the appropriate technology; FIs may wish to consider other forms of AI that produce more intuitive explainations.

B.4 Fairness

Incremental challenges of generative AI

In the 2021 paper, we stated that fairness is integral to prevent challenges that can arise when algorithms and learning models receive data that presents ingrained human flaws and biases, or when human decisions could make an algorithm discriminate unfairly. Fairness in AI protects FIs from deploying technology that undermines their codes of conduct and ethical values and conflicts with social values. This continues to be true. Unfair bias can be introduced into AI systems in a number of ways, for example low quality of, or inadequately representative, training data.

[19] Sourab Mangrulkar and Sayak Paul, 'PEFT: Parameter-Efficient Fine-Tuning of Billion-Scale Models on Low-Resource Hardware', February 2023, https://huggingface.co/blog/peft

- The challenges of unfair bias are particularly applicable to generative AI given its accessibility to a wide range of public data, as well as broad public interest, and the fact that many users may not necessarily be able to discern fact from fiction.

- Generative AI models can potentially generate harmful content, represent and amplify biases, and be at odds with generally shared values. Ethical failures in generative AI systems can also create reputational challenges by violating social norms and values.

- For example, FIs face growing challenges to meet the Monetary Authority of Singapore (MAS) "Fairness" FEAT principles because of the growing difficulty of identifying and responding to bias or prejudice in the inputs and outputs characteristic of generative AI systems.[20] This is due to the fact that it is challenging to assess for bias in sets of unstructured data, images etc, which are included in the datasets of generative AI systems.

Mitigants

Below are some suggested mitigants of Fairness challenges:

- The mitigation of unnecessary biased data
The removal or remediation of biased data from the dataset used to train the models results in an output that is provided only upon the meeting of specifically pre-defined supporting conditions.

- Standards
A defined standard for fairness with a robust process for identifying at-risk groups and potential proxies.

- Tried and true
Recommendations of trusted algorithmic methods to identify and reduce unfairness.

- Measurement
Fairness metrics to be put in place which will help to quantitatively assess fairness in the data set as well as AI outputs.

- Anonymity
Using anonymised data so as not to use distinguishing characteristics that could be used to identify and trigger bias upon output (e.g., replacement of Chairman with Chairperson).

- Weighting of parameters
During the pre-training of the model phase, assigning higher or lower weight to parameters related to majority or minority groups.

- Fair representation
Ensure that the dataset used to train the models is a comprehensive and fair representation of the population.

- Human oversight
Have humans interjection in during review process and to remove any bias and reinforcement learning from human feedback. These humans will need adequate training and expertise to identify discrepancies and be aware of generative AI's limitations. The people designing or implementing the models also need to evaluate and validate the systems to correct for bias and potential harm.

- Right to refusal
Some types of biases can also be mitigated via training for refusals, by getting the model to refuse responding to certain questions. However, these refusals can in practice also amplify biases and there are often publicised workarounds that can reduce their effectiveness.

- Collaboration
Regulator-industry-academic partnerships to create guidance on testing methodologies for fairness would be useful (e.g. the Hong Kong Monetary Authority guidance on testing for fairness in its white paper on "Reshaping Banking with AI", the MAS-led Veritas consortium).

[20] MAS, 'Emerging Risks and Opportunities of Generative AI for Banks: A Singapore Perspective', November 2023, https://www.mas.gov.sg/-/media/mas/news/media-releases/2023/executive-summary---emerging-risks-and-opportunities-of-generative-ai-for-banks.pdf

C. Third-Party Risk Management and Resilience

Incremental challenges of generative AI

Our 2021 paper notes that "Third-party risk management and the management of interdependencies is closely linked to the topic of (operational) resilience". Given the potential for generative AI to be used in different use cases in the financial industry, combined with the limited transparency of the FMs, this might lead to higher resilience challenges. This is potentially exacerbated by limited substitutability due to the fact that it is less straightforward to replicate training resources.

- **Increased reliance on third parties developing the FMs**
  With generative AI, reliance by FIs on third parties is increasing, because FMs are generally developed by third parties. There is a distinction between accountability and responsibility, where FIs must remain accountable, but third parties may be responsible for certain parts of the AI value chain.

- **Concentration challenges**
  Given the need for large volumes of data to develop a LLM or FM, the ability to create and run such a model is going to be limited to organisations with the performance, data, and storage capacity to operate this, leading to concentration challenges.

Mitigants

Below are some suggested mitigants of Third-Party Risk Management and Resilience challenges[21]:

- **Minimise shadow AI[22]**
  It is easier to manage what you can measure. In order to effectively mitigate the risks of third-party AI tools, organisations need a clear view and inventory of uses of AI within their operations.

- **Ensure that third-party risk mitigation is part of your Responsible AI (RAI) program**
  Full awareness of AI is not enough. In order to effectively detect and mitigate third-party risks, an FI's RAI program should extend to all uses of AI across the organisations, whether internally or externally built or developed, including through any relevant procurement and post-procurement policies

- **Uplift the organisation's third-party supply chain management process**
  Enhance third-party sourcing processes, AI and generative AI threats and controls assessment and required contract clauses, bridging attestation frameworks gaps.

- **Continually update and iterate to address new risks**
  As with all other components of RAI, third-party risk mitigation is not a one-off exercise. Whether due to advancements in AI or legal and regulatory developments, FIs should plan to regularly revisit their approach to third-party risk mitigation in their RAI program. There should also be clear contractual arrangements between FIs and third-party providers, including obligations around transparency and explainability, at least for critical activities.

- **Establish shared responsibility model**
  FIs should negotiate a shared responsibility model in third-party agreements, considering the service materiality and the service reliance on the third-party generative AI models and reviewing of models not covered under attestation frameworks. In the absence of any regulatory guidelines or industry framework on shared responsibility between AI developers and deployers, several generative AI service providers have proactively developed measures to address concerns by AI deployers. As an example, Google Cloud indemnifies its customers against liabilities arising from an infringement by its services of a third party's IP rights.[23] Microsoft is also committed to defend its customers from IP infringement claims arising from the customer's use and distribution of the output content generated by Microsoft's Copilot services.[24] FI deployers can also review the risk assessment and AI governance frameworks of the third-party developers and any gaps can be addressed in the contractual clauses. The responsibility of the various controls will depend on the deployment model of the generative AI tool and use case. We recommend that developers, deployers and the regulatory community come together to establish a mutual understanding of a shared responsibility framework.

[21] Elizabeth M. Renieris et al., 'Responsible AI at Risk: Understanding and Overcoming the Risks of Third-Party AI', April 2023, https://sloanreview.mit.edu/article/responsible-ai-at-risk-understanding-and-overcoming-the-risks-of-third-party-ai/\
[22] Shadow AI - the unauthorised use or implementation of AI that is not controlled by, or visible to, an organisation's IT department.
[23] Neal Suggs and Phil Veneables, 'Shared fate: Protecting customers with generative AI indemnification', October 2023, https://cloud.google.com/blog/products/ai-machine-learning/protecting-customers-with-generative-ai-indemnification
[24] Microsoft, 'Introducing the Microsoft Copilot Copyright Commitment', September 2023, https://www.microsoft.com/en-us/licensing/news/microsoft-copilot-copyright-commitment

D. Cybersecurity

Incremental challenges of generative AI

The cybersecurity challenges associated with the use of generative AI are a mixture of new threats such as malicious content generation and ad-hoc exploit creation, as well as increased exposure to existing cybersecurity challenges such as the adversarial attacks, data and model poisoning, and data and access control management. Cybersecurity professionals will be required to upskill their ability and adapt newer strategies as needed.

There are certain areas that require particular attention and consideration with respect to generative AI:

- **Standardisation of development lifecycle**
  The lack of a standardised development lifecycle for generative AI due to its relative novelty, potentially leaves gaps on applying security standards and principles in the design phase.

- **Lack of commercial / open-source for generative AI security testing**
  Security tools, to test against generative AI challenges, to provide verification and assurance, are still maturing.

- **Security controls bypass via generative AI inputs/outputs manipulation**
  Use of prompt engineering by skilled adversaries to bypass model controls built by the model or FM developers, to obtain information and/or perform activities that would otherwise be unauthorised.[25]

- **Third-parties attestation framework gaps**
  Assessor and attestation frameworks (such as International Organisations for Standards (ISO)/ International Electrotechnical Commission (IEC) 27001, ISO 28000, ISO/IEC 15408 and SOC 2 type 2) for third parties have yet to be updated to address the scope of AI and specifically generative AI challenges. This increases the complexity on both the third party and the FIs in supply chain cybersecurity risk management.

- **Deep Fake and impersonation attempts**
  FIs rely on biometric identification and authentication mechanisms (e.g. voice, images, video) for customer identification and authentication. Generative AI models could allow malicious actors to quickly train and generate high quality avatars based on publicly available data to impersonate and even deceive security controls resulting in an increase of fraudulent activities.[26]

Mitigants

Below are some suggested mitigants of Cybersecurity challenges:

- **Uplift the cybersecurity risk management framework**
  Assess the impact that AI and generative AI threats have on policies, technical standards, processes, and control implementation.

- **Cybersecurity assessment against AI and generative AI use cases**
  As part of the use case evaluation process, qualify use cases with the agreed risk appetite and establish effective governance mechanisms as part of the organisational cybersecurity risk framework.

- **Uplift the organisation's SDLC**
  Extending processes to cater to AI and generative AI specific concerns including data life cycle management (data sourcing, cleaning, maintaining linage between data model and training data).

- **Develop and/or source security training and awareness programs**
  This should be considered for all users, prioritising developers, security testing and assurance personnel.

[25] Andy Zou et al., 'Universal and Transferable Adversarial Attacks on Aligned Language Models', July 2023, https://arxiv.org/pdf/2307.15043.pdf

[26] (1) Thomas Brewster, 'Fraudsters Cloned Company Director's Voice In $35 Million Heist, Police Find', October 2021, https://www.forbes.com/sites/thomasbrewster/2021/10/14/huge-bank-fraud-uses-deep-fake-voice-tech-to-steal-millions/ ; (2) Bloomberg, 'Deepfake Imposter Scams Are Driving a New Wave of Fraud', August 2023, https://www.bloomberg.com/news/articles/2023-08-21/money-scams-deepfakes-ai-will-drive-10-trillion-in-financial-fraud-and-crime?leadSource=uverify%20wall

- **Raising awareness of the potential benefits and risks from the growing adoption of generative AI**

  As more FIs leverage generative AI to enhance their systems and business processes, there is an increasing need to guard against potential challenges, including leakage of sensitive information and data poisoning. FIs would need to manage these challenges by establishing guardrails, for example, by raising employee awareness on the safe use of generative AI and establishing comprehensive data handling policies.[27]

NB: Use of Generative AI as additional approach to mitigate cybersecurity threats

Generative AI can also serve as an additional tool in the toolbox for FIs to defend against cybersecurity threats in areas such as:

- Performing initial incident analysis.
- Intelligent threat hunting to identify security vulnerabilities.
- Dissect malware code samples to detect trends in malware software, such as polymorphic code, that makes such code more easily identifiable even when malware is constantly changing.
- Identiy communication patterns to detect a shift in sentiment that may suggest social engineering.
- Deploying AI-enabled solutions in areas such as secure code development, security monitoring, threat hunting and red-teaming to improve the effectiveness and robustness of their cyber defence.
- Generate synthetic data for cybersecurity training.

E. Expertise

Incremental challenges of generative AI

While the fundamental challenges associated with the use of generative AI by employees are the same as outlined in the 2021 ASIFMA AI paper, there are certain areas that require particular attention and consideration with respect to generative AI:

- Increased dissemination of false (e.g., deepfakes), harmful, or sensitive information due to lack of awareness and expertise.
- Misuse of generative AI applications due to lack of competency.
- Lack of AI experts within FIs risk and compliance teams to truly understand the capabilities and limitations of generative AI, and who are then able to effectively oversee and challenge generative AI implementation.
- Understanding that the flexibility/adaptability of generative AI means it is not possible to anticipate or prevent all possible misuse.

Mitigants

Below are some suggested mitigants of Expertise challenges:

- **Technical Training**

  Providing technical staff with a deep technical understanding and education in generative AI. For non-technical staff, provide education and training to deepen understanding and raise awareness, which can include regular updates and internal assessments. Training should be focused on real world scenarios, where AI can go wrong, the implications, how to improve governance, risk assessment and validation report being reviewed by a diverse group of people.

- **Education**

  Wider education and awareness raising for FI industry (and its users) on the applications and impacts of generative AI is needed, including various risk factors and regulations. For example, even where there is a human in the loop, this entails upskilling and training across employees to provide a consistent human review of output from the models , which is also referenced as a mitigant of hallucination as mentioned earlier in the paper. (See Section '7B.1 Hallucination' above).

- **Engage people from the user community to understand harm**

  Provide instructions for use or a user manual that has the right level of information to interpret the model's outcomes. Simple and easy language will help a broad range of users to understand.

- **Oversight**

  Ensuring AI experts are involved in the process of onboarding/using generative AI and the relevant governance committees to ensure appropriate oversight.

[27] MAS, 'MAS Cyber Security Advisory Panel Proposes Ways to Tackle Mobile Malware Scams and Generative AI Risks for the Financial Sector', October 2023, https://www.mas.gov.sg/news/media-releases/2023/csap-2023

F. Environmental Implications

Incremental challenges of generative AI

New research suggests that AI systems can have serious environmental impacts.[28] This includes its energy consumption, water consumption for cooling, carbon footprint, and contribution to e-waste as a result of the increased use of hardware as compared to other types of computing. As the generative AI models become larger, more computational power is needed and as such, more resources are required.

Mitigants

Below are some suggested mitigants of Environmental Implications:

- **Reduce**
  FIs should assess the sustainability of their own and their third-party data centers and their use of renewable energy and incorporate into relevant internal and external reporting.

- **Reuse**
  Reuse models and resources, particularly by leveraging global and vendor investments in FMs.[29]

8

ASIFMA PROPOSED RECOMMENDATIONS FOR A FUTURE-PROOF, RISK-BASED REGULATORY FRAMEWORK FOR GENERATIVE AI

Backdrop: Evolving regulatory environment

With the explosive growth of AI applications and the associated challenges coming to light, governments and regulators across the world, including in Asia, have become more proactive in developing requirements to govern the use of generative AI recently. In China, the Cybersecurity Administration of China (CAC) published the "Interim Measures for the Management of Generative AI Services" in July 2023 with more concrete AI regulatory requirements compared to the previous "Provision of AI Services" issued. This was followed by the proposed technical standards on generative AI issued by China's National Information Security Standardisation Committee, which provides specifics on complying with the CAC measures.[30] In India, the Ministry of Electronics and Information Technology indicated in May 2023 that AI regulations will be coming after earlier stating that they had no intention to issue AI regulations not long before. In South Korea, the government had announced the "Digital Bill of Rights" in September 2023, creating a set of guidelines and a base standard for future AI legislation.[31] In Hong Kong, the Chief Executive Officer of the Securities and Futures Commission highlighted the importance of addressing challenges associated with AI models in her recent speech at the Hong Kong Investment Funds Association in June 2023.[32] In Australia, the government issued a consultation in June 2023 exploring the governance approach of responsible AI.[33] In Singapore, the government published the "Singapore National AI Strategy 2.0", which is an update from the 2019 AI strategy, highlighting the need to take a pragmatic approach by supporting experimentation and innovation, while still ensuring that AI is developed and used responsibly, in line with the rule of law and the safeguards they have put in place.[34] In the European Union (EU), the European Council and the Parliament have reached an agreement on the provisional rules on the EU AI Act in December 2023, which aims to ensure that fundamental rights, democracy, the rule of law and environmental sustainability are protected from high risk AI, while boosting innovation and making Europe a leader in the field.[35] At the global level, the G7 have established the Hiroshima Process International Guiding Principles for Organisations Developing Advanced AI Systems and the Hiroshima Process International Code of Conduct for Organisations Developing Advanced AI Systems in September 2023, which was done in partnership with the Organisation for Economic Co-operation and Development (OECD) and Global Partnership on Artificial Intelligence, which seeks to promote the development of secure and trustworthy AI systems internationally and address the challenges associated with the technology.[36] In addition, the OECD had also published a paper on Generative Artificial Intelligence in Finance in December 2023, presenting the recent evolutions in generative AI and its slow-paced deployment in finance, highlighting the potential challengerisks from a wider use of generative AI tools by financial market participants, and discusses the potential policy implications.

Recommendations for regulatory approaches towards generative AI

As mentioned in our "Enabling an Efficient Regulatory Environment for AI" published in June 2021, we recommended that regulators take a principles-and risk-based approach to AI. The principles include supporting public-private collaboration, allowing FIs to take a risk-based approach to manage AI-related risks, taking materiality of the use case and stakeholders into account, leveraging existing regulatory frameworks, striving for regional and international harmonisation, promoting, and facilitating cross-border data flow, and engaging with the industry on areas that need further discussion.

Generative AI can potentially be used in a whole range of functions across financial markets, to augment existing activities, to replace them, or to perform complex and intensive tasks that were not previously feasible.

[30] TC260, 'Basic security requirements for generative artificial intelligence service', October 2023, https://www.tc260.org.cn/upload/2023-10-11/1697008495851003865.pdf
[31] Ministry of Science and ICT, 'South Korea presents a new digital world order to the world!', September 2023, https://www.msit.go.kr/eng/bbs/view.do?sCode=eng&mId=4&mPid=2&pageIndex=&bbsSeqNo=42&nttSeqNo=878&searchOpt=ALL&searchTxt=
[32] SFC, 'Reflect, Reset and Refocus: Game Plan for Hong Kong as an Asset Management Hub', June 2023, https://www.sfc.hk/-/media/EN/files/COM/Speech/HKIFA-Keynote---Eng_20230605.pdf?rev=e6f1d40a460049f8a3db13c4a0d34ac4&hash=9D7136A93B2E75DAC7DA336B0F3B1B7B
[33] Australian Government: Department of Industry, Science and Resources, 'Responsible AI in Australia: have your say', June 2023, https://www.industry.gov.au/news/responsible-ai-australia-have-your-say
[34] Singapore Government, 'Singapore National AI Strategy 2.0', December 2023, https://file.go.gov.sg/nais2023.pdf
[35] European Parliament, 'Artificial Intelligence Act: deal on comprehensive rules for trustworthy AI', December 2023, https://www.europarl.europa.eu/news/en/press-room/20231206IPR15699/artificial-intelligence-act-deal-on-comprehensive-rules-for-trustworthy-ai
[36] Ministry of Foreign Affairs Japan, 'G7 Leaders' Statement on the Hiroshima AI Process', October 2023, https://www.mofa.go.jp/files/100573466.pdf
[37] OECD, 'General artificial Intelligence in Finance'. December 2023, https://www.oecd-ilibrary.org/docserver/ac7149cc-en.pdf?expires=1704356849&id=id&accname=guest&checksum=394C7A5213BF1D9F34ADA3B42FEA8F37

Each use case will have its own risk profile and key stakeholders. FIs should be allowed to adopt a risk-based approach that is proportionate to the risk of the particular use case and its potential impact on stakeholders and measured against the performance of comparable current processes (if existing) or an available human-powered alternative.

We continue to promote this approach and put forward the additional recommendations and nuances for a balanced regulatory and industry approach for generative AI:

Principle 1

Leverage existing regulations

The development and use of AI and generative AI are covered by a myriad of existing regulations in the financial sector in the areas of cybersecurity, governance, third-party management, privacy etc. In addition, regulators around the world have been providing guidance – and in some cases, drafting regulation – around the use of AI. When assessing the incremental challenges of generative AI, we urge regulators to leverage these existing frameworks and mitigants where appropriate, conduct a gap analysis to identify any challenging areas pertaining to the use of generative AI that are not addressed by the existing regulatory framework, and focus any additional regulatory guidance on addressing these gaps.

As identified in section 7.C above, the importance of robust third-party risk management will continue to grow with the adoption of generative AI. Regulators and FIs should leverage existing outsourcing, operational resilience and technology risk management guidelines and emerging frameworks on critical third parties. Whilst the ultimate responsibility for the use of third-party developed generative AI tools will lie with the regulated FIs, regulators are exploring how the existing frameworks might have to be amended or expanded to address challenges across the (generative) AI lifecycle. As mentioned, accountability of the various controls in the generative AI lifecycle will depend on the deployment model, and we suggest that responsibility (and liability) should lie with the party who has control over the specific element of the lifecycle. We suggest regulators, developers and FIs come together to agree on an appropriate and enforceable shared responsibility model for generative AI. ASIFMA and its members look forward to inputting into this important policy debate.

Principle 2

Maintain governance and accountability

Organisations must establish a comprehensive governance framework that reflects legal, compliance and ethical considerations, and transparency to harness the governance and accountability pertaining to the use of generative AI. We recognise that many organisations have already invested in these frameworks to manage the challenges of "traditional" AI, whether as new structures or within existing frameworks. Hence, we suggest that such frameworks should be validated against generative AI specific challenges to validate coverage, but do not need rebuilding from scratch.

In line with the Securities Industry and Financial Markets Association (SIFMA) response to the National Telecommunications and Information Administration request for public comment on AI system accountability measures and policies, we suggest that a risk-based AI governance framework should include the following components:[37]

1. **Scoping:** Companies should determine which AI applications are in scope of the framework when building their governance programs.
2. **Inventory:** Companies should prepare and maintain an inventory of their AI applications with sufficient detail to allow them to be risk rated.
3. **Risk Rating:** Companies should have a process for identifying their highest-risk AI applications. The risks considered would include legal and regulatory risks, including operational, reputational, contractual, discrimination, cybersecurity, privacy, consumer harm, lack of transparency, and confidentiality risks.

[37] SIFMA, 'Response to Request for Comments on AI Accountability Measures and Policies', June 2023, https://www.sifma.org/wp-content/uploads/2023/06/Response-to-Request-for-Comments-on-AI-Accountability-

4. **Responsible Persons or Committees:** Companies should designate one or more individuals or committees who are responsible for identifying and assessing their highest-risk AI applications, and either accepting those risks, mitigating them, or abandoning the particular AI application because the risks are too high.
5. **Training:** Companies should develop training programs to ensure that stakeholders are able to identify the challenges associated with their AI use and the various options for reducing risk.
6. **Documentation:** Companies should maintain documentation sufficient for an audit of the risk assessment program.
7. **Audit:** Companies should conduct periodic audits that focus on the effectiveness of the risk assessment program, rather than on individual AI applications. Companies should be permitted to determine how and when audits should be conducted, and who can conduct those audits.
8. **Third-Party Risk Management:** Companies should use the same risk-based principles that are applied to in-house AI applications to evaluate third-party AI applications, and mitigate those risks through diligence, audits, contractual terms, and internal testing. There should be referenced to adapting the risk-based principles to the context of third-party models and applications. (See Section '7B.3 Transparency, Explainability, Traceability' above)

This proposed framework could be incorporated into existing governance and compliance programs in related areas such as model risk, data governance, privacy, cybersecurity, vendor management, and product development, with further guidance from applicable sectoral regulators as needed. Further, having qualified persons identify, assess, and mitigate the challenges associated with the highest-risk AI uses improves accountability, appropriate resource allocation, and employee buy-in through clearly defined and fair processes.

Such risk-based approach would provide a valuable, flexible framework through which FIs and their sectoral regulators can build tailored AI governance and compliance programs that ensure accountability and trust without stifling innovation or wasting time or resources on low-risk generative AI applications.

These AI governance principles also apply to the use of generative AI. Robust governance and accountability frameworks are more crucial than ever given the accessibility and democratisation of generative AI, its potential widespread use, and the limited substitutability of the FMs.

Principle 3

Provide the appropriate level of transparency

Transparency in AI refers to the level and quality of disclosure provided regarding the application of AI in services and/or products, including the challenges that may be involved in AI usage. Explainability typically refers to the extent to which workings of a model can be understood. In line with our 2021 paper, we suggest the regulatory focus should be on transparency of generative AI models instead of explainability which will allow a firm to demonstrate how the AI application has been developed, how it will be used and monitored, and how it can stand up to scrutiny and challenge. Within these broad themes, transparency should meet the varied needs of individual types of stakeholders, both inside and outside the firm.

Trustworthy generative AI starts with transparency on two levels, to supplement our existing guidance: transparency from the developer towards the deployer, and transparency from the deployer towards the consumer.

Transparency from developers towards deployers to manage risks from AI-generated systems:

- Transparency on type of data sets used.
- Transparency around model development, limits, and testing.
- Transparency around the collection and use of personal data.
- Transparency on ownership of data and IP with a particular focus on firm-specific data sources rather than base LLMs.
- Clear guidelines are needed to determine the responsibility of platforms and content creators to label AI-generated content.

Transparency from deployers towards consumers:

- Labelling AI-generated content will allow consumers to make more informed decisions and is of particular relevance to generative AI as it will help inform future uses of this data, including by other generative AI models.

Principle 4

Continue to adopt a risk-based approach

Policymakers should adopt a governance framework for generative AI that is risk-based and not overly prescriptive. A generative AI governance framework should treat AI models, algorithms, applications, and systems (collectively, "AI applications") differently depending on the likelihood or severity of the potential harm they might cause. The concept of same business, same risks, same rules should continue to apply in the context of generative AI.

Such risk-based approach to generative AI governance provides the necessary flexibility to balance the potential challenges with the potential benefits and opportunities in deploying generative AI. There are several considerations associated with a risk-based governance framework for AI, including (1) which specific challenges a company should consider when deciding which AI applications are high-risk, (2) how best to mitigate the challenges associated with high-risk AI applications, and (3) which AI applications carry unacceptable challenges and should not be pursued. The granular determinations of such considerations are best made by the company's management, with guidance from its applicable sectoral regulators. Regulators therefore should primarily guide companies to focus their efforts on identifying their highest-risk AI uses, and on mitigating the challenges they present, as well as define what "high-risk" means for firms to focus on instead of bucketing "high-risk" use cases without clear criteria. Such determinations and governance will then already exist for other high-risk models that use traditional AI, or other forms of models.

When assessing the challenges and deciding on the appropriate control framework of a particular use case, firms should take into account factors such as the potential impact of its use on the firm's resilience, on financial stability, and on end-consumers, whether the generative AI application/use case uses public data or proprietary data, and whether it is an enterprise vs a consumer application.

Principle 5

Continue to adopt a technology-agnostic regulatory approach

We urge regulators to continue to adopt a technology-agnostic approach when assessing and addressing any incremental risks associated with generative AI. The technology is fast-evolving and is therefore important to avoid designing regulation based on a particular technology and to avoid directly or indirectly dictating the use of any one type of technology over another. Such technology-agnostic approach to regulation will accommodate future innovation without requiring regulatory reforms each time new technology is implemented and will allow the market to innovate and capture risk appropriately. In contrast, a technology-specific regulation may run the risk of subsuming technology used in traditional financial activities into incongruous regulatory perimeters. Moreover, existing regulations that only permit certain technologies to be used should be updated to accommodate the changing technology landscape.

Generative AI and FMs are in their early development stage and continue to evolve. There are no universally agreed definitions which we believe is appropriate given the evolving nature of the technology and to ensure regulations remain future proof. Any definitions of these concepts in any future regulation should be broad enough to be adaptable. Regulators should focus on regulating the input and output, rather than the technology itself and should not favour one technology over another.

[38] CourtListener, THALER v. PERLMUTTER, 1:22-cv-01564, (D.D.C.), August 2023, https://www.courtlistener.com/docket/63356475/thaler-v-perlmutter/Measures-and-Policies.pdf

Principle 6

Address any IP Protection challenges

FIs using generative AI are at risk of unintentional IP infringement given the limited control over the input and output of the FM. The existing legal framework may need be clarified, broadened, or strengthened to reflect the increased use of generative AI. For example, clarity is needed on whether fair use exceptions apply to training data.

It is an open debate as to whether current IP and copyright laws are still fit for purposes in the context of generative AI and its related content. This is both in terms of the input data, as well as the ownership of the content generated by AI. In August 2023, a US federal judge has ruled that works produced entirely by generative AI cannot be copyrighted.[38] Policymakers need to provide clarifications on copyright in their respective regulations.

There is also need for greater international alignment given the existing different approaches being taken. E.g. AI training requires the extraction of information from datasets which may contain copyright-protected works – which may be copyright infringement unless this falls under an exception (transient and temporary copying or text and datamining exceptions in the EU or fair use in the US. We also note that Japan currently takes the approach that datasets for training AI models do not violate copyright law. Singapore amended its Copyright Act to introduce a new exception to copyright infringement for use of works for text and data mining, which is expected to significantly increase the availability of training data for AI. In addition, China's recent proposed technical standard on generative AI devoted some attention to IP protection (e.g. operators to establish mechanisms to address complaints from IP owners). According to the guidance, the process should enable complainants to query the entire training corpus to learn whether it includes proprietary material etc. On the other hand, the United Kingdom did not follow this approach and the EU AI Act requires publication of copyright data used for training.

Principle 7

Strive for regulatory certainty and a harmonised framework

Leveraging the potential of generative AI and addressing the challenges demands a coordinated response. There is a need for collaboration between all stakeholders to achieve a harmonised approach which will ensure that challenges can be sensibly mitigated, and the full potential achieved. This includes close collaboration between regulators and FIs including through consultations before any new or amended regulations or regulatory guidance is issued. Likewise, regulators should promote consistency across jurisdictions through globally recognised standards. Financial services regulators should also consider collaboration with regulators in other areas of the economy. There is a risk that standard setting in the wider economy could impose undue constraints on financial services innovation and inclusion.

Evolving technology (and hence challenges) as well as evolving regulatory frameworks make it challenging for the financial services industry to adopt generative AI (given the investment it takes to research, develop, govern, and then and roll it out) and the appropriate governance structure to go with it. Therefore, we need a risk-based, transparent regulatory framework to allow adoption that should offer wider benefits in efficiency and inclusion. Regulatory fragmentation exacerbates the cost and therefore we need national and international coordination and alignment. Sectoral regulators overseeing AI should cooperate to ensure that firms are not subject to conflicting obligations. Any new specific AI regulations and guidelines should be applied consistently across bank and non-bank FIs to ensure that consumers remain protected wherever they choose to receive their financial services.

9

GLOSSARY OF ACRONYMS

AI	Artificial Intelligence
API	Application Programming Interface
ASIFMA	Asia Securities Industry and Financial Markets Association
CAC	Cybersecurity Administration of China
EU	European Union
FI	Financial Institution
FM	Foundation Model
GPT	Generative Pre-trained Transformer
IEC	International Electrotechnical Commission
IP	Intellectual Property
ISO	International Organisation for Standardisation
LLM	Large Language Model
MAS	Monetary Authority of Singapore
ML	Machine Learning
OECD	Organisation for Economic Co-operation and Development
RAI	Responsible Artificial Intelligence
SDLC	Software Development Lifecycle
SIFMA	Securities Industry and Financial Markets Association
US	United States