4 October 2024

Department of Industry, Science and Resources

By upload

Dear AI Team

Mandatory Guardrails for AI in High-risk Settings: Proposals Paper

AFMA welcomes the opportunity to comment on the proposed 'mandatory guardrails' for AI in high-risk settings.

AFMA's positioning

Broadly our position is Australia should aim to gain the benefits of recent developments in AI while managing the risks for vulnerable people(s).

The drafting should seek to avoid creating disproportionate challenges for firms seeking to innovate in good faith with a sophisticated new technology for the benefit of their clients in the wholesale space. Industry is unlikely to fully realise the benefits of the AI revolution in Australia unless there is a low-risk path for firms seeking to develop and deploy the technology.

In the event that AI cannot readily be deployed in Australia these financial market functions will likely be provided from jurisdictions with more accommodative regulatory settings.

Utilise existing frameworks

AFMA continues to support careful, targeted, and proportionate regulation aimed at minimising harms from AI in areas where these risks exist and are not already managed by existing risk management frameworks and practices.

Australian Financial Markets Association
ABN 69 793 968 987
Level 18, 45 Clarence Street GPO Box 3655 Sydney NSW 2001
Tel: +612 9776 7993 Email: secretariat@afma.com.au

This should be done using existing regulatory frameworks wherever possible through uplift and adjustments that can leverage existing compliance structures and we do not at this stage see a need for a separate broad-based AI Act.

We do not see the need for AI-specific legislation and in particular would not support an EU-style Bill but do support consistent treatment across the breadth of existing legislation.

Internationally compatible but not an overly restrictive regime

AFMA supports Australia having AI standards that are interoperable with international standards, use globally consistent terminology, leverage international regulatory outcomes, and that are careful to preserve and foster the potential efficiency gains that AI is enabling. However, this would not suggest following the lead of the restrictive jurisdictions. Excessive regulation in a small open economy risks creating barriers to trade, commerce, innovation and the growth of the economy.

We strongly support a risk-based approach, and in this regard the framework proposed by the Government could be calibrated to achieve these outcomes.

Industry plans to build on longstanding AI practices for the benefit of investors

We note that most AI is not new, and in financial services the risks and challenges are like those experienced before that have been successfully managed.

In GenAI, firms see significant opportunities to add value and improve their operations for clients' benefit. Firms are being careful and deliberate when deploying GenAI, and they have adapted and continue to adapt their internal governance to make sure the potential risks are appropriately addressed.

Summary

We encourage the Government to look to adopt a more pro-innovation approach as is the case in the UK and US rather than one more aligned with the restrictive EU and Canadian regulatory regimes. The more restrictive regulatory approaches that have influenced the proposed drafting could readily create significant impediments to the uptake of this critical technology.

We trust our responses are of assistance.

Yours sincerely

Damian Jeffree
Head of Financial Markets, Exchange and Digital

1. Do the proposed principles adequately capture high-risk AI? Are there any principles we should add or remove? Please identify any:
   • low-risk use cases that are unintentionally captured
   • categories of uses that should be treated separately, such as uses for defence or national security purposes.

Capturing high-risk AI

We welcome the opportunity to provide feedback to the Australian Government on this important topic. AFMA supports that the framework to separate out "high-risk" be built through a collaborative, robust stakeholder process. We note that the compressed timeframe for the current consultation from 5 September to 4 October compromises the potential to meet this objective.

We support policymakers focussing on high-risk and potential outcomes associated with deploying AI models and systems in specific contexts, while avoiding broad classifications of risk for entire sectors, categories of AI or uses of AI.

As the field matures internationally, alignment with key allied jurisdictions on what matters are in scope for high-risk AI will be important and to achieve this it is important that the framework locally retains sufficient flexibility and an international orientation. However, in the meantime Australia should not emulate restrictive regimes such as those found in the EU.

Firms best placed to assess AI risks

Ultimately, we believe that the organizations are responsible and best positioned to determine whether that specific AI use cases deployed in a specific context is high-risk or not.

The most important considerations for understanding the nature and degree of AI risk — and for distinguishing truly high-risk use cases for regulatory purposes — are the severity, scale and likelihood of potential harm to individuals or society, and whether the harmful impact could be effectively remediated or reversed.

Existing and modified existing regulations should be the primary restraint on AI in financial markets

AI systems are already constrained by a range of existing regulations we discuss in following sections.

The risks associated with a particular AI model or tool will be higher or lower depending on the specific use case. Where AI is the principal basis for making consequential decisions, assessments of the type described in the guardrails can be appropriate where the existing constraints are insufficient.

The paper proposes firms will be required to use the following principles (the Principles) to identify high-risk AI:

> In designating an AI system as high-risk due to its use, regard must be given to:
>
> a. The risk of adverse impacts to an individual's rights recognised in Australian human rights law without justification, in addition to Australia's international human rights law obligations
>
> b. The risk of adverse impacts to an individual's physical or mental health or safety
>
> c. The risk of adverse legal effects, defamation or similarly significant effects on an individual
>
> d. The risk of adverse impacts to groups of individuals or collective rights of cultural groups
>
> e. The risk of adverse impacts to the broader Australian economy, society, environment and rule of law
>
> f. The severity and extent of those adverse impacts outlined in principles (a) to (e) above.

AFMA supports the view that AI does have the potential to exacerbate existing risks, both in terms of speed and scale, we are also of the view that new issues may arise as the use of AI in financial services grows.

Large and potentially high-risk problem space for firms

As drafted, the Principles require firms to consider legal impacts, defamation or other 'significant effects' on individuals, as well as adverse effects on cultural groups, groups of individuals, the economy, society, the environment, the rule of law, and 'Australian human rights law'. It also imports and will create a legislative requirement for firms to consider potential impacts on 'Australia's international human rights law obligations'.

As *voluntary* guidelines these would be broadly helpful in ensuring firms have turned their minds to relevant categories of harm that should be considered.

However, as we understand the proposal, firms would be mandated under enforceable provisions to ensure they have identified any and all potential adverse outcomes. For example, principle (e) condenses many impacts into one principle and may require matters as diverse as the significant water and energy consumption associated with training large AI models to be considered. We query whether this breadth of considerations is intended or appropriate for each deployment of an AI system in the economy.

Depending on the enforcement approach adopted, firms could face significant penalties for failing to properly categorise a complex technology over a very large problem space. The alternative of working on

the basis that nearly all AI should be treated as high-risk could be cost prohibitive and may impede the utilization of the technology in Australia.

Much will depend on the details of the rules and enforcement regime to avoid this requirement creating a high-risk environment for innovation.

The importation of Australia's international human rights law obligations *in toto* into a domestic legislated and enforceable requirement for firms to evaluate when developing and deploying AI products without intervening domestic legislation appears novel and warrants careful review and consideration.

Principle (f) is listed in the Principles as a separate and additional consideration on top of those from (a) to (e). If it is the intention of the Government to limit the categorisation of AI to high-risk then it would be appropriate to (1) move (f) to the top as a qualifier for all the subsequent principles and (2) include drafting to make it a clear materiality qualifier. The current drafting 'severity and extent' does not function as a materiality qualifier and may be inconsistent with standard risk management drafting. All risks have a potential severity and extent even if they are both *de minimis*.

AFMA also holds that firms should be able to take into account technical indicators when assessing the risks of AI.

Inappropriate test for the large stock of existing AI systems and similar systems

AI systems with underlying technologies such as logistic regression are mature and well handled by the existing internal model risk management processes. More generally, additional examples of systems which would not be considered AI and which might be excluded from the high-risk use case category due to their level of simplicity would be of assistance.

Ancillary systems

We would welcome more clarity about the regulatory treatment towards outputs from ancillary systems that contribute to an AI automated decision, as we believe that they should not classify as high-risk.

While these systems might entail some risks from output errors such as wrong data presentation, they have no direct influence on decision-making outcomes but instead are involved in more narrow procedural tasks, are used for monitoring and reporting to ensure compliance and system integrity, or are used to perform a preparatory task for the purpose of then making a decision on a final outcome. For example, the evaluation of the creditworthiness of an individual and the process to establish their credit score both involve several ancillary systems. These could include data preparation, data insights, documentation analysis and monitoring and reporting. While the use of AI to make the final

decision/assessment could be captured under the definition of high-risk, the outputs from ancillary systems that contribute to an overall credit score or assessment of credit worthiness should be excluded.

2. Do you have any suggestions for how the principles could better capture harms to First Nations people, communities and Country?

We have no specific comments but note work that assists banks implement Fairness, Ethics, Accountability and Transparency done by the Veritas Consortium, led by the Monetary Authority of Singapore (MAS). The toolkit is designed to increase understanding of emerging risks across different use cases; the need for internationally compatible risk-based approach to determine appropriate governance for responsible AI and the creation of more appropriate guidance and oversight. We note also 'Project MindForge' which is currently assessing the applicability of the Assessment Methodologies for Generative AI (GenAI).

3. Do the proposed principles, supported by examples, give enough clarity and certainty on high-risk AI settings and high-risk AI models? Is a more defined approach, with a list of illustrative uses, needed?
   • If you prefer a list-based approach (similar to the EU and Canada), what use cases should we include? How can this list capture emerging uses of AI?
   • If you prefer a principles-based approach, what should we address in guidance to give the greatest clarity?

AFMA supports a principles-based approach

AFMA strongly prefers a principles-based approach, rather than the list-based approach of the EU and Canada as it is more sustainable and simpler to understand and comply with.

A principles-based approach to the regulation of emerging technologies such as AI allows organizations to take a risk-based approach to managing any new or evolving risks and adapt existing risk management frameworks accordingly. This encourages responsible innovation and ensures that consumers are protected irrespective of the technology that is used.

Whether a use case is high-risk will inevitably be fact-specific and will depend on the user, how the specific AI is being deployed and for what purpose. As the AI use cases alongside the related technology evolve, an explicit list of illustrative cases would not be efficient as it would need to be constantly updated. Instead, introducing a list of principles that would capture high-risk AI would ensure a more "future-proof" approach.

AI risks are shaped by many factors, including how the AI application is used; the environment in which it is deployed; the type of data processed or used to create models and tools; interactions with other AI systems; and the user characteristics, such as level of experience with or training on AI. The approach

adopted by financial institutions (or any entities) to address AI risk is context-specific and depends on their policies, risk appetite and governance structure.

A principles-based approach to regulation should allow financial institutions to ensure all relevant risks are effectively addressed while providing banks with the flexibility to structure their internal processes in accordance with their business model, risk profile and other characteristics. Existing risk management frameworks, such as model risk and third-party risk management frameworks, are designed to be able to adapt to changes in technology and business models, including those stemming from emerging technologies such as AI and generative AI.

We note that illustrative use cases can complement a principles-based approach in aiding understanding and compliance with the requirements.

Criteria for deeming low-risk can also assist

If a system does not meet the criteria for a high-risk system then it should be by default a low-risk system. However, it still may be useful a principles-based to set out criteria for AI systems which will automatically deem them low risk, see Art 6(3) of the EU AI Act. For example, it performs a narrow procedural task, intended to improve results of human output, detects patterns from prior decisions, preparatory task for high-risk use cases.

Sectoral regulators can bring sectoral knowledge to principles

Selecting sectoral regulators as the designated authority may provide sectoral knowledge on how to interpret high level principles within their sectors. Banking regulators in Hong Kong and Singapore, for example, are already starting to take this approach (for e.g. HKMA circular on GenAI or MAS FEAT principles notes customer-facing applications to be a matter of concern). It also allows institutions to fit AI into existing compliance structures and relationships.

4. Are there high-risk use cases that government should consider banning in its regulatory response (for example, where there is an unacceptable level of risk)? If so, how should we define these?

Our focus is solely on the wholesale financial markets. In this context we are not currently aware of any use cases that should be banned.

5. Are the proposed principles flexible enough to capture new and emerging forms of high-risk AI, such as general-purpose AI (GPAI)?

The current definition of GPAI may be non-optimal. The proposed definition may unintentionally apply to a wide range of statistical techniques, many of which are not new. The definition needs to be made more specific to select the appropriate technologies.

The European Parliament amendments to the AI Act propose a definition for GPAIS as "an AI system that can be used in and adapted to a wide range of applications for which it was not intentionally and specifically designed" (EP 2023 Amendment 169 Article 3 paragraph 1 point 1d), a definition for the more capable and larger-scale subset of GPAIS identified as foundation models as "an AI system model that is trained on broad data at scale, is designed for generality of output, and can be adapted to a wide range of distinctive tasks" (EP 2023 Amendment 169 Article 3 paragraph 1 point 1c).

6. Should mandatory guardrails apply to all GPAI models?

AFMA does not support mandatory guardrails to all GPAI models

Mandatory guardrails should only apply to high-risk AI cases not to all GPAI models.

GPAI, while general-purpose, does not automatically equate to high-risk.

The risk argument presented in the consultation paper for deeming all GPA high-risk is: "Given GPAI models pose unforeseeable risks, the Australian Government proposes to apply mandatory guardrails to all GPAI models".

AFMA holds this claim may not be sufficient to support the proposed response and that the case is yet to be made for deeming all GPA high-risk.

GPAI is the one of most important developments in the field with the potential and a blanket high-risk categorisation goes directly against a risk-based approach as it does not meet the minimum risk analysis required to fit a risk-based characterisation.

International alignment need not mean overly restrictive settings

The argument based on international consistency given in the paper is: "Since most highly capable GPAI models are not currently developed domestically, Australia's alignment with other international jurisdictions is important to reduce the compliance burden for both industry and government and enables pro-innovation regulatory settings."

AFMA supports international alignment but care must be taken to align with jurisdictions that find the right balance that enables and supports innovation. Alignment with restrictive jurisdictions could create a substantial barrier to domestic development of GPAI models. International alignment should be developed gradually and carefully to ensure competitiveness in the Australian economy.

Risk considerations for AI systems

Please also refer to our comments on the importance of using risk analysis in answer to question 3, these also apply to GenAI risk assessment.

7. What are suitable indicators for defining GPAI models as high-risk? For example, is it enough to define GPAI as high-risk against the principles, or should it be based on technical capability such as FLOPS (e.g. 10^25 or 10^26 threshold), advice from a scientific panel, government or other indicators?

We do not recommend defining GPAI as high-risk based solely on FLOPS. Instead, a risk-based and principles-based approach should be adopted to assess the risk level of specific use case.

Higher FLOPS usage does not necessarily correlate with higher risk. For example, AI models for creating presentation slides or videos may require substantial computational power, thus using more FLOPS, but typically pose lower risks. The focus should be on evaluating the potential impact and consequences of each application rather than the computational resources it consumes. For systems near the range of any FLOPS cut-off there could be incentives for developers to create models and computing clusters that fall just below the established thresholds.

The pace of AI innovation combined with our evolving understanding of the risks and harms to the community means even the most well-intentioned regulatory approach is likely to lag industry. As such, AFMA would prefer a partnership-based model approach is adopted that focuses on enabling the technology to be used safely.

To this end, articulating, maintaining and updating specific AI in-principle outcomes the government wishes to avoid and high-risk use-cases which could enable such outcomes is a sensible approach rather than attempting to place any risk weighting on the technology mechanisms themselves.

This will involve industry working with the sectoral regulators to implement more general principles, such as those put forward by the UK Government, and make them applicable to each sector. For example, a fairness principle might mean very different things in the context of a financial market versus a social media company. In the case of financial markets there is an established set of norms that can and should be leveraged in a consistent manner for AI.

8. Do the proposed mandatory guardrails appropriately mitigate the risks of AI used in high-risk settings? Are there any guardrails that we should add or remove?

AFMA would welcome further consultation with sectoral regulators as AI policies are implemented.

A review cycle for any policies or guardrails with appropriate frequency should be put in place, given the rapid advancements and change in this area.

With regard to the Guardrails listed, we note the following concerns:

Guardrail 1:

> Organisations must also make their accountability processes publicly available and accessible to improve public confidence in AI products and services.

We do not object to firms being required to have accountability processes but these are internal matters for firms and their regulator. If public statements are required these should be high-level only.

Guardrail 2:

> Deployers will be responsible for following instructions for use set by developers and managing risks specific to the use case.

There appears to be an assumption that developers can foresee the deployment risks better than deployers. Deployers often have direct sight of the risks of a particular deployment. Deployers will have their own risk management approach that should not be unable to override those of the developers who will often have no visibility or involvement in the final deployment.

Guardrail 3:

> "data source must be disclosed"

This should only be for developers and again only at a high level. If a custom data set is used this is proprietary information and firms would not allow use in jurisdictions that required disclosure.

While we would support public policy measures that would facilitate voluntary data sharing for additional use cases, this would have to be developed and executed very carefully. There are several risks from cross-industry data sharing, including negative impacts to competition, a potential increase in data leaks and cybersecurity breaches and data manipulation. At the same time, while AFMA supports in principle free flow of data, we note that data localisation requirements have been increasing globally; as a result, voluntary data sharing might subsequently increase regulatory risk for entities.

Therefore, any data sharing initiative would have to be carefully designed to ensure that it did not inadvertently increase the likelihood of such risks occurring. If developing any public policy measures, it would be helpful if these would explicitly confirm that such data sharing initiatives would not breach applicable data privacy laws.

In order to ensure a clear distinction between the terms “Transparency” and “Explainability”, we suggest adopting clear definitions of the two terms. Consideration could be given to existing literature such as the ASIFMA’s 2024 Gen AI Paper which defines the terms as follows:

a. 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.

b. Explainability typically refers to the extent to which workings of a model can be understood.

At the same time work could continue with industry and academia on how to best address concerns over the explainability of AI models. The Government could also draw reference from a number of robust explainability techniques in use and Bank Policy Institute (BPI) has offered a comprehensive outline of how banks identify and address issues around explainability, and ensure that the outputs of AI models meet their organizational tolerances and regulatory requirements.

We caution that this is an evolving technical field and recent research using Sparse Auto-Encoders (SAE’s) suggests that multiple encoding of concepts is happening on single nodes simultaneously across multiple eigenbases. The concepts being encoded may not map to ones preferred by humans, and their meaning may be inherently difficult to explain.

AFMA has led work to suggest that Governments should allow firms to use the same risk management tools they have long used with employees.

These typically include; checking some structure training has been successfully completed (in addition to unstructured training), testing candidates, background checks, monitoring and consequence management. Incorporating and adapting these types of methods that have long been used for employees would appear a sensible and lower risk path forward.

Guardrail 5:

“Enable human control or intervention in an AI system to achieve meaningful human oversight.”

This guardrail states that organisations must ensure that humans can effectively understand a high-risk AI system. This guardrail must be appropriately described to allow for complex systems, as these systems may not be readily understood.

Most operators of modern computing systems do not understand their internal operations yet can use them safely unsupervised.

Human control and intervention in real-time systems can be counterproductive to output quality and other factors, for example crash avoidance systems can override erroneous input quicker than human reaction times. In other systems real time meaningful human oversight may not scale efficiently and may compromise privacy and other factors.

There are risks of adopting an overly risk averse approach with this guideline depending on its final form and implementation.

Guardrail 8:

> "Be transparent with other organisations across the AI supply chain about data, models and systems to help them effectively address risks."

Guardrail 8 requires developers to provide substantial documentation. This requirement needs to take into consideration the willingness of the developer to share what is often proprietary information, for example the requirement that data sources must be disclosed. A deployer can become a developer if the deployer undertakes any activities to develop, customise, refine or enhance externally built models, such as fine-tuning a pre-trained AI LLM or applying retrieval augmented generation (RAG). It is debateable how sustainable and helpful it will be to maintain a system for deployers to provide detailed information back to the developers. A single LLM may have tens or even hundreds of millions of users globally. A requirement to contact developers outside of their structured feedback and beta testing programs may be counterproductive.

We suggest this Guardrail be reconsidered in the context of these concerns.

9. How can the guardrails incorporate First Nations knowledge and cultural protocols to ensure AI systems are culturally appropriate and preserve ICIP?

See our response to Question 2.

10. Do the proposed mandatory guardrails distribute responsibility across the AI supply chain and throughout the AI lifecycle appropriately? For example, are the requirements assigned to developers and deployers appropriate?

AFMA is supportive of ensuring that AI guardrails distribute responsibility across the AI supply chain and lifecycle, thereby ensuring that developers and deployers share accountability for managing AI-related risks on a proportionate basis. However, it is again important to note that existing risk management frameworks and practices are adequate for addressing and adapting to any emerging risks and challenges presented by AI. Leveraging established third-party risk management (TPRM) frameworks, operational resilience practices and other risk management standards will effectively mitigate risks throughout the AI lifecycle and through contractual arrangements which ensure that risk-management practices and regulatory obligations cascade down the entire supply chain.

It is therefore important to ensure that any proposed 'mandatory guardrails' will avoid an undue governance burden and take a principles-based approach that will allow organizations to tailor their supply chain oversight accordingly. A more specific GPAI definition, and refined approach into applying the guardrails at scale, may assist in reducing friction.

Guardrail 8 establishes guidance aimed at addressing challenges associated with the explainability of models and transparency across the supply chain. It is important to ensure distinction between these concepts, specifically between transparency in terms of the model risk context, and transparency in the broader context of third-party risk management.

Transparency in the model risk context concerns the clarity and interpretability of the model's inner workings and decision-making processes and implementing robust model validation processes and testing procedures in respect of third-party models can be challenging. Transparency in the context of third-party risk management, on the other hand, relates to the ability for a bank to manage the risks of the arrangement, including the existing or planned usage of AI, and banks' ability to do due diligence on a third-party's control environment, their compliance with regulations and adherence to contractual obligations.

This will be reliant on the ability for third parties to explain why, when and how AI is being used, information about data inputs and outputs, and with what governance and risk management measures. This challenge extends to having sight of potential vulnerabilities and compliance issues across the supply chain to ensure firms are appropriately protected from the risks that AI introduces.

Third party models

It is not uncommon that banks rely on third-party models. Assessing the explainability or fairness of third-party AI models presents additional challenges as such third parties, which are typically not subject to regulations of the same depth and breadth as banks, can refuse to disclose their proprietary information (including training data and information relating to the operation of their algorithms), making it challenging for banks to fully evaluate the explainability or fairness of these solutions. Generative AI models provided by third parties present further complexity with respect to explainability, as these models are predominately built on very large and typically largely unstructured datasets.

This is not inherently undesirable as long as more structured training data overlays provide executive level awareness within the model of how to avoid bias etc., and indeed extremely large data sets have been, along with large increase in compute, critical in the latest leap in AI capabilities.

Such information could include, among others, the type and timeframe of data used to train the system, limitations of the data or system, appropriate and inappropriate use cases, what the system does with data presented to it (e.g., whether input data is used for secondary purposes), as well as other technical information helpful to evaluating the model and other key risks. Larger and more complex third-party

ecosystems introduce added challenges with obtaining visibility into potential vulnerabilities and compliance issues across the supply chain.

Contractual agreements could be used to ensure third parties notify banks of their use of AI and to provide the necessary contractual protections regarding the use of bank data. However, the ability to secure necessary or optimal contractual obligations and protections may be dependent on a firm's negotiating power with large global providers.

11. Are the proposed mandatory guardrails sufficient to address the risks of GPAI? How could we adapt the guardrails for different GPAI models, for example low-risk and high-risk GPAI models?

We would suggest the focus should be on the risk of using an agreed industry framework of the application of the AI, rather than the model risk in isolation.

12. Do you have suggestions for reducing the regulatory burden on small-to-medium sized businesses applying guardrails?

AFMA supports interoperability and global standards but only if done with jurisdictions that are supportive of innovation. Alignment of Australia's AI regulatory regime with the more restrictive EU and Canadian approaches will increase costs for small-to-medium businesses and make Australia a less attractive venue for the development of AI.

Small and medium businesses that are resource-constrained risk being penalised or even locked out of AI if the regulatory settings are not right-sized.

Any 'mandatory guardrails' should be designed such that the regulatory burden is commensurate with the risk and is manageable, regardless of the size of the institution. Exempting businesses based on their size does not seem appropriate.

Adaption of Australia's existing legal and regulatory frameworks will be the best address of the use of AI in high-risk settings.

Artificial intelligence is an established technology utilized by the financial services industry. New advancements, such as GenAI and PredAI, have led to increased focus on the potential opportunities of use cases, for example, to serve clients directly or indirectly. As AI technology evolves and new use cases continue to develop, it is integral that a technology neutral, principles-based, and outcomes-focused approach is prioritized.

Australian authorities can apply and adapt existing standards and frameworks where applicable, rather than create new AI-specific standards that could lead to conflicts of law for technology solutions

implemented by financial services firms or create undue costs and burden for implementing and monitoring AI use cases.

Existing standards have proven effective, while remaining technology-neutral, and promote outcomes-based regulation. Should gaps in existing standards be identified as new AI use cases gain prominence, from a financial stability perspective, standard setters should explore whether it would be sufficient to update existing governance frameworks or if new guidance may be necessary to fill in any gaps. After such analysis, if these options are insufficient, only then should new standards be considered, provided that they complement existing processes and procedures for technological innovations.

Policymakers should align legislative and regulatory proposals with existing, effective domestic and international policies and industry risk management strategies to promote a harmonized approach and avoid introducing uncertainty and conflicting compliance requirements.

13. Which legislative option do you feel will best address the use of AI in high-risk settings? What opportunities should the government take into account in considering each approach?

Any binding regulation of AI should apply across industries and to the entire economy and public space. This can also be achieved by making targeted amendments to existing non-AI-related laws and regulations. The goal should be to have a technology-neutral approach to a regulatory framework that is results-oriented and principles-based. The consequence of this approach would also be that there is no need to define AI or prohibited / high-risk uses. Many of the types of risks presented by AI applications are common to the use of technology in general. While Generative AI at scale is a relatively emerging technology, AI itself is not really new: narrow AI and machine learning have long histories in the financial services industry and are already subject to existing regulations and risk frameworks. Many sophisticated analytical systems and machine learning engines have been successfully in operation for years, including in credit risk scoring and high-frequency trading. However, the new models have novel characteristics that need to be better understood and, in time, may warrant targeted adjustments to the existing regulatory framework.

There is no need for financial market-specific AI regulation: As indicated, AI applications in finance are already subject to regulation through sectoral or cross-sectoral regulations that are technology-neutral and apply to the use of any general-purpose technology, like AI. Examples include:

- financial services laws and regulations such as the APRA prudential standards which impose a range of obligations on APRA-regulated entities, including with respect to data, governance, outsourcing and operational risks generally. It is worth noting there are a range of safety, security and transparency obligations already placed on businesses operating within other sectors too (e.g. motor vehicles, airlines and medical devices);
- cross-sectoral laws and regulations, including:

- privacy laws which place obligations on businesses handling personal information, including when developing and using AI;
- intellectual property laws which place a range of obligations on businesses, including with respect to copyright, to assist in the protection of AI systems;
- online safety laws which are intended to address, minimize and prevent harm relating to illegal and restricted material online, including in the AI space; and
- competition and consumer protection laws which would, amongst other things, assist in protecting consumers against unfair commercial / lending practices and terms, including in the AI space.

- anti-discrimination guardrails are also implicitly woven through Australia's legal system. For example, all government and non-government bills must contain a statement of compatibility that the bill or legislative instrument in question is compatible with the rights and freedoms recognized in the seven core international human rights treaties which Australia has ratified.

If proven necessary, targeted regulatory adjustments could be considered. This may include, for example, supplementing the regulations on the transparency of AI systems for all market participants, further enhancing data protection laws, adapting product liability law and general civil law (e.g., consumer protection, bias and discrimination, etc.) in order to create the necessary regulation for the use of AI systems. However, these adjustments should as far as possible be made in a technology-neutral and principles-based manner rather than being AI specific and overly prescriptive. Under existing rules, APRA regulated entities (such as banks and insurers), operate under prudential requirements that require them to develop their own sophisticated risk management governance frameworks, systems, and controls. Financial institutions keep such arrangements under constant review and make adjustments as needed, including proactively. This assists in driving innovation efforts in a safe and compliant manner. The industry starts from a very strong risk management foundation – one that is commensurate with and contributes to the high trust placed in financial institutions by their clients.

To address the risks that AI entails, the adoption of AI legislation (Option 3 in the consultation document) based on the European model is disproportionate at this point in time, in particular with regards to international competitiveness. It would also introduce an additional level of complexity and potential duplication with Australia's existing regulatory framework. As indicated above, the targeted adaption of existing regulatory frameworks based on a gap analysis is our preferred approach. The gap analysis will need to show whether a domain-specific approach with a targeted adaptation of existing regulatory frameworks, including financial market law (Option 1), or, a framework approach (Option 2), are best suited to meet the goals of a technology neutral and principles-based approach that avoids creating unnecessary sectoral distinctions.

AFMA supports authorities maintaining a technology-neutral, risk-based and outcomes-focused approach to the regulation of AI. This will best ensure a responsible approach to AI development and deployment

by finding the optimal balance between addressing risks and encouraging innovation. Australia should follow international standards as broad consensus emerges.

The types of risks identified in financial services are not unique to AI. These can be addressed via existing regulation (both cross-industry, such as data privacy or consumer protection, and FS-specific) and financial services risks and controls frameworks. Organisations may need to modify existing processes, policies and procedures to adapt existing frameworks to meet the risks posed by AI, but we believe that the current regulatory framework does not provide any barriers to the responsible development and deployment of AI. We therefore suggest that the Government adopts a domain-specific approach and introduces guardrails within existing regulatory frameworks as needed.

This would reduce the risk of regulatory duplication, or worst still, conflict, ensure better coordination and mitigate the risk of regulatory siloes. Regulatory fragmentation is neither new nor specific to AI but it poses similar challenges, risks and costs, from inhibiting innovation and competition to posing risks to operational resilience.

14. Are there any additional limitations of options outlined in this section which the Australian Government should consider?

Particularly in the financial services industry, the Australian Government should consider the complexity of the existing regulatory landscape. We hold that AI is best regulated by existing regulators rather than introducing any new regulatory bodies, and that the approach should be aligned with existing governance approaches in financial services which are mature, and generally considered to be successful.

15. Which regulatory option/s will best ensure that guardrails for high-risk AI can adapt and respond to step-changes in technology?

AFMA supports authorities maintaining a technology-neutral, risk-based and outcomes-focused approach to the regulation of AI. This will best ensure a responsible approach to AI development and deployment by finding the optimal balance between addressing risks and encouraging innovation. As far as possible, this should be addressed at the global level to maximise regulatory alignment/harmonisation.

As a result, we recommend that the Australian Government considers that both the technology, industry usage and government's concerns will change and should build any regulatory approach to be adaptable.

16. Where do you see the greatest risks of gaps or inconsistencies with Australia's existing laws for the development and deployment of AI? Which regulatory option best addresses this, and why?

Specifically for the financial services industry, there is a history of applying overlapping and duplicative regulatory requirements. To avoid this in relation to AI sectoral regulators should be working to implement the same principles, and where appropriate across related sectors, e.g. banking and finance, using the same rules.