FAIRLY

THE THREE LINES OF DEFENSE AGAINST MODEL RISK

What Organizations Adopting AI Can Learn From Financial Institutions

Roundstone Bay, Galway Ireland, 2004

If a simple modeling error could send a heavy salvage crane crashing to the bottom of Roundstone Bay, imagine the impact unmitigated model risk might have on your business.

TABLE OF CONTENTS:

MODELS ARE GROWING INCREASINGLY COMPLEX

As technology evolves towards ever greater complexity, there is a parallel evolution towards increasingly complex models to help implement and manage the innovative technologies. This progression is hardly limited to financial applications. The number of models developed for finance, aviation, medicine, spaceflight, and transportation—as well as their complexity—increases year after year. Applications of data science and artificial intelligence have enabled many firms to streamline their business models and reduce overhead to increase profitability. But the increased complexity of these methods also introduces an increased level of susceptibility to model risk. It follows that the consequences of unmitigated model risk within model-dependent industries increases correspondingly and at exponential rates due to inter-dependencies between models.

WHAT IS MODEL RISK?

> *Model risk is the potential for adverse consequences from decisions based on incorrect or misused model outputs and reports. Model risk can lead to financial loss, poor business and strategic decision making, or damage to a bank's reputation. – From SR11-7, page 3*

The first necessary step for any firm intent on creating a robust and efficient model risk discipline is to recognize the presence of model risk throughout all of its various lines of business. A second equally important step is to find ways to communicate the importance of model risk to a firm's senior management, especially with respect to how unmitigated model risk can affect a firm's business performance.

Model Risk Arises From Two Primary Sources:

- A model may have fundamental errors that may result in inaccurate outputs when viewed against the design objective and intended business uses.
- A model may be used incorrectly or inappropriately. Even a fundamentally sound model producing accurate outputs consistent with the design objective of the model may exhibit high model risk if it is misapplied or misused.
- As George Box presciently observed in 1987, although model risk can never be completely eliminated, it can be substantially mitigated through the consistent application of sound model risk mitigation best practices.

WHY SHOULD ORGANIZATIONS CARE ABOUT MODEL RISK?

> *Because the very nature of a model is a simplified and idealized representation of something, all models will be wrong in some sense. Models will never be "the truth" if truth means entirely representative of reality.*

- George Box, 1987, from "Empirical Model Building and Response Surfaces".

Complicated computerized models [1] , [2] and their myriad quantitative financial applications are fundamentally important for today's financial services industry. Indeed, no contemporary financial institution with any level of sophistication can perform its business functions effectively without a suite of models, from quantitative investment models required to manage investment portfolios, to models used to underwrite loans or monitor for money laundering or other undesirable types of behavior.

Unfortunately, the benefits that accrue to institutions that employ a portfolio of quantitative models also come with an unavoidable downside in the form of several types of risk that can be collectively categorized under the rubric of **"model risk."** Organizations from other industries can learn from these financial institutions.

[1] What is a model? For the purposes of this document, the term 'model' will refer to any type of quantitative methodology or algorithm, implemented in computer software, that applies statistical, economic, financial, or mathematical theories, techniques and assumptions to convert quantitative and/or qualitative input data into useful quantitative estimates. - paraphrased from SR11-7, page 3

[2] "A model consists of three components: an information input component, which delivers assumptions and data to the model; a processing component, which transforms inputs into estimates; and a reporting component, which translates the estimates into useful business information." – SR11-7, page 3

THE THREE LINES OF DEFENSE (LODS) AGAINST MODEL RISK ACCORDING TO SR11-7/OCC2011-12

> *The biggest liability is a failure of imagination*
> *– Senior Model Risk Analyst, US Federal Reserve Bank*

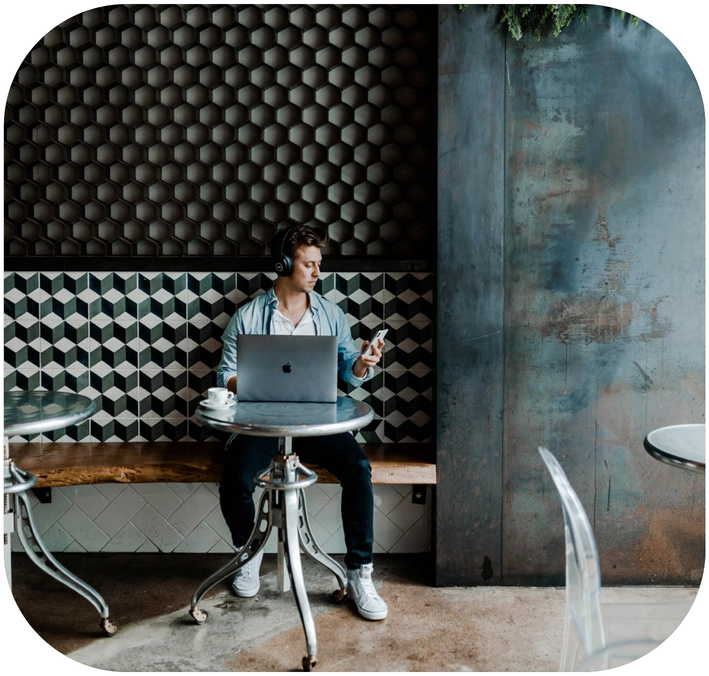
There is likely no other discipline for which the need for tools to mitigate model risk seems more apparent than in financial applications, as modern finance is almost entirely dependent on an ever-increasing array of complex quantitative models. [3]

Nearly all financial institutions doing business in the United States strive to attain compliance with the requirements of the ground-breaking baseline guidance for identifying and managing a rigorous model risk discipline issued jointly by the Federal Reserve Bank and the Office of the Comptroller of the Currency in April of 2011 as SR11-7 and OCC 2011-12 (hereafter referred to as SR11-7). This 21-page document was an immediate game-changer that both raised the bar and set the current standard for effective Model Risk Management (hereafter abbreviated as MRM).

Because SR11-7 is thoroughly comprehensive and well-written, it has also been adopted as a de facto standard for rigorous MRM by most European, British and Asian financial institutions.

[3] And nowhere within the universe of quantitative financial models currently in vogue are model-associated risks more challenging than for those that are based on artificial intelligence (AI) and machine learning (ML) methodologies.

THE SR11-7 REGULATORY GUIDANCE FOR MRM PRESCRIBED THREE LINES OF DEFENSE AGAINST MODEL RISK



SR11-7 identified and clearly described for the first time a requirement for three independent Lines of Defense (LOD) against model risk to be implemented and maintained by conforming financial institutions. Nearly all conforming institutions have accomplished through a three-tier vertical hierarchy of LODs.

Although the phrase 'Line of Defense' does not actually appear in SR11-7, it is unambiguously implied by the document's organization through the description of requirements for model development, model validation and model audit functions in separate sections of the document. In practice, the acronym LOD has become a kind of lingua franca among financial institutions to reference any one of the vertical layers of a 3-tier infrastructure for managing model risk as described in the SR11-7 guidance.

The following are verbatim excerpts from SR11-7 that broadly outline the three LODs against model risk:

1st LOD

Model identification, development and implementation

"Model risk management should include disciplined and knowledgeable development and implementation processes that are consistent with the situation and goals of the model user and with bank policy. …… An effective development process begins with a clear statement of purpose to ensure that model development is aligned with the intended use. The design, theory, and logic underlying the model should be well documented and generally supported by published research and sound industry practice." – page 5, SR11-7

2st LOD

Model validation and ongoing monitoring, independent of the 1st LOD

"Model validation is the set of processes and activities intended to verify that models are performing as expected, in line with their design objectives and business uses. Effective validation helps ensure that models are sound. It also identifies potential limitations and assumptions and assesses their possible impact. As with other aspects of effective challenge, model validation should be performed by staff with appropriate incentives, competence, and influence."

"Staff doing validation should have the requisite knowledge, skills, and expertise. A high level of technical expertise may be needed because of the complexity of many models, both in structure and in application."

"Staff conducting validation work should have explicit authority to challenge developers and users and to elevate their findings, including issues and deficiencies."

– excerpts from Section V of SR11-7

3rd LOD

Internal Audit, responsible for independent oversight of the 1st and 2nd LODs.

"A bank's internal audit function should assess the overall effectiveness of the model risk management framework, including the framework's ability to address both types of model risk (model errors; inappropriate use of models), for individual models and in the aggregate.

Internal audit also has an important role in ensuring that validation work is conducted properly, and that appropriate effective challenge is being carried out. It should evaluate the objectivity, competence, and organizational standing of the key validation participants, with the ultimate goal of ascertaining whether those participants have the right incentives to discover and report deficiencies.

Internal audit should verify that acceptable policies are in place and that model owners and control groups comply with those policies. Internal audit should also verify records of model use and validation to test whether validations are performed in a timely manner and whether models are subject to controls that appropriately account for any weaknesses in validation activities. "- page 19, SR11-7

Note that as a baseline document for model risk management in finance, SR11-7 is descriptive rather than prescriptive. SR11-7 explains what is expected of conforming financial institutions in order to achieve a rigorous model risk discipline without specifying how financial institutions should go about implementing the guidelines. This approach gives firms significant latitude in designing a model risk infrastructure that is compatible with their internal systems and data resources.

In finance, the cost of model risk is primarily measured in financial loss and reputational harm. In other industries unmitigated model risk may also be measured in lives lost (witness the 2017-18 Boeing Max 737 MAX 8 disasters).

Model risk is particularly punitive when associated with any form of bias due to the heightened levels of visibility generated by such occurrences. This has been especially costly for some financial institutions in recent years when biases in lending practices were exposed and publicized by regulators and investigative journalism. AI/ML models that were employed by a large commercial bank to assess credit worthiness of applicants were to have used training data that reflected biases against applicants living in area codes that had higher than average incidences of default as well as applicants with Hispanic-sounding surnames. Public furor over the biases resulted in substantial regulatory fines levied against the offending institutions, forced resignations or dismissals of senior managers, including several CEOs, and long-lasting damage to the bank's reputation.

The risk arising from flawed or misused models is far from theoretical. In finance the damage caused by model risk is financially punitive but can potentially be remediated. In other industries, especially those involving travel such as aviation, nautical and transportation, poorly managed model risk can result in the loss of irreplaceable human lives. For these industries, an effective MRM is even more critical, yet they tend to lag behind the financial industry, which acquired some very painful and bitter lessons about model risk during the global financial crises of 1987 and 2008, and more recently in 2012.[4]

THE RISK OF RISK MODELS

In order to effectively identify and manage risks, firms often create models of risks. But what happens when those very models turn out to be flawed?

This is a key part of what occurred at J.P. Morgan in 2012, in a financial debacle that has become known throughout the financial industry simply as "The London Whale". At the center of the 2012 London Whale was a seriously flawed risk model, one that was required by regulators to address deficiencies in risk models exposed by the 2008 global market crisis. The model, known as the

[4] We note as an aside that nonfinancial industries would do well to take advantage of the hard-won lessons of financial firms by adopting their advanced MRM practices - practices whose development has been guided by painful experiences in order to mitigate and manage model risk to the greatest extent possible.

Comprehensive Risk Measure, or CRM, was a type of Value-at-Risk, or VaR model [5] intended to address correlation risks. Correlation risks feature prominently among the blind spots in historical VaR models exposed by the 2008 global crisis. A flawed implementation of the CRM risk model, compounded by dysfunctional management and unrestrained trader overreach in JPM's Central Investment Office (CIO), resulted in the realization of 6.2 billion USD in trading losses and another 900 million in regulatory fines that were subsequently assessed against the firm.

Based on guidance from the Federal Reserve, the FDIC and other regulators, financial service firms have developed a variety of effective tools to identify, measure and mitigate most model risks to the greatest extent possible, as noted in the previous sections describing the three Lines of Defense against model risk. But the regulatory guidance predates the Artificial Intelligence (AI) renaissance, and with the advance of big data, artificial intelligence and machine learning, potential model risks increase, and the controls needed to manage those risks to comply with regulatory and contractual obligations deserve additional attention.

For example, the Federal Reserve's SR11-7 baseline guidance for model risk management first introduced a concept called "posing effective challenge" to the model, which requires critical analysis by objective, and informed parties who can identify model limitations and implement appropriate changes.

Such effective challenge should include (among many other possibilities):

- testing the theory and logic underlying the model design
- validating the model as well as the integrity of data it uses
- testing the performance of the model over a range of inputs
- assessing performance of the model with severely stressed inputs
- implementing a governance model that permits independent review and assessment.

[5] Value at risk models are commonly used by financial institutions to estimate limits on the amount of money that a firm might lose over a given period of time (usually one year) at a specified confidence level, typically 99%. A daily VaR of 100 million at 99% confidence would estimate that on 99 days out of 100, the losses would be less than 100 million. VaR estimates are commonly based on the statistical properties of a firm's daily P&L (profit and loss) history over a period of between one and 5 years. One of the characteristic weaknesses of the VaR methodology is the implicit assumption that future (i.e., tomorrow's) losses can be estimated from past historical data. This assumption tends to fail badly during market meltdowns. A second weakness is that a 99% daily VaR can place no limits on how much money might be lost on the 1 day out of a 100 that the VaR limit is breached, due to insufficient relevant data to estimate losses accurately beyond the 99th percentile.

Under an AI regime in which models work by identifying patterns in large data sets and making decisions based on those patterns, replication of the model's output (let alone reviewing performance across a range of inputs) becomes far more difficult. Further, when AI models apply ML paradigms to very large data sets, often from multiple sources, validating the integrity of the many alternate data becomes exponentially more challenging.

A common dilemma that results if a model's output is generated by a vendor's 'black box'[6] model is that the ability of independent reviewers to effectively assess the accuracy of model output becomes substantially more limited. The black box validation paradox is one that can bedevil even the most proficient validation teams as model vendors loath to explain the inner workings of their models due to the risk of giving away intellectual property (IP). These difficulties are also common to the use of AI/ML models due to their well-known lack of transparency and consequent difficulties in unraveling and explaining their operation.

Regardless of industry, the key questions all firms that rely on models should consider are how to determine:

- whether any defects in the model design exist
- whether any model defect resulted in adverse decisions
- which party—among the model developer (business owners, model developers and supervisors), model risk managers (model validation), and the model's end users — will assume the risk of any errors or defects
- finally, the amount of financial and/or reputational damage resulting from model errors and defects, or misapplication of the model for purposes beyond the intent of the model developers.

These are a minimal set of core model risk concerns that are common to any class of model, whether they are realized by traditional hand-written code or AI/ML methodologies. It is imperative for highly model-dependent firms to develop an effective MRM infrastructure that is capable of addressing these core concerns at all three LODs in order to confidently conduct their business.

Which leads quite naturally to the next question:

[6] 'Black Box' is a commonly used metaphor that refers to the opacity of models whose inner workings are unknown to users. Examples are 3rd party models provided by vendors who are protective of their intellectual property, and nearly all AI/ML models.

WHO OWNS MODEL RISK IN A MACHINE LEARNING WORLD?

The simple short answer to this often vexing question is: everyone who interacts with models owns at least some part of the risk associated with using them: developers, validators and users.[7] Or, in other words, everyone who is involved with the three LODs, as described previously, are the joint owners of model risk. Determination of ownership is critically important if model defects (owner responsibility) or model misapplication (user responsibility) result in financial loss or reputational damage to a firm.

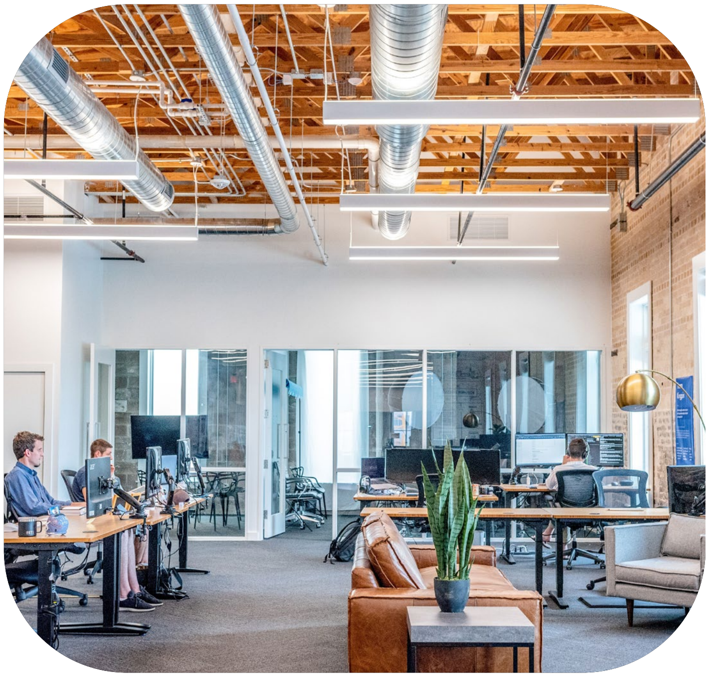
At leading financial institutions, the widespread introduction of AI/ML models occurred well after SR11-7 was issued in April of 2011, beginning somewhere around 2015. While the basic concepts and requirements for a sound model risk discipline still apply to this new class of models, the requirements for model identification, development, testing, validation and implementation into production are specific to this class of model and present an additional bespoke set of challenges for all three LODs against AI/ML model risk.

Model risk

can never be completely eliminated because, as noted in George Box's comment at the beginning of this document, all models are eventually wrong at different levels of precision. This truth is especially relevant in finance since, unlike engineering or the sciences, financial models are not derived from first principles of nature, such as Isaac Newton's laws of motion or the laws of thermodynamics. However, while model risk cannot be completely eliminated because of the very nature of models, it can be mitigated and managed under a rigorous model risk discipline.

[7] This is similar to the question of who owns the risk for operating a motor vehicle? The answer determined over many years of lawsuits and trials in US courts is this: everyone involved in the creation, maintenance and operation of a motor vehicle. The manufacturer (the developers), the dealers who sell and perform repair and maintenance work (the validators) and the individuals who choose to drive a vehicle on public roads (the users).

AI/ML MODELS CAN ALSO HARBOR DEFECTS



If AI or ML models are applied to business needs and an adverse result arises—such as the poor performance of a loan or investment portfolio—the first question to answer is whether the model itself was flawed, whether it received faulty input data, or if it was simply an incorrect application of a valid model. Model outcomes may out-perform or under-perform relative to a benchmark and yet still be operating exactly as intended by the model's developers.

In some instances, model defects may be readily verifiable such as computational errors, use of incorrect variables, violation of the model's underlying assumptions[8] or the use of incorrect units of input data. In other instances, AI/ML defects may result from a misinterpretation of underlying data, or reliance on coincidental correlations that appear in data but without causal connection. These errors may be much more difficult to detect and diagnose. In still other instances, a model developer may make certain simplifying underlying assumptions [8] that may directly impact the model's performance. Such simplifying assumptions are the unavoidable core part of the combined art and science of attempting to model reality.

A key distinguishing feature of any successful model, whether it may be a traditional hand-tooled 'quant model' or an AI/ML model, is that in order to be useful it must be able to ***reduce complexity.*** All successful (i.e. useful) models must share this singular common function. But in order to reduce complexity, all useful models must rely on underlying simplifying assumptions. There is no other practical way to reduce complexity. Whether or not a model's simplifying assumptions will result in a "defect" or "error" may depend significantly on the representations made about the model and the context in which it is applied. A model that works perfectly well under one regime may fail when applied to another regime that, at least superficially, seems to be similar. The long history of model risk is replete with many such examples.[9]

[8] Some classic risky examples are: history is the best predictor of future outcomes; sovereign bonds will never default; interest rates will always be positive; housing prices will always increase; correlations will remain stable in an economic meltdown

[9] Visit this YouTube link for a seminar on The History of Model Risk recorded by Dr. Jon Hill: (16) MRMIA Broadcast #6: History of Model Risk with Jon Hill, Ph. D. - YouTube

AI/ML DATA QUALITY

Because AI/ML models share a propensity for acquiring data from many alternative resources of varying quality and reliability, the importance of examining and assessing the quality of data received from alternative sources is even greater than it is for traditional quant models. Unrestrained use of alternative data of questionable quality potentially compromises any results produced by AI/ML models. As a result, the growing practice of incorporating alternative data is currently an area of special focus by bank examiners from federal regulatory agencies.

During a financial regulatory bank exam on model risk practices, all three LODs should expect intense scrutiny of the exercise of their responsibilities vis a vis AI/ML data quality:

- 1st LOD developers must be able to demonstrate that due diligence reviews have been performed on all of the alternative data sources used by their AI/ML models and attest to their sufficiency to support their models.
- 2nd LOD model risk managers must be able to demonstrate that they have independently reviewed and challenged the data quality assessments performed by the developers.
- 3rd LOD internal auditors must be able to demonstrate that they have independently reviewed and challenged the data quality assessments and attestations performed by the 1st and 2nd LODs.

CHIEF RESPONSIBILITIES OF THE LODS

The primary responsibilities of the three LODs[10] as outlined in SR11-7 are typically assigned in leading financial institutions in the following manner:

First LOD

model identification, development and acceptance testing within any business unit (BU) that relies on models. First LOD personnel include model supervisors, model developers, owners, and users.

Second LOD

model risk management, model risk governance including developing policies and procedures, validations, assigning model risk ratings, ongoing monitoring and model inventory management. Personnel are chartered as model risk managers. [11]

Third LOD

Internal Audit, which is tasked with ensuring that the first and second LODs within a firm verifies that appropriate model policies are in place and that both first and second LODs comply with their relevant policies and responsibilities. Personnel are chartered as internal auditors.

[10] Refer to Appendix A for a more detailed description of the expectations and responsibilities of each LOD

[11] Refer to Appendix B for a description of an effective SR11-7 compliant playbook approach to establishing a rigorous second LOD MRM discipline within a financial firm

WHERE SHOULD THE THREE LODS AGAINST MODEL RISK RESIDE WITHIN THE CORPORATE HIERARCHY?

Since regulatory guidance mandates that each LOD be as independent as possible within the corporate entity, the answer to the question of residence may vary from firm to firm depending on how it structures its vertical MRM platforms. Here are the most common configurations that leading financial firms have employed:

First LOD

The first LOD will invariably be found within a firms' business units (BU) as the BUs are responsible for identifying model requirements (what should this model do, what input data would it require, etc.), assigning resources to research different modeling approaches, to prototype and eventually develop software to implement the model. The first LOD would also typically have responsibility for developing Challenger Models as benchmarks to assess the performance and accuracy of the Champion Models intended for production.

As an aside, most model developers build their models using higher-level languages like Python, JAVA, VBA, Fincad, S+ or SAS. After a model has been validated by the second LOD, it will typically be converted into more efficient production code such as C or C++ by Information Technology (IT) staff within a BU, who will bear responsibility for putting the model into a full production platform and monitoring its performance. The management of a BU's first LOD should report into the senior management of the BU.

Model users are also considered part of the first LOD as they will typically work within the same BU as the model owners and developers. However, it is also possible and perfectly acceptable for model users to be part of other BUs within the firm.

Second LOD

The second LOD is typically called Model Validation or, more comprehensively, Model Risk Management (MRM) at most financial firms although this latter name is a bit of a misnomer as all three LODs have roles to play in the firmwide management and mitigation of model risk. Second LOD MRM will typically include model governance as well as validation.[12] The placement of the second LOD within a financial firm's corporate structure has some variability from firm to firm as there is no clear mandate from SR11-7 about lines of ownership. In most financial firms, the second LOD MRM

[12] For a detailed discussion of the differences between MRM, model validation and model risk governance, see the references section of this document: Hill, The Top Fourteen Challenges for Today's Model Risk Managers, 2019.

function will reside somewhere within the corporate Risk Management division, ultimately reporting into the firm's Chief Risk Officer (CRO). But since a firm's risk models are also owned and developed with the CRO's organization, great care must be taken to ensure adequate independence between model validation and model development. The degree of independence between the three LODs is routinely assessed by regulators during bank exams on model risk discipline.

In some firms MRM may coexist with model development groups within the same risk management organization, separated by divergent lines of reporting into different managers who then report into the CRO. This arrangement, though common, may not provide sufficient independence to satisfy regulatory reviews for SR11-7 compliance, which is a concern that should be addressed by management. One way for a firm to ensure greater independence of MRM from risk model development with the CRO's organization would be to appoint a Chief Model Risk 0fficer (CRMO) who reports directly into the CRO. Under this structure, all second LOD model risk staff would report into the CRMO. This is currently a common solution to the independence issue adopted by many leading financial institutions.

Third LOD

The third LOD against model risk is also known as Internal Audit (IA). This is the most clearly differentiated LOD in terms of corporate residence as IA departments are tasked with reviewing every part of a financial firm's business and therefore must function independently of all of them. IA should be a completely separate department without sharing reporting lines into the management of any other operational departments that comprise a financial firm's business units.

AI/ML MODELS MAY WARRANT SPECIAL CONSIDERATION

A few of the more common challenges encountered by the three LODs during the development, validation and auditing of AI/ML models are:

- the lack of transparency and the resulting difficulties in explaining the relationship between input and output
- model stability under changing conditions
- availability of fair, unbiased and trustworthy training datasets
- the capacity of AI/ML models for incorporating information from multiple very large and varied datasets, especially when alternative data sources with unestablished lineage or reliability are sourced for training the models.

AN OVERVIEW OF INTERNAL AUDIT'S ROLES AND RESPONSIBILITIES

Scope

The scope of Internal Audit (IA) within a financial firm is often quite broad. At a high level the scope can include topics such as governance, risk management (including model risk), management controls over the effectiveness and efficiency of operations, compliance with federal regulations and the quality and accuracy of financial and managerial reporting.

Due to the SR11-7 strict requirement for independence, Internal Auditors should never be involved in the execution of company activities. The main responsibility of IA is to advise management and the Board of Directors (BOD), and/or other oversight bodies, on how to improve upon execution of their responsibilities. Because of the breadth of the scope of its involvement, IA staff will typically have a variety of professional backgrounds and have attained higher levels of graduate education, achieving masters and even doctorate degrees.

REQUIREMENTS FOR INDEPENDENCE

Professional internal auditors are governed by the standards of the Institute of Internal Auditors (IIA), the international standard setting body for the internal audit profession. The IIA standards mandate IA staff to be as completely independent of the business activities they audit as possible, even though they are employees of the same company (unless external auditors are engaged for certain types of reviews). Independence and objectivity are the mainstays of the IIA standards are independence and objectivity. These mandates are discussed in detail in the standard and the supporting practice guides and advisories.

Even though employed by the same company as the subjects of their audits, IA can maintain independence through its organizational placement and carefully designed reporting lines. In the US, internal auditors of publicly traded companies are required to report functionally to the Board of Directors (BOD), most often through an audit subcommittee of the BOD. To maintain as much independence as possible, IA should never report into C-suite management except for administrative purposes.

CORE RESPONSIBILITIES OF INTERNAL AUDIT

The primary customer of all IA activity is the audit committee, which is the entity responsible for oversight of management activities. The audit committee is typically a sub-committee of the BOD. Thus, organizational independence is achieved by a relationship that ensures the head of IA reports functionally into the BOD. Some examples of functional reporting to the BOD includes review and approval of:

- The IA charter
- A risk-based IA plan
- Adequacy of IA resources and budget
- Line of reporting to receive communications from the head of IA on IA's execution of the plan
- Compensation of the head of IA and other senior IA staff
- Appointment or removal of the head of IA
- Audit reports, findings and recommendations for improvement of business activities

IA activity is primarily directed at evaluating internal controls.

MANAGEMENT, INTERNAL AUDIT, AND INTERNAL CONTROLS

An internal control is broadly defined as a process performed by a firm's BOD, management and other personnel that is designed to provide assurance that the following core objectives are achieved by all of the firm's business units.

- Effectiveness and efficiency of business operations
- Reliability of financial and management reporting
- Compliance with regulatory requirements and applicable law
- Protection of assets

Management is responsible for internal controls which are comprised of the following five critical components:

- The control environment
- Risk assessment
- Risk-focused control activities
- Information and communication
- Monitoring activities

Managers establish policies, processes and practices within these five areas of management control to help the organization achieve the four specific objectives listed above.

WHAT IS INTERNAL AUDIT'S ROLE IN MODEL RISK MANAGEMENT?

Risk management is the process by which an organization strives to identify, analyze, monitor and gather information about strategic risks that could impact the organization's ability to execute its business plans and achieve its related objectives. IIA standards require the IA function to evaluate the effectiveness of the organization's risk management activities. To accomplish this, IA performs audits to evaluate whether the five components of management control are present and operating effectively within the 1st and 2nd LODs, as well as to provide recommendations to address and improve upon any deficiencies identified.

The Internal Audit function is primarily responsible for independently testing all of management's control assertions for both first and second LODs, and for reporting its findings to the audit subcommittee of the BOD. This requirement applies as much to IA audits of AI/ML models as it does to traditional quant models.

A sound MRM discipline is indisputably the most effective (and actually only) defense against the types of modeling errors described throughout this document. History has demonstrated repeatedly that failures of MRM can result in intense media coverage and inflict serious reputational damage on a firm (witness the 2012 London Whale scandal at JPM or the two Boeing 737 MAX disasters in 2018 and 2019), while little or no credit is awarded for modeling disasters that have been prevented by and effective MRM culture.

A Well-Designed AI/ML Model Validation platform would offer client firms a shortcut to achieving a complete AI/ML model risk solution by applying the methodologies of ML to substantially reduce the manual overhead required for a rigorous SR11-7 compliant validation. Some of the more time-consuming aspects of model risk management as described above that are immediately amenable to ML automation are:

1. Assessment of model documentation for clarity, completeness and accuracy using AI based Natural Language Processing (NLP)
2. Assessment and/or remediation of input data quality and suitability for the model's intended application(s)
3. Generation of a suite of validation benchmarks to assess the model's performance, accuracy, and fitness for purpose
4. Generation of suites of tests to determine the accuracy of a model's output
5. Generation of suites of stress tests to determine a model's stability over a wide range of input values and determine any modes of failure
6. ML models may also be trained to execute the test suites outlined in #4 and #5 above and to analyze the results

THE ROAD AHEAD FOR MODEL RISK MANAGEMENT

> **Imagination is more important than knowledge**
> **– *Albert Einstein***

MRM practices have effectively remained unchanged over the last 15 years, with many processes that are still manual such as inventory attestation. This period of stagnation promises to soon come to an end. Imagination is beginning to transform the way that MRM will be practiced at leading firms over a five-year horizon due to the convergence of Big Data, Machine Learning (ML) and Blockchain (BC) digitization and commoditization of models. The day-to-day activities of MRM managers in the near future will likely be substantially different from today's standards due to the above-mentioned disruptions that are already starting to impact the profession.

Leading edge financial firms are already well along in applying AI/ML methods both to model development and MRM. AI/ML models have a vast capacity for assimilating large amounts of data, breaking "the curse of dimensionality" that limits traditional quant models. With increased amounts of data, along with increasing use of 'alternative' data sources, there is necessarily an increase in data risk due to incomplete, erroneous or unreliable data. AI/ML models are also much less transparent than traditional quant models and must often be treated as 'black box' models. One somewhat surprising twist at a few imaginative financial institutions is the use of ML models to validate other ML models, a clever way to circumvent the black box validation conundrum. One example is the use of ML adversarial models to pose effective challenge to detect and mitigate bias.

FAIRNESS, BIAS MITIGATION AND TRUSTWORTHINESS IN AI/ML APPLICATIONS

Certain concepts, expressed by terms such as fairness, bias, or trustworthiness, play a more significant role in the types of AI/ML models often used in finance than they do in traditional quant pricing or risk models. This is partly because many of the current financial applications of AI/ML involve assessments of the credit worthiness (i.e., probability of default (PD), loss given default (LGD)) of loan and credit card applicants. AI/ML models are also heavily used for Anti Money Laundering (AML) and fraudulent transaction monitoring applications.

The training datasets used for these applications may include, for example, embedded correlations between PD/LGD and ethnicity (by name association) or income levels that may be associated with different ZIP codes. ML programs possess a much greater capacity to consume large and seemingly disparate alternative data sources, analyze them and identify many types of subtle and unexpected correlations that may not be apparent to human reviewers. As a result, the elimination of the various forms of bias that are known to appear in ML financial models has recently become a current area of focus in ML model design, implementation, and validation. The goal is to improve both the trustworthiness and fairness of ML models used in finance.

Fairness is a multifaceted, context-dependent social concept that does not lend itself to a single, simple definition. For the purposes of this paper, we will use the term ‘fairness’ as being equivalent to ‘unbiased’.

Bias mitigation algorithms attempt to improve fairness metrics by modifying the training data, the learning algorithm, or the predictions. These algorithms fall into three general categories: pre-processing, in-processing, and post-processing.

Bias can be measured and mitigated at different points in a machine learning pipeline: either on the training data, the learned model, or the model output. These three opportunity areas for bias measurement and mitigation correspond to the pre-processing, in-processing, and post-processing categories of bias mitigation algorithms.

FINAL THOUGHTS

VALUE ADDED BY AN EFFECTIVE MRM AI/ML FRAMEWORK BENEFITS AN ENTIRE FIRM

Forward-looking model risk managers who recognize the likelihood of pending disruptions and strive to be thought leaders in their field will find themselves positioned to play important roles in the evolution of AI/ML models and the MRM discipline required for managing the associated risks over the next five years. Firms that employ these thought leaders will benefit from an improved model ecosystem that requires far less manual overhead, and in this way help to improve their firms' profitability. The most important requirement for these projections to become reality is having the courage to pursue imaginative ideas.

Such ideas involving the use of machine learning, smart models, digitization, and commoditization of a firm's models on blockchain platforms can open up new lines of revenue for firms that have invested resources in model development.

When all is said and done, the single most important question that remains is: what is the actual value added to a corporation that makes the substantial investment required to design and implement an effective MRM AI/ML framework? Here are four benefits that will be realized by any firm willing to make such an investment in a rigorous MRM discipline:

1. **An effective MRM framework enabled with AI/ML will reduce long-term inherent model risk.** Reduced model risk will always translate in the long run to reduced overhead and increased profitability.

2. **An effective AI/ML enabled model development and validation MRM framework will reduce manual overhead, increase accuracy and substantially reduce time to market.** These advantages will be reflected in a practitioner's firm reduced headcount and costs for both model development and validation.

3. **An effective AI/ML enabled MRM framework will reduce regulatory risk.** Among other consequences, regulatory sanctions can result in fines, reputational damage, and prohibition from paying dividends or stock buybacks.

4. An MRM framework enhanced by AI/ML will reduce reputational risk.

Long term reputational damage is perhaps the most serious of the many consequences of a failed MRM discipline. Many leading firms have suffered extensive and financially punitive reputational damage as a direct result of a flawed MRM discipline. The two most prominent recent examples, the 2012 London Whale and the 2018-19 Boeing MAX 737 disasters lead to regulatory sanctions, multi-billions of dollars in fines and business losses and severe reputational damage that, in the case of Boeing, will translate to long-term enduring financial losses.

Finally, there are three primary corporate culture mandates that must be realized in order for a firm to allocate the necessary funding and undertake the effort required to build out an effective MRM for AI/ML discipline:

1. **Persuasive MRM managers** who are thought leaders in the field, understand the vast potential of incorporating AI/ML models into their business. They must also have the ability to communicate the value added and the requirements for customized risk mitigation of these leading-edge models to senior management.
2. **A core group of farsighted senior managers** who can look beyond the next quarter's financial statement to realize the value of long-term returns on a worthwhile investment.
3. **A corporate culture** that is, at a minimum, tolerant of change and open to the application of advanced AI/ML models to their business needs and cognizant of the MRM discipline necessary to mitigate the risks associated with them.

A firm that has the determination, foresight, and resources necessary to bring all of these corporate stars into alignment will be well on its way to achieving an industry best practice for a leading-edge AI/ML corporate framework.

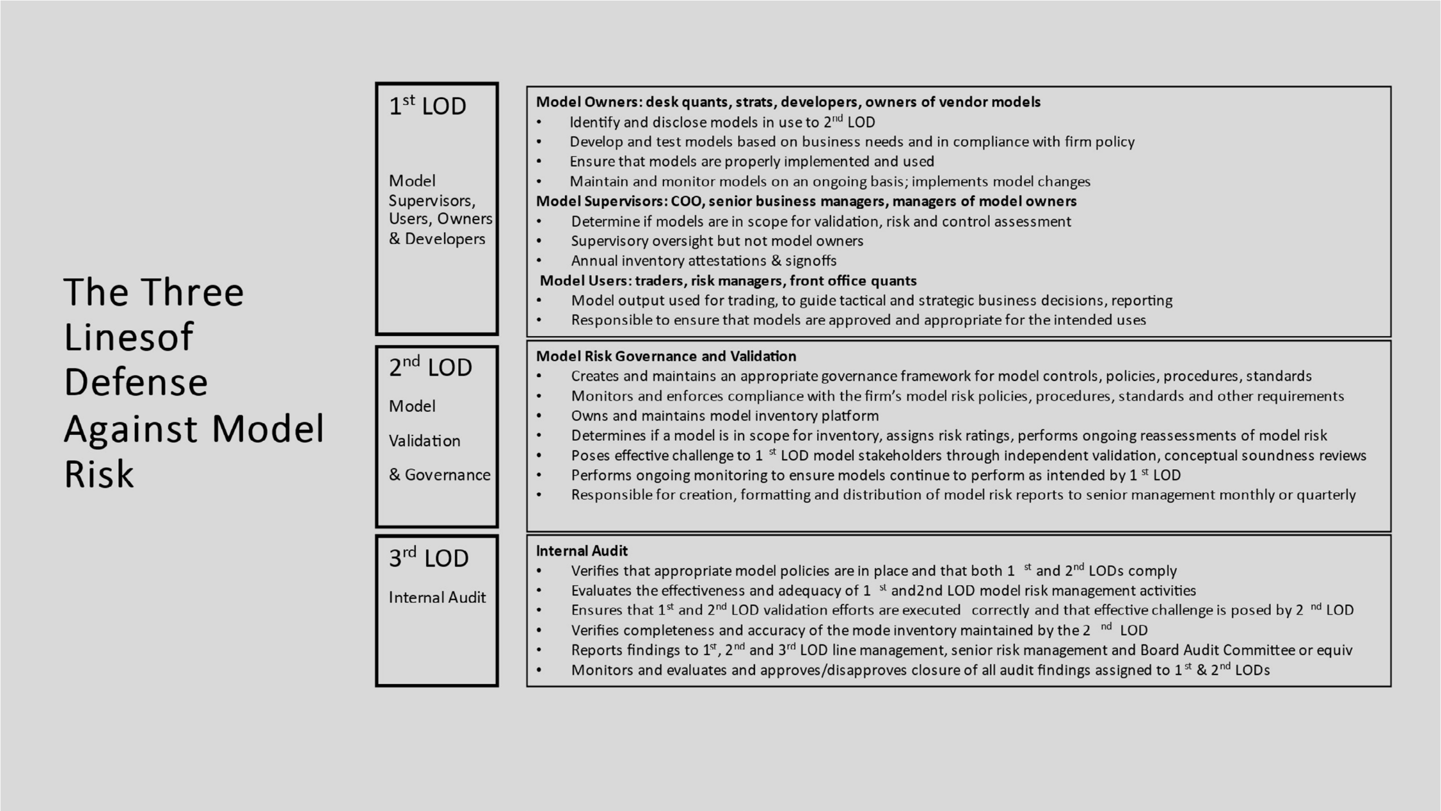
PUBLICATIONS CITED

Hill, J. R. (2019). The Top Fourteen Challenges for Today's Model Risk Managers. *Journal of Risk Management in Financial Institutions Vol. 12, 2, 146-167.*

Hill, J. R. (2020). A Smarter Model Risk Discipline Will Follow from Building Smarter Models. *Journal of Risk Management in Financial Institutions Vol. 13 1, 24-34*.

APPENDIX A

DETAILED EXPECTATIONS AND RESPONSIBILITIES OF EACH LOD ARE DESCRIBED IN TABLE 1 BELOW:



The Three Linesof Defense Against Model Risk	1st LOD Model Supervisors, Users, Owners & Developers	**Model Owners: desk quants, strats, developers, owners of vendor models** • Identify and disclose models in use to 2nd LOD • Develop and test models based on business needs and in compliance with firm policy • Ensure that models are properly implemented and used • Maintain and monitor models on an ongoing basis; implements model changes **Model Supervisors: COO, senior business managers, managers of model owners** • Determine if models are in scope for validation, risk and control assessment • Supervisory oversight but not model owners • Annual inventory attestations & signoffs **Model Users: traders, risk managers, front office quants** • Model output used for trading, to guide tactical and strategic business decisions, reporting • Responsible to ensure that models are approved and appropriate for the intended uses
	2nd LOD Model Validation & Governance	**Model Risk Governance and Validation** • Creates and maintains an appropriate governance framework for model controls, policies, procedures, standards • Monitors and enforces compliance with the firm's model risk policies, procedures, standards and other requirements • Owns and maintains model inventory platform • Determines if a model is in scope for inventory, assigns risk ratings, performs ongoing reassessments of model risk • Poses effective challenge to 1st LOD model stakeholders through independent validation, conceptual soundness reviews • Performs ongoing monitoring to ensure models continue to perform as intended by 1st LOD • Responsible for creation, formatting and distribution of model risk reports to senior management monthly or quarterly
	3rd LOD Internal Audit	**Internal Audit** • Verifies that appropriate model policies are in place and that both 1st and 2nd LODs comply • Evaluates the effectiveness and adequacy of 1st and2nd LOD model risk management activities • Ensures that 1st and 2nd LOD validation efforts are executed correctly and that effective challenge is posed by 2nd LOD • Verifies completeness and accuracy of the mode inventory maintained by the 2nd LOD • Reports findings to 1st, 2nd and 3rd LOD line management, senior risk management and Board Audit Committee or equiv • Monitors and evaluates and approves/disapproves closure of all audit findings assigned to 1st & 2nd LODs

Table 1: A detailed summary of the responsibilities and expectations for each LOD against model risk as described in SR11-7/OCC2011-12

APPENDIX B

AN SR11-7 COMPLIANT PLAYBOOK FOR ESTABLISHING AND EXECUTING A RIGOROUS MRM DISCIPLINE

1. Establish an MRM framework across the firm's entire model ecosystem covering the entire model lifecycle.
2. Establish rigorous model governance, validation[13] and inventory processes.
    a. SR11-7 from the FRB can serve as a useful guide
    b. Standardized templates for both model and validation reports
    c. Special attention should be given to the application of AI/ML models by both first and second Lines of Defense (model development and MRM), given the issues of large amounts of data used by such models and the lack of transparency into their operation.
3. Develop a process for classifying quantitative software implementations as true models, near-models, or tools.
    a. All true models should be assigned unique, non-recycled model IDs
    b. This can be extended to near-models and quantitative tools
4. Develop metrics for estimating model risk.
    a. Most firms classify models into one of three risk buckets: high, medium, and low as a function of complexity, materiality, and model ecosystem inter-dependency (sometimes called prevalence).
5. Develop methods for aggregating model risk to portfolio, business unit and firmwide levels and for creating meaningful model risk reports for consumption by the Board Risk Committee and senior management.
    a. This requires summarizing firmwide model risk in ways that are easily assimilated by senior managers who may have little time for technical details. Model risk heat maps that use color coding to identify concentrations of model risk are one such management-friendly form. (Hill, The Top Fourteen Challenges for Today's Model Risk Managers, 2019)
6. Financial institutions should strive to raise awareness of model risk to the same level of visibility as the other traditional legs of risk management (i.e., market, credit, operational and liquidity risk).
7. Three important resource challenges that must be met for a fully compliant rollout of an effective MRM infrastructure:
    a. Obtaining the necessary data for effective model validation
    b. Obtaining qualified MRM staffing resources in the second and third LODs to implement the steps listed above over a reasonable time horizon.
    c. Obtaining qualified IT staffing support to create and implement a technology platform that can support the MRM infrastructure.

For a more detailed and comprehensive description of the requirements for an effective and forward-looking MRM discipline and learn how the FAIRLY solution can help, please contact info@fairly.ai.

ABOUT FAIRLY

FAIRLY is an award-winning on-demand AI Audit platform on a mission to accelerate the broad use of fair and responsible AI by helping organizations bring safer, faster and compliant AI models to market. FAIRLY makes it easy to apply policies and controls early in the development process and adhere to them throughout the entire model lifecycle. Our automation platform decreases subjectivity, giving technical and non-technical users the tools they need to meet policy and audit requirements while providing all stakeholders with confidence in model performance.

www.fairly.ai

Disclaimer

All information provided reflects the status of this project at the time of writing and may be subject to change.

Neither Fairly AI Inc. nor its partners or subsidiaries as a whole, nor any single party within Fairly AI Inc.'s network warrant that the information contained in this document is capable of use, nor that the use of such information is free from risk. Neither Fairly AI Inc. nor its partners and subsidiaries accepts any liability for loss or damage suffered by any person using the information.

[13]There exists some confusion about the difference between model risk governance and model validation. Model risk governance impacts all phases of the model life cycle, from identification and development through to ongoing monitoring and retirement. Validation is one of the central phases of the model life cycle and is therefore one central phase of model governance. For a more detailed discussion of model governance and validation see the following reference: (Hill, The Top Fourteen Challenges for Today's Model Risk Managers, 2019)