Humans keeping AI in check – emerging regulatory expectations in the financial sector[1]

Executive summary

Artificial intelligence (AI) including machine learning (ML) can offer significant potential to improve the delivery of financial services and operational and risk management processes. Technology is now part and parcel of financial services and there is no question that it will continue to drive profound changes for consumers and financial institutions. This trend has been supported by financial authorities' efforts to promote innovation and use of new technologies in the financial sector. In doing so, sound regulatory frameworks are essential to optimise benefits and minimise risks from these new technologies.

There are AI governance frameworks or principles that apply across industries and, more recently, several financial authorities have initiated development of similar frameworks for the financial sector. Within these frameworks, several common themes converge on general guiding principles on reliability, accountability, transparency, fairness and ethics. Other guiding principles mentioned in certain frameworks relate to data privacy, third-party dependency and operational resilience. While such high-level principles are useful in providing a broad indication of what firms should consider when using AI technologies, there are growing calls for financial regulators to provide more concrete practical guidance. One approach to meet this industry need is for regulators to provide compilation of emerging industry best practices on AI governance, where available, for each of these generally accepted principles.

Existing requirements on governance, risk management, as well as development and operation of traditional models also apply to AI models. These include governance requirements placing responsibility on the use of such models on boards and senior management of financial institutions. More specifically, firms are typically required to have sound model validation processes in place to ascertain the reliability of modelling results. Importantly, supervisors expect such models to be transparent, not only as part of good risk management practice, but also to enable supervisory review of models. Moreover, existing laws, standards or regulatory guidance cover data privacy, third-party dependency and operational resilience, including in the use of models.

While most of the issues arising from the use of AI by financial institutions are similar to those for traditional models, the perspective might be different. In the context of AI models, some of the common guiding principles identified above are viewed from the perspective of fairness. For example, ensuring reliability/soundness of AI models aims to avoid causing discrimination due to inaccurate decisions. Moreover, ensuring accountability and transparency in the use of AI includes ascertaining that data subjects are aware of data-driven decisions, and have channels to inquire about and challenge these decisions.

The stronger emphasis on fairness in the use of AI results in calls for more human intervention. The general thrust of AI pronouncements by regulatory bodies seems to focus on undesirable results, including unintended bias that lead to discriminatory outcomes that can arise from automations and lack of transparency in AI models. Although human-based modelling can suffer from similar weaknesses of AI models – for example bias or errors – a key distinguishing feature of the former is that humans can be held unambiguously accountable. Use of AI, however, can lead to philosophical

[1] Jermy Prenio (Jermy.Prenio@bis.org) and Jeffery Yong (Jeffery.Yong@bis.org), Bank for International Settlements. We are grateful to Douglas Araujo, Julian Arevalo, Orlando Fernandez Ruiz, Denise Garcia Ocampo, Xuchun Li and Oliver Thew for helpful comments. Luciana D'Agnone provided valuable administrative support.

reflection of the demarcation between machines and human beings. This challenge is compounded by a number of factors, including (i) the speed and scale of AI adoption by financial institutions; (ii) technical construct of AI algorithms; and (iii) lack of model explainability. From a regulatory standpoint, placing accountability firmly and clearly on responsible persons within a firm is key to operationalising sound AI regulatory frameworks. However, trade-offs will need to be made between reaping the benefits from large scale machine automation against the need for human input and oversight.

There is scope to further define fairness to support sound AI governance. Fairness, as described in the policy documents covered in the paper, relates to avoiding discriminatory outcomes. However, non-discrimination may not be explicit in consumer protection laws in some jurisdictions. Making non-discrimination objectives explicit may help provide a good foundation for defining fairness in the context of AI, provide a legal basis for financial authorities to issue AI-related guidance and, at the same time, ensure that AI-driven, traditional model-driven and human-driven decisions in financial services are assessed against the same standard.

The challenges and complexity presented by AI call for a proportional and coordinated regulatory and supervisory response. This requires differentiating the regulatory and supervisory treatment on the use of AI models, depending on the conduct and prudential risks that they pose. AI models whose results have significant implications on conduct and prudential risks will need to be subject to more stringent regulatory and supervisory treatment than those with less significant implications. In addition, use of AI by financial institutions will have implications for profitability, market impact, consumer protection and reputation. This calls for more coordination between prudential and conduct authorities in overseeing the deployment of AI in financial services.

Given emerging common themes on AI governance in the financial sector, there seems to be scope for financial standard-setting bodies to develop international guidance or standards in this area. Authorities' views on how these common themes should be implemented are still evolving. A continued exchange of views and experiences at the international level could eventually lead to the development of international standards. Such international standards could be helpful particularly to jurisdictions that are just starting their digital transformation journey. They can also serve as a minimum benchmark in guiding orderly deployment of AI technologies within the financial sector. As more specific regulatory approaches or supervisory expectations emerge on specific aspects of AI use cases, the standard-setting bodies can identify such common "best practice" that will be useful for other jurisdictions to consider. At the same time, given the evolving technology trends, principles-based guidance continue to have its benefits and can complement such best practice approach.