Introduction to "Beyond SR 11-7: Model Risk Management in the Age of AI/ML"

By Ben Peterson, Data Lead, EMEA

November 2023

Regulators are acting quickly as artificial intelligence/machine learning (AI/ML) models become pervasive in finance. In 2023, UK and U.S. regulators made significant steps toward tighter regulation, and the UK PRA published its new SS 1/13 regulation on Model Risk Management.

This article summarizes our November 2023 whitepaper on Model Risk Management in the age of AI, taking in the general and specific risks that regulators perceive in AI/ML models, the regulatory position across jurisdictions, and the ways MRM will need to change in the near term to align with the new consensus.

Introduction

Model Risk Management (MRM) is an important aspect of any bank's activity, but it is not an area in which regulation and best practice have tended to change frequently. Until recently, not even the rapid spread of AI/ML modeling techniques across all areas of banking has provoked clear change in regulations or best-practice frameworks.

This summer's publication of the UK's SS 1/23, the first regulation specific to AI/ML model risk in banking, is thus a key milestone and an opportunity to learn about and prepare for the global regulatory response to AI/ML. By analyzing SS 1/23, the non-regulatory frameworks being produced by organizations such as the White House, and the increasing volume of regulatory statements from other jurisdictions, we can ensure we are well prepared for the next major MRM regulation.

New Models and New Risks

AI/ML techniques bring both specific and general risks to a bank's model estate. Regulators have demonstrated a good understanding of these risks and recent regulatory statements highlight several areas of risk that will require increased attention from market participants.

General Risks

The general risks are independent of the particular AI/ML approaches and tools used; they are characteristic of the widespread adoption of advanced decision-making analytics regardless of the specific approach. Such risks include:

- **Responsibility:** This category of risk relates to the wellbeing and fair treatment of individuals whose lives may be affected by models. Responsibility includes:
  - Fairness - the risk that a model will produce an unfair outcome.
  - Bias - the risk that a model will exhibit systematic bias.
  - Ethical risk - the risk that behavior based on the model will not represent good social values.
- **Complexity:** This interesting group of risks relates to the fact that AI/ML models are often more complex and opaque than the models they supersede. Regulators have focused particularly on the explainability and interpretability of models, but this group of risks also includes risk arising from the complexity of input data.
- **Breadth:** AI/ML model use cases often span silos, datasets, and even business entities (many such models have multiple parties involved in creation and training). This creates new risks and regulators have particularly focused on governance of models operated by third parties.

Specific Risks

In addition to the general risks characteristic of large scale deployment of complex models, AI/ML brings various specific risks that relate to the demands of each model type. These risks are characteristics of individual AI modeling approaches that need to be understood, exposed, mitigated, and evaluated for each use case.

Regulators have generally refrained from making detailed technical statements about these specific risks, but model owners will still need to document and mitigate these risks, which can be conceptually complex. For example, some model families are very sensitive to the profile of input data, requiring uplift in data quality processes, and many behave differently depending on complex feature engineering and tuning decisions, which need to be recorded and reproduced as part of the validation process.

The MRM Regulatory Landscape

The current regulatory environment includes financial MRM regulations, general AI model regulations, and a set of best practice frameworks which are not regulatory, but which can be used to promote compliance and demonstrate diligence.

> ...these risks are primarily driven by the operating speed of the systems supporting the technology, the opacity, and complexity of the underlying models, the ability for continuous learning and dynamic recalibration, and data risks stemming from the use of larger datasets, including alternative or unstructured data.
>
> —PRA consultation paper, June 2022

Jurisdictions vary widely in terms of their maturity and operational approach, although they show a strong convergence in terms of their priorities. The UK has made a conscious decision to lead in AI governance and has produced a new MRM regulation in 2023 which essentially represents an AI/ML aware update to the U.S.'s SR 11-7. The U.S. has not yet produced a new MRM regulation, though the main financial regulators issued a Request for Information in 2021 which likely presages new regulation; on the other hand, the U.S. has been very active in promoting frameworks and high-level guidance. In particular, the National Institute of Standards and Technology's (NIST) AI Risk Management Framework (AI RMF) is influential and likely reflects future practices. The AI RMF is complemented by cross-jurisdiction initiatives such as the Enterprise Data Management Council's upcoming CDMC+ Analytics Checklist which focuses on the management of data and model together as assets.

The White House's 2023 Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence (White House's 2023 Executive Order) is likely to drive broad, high-level regulation similar to the EU's EU AI Act; but the Executive Order shows much more concern with prudential risk compared to the EU's primarily consumer-focused regulation. Other jurisdictions, including Canada and Japan, have also made new statements on MRM and AI.

The regulatory landscape then is complex with varying approaches from different regulatory and standards bodies; our whitepaper contains a more detailed overview. However, in practice, similar themes occur across all these initiatives, and these themes have direct implications for practical MRM.

Regulatory Trends and Priorities

This section summarizes the areas of consensus across recent regulations and regulatory statements, with a particular focus on the implications for SS 1/23. For a more detailed view, please see our whitepaper.

Definition and Scope of a 'Model'

A key area of consensus is the broadening of the definition of a 'model'. SR 11-7 defines a model as specifically 'quantitative' whereas almost all recent regulatory pronouncements make it clear that models with quantitative or generative outputs are in-scope. SS 1/23 in particular goes further by:

- Extending the scope downstream, stating that post-model adjustments are be part of the 'model'.
- Broadening the scope horizontally, to include all use cases 'relevant to the safety and soundness of firms'.

This tendency to see the 'model' as everything that contributes to a meaningful decision is typical. But the concept of 'proportionality' is also foregrounded, acknowledging that not every model needs to be governed minutely. Again, SS 1/23 goes into more detail on this than SR 11-7 did and gives some details on model 'tiering' —the allocation of models to different 'tiers' of risk significance. SS 1/23 goes further than previous regulations in proposing an approach to tiering which includes the twin criteria of materiality and complexity, explicitly referring to AI/ML risks such as data sensitivity and interpretability.

Model Lifecycle and Governance Concerns

SR 11-7 was groundbreaking in its day by describing a model lifecycle that was more than just a software development lifecycle with activities such as back-testing attached to it. SS 1/23 takes another step by clearly separating governance concerns from the steps of the model lifecycle and indeed showing that those concerns relate to the lifecycle in a complex way.

In this respect, SS 1/23 is well aligned to the newest offerings from many execution platform vendors and to frameworks such as NIST AI RMF which increasingly separate the procedural steps in the model lifecycle from the policy-based risk management activities of MRM.

Specific Risks

While traditional MRM focused on quantitative financial models that presented a limited (though critical) spectrum of risks, newer regulations are identifying and addressing more specific risks that are unique to complex AI/ML models. For instance, SS 1/23 references model explainability specifically in its tiering criteria. The joint U.S. regulator's RFI of 2021 in particular highlights explainability and other specific AI risks. Several statements highlight the importance of data quality and data-associated risk: SS 1/23 highlights data quality management and data 'relevance' while the U.S. regulator's RFI highlights 'broader and more intensive data usage' as an emerging risk.

Accountability

At the same time, many regulatory statements seem to move in the direction of integrating MRM more closely with an organization's global compliance and accountability structures. SS 1/23 for example connects MRM to the UK's Senior Management Functions structure, while Japan's FSA Principles for Model Risk Management clearly articulate how MRM should interact with the three lines of defense. The trend is to link MRM outcomes and risks extremely closely to data risks and to the overall accountability framework of the enterprise.

MRM Requirements

To meet emerging risk management needs—both regulatory and internal—three capabilities within the enterprise will need to adapt: the actual MRM process and tooling, the physical model execution and deployment environment, and the data supply chain. Additionally, the delivery of models will need to continue to move in the direction of productization.

Productization

It is clear that in future AI/ML models will often need to be accompanied by expanded sets of documentation and metadata—including information about intended use, provenance, performance, and justification. In many ways, this is a form of productization; just as a 'data product' consists of data and rich accompanying information, so a 'model product' could be seen as consisting of the model and enough metadata to guarantee compliance and appropriate use.

> AI may present particular risk management challenges to financial institutions in the areas of explainability, data usage, and dynamic updating.
>
> —Joint RFI by FRB, FDIC et al, 2021

Data Pipeline

Regulators and practitioners alike are focusing heavily on **the importance of data quality** in ensuring model quality. Some practitioners, such as EDMC, go as far as to consider models and data as two types of 'analytics asset' that should be governed under the same framework. Whether a firm takes this view or not, it is clear that significant uplift in data quality and data governance will be required to de-risk AI/ML models. This uplift is likely to include:

- Improvements in data tagging, labelling, and discoverability.
- Clarification and strengthening of data ownership.
- Further automation of data quality and data lineage.
- Uplift of root cause analysis and data exception handling processes.
- Increased governance of third-party and internal data providers.

Firms will need to be able to make guarantees about the quality, usability, and meaning of **third-party data and analytics assets**. For GDPR-regulated firms, this is not an entirely new requirement, and techniques used in GDPR compliance may prove applicable to MRM—for example, the use of third party guarantees about data flow.

AI/ML regulation will require new quality metrics as part of the overall data quality and control process. For example, regulators have shown a strong understanding of the ways in which **novel combinations of data pose** new risks, a point which is called out by the White House's 2023 Executive Order.

Maintaining Reproducibility

Metadata and metrics are only useful when correct (a fact repeated in several recent regulatory statements) and demonstrating that model metadata is correct has historically often been very difficult. For example, an organization may state that a given set of hyperparameters were used to produce certain model outputs, but this is difficult to verify without a re-run of the model.

As a result, enterprises have tended to adopt one of two main approaches to governing the model execution platform itself:

- Developing highly-governed, highly-reproducible model execution platforms in-house.
- Moving model execution to large, usually cloud-based vendors who offer a degree of reproducibility.

The adoption of AI/ML at scale puts considerable strain on the requirement for reproducibility, because both data sets and modeling processes are more complex. Another characteristic of the AI/ML execution platform is that it is likely to span multiple entities, with features such as Natural Language Processing being farmed out to partners while the financial institution continues to execute and orchestrate other aspects of the information flow.

The result is a demanding set of aspirational requirements for physical model execution platforms, and a focus on analytics operations that is likely to continue and expand the trend begun with DataOps and MLOps.

Conclusion

In this article, we have summarized the risks posed by AI/ML, the recent regulatory responses to those risks, and the implications for future MRM processes and tools. One lesson seems to be that the rise of AI/ML will require a significant uplift in governance and risk processes in order to maintain compliance; but that uplift will itself bring considerable benefits, since to get good results from complex models requires a strong data pipeline and validation framework regardless of compliance needs. A more detailed exploration of these topics is in our full whitepaper Beyond SR 11-7: Model Risk Management in the Age of AI/ML available on our website.