Global Research
& Risk Solutions

CRISIL
An S&P Global Company

MRM in times of AI

Model risk management for banks in the AI paradigm takes off from the traditional craft

June 2023

Analytical contacts:

Anshuman Prasad
Senior Director
Global Head, Risk Analytics
anshuman.prasad@crisil.com

Shashi Sharma
Director
Quantitative Services – Model Risk
shashi.sharma@crisil.com

Artificial intelligence and machine learning (AI/ML) are expanding the frontiers of finance. Over the next few years, we foresee a proliferation of AI/ML use cases in the back, middle and front office functions at banks.

Surveys show growing recognition among finance executives that adoption of AI/ML and use of advanced analytics will be the next big differentiator among competitors.

But amid the hype, banks need to tread with caution. For, AI/ML models come loaded with risk and complexity. The traditional or existing model risk management (MRM) frameworks are plainly insufficient to handle them.

This white paper uses the foundational building blocks of a traditional MRM framework as the point of departure to identify the precise areas where the framework must adapt to new AI/ML model risks across the model lifecycle.

It also recommends specific techniques where adjustments are due at each stage, in order to validate such models.

The MRM framework

The very first step to effectively augment the MRM framework of banks where AI/ML methodologies are used, it to understand what goes into the basic framework.

The fundamental elements of any MRM framework include:

1. **Model identification**: This involves establishing relevant criteria for identifying the model

2. **Model risk assessment:** This entails evaluating the magnitude and significance of the identified risks, considering factors such as the complexity and criticality of the models, potential impact on business decisions and regulatory requirements, eventually leading to model risk classification

3. **Model risk mitigation:** This includes implementing controls and measures to reduce the identified risks to an acceptable level. It may involve model validation, robust model development and implementation processes, data quality assurance and model governance frameworks

4. **Model risk monitoring and maintenance:** Models need to be continuously monitored and reviewed for their performance over time to detect any emerging risks or deviations from expected behaviour. This includes ongoing model validation, performance monitoring and periodic reassessment of risks

5. **Reporting and governance:** This refers to delineating responsibility, accountability, and governance in MRM. It includes defining roles and responsibilities, documenting policies and procedures, and ensuring effective communication and reporting to stakeholders, including senior management and regulators

Adapting the building blocks of MRM to AI/ML model risks

How can the above framework constituents be tuned to factor in AI/ML specific model risks?

Model identification: Every model requires a clear definition. Nearly all financial institutions have successfully transposed the SR 11-7's model definition guidance into a template that helps them distinguish a model from a non-model. However, no standard definition for AI/ML models is available yet. The widely varying definitions could be attributed to the lack of a clear dividing line between traditional and AI models

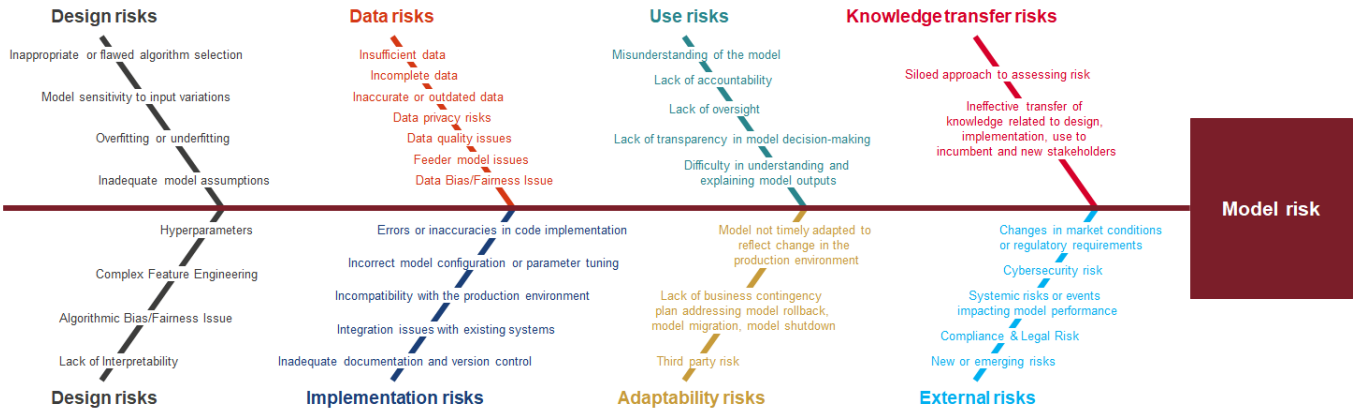
CRISIL believes the most the 'practical' definition of AI model should allow for this segmentation, by calling out the advanced statistical/analytical techniques involving either or a combination of deep learning (including artificial neural network, or ANN), ensemble learning (for e.g., boosting), reinforcement learning, natural language processing. Additionally, it should be inclusive of both, autonomous and semi-autonomous ways of model learning. This bifurcation would be the starting point of tailoring the governance procedures between the two segments.

Model risk assessment: Model risk assessment requires that all sources of risk are identified and assessed for severity and likelihood before assigning a model risk rating.

CRISIL maintain that any approach that does not account for model-lifecycle stages in risk identification has the potential of missing some of the model risk sources, whether in traditional or AI/ML models.

The fishbone chart below gives a high-level view of AI/ML model specific risk sources.

Sources of AI/ML model risks



Model risk classification/tiering comprises a scorecard-based approach to come up the model risk rating. The scorecard methodology evaluates various quantitative and qualitative factors within each criteria. The criteria could relate to model complexity, materiality, and reliance.

The model risk rating is a very important component, as it determines the rigour, sophistication and prioritisation of model risk governance and management activities, including model development, validation, ongoing monitoring and risk reporting.

Like traditional models, AI models contain intrinsic uncertainties that come with data, algorithms, interpretation and use. However, the presence of complex modelling algorithms, increased data dimensionality involving use of alternative datasets, and automation of business processes, not only lead to manifestation of new risk types/sources but also make their identification harder as they may manifest in unfamiliar ways.

CRISIL opines that till a financial institution's AI governance processes are mature enough, existing model risk assessment and classification criteria used for traditional models should be augmented by a separate AI/ML model

risk assessment. Supported by extensive guidance, this separation will allow, without the constraints of traditional risk management, a deeper dive into the risk drivers unique to AI/ML models.

Furthermore, the evaluation can be done from model ideation through model development. Rooting this assessment at the ideation level will serve as a checkpoint to critically assess the use of AI models and their potential benefits and risks, including compliance, operational and reputational risks before initiating the development process. This understanding of risk can also serve to inform the various data pre-processing requirements and identify model testing procedures.

Model risk mitigation: This step involves developing procedures and guidelines to mitigate risk across the model lifecycle. These procedures serve to provide a structured and standardised framework in which consistent and robust practices of data pre-processing, model development, validation and implementation are followed. Doing so reduces the likelihood of errors, biases, or misinterpretations that could lead to adverse outcomes or inaccurate results.

CRISIL suggests applying the following mitigation guidance to address the key sources of AI/ML model risks, will substantially help in establishing trustworthy and responsible AI/ML modelling. Some of these risk areas are inter-related and should be evaluated holistically by respective business and risk management functions, in addition to independent model validation.

Risk types	Mitigation guidance
Complex or black-box models /Interpretability /Explainability	• Set policy that establishes explainability-related assessment procedures in context of model use, materiality, model structure and data availability • Collect justification of each stakeholder who may or may not have a concern regarding explainability • Identify the granularity of explanation needed, i.e., for individual observation or in aggregate • Outline the trade-offs between explainability and model performance • Understand the key characteristics and limitations of the methodologies chosen to generate local/global level explainability.
Bias/Fairness	• Set policy that establishes bias assessment procedures depending on the context of model use, materiality, model structure • Identify all possible outcomes of model use including most adverse outcome to the subject • Recognise the potential consequences of model errors and assess whether these errors could disproportionately affect specific groups more frequently or with greater magnitude. **Data bias** • Reduce data bias via data lineage, data quality and relevance testing. Cover areas of data design, collection, preparation, pre-processing, and analysis • Use descriptive statistics to understand data structure and detect if under-represented groups are present • Employ various statistical tests to identify difference in representation • Test for collinearity between model features and prohibited/sensitive variable that can serve as proxies • Review bias introduced from data exclusion • Review if outliers are more prevalent in certain groups

Risk types	Mitigation guidance
	Algorithm bias • Assess sampling/training bias • Review the chosen cost function and how it may lead to bias • Perform counterfactual fairness testing • Use surrogate models to evaluate degree of differential treatment • Assess biasedness using various measures of disparity impact
Hyper parameters	• Use combination of expert judgement and quantitative approach to set the hyper parameters • Determine the limitations of the hyperparameter tuning method as the curse of dimensionality may render certain tuning methods impractical • Delineate if all hyperparameters are not equally important specific to the ML algorithm and focus on testing the critical ones • Determine the balance between prediction accuracy and model complexity to assess if hyper parameters are prudently configured • Evaluate the adequacy of hyper parameters where transfer learning is used
Feature engineering	• Establish policy that informs the level of support required to establish the conceptual soundness of features in context of model use, materiality, model structure • Assess if features selected for obtaining acceptable model performance possess a meaningful relationship with target variable • Employ various feature predictive power tests, target-feature correlation tests • Assess variable clustering techniques used for variable reduction • Assess dimensional reduction appropriateness via performance on out-of-sample • Assess soundness of ML techniques (many automated) for feature engineering • Assess statistical features (such as TF-IDDF, word embedding, count/density, bag-of-words) related from extracting unstructured data into numerical data

Relevant stakeholders from compliance, operations, legal, and technology departments should be actively consulted to ensure that any reputational risk, compliance risk, third party and cyber security risks related to use of AI/ML models are appropriately owned and investigated.

Model risk monitoring and maintenance: This involves systematically assessing and tracking the model performance and refreshing the model on time to account for any model and data drift. Some categories of AI/ML models necessitate dynamic or highly frequent adjustment of their parameters to effectively capture the emerging patterns in the data.

CRISIL believes the following guidance will help address uncertainty that come from the lack of a clear line separating model recalibration from model change:

- Identify conditions, testing metrics that will trigger ongoing recalibration
- Identify thresholds to limit what qualifies as ongoing recalibration vs model change

- Examine what aspects of the model will be open to change under model recalibration (such as data refresh, statistical or configuration parameters)
- Identify expected frequency of recalibration
- Identify scope of governance, including testing and review as part of model recalibration
- Identify criteria that could trigger legal review

Reporting and governance: Finally, financial institutions can augment their model governance frameworks by implementing various tollgates/checkpoints across the model lifecycle to ensure that AI/ML model risk is appropriately identified, assessed, and approved.

If the model risk appetite metrics are only validation-related, they need to be expanded to include other metrics such as models breaching performance thresholds, models with high operational risk issues such as data privacy issues with overdue attestation, or models that are not on the inventory to name a few. Specific risk limits against these metrics need to reflect the institution's appetite for AI/ML model risk.

Moreover, key people in analytics teams and related risk management roles must be identified, and their roles within the risk management framework and their mandate and responsibilities in relation to AI controls defined.

Ongoing training and guidance in terms of playbooks should also be provided to risk managers to ensure they develop knowledge beyond their previous experience with traditional analytics.

Conclusion

The emergence of AI/ML model risk brings forth unfamiliar challenges, and introduces new sources of risk.

As financial institutions increasingly adopt AI/ML technologies, it is becoming clear that addressing relevant risks will require a structured and disciplined approach, rather than ad hoc moves.

That's why it is crucial to augment and adapt all elements of the Model Risk Management (MRM) framework to ringfence.

In this paper, CRISIL has shared its opinion on how to enhance a bank's MRM framework, enabling the proper identification, assessment, monitoring, and mitigation of AI/ML model risk throughout the model lifecycle.

We believe these reflections could serve as a guide for institutions to navigate the complexities and uncertainties associated with AI/ML models, ensuring the resilience and effectiveness of their risk management practices

About CRISIL Limited

CRISIL is a leading, agile and innovative global analytics company driven by its mission of making markets function better.

It is India's foremost provider of ratings, data, research, analytics and solutions with a strong track record of growth, culture of innovation, and global footprint.

It has delivered independent opinions, actionable insights, and efficient solutions to over 100,000 customers through businesses that operate from India, the US, the UK, Argentina, Poland, China, Hong Kong and Singapore.

It is majority owned by S&P Global Inc, a leading provider of transparent and independent ratings, benchmarks, analytics and data to the capital and commodity markets worldwide.

About CRISIL Global Research & Risk Solutions

Global Research & Risk Solutions is the world's largest and top-ranked provider of high-end research and analytics services. We are the world's largest provider of equity and fixed income research support to banks, and the foremost provider of end-to-end risk and analytics services to trading and risk management functions at world's leading financial institutions. We provide corporate research and analytics solutions to operations, strategy, and sales and marketing teams of corporations globally. Coalition provides analytics and business intelligence to 14 leading global investment banks. We operate from 8 research centers in Argentina, China, India and Poland, working with clients across time zones and languages. Being part of CRISIL enables us to attract and retain top quality talent. We have over 2,300 employees, 75% of whom hold advanced degrees in finance, accounting and management. We employ the largest number of CFAs and CFA aspirants in India. We have won top honours at the World HR Congress on Talent Management and HR Project for the year 2015. We have also won the NASSCOM Exemplary Talent Practices Award (NExT Practices) for skill development for two years in succession in 2011 and 2012. The award recognizes us as a firm that has the vision to proactively invest in its people and get them future-ready.

We are committed to delivering cutting-edge analysis, opinions, and solutions. This underscores our proposition of being the best people to work with.

CRISIL Privacy Notice

CRISIL respects your privacy. We may use your contact information, such as your name, address, and email id to fulfil your request and service your account and to provide you with additional information from CRISIL. For further information on CRISIL's privacy policy please visit www.crisil.com/privacy.