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

Response to Consultation

Financial Sector Conduct Authority

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?**

   Broadly yes, but there are some gaps worth flagging - particularly from an EMDE and conduct perspective:

   On benefits - the report is largely silent on financial inclusion as a potential upside of AI. In South Africa and similar contexts, alternative data credit scoring has real potential to bring previously excluded consumers into the formal financial system. That deserves explicit recognition. The report also says nothing about the benefits of AI for supervisors themselves - suptech is a real and growing area and EMDE regulators with thin teams need this acknowledged.

   The consultation report could more comprehensively address hyper-personalisation as a significant benefit and use case of AI within the financial sector. While the report references tailored advice under customer-focused applications, it does not sufficiently explore the broader capabilities of AI-driven predictive analytics and machine learning in enabling highly personalised financial services.

   Financial institutions are increasingly leveraging AI to analyse customer behaviour, transaction data, risk profiles, financial goals, and life-stage information to deliver tailored products and services across a wide range of financial activities. This includes personalised product recommendations relating to credit, insurance, savings, and investment products, allowing institutions to better align offerings with individual customer needs, preferences, and risk appetites.

   In addition, AI enables the delivery of dynamic and hyper-personalised customer engagement, including real-time alerts and recommendations relating to spending patterns, expenditure trends, budget management, cash-flow optimisation, and savings opportunities. These capabilities can enhance financial literacy, improve customer outcomes, and support more informed financial decision-making.

   The report could also recognise the growing use of AI in the design and administration of personalised rewards and loyalty programmes. For example, financial institutions increasingly customise credit and debit card rewards based on an individual's spending

behaviour, offering targeted cashback incentives at frequently used merchants, tailored grocery benefits, travel-related discounts, or other rewards aligned to customer preferences. Such AI-driven personalisation can enhance customer value propositions, strengthen engagement, and improve customer retention.

On risks - the followidoes are missing or underdeveloped:

Data scarcity. EMDE institutions often work with limited, unrepresentative, incomplete or fragmented datasets. The risks of overfitting and poor model generalisation in these environments are meaningfully different from what a large international bank faces and the report do not distinguish between them.

The consultation paper would benefit from more explicit consideration of the context-specific risks and structural challenges faced by financial institutions in EMDEs.

In particular, greater emphasis could be placed on data-related constraints. In many EMDEs, financial institutions operate in environments characterised by incomplete and fragmented datasets, often reflecting the presence of large unbanked and informal economic segments. The use of AI in such contexts may therefore lead to systematic biases in model outputs, with a heightened risk of reinforcing financial exclusion and exacerbating existing socioeconomic inequalities, especially in credit underwriting and customer profiling applications.

Furthermore, the paper could more fully acknowledge the disproportionate reliance on third-party and external technology providers in EMDEs, relative to in-house AI development capabilities.

This reliance introduces a number of important governance and risk management challenges, including:

Limited transparency and visibility into model design and underlying methodologies;

Reduced institutional control over risk management frameworks and model performance monitoring; and

Increased concentration and systemic risks, particularly where multiple institutions depend on a small number of global service providers.

Finally, it is important to recognise that AI governance frameworks in EMDEs are often constrained by broader structural and technological limitations, including legacy IT infrastructure, limited data ecosystems, and capacity constraints. These factors may undermine the effective implementation of otherwise robust governance principles and should therefore be explicitly reflected in the proportionality and guidance set out in the paper.

Incorporating these considerations would enhance the relevance and applicability of the proposed practices and support a more inclusive and risk-sensitive approach to AI governance and risk management.

Geopolitical concentration. The report flags commercial concentration risk in AI supply chains but does not go far enough. Most frontier models originate from a handful of jurisdictions. For EMDE regulators this creates a sovereignty dimension that goes beyond vendor lock-in.

Consumer vulnerability. The consumer protection section is present but thin. Consumers in EMDE markets often have lower financial literacy, fewer alternative products, and limited practical access to redress. This changes the risk profile of AI-driven decisions significantly and the report treats all consumers as broadly equivalent.

Cross-border regulatory arbitrage. Where AI systems are built and hosted offshore but deployed into EMDE markets, local regulators may have very limited visibility or reach. This is not addressed.

While the consultation paper appropriately recognises herding dynamics within the context of third-party dependencies and concentration risk, particularly noting that reliance on common AI models, datasets, or infrastructure can produce correlated behaviours and amplify procyclicality, it does not sufficiently consider the broader structural implications for financial markets.

Specifically, the analysis does not fully address how the widespread adoption of common AI systems may reshape market dynamics themselves. The use of similar models, data inputs, and optimisation techniques across institutions can contribute to a reduction in behavioural and strategic diversity, thereby weakening market heterogeneity. Over time, this may increase systemic fragility, as markets become more susceptible to collective responses to shocks, particularly under conditions of stress.

In this regard, the issue extends beyond correlated outcomes at the firm level to encompass a potential reconfiguration of market structure. A more explicit treatment of these dynamics, particularly the interplay between technological convergence and systemic resilience, would strengthen the paper’s assessment of AI-related financial stability risks.

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

The practices make sense for large, well-resourced institutions in advanced economies. For EMDE contexts - and especially for a conduct regulator overseeing thousands of entities with a small team - there are some real practical gaps.

Proportionality is mentioned but not operationalised. While the current framing of the sound practices is sufficiently high level and flexible, there is no meaningful guidance on what a smaller insurer or intermediary is actually expected to do versus a large bank, for example. Saying 'apply proportionality' without giving any worked examples or tiered expectations is not very useful for institutions or for supervisors trying to assess compliance. Additional guidance focused on “how to do proportionality in practice” would significantly enhance usability. An example of practical enhancements for instance in respect of governance and risk management is to provide guidance on what a clearly defined baseline control framework entails for smaller financial institutions. Furthermore, it would be helpful to provide guidance on how the lifecycle management emphasised in the sound practices

could be scaled for smaller financial institutions for example by allowing simpler validation approaches.

The conduct dimension is underweighted. The practices lean heavily toward prudential and risk management framing. For the FSCA, what matters most is whether the accountability chain for AI-driven consumer decisions is traceable and whether consumers can challenge outcomes. This gets a mention in Sound Practice 10, but it is not given the prominence it deserves given how much of the risk section deals with consumer harm.

There is nothing here for supervisors. The entire document is addressed to financial institutions. A conduct regulator with limited capacity trying to assess whether thousands of regulated entities are meeting these practices has no guidance on how to prioritise, what to look for, or how to develop supervisory tools. A companion piece for regulators would be very useful.

While Sound Practice 1 appropriately emphasises the importance of strategic direction and oversight by the board and senior management and mention in passing that appropriately resources to support AI adoption include appropriate competencies and skills at board and senior management level, it could be strengthened by explicitly recognising that effective governance of AI requires not only oversight structures, but also the necessary competencies, diversity of perspectives, and access to specialised expertise at governing body level.

For effective oversight, boards are required to develop a sufficient understanding of AI to effectively oversee its adoption, evaluate associated risks and opportunities, and guide its responsible use within the organisation. This includes assessing the strategic importance and criticality of AI to the institution's business model, products, services, and operations, as well as ensuring that appropriate monitoring, reporting, and risk management mechanisms are in place.

The consultation paper could therefore benefit from acknowledging that a lack of technological expertise at board level may impair the board's ability to fulfil its fiduciary and oversight responsibilities. Effective AI governance requires boards to collectively possess an appropriate mix of skills, experience, cognitive diversity, and independent perspectives to critically assess AI-related decisions and challenge management where necessary.

Importantly, while not every board member needs to be an AI specialist, boards should collectively maintain a foundational understanding of AI technologies and their implications.

In addition, governing bodies should have the capability to assess the business impact of AI and determine where AI applications are mission-critical or material to the institution's operations, customer interactions, risk management processes, or decision-making frameworks. Such assessments are essential to ensure that oversight and reporting structures are proportionate to the significance of the AI use case and the associated risk profile.

The consultation paper may also wish to recognise that, where material AI systems are deployed, financial institutions should consider establishing dedicated governance mechanisms to supplement board oversight. This may include the creation of board-level

technology or AI committees, advisory panels, or access to independent technical experts who can provide specialised advice on complex AI-related matters.

From a governance perspective, the consultation paper appears to omit consideration of a foundational layer: the ability of institutions to maintain visibility over the deployment of artificial intelligence within their operations.

In practice, the use of AI is not always readily identifiable. This may arise where third-party software solutions incorporate embedded AI functionalities, or where internal teams independently utilise generative AI tools for activities such as summarisation, drafting communications, or ad hoc experimentation. These forms of usage may fall outside formal governance channels and therefore remain insufficiently monitored.

In this regard, it is essential to establish a robust AI inventory framework as a core governance control. Such a framework should, at a minimum, encompass mechanisms for Shadow AI detection, a comprehensive use case register, system classification protocols, risk-tiering methodologies, and a centralised model registry. Collectively, these elements would enhance transparency, support effective oversight, and enable proportionate risk management across the institution's AI landscape.

The discussion on oversight and human-in-the-loop arrangements would benefit from greater specificity. While the paper emphasises the importance of maintaining human oversight, it does not clearly articulate what constitutes "sufficient" or "effective" oversight in contexts where agentic AI systems operate with a high degree of autonomy in live production environments, potentially for extended periods. In particular, there is limited guidance on the frequency, depth, and nature of human intervention required, as well as how accountability and control should be maintained when real-time decision-making is largely delegated to automated systems. A more granular framework outlining supervisory expectations in such scenarios would enhance clarity and practical applicability.

Furthermore, with respect to accountability, the paper would benefit from a more explicit delineation of roles and responsibilities across the different layers of AI systems. Clear allocation of accountability at the data, model, and application layers is essential to ensure effective governance, risk management, and auditability. This includes identifying responsible functions for data quality and provenance, model development and validation, and the deployment and operation of AI-enabled applications. Strengthening this aspect would support clearer lines of responsibility and facilitate more robust oversight within financial institutions.

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?**

Reasonable in principle but the balance tips toward GenAI and Agentic AI in a way that does not reflect where most EMDE institutions actually are right now.

In the South African financial sector, traditional ML is still the dominant form of AI in active deployment - used for credit scoring, fraud detection, transaction monitoring. The risks that

matter most day-to-day are bias in those models, explainability gaps, and model drift. These get comparatively less attention in the practices than the emerging technology risks.

The agentic AI treatment is thorough and forward-looking, which is fine. But the level of detail creates a compliance expectation that is disproportionate to actual deployment levels in most EMDE markets. Institutions may feel pressure to address agentic AI governance before they have even got the basics right for traditional ML.

From a conduct lens - the GenAI risks that are most immediate are in customer-facing applications: AI chatbots giving advice, automated product recommendations, AI-generated disclosure documents. The report touches on this but does not give enough concrete guidance on what financial institutions are expected to do. When must a consumer be told they are interacting with AI? What does adequate AI disclosure look like in an advice context? These are live questions that the practices leave open.

Sound Practice 8 on explainability stops short of saying what it should: that in consumer-facing, high-impact contexts, limited explainability is not just a technical challenge - it can be a regulatory breach. A conduct regulator needs to be able to point to that.

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

The menu approach is sensible given how fast the technology is moving. No argument with the principle.

The concern is that flexibility without supervisory capacity creates a gap that is hard to close. In a jurisdiction where regulatory teams are stretched, a non-prescriptive framework can effectively become voluntary in practice - institutions select the easiest practices, there is no supervisory floor, and meaningful oversight is difficult to sustain.

The FSCA would find it useful if the FSB identified a minimum baseline - a set of practices below which proportionality is not a valid reason for non-compliance. Even a short list of non-negotiables would help conduct regulators hold the line.

There is also no guidance on how regulators should prioritise oversight across a large population of regulated entities. Which practices matter most from a consumer protection standpoint? Where should a regulator with limited capacity concentrate its attention? This would be genuinely useful for EMDE regulators, and it is absent from the document.

On future-proofing - the commitment to update the practices as technology evolves is welcome but updating guidance takes time and capacity that EMDE regulators often do not have. Some thought about how to support EMDE regulators in keeping pace would be appreciated.

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.**

Not sufficiently, no. The case studies are almost entirely drawn from large internationally active banks and a handful of global insurers. There is nothing that reflects the experience of the kinds of institutions the FSCA regulates.

Specifically missing: short-term and long-term insurers using AI in underwriting or claims; CIS managers using AI in portfolio management or compliance; financial services intermediaries using AI-assisted advice tools; payment providers in high-volume low-value markets. These are not edge cases - they are core to the financial sectors that conduct regulators like the FSCA oversee.

There are also no EMDE case studies. All the examples come from institutions in advanced economies. The data environments, governance maturity, and risk profiles in those contexts are genuinely different from what EMDE institutions face and the case studies do not help institutions or regulators in those markets understand what 'good' looks like for them.

One other gap: all the case studies are success stories. For supervisory purposes, examples of things that went wrong and how institutions course-corrected would be considerably more instructive.

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

Partially, but with some caveats.

The scale of most case studies is not relatable for the majority of South African financial institutions. Metrics like 400,000 manual processing instances eliminated or six million lines of code generated in a year reflect the resources of G-SIBs, not regional insurers or smaller investment managers. These examples may inadvertently set benchmarks that feel unachievable.

The case studies also focus almost entirely on efficiency and productivity gains. From a conduct regulator's perspective, what matters is whether AI improved consumer outcomes - fairer decisions, faster claims, better-suited products. That angle is largely absent.

It is not always clear which sound practices a given case study is meant to illustrate. Mapping case studies explicitly to the relevant practices would make them more useful as implementation guidance.

A layered approach would help - showing what good looks like at different levels of institutional size and AI maturity, not just at the frontier.

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

Generally clear and fit for purpose. A few gaps worth noting.

Agentic memory poisoning is described in the body of the report (footnote 31) and referred to as a distinct threat but there is no entry in the glossary. Given how novel this concept is, a definition would be useful.

The definition of criticality is circular - it essentially says criticality means importance. This is not helpful for operationalisation, particularly in EMDE contexts where what counts as a critical operation may look very different from the G-SIB context the definition implicitly assumes.

Several terms that are directly relevant to a conduct regulator are missing entirely: automated advice, suitability, robo-advice, and contestability. The last one appears in Sound Practice 10 as a form of human oversight but is never defined.

Traditional AI is defined residually as techniques that pre-date recent advances like GenAI. This will date badly and does not give institutions or supervisors a meaningful definitional boundary to work with.

Data drift and model drift are correctly distinguished, which is good. But neither definition notes the conduct implications - that drift in a consumer-facing model can produce discriminatory or unsuitable outcomes. This is worth adding.

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

A few broader observations that did not fit neatly into the specific questions.

The entire document is addressed to financial institutions. There is no acknowledgement that supervisors - particularly those in EMDE jurisdictions with limited capacity - also need guidance. Overseeing AI adoption across a large and diverse regulated population requires supervisory tools, prioritisation frameworks, and capacity that most EMDE regulators do not currently have. A companion framework or even a section on supervisory expectations would significantly improve the practical utility of the document.

While the consultation paper, with its objective of facilitating the responsible adoption of artificial intelligence (AI) by financial institutions, is welcomed, there is also a need to more explicitly support the alignment of regulatory approaches across jurisdictions.

In this regard, further work at an international level, similar to the FSB development of high-level recommendations for crypto-assets, would be beneficial. The development of principles-based guidance directed at financial regulators, in addition to institution-focused sound practices, would help promote greater consistency in regulatory expectations and frameworks globally.

This is particularly important given that financial systems are deeply interconnected, and fragmented regulatory approaches to AI may give rise to systemic vulnerabilities that could threaten global financial stability. Divergent frameworks can create opportunities for regulatory arbitrage, obscure the build-up of risks across jurisdictions, and complicate effective cross-border supervisory cooperation.

More fundamentally, the case for global alignment is amplified by the borderless nature of AI systems and their rapid and exponential growth. AI models, data flows, and supporting infrastructure often operate across multiple jurisdictions simultaneously, while reliance on shared technologies and service providers increases interdependencies across financial systems.

South Africa operates a Twin Peaks model. Many of the sound practices sit at the intersection of conduct and prudential responsibilities - model risk management, data governance, third-party risk. The FSB framework does not address how these responsibilities should be divided in shared regulatory architectures. This is a gap that affects how EMDE twin peaks regulators can actually implement the practices domestically and coordinate their supervisory responses.

Cross-border AI deployment is under addressed. Many South African financial institutions use AI tools and cloud infrastructure hosted offshore. This raises data sovereignty concerns under POPIA and the Cybercrimes Act that Sound Practice 12 does not adequately cover. There needs to be more explicit guidance on the supervisory challenges of overseeing AI systems that process data across jurisdictional boundaries.

Financial inclusion does not feature as a regulatory objective anywhere in the document. For EMDE regulators and the markets they oversee, AI has real potential to advance or undermine inclusion - through alternative data credit scoring, through bias in historical training data, through language barriers in customer-facing AI. This deserves explicit recognition in the final report.

On language bias specifically - this is a material conduct risk in multilingual markets that the report ignores entirely. Customer-facing AI built primarily on English-language training data will systematically disadvantage non-English speakers. In South Africa this is not a marginal concern.