FSB FINANCIAL STABILITY BOARD

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

Response to Consultation

Sergio F (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?**

I agree with the taxonomy in Section 1.3 and with the four vulnerabilities carried over from the 2024 report. I would add one risk that is missing and adjust the framing of a second. Both come from the same theme, which runs through all of my answers: the report treats AI risk as a property of an institution, and for internationally active banks a large part of AI risk is a property of a jurisdiction. I develop that point in my answer to Question 8.

A missing risk: correlated failure in financial crime detection.

Section 1.3 discusses homogeneous behaviours and correlated outcomes arising from reliance on common AI models, datasets, or infrastructure. Every consequence it lists is a markets consequence: herding, procyclicality, market stress, liquidity crunches, asset price vulnerabilities. Those are real. There is a second channel with a different shape, and the report does not name it.

Screening and transaction monitoring in most banks is not built in house. It is licensed from a small number of vendors, and those vendors calibrate against reference data that reflects some jurisdictions far better than others. An institution that buys the capability also buys the vendor's blind spots. If a vendor's name matching handles compound Spanish surnames poorly, then every institution using that vendor handles them poorly, in the same way, at the same time.

I use that example because it is the one I have watched. Across Latin America a person ordinarily carries a paternal and a maternal surname, the two are transposed between systems often enough to be routine, and accents and particles survive in one database and vanish in another. None of this is exotic where I work. It is the default. In the reference data most vendors calibrate on, it is an edge case, and it is treated as one.

What separates this from market herding is that the failure is silent. Herding shows up in prices. A monitoring model that systematically fails to alert on a class of activity produces no signal at all, because an absent alert looks exactly like an absent problem. Institutions measure the false positives they can see. Nobody measures the false negatives they never

generated. The failure surfaces years later, through an enforcement action, or through a case that reaches the bank from law enforcement rather than the other way round.

I would suggest the report name this alongside the market channel: concentration among AI providers can produce correlated detection failure in the controls that are supposed to protect the integrity of the system, and that class of failure is harder to spot than the market class because there is no price to look at.

A framing to adjust: geographic concentration is treated as a continuity problem.

Section 1.3 says there can be geographic concentration across key points in the AI supply chain, which could be affected by physical events like natural disasters. That treats geography as a business continuity question. There is a second sense that matters more. Providers concentrated in a few jurisdictions build models on data generated in those jurisdictions. Losing the provider is one risk. The other is that the provider's model was never very good in your market, and that every competitor's provider has the same weakness. Concentration in that sense does not threaten availability. It quietly standardises a gap.

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

Three points.

1. Region is documented in Sound Practice 3 and then never used again.

Sound Practice 3 asks institutions to document, among other attributes of a use case, the business lines, regions, products, customers, and market participants impacted. The report is right to include regions there.

Nothing downstream uses the field. Sound Practice 9 says institutions specify thresholds across all relevant dimensions of performance, and lists accuracy, range of output, robustness, stability, and sensitivity or stress testing. Each is stated at the level of the use case as a whole. Sound Practice 7 lists what an effective data governance framework should address, and none of those considerations is jurisdictional.

This matters because of how internationally active groups actually build. A model is developed centrally, validated centrally against pooled data, and deployed across the group. Aggregate accuracy clears the threshold. Inside that aggregate, the two or three largest markets carry the number, and performance in a market representing two per cent of group exposure can be poor without ever moving it. That market is two per cent of the group. It is one hundred per cent of the local legal entity, whose board carries local legal accountability for the outcome and which usually did not build the model, cannot inspect it, and has no realistic way to challenge it.

Suggestion: Sound Practice 9 should say that where a use case has been documented under Sound Practice 3 as affecting several regions, performance should be assessed along those same regions as far as the data allows, and that thresholds should be set by reference to materiality at the entity that bears the consequence, not only materiality to the group. Sound Practice 7 could make the parallel point. Data quality should be assessed per jurisdiction, because it varies per jurisdiction, and an institution measuring it only in

aggregate will not know which of its markets is being carried by the others. This is a cheap change to make. The report already collects the field.

2. The model risk anchor in Sound Practice 5 has just moved.

Sound Practice 5 tells institutions they can build their AI materiality and risk assessment on existing regulation or guidance on model risk management, and cites several jurisdictions.

In the United States, that anchor changed on 17 April 2026. The Federal Reserve, the OCC and the FDIC issued revised interagency guidance on model risk management, published as SR 26-2, OCC Bulletin 2026-13 and FDIC FIL-15-2026. It rescinded and replaced the 2011 guidance, and it also rescinded the 2021 interagency statement on model risk management for bank systems supporting Bank Secrecy Act and anti money laundering compliance. Three features are relevant here. It narrows the definition of a model. It states that generative AI and agentic AI models are not within its scope. And the agencies said they intend to issue a request for information addressing model risk management and banks' use of AI, including generative and agentic AI.

So an institution in the largest FSB jurisdiction that follows Sound Practice 5 and reaches for its model risk framework will find that framework silent on precisely the AI types this report spends the most effort on. It is silent by design, pending a consultation that has not yet opened.

I do not think the FSB should try to resolve that. But the report should say it out loud. As drafted, Sound Practice 5 implies the model risk anchor is both available and stable. In at least one major jurisdiction it is currently neither. A sentence acknowledging that model risk frameworks are under revision in some jurisdictions, and that generative and agentic AI presently sit outside those frameworks in some of them, would tell an institution something it needs to hear: the distance between "AI is out of scope of our model risk guidance" and "AI is ungoverned" is theirs to close, and for the moment this report is the best instrument they have for closing it.

There is a timing point worth flagging too. The final report is expected in October 2026. If the United States request for information runs on any normal schedule, the two will overlap. Some coordination between them would spare institutions that have to answer to both.

3. Proportionality has no data environment axis.

At the outreach event on 7 July, Vice Chair Bowman asked for feedback on whether the report strikes the right balance on proportionality. This is my answer.

Every proportionality axis in the report sits on the institution side: size, complexity, interconnectedness, role in the financial system, structure, criticality of the function, and the materiality and risk of the use case. All sensible. All incomplete.

Take two banks of identical size and complexity deploying an identical use case. One operates in a market with full credit bureau coverage, a national digital identity scheme, and mostly electronic payments. The other operates in much of Latin America, where bureau coverage is thinner, identity records are more fragmented, and cash still carries a large

share of activity. Same institutional profile, same use case, materially different residual risk. Proportionality as drafted cannot see the difference.

The consequence lands in Sound Practice 7, which asks for data that is accurate, complete, consistent, reliable and secure. In some markets, complete is not reachable at any level of institutional effort, because the underlying records do not exist. Sound Practice 7 reads as though data quality is a function of discipline. Sometimes it is a function of the country. An institution that treats a structural data gap as a data preparation problem will spend real money and arrive back where it started.

Suggestion: add the maturity of the data environment as an explicit proportionality consideration, and add a line to Sound Practice 7 to the effect that where data limitations are structural rather than remediable, the appropriate response is to document the limitation and constrain the use case accordingly, rather than proceed on the assumption that more preparation will close the gap. Those are different decisions, taken by different people, and the report should say so.

This is also where proportionality can cut the other way. As drafted it only scales requirements down for smaller and simpler firms. A small firm operating in a hard data environment may need more, not less.

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?**

Section 3.7 handles GenAI and agentic AI better than I expected. The forms of oversight it sets out are useful, and human in command and the kill switch are the right instincts for high autonomy. My comment concerns one asymmetry inside Sound Practice 10.

The report accepts that AI will oversee AI. AI in the loop is listed as a form of human oversight, described as potentially warranted as institutions scale the number of use cases, and Section 3.7 says monitoring may need to be continuous or performed by AI. Section 1.3 says the same from the risk side: effective monitoring and detection may require augmentation with another AI agent or other forms of AI.

The report then closes the accountability question and leaves the control question open. It says that even when active monitoring is primarily undertaken by machines, financial institutions and individuals retain ultimate accountability. That is correct and I would not change it. But accountability is not a control. It settles who answers after a failure. It does not settle what was supposed to stop the failure.

Two gaps follow, and both are visible in Sound Practice 10 itself.

The first is effectiveness. Sound Practice 10 asks institutions to assess whether human overseers are providing effective oversight, and it returns to the point later, asking them to analyse oversight patterns for rubber stamping or for override patterns suggesting pervasive misalignment. Both are good. Neither has a machine counterpart. Where the overseer is a human, the report asks whether the oversight is any good. Where the overseer is a machine,

it does not ask. The party whose effectiveness gets scrutinised is the one that scales worst, and it is being displaced for exactly that reason.

The second is independence. Sound Practice 10 says responsibility for monitoring a use case after deployment should sit with staff who were not part of design and development, to promote objectivity. That is segregation of duties and it is right. There is no equivalent for an AI overseer. Nothing in the report suggests that a monitoring system should be independent of the system it monitors: a different provider, a different foundation model, different training data. Where both come from the same foundation model and the same vendor, the monitor inherits the blind spot of the thing it is watching. A control that fails under the same conditions as the process it controls is not a control. It is a second copy of the exposure, reported as assurance. This connects to my answer on Question 1. Given how concentrated the provider layer is, the monitor and the monitored will often come from the same place by default rather than by choice.

The report supplies its own answer in the last bullet of Section 3.7, where it says that as agent use increases, institutions should adapt human resources controls and processes to AI agents in a way that treats them as synthetic employees. I would take that metaphor seriously, because it resolves this. You would not let an employee audit their own work, and you would not staff the checker and the checked from the same team and call the result an independent review. Applied to agents, the report's own analogy requires the thing that is missing from it.

Suggestion: state that where AI is used as a control over other AI, the monitoring system is itself an AI use case within scope of Sound Practices 5 and 9; that its materiality should be inherited from the process it oversees rather than assessed on its own terms, since a monitor makes no business decisions and will otherwise be rated low materiality while carrying the consequence of whatever it fails to catch; and that institutions should weigh provider and model independence between overseer and overseen where the materiality of the underlying use case warrants it.

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

The seven stage lifecycle in Section 3.1 is a reasonable teaching device, and the report is careful to say there is no single set of stages that fits every case. My concern is narrower. The lifecycle assumes the institution controls the transitions between stages.

For a use case built on an externally maintained model, it does not. The report says as much in Section 1.3: third party providers may change the underlying models powering their AI systems without adequate notice, leaving institutions unaware of changes that could affect performance or risk outcomes. When that happens, the use case has moved from operation and monitoring back into design and development. The institution took no action. In the case that worries me, the institution does not know.

Sound Practice 3 already reaches for the right control. It says that where rollback or configuration access sits with a vendor, institutions should establish contractual arrangements for the vendor to provide timely information, together with compensating controls. That instinct should be carried into Section 3.1 and into re evaluation. The triggers

for re evaluation should include provider side version changes, and the report should be explicit that for externally maintained models, an institution's ability to know it has re entered an earlier stage is a contractual matter settled at procurement, not a monitoring matter settled afterwards.

That is also my answer on flexibility. The practices flex reasonably well against new types of AI. What they flex less well against is the possibility that the thing being governed is changed by someone else while it is in production.

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.**

The case studies are the best and the weakest part of the report.

The best, because they are concrete, and concreteness is rare in international guidance. A practitioner can read the fraud detection study and recognise their own situation in it.

The weakest, because every one of them worked. Over ninety five per cent of non performing SME loans identified three months early. Over eighty per cent of at risk borrowers kept out of default. Fraud losses down by more than twenty per cent. Preparation time halved, with roughly three times more client dialogue. Around four hundred thousand manual processing instances eliminated. A business day compressed to about forty five seconds. Six million lines of code.

Sound Practice 4 asks institutions to integrate insights from incidents, drift events, near misses, testing outcomes and validation findings into training and behavioural expectations. The report's own evidence base contains none of those things.

I would suggest including at least one case study of a use case that failed validation, was rolled back, or was deliberately not scaled, described with the same specificity as the successes. Failure material is the most transferable thing in risk management and the least likely to be volunteered by any single firm, which is exactly why an international body is better placed to collect it than a trade association or a vendor. If confidentiality is the obstacle, the studies are already anonymised. I would rather read one honest account of a model that looked fine in aggregate and was quietly withdrawn from a market than five more accounts of uplift.

I am not in a position to contribute a case study of my own. Anything I could describe from direct experience belongs to a former employer, and I am not willing to put it into a public document.

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

Partly. The case studies show what good looks like once it has arrived. They do not show the road, and a practitioner cannot act on an outcome.

One further point. The metrics appear to be reported by the institutions themselves, and the report does not say whether any of them were independently verified. That deserves a footnote. A board reading this document may reasonably take a twenty per cent reduction in fraud losses as a benchmark. It is a figure from one firm, with its own baseline, its own portfolio, and its own reasons to describe the result favourably. The report should say what it is.

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

I have no comments on the glossary definitions.

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

I am responding in a personal capacity. The views here are my own. I am not submitting on behalf of any institution.

My background is in financial risk, over roughly fourteen years, mostly in model risk governance and in the controls banks use against money laundering and other financial crime. I have worked on both sides of the supervisory line. I began at the Comisión Nacional Bancaria y de Valores, Mexico's banking supervisor, during the period when Mexico was implementing Basel III, and I later held risk governance roles at Mexican banks and at the Latin American hub of a global systemically important bank. I now work independently on model and AI risk governance for institutions operating across the United States and Latin America, and I teach market risk and Basel frameworks in a graduate programme at the Universidad Nacional Autónoma de México.

The report is a useful document, and I think institutions will actually pick it up and use it. That is not something I say about most international guidance. My comments are offered in that spirit.

One theme runs through all of my answers. The report treats AI risk as a property of an institution. For the firms it identifies as the most advanced adopters, internationally active banks, a large part of AI risk is a property of a jurisdiction.

That distinction is why I think this belongs in front of the FSB rather than a national supervisor. A weakness that sits in a jurisdiction rather than in a firm behaves differently from one that sits in a firm. It does not show up in group level metrics, because the jurisdiction is too small to move them. It is shared by every institution operating in that market on the same vendor stack, so it does not get competed away. And it accumulates for as long as nobody is looking, which means it does not arrive gradually at one firm. It arrives at once, across all of them. That is the shape of a stability problem, and it is the reason I am writing about geography in a report about technology.

The report notices the jurisdictional dimension once. Sound Practice 3 lists the regions affected by a use case among the attributes an institution should document. After that, geography drops out. Data governance, performance thresholds, and effective challenge are all specified at the level of the firm as a whole. Several of my suggestions ask only that the jurisdictional dimension be carried into the practices where it would do the most work.

A closing observation. Two of my points, the missing risk under Question 1 and the data environment axis under Question 2, come from the same place. Most of what I know about model governance I learned in Latin America, in markets small enough to be a rounding error in a group validation report and large enough to matter a great deal to the people who live in them. That perspective is thin in consultations like this one. Not because anyone excludes it, but because the firms with the resources to respond are concentrated in the jurisdictions where the data is already good, and they answer honestly about the conditions they know.

The FSB asked for input from a broad range of stakeholders. I would encourage the Secretariat to look at how much of what it receives comes from institutions whose home markets already work, and to weigh the responses accordingly.

I am happy to expand on any of the points above. I have also sent the full response as a single formatted document to fsb@fsb.org.

Response to the consultation report

Sound Practices for Responsible Adoption of Artificial Intelligence (AI)

Financial Stability Board. Consultation report of 10 June 2026. Comment period closing 22 July 2026.

Respondent: Sergio Federico Carrillo Farfán. Submitted in a personal capacity.

Mexico. July 2026.

About this response

I am responding in a personal capacity. The views here are my own. I am not submitting on behalf of any institution, and I have no objection to publication.

My background is in financial risk, over roughly fourteen years, mostly in model risk governance and in the controls banks use against money laundering and other financial crime. I have worked on both sides of the supervisory line. I began at the Comisión Nacional Bancaria y de Valores, Mexico's banking supervisor, during the period when Mexico was implementing Basel III, and I later held risk governance roles at Mexican banks and at the Latin American hub of a global systemically important bank. I now work independently on model and AI risk governance for institutions operating across the United States and Latin America, and I teach market risk and Basel frameworks in a graduate programme at the Universidad Nacional Autónoma de México.

The report is a useful document, and I think institutions will actually pick it up and use it. That is not something I say about most international guidance. My comments are offered in that spirit.

One theme runs through most of what follows. The report treats AI risk as a property of an institution. For the firms it identifies as the most advanced adopters, internationally active banks, a large part of AI risk is a property of a jurisdiction.

That distinction is why I think this belongs in front of the FSB rather than a national supervisor. A weakness that sits in a jurisdiction rather than in a firm behaves differently from one that sits in a firm. It does not show up in group level metrics, because the jurisdiction is too small to move them. It is shared by every institution operating in that market on the same vendor stack, so it does not get competed away. And it accumulates for as long as nobody is looking, which means it does not arrive gradually at one firm. It arrives at once, across all of them. That is the shape of a stability problem, and it is the reason I am writing about geography in a report about technology.

The report notices the jurisdictional dimension once. Sound Practice 3 lists the regions affected by a use case among the attributes an institution should document. After that, geography drops out. Data governance, performance thresholds, and effective challenge are all specified at the level of the firm as a whole. Several of my suggestions ask only that the jurisdictional dimension be carried into the practices where it would do the most work.

Question 1. Benefits and risks described in the report

The taxonomy in Section 1.3 is broadly right, and I agree with the four vulnerabilities carried over from the 2024 report. I would add one risk that is missing and adjust the framing of a second.

A missing risk: correlated failure in financial crime detection

Section 1.3 discusses homogeneous behaviours and correlated outcomes arising from reliance on common AI models, datasets, or infrastructure. Every consequence it lists is a markets consequence: herding, procyclicality, market stress, liquidity crunches, asset price vulnerabilities. Those are real. There is a second channel with a different shape, and the report does not name it.

Screening and transaction monitoring in most banks is not built in house. It is licensed from a small number of vendors, and those vendors calibrate against reference data that reflects some jurisdictions far better than others. An institution that buys the capability also buys the vendor's blind spots. If a vendor's name matching handles compound Spanish surnames poorly, then every institution using that vendor handles them poorly, in the same way, at the same time.

I use that example because it is the one I have watched. Across Latin America a person ordinarily carries a paternal and a maternal surname, the two are transposed between systems often enough to be routine, and accents and particles survive in one database and vanish in another. None of this is exotic where I work. It is the default. In the reference data most vendors calibrate on, it is an edge case, and it is treated as one.

What separates this from market herding is that the failure is silent. Herding shows up in prices. A monitoring model that systematically fails to alert on a class of activity produces no signal at all, because an absent alert looks exactly like an absent problem. Institutions measure the false positives they can see. Nobody measures the false negatives they never generated. The failure surfaces years later, through an enforcement action, or through a case that reaches the bank from law enforcement rather than the other way round.

I would suggest the report name this alongside the market channel: concentration among AI providers can produce correlated detection failure in the controls that are supposed to protect the integrity of the system, and that class of failure is harder to spot than the market class because there is no price to look at.

A framing to adjust: geographic concentration is treated as a continuity problem

Section 1.3 says there can be geographic concentration across key points in the AI supply chain, which could be affected by physical events like natural disasters. That treats geography as a business continuity question. There is a second sense that matters more.

Providers concentrated in a few jurisdictions build models on data generated in those jurisdictions. Losing the provider is one risk. The other is that the provider's model was never very good in your market, and that every competitor's provider has the same weakness. Concentration in that sense does not threaten availability. It quietly standardises a gap.

Question 2. Comprehensiveness and clarity of the sound practices

Three points.

Region is documented in Sound Practice 3 and then never used again

Sound Practice 3 asks institutions to document, among other attributes of a use case, the business lines, regions, products, customers, and market participants impacted. The report is right to include regions there.

Nothing downstream uses the field. Sound Practice 9 says institutions specify thresholds across all relevant dimensions of performance, and lists accuracy, range of output, robustness, stability, and

sensitivity or stress testing. Each is stated at the level of the use case as a whole. Sound Practice 7 lists what an effective data governance framework should address, and none of those considerations is jurisdictional.

This matters because of how internationally active groups actually build. A model is developed centrally, validated centrally against pooled data, and deployed across the group. Aggregate accuracy clears the threshold. Inside that aggregate, the two or three largest markets carry the number, and performance in a market representing two per cent of group exposure can be poor without ever moving it. That market is two per cent of the group. It is one hundred per cent of the local legal entity, whose board carries local legal accountability for the outcome and which usually did not build the model, cannot inspect it, and has no realistic way to challenge it.

Suggestion: Sound Practice 9 should say that where a use case has been documented under Sound Practice 3 as affecting several regions, performance should be assessed along those same regions as far as the data allows, and that thresholds should be set by reference to materiality at the entity that bears the consequence, not only materiality to the group. Sound Practice 7 could make the parallel point. Data quality should be assessed per jurisdiction, because it varies per jurisdiction, and an institution measuring it only in aggregate will not know which of its markets is being carried by the others.

This is a cheap change to make. The report already collects the field.

The model risk anchor in Sound Practice 5 has just moved

Sound Practice 5 tells institutions they can build their AI materiality and risk assessment on existing regulation or guidance on model risk management, and cites several jurisdictions.

In the United States, that anchor changed on 17 April 2026. The Federal Reserve, the OCC and the FDIC issued revised interagency guidance on model risk management, published as SR 26-2, OCC Bulletin 2026-13 and FDIC FIL-15-2026. It rescinded and replaced the 2011 guidance, and it also rescinded the 2021 interagency statement on model risk management for bank systems supporting Bank Secrecy Act and anti money laundering compliance. Three features are relevant here. It narrows the definition of a model. It states that generative AI and agentic AI models are not within its scope. And the agencies said they intend to issue a request for information addressing model risk management and banks' use of AI, including generative and agentic AI.

So an institution in the largest FSB jurisdiction that follows Sound Practice 5 and reaches for its model risk framework will find that framework silent on precisely the AI types this report spends the most effort on. It is silent by design, pending a consultation that has not yet opened.

I do not think the FSB should try to resolve that. But the report should say it out loud. As drafted, Sound Practice 5 implies the model risk anchor is both available and stable. In at least one major jurisdiction it is currently neither. A sentence acknowledging that model risk frameworks are under revision in some jurisdictions, and that generative and agentic AI presently sit outside those frameworks in some of them, would tell an institution something it needs to hear: the distance between "AI is out of scope of our model risk guidance" and "AI is ungoverned" is theirs to close, and for the moment this report is the best instrument they have for closing it.

There is a timing point worth flagging too. The final report is expected in October 2026. If the United States request for information runs on any normal schedule, the two will overlap. Some coordination between them would spare institutions that have to answer to both.

Proportionality has no data environment axis

At the outreach event on 7 July, Vice Chair Bowman asked for feedback on whether the report strikes the right balance on proportionality. This is my answer.

Every proportionality axis in the report sits on the institution side: size, complexity, interconnectedness, role in the financial system, structure, criticality of the function, and the materiality and risk of the use case. All sensible. All incomplete.

Take two banks of identical size and complexity deploying an identical use case. One operates in a market with full credit bureau coverage, a national digital identity scheme, and mostly electronic payments. The other operates in much of Latin America, where bureau coverage is thinner, identity records are more fragmented, and cash still carries a large share of activity. Same institutional profile, same use case, materially different residual risk. Proportionality as drafted cannot see the difference.

The consequence lands in Sound Practice 7, which asks for data that is accurate, complete, consistent, reliable and secure. In some markets, complete is not reachable at any level of institutional effort, because the underlying records do not exist. Sound Practice 7 reads as though data quality is a function of discipline. Sometimes it is a function of the country. An institution that treats a structural data gap as a data preparation problem will spend real money and arrive back where it started.

Suggestion: add the maturity of the data environment as an explicit proportionality consideration, and add a line to Sound Practice 7 to the effect that where data limitations are structural rather than remediable, the appropriate response is to document the limitation and constrain the use case accordingly, rather than proceed on the assumption that more preparation will close the gap. Those are different decisions, taken by different people, and the report should say so.

This is also where proportionality can cut the other way. As drafted it only scales requirements down for smaller and simpler firms. A small firm operating in a hard data environment may need more, not less.

Question 3. Balance between all forms of AI and emerging forms

Section 3.7 handles GenAI and agentic AI better than I expected. The forms of oversight it sets out are useful, and human in command and the kill switch are the right instincts for high autonomy. My comment concerns one asymmetry inside Sound Practice 10.

The report accepts that AI will oversee AI. AI in the loop is listed as a form of human oversight, described as potentially warranted as institutions scale the number of use cases, and Section 3.7 says monitoring may need to be continuous or performed by AI. Section 1.3 says the same from the risk side: effective monitoring and detection may require augmentation with another AI agent or other forms of AI.

The report then closes the accountability question and leaves the control question open. It says that even when active monitoring is primarily undertaken by machines, financial institutions and individuals retain ultimate accountability. That is correct and I would not change it. But accountability is not a control. It settles who answers after a failure. It does not settle what was supposed to stop the failure.

Two gaps follow, and both are visible in Sound Practice 10 itself.

The first is effectiveness. Sound Practice 10 asks institutions to assess whether human overseers are providing effective oversight, and it returns to the point later, asking them to analyse oversight patterns for rubber stamping or for override patterns suggesting pervasive misalignment. Both are good. Neither has a machine counterpart. Where the overseer is a human, the report asks whether the oversight is any good. Where the overseer is a machine, it does not ask. The party whose effectiveness gets scrutinised is the one that scales worst, and it is being displaced for exactly that reason.

The second is independence. Sound Practice 10 says responsibility for monitoring a use case after deployment should sit with staff who were not part of design and development, to promote objectivity. That is segregation of duties and it is right. There is no equivalent for an AI overseer. Nothing in the report suggests that a monitoring system should be independent of the system it monitors: a different provider, a different foundation model, different training data. Where both come from the same foundation model and the same vendor, the monitor inherits the blind spot of the thing it is watching. A control that fails under the same conditions as the process it controls is not a control. It is a second copy of the exposure, reported as assurance. This connects to my answer on Question 1. Given how concentrated the provider layer is, the monitor and the monitored will often come from the same place by default rather than by choice.

The report supplies its own answer in the last bullet of Section 3.7, where it says that as agent use increases, institutions should adapt human resources controls and processes to AI agents in a way that treats them as synthetic employees. I would take that metaphor seriously, because it resolves this. You would not let an employee audit their own work, and you would not staff the checker and the checked from the same team and call the result an independent review. Applied to agents, the report's own analogy requires the thing that is missing from it.

Suggestion: state that where AI is used as a control over other AI, the monitoring system is itself an AI use case within scope of Sound Practices 5 and 9; that its materiality should be inherited from the process it oversees rather than assessed on its own terms, since a monitor makes no business decisions and will otherwise be rated low materiality while carrying the consequence of whatever it fails to catch; and that institutions should weigh provider and model independence between overseer and overseen where the materiality of the underlying use case warrants it.

Question 4. Flexibility to accommodate newer types of AI over time

The seven stage lifecycle in Section 3.1 is a reasonable teaching device, and the report is careful to say there is no single set of stages that fits every case. My concern is narrower. The lifecycle assumes the institution controls the transitions between stages.

For a use case built on an externally maintained model, it does not. The report says as much in Section 1.3: third party providers may change the underlying models powering their AI systems without adequate notice, leaving institutions unaware of changes that could affect performance or risk outcomes. When that happens, the use case has moved from operation and monitoring back into design and development. The institution took no action. In the case that worries me, the institution does not know.

Sound Practice 3 already reaches for the right control. It says that where rollback or configuration access sits with a vendor, institutions should establish contractual arrangements for the vendor to provide timely information, together with compensating controls. That instinct should be carried into Section 3.1 and into re evaluation. The triggers for re evaluation should include provider side version changes, and the report should be explicit that for externally maintained models, an

institution's ability to know it has re entered an earlier stage is a contractual matter settled at procurement, not a monitoring matter settled afterwards.

That is also my answer on flexibility. The practices flex reasonably well against new types of AI. What they flex less well against is the possibility that the thing being governed is changed by someone else while it is in production.

Questions 5 and 6. Case studies

The case studies are the best and the weakest part of the report.

The best, because they are concrete, and concreteness is rare in international guidance. A practitioner can read the fraud detection study and recognise their own situation in it.

The weakest, because every one of them worked. Over ninety five per cent of non performing SME loans identified three months early. Over eighty per cent of at risk borrowers kept out of default. Fraud losses down by more than twenty per cent. Preparation time halved, with roughly three times more client dialogue. Around four hundred thousand manual processing instances eliminated. A business day compressed to about forty five seconds. Six million lines of code.

Question 6 asks whether the case studies give actionable insight. Partly. They show what good looks like once it has arrived. They do not show the road, and a practitioner cannot act on an outcome. Sound Practice 4 asks institutions to integrate insights from incidents, drift events, near misses, testing outcomes and validation findings into training and behavioural expectations. The report's own evidence base contains none of those things.

Two suggestions follow.

On Question 5, include at least one case study of a use case that failed validation, was rolled back, or was deliberately not scaled, described with the same specificity as the successes. Failure material is the most transferable thing in risk management and the least likely to be volunteered by any single firm, which is exactly why an international body is better placed to collect it than a trade association or a vendor. If confidentiality is the obstacle, the studies are already anonymised. I would rather read one honest account of a model that looked fine in aggregate and was quietly withdrawn from a market than five more accounts of uplift.

On Question 6, the metrics appear to be reported by the institutions themselves, and the report does not say whether any of them were independently verified. That deserves a footnote. A board reading this document may reasonably take a twenty per cent reduction in fraud losses as a benchmark. It is a figure from one firm, with its own baseline, its own portfolio, and its own reasons to describe the result favourably. The report should say what it is.

Question 7. Glossary

I have no comments on the glossary definitions.

A closing observation

Two of the points above, the missing risk under Question 1 and the data environment axis under Question 2, come from the same place. Most of what I know about model governance I learned in Latin America, in markets small enough to be a rounding error in a group validation report and large enough to matter a great deal to the people who live in them. That perspective is thin in

consultations like this one. Not because anyone excludes it, but because the firms with the resources to respond are concentrated in the jurisdictions where the data is already good, and they answer honestly about the conditions they know.

The FSB asked for input from a broad range of stakeholders. I would encourage the Secretariat to look at how much of what it receives comes from institutions whose home markets already work, and to weigh the responses accordingly.

I am happy to expand on any of the points above.

Sergio Federico Carrillo Farfán
Mexico