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

Response to Consultation

University of São Paulo Center for Banking Law

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?**

We agree with the benefits and risks of AI adoption by financial institutions identified in the report. We have included additional case studies in response to Question 5 below, which further illustrate some of the benefits observed in the Brazilian market, including measurable improvements in fraud detection, credit underwriting, pricing, compliance processes, operational efficiency, and decision-making quality across different types of financial institutions, particularly nonbank entities. In addition, we believe the report could address certain substantive risks in greater detail, particularly those relating to (i) regulatory perimeter risk, (ii) reliance on AI by senior management, (iii) data governance and treatment risk, (iv) social engineering and misinformation risk, and (v) emerging markets exposure.

Regulatory Perimeter Risk

Several risks identified in the report, particularly those relating to transparency, explainability, accountability and operational resilience, are amplified by the fact that many AI providers operate outside the regulatory perimeter applicable to financial institutions. These risks may be further heightened in the context of cross-border service provision, where multiple jurisdictions, regulatory regimes, and supervisory authorities are involved.

Financial regulators and supervisors are generally empowered to regulate supervised entities, but not necessarily the underlying AI providers whose systems may be used in critical financial activities. As a result, financial institutions may be required to comply with governance, transparency and explainability obligations without having access to sufficient information regarding the underlying models.

This challenge is particularly relevant in jurisdictions where no specific regulatory regime applies to AI providers. For example, although the Central Bank of Brazil ("BCB") has broad powers over entities operating within the National Financial System and the Brazilian Payments System, those powers generally do not extend to AI providers merely because their products are used by supervised institutions.

Accordingly, there may be a growing mismatch between the increasing reliance of financial institutions on AI-enabled services and the ability of financial regulators to oversee the

providers of those services. This may limit regulators' ability to effectively mitigate certain AI-related risks. To address this issue, regulators, supervisors, self regulatory bodies and AI providers should cooperate to ensure that financial institutions have access to sufficient information to understand, assess and monitor the AI systems they use, while preserving the confidentiality of proprietary information and avoiding unnecessary disclosure to end customers.

Reliance on AI by Senior Management

The report correctly identifies the risk of “shadow AI” and appropriately emphasizes the importance of human oversight. However, it devotes limited attention to the governance implications arising from the increasing use of AI by senior management and boards of directors.

As AI tools become more sophisticated, senior decision-makers may increasingly rely on AI-generated analyses, forecasts or recommendations when making strategic decisions. Excessive reliance on such outputs may create governance risks, particularly where the underlying models lack sufficient transparency or explainability. Such reliance may also contribute to automation bias, whereby users place undue confidence in AI-generated outcomes.

The report could also consider the value of independent challenge mechanisms for AI-generated outputs used by senior management and boards. As business decisions become increasingly supported by AI, it may no longer be realistic to expect decision-makers to manually verify the underlying data or analyses. Instead, institutions could require material AI-assisted recommendations to be subject to review through independent validation functions, alternative models, external benchmarking, or separate AI tools that are not trained on the same datasets, assumptions, or institutional objectives. Such mechanisms could help mitigate automation bias and reduce the risk of management decisions becoming overly dependent on a single AI-generated perspective.

The report could therefore further explore governance mechanisms designed to ensure that AI remains a decision-support tool rather than a substitute for managerial judgment. In this regard, institutions may benefit from adopting organization-wide AI governance policies that establish clear rules regarding the permissible use of AI in decision-making processes, including by senior management and governing bodies.

Data Governance and Treatment Risk

AI systems often require large volumes of data for training, testing and operation. While the report discusses data-related considerations, additional attention could be given to risks associated with the collection, processing and use of customer information.

Financial institutions may face legal, operational and reputational risks where personal data are used in a manner inconsistent with applicable privacy and data protection requirements. Additional concerns include data quality issues, inaccuracies in training datasets, cross-border transfers of information, difficulties in complying with data deletion requirements and the potential exposure of confidential information during model development or deployment.

Given the central role that data plays in AI systems, robust data governance frameworks are likely to remain a critical component of effective AI risk management.

Social Engineering And Misinformation Risk

The report discusses cybersecurity risks, but could place greater emphasis on the ability of AI systems to facilitate social engineering, misinformation and fraud at scale.

AI may significantly increase the sophistication and effectiveness of phishing campaigns, impersonation attacks, deepfakes and other forms of fraud targeting financial institutions, their employees and their customers. These developments may challenge existing fraud prevention and cybersecurity frameworks.

More broadly, AI-generated misinformation may affect market confidence and public perceptions regarding the financial condition of individual institutions. In extreme scenarios, false information disseminated at scale could contribute to liquidity stress or other forms of market disruption. The International Monetary Fund has identified such risks in IMF Note 2026/005, Artificial Intelligence and Cybersecurity in the Financial Sector.

Emerging Markets Exposure

The report could also further consider the specific implications of AI adoption for emerging markets. According to IMF Note 2026/005, financial institutions in emerging markets frequently have fewer cybersecurity resources and more limited access to advanced defensive technologies than their counterparts in advanced economies. At the same time, they may be more dependent on smaller or less mature technology providers for specialized financial applications.

These circumstances may increase operational and cybersecurity vulnerabilities and contribute to a widening gap in technological capabilities across jurisdictions. Over time, this dynamic could contribute to greater fragmentation within the global financial system, particularly if access to advanced AI technologies becomes concentrated among a small number of jurisdictions or providers.

As noted in our response to Question 5, this challenge may be particularly acute for nonbank financial institutions, which often rely more heavily on external providers of cloud infrastructure and foundation models and may face greater practical difficulties in obtaining meaningful access, monitoring, and audit rights over critical third-party services. The experience reported by market participants in our survey suggests that concentration and dependency risks may therefore have a disproportionately significant impact in emerging markets, where local alternatives to globally dominant technology providers are often more limited.

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

While we welcome the FSB's effort to articulate a technology-neutral and proportionate set of sound practices, in our view the practices are not yet sufficiently clear to enable consistent and proportionate responsible AI adoption across the financial sector.

First, although the report formally distinguishes between traditional AI, GenAI and agentic AI, and states that all practices apply to all forms of AI subject to proportionality, the substantive guidance, the risk taxonomy and the case studies are weighted heavily toward more advanced and higher-autonomy forms of AI. The most detailed expectations (for example, around enhanced documentation and logging, security risks related to prompt injection, and human oversight of autonomous agents) are framed primarily by reference to GenAI and agentic systems. As drafted, this emphasis risks being read as a general baseline rather than as guidance calibrated to higher-risk use cases.

In this case, institutions whose AI footprint is predominantly traditional machine learning (a technology in use across the financial sector for more than a decade and, in many jurisdictions, already governed by established rules including consumer protection requirements) may therefore be uncertain around the implementation of the sound practices, as well as to the extent that they apply to their AI-related products and services. As a consequence, these institutions may default to implementing controls designed for higher-risk models. The result would be guardrails more onerous than the materiality and risk of the underlying use case warrants, which is an outcome at odds with the very proportionality principle the report seeks to advance.

We suggest that the report signal more explicitly, for each practice, which elements are most relevant to traditional AI as opposed to emerging forms (for instance through a mapping table or clearer in-text signposting), so that proportionality becomes more operational rather than merely stated.

Second, the organisation-wide governance practices (Sound Practices 1–4) are described as cross-cutting and intended to inform the remaining practices. In several places, however, they go beyond setting transversal principles and instead restate operational expectations that are then developed in full elsewhere.

The clearest example is materiality and risk assessment: Sound Practice 3 expressly includes processes for "materiality and risk assessment", which is the entire subject of Sound Practice 5; a comparable, if narrower, overlap arises in relation to data quality and governance, addressed originally in Sound Practice 2 and better tackled under Sound Practice 7. This duplication creates ambiguity as to which practice is operative, where the authoritative expectation resides, and how institutions should avoid double-counting the same control across governance and lifecycle requirements.

We suggest that Sound Practices 1–4 be confined to genuinely cross-cutting governance principles, with operational detail consolidated in the relevant lifecycle practice and incorporated by cross-reference.

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?**

We do not consider that the report strikes an appropriate balance between managing risks across all forms of AI and addressing the specific risks of GenAI and agentic AI. While proportionality is mentioned at various points, the substantive recommendations undervalue the existing regulation and are built around scenarios typical of GenAI and agentic systems,

failing to adequately reflect either the longstanding experience of financial institutions with traditional AI or the regulatory frameworks that already govern it.

Existing Experience and Technology-neutral Regulation

Many financial institutions began adopting what we now call traditional AI more than a decade ago, progressively replacing earlier logistic-regression systems with decision-tree models and other machine-learning techniques. Regulation evolved alongside this shift and, in several jurisdictions, established rules that are effectively technology-neutral and apply equally to the earlier statistical models and to more recent machine-learning approaches.

Recent Brazilian rules illustrate this point. Transparency in the relationship with customers is already a regulatory requirement across the financial system: Resolution CMN 4,949/2021, of the National Monetary Council ("CMN", the governing policy body in charge of regulating financial institutions alongside the BCB) establishes it as a foundational principle for financial institutions, while Resolution BCB 155/2021 sets out equivalent obligations for broker-dealers and payment institutions. At the level of the individual data subject, Article 20 of the Brazilian General Data Protection Law (Federal Law 13,709/2018) establishes the right to request review of decisions taken solely on an automated basis, including those defining a person's credit profile.

Additionally, statutory rights regarding individual access to risk-evaluation data and explainability around credit scoring criteria have long been established in the Brazilian legal framework, specifically through the Consumer Protection Code (Article 43 of Federal Law 8,078/1990) and the Credit Information Law (Article 5(iv) of Federal Law 12,414/2011). Although enacted prior to recent advancements in machine learning, these technology-neutral provisions already function effectively as a baseline for AI-driven risk assessments and customer rights.

National rules also reach beyond the direct customer relationship and set internal requirements that both encourage and condition the use of AI-based processes. The expected-loss provisioning model, for example, encourages forward-looking estimation of the probability of default (Resolution CMN 4,966/2021) and where internal models are used to drive capital requirements, regulation already goes considerably deeper.

The strictest example, however, would be internal credit risk classification systems used under the IRB approaches. In Brazil, IRB is an opt-in regime and conditional on prior authorisation by the BCB and is subject to detailed requirements as to their structure, validation and governance, including that the classification methodology be consistent, justifiable and verifiable (Resolution BCB 303/2023).

Requirements of this kind, which demand that institutions be able to interpret and justify model outputs to supervisors, have sometimes been characterised as a constraint on the use of machine learning. The European Banking Authority's ("EBA") 2021 discussion paper and 2023' follow-up report on machine learning for IRB models reflected precisely this concern, noting that the use of such techniques in the IRB context remains limited because their complexity makes their results difficult to explain and justify to supervisors.

However, we would read this evidence differently: instead of revealing an absence of supervisory discipline, it shows that the existing prudential framework already reaches machine-learning use cases and requires explainability, encouraging regulated institutions to adopt a cautious approach. As can be seen on EBA's 2023 follow-up report, a set of rules with a principle-based approach, instead of being seen as a barrier, is welcomed by the industry, although the report also notes a lack of clarity on how to comply and flags that the (then-proposed) AI Act would benefit from further clarification to avoid legal uncertainty in its interaction with the prudential framework.

In other words, the level of supervisory discipline that the report tends to treat as a gap already exists or is under discussion, actively shaping how AI is implemented. Taken together, these frameworks already impose relevant governance, explainability and review obligations on traditional AI use cases, independently of the forms of AI on which the report concentrates.

Traditional AI is Inherently more Explainable

Traditional AI used in core financial applications is also inherently more explainable. Models such as decision trees, gradient-boosting methods and logistic regression operate over a comparatively limited set of features with an inspectable structure, allowing a relatively direct mapping between inputs and the resulting output, and producing reproducible results for a given input.

By contrast, GenAI systems differ in two material respects: they involve orders of magnitude, more parameters, and their outputs are generated through stochastic sampling at inference, so that the same input may yield different outputs. These characteristics make the explainability and reproducibility challenges that the report emphasises far more acute for GenAI and agentic systems than for the traditional models that still underpin most regulated decision-making.

For this reason, the recommended sound practices should address the difference between these architectures, making clear that GenAI demands greater control and supervision, while many traditional AI use cases are already disciplined by existing regulation.

A Balanced Approach Should Focus on Quality and Performance, not on Prescribing Controls by AI Type

In our view, a balanced framework should recognise that, while model performance metrics are mathematical, the selection of benchmarks and tolerance thresholds can vary significantly among institutions and jurisdictions, creating a degree of subjectivity in supervision, making audit and supervision by authorities more difficult.

This matters directly to the balance the report seeks to strike: a focus on standardised, outcome-based measurement applies proportionately to every form of AI according to its materiality and risk, whereas prescribing controls by AI type tends to default to the stricter requirements appropriate for GenAI and agentic systems, extending them to traditional models that do not warrant them.

In this sense, standardisation is the most effective response to this perceived problem. We would therefore encourage the FSB to promote international standardisation of methods for measuring performance and drift, so that these metrics can be compared across different types of institutions (with different levels of risk exposure) and improve comparability across jurisdictions, where local conventions can diverge considerably.

In Brazil, for example, the maximum Kolmogorov–Smirnov (KS) statistic is widely used. This practice of separating clearly performing from clearly non-performing borrowers derives from the local development of credit scores based on registries of defaulted borrowers ("cadastro de negativados") and it is not necessarily the most informative measure available. More modern measures, such as the Brier score, ROC AUC and PR AUC, can characterise model performance more completely and a common methodological baseline would strengthen supervisors' ability to mandate the specific measurement technique required to evaluate the models.

Difference Between Explainability and Transparency

On transparency, we identified a structural problem. AI providers treat their training data curriculum, meaning the sources used and the way those sources are processed by the model, as critical trade secrets, given the risk that their clients could become their competitors. For this reason, when they have sufficient market power, it is unlikely that the providers would grant full transparency into their models.

This tension is already acknowledged in existing regulation. Article 20, §1 of the Brazilian General Data Protection Law (Federal Law 13,709/2018) requires data controllers to provide clear and adequate information about the criteria and procedures used in automated decisions, while expressly safeguarding commercial and industrial secrets. The same reasoning can be seen in the Credit Information Law (Federal Law 12,414/2011), in which Article 5 establishes the providers' right to protect their trade secrets.

These provisions reflect a deliberate legislative choice to privilege explanation of the decision criteria over full disclosure of how a model is built, allowing greater explainability while protecting intellectual property and trade secrets.

We understand that the same logic applies to supervision: transparency in itself is of limited value to banking authorities, because it shows how a model is constructed but not how its decisions are actually reached.

A greater focus on explainability, meaning the ability to articulate how specific inputs drive a given output or decision, would be more productive. Explainability allows an institution to understand which factors are driving outcomes, to make adjustments where necessary to align with internal policies, and to detect and control unlawful bias. It delivers the supervisory and accountability benefits that transparency is often assumed to provide, without requiring providers to surrender proprietary information.

Furthermore, explainability is typically achieved through post-hoc methods like SHapley Additive exPlanations ("SHAP") and counterfactual checks. By leveraging these techniques, financial institutions can design validation processes tailored to their specific market, while

AI providers can verify that their products adhere to financial sector standards without forcing them to expose sensitive information.

There is also a competition dimension. Requiring transparency, when the most powerful and established AI providers have the bargaining power to resist it, would in practice mainly raise barriers for new entrants, reinforcing precisely the technological concentration that the report identifies as a risk. An explainability-based approach avoids that effect, because it can be satisfied without disclosing the commercially sensitive internals of a model.

This too speaks to the balance between forms of AI: explainability is readily achievable for the traditional models that underpin most regulated decision-making, while a heavier emphasis on oversight is warranted for GenAI and Agentic AI. In this sense, calibrating the expectation to the explainability profile of each architecture, rather than applying a uniform transparency requirement, is what an appropriately balanced framework would do.

Hardware Concentration Deserves at least as much Attention as Software

Finally, we would highlight that concentration risk in AI is not limited to software. As illustrated in our response to Question 5, these concentration dynamics may be particularly significant for fintechs and other nonbank financial institutions that rely heavily on third-party providers. The resulting dependencies can expose them not only to software-related concentration risks, but also to upstream hardware constraints that remain largely outside their control, potentially affecting their ability to scale, diversify providers, or maintain operational resilience.

Open-source and in-house alternatives to proprietary AI models do exist, however the supply of specialised hardware is different: it is provided by few vendors, with limited substitutability and significant geographic concentration.

As a result, any disruption to hardware supply, whether commercial, geopolitical or physical, could have a disproportionate and sector-wide impact on the institutions that have embedded AI in their processes. We encourage the FSB to give this dimension of the AI supply chain at least as much attention as software and model concentration, including in its analysis of correlated outcomes and operational resilience.

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

In summary, the sound practices are conceptually flexible enough to accommodate newer types of AI over time, but their current level of abstraction and implementation focus create practical challenges.

In support of this conclusion on the flexibility of the sound practices, we make the following observations:

The proposed sound practices are based on a framework intended to establish high-level principles, adopting a technology-neutral and risk-based approach. This favors the development of sound practices that are capable of accommodating and addressing new AI technologies over time, rather than prescribing controls tied to specific AI models or architectures.

The report adopts a broad and technology-agnostic perspective, encompassing different types of AI (e.g., neural networks and random forest algorithms) without attempting to provide an exhaustive taxonomy. Instead, it acknowledges that AI technologies will continue to evolve and that new approaches are likely to emerge over time.

This conclusion is further reinforced by Sound Practice 4, which encourages financial institutions to continuously learn, adapt and adjust their oversight, governance, risk management practices and capabilities as AI evolves. In doing so, the report appropriately recognizes that AI governance should be dynamic and capable of evolving alongside technological developments.

Furthermore, while it acknowledges emerging technologies such as GenAI and agentic AI, it avoids prescribing technology-specific requirements. Instead, it focuses on fundamental governance and risk management principles, that are applicable across current and future generations of AI systems.

On the other hand, we have the following observations regarding the approach adopted in the report:

1. While the principles-based approach provides the necessary flexibility, the high level of generality of the proposed sound practices may reduce their practical effectiveness. We believe that greater specificity would facilitate implementation and encourage broader and more consistent adoption by financial institutions.

One way to achieve this goal without compromising the flexibility discussed above would be to organize the sound practices around specific AI use cases or business processes. For example, the report could include dedicated sections addressing AI use in credit underwriting (potentially distinguishing among different stages, such as credit scoring and cash flow monitoring), fraud prevention, customer service and other key financial activities.

Such an approach would be consistent with the report's risk-based methodology, as different activities present distinct objectives, risk profiles, governance requirements, and supervisory expectations.

It would also facilitate a more contextual assessment of the appropriateness of particular sound practices to each activity. For example, with respect to Sound Practice 10 (Human Oversight), additional guidance could clarify under which circumstances human oversight is expected to generate meaningful risk mitigation and when it may instead impose unnecessary operational costs without corresponding benefits to customers or financial stability. Likewise, for specific use cases such as credit scoring, practical examples of effective human oversight mechanisms would help institutions implement the principle in a proportionate manner.

The same consideration applies to documentation requirements. Although the report only establishes a general expectation that AI-related processes be documented (for example, under Sound Practice 3), the scope and level of documentation should be proportionate to the particular activity, its materiality, and its associated risks.

These considerations are particularly important in emerging markets such as Brazil, where access to credit remains a significant challenge for a substantial portion of the population. AI governance requirements should therefore avoid creating unnecessary costs and barriers. Instead, they should support the development of more efficient and inclusive credit markets by mitigating risks such as algorithmic discrimination while enabling more accurate and data-driven credit risk assessments.

2. One recommendation that emerged from our discussions with Brazilian market participants concerns the convenience of distinguishing between sound practices applicable to AI models over which financial institutions exercise substantial control (such as internally developed or fine-tuned models) and those applicable to externally provided models, including foundation models and AI services accessed through public application program interfaces (APIs). While the underlying AI risks may be broadly similar, the institution's ability to identify, assess, validate and mitigate those risks differs significantly depending on the degree of control it has over the model. Accordingly, governance expectations should reflect these differences, with greater emphasis on model development, data selection and validation for internally controlled models, and on third-party risk management, due diligence and ongoing monitoring for externally provided AI systems.

3. The sound practices should also be sufficiently flexible to accommodate transparency obligations toward consumers without disregarding the legitimate discretion of financial institutions to determine whether, and to whom, they will offer financial products and services, subject to applicable law. In this regard, sound practices relating to explainability and transparency, such as Sound Practice 8, would benefit from expressly acknowledging that the use of AI does not, by itself, alter the institution's contractual freedom or create an obligation to offer products or services where none would otherwise exist. Any transparency requirements should therefore be implemented in a manner that is consistent with the applicable legal framework governing financial institutions' commercial autonomy and consumer protection obligations.

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.

Within the Brazilian market, we sought to assess how the case studies in the report compare with actual industry practice by engaging with local players and industry associations regarding their use of AI in day-to-day operations, including through an anonymous survey of financial institutions of varying size and nature operating in Brazil and meetings with players from the sector. Respondents ranged from large multiple and commercial banks to broker-dealers, payment institutions, and smaller regulated entities still at the exploratory stage, and their answers point to a genuinely uneven landscape. While some institutions report AI already embedded in credit, pricing, fraud, and compliance decisions with measurable gains, others remain at proof-of-concept stage and are still unable to formally assess impact on end clients.

Despite this variation, the same challenges are mentioned across nearly every response, namely explainability and transparency, bias and discriminatory treatment, cybersecurity

risks specific to AI, data quality and governance, dependence on cloud and foundation model providers, integration with legacy systems, and a shortage of qualified personnel.

Most respondents' approach to human oversight also mirrors the models described in the report, such as human-in-the-loop and AI-in-the-loop, in which they either treat AI output as auxiliary input with no direct decision-making effect, or require mandatory human review before any consequential decision. A broker-dealer respondent, for instance, reserves human review for critical cases, exceptions, or samples, having reported very positive, measurable gains from automated decision-making, while a payment institution limits AI's role in client interactions to non-material communications, signaling a more formal assessment of downstream impact. This confirms our broader observation that the report's case studies, drawn almost entirely from large, internationally active banks and G-SIBs, leave less visibility into how fintechs, digital banks, broker-dealers, and other nonbank regulated financial institutions apply the sound practices in practice, given that their business models, risk profiles, and governance structures differ significantly from those of traditional banks, and given their growing systemic relevance in jurisdictions like Brazil.

One example that helps illustrate this diversity of practice, without being taken as representative of the sector as a whole, is a large Brazilian fintech operating in payments and credit, which relies primarily on traditional AI, rather than GenAI or agentic AI, for its core fraud prevention and credit underwriting functions. Its principal safeguard against algorithmic discrimination is a counterfactual-check methodology, under which transactions or applicants flagged as fraudulent or non-creditworthy are reverse-tested to assess what would have happened had the model's decision gone the other way. A flagged case is treated as a potential indicator that the model may be producing systematically different outcomes across customer segments, prompting further investigation into possible proxy variables, feature interactions, or unintended discriminatory effects.

New or recalibrated models are also rolled out gradually, starting with small population samples, and the firm selects performance metrics case by case rather than committing to a single fixed measure, recognizing that a widely used statistic such as the maximum Kolmogorov–Smirnov statistic can understate real-world effectiveness if applied mechanically. Finally, as a nonbank financial institution, the firm faces third-party concentration risk more acutely than larger banks, given its dependence on third-party foundation models for customer-facing agents and the difficulty of securing the vendor access and audit rights contemplated by Brazilian outsourcing rules for financial industries, compounded by geopolitical constraints on hardware access.

Taken together, the diversity of experience surfaced in our survey, which considered institutions of different sizes, different business models, and different stages of AI maturity, together with the fintech example described above, is the sort of evidence we believe would meaningfully round out the report's existing case studies and give a fuller picture of responsible AI adoption across the financial sector.

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

Yes. While the report primarily focuses on internationally active banks and G-SIBs - often overlooking how fintechs, digital banks, broker-dealers, and other nonbank institutions apply

these practices - the case studies still offer highly actionable, pragmatic, and strategically valuable insights for all financial institutions navigating responsible AI adoption.

The report presents a comprehensive, macro-prudential analysis of the rapid integration of AI within the global financial system. It synthesizes the dual nature of AI: its capacity for profound operational transformation versus its potential to introduce or exacerbate systemic and localized vulnerabilities.

Based on this analysis, we have identified the following actionable insights derived from the case studies:

Organization-Wide AI Governance (Sound Practices 1 - 4)

To ensure AI adoption aligns with broader institutional strategies and risk appetites, financial institutions must implement robust governance structures at the highest levels.

Establish Strategic Boundaries, Risk Appetite, and Capability Building: As can be learned from the case study "Aligning skills and capabilities to AI strategy" (pages 19-20), the board of directors and senior management must actively define the outcomes expected from AI adoption and set explicit boundaries, which may include explicitly prohibiting fully automated decision-making in highly critical business areas.

This case study demonstrates how boards and senior management treat AI as a strategic pillar by investing in executive master classes and staff "AI bootcamps" to create hybrid professionals. It highlights the actionable step of setting concrete metrics for cost savings and capacity enhancements, and for less mature institutions, establishing strict boundaries by restricting staff to pre-approved tools.

Enforce Clear Accountability Structures: As provided by the case study "Documentation of AI use cases" (pages 22-23), institutions must define clear roles using frameworks such as the "three lines of defence," ensuring that business owners, risk managers, and independent auditors have distinct, non-overlapping responsibilities. Cross-functional coordination mechanisms must be established to integrate perspectives from compliance, legal, technology, and business units.

This case study details how large internationally active banks abandon static spreadsheets for dedicated software systems to maintain an enterprise-wide AI footprint. Actionable steps include tracking dependencies, risk materiality, and for agentic AI, implementing a formal "certification" process that records the agent's permitted scope and tool access.

Maintain Granular AI Inventories and Clear Accountability: The case study "Documentation of AI use cases" (pages 22-23) also demonstrates how organizations must expand their risk management documentation to track the entire lifecycle of AI use cases. This requires maintaining detailed, centralized inventories that document a model's purpose, materiality, data dependencies, known limitations, and third-party vendor relationships. For advanced models like Generative AI (GenAI), actionable documentation includes "prompt versioning" to enable rollbacks to previous, stable configurations.

Additionally, the case study "How an adaptive AI strategy delivered industry leadership" (pages 25-26) illustrates the actionable item of dynamically updating the organization's

governance. A G-SIB transitioned from a decentralized, research-first approach to a centralized AI group acting as an enterprise accelerator once AI use scaled to tens of thousands of employees, ensuring consistent application of risk and security standards.

Cultivate Organizational Adaptability: Institutions must continuously upskill their workforce to create a risk-aware culture capable of recognizing AI-specific failures, such as automation bias or data drift. Organizations should actively monitor external developments and participate in public-private consortia and information-sharing groups to stay abreast of novel AI threat vectors.

The case study "Learning from the AI ecosystem through public-private sector collaboration" (pages 24-25) maps to the directive to monitor external developments. Financial institutions are encouraged to participate in platforms like the MAS Project MindForge, the BoE/FCA AI Consortium, and the HKMA GenAI Sandbox to pilot use cases safely and build a shared understanding of emerging best practices.

AI Lifecycle Risk Management (Sound Practices 5 - 10)

Financial institutions must embed rigorous controls throughout the specific stages of AI development, from inception to deployment and eventual retirement.

Conduct Dynamic Materiality and Risk Assessments: The risk profile of an AI system is not static. Institutions must evaluate both the inherent and residual risks of an AI use case at its inception and systematically reassess these risks as the model scales or as real-world conditions change.

Prioritize Purpose-Driven Selection: When selecting an AI model, organizations should weigh the trade-offs between complex, high-performing algorithms and simpler models. A less complex, highly explainable model may be preferable if it meets operational needs with lower associated risk.

Enforce Strict Data Lineage and Quality: Because AI performance is fundamentally dependent on input data, institutions must establish robust data governance. Actionable items include classifying data sensitivity, validating third-party data, and maintaining strict "data lineage" to track exactly how data is cleaned, enriched, and transformed before it reaches the AI.

Address the Explainability Continuum: For complex "black box" algorithms, full transparency into how the AI maps inputs to outputs may be impossible. Actionable governance requires deploying compensating controls when explainability is limited; these include utilizing challenger models, conducting extreme stress testing, or imposing stricter guardrails on the model's operational scope.

The case study "Project Noor" (page 35) relates to the actionable item of deploying Explainable AI (XAI) tools to overcome the "black box" problem. Led by the BIS Innovation Hub, this initiative prototypes techniques to convert complex model logic into plain language and visual frameworks, allowing internal validators and supervisors to verify conceptual soundness.

Execute Proportional Performance Testing: Before deployment, developers must conduct rigorous testing to identify overfitting, instability, and unacceptable statistical bias. Post-deployment, institutions must enforce ongoing monitoring to catch "data drift" (when real-world data diverges from training data) and perform "red-teaming" (actively trying to break or manipulate the AI to uncover vulnerabilities).

In this context, the case study "Developmental testing" (page 39) emphasizes the actionable step of enforcing rigorous testing prior to deployment. A large bank halted the rollout of a machine learning trading model because testing revealed it was unstable, relied on unrepresentative data, and failed to consistently outperform its existing non-AI model.

Implement Meaningful Human Oversight (Sound Practice 10): Financial institutions must combat automation bias by aligning incentives for critical human review. Oversight mechanisms must match the AI's autonomy, adopting frameworks such as "Human-in-the-loop" (requiring explicit human approval for all actions), "Human-on-the-loop" (periodic intervention), or "Human-in-command" (macro-level management with immediate "kill switch" authority).

This specific actionable item is treated under two case studies: "Machine learning for credit risk management" (pages 7-8)" and "Strengthening fraud detection with agentic AI" (page 9). The first one presented a case where, for automating credit narratives, a bank ensured human credit approvers retained the responsibility of reviewing and validating the AI-generated content to maintain decision-making consistency. Meanwhile, the second one exemplified deploying "Human-in-the-loop" oversight for highly autonomous systems. While the bank's agentic AI autonomously proposes new fraud rules based on real-time data, human fraud analysts are required to review and approve all rules prior to implementation.

Cyber Defense and Third-Party Ecosystem Management (Sound Practices 11–12)

Upgrade Cyber Defenses for AI-Specific Threats: Traditional cybersecurity must be updated to counter novel AI attack vectors, such as prompt injections, model extraction, and data poisoning. Actionable steps include applying "secure-by-design" principles, strictly limiting an AI agent's permissions (the principle of "least privilege"), and incorporating AI-driven threat scenarios into mandatory operational resilience testing.

This actionable item is brought by the case study "Strengthening fraud detection with agentic AI" (pages 9-10), which details the actionable step of tackling offensive AI with defensive AI. A digital bank upgraded its facial recognition software with object detection and background similarity analysis to catch deepfakes and mule accounts.

Additionally, case study "UK Cross-Market Operational Resilience Group" (page 47) illustrates the requirement to incorporate AI into cyber testing. Financial institutions use the group's Dynamic Scenario Library to test resilience against threat actors using GenAI and agentic AI to bypass internal controls.

Demand Vendor Transparency and Manage Third-Party Risk: Institutions must hold third-party AI vendors to the same risk management standards as in-house developers. Actionable items include mandating standardized disclosures (such as "AI model cards" detailing the model's limitations and biases), securing appropriate legal indemnities if

transparency is withheld, and establishing clear exit strategies and data portability clauses to prevent vendor lock-in.

In this context, case study "Financial institutions' management of third-party AI risks" (pages 50-51) outlines actionable steps for managing vendor dependencies, such as standardizing vendor disclosures ("AI Cards"), monitoring vendor release notes for unapproved features, updating contracts for IP and cybersecurity, and applying compensatory testing or indemnities when vendor transparency is limited.

This is strictly governed by Resolution CMN 4,893/2021 and Resolution BCB 85/2021, which establish rigorous requirements for cybersecurity policies and the contracting of relevant data processing and cloud computing services by financial institutions and other licensed institutions by the BCB. Financial institutions are held ultimately responsible for the operational resilience and data security of their third-party AI vendors, necessitating explicit audit rights and exit strategies in vendor contracts.

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

We welcome the Financial Stability Board's (FSB) comprehensive effort to establish a standardized lexicon for AI in the financial sector. A clear, shared vocabulary is essential for effective risk management, cross-border regulatory alignment, and sound compliance frameworks.

In our view, the glossary is clear, conceptually sound, and well-aligned with existing international standards (such as the OECD definition of AI). However, to fully align with the most recent and rapid developments in industry practices (specifically the shift toward agentic workflows, natively multimodal models, and specialized enterprise risk controls), we recommend targeted updates to several key definitions that were sent via email to the FSB.

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

CDB | USP welcomes the opportunity to respond to the Financial Stability Board's ("FSB") Public Consultation Report on Sound Practices for Responsible Adoption of Artificial Intelligence (AI).

Founded in 2020 by law students at the University of São Paulo, CDB | USP is an academic think-tank dedicated to: (i) educating law students in Brazil on foundational legal and policy concepts in financial regulation, foreign exchange, digital assets, and emerging financial technology; and (ii) contributing independent research and policy analysis to local and international debates on cutting-edge financial regulation.

The group is registered as an official Extracurricular Activity at the University of São Paulo Law School and operates under the leadership of Associate Professor Carlos Pagano Botana Portugal Gouvêa and Professor Ruy Pereira Camilo Júnior.

UNIVERSITY OF SÃO PAULO

LAW SCHOOL

COMMERCIAL LAW DEPARTMENT

Center for Banking Law ("CDB | USP")

COMMENTS ON THE FINANCIAL STABILITY BOARD'S CONSULTATION REPORT "SOUND PRACTICES FOR RESPONSIBLE ADOPTION OF ARTIFICIAL INTELLIGENCE (AI)"

Faculty Leads: Associate Professor Carlos Pagano Botana Portugal Gouvêa and Professor Ruy Pereira Camilo Júnior.

Researchers:

Aaron Papa de Morais

Fabio Kupfermann Rodarte

Gabriela Trevisan

Matheus Cangussu

Pedro Campos Ferraz

Thássila Victória Nogueira

COVER NOTE

CDB | USP welcomes the opportunity to respond to the Financial Stability Board's ("FSB") Public Consultation Report on Sound Practices for Responsible Adoption of Artificial Intelligence (AI).

Founded in 2020 by law students at the University of São Paulo, CDB | USP is an academic think-tank dedicated to: (i) educating law students in Brazil on foundational legal and policy concepts in financial regulation, foreign exchange, digital assets, and emerging financial technology; and (ii) contributing independent research and policy analysis to local and international debates on cutting-edge financial regulation.

The group is registered as an official Extracurricular Activity at the University of São Paulo Law School and operates under the leadership of Associate Professor Carlos Pagano Botana Portugal Gouvêa and Professor Ruy Pereira Camilo Júnior.

CDB | USP
CENTRO DE DIREITO BANCÁRIO | USP

COMMENTS TO FSB'S "SOUND PRACTICES FOR RESPONSIBLE ADOPTION OF ARTIFICIAL INTELLIGENCE (AI)" CONSULTATION REPORT QUESTIONS[1]

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?

We agree with the benefits and risks of AI adoption by financial institutions identified in the report. We have included additional case studies in response to Question 5 below, which further illustrate some of the benefits observed in the Brazilian market, including measurable improvements in fraud detection, credit underwriting, pricing, compliance processes, operational efficiency, and decision-making quality across different types of financial institutions, particularly nonbank entities. In addition, we believe the report could address certain substantive risks in greater detail, particularly those relating to (i) regulatory perimeter risk, (ii) reliance on AI by senior management, (iii) data governance and treatment risk, (iv) social engineering and misinformation risk, and (v) emerging markets exposure.

Regulatory Perimeter Risk

Several risks identified in the report, particularly those relating to transparency, explainability, accountability and operational resilience, are amplified by the fact that many AI providers operate outside the regulatory perimeter applicable to financial institutions. These risks may be further heightened in the context of cross-border service provision, where multiple jurisdictions, regulatory regimes, and supervisory authorities are involved.

Financial regulators and supervisors are generally empowered to regulate supervised entities, but not necessarily the underlying AI providers whose systems may be used in critical financial activities. As a result, financial institutions may be required to comply with governance, transparency and explainability obligations without having access to sufficient information regarding the underlying models.

[1] These comments were informed by a targeted survey conducted with key market participants, including Banco BTG Pactual, CloudWalk, Avra, and Ancord. The views expressed herein represent the independent opinion of CDB | USP and do not necessarily reflect the institutional stance of the surveyed participants.

This challenge is particularly relevant in jurisdictions where no specific regulatory regime applies to AI providers. For example, although the Central Bank of Brazil ("BCB") has broad powers over entities operating within the National Financial System and the Brazilian Payments System, those powers generally do not extend to AI providers merely because their products are used by supervised institutions.

Accordingly, there may be a growing mismatch between the increasing reliance of financial institutions on AI-enabled services and the ability of financial regulators to oversee the providers of those services. This may limit regulators' ability to effectively mitigate certain AI-related risks. To address this issue, regulators, supervisors, self regulatory bodies and AI providers should cooperate to ensure that financial institutions have access to sufficient information to understand, assess and monitor the AI systems they use, while preserving the confidentiality of proprietary information and avoiding unnecessary disclosure to end customers.

Reliance on AI by Senior Management

The report correctly identifies the risk of "shadow AI" and appropriately emphasizes the importance of human oversight. However, it devotes limited attention to the governance implications arising from the increasing use of AI by senior management and boards of directors.

As AI tools become more sophisticated, senior decision-makers may increasingly rely on AI-generated analyses, forecasts or recommendations when making strategic decisions. Excessive reliance on such outputs may create governance risks, particularly where the underlying models lack sufficient transparency or explainability. Such reliance may also contribute to automation bias, whereby users place undue confidence in AI-generated outcomes.

The report could also consider the value of independent challenge mechanisms for AI-generated outputs used by senior management and boards. As business decisions become increasingly supported by AI, it may no longer be realistic to expect decision-makers to manually verify the underlying data or analyses. Instead, institutions could require material AI-assisted recommendations to be subject to review through independent validation functions,

alternative models, external benchmarking, or separate AI tools that are not trained on the same datasets, assumptions, or institutional objectives. Such mechanisms could help mitigate automation bias and reduce the risk of management decisions becoming overly dependent on a single AI-generated perspective.

The report could therefore further explore governance mechanisms designed to ensure that AI remains a decision-support tool rather than a substitute for managerial judgment. In this regard, institutions may benefit from adopting organization-wide AI governance policies that establish clear rules regarding the permissible use of AI in decision-making processes, including by senior management and governing bodies.

Data Governance and Treatment Risk

AI systems often require large volumes of data for training, testing and operation. While the report discusses data-related considerations, additional attention could be given to risks associated with the collection, processing and use of customer information.

Financial institutions may face legal, operational and reputational risks where personal data are used in a manner inconsistent with applicable privacy and data protection requirements. Additional concerns include data quality issues, inaccuracies in training datasets, cross-border transfers of information, difficulties in complying with data deletion requirements and the potential exposure of confidential information during model development or deployment.

Given the central role that data plays in AI systems, robust data governance frameworks are likely to remain a critical component of effective AI risk management.

Social Engineering And Misinformation Risk

The report discusses cybersecurity risks, but could place greater emphasis on the ability of AI systems to facilitate social engineering, misinformation and fraud at scale.

AI may significantly increase the sophistication and effectiveness of phishing campaigns, impersonation attacks, deepfakes and other forms of fraud targeting financial

institutions, their employees and their customers. These developments may challenge existing fraud prevention and cybersecurity frameworks.

More broadly, AI-generated misinformation may affect market confidence and public perceptions regarding the financial condition of individual institutions. In extreme scenarios, false information disseminated at scale could contribute to liquidity stress or other forms of market disruption. The International Monetary Fund has identified such risks in IMF Note 2026/005, Artificial Intelligence and Cybersecurity in the Financial Sector.

Emerging Markets Exposure

The report could also further consider the specific implications of AI adoption for emerging markets. According to IMF Note 2026/005, financial institutions in emerging markets frequently have fewer cybersecurity resources and more limited access to advanced defensive technologies than their counterparts in advanced economies. At the same time, they may be more dependent on smaller or less mature technology providers for specialized financial applications.

These circumstances may increase operational and cybersecurity vulnerabilities and contribute to a widening gap in technological capabilities across jurisdictions. Over time, this dynamic could contribute to greater fragmentation within the global financial system, particularly if access to advanced AI technologies becomes concentrated among a small number of jurisdictions or providers.

As noted in our response to Question 5, this challenge may be particularly acute for nonbank financial institutions, which often rely more heavily on external providers of cloud infrastructure and foundation models and may face greater practical difficulties in obtaining meaningful access, monitoring, and audit rights over critical third-party services. The experience reported by market participants in our survey suggests that concentration and dependency risks may therefore have a disproportionately significant impact in emerging markets, where local alternatives to globally dominant technology providers are often more limited.

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

While we welcome the FSB's effort to articulate a technology-neutral and proportionate set of sound practices, in our view the practices are not yet sufficiently clear to enable consistent and proportionate responsible AI adoption across the financial sector.

First, although the report formally distinguishes between traditional AI, GenAI and agentic AI, and states that all practices apply to all forms of AI subject to proportionality, the substantive guidance, the risk taxonomy and the case studies are weighted heavily toward more advanced and higher-autonomy forms of AI. The most detailed expectations (for example, around enhanced documentation and logging, security risks related to prompt injection, and human oversight of autonomous agents) are framed primarily by reference to GenAI and agentic systems. As drafted, this emphasis risks being read as a general baseline rather than as guidance calibrated to higher-risk use cases.

In this case, institutions whose AI footprint is predominantly traditional machine learning (a technology in use across the financial sector for more than a decade and, in many jurisdictions, already governed by established rules including consumer protection requirements) may therefore be uncertain around the implementation of the sound practices, as well as to the extent that they apply to their AI-related products and services. As a consequence, these institutions may default to implementing controls designed for higher-risk models. The result would be guardrails more onerous than the materiality and risk of the underlying use case warrants, which is an outcome at odds with the very proportionality principle the report seeks to advance.

We suggest that the report signal more explicitly, for each practice, which elements are most relevant to traditional AI as opposed to emerging forms (for instance through a mapping table or clearer in-text signposting), so that proportionality becomes more operational rather than merely stated.

Second, the organisation-wide governance practices (Sound Practices 1–4) are described as cross-cutting and intended to inform the remaining practices. In several places, however, they go beyond setting transversal principles and instead restate operational expectations that are then developed in full elsewhere.

The clearest example is materiality and risk assessment: Sound Practice 3 expressly includes processes for "materiality and risk assessment", which is the entire subject of Sound Practice 5; a comparable, if narrower, overlap arises in relation to data quality and governance, addressed originally in Sound Practice 2 and better tackled under Sound Practice 7. This duplication creates ambiguity as to which practice is operative, where the authoritative expectation resides, and how institutions should avoid double-counting the same control across governance and lifecycle requirements.

We suggest that Sound Practices 1–4 be confined to genuinely cross-cutting governance principles, with operational detail consolidated in the relevant lifecycle practice and incorporated by cross-reference.

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?

We do not consider that the report strikes an appropriate balance between managing risks across all forms of AI and addressing the specific risks of GenAI and agentic AI. While proportionality is mentioned at various points, the substantive recommendations undervalue the existing regulation and are built around scenarios typical of GenAI and agentic systems, failing to adequately reflect either the longstanding experience of financial institutions with traditional AI or the regulatory frameworks that already govern it.

Existing Experience and Technology-neutral Regulation

Many financial institutions began adopting what we now call traditional AI more than a decade ago, progressively replacing earlier logistic-regression systems with decision-tree models and other machine-learning techniques. Regulation evolved alongside this shift and, in several jurisdictions, established rules that are effectively technology-neutral and apply equally to the earlier statistical models and to more recent machine-learning approaches.

Recent Brazilian rules illustrate this point. Transparency in the relationship with customers is already a regulatory requirement across the financial system: Resolution CMN 4,949/2021, of the National Monetary Council ("CMN", the governing policy body in charge of regulating financial institutions alongside the BCB) establishes it as a foundational principle for financial institutions, while Resolution BCB 155/2021 sets out equivalent obligations for broker-dealers and payment institutions. At the level of the individual data subject, Article 20 of the Brazilian General Data Protection Law (Federal Law 13,709/2018) establishes the right to request review of decisions taken solely on an automated basis, including those defining a person's credit profile.

Additionally, statutory rights regarding individual access to risk-evaluation data and explainability around credit scoring criteria have long been established in the Brazilian legal

framework, specifically through the Consumer Protection Code (Article 43 of Federal Law 8,078/1990) and the Credit Information Law (Article 5(iv) of Federal Law 12,414/2011). Although enacted prior to recent advancements in machine learning, these technology-neutral provisions already function effectively as a baseline for AI-driven risk assessments and customer rights.

National rules also reach beyond the direct customer relationship and set internal requirements that both encourage and condition the use of AI-based processes. The expected-loss provisioning model, for example, encourages forward-looking estimation of the probability of default (Resolution CMN 4,966/2021) and where internal models are used to drive capital requirements, regulation already goes considerably deeper.

The strictest example, however, would be internal credit risk classification systems used under the IRB approaches. In Brazil, IRB is an opt-in regime and conditional on prior authorisation by the BCB and is subject to detailed requirements as to their structure, validation and governance, including that the classification methodology be consistent, justifiable and verifiable (Resolution BCB 303/2023).

Requirements of this kind, which demand that institutions be able to interpret and justify model outputs to supervisors, have sometimes been characterised as a constraint on the use of machine learning. The European Banking Authority's ("EBA") 2021 discussion paper and 2023' follow-up report on machine learning for IRB models reflected precisely this concern, noting that the use of such techniques in the IRB context remains limited because their complexity makes their results difficult to explain and justify to supervisors.

However, we would read this evidence differently: instead of revealing an absence of supervisory discipline, it shows that the existing prudential framework already reaches machine-learning use cases and requires explainability, encouraging regulated institutions to adopt a cautious approach. As can be seen on EBA's 2023 follow-up report, a set of rules with a principle-based approach, instead of being seen as a barrier, is welcomed by the industry, although the report also notes a lack of clarity on how to comply and flags that the (then-proposed) AI Act would benefit from further clarification to avoid legal uncertainty in its interaction with the prudential framework.

In other words, the level of supervisory discipline that the report tends to treat as a gap already exists or is under discussion, actively shaping how AI is implemented. Taken together, these frameworks already impose relevant governance, explainability and review obligations on traditional AI use cases, independently of the forms of AI on which the report concentrates.

Traditional AI is Inherently more Explainable

Traditional AI used in core financial applications is also inherently more explainable. Models such as decision trees, gradient-boosting methods and logistic regression operate over a comparatively limited set of features with an inspectable structure, allowing a relatively direct mapping between inputs and the resulting output, and producing reproducible results for a given input.

By contrast, GenAI systems differ in two material respects: they involve orders of magnitude, more parameters, and their outputs are generated through stochastic sampling at inference, so that the same input may yield different outputs. These characteristics make the explainability and reproducibility challenges that the report emphasises far more acute for GenAI and agentic systems than for the traditional models that still underpin most regulated decision-making.

For this reason, the recommended sound practices should address the difference between these architectures, making clear that GenAI demands greater control and supervision, while many traditional AI use cases are already disciplined by existing regulation.

A Balanced Approach Should Focus on Quality and Performance, not on Prescribing Controls by AI Type

In our view, a balanced framework should recognise that, while model performance metrics are mathematical, the selection of benchmarks and tolerance thresholds can vary significantly among institutions and jurisdictions, creating a degree of subjectivity in supervision, making audit and supervision by authorities more difficult.

This matters directly to the balance the report seeks to strike: a focus on standardised, outcome-based measurement applies proportionately to every form of AI according to its

materiality and risk, whereas prescribing controls by AI type tends to default to the stricter requirements appropriate for GenAI and agentic systems, extending them to traditional models that do not warrant them.

In this sense, standardisation is the most effective response to this perceived problem. We would therefore encourage the FSB to promote international standardisation of methods for measuring performance and drift, so that these metrics can be compared across different types of institutions (with different levels of risk exposure) and improve comparability across jurisdictions, where local conventions can diverge considerably.

In Brazil, for example, the maximum Kolmogorov–Smirnov (KS) statistic is widely used. This practice of separating clearly performing from clearly non-performing borrowers derives from the local development of credit scores based on registries of defaulted borrowers ("***cadastro de negativados***") and it is not necessarily the most informative measure available. More modern measures, such as the Brier score, ROC AUC and PR AUC, can characterise model performance more completely and a common methodological baseline would strengthen supervisors' ability to mandate the specific measurement technique required to evaluate the models.

Difference Between Explainability and Transparency

On transparency, we identified a structural problem. AI providers treat their training data curriculum, meaning the sources used and the way those sources are processed by the model, as critical trade secrets, given the risk that their clients could become their competitors. For this reason, when they have sufficient market power, it is unlikely that the providers would grant full transparency into their models.

This tension is already acknowledged in existing regulation. Article 20, §1 of the Brazilian General Data Protection Law (Federal Law 13,709/2018) requires data controllers to provide clear and adequate information about the criteria and procedures used in automated decisions, while expressly safeguarding commercial and industrial secrets. The same reasoning can be seen in the Credit Information Law (Federal Law 12,414/2011), in which Article 5 establishes the providers' right to protect their trade secrets.

These provisions reflect a deliberate legislative choice to privilege explanation of the decision criteria over full disclosure of how a model is built, allowing greater explainability while protecting intellectual property and trade secrets.

We understand that the same logic applies to supervision: transparency in itself is of limited value to banking authorities, because it shows how a model is constructed but not how its decisions are actually reached.

A greater focus on explainability, meaning the ability to articulate how specific inputs drive a given output or decision, would be more productive. Explainability allows an institution to understand which factors are driving outcomes, to make adjustments where necessary to align with internal policies, and to detect and control unlawful bias. It delivers the supervisory and accountability benefits that transparency is often assumed to provide, without requiring providers to surrender proprietary information.

Furthermore, explainability is typically achieved through post-hoc methods like SHapley Additive exPlanations ("SHAP") and counterfactual checks. By leveraging these techniques, financial institutions can design validation processes tailored to their specific market, while AI providers can verify that their products adhere to financial sector standards without forcing them to expose sensitive information.

There is also a competition dimension. Requiring transparency, when the most powerful and established AI providers have the bargaining power to resist it, would in practice mainly raise barriers for new entrants, reinforcing precisely the technological concentration that the report identifies as a risk. An explainability-based approach avoids that effect, because it can be satisfied without disclosing the commercially sensitive internals of a model.

This too speaks to the balance between forms of AI: explainability is readily achievable for the traditional models that underpin most regulated decision-making, while a heavier emphasis on oversight is warranted for GenAI and Agentic AI. In this sense, calibrating the expectation to the explainability profile of each architecture, rather than applying a uniform transparency requirement, is what an appropriately balanced framework would do.

Hardware Concentration Deserves at least as much Attention as Software

Finally, we would highlight that concentration risk in AI is not limited to software. As illustrated in our response to Question 5, these concentration dynamics may be particularly significant for fintechs and other nonbank financial institutions that rely heavily on third-party providers. The resulting dependencies can expose them not only to software-related concentration risks, but also to upstream hardware constraints that remain largely outside their control, potentially affecting their ability to scale, diversify providers, or maintain operational resilience.

Open-source and in-house alternatives to proprietary AI models do exist, however the supply of specialised hardware is different: it is provided by few vendors, with limited substitutability and significant geographic concentration.

As a result, any disruption to hardware supply, whether commercial, geopolitical or physical, could have a disproportionate and sector-wide impact on the institutions that have embedded AI in their processes. We encourage the FSB to give this dimension of the AI supply chain at least as much attention as software and model concentration, including in its analysis of correlated outcomes and operational resilience.

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

In summary, the sound practices are conceptually flexible enough to accommodate newer types of AI over time, but their current level of abstraction and implementation focus create practical challenges.

In support of this conclusion on the flexibility of the sound practices, we make the following observations:

The proposed sound practices are based on a framework intended to establish high-level principles, adopting a technology-neutral and risk-based approach. This favors the development of sound practices that are capable of accommodating and addressing new AI technologies over time, rather than prescribing controls tied to specific AI models or architectures.

The report adopts a broad and technology-agnostic perspective, encompassing different types of AI (e.g., neural networks and random forest algorithms) without attempting to provide an exhaustive taxonomy. Instead, it acknowledges that AI technologies will continue to evolve and that new approaches are likely to emerge over time.

This conclusion is further reinforced by Sound Practice 4, which encourages financial institutions to continuously learn, adapt and adjust their oversight, governance, risk management practices and capabilities as AI evolves. In doing so, the report appropriately recognizes that AI governance should be dynamic and capable of evolving alongside technological developments.

Furthermore, while it acknowledges emerging technologies such as GenAI and agentic AI, it avoids prescribing technology-specific requirements. Instead, it focuses on fundamental governance and risk management principles, that are applicable across current and future generations of AI systems.

On the other hand, **we have the following observations regarding the approach adopted in the report:**

1. While the principles-based approach provides the necessary flexibility, the high level of generality of the proposed sound practices may reduce their practical effectiveness. We believe that greater specificity would facilitate implementation and encourage broader and more consistent adoption by financial institutions.

One way to achieve this goal without compromising the flexibility discussed above would be to organize the sound practices around specific AI use cases or business processes. For example, the report could include dedicated sections addressing AI use in credit underwriting (potentially distinguishing among different stages, such as credit scoring and cash flow monitoring), fraud prevention, customer service and other key financial activities.

Such an approach would be consistent with the report's risk-based methodology, as different activities present distinct objectives, risk profiles, governance requirements, and supervisory expectations.

It would also facilitate a more contextual assessment of the appropriateness of particular sound practices to each activity. For example, with respect to Sound Practice 10 (Human Oversight), additional guidance could clarify under which circumstances human oversight is expected to generate meaningful risk mitigation and when it may instead impose unnecessary operational costs without corresponding benefits to customers or financial stability. Likewise, for specific use cases such as credit scoring, practical examples of effective human oversight mechanisms would help institutions implement the principle in a proportionate manner.

The same consideration applies to documentation requirements. Although the report only establishes a general expectation that AI-related processes be documented (for example, under Sound Practice 3), the scope and level of documentation should be proportionate to the particular activity, its materiality, and its associated risks.

These considerations are particularly important in emerging markets such as Brazil, where access to credit remains a significant challenge for a substantial portion of the population. AI governance requirements should therefore avoid creating unnecessary costs and barriers. Instead, they should support the development of more efficient and inclusive credit

markets by mitigating risks such as algorithmic discrimination while enabling more accurate and data-driven credit risk assessments.

2. One recommendation that emerged from our discussions with Brazilian market participants concerns the convenience of distinguishing between sound practices applicable to AI models over which financial institutions exercise substantial control (such as internally developed or fine-tuned models) and those applicable to externally provided models, including foundation models and AI services accessed through public application program interfaces (APIs). While the underlying AI risks may be broadly similar, the institution's ability to identify, assess, validate and mitigate those risks differs significantly depending on the degree of control it has over the model. Accordingly, governance expectations should reflect these differences, with greater emphasis on model development, data selection and validation for internally controlled models, and on third-party risk management, due diligence and ongoing monitoring for externally provided AI systems.

3. The sound practices should also be sufficiently flexible to accommodate transparency obligations toward consumers without disregarding the legitimate discretion of financial institutions to determine whether, and to whom, they will offer financial products and services, subject to applicable law. In this regard, sound practices relating to explainability and transparency, such as Sound Practice 8, would benefit from expressly acknowledging that the use of AI does not, by itself, alter the institution's contractual freedom or create an obligation to offer products or services where none would otherwise exist. Any transparency requirements should therefore be implemented in a manner that is consistent with the applicable legal framework governing financial institutions' commercial autonomy and consumer protection obligations.

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.

Within the Brazilian market, we sought to assess how the case studies in the report compare with actual industry practice by engaging with local players and industry associations regarding their use of AI in day-to-day operations, including through an anonymous survey of financial institutions of varying size and nature operating in Brazil and meetings with players from the sector. Respondents ranged from large multiple and commercial banks to broker-dealers, payment institutions, and smaller regulated entities still at the exploratory stage, and their answers point to a genuinely uneven landscape. While some institutions report AI already embedded in credit, pricing, fraud, and compliance decisions with measurable gains, others remain at proof-of-concept stage and are still unable to formally assess impact on end clients.

Despite this variation, the same challenges are mentioned across nearly every response, namely explainability and transparency, bias and discriminatory treatment, cybersecurity risks specific to AI, data quality and governance, dependence on cloud and foundation model providers, integration with legacy systems, and a shortage of qualified personnel.

Most respondents' approach to human oversight also mirrors the models described in the report, such as human-in-the-loop and AI-in-the-loop, in which they either treat AI output as auxiliary input with no direct decision-making effect, or require mandatory human review before any consequential decision. A broker-dealer respondent, for instance, reserves human review for critical cases, exceptions, or samples, having reported very positive, measurable gains from automated decision-making, while a payment institution limits AI's role in client interactions to non-material communications, signaling a more formal assessment of downstream impact. This confirms our broader observation that the report's case studies, drawn almost entirely from large, internationally active banks and G-SIBs, leave less visibility into how fintechs, digital banks, broker-dealers, and other nonbank regulated financial institutions apply the sound practices in practice, given that their business models, risk profiles, and

governance structures differ significantly from those of traditional banks, and given their growing systemic relevance in jurisdictions like Brazil.

One example that helps illustrate this diversity of practice, without being taken as representative of the sector as a whole, is a large Brazilian fintech operating in payments and credit, which relies primarily on traditional AI, rather than GenAI or agentic AI, for its core fraud prevention and credit underwriting functions. Its principal safeguard against algorithmic discrimination is a counterfactual-check methodology, under which transactions or applicants flagged as fraudulent or non-creditworthy are reverse-tested to assess what would have happened had the model's decision gone the other way. A flagged case is treated as a potential indicator that the model may be producing systematically different outcomes across customer segments, prompting further investigation into possible proxy variables, feature interactions, or unintended discriminatory effects.

New or recalibrated models are also rolled out gradually, starting with small population samples, and the firm selects performance metrics case by case rather than committing to a single fixed measure, recognizing that a widely used statistic such as the maximum Kolmogorov–Smirnov statistic can understate real-world effectiveness if applied mechanically. Finally, as a nonbank financial institution, the firm faces third-party concentration risk more acutely than larger banks, given its dependence on third-party foundation models for customer-facing agents and the difficulty of securing the vendor access and audit rights contemplated by Brazilian outsourcing rules for financial industries, compounded by geopolitical constraints on hardware access.

Taken together, the diversity of experience surfaced in our survey, which considered institutions of different sizes, different business models, and different stages of AI maturity, together with the fintech example described above, is the sort of evidence we believe would meaningfully round out the report's existing case studies and give a fuller picture of responsible AI adoption across the financial sector.

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

Yes. While the report primarily focuses on internationally active banks and G-SIBs - often overlooking how fintechs, digital banks, broker-dealers, and other nonbank institutions apply these practices - the case studies still offer highly actionable, pragmatic, and strategically valuable insights for all financial institutions navigating responsible AI adoption.

The report presents a comprehensive, macro-prudential analysis of the rapid integration of AI within the global financial system. It synthesizes the dual nature of AI: its capacity for profound operational transformation versus its potential to introduce or exacerbate systemic and localized vulnerabilities.

Based on this analysis, we have identified the following actionable insights derived from the case studies:

a) Organization-Wide AI Governance (Sound Practices 1 - 4)

To ensure AI adoption aligns with broader institutional strategies and risk appetites, financial institutions must implement robust governance structures at the highest levels.

i) Establish Strategic Boundaries, Risk Appetite, and Capability Building: As can be learned from the case study "Aligning skills and capabilities to AI strategy" (pages 19-20), the board of directors and senior management must actively define the outcomes expected from AI adoption and set explicit boundaries, which may include explicitly prohibiting fully automated decision-making in highly critical business areas.

This case study demonstrates how boards and senior management treat AI as a strategic pillar by investing in executive master classes and staff "AI bootcamps" to create hybrid professionals. It highlights the actionable step of setting concrete metrics for cost savings and capacity enhancements, and for less mature institutions, establishing strict boundaries by restricting staff to pre-approved tools.

ii) Enforce Clear Accountability Structures: As provided by the case study "Documentation of AI use cases" (pages 22-23), institutions must define clear roles using frameworks such as the "three lines of defence," ensuring that business owners, risk managers, and independent auditors have distinct, non-overlapping responsibilities. Cross-functional coordination mechanisms must be established to integrate perspectives from compliance, legal, technology, and business units.

This case study details how large internationally active banks abandon static spreadsheets for dedicated software systems to maintain an enterprise-wide AI footprint. Actionable steps include tracking dependencies, risk materiality, and for agentic AI, implementing a formal "certification" process that records the agent's permitted scope and tool access.

iii) Maintain Granular AI Inventories and Clear Accountability: The case study "Documentation of AI use cases" (pages 22-23) also demonstrates how organizations must expand their risk management documentation to track the entire lifecycle of AI use cases. This requires maintaining detailed, centralized inventories that document a model's purpose, materiality, data dependencies, known limitations, and third-party vendor relationships. For advanced models like Generative AI (GenAI), actionable documentation includes "prompt versioning" to enable rollbacks to previous, stable configurations.

Additionally, the case study "How an adaptive AI strategy delivered industry leadership" (pages 25-26) illustrates the actionable item of dynamically updating the organization's governance. A G-SIB transitioned from a decentralized, research-first approach to a centralized AI group acting as an enterprise accelerator once AI use scaled to tens of thousands of employees, ensuring consistent application of risk and security standards.

iv) Cultivate Organizational Adaptability: Institutions must continuously upskill their workforce to create a risk-aware culture capable of recognizing AI-specific failures, such as automation bias or data drift. Organizations should actively monitor external developments and participate in public-private consortia and information-sharing groups to stay abreast of novel AI threat vectors.

The case study "Learning from the AI ecosystem through public-private sector collaboration" (pages 24-25) maps to the directive to monitor external developments. Financial

institutions are encouraged to participate in platforms like the MAS Project MindForge, the BoE/FCA AI Consortium, and the HKMA GenAI Sandbox to pilot use cases safely and build a shared understanding of emerging best practices.

b) AI Lifecycle Risk Management (Sound Practices 5 - 10)

Financial institutions must embed rigorous controls throughout the specific stages of AI development, from inception to deployment and eventual retirement.

i) Conduct Dynamic Materiality and Risk Assessments: The risk profile of an AI system is not static. Institutions must evaluate both the inherent and residual risks of an AI use case at its inception and systematically reassess these risks as the model scales or as real-world conditions change.

ii) Prioritize Purpose-Driven Selection: When selecting an AI model, organizations should weigh the trade-offs between complex, high-performing algorithms and simpler models. A less complex, highly explainable model may be preferable if it meets operational needs with lower associated risk.

iii) Enforce Strict Data Lineage and Quality: Because AI performance is fundamentally dependent on input data, institutions must establish robust data governance. Actionable items include classifying data sensitivity, validating third-party data, and maintaining strict "data lineage" to track exactly how data is cleaned, enriched, and transformed before it reaches the AI.

iv) Address the Explainability Continuum: For complex "black box" algorithms, full transparency into how the AI maps inputs to outputs may be impossible. Actionable governance requires deploying compensating controls when explainability is limited; these include utilizing challenger models, conducting extreme stress testing, or imposing stricter guardrails on the model's operational scope.

The case study "Project Noor" (page 35) relates to the actionable item of deploying Explainable AI (XAI) tools to overcome the "black box" problem. Led by the BIS Innovation Hub, this initiative prototypes techniques to convert complex model logic into plain language and visual frameworks, allowing internal validators and supervisors to verify conceptual soundness.

v) **Execute Proportional Performance Testing:** Before deployment, developers must conduct rigorous testing to identify overfitting, instability, and unacceptable statistical bias. Post-deployment, institutions must enforce ongoing monitoring to catch "data drift" (when real-world data diverges from training data) and perform "red-teaming" (actively trying to break or manipulate the AI to uncover vulnerabilities).

In this context, the case study "Developmental testing" (page 39) emphasizes the actionable step of enforcing rigorous testing prior to deployment. A large bank halted the rollout of a machine learning trading model because testing revealed it was unstable, relied on unrepresentative data, and failed to consistently outperform its existing non-AI model.

vi) **Implement Meaningful Human Oversight (Sound Practice 10):** Financial institutions must combat automation bias by aligning incentives for critical human review. Oversight mechanisms must match the AI's autonomy, adopting frameworks such as "Human-in-the-loop" (requiring explicit human approval for all actions), "Human-on-the-loop" (periodic intervention), or "Human-in-command" (macro-level management with immediate "kill switch" authority).

This specific actionable item is treated under two case studies: "Machine learning for credit risk management" (pages 7-8)" and "Strengthening fraud detection with agentic AI" (page 9). The first one presented a case where, for automating credit narratives, a bank ensured human credit approvers retained the responsibility of reviewing and validating the AI-generated content to maintain decision-making consistency. Meanwhile, the second one exemplified deploying "Human-in-the-loop" oversight for highly autonomous systems. While the bank's agentic AI autonomously proposes new fraud rules based on real-time data, human fraud analysts are required to review and approve all rules prior to implementation.

c) Cyber Defense and Third-Party Ecosystem Management (Sound Practices 11–12)

i) Upgrade Cyber Defenses for AI-Specific Threats: Traditional cybersecurity must be updated to counter novel AI attack vectors, such as prompt injections, model extraction, and data poisoning. Actionable steps include applying "secure-by-design" principles, strictly limiting an AI agent's permissions (the principle of "least privilege"), and incorporating AI-driven threat scenarios into mandatory operational resilience testing.

This actionable item is brought by the case study "Strengthening fraud detection with agentic AI" (pages 9-10), which details the actionable step of tackling offensive AI with defensive AI. A digital bank upgraded its facial recognition software with object detection and background similarity analysis to catch deepfakes and mule accounts.

Additionally, case study "UK Cross-Market Operational Resilience Group" (page 47) illustrates the requirement to incorporate AI into cyber testing. Financial institutions use the group's Dynamic Scenario Library to test resilience against threat actors using GenAI and agentic AI to bypass internal controls.

ii) Demand Vendor Transparency and Manage Third-Party Risk: Institutions must hold third-party AI vendors to the same risk management standards as in-house developers. Actionable items include mandating standardized disclosures (such as "AI model cards" detailing the model's limitations and biases), securing appropriate legal indemnities if transparency is withheld, and establishing clear exit strategies and data portability clauses to prevent vendor lock-in.

In this context, case study "Financial institutions' management of third-party AI risks" (pages 50-51) outlines actionable steps for managing vendor dependencies, such as standardizing vendor disclosures ("AI Cards"), monitoring vendor release notes for unapproved features, updating contracts for IP and cybersecurity, and applying compensatory testing or indemnities when vendor transparency is limited.

This is strictly governed by Resolution CMN 4,893/2021 and Resolution BCB 85/2021, which establish rigorous requirements for cybersecurity policies and the contracting of relevant data processing and cloud computing services by financial institutions and other licensed

institutions by the BCB. Financial institutions are held ultimately responsible for the operational resilience and data security of their third-party AI vendors, necessitating explicit audit rights and exit strategies in vendor contracts.

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

We welcome the Financial Stability Board's (FSB) comprehensive effort to establish a standardized lexicon for AI in the financial sector. A clear, shared vocabulary is essential for effective risk management, cross-border regulatory alignment, and sound compliance frameworks.

In our view, the glossary is clear, conceptually sound, and well-aligned with existing international standards (such as the OECD definition of AI). However, to fully align with the most recent and rapid developments in industry practices (specifically the shift toward agentic workflows, natively multimodal models, and specialized enterprise risk controls), we recommend targeted updates to several key definitions.

Below is a detailed analysis of terms where refinement or addition could enhance the clarity and practical utility of the glossary for financial institutions and supervisors:

CDB | USP
CENTRO DE DIREITO BANCÁRIO | USP

Term	Current FSB Glossary Definition	Proposed New Language (Draft for FSB Glossary)	Rationale & Impact
Agentic AI	Systems designed to autonomously perform complex and extended tasks, often making decisions and taking actions with limited human oversight. These systems are distinct from GenAI but may incorporate generative capabilities to enhance content generation and decision making..	"AI systems designed to autonomously coordinate reasoning, planning, and tool-use (such as APIs, database queries, and external software) to execute complex, multi-step workflows. While distinct in operational flow, modern agentic systems typically leverage foundation models as their core cognitive engine."	Clarifies the technology stack. It alerts regulators that the security vulnerabilities and model risks of the underlying foundation model directly propagate to the autonomous agent.
Large Language Models (LLMs)	A type of foundation model that is trained on and designed to perform tasks with natural language. Key tasks LLMs perform include text generation, document classification, summarisation, question-and-answer, and sentiment analysis, among other tasks.	“Large Language & Multimodal Foundation Models - A class of foundation models trained on massive, diverse datasets to process, interpret, and generate outputs across multiple data modalities (including text, images, video, audio, and code natively)"	Reflects the industry state-of-the-art. Modern use cases in financial markets — such as insurance claim photo audits or biometric liveness checks for KYC — rely on native multimodality, not just text. We understand that this broader concept is included in the definition of 'GenAI.' If this is not the case, or if the FSB considers it beneficial, we suggest clarifying this point in the definition of “Large Language Models (LLMs), as suggested on the left.

Open-Source Models	AI models where the full training code, and in some cases the training data or its composition, is made publicly available for use and modification. These models offer customisability and reduce vendor lock-in but may introduce additional risks, such as security vulnerabilities, data privacy, or data quality concerns.	Replace “Open-Source Models” to ‘Open Models (Open-Weights vs. Open-Source)’ - “A class of AI models characterized by varying degrees of public access. A distinction is made between: (i) Fully Open-Source Models, where the model weights, complete training pipeline code, and datasets are publicly shared; and (ii) Open-Weights Models, where the pre-trained model parameters (weights) are shared publicly for local execution and fine-tuning, while the proprietary training code, datasets, and compute infrastructure remain private."	Critical for third-party risk management (Sound Practice 12). Open-weights models offer massive compliance benefits - allowing local hosting inside a bank's secure private cloud - but have different license agreements than open-source software.

LLM Hallucinations	Occur when a large language model (LLM) generates seemingly confident but inaccurate outputs, or fabricates nonsensical outputs in response to user inputs.	"Occur when a foundation model generates outputs that are ungrounded in its training data or provided context, resulting in seemingly confident outputs, but factually incorrect or representing fabricated claims."	**Broader Technological Scope:** Replacing "large language model (LLM)" with **"foundation model"** ensures technology neutrality and future-proofs the definition, capturing multimodal models (e.g., vision-language models) alongside text-based systems. **Technical Accuracy & Root Cause:** Rather than merely describing the outcome, the revised definition explicitly identifies the underlying mechanism: output that is **"ungrounded in its training data or provided context."** This distinction is critical in financial and operational contexts, particularly where systems use Retrieval-Augmented Generation (RAG) or external context windows. **Regulatory Rigor:** Replacing potentially challengeable phrasing such as "nonsensical outputs" with **"ungrounded"** and **"fabricated claims"** provides clearer, objectively verifiable standards for compliance assessment and risk management.
Prompt Injection *(New Addition)*	*Currently omitted from the glossary.*	"A security vulnerability where an attacker manipulates the input prompts of a generative AI system to override its system-level guidelines, causing it to perform unauthorized actions or leak restricted data".	Aligns the glossary directly with Sound Practice 11 (Cyber and ICT Risk). These are the most common and dangerous security threat vectors for generative and agentic AI in banking.

Jailbreaking *(New Addition)*	"A specialized form of prompt injection designed to completely bypass the safety, compliance, or ethical guardrails embedded in a model by its developers."	Aligns the glossary directly with Sound Practice 11 (Cyber and ICT Risk). These are the most common and dangerous security threat vectors for generative and agentic AI in banking.


CDB | USP
CENTRO DE DIREITO BANCÁRIO | USP