FSB FINANCIAL STABILITY BOARD

# Sound Practices for Responsible Adoption of Artificial Intelligence (AI): Consultation report

## Response to Consultation


## Shankar S (independent response)


1. **Do you agree with the benefits and risks of AI adoption by financial institutions described in this report? Are there any substantive benefits or risks not covered?**

2. **Are the sound practices sufficiently comprehensive and clear to enable financial institutions' responsible AI adoption?**

   A menu of practices enables adoption. It does not yet define when reliance is warranted.

   The twelve sound practices are individually sound and collectively well organised. The report describes them as "a menu of sound practices that financial institutions could use to enable responsible AI adoption" and is explicit that they are "not intended to establish an international standard" (p. 1). Within that self-description sits the one structural gap this submission asks the FSB to consider before the report is finalised. Each sound practice describes what a financial institution should have or should do: a governance framework, a risk assessment process, data governance, oversight, monitoring. These are governance inputs. None of the twelve - individually or together - defines the condition under which the institution's reliance on a consequential AI-enabled decision or action becomes defensible: the point at which the institution can show, rather than assert, that governance operated when the decision was generated or the action was taken.

   The report comes closest in two places, and both illustrate the gap rather than close it. Section 3.2 treats "the feasibility of human oversight and level of reliance on the AI use case's outputs" as factors in materiality and risk assessment (p. 29) - reliance appears as an input to be scored, not as a state to be warranted. Section 3.7 requires oversight that is "meaningful", involving "sufficient ability, authority, and incentive to intervene, as opposed to oversight that is nominal or driven by 'tick-the-box' compliance" (p. 41). That sentence names precisely the failure mode that a menu of practices, without a threshold, cannot prevent: an institution can adopt every practice in Annex 1 - as documented intent - and still be unable to show that governance was in effect for any specific decision or action. The distinction that matters is between governance intent and Governance in Effect. Governance intent means what the institution meant to apply and what the institution

approved and documented. Governance in Effect means governance that is shown to have operated in practice as decisions were generated, actions were taken and consequences emerged.

This is not an argument against proportionality. The report is right that "more robust practices may be appropriate" for institutions that are large, complex and highly connected (p. 1). But proportionality describes how much governance to apply, and it presupposes a floor beneath which no proportionate reduction reaches. The standard the report needs is a threshold, not a ceiling: the minimum that must be demonstrably true before reliance on a consequential AI-enabled decision or action is warranted - not an ideal of governance perfection. A threshold gives proportionality something to be proportionate to.

Proposed refinement. The final report should acknowledge the distinction between practices adopted and governance demonstrably in effect at the point of decision or action, and should encourage financial institutions to define, for each material AI use case, the demonstrable basis on which the institution relies on that use case's outputs - the evidence that would allow it to stand behind a specific decision or action if tested by a supervisor, a counterparty or an affected customer.

3. **Do the sound practices strike an appropriate balance between managing risks relating to all forms of AI, and addressing some of the risks relating to emerging and new complex forms of AI, such as GenAI and agentic AI?**

The balance across forms of AI is broadly right. The criterion for oversight is missing - and it is the criterion that scales.

The report's treatment of GenAI and agentic AI is practical and current. The additional considerations at pp. 43-44 - boundaries on agent actions, individual identifiers for agents, red-teaming of agent behaviour, controls on transaction execution, monitoring of intermediate steps - are the right ground. The concern this submission raises is therefore not the balance between forms of AI. It is that Sound Practice 10's oversight taxonomy lacks the one criterion that holds across all forms and becomes more load-bearing as autonomy increases: whether the oversight is independent of the system it oversees.

Consider where the report's own examples locate the trigger for human intervention:

- 'human-on-the-loop' intervention occurs "when the AI's outputs have a low confidence score" (p. 42);

- in the illustrative customer-service deployment, "human agents [are] alerted to intervene only when sentiment analysis detects frustration" (p. 42);

- footnote 72 contemplates agentic systems escalating to humans when accuracy metrics are breached or when users request it (p. 42); and

- at scale, "monitoring may need to be continuous or performed by AI" (p. 44).

In each case, the trigger for human involvement, the information channel available to the human, or both, are produced by the system under oversight - or by another AI system standing in for it. A confidence score is the system's testimony about itself. Sentiment analysis performed by the deployment being overseen is self-referential in the same way. Oversight of this kind fails exactly when it is most needed, because a system that is confidently wrong will not summon its own supervisor. The report already names the underlying misconception - the assumption "that human involvement ensures effective oversight" (p. 14) - but the taxonomy does not operationalise the safeguard against it.

The condition this submission draws on from Warranted AI - Oversight Independence - states the criterion directly: reliance requires that the system could be observed, challenged, interrupted or overridden through a capability that does not depend on the system's own self-assessment. The Warranted AI paper states the test in assessable form (Chapter 4): whether 'the trigger for oversight attention, the information available to the overseer and the capability to interrupt or override are each independent of the system being overseen'. Elements of that criterion already appear in the report. Allocating post-deployment monitoring to staff who were not part of design and development "to promote objectivity" (p. 43), red-teaming and the analysis of human oversight patterns for rubber-stamping and for overrides suggesting pervasive misalignment (p. 43) are all independence mechanisms. What is absent is the criterion itself, stated once and applied to every form in the taxonomy. Nor is the criterion specific to operational oversight: it applies at every governance altitude - up to the information on which boards themselves rely when overseeing AI adoption - though this submission confines itself to the taxonomy at p. 42.

The criterion also matters at the level of the FSB's mandate, not only at the level of the individual institution. The report identifies "reliance on common AI models, datasets, or infrastructure" as a source of "correlated behaviours across financial institutions" (p. 16). The same correlation operates one layer up. Where many institutions deploy oversight whose triggers are the model's own self-assessment - confidence scores, self-reported metrics, model-generated alerts - a miscalibration in a widely used foundation model degrades not only outputs across institutions but every institution's capacity to notice, simultaneously. Independent oversight is what prevents oversight failure from scaling with adoption. That makes oversight independence a financial-stability property, not only a use-case control.

Proposed amendment. The supporting text to Sound Practice 10 should direct financial institutions, for each form of human oversight selected, to assess whether (a) the trigger for human intervention, (b) the information available to the human overseer, and (c) the capability to interrupt or override are each independent of the AI model or system being overseen. Where they are not, the institution should treat that oversight arrangement as part of the system rather than as oversight of it and compensate accordingly. The graduated forms at p. 42 can remain unchanged; the assessment applies across all of them and remains applicable as newer forms emerge.

4. **Are the sound practices sufficiently flexible to accommodate and address newer types of AI and responsible adoption over time?**

5. **Do the case studies in this report sufficiently highlight how different types of financial institutions can benefit from responsible AI adoption? Are there additional case studies for inclusion in the report? If so, please provide such case studies, particularly for nonbanks.**

6. **Do the case studies in this report provide actionable insights for financial institutions in their responsible AI adoption?**

7. **Are the definitions in the glossary clear and aligned with industry sound practices, including recent developments in AI?**

   Two observations. First, the glossary does not define human oversight, although it is the operative term of Sound Practice 10, and the forms it takes - 'human-in-the-loop', 'AI-in-the-loop', 'human-on-the-loop', 'human-in-command', 'kill switch' and 'contestability' - are defined only in the body text at p. 42. A glossary entry would be a natural home for the criterion proposed above: human oversight defined not by the presence of a human, but by the presence of a capability to observe, challenge, interrupt or override that does not depend on the system's own self-assessment.

   Second, the definition of agentic AI - "systems designed to autonomously perform complex and extended tasks, often making decisions and taking actions with limited human oversight" (p. 56) - is right to define agency by decisions and actions together. Governance frameworks that reach only decisions will not reach the acting layer of agentic systems. The phrase "limited human oversight" would carry more weight if the glossary made human oversight itself assessable, per the first observation.

8. **Are there any comments you would like to make on this report? Please fill in.**

   About this submission. Shankar Siva is a governance and risk adviser and the author of Warranted AI: Seven Conditions for Defensible Reliance on Consequential AI-Enabled Decisions and Actions in Regulated Industries (Working Paper, July 2026), SSRN Abstract ID 7137118, available at https://ssrn.com/abstract=7137118 (in processing).

Warranted AI defines a Warranting Standard for organisational reliance on consequential AI-enabled decisions and actions. It is a threshold test for when reliance can be justified, defended and stood behind. It is not an AI ethics framework, a model explainability method or a compliance checklist. Warranted AI is not proof of legality. It is the evidentiary condition for defensible organisational reliance. The examples and implementation observations in this submission draw on direct experience working with APRA-regulated entities navigating governance challenges in operational settings.

The Warranted AI paper engages this consultation report in its regulatory-alignment analysis. This submission builds on and does not repeat the author's June 2026 submission to the Office of the Australian Information Commissioner on automated decision-making transparency (Shankar Siva, Submission to the Office of the Australian Information Commissioner, Automated Decision-Making Transparency Obligation (APP 1) - Issues Paper, 11 June 2026), where Oversight Independence appeared under its earlier name, Supervisory Independence; the substance is unchanged.

This submission addresses Questions 2, 3 and 7. Questions 1, 4, 5 and 6 are not addressed. It proposes one refinement to the final report's framing of the sound practices (Question 2), one amendment to Sound Practice 10 applying the Oversight Independence condition from the paper (Question 3) and two observations on the glossary (Question 7).

A formatted copy of this response will be provided by email to fsb@fsb.org as supplementary material for the public record.

**Shankar Siva | Governance and risk adviser | Author, Warranted AI**

# Submission to the Financial Stability Board

## Sound Practices for Responsible Adoption of Artificial Intelligence (AI) - Consultation Report (10 June 2026)

21 July 2026

Submitted via: FSB online consultation form

---

**About this submission**

Shankar Siva is a governance and risk adviser and the author of *Warranted AI: Seven Conditions for Defensible Reliance on Consequential AI-Enabled Decisions and Actions in Regulated Industries* (Working Paper, July 2026), SSRN Abstract ID 7137118, available at https://ssrn.com/abstract=7137118 (in processing).

Warranted AI defines a Warranting Standard for organisational reliance on consequential AI-enabled decisions and actions. It is a threshold test for when reliance can be justified, defended and stood behind. It is not an AI ethics framework, a model explainability method or a compliance checklist. Warranted AI is not proof of legality. It is the evidentiary condition for defensible organisational reliance. The examples and implementation observations in this submission draw on direct experience working with APRA-regulated entities navigating governance challenges in operational settings.

The Warranted AI paper engages this consultation report in its regulatory-alignment analysis. This submission builds on and does not repeat the author's June 2026 submission to the Office of the Australian Information Commissioner on automated decision-making transparency [1], where Oversight Independence appeared under its earlier name, Supervisory Independence; the substance is unchanged.

This submission addresses Questions 2, 3 and 7. Questions 1, 4, 5 and 6 are not addressed. It proposes one refinement to the final report's framing of the sound practices (Question 2), one amendment to Sound Practice 10 applying the Oversight Independence condition from the paper (Question 3) and two observations on the glossary (Question 7).

---

[1] Shankar Siva, Submission to the Office of the Australian Information Commissioner, Automated Decision-Making Transparency Obligation (APP 1) - Issues Paper, 11 June 2026.

**Question 2 - Are the sound practices sufficiently comprehensive and clear to enable financial institutions' responsible AI adoption?**

***A menu of practices enables adoption. It does not yet define when reliance is warranted.***

The twelve sound practices are individually sound and collectively well organised. The report describes them as "a menu of sound practices that financial institutions could use to enable responsible AI adoption" and is explicit that they are "not intended to establish an international standard" (p. 1). Within that self-description sits the one structural gap this submission asks the FSB to consider before the report is finalised. Each sound practice describes what a financial institution should have or should do: a governance framework, a risk assessment process, data governance, oversight, monitoring. These are governance inputs. None of the twelve - individually or together - defines the condition under which the institution's reliance on a consequential AI-enabled decision or action becomes defensible: the point at which the institution can show, rather than assert, that governance operated when the decision was generated or the action was taken.

The report comes closest in two places, and both illustrate the gap rather than close it. Section 3.2 treats "the feasibility of human oversight and level of reliance on the AI use case's outputs" as factors in materiality and risk assessment (p. 29) - reliance appears as an input to be scored, not as a state to be warranted. Section 3.7 requires oversight that is "meaningful", involving "sufficient ability, authority, and incentive to intervene, as opposed to oversight that is nominal or driven by 'tick-the-box' compliance" (p. 41). That sentence names precisely the failure mode that a menu of practices, without a threshold, cannot prevent: an institution can adopt every practice in Annex 1 - as documented intent - and still be unable to show that governance was in effect for any specific decision or action. The distinction that matters is between governance intent and Governance in Effect. Governance intent means what the institution meant to apply and what the institution approved and documented. Governance in Effect means governance that is shown to have operated in practice as decisions were generated, actions were taken and consequences emerged.

This is not an argument against proportionality. The report is right that "more robust practices may be appropriate" for institutions that are large, complex and highly connected (p. 1). But proportionality describes how much governance to apply, and it presupposes a floor beneath which no proportionate reduction reaches. The standard the report needs is a threshold, not a ceiling: the minimum that must be demonstrably true before reliance on a consequential AI-enabled decision or action is warranted - not an ideal of governance perfection. A threshold gives proportionality something to be proportionate to.

**Proposed refinement.** The final report should acknowledge the distinction between practices adopted and governance demonstrably in effect at the point of decision or action, and should encourage financial institutions to define, for each material AI use case, the demonstrable basis on which the institution relies on that use case's outputs - the evidence that would allow it to stand behind a specific decision or action if tested by a supervisor, a counterparty or an affected customer.

**Question 3 - Do the sound practices strike an appropriate balance between managing risks relating to all forms of AI, and addressing some of the risks relating to emerging and new complex forms of AI, such as GenAI and agentic AI?**

***The balance across forms of AI is broadly right. The criterion for oversight is missing - and it is the criterion that scales.***

The report's treatment of GenAI and agentic AI is practical and current. The additional considerations at pp. 43-44 - boundaries on agent actions, individual identifiers for agents, red-teaming of agent behaviour, controls on transaction execution, monitoring of intermediate steps - are the right ground. The concern this submission raises is therefore not the balance between forms of AI. It is that Sound Practice 10's oversight taxonomy lacks the one criterion that holds across all forms and becomes more load-bearing as autonomy increases: whether the oversight is independent of the system it oversees.

Consider where the report's own examples locate the trigger for human intervention:

- 'human-on-the-loop' intervention occurs "when the AI's outputs have a low confidence score" (p. 42);
- in the illustrative customer-service deployment, "human agents [are] alerted to intervene only when sentiment analysis detects frustration" (p. 42);
- footnote 72 contemplates agentic systems escalating to humans when accuracy metrics are breached or when users request it (p. 42); and
- at scale, "monitoring may need to be continuous or performed by AI" (p. 44).

In each case, the trigger for human involvement, the information channel available to the human, or both, are produced by the system under oversight - or by another AI system standing in for it. A confidence score is the system's testimony about itself. Sentiment analysis performed by the deployment being overseen is self-referential in the same way. Oversight of this kind fails exactly when it is most needed, because a system that is confidently wrong will not summon its own supervisor. The report already names the underlying misconception - the assumption "that human involvement ensures effective oversight" (p. 14) - but the taxonomy does not operationalise the safeguard against it.

The condition this submission draws on from Warranted AI - Oversight Independence - states the criterion directly: reliance requires that the system could be observed, challenged, interrupted or overridden through a capability that does not depend on the system's own self-assessment. The Warranted AI paper states the test in assessable form (Chapter 4): whether 'the trigger for oversight attention, the information available to the overseer and the capability to interrupt or override are each independent of the system being overseen'. Elements of that criterion already appear in the report. Allocating post-deployment monitoring to staff who were not part of design and development "to promote objectivity" (p. 43), red-teaming and the analysis of human oversight patterns for rubber-stamping and for overrides suggesting pervasive misalignment (p. 43) are all independence mechanisms. What is absent is the criterion itself, stated once and applied to every form in the taxonomy. Nor is the criterion specific to operational oversight: it applies at every governance altitude - up to the information on which boards themselves rely when overseeing AI adoption - though this submission confines itself to the taxonomy at p. 42.

The criterion also matters at the level of the FSB's mandate, not only at the level of the individual institution. The report identifies "reliance on common AI models, datasets, or infrastructure" as a source of "correlated behaviours across financial institutions" (p. 16). The same correlation operates one layer up. Where many institutions deploy oversight whose triggers are the model's own self-assessment - confidence scores, self-reported metrics, model-generated alerts - a miscalibration in a widely used foundation model degrades not only outputs across institutions but every institution's capacity to notice, simultaneously. Independent oversight is what prevents oversight failure from scaling with adoption. That makes oversight independence a financial-stability property, not only a use-case control.

**Proposed amendment.** The supporting text to Sound Practice 10 should direct financial institutions, for each form of human oversight selected, to assess whether (a) the trigger for human intervention, (b) the information available to the human overseer, and (c) the capability to interrupt or override are each independent of the AI model or system being overseen. Where they are not, the institution should treat that oversight arrangement as part of the system rather than as oversight of it and compensate accordingly. The graduated forms at p. 42 can remain unchanged; the assessment applies across all of them and remains applicable as newer forms emerge.

---

**Question 7 - Are the definitions in the glossary clear and aligned with industry sound practices, including recent developments in AI?**

Two observations. First, the glossary does not define human oversight, although it is the operative term of Sound Practice 10, and the forms it takes - 'human-in-the-loop', 'AI-in-the-loop', 'human-on-the-loop', 'human-in-command', 'kill switch' and 'contestability' - are defined only in the body text at p. 42. A glossary entry would be a natural home for the criterion proposed above: human oversight defined not by the presence of a human, but by the presence of a capability to observe, challenge, interrupt or override that does not depend on the system's own self-assessment.

Second, the definition of agentic AI - "systems designed to autonomously perform complex and extended tasks, often making decisions and taking actions with limited human oversight" (p. 56) - is right to define agency by decisions and actions together. Governance frameworks that reach only decisions will not reach the acting layer of agentic systems. The phrase "limited human oversight" would carry more weight if the glossary made human oversight itself assessable, per the first observation.

---

**Shankar Siva**
Governance and risk adviser
Author, *Warranted AI: Seven Conditions for Defensible Reliance on Consequential AI-Enabled Decisions and Actions in Regulated Industries*