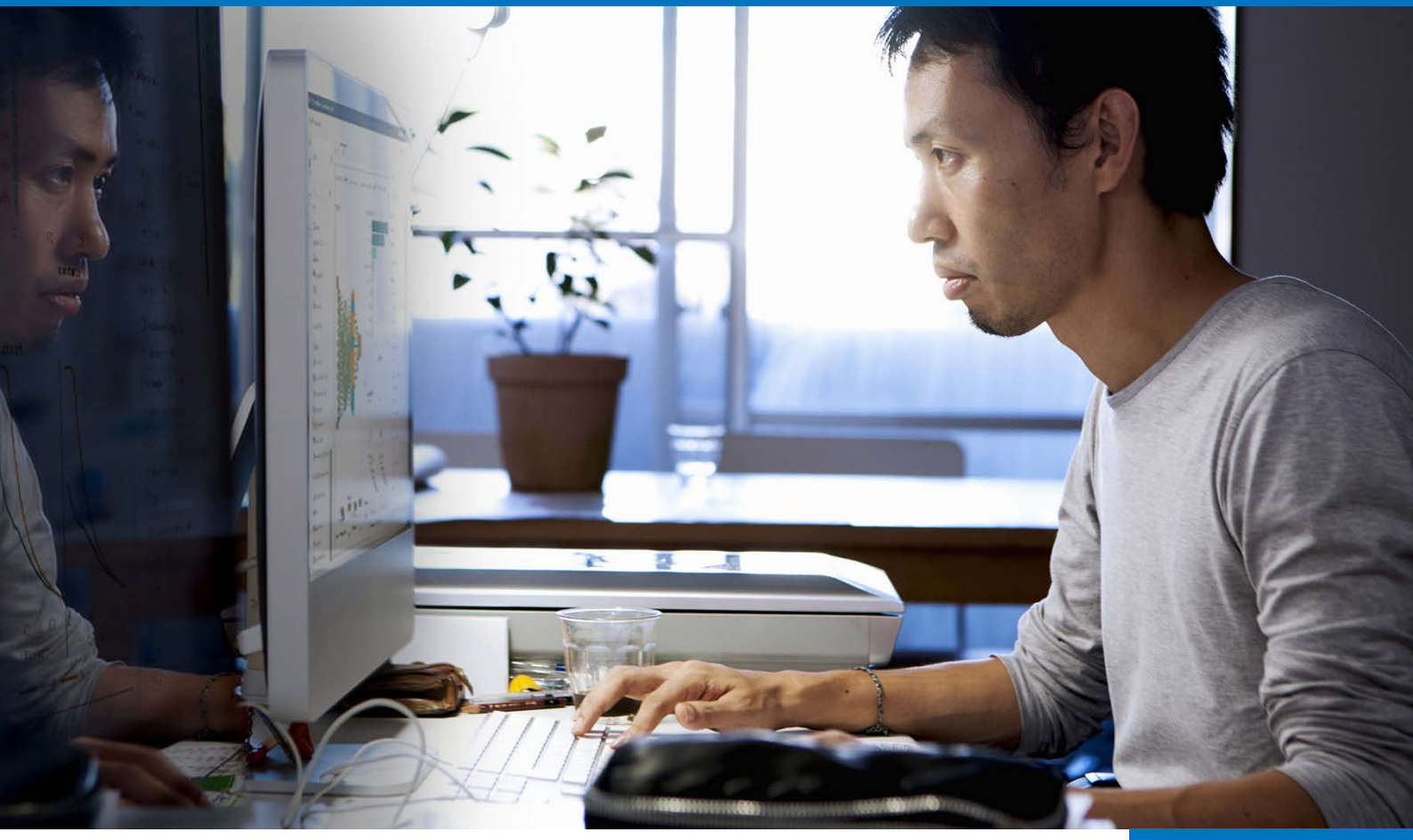
WHITE PAPER

Machine Learning Model Governance



sas
THE POWER TO KNOW.

Contents

By David Asermely, Global Lead, SAS Model Risk Management

Why Is Model Governance Needed?

Machine learning (ML) models need governance just like other models, only more so. This is particularly true of ML models designed to improve automatically through experience. Their ability to "learn" not only enables greater accuracy and predictability, but can also greatly increase model risk and result in ethical biases. So it's essential to establish rigorous governance processes that can quickly identify when a model begins to fail, complete with defined operating controls on inputs (data) and output (model results). The dynamic nature of ML models also means they require more frequent performance monitoring, constant data review and benchmarking, better contextual model inventory understanding, and actionable contingency plans.

The time to implement effective ML model governance is now – before today's global, multidimensional marketplace and massive data volumes overload traditional model risk guardrails and governance practices.

The Data Modeling Imperative

Today, there is great interest in harnessing ML to turn the massive volumes of data – including nontraditional data – into new insights and information. Unlike traditional statistical models, ML models are not limited by the number of dimensions they can effectively access. ML models can ingest vast amounts of unstructured data, identify patterns, and translate these patterns into actionable information.

The predictive power of these modeling techniques, in combination with the availability of big data and increasing computational power, will continue to be a source of competitive advantage for smart organizations. Those who fail to embrace ML face increasing competition and unsustainable operating models.

> The "use of AI and machine learning risks creating 'black boxes' in decision-making that could create complicated issues, especially during tail events."[1]

[1] Financial Stability Board (FSB), Artificial intelligence and machine learning in financial services. Market developments and financial stability implications. Nov. 1, 2017.

Regulatory Expectations and Implications of Ineffective Model Risk Management

The benefits of data modeling come with new risks and evolving regulations. Most notably, there are many ethical and legal questions around AI and data privacy. Analyzing nontraditional data using less transparent ML models (for example, to make better predictions) has also raised major financial, reputational and regulatory concerns.

Given the speed of industry adoption, government regulatory bodies are stepping up by crafting regulations and guidelines on everything from data privacy (GDPR)[2] usage and self-driving cars[3] to numerous ethics guidelines[4] [5] and specific model risk management frameworks.[6] [7]

However, there's no one-size-fits-all approach. The conversation regulators are having about AI varies dramatically by application and use case. For example, although regulators have actively promoted the use of AI technologies to enhance fraud detection,[8] they are concerned about its potential to introduce unfair bias in credit sanctioning.

> "To facilitate the application of AI to financial services, it would be important for relevant entities to establish reliable structures for effective governance and responsibility in case of tail events to ensure public trust to innovative financial services."[9]

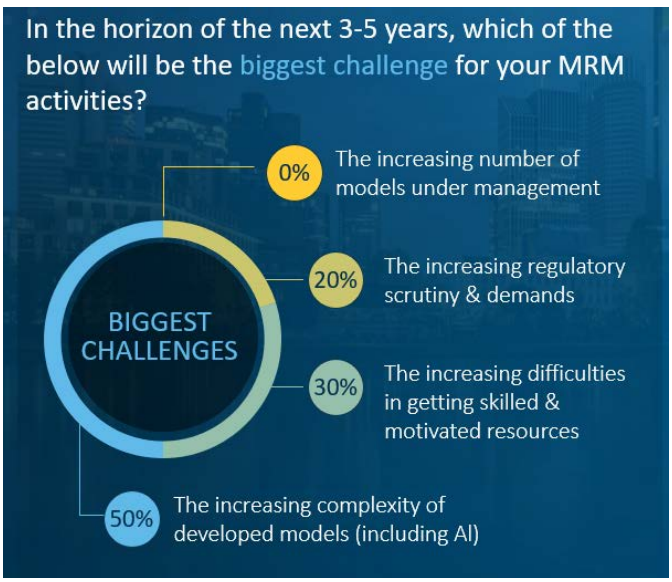


Figure 1: The biggest challenges banks face in the next three to five years around model risk.

[2] https://gdpr-info.eu/

[3] https://www.nhtsa.gov/technology-innovation/automated-vehicles-safety

[4] High-Level Expert Group on Artificial Intelligence. European Commission.

[5] https://flia.org/notice-state-council-issuing-new-generation-artificial-intelligence-development-plan/

[6] https://www.federalreserve.gov/supervisionreg/srletters/sr1107.htm

[7] https://www.opengovasia.com/singapore-releases-first-artificial-intelligence-ai-governance-framework-in-asia/

[8] https://www.federalreserve.gov/newsevents/speech/brainard20181113a.htm

[9] https://www.boj.or.jp/en/announcements/press/koen_2017/ko170413a.htm/

Why the heightened focus on governance from corporate and regulatory stakeholders? Because traditional model risk "guardrails" and governance practices such as manual back testing and annual model reviews aren't sufficient to handle the constant rate of change within ML models as they learn. Relying on these old-school practices will ultimately result in the breeding of "black swans" — events that come as a surprise, have a major effect, and are often inappropriately rationalized after the fact with the benefit of hindsight.

As explored in this paper, the best way to mitigate this risk is through smart governance and automated tools. "Banks should ensure that they have effective governance structures and risk management processes in order to identify, manage and monitor risks associated with the use of enabling technologies and the emergence of new business models and entrants into the banking system brought about by fintech developments."[11]

> "Model risk increases with greater model complexity, higher uncertainty about inputs and assumptions, broader use, and larger potential impact."[10]

Modernized Model Governance Is the Answer

Given the concerns regarding the transparency and potential misuse of ML models, it's vital that organizations implement a robust and modernized model governance system. In response, smart model risk management teams are investing significant time and resources to determine how to best manage these models.

With the right governance in place, your organization will be positioned to safely deploy and use big data and ML modeling techniques — and scale them with ease while retaining proper controls. You'll also be able to create and run them more effectively, which is no small feat. Increasingly, companies are learning that, "Developing and deploying ML systems is relatively fast and cheap, but maintaining them over time is difficult and expensive."[12]

You'll know your governance process is effective when it enables you to answer the following questions with specificity and in accordance with the best practices shared in each answer.

1: How do you define and classify your ML models?

Let's start with some of the most basic questions about ML models — how do you define machine learning models? How many machine learning models do you have in your inventory? And what are those models used for?

These seem like simple questions — but many model risk teams struggle to answer them. Smart, effective model risk management requires focusing on exactly the right models at the right time. It's essential to know where you have ML models, how they are being used, and how they fit within your enterprisewide model ecosystem.

[10] The Federal Reserve's "Guidance on Model Risk Management" (SR Letter 11-7).

[11] "Implications of fintech developments for banks and bank supervisors - consultative document." BCBS, August 2017.

[12] "Hidden Technical Debt in Machine Learning Systems" https://papers.nips.cc/paper/5656-hidden-technical-debt-in-machine-learning-systems.pdf

Classification requires clear definitions — and because machine-learning and AI can mean different things depending on the industry and institution, every organization must develop its own definitions and consistently apply them across its enterprise model inventory. In addition, it's important to understand an ML model's business purpose (for example, process automation and fraud detection) and purpose within a specific AI initiative (for instance, feature selection, data preparation, benchmarking or validation). Once ML models have been properly identified and classified within your inventory, you can provide the attention required by these highly dynamic model types.

2: Why was this ML technique or approach chosen?

For all models, it is important to document the rationale for using a given approach — whether it's a type of ML or otherwise – including its key features, available data and risks. The rationale for using complex ML models should always be considered against the use of traditional statistical models as a baseline; this forces the team to make a compelling case to justify use of the more complex and less transparent option. In addition, the rationale for using a certain type of model must address the potential contagion risk it could bring to other models within its ecosystem and how this risk will be mitigated.

Asking the first and second lines of defense (model developers and risk groups, respectively) to justify their ML approach using traditional statistical models as a baseline increases their governance workload. This cost should always be considered when deciding between techniques. But you can simplify this work by deploying a model management system that lets users simply check off their reasons, or rationales, for why they used a certain model approach. Reasons to chose a certain model approach could include the following:

- Involves unstructured data.
- Better predictability.
- Best fit for the problem.
- Low materiality model.

For every model, the ML approach rationale should be documented and approved by the model owner and risk division. Ideally, the decision to use — or not use — an advanced ML model should happen as early as possible within the model life cycle to prevent wasted work cycles. This decision should then be approved by key stakeholders. In some costly instances, teams have developed and validated an ML model that worked as designed and performed well, but it was never put into production because the risk division was not comfortable with the model's approach.

As a best practice, use a customizable workflow to enforce approval steps earlier in the process, as dictated by corporate policy. Automated systems and processes can streamline this process and provide greater transparency for model developers and stakeholders.

3: In what context is the ML model being used? Do you understand your ML model's interconnected risk?

Models do not exist in isolation. Often, data flows into models that produce output data, and this output data is then fed into other models. For example, if you have a model that projects future interest rates, it will likely have output that feeds many downstream models taking interest rates into consideration. These interdependencies create unintended consequences – for instance, when data or models are changed upstream, the downstream models can be disrupted. Note that:

> Model risk increases with greater model complexity, higher uncertainty about inputs and assumptions, broader use and larger potential impact. Banks should consider risk from individual models and in the aggregate. Aggregate model risk is affected by interaction and dependencies among models; reliance on common assumptions, data or methodologies; and any other factors that could adversely affect several models and their outputs at the same time. With an understanding of the source and magnitude of model risk in place, the next step is to manage it properly.[13]

As shown in Figure 2, visualizing data and model flows can help model risk managers put the model in proper context. Data and model map visualizations are important for all models, but the need for them is heightened with ML models due to their higher risk-to-reward range and dynamic use of different variables within data sources. Model map visualizations can help you highlight the less transparent and more dynamic ML models within the system that may need even closer attention.

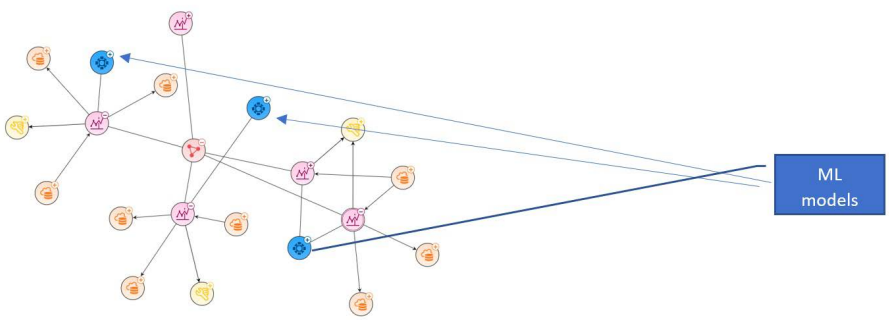


Figure 2: Model and data maps help visualize ecosystem risk.

"…developing a complete firm-wide model inventory and identifying interdependencies between models, has increased awareness of model risks that do not arise from any one model, but of those that exist between and external to models within what has been described most effectively as the model ecosystem."[14]

[13] The Federal Reserve's "Guidance on Model Risk Management" (SR Letter 11-7)

[14] "The top 14 challenges for today's model risk managers: Has the time come to think about going beyond SR11-7," Journal of Risk Management in Financial Institutions

4: What data is being used in your ML models?

Today, there's great interest in harnessing ML to turn the massive volumes of data – including nontraditional data – into new insights and information. In contrast to traditional statistical models, which are limited in the number of dimensions they can effectively access, ML models overcome these limitations. ML models can ingest vast amounts of unstructured data, identify patterns and translate them into practical information.

ML models can also ingest data from nontraditional data sources in diverse formats. For example, the US Department of the Treasury June 2018 report highlighted some of the alternative data sources being used within alternative credit models today. Specifically, the report states that, "New models and data may also unintentionally run the risk of producing results that arguably risk violating fair-lending laws if they result in a 'disparate impact' on a protected class or because the FTC or the Bureau might find the use of such models and data to be a violation of UDAP or UDAAP, respectively."[15]

Model and data governance often intersect. However, given the opaque nature of many ML techniques, their ability to use more dimensions and nontraditional data sets, and the need to recalibrate ML models more frequently, the burden of governance has shifted to the data.

While this increased reliance on data can boost model accuracy, it can also introduce biases and inaccuracies. Consider the implications of biases on fair lending compliance. "Fair lending" regulations are designed to protect consumers from unfair or discriminatory lending practices – in part by limiting the types of data that a bank can use to determine whether a loan should be made. For example, a bank cannot reject a loan application because of the applicant's age, race, sex, ZIP code, religion, marital status, etc. However, if a bank is using alternative data sources and ML techniques to make lending decisions, it's possible that its models will inadvertently consume variable inputs that directly indicate a protected class. An ML model might identify a relationship between individuals who listen to classic rock music stations and their loan default rates. This ML model could trigger a fair lending complaint due to age discrimination. As this example illustrates, ML models may implicitly infer patterns and correlations from the benign fields that effectively identify a protected class, necessitating closer and ongoing governance.

Furthermore, ML models are known to introduce unfair bias by overfitting to the training data. In a classic example, between 2014 and 2017 Amazon built an ML model to screen resumes and suggest likely successful hires. Amazon used 10 years of hiring and job performance data to train the model. Unfortunately, the data had a negative bias against women and ranked female applicants lower. The data set overfit the model toward male applicants because of the Amazon workforce gender split. Overfitting to your training data set can introduce unfair model bias and is a major hurdle in ML validation and governance. Techniques such as local interpretable model-agnostic explanations (LIME), synthetic construction of test data, and relative variable importance can be used to mitigate model bias.

> "Given the large data sets involved with most AI approaches, it is vital to have controls around the various aspects of data – including data quality as well as data suitability."[16]

[15] Financial System That Creates Economic Opportunities: Nonbank Financials, Fintech, and Innovation. US Federal Reserve report, p. 137. 2018. https://home.treasury.gov/sites/default/files/2018-07/A-Financial-System-that-Creates-Economic-Opportunities---Nonbank-Financi....pdf

[16] Fed Board of Governors speech 2018: What Are We Learning About Artificial Intelligence in Financial Services? https://www.federalreserve.gov/newsevents/speech/brainard20181113a.htm

Therefore, as part of your model governance, it's vital to properly assess the data sources and variables available for use by ML modeling techniques so you can answer the following questions:

- What data is available for the model?
- Are we comfortable with the model making decisions based upon that data?
- Would you be comfortable telling a customer a decision was made because of that data?
- Do the data inputs directly or indirectly violate any regulations (for example, fair lending)?
- How have you mitigated model bias?
- How frequently are new data fields added?

These questions are pertinent because ML models often overfit to the training data, introduce unfair bias, and quickly degrade as the data changes. Because of this risk, it's prudent to measure the difference between the data used to train and test an ML model versus the production data; when the production data form and structure are too different from the training data, the model should be reevaluated.

From a model management perspective, all of this should be documented and considered during the model methodology rationale phase. Given the complexity of this work, many companies are using modern model management systems to document what data sources are available to which models, as well as provide early warning signs when the data form and structure change beyond acceptable limits. Performance monitoring on your ML data inputs will provide the additional transparency needed to make informed judgments about a model's accuracy and fitness for purpose.

5: Are the ML model's results explainable?

Regulations, including SR 11-7 and OCC 2011-12, consider ML models as models. Global regulations require model users to understand each model's limitations, intent and output. Thus, your ML models must be explainable. In many cases, regulations also compel lenders to explain, in simple terms, why a decision was made. A bank's procedures for monitoring fair lending must appropriately flag bias classifications. This regulatory requirement is a perfect example of *why* ML/AI model outputs must be interpretable and explainable.

Some forms of ML do not provide visibility into what drives a model's prediction. This lack of transparency is a serious issue, as explained by an ECB Supervisory Board member in 2018: "The algorithms underlying AI must be carefully designed – and the decisions embedded within these algorithms must remain well understood – both by banks' leaders and by supervisors."[17]

> "Mistakes have been made too many times in the past when financiers have become too excited about some new, complex way of doing things. AI and machine learning must not be added to the list of financial crazes where executives have allowed the risks taken by their 'quants' to run ahead of their ability to understand and control what is going on."[17]

[17] ECB Supervisory Board member speech 2018, "The digitalisation of banking – supervisory implications." https://www.bankingsupervision.europa.eu/press/speeches/date/2018/html/ssm.sp180606.en.html

The "explainability" limitations of ML have stopped many banks from taking advantage of new, nontraditional data sources. But this limits the predictive power of ML models; studies show that if they do not use these new data sources, ML models often cannot outperform historically tuned statistical models. Organizations must determine if the predictive lift is worth moving from well understood and explainable models to more complex and less explainable ML models.

Drivers for explainability go beyond regulatory compliance. When consuming large volumes of data through ML models, end users need to understand – and be able to communicate – the results so decision makers trust them. These issues only intensify when working with complex black-box ML models such as deep neural networks; in these cases, business users, development teams, validators, auditors and regulators need transparency in order to trust and explain the model's output.

Techniques such as partial dependence plots, Shapley, individual conditional expectation plots, variable importance and LIME provide some visibility into the important features of an ML model and its sensitivity to certain variables. The industry continues to improve these techniques, but they are not yet sufficient to truly demystify ML models.

In light of these challenges, each ML model should have a documented approach on model explainability. Ideally, you have a way to centrally store and maintain this information over time, including:

- The business context.
- The selected interpretability technique and approach.
- The rationale for taking this approach.
- The approach limitations.

6: How are your ML models performing?

According to the Federal Reserve, "A firm should use measures to assess model performance that are appropriate for the type of model being used."[18] Performance monitoring provides insight into how well your models are making predictions against actual outcomes. Performance data is often used in audit reports, as early warning signals so proper contingency plans can be executed, and as guidelines for retraining or decommissioning models. Typically, the thresholds for these cutoffs are a discussion between multiple parties, including the model risk management group, the developer, model owner and the validator. These cutoffs should be documented and signed off on according to the workflow of your formal model governance process. A model governance system can help automate this process and document approvals.

Traditionally, the periodicity of these quantitative tests is set according to model materiality and recourse availability. The dynamic nature of some ML techniques (for example, reinforced learning) necessitates more frequent performance monitoring on these models. With a model governance system that automates this work, you can properly

[18] https://www.federalreserve.gov/supervisionreg/srletters/sr1518.htm. Federal Reserve, Attachment SR 15-18

run performance monitoring weekly, daily or even multiple times per day. Ideally, results are generated systematically and automatically fed through a model risk management system that compares the results against the agreed-upon thresholds to generate model risk findings and associated alerts. Note that in environments with low business volume, producing such reports too frequently will give unstable results. As such, these reports should only be run as frequently as the volume justifies.

The dynamic nature of some ML techniques (for example, reinforced learning), necessitates more frequent performance monitoring on these models.

7: What, when and how are machine learning models being used?

Most organizations don't have complete and up-to-date records on what models are being used, as well as how, where and by whom. In most cases, this data is captured manually and inconsistently at best. This is in direct conflict with best practices and regulatory guidelines, which state that "the extent and sophistication of a bank's governance function is expected to align with the extent and sophistication of model usage."[19]

Put simply, without proper model usage data, you will not be able to answer even simple questions about your model inventory. Modernized model risk management captures usage information consistently via automated design.

"It is an uncomfortable truth that today most financial firms cannot claim to have a complete and accurate inventory of all their active models, even though this is a regulatory requirement under increasing scrutiny in recent bank exams. It is even more uncomfortable that these firms cannot answer with much accuracy such questions as 'how many times was this model actually used during the last year?' or 'which models exhibit significant seasonality' or 'in what geographic regions, or Legal Entities, is this model used?' or 'are there any validated models with active status in your inventory that were not executed last year?"[20]

[19] The Federal Reserve's "Guidance on Model Risk Management" (SR Letter 11-7)

[20] Source: Hill, J. R. (2018) "Shouldn't a model 'know' its own ID?" The Journal of Structured Finance, Fall, pp.89–98.

ML techniques have increased the number of models used by organizations. However, without a model management system to automate the proper capture of model usage data, you will be unable to properly manage model risk across your growing model inventory. Consistent model usage capture leads to a treasure trove of additional information on your model inventory. This model usage data, in combination with other qualitative and quantitative data points, will be the backbone to using AI/ML techniques to manage model risk.

8: How is each machine-learning model recalibrated?

As with other models, ML models must be periodically recalibrated — and each recalibration has the potential to dramatically change model performance. To control and understand these changes, you need a way to clearly and consistently define, review and obtain signoff on the recalibration plan for each model. Changes from this plan should be similarly controlled. Because such controls are difficult to operationalize and track manually, it's recommended that you automate related processes.

9: How are your ML models performing compared to challenger statistical models? What is your contingency plan?

For high-materiality and low-transparency ML models, it is best to assign a statistical version as a benchmark, or champion, so you can monitor its performance. As stated in US Fed SR 15-18, "Benchmark models that are developed and run independently of primary models can be used to more effectively calibrate the firm's final estimates. For example, a firm can use benchmark model outputs to substantiate model overlays, given differences in risk capture between primary and benchmark models. This type of 'triangulation' is especially suitable for those areas of modeling that present considerable uncertainty."[21]

> "The most effective challenger/benchmark models are those that implement a different methodology from that of the champion."[22]

Benchmarking enables you to compare and assess your ML model and statistical model results and performance. If the ML model's performance degrades, you can replace it with the statistical champion version. The contingency plan should be well defined, specify the criteria that must be met to justify the switch, and be approved within your official model life cycle management process – all best orchestrated using a model management system. Ideally, challenger models will be fully validated and in production, as this will facilitate a quick and seamless switch.

[21] https://www.federalreserve.gov/supervisionreg/srletters/sr1518.htm. Federal Reserve, Attachment SR 15-18

[22] The top 14 challenges for today's model risk managers: Has the time come to think about going beyond SR11-7, Journal of Risk Management in Financial Institutions

10: Are you able to quickly adjust and respond to change?

Business, market and regulatory requirements will change – that's a certainty. The question is, how swiftly and efficiently can your business adapt to these changes? Increasing ML adoption is challenging existing model governance policy and systems. Today, much model risk data is buried in emails, local drives, hallway conversations and thick documents.

Given this rate of change, it's important to have the tools and talent to translate vast amounts of data now available to your business into timely insights to improve your model governance. Armed with this information, your organization can deploy iterative model risk management practices via a data-driven approach to assess model risks, identify gaps, review and update policies, and fine-tune processes. Machine adoption will increase the pace of model governance change required to successfully listen and adjust to your data.

The Future of Model Risk Management: Models Validating Models

Are you well equipped to answer all these questions about your ML models? Will you be able to do so as your organization's use of ML models rapidly accelerates?

If not, you're not alone. In SAS' 2018 MRM customer survey, we asked participants, "What, in your view, will be the impact of artificial intelligence on MRM?" As shown in Figure 3, 60% of the respondents replied that AI models will make their work more difficult, while 30% thought that "AI models can be used to improve or automate MRM activities."[23]

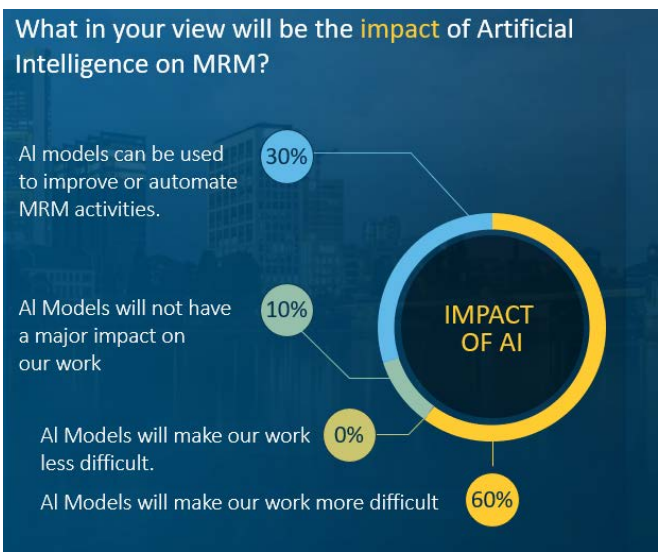


Figure 3: The expected impact of AI on MRM.

[23] SAS' 2018 SAS MRM Customer Survey

As ML models become the norm, your business will likely be using thousands of models simultaneously; in fact, some large enterprises already are. So the question is, how can you operationalize and scale rigorous supporting model governance?

The answer could be to use "smart" ML models built to validate and govern ML models.

This is entirely feasible and logical with a modernized model risk management platform that will capture and store troves of MRM data and information, including:

- Proper inventory classifications and hierarchies.
- More frequent and consistent performance monitoring quantitative data.
- Benchmark comparisons.
- Model usage data.
- Model interconnectedness, interpretability and variable sensitivity information.
- Input data quality metrics.
- Consistent documentation.
- Model metadata.
- Highly skilled human validation results.
- And much more.

SAS envisions that machine-to-machine enabled model governance – powered by analysis of data – will play a significant role validating individual ML models and even entire model inventories. Of course, having models validate models does add the potential of survivor bias propagating "lying" models. However, it also creates a way to automate model validation and even identify risks that humans alone would not be able to see.

Learn More

As your company seriously considers replacing well-understood statistical models with complex AI/ML black box models, it's time to think about how you will operationalize and scale comprehensive model governance. ML models offer the promise of better predictions, but they may introduce unknown ethical biases and increased model risk. Model risk professionals are grappling with how to best reduce this risk with automation, technology and best practices. By utilizing the best practices discussed above – rationale, mapping, data governance, performance monitoring, recalibration, interpretability, benchmarking and contingency planning – financial organizations can better satisfy increased regulatory demands when implementing a robust, reliable and automated ML model governance infrastructure.

To learn more about how SAS can help operationalize these best practices, please visit sas.com/mrm.

To contact your local SAS office, please visit: sas.com/offices