4 October 2024

Anthony Murfett
Head of Technology and Digital
Department of Industry, Science and Resources
GPO Box 2013,
Canberra, ACT, 2601
Via online submission portal.

Dear Mr Murfett,

RE: Consultation on Proposals Paper for introducing mandatory guardrails for AI in high-risk settings.

The FSC welcomes the opportunity to make a submission to the Department of Industry, Science, and Resources consultation on the mandatory guardrails for AI in high-risk settings.

The FSC is supportive of a risk and principles-based approach to governing AI use in Australia with only the most obviously harmful AI products outright banned. Australia's approach to AI governance should utilise a pro-AI approach which encourages innovation in Australia whilst maintaining important consumer protections. This in turn will encourage more and more consumers to interact with the technology and build trust in its use cases.

The FSC supports Option 2, outlined in the Proposals Paper, which would create framework legislation without creating an onerous extra layer of regulation through an enforceable legislative framework but without requiring individual regulatory regimes to update for the AI framework individually.

Clear guidance is required to ensure that there is consistent application of the guardrails across industry and between parties within the AI lifecycle, but care should be taken to always strike a balance between consumer protection and innovation.

Answers to specific questions are outlined below.

Summary of Recommendations

1. The FSC is supportive of developing principles but is concerned that as currently drafted, the principles may be too broad and will unintentionally capture all use cases. Principles should point to factors which indicate whether a use case is high risk. Government should provide non-prescriptive guidance about the application and leave room for industry to provide self-regulatory mechanisms in the interpretation of the principles.
2. Clear, but non-prescriptive guidance regarding the weighting of the principles and how they should be applied is required to ensure consistent coverage across use-cases.

3. Further clarity is required as to the application of the first Principle relating to the adverse impact on human rights.

4. Guidance should include clear definitions of deployers and developers and how the interaction of this relationship should work in relation to both parties' responsibilities in the application of the principles and guardrails.

5. The proposed framework should include clear triggers for review, encompassing both time passed and technological advancement, to ensure that they remain fit for purpose.

6. Non-prescriptive guidance is required to provide clarity of obligations between deployers and developers where one is a foreign entity.

7. Further guidance needs to be provided for entities who are both the deployer and the developer.

8. The Principles and Guardrails should apply by use-case and not to specific models of AI. This would mean that not all models are theoretically captured but the application of Guardrails tempered by the use case.

9. In determining suitable indicators for defining high-risk models, weight should be given to the use-case, rather than the specific model.

10. Non-prescriptive guidance is required in relation to how Australia intends to define the consideration of human rights within the Guardrails, given the lack of clear legislation in this area.

11. Clarity is required as to how a person would be expected to go about requesting access to personal information a company may have collected in relation to GPAI training data if they are not aware that it has been collected.

12. Guidance is required regarding the interaction of proposed Privacy Law reforms, including the right to be forgotten, and the proposed Guardrails.

13. Further clarity is needed to clearly define the threshold for when informing to a client is required.

14. Guidance is required in relation to how the process of contesting an AI decision would work in practice, particularly in relation to how an entity who has developed a product might address such a challenge.

15. Guidance is required to understand how third parties deploying AI through software are captured under the Guardrails where they are not the primary developer or deployer.

16. Non-prescriptive guidance, perhaps in the form of templates, would be useful for all businesses, including SMEs, to understand how best to communicate with end-users under the Guardrail obligations.

17. Option 2 – Framework Legislation is the most appropriate legislative approach to regulating AI in Australia. However, framework legislation should be designed with a light touch, focusing on principles rather than prescriptive rules, and with ample room for industry self-regulation.

Financial Services Council Ltd
Suite 30.01, Level 30, 9 Castlereagh Street
Sydney NSW 2000

E. info@fsc.org.au
W. fsc.org.au
ABN 82 080 744 163

About the Financial Services Council

The FSC is a peak body which sets mandatory Standards and develops policy for more than 100 member companies in one of Australia's largest industry sectors, financial services. Our Full Members represent Australia's retail and wholesale funds management businesses, superannuation funds, and financial advice licensees.

The financial services industry is responsible for investing more than $3 trillion on behalf of over 15.6 million Australians. The pool of funds under management is larger than Australia's GDP and the capitalisation of the Australian Securities Exchange and is one of the largest pools of managed funds in the world.

AI in Financial Services

AI has myriad uses in the Australian financial services context and many organisations are considering ways to responsibly incorporate AI into their systems and processes. Use cases seen in other jurisdictions around the world include assessing financial risks, crafting investment portfolios, and reducing and preventing fraud and combating financial crime. Other, non-financial services specific uses include helping manage customer service and improving efficiencies within the workplace.

Equally, FSC members understand that AI presents a risk to their businesses, particularly in relation to fraud, scams, and cybersecurity. The sophistication of AI forgeries make it even harder for organisations like superannuation funds to check and confirm the identity of their members and FSC members are keen to see appropriate risk-based mitigation measures put in place in this regard.

FSC members are fully alive to the risks associated both internally and externally of using AI within their organisations. The FSC is therefore overall supportive of the work of government to provide certainty for industry with regard to AI use, encouraging innovation in the future and providing organisations with appropriate risk-based guidelines for the use of AI technology in Australia.

Proposals Paper Question Responses

Question 1. Do the proposed principles adequately capture high-risk AI? Are there any principles we should add or remove?

The FSC supports the development of AI principles, however, there are concerns that the broad nature of the principles may lead to all use-cases for AI being captured, leading to the potential for innovation to be stifled.

The FSC is supportive of a principles-based approach to AI regulation that point to factors which indicate whether a use case may be high-risk. The provision of robust principles for considering whether a use case is high-risk should aim to provide certainty for industry as they look to implement solutions within their organisations.

That said, the interpretation of the principles is what is most important and due to their current broad nature, the FSC is concerned that they do not provide certainty and instead will capture more than intended. The FSC encourages Government to provide non-prescriptive guidance about the application of the principles to narrow the scope.

Equally though, industry should be allowed to show leadership in relation to the deploying of AI in Australia. The FSC encourages Government to also leave room for self-regulatory mechanisms in the interpretation of the principles, to allow for industry leadership and expertise to support innovation.

Financial Services Council Ltd
Suite 30.01, Level 30, 9 Castlereagh Street
Sydney NSW 2000
E. info@fsc.org.au
W. fsc.org.au
ABN 82 080 744 163

RECOMMENDATION 1

The FSC is supportive of developing principles but is concerned that as currently drafted, the principles may be too broad and will unintentionally capture all use cases. Principles should point to factors which indicate whether a use case is high risk. Government should provide non-prescriptive guidance about the application and leave room for industry to provide self-regulatory mechanisms in the interpretation of the principles.

Question 2. 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?

The FSC is supportive of a principles-based approach to the regulation of AI in Australia. This is because striking the right balance between consumer safety and encouraging innovation is paramount to making AI use a success in the Australian market. The FSC does not support a more defined approach which would mean that the regime is both rigid and not future proofed.

Notwithstanding the answer to question 1, it is important that guidance is provided as to the intention of the application of these principles so that they are applied consistently between use cases. There is a high risk of unintended consequences in having broad based principles with little guidance on how they might apply.

Despite the examples provided, more guidance is necessary on how the principles should be applied in practice. The paper states that '[t]he principles should be considered as a whole when assessing whether the use of the AI system is high-risk.' However, it is not clear how many of the principles should apply before the use case is considered high-risk, or if only a single of the principles need to be met. For instance, even if all principles potentially apply to a use case, but the extent and severity of those adverse impacts is low, should the use case be deemed low or medium risk and therefore not subject to the mandatory guardrails?

Further, the principles themselves are broad, for example, the risk of adverse impacts to an individual's physical or mental health or safety may implicitly (or explicitly) include a person's financial security which would capture a wide range of use cases in the financial services sector without reference to whether this is the Government's intention.

Guidance is necessary to ensure clarity and consistency in application across industry without becoming so prescriptive that it stifles innovation.

RECOMMENDATION 2

Clear, but non-prescriptive guidance regarding the weighting of the principles and how they should be applied is required to ensure consistent coverage across use-cases.

Notwithstanding the above, the FSC believes the first principle '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' requires significant further clarity.

This principle gives rise to the application of several different treaties and agreements that Australia is party to, making it potentially onerous to reliably apply. The purpose of the Principles and associated guardrails should be to provide certainty and clarity to both

consumers and organisations looking to offer AI based products and services. Therefore the application of the principles should be accessible.

This principle in particular will require non-prescriptive guidance regarding its application.

RECOMMENDATION 3

Further clarity is required as to the application of the first Principle relating to the adverse impact on human rights.

There also needs to be clearer guidance about the definition of deployers and developers and the transparency required in the relationship between the two as it pertains to the application of both the principles and the guardrails.

Clear expectations should be applied to both parties to avoid scenarios where the deployer has assumed the developer has applied the correct principles and vice-versa, including what the implications would be for a deployer if they relied on representations from the developer about its processes.

Further, consideration needs to be given to how a deployer might ascertain the governance processes of the developer, especially where they are not captured by the Australian framework or another jurisdiction that has a robust framework for disclosure, such as the EU.

RECOMMENDATION 4

Guidance should include clear definitions of deployers and developers and how the interaction of this relationship should work in relation to both parties' responsibilities in the application of the principles and guardrails.

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

Whilst the broad nature of the proposed principles will enable the capture of many new and emerging forms of high-risk AI, the potential impacts from future advances in AI, including GPAI, are unknown, and as a result it is unrealistic to expect current regulation to effectively capture all new forms of AI-based capability.

Hence, the regulation should include clear triggers for review, including both timelines as well as potential technical advancement triggers such as the development of viable GPAI capability. Early review of these regulations will enable Australia to ensure balance in the protection of the rights of its citizens, as well as investment and the economy.

RECOMMENDATION 5

The proposed framework should include clear triggers for review, encompassing both time passed and technological advancement, to ensure that they remain fit for purpose.

Further, as the Principles and Guardrails appear to be silent on their application to foreign versus domestic entities, it is reasonable to assume that this means that the Guardrails will apply regardless of domicile. This would appear, in practice, to be unworkable and places entities in a confusing position as to how the obligations would apply to them if they were the deployer of a product developed overseas. In some instances this would be two completely

different organisations but, in some instances, including within the FSC membership, they may be connected entities but where one is an Australian owned subsidiary.

Non-prescriptive guidance should be provided to entities to help them understand the obligations between deployer and developer where one party is a foreign entity.

RECOMMENDATION 6

Non-prescriptive guidance is required to provide clarity of obligations between deployers and developers where one is a foreign entity.

Additionally, further guidance is required in relation to the governance arrangements where an entity is both the deployer and the developer. For example, it is not clear whether there is a need for segregation between teams working on development or deployment of AI, or if further risk mitigation measures are needed in this situation as compared to where the developer and deployer are separate entities.

RECOMMENDATION 7

Further guidance needs to be provided for entities who are both the deployer and the developer.

Question 6. Should mandatory guardrails apply to all GPAI models?

The question as to whether AI is high-risk should apply to the use-case and not the model itself. This means that when considering whether something is high-risk and therefore the guardrails would apply, an organisation would have to consider not just which model was used but how it will be used. This provides both better security and flexibility because not all GPAI models should be considered high-risk, but the question as to risk should also capture AI generated code which has the potential to be used maliciously. Consideration should also be given to whether a model is open source, given the lack of control of the model.

RECOMMENDATION 8

The Principles and Guardrails should apply by use-case and not to specific models of AI. This would mean that not all models are theoretically captured but the application of Guardrails tempered by the use case.

Question 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?

The indicators for high-risk should be based on the use case, rather than on the specific technology of the model. In the case of GPAI, the use cases to which it can be applied would be the primary trigger.

In particular, consideration should be given to the degree of autonomous decision-making provided to a GPAI model.

RECOMMENDATION 9

In determining suitable indicators for defining high-risk models, weight should be given to the use-case, rather than the specific model.

Question 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?

The FSC largely is supportive of the proposed guardrails. Specific commentary is below.

Guardrail 1

The application of the Guardrails, with regard to the impact on human rights, is unclear given the mixed architecture of Australia's human rights laws (such as international treaties and various pieces of legislation such as anti-discrimination legislation) and warrants further guidance. Such additional clarity could also include areas such as impacts to democratic rights and other freedoms not necessarily codified in Australian law.

RECOMMENDATION 10

Non-prescriptive guidance is required in relation to how Australia intends to define the consideration of human rights within the Guardrails, given the lack of clear legislation in this area.

Guardrail 3

Guardrail 3 poses some significant questions in relation to jurisdictional issues. An example is a scenario where an international organisation uses publicly available personal information of Australians to train its GPAI. It is not clear if this act is covered by the Privacy Act 1988 as they are not likely carrying on a business in Australia. Even if a business has an Australian connection, individuals are unlikely to have been notified of the collection of use of their information under APP 5, and practically would not know their information was being used by the GPAI model.

Clarity is required as to how a person would be expected to go about requesting access to personal information a company may have collected in relation to GPAI training data if they are not aware that it has been collected.

RECOMMENDATION 11

Clarity is required as to how a person would be expected to go about requesting access to personal information a company may have collected in relation to GPAI training data if they are not aware that it has been collected.

Further, given the significant amount of proposed privacy reform, clarity is required as to how the proposed right to be forgotten would interact with GPAI models which have been trained on data that an individual has requested that entity no longer holds.

RECOMMENDATION 12

Guidance is required regarding the interaction of proposed Privacy Law reforms, including the right to be forgotten, and the proposed Guardrails.

Guardrail 6

Further clarity is needed to clearly define the threshold for when informing a client is required. Consideration should be given around aspects like AI-augmented human interaction (e.g. AI translation both for language and reading-level purposes), or the use of AI-generated content that has been human-reviewed.

RECOMMENDATION 13

Further clarity is needed to clearly define the threshold for when informing to a client is required.

Guardrail 7

The FSC notes Guardrail 7 requires establishing processes for people impacted by AI systems to challenge use or outcomes. Further non-prescriptive guidance is needed to outline how such a challenge can be implemented in practice, including in circumstances where AI may be used in the design of a product. In such circumstances, it would be very difficult for an entity who has developed a product with the use of AI to address such a challenge and it is unclear from the position paper what kind of action may be required in response to such a challenge.

RECOMMENDATION 14

Guidance is required in relation to how the process of contesting an AI decision would work in practice, particularly in relation to how an entity who has developed a product might address such a challenge.

Question 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?

As mentioned, clear guidelines should be given about the delineation between the role of the deployer and the developer and consideration as to the transparency required between the two. The proposed definition of 'deployer' includes internal use cases, indicating that an organisation might become a deployer if it used an AI function within a software package off the shelf.

A clearer definition of deployer is also required to understand how third or fourth parties and other outsourced providers are captured in the framework. For example, if an organisation purchases a piece of software off the shelf that will have AI use cases, clarity is required as to what a third party's responsibility would be as neither the developer nor the primary deployer.

Additionally, it is important to have clarity particularly on the accountability around transparency across parties. If, for example, a third party creates content in a setting that is not high-risk, and therefore does not disclose that to a financial services firm who unknowingly uses it in a high-risk setting, requiring disclosure, where does that accountability lie?

RECOMMENDATION 15

Guidance is required to understand how third parties deploying AI through software are captured under the Guardrails where they are not the primary developer or deployer.

Financial Services Council Ltd
Suite 30.01, Level 30, 9 Castlereagh Street
Sydney NSW 2000

E. info@fsc.org.au
W. fsc.org.au
ABN 82 080 744 163

Question 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?

The models themselves are unlikely to be the determinant of risk, but rather the use case to which they are applied.

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

The Guardrails require that end-users of high-risk AI use cases be informed about how they have been impacted by the AI. The FSC submits that guidance would be useful to assist all businesses, including SMEs about how best to approach this. This could be in the form a customisable (but not prescriptive template) or general guidance.

RECOMMENDATION 16

Non-prescriptive guidance, perhaps in the form of templates, would be useful for all businesses, including SMEs, to understand how best to communicate with end-users under the Guardrail obligations.

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

Australia's approach to AI regulation should focus on balancing the potential for consumer harm with the benefits that new technology brings to both industry and consumers alike.

Although there are several international examples of regulatory frameworks, it is important that Australia focus on a specific framework that works for its economic circumstances. While international cooperation is valuable, Australia should prioritise developing AI policies tailored to its unique economic and social context, rather than importing potentially restrictive overseas policies.

Legislation should be flexible and adaptive and not create an extra layer of complicated regulatory and enforcement burden or be so prescriptive that it is essentially out of date immediately after publishing.

For this reason, the FSC is supportive of the approach outlined in Option 2, requiring an overarching legislative framework that recognises existing industry specific legislation but does not necessarily create a separate onerous enforcement regime.

The financial services industry is already serviced by a number of strict regulatory and prudential rules and is overseen by several enforcement agencies, including APRA and ASIC. AI legislation should be able to slot in reasonably seamlessly into the existing frameworks and enforcement capabilities.

Proposed option 2 strikes the best balance between ensuring a consistent approach to AI regulation across the economy while having deference to the individual needs of each industry. Any option undertaken would need a clear timeframe for review given the pace at which innovation in this sector is occurring.

The proposed Option 2 allows for the regulatory framework around AI to change in a consistent way without creating overlapping and confusing layers of regulation and legislation.

RECOMMENDATION 17

Option 2 – Framework Legislation is the most appropriate legislative approach to regulating AI in Australia. However, framework legislation should be designed with a light touch, focusing on principles rather than prescriptive rules, and with ample room for industry self-regulation.

If you have any questions about this submission, please do not hesitate to contact me.

Yours sincerely,

Kirsten Samuels
Policy Director, Superannuation and Innovation

Financial Services Council Ltd
Suite 30.01, Level 30, 9 Castlereagh Street
Sydney NSW 2000

E. info@fsc.org.au
W. fsc.org.au
ABN 82 080 744 163