The imperative of expanding the traditional MRM function

How implementing an MLOps solution can help manage the risk and complexity associated with ML models proliferation.

Financial institutions and non-bank financial technology companies (FinTechs) alike make extensive use of various machine learning models (MLOps) in core and non-core areas of their business.

Banks, for example, rely on such models for a range of risk assessments, including predictive underwriting, credit risk management, suspicious and/or fraudulent activity management, fair lending compliance, derivative and financial instrument pricing and valuation, securitisation risks associated with trading and financial reporting. Developed in Python, R, MatLab or Excel, these powerful models are broadly leveraged by business users to support complex business needs.

Managing an increasingly complex model environment creates challenges for the modelling, risk and compliance teams, for senior management, as well as for auditors.

Regulatory considerations

Banking institutions have a regulatory framework that provides supervision and guidance. A model risk management (MRM) governance function must be implemented to help prevent making model-based decisions with damaging consequences for the business, should models prove to be inaccurate, flawed, or misused. If not properly managed in production, these models could trigger adverse commercial, operational, reputational, or regulatory outcomes.

Frameworks such as SR 11-7, SS3/8 are in place to ensure transparency and auditability of such models. The FDIC's requirements for institutions with over $1bn in assets are the following:

- Implement a disciplined and knowledgeable model development process that is well documented and conceptually sound
- Set up controls to ensure proper implementation
- Implement processes to ensure correct and appropriate use
- Implement effective validation processes
- Ensure strong governance, policies, and controls.

In SR 11-7, the Federal Reserve and the Office of the Comptroller of the Currency (OCC), broadly define "model" and "model risk" as a quantitative method, system, or approach that applies statistical, economic, financial, or mathematical theories, techniques, and assumptions to process input data into quantitative estimates.

These regulatory requirements generate costs for institutions incrementally to those associated with the models' development and implementation. Indeed, one way to avoid the "uncontrolled" proliferation of "pet spreadsheets and models" throughout the organisation is to monitor the applications in a centralised fashion.

The objective is not to impose specific models but rather to independently review and assess their build robustness, the value they provide and their evolution over time.

Focusing on applicable and upcoming regulations and monitoring model applications centrally can help financial institutions and FinTechs manage an increasingly complex MRM landscape.

The imperative of expanding the traditional MRM function

The changing shape of model risk management (MRM)

Ensuring visibility and relevance is especially critical in an environment where hundreds - if not thousands - of models must coexist and be used across the organisation. Traditionally, model risk management has been deployed to help validate the reliability, consistency, and robustness of models used by financial institutions.

The shift to external data

Today, the nature of the data consumed by these models is changing: the mix between internal and external data is rapidly shifting towards more external than ever before. Reliance on these models and consumption of the outputs and insights is increasingly exclusive to business units, which positions them outside the scope of the traditional corporate IT function. In other words, the lack of joint oversight from corporate IT on BU-specific models constitutes a risk exposure that must be addressed.

As model portfolios proliferate, so does the complexity of these models. Modelling teams have begun to incorporate Machine Learning (ML) tools and algorithms, for instance, to add predictive capabilities. This shift from easily replicable (for validation purposes) to more opaque, less explainable ML models creates new challenges for financial institutions and regulatory authorities. In a resource and cost-constrained environment, this factor contributes to a worsening model backlog and pipeline processing issue.

The need for adequate governance and oversight

Specifically for the growing share of ML models versus traditional models, the need for adequate governance and oversight becomes more pressing. Input data needs to be checked for quality and relevance, models' workflow needs to be managed, users need to set the level of automation required to feed data into the models, run and train them, and expose the outputs.

To keep risk under control, model owners must be clearly designated to guarantee the organisation's compliance with an end-to-end governance process

Benefits of an expanded and improved MRM function

Benefits from implementing an expanded and improved MRM function include providing a central repository for all models, tools, and other engines and the ability to validate and calibrate each step of the model lifecycle. This includes input data to model development, assess robustness, evolution and changes, as well as the reliability of outputs.

Further benefits of an expanded and improved MRM function are defining consistent model documentation standards and flexibly controlling accesses and roles. In addition, successfully leveraging technology, such as automation, advanced analytics and machine learning, can improve performance and cost-effectiveness and reduce complexity.

Finally, as a minimum, an effective MRM function improves the assessment of key ML model risk factors such as data relevance and reliability, model explainability and transparency. Significantly, it improves data privacy and security.

Compared to "traditional" models, such as those built-in Excel, ML models introduce new risks that need to be addressed in a specific way. Managing machine learning life cycles at scale is, therefore, more challenging.

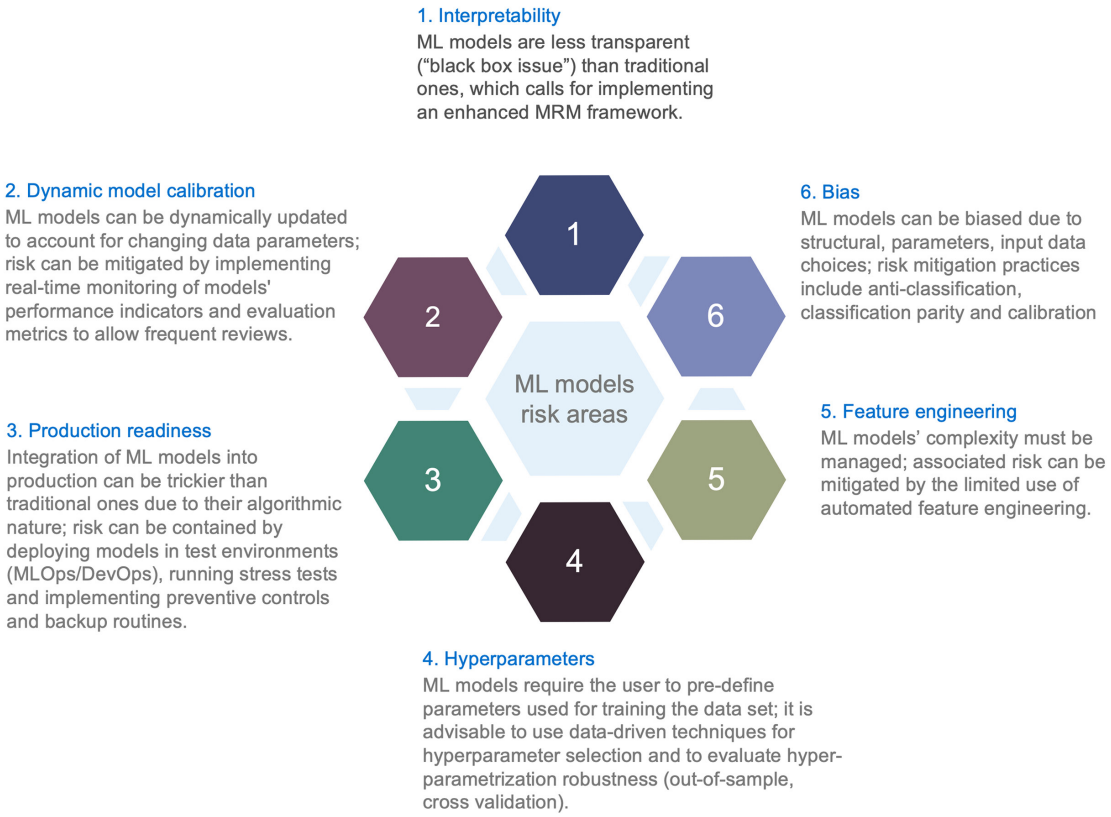
Firstly, ML models' performance depends mostly upon the data they are fed and trained with. Data drift can occur because of a change in the data collection phase or the context of the model evolving, feeding the model with potentially new unseen data, eventually degrading the model performance.

Even though the ML life cycle involves people from the business, data science, and IT teams, none of these groups uses the same tools or even, in many cases, share the same fundamental skills to serve as a baseline for communication.

Plus, while data scientists are specialised in data analysis and model building, different skills and tools are needed to deploy and maintain models in production. The complexity quickly becomes overwhelming when factoring in the data teams' staff turnover because data scientists end up managing models they did not create.

The imperative of expanding the traditional MRM function

Machine learning also presents risks to institutions that the principles embedded in SR 11-7 for model risk management can help to address. Below are some of the major considerations that financial institutions need to think about before utilising ML/artificial intelligence in their modelling framework:



MLOps: what it is and how can it help financial institutions?

Machine Learning Operations (MLOps) is broadly defined as a set of practices for collaboration and communication between data scientists and operations professionals; it is conceptually similar to DevOps, which is used to streamline the practice of software changes and updates. They both centre around automation of the end-to-end life cycle, continuous, high-quality delivery, and collaboration and communication between teams.

DevOps is not fully applicable in the data science context because deploying software code into production is different from deploying machine learning models into production: while software code is static, machine learning models are dependent upon training, testing, validation and everchanging input data, therefore exposing them to risks inherent to training data management and model re-training.

Helping to address banking challenges

Essentially, MLOps provides costs savings, improved performance and reliability through automation. Its embedded tracking capability makes model versioning more efficient and the end-to-end model lifecycle management more robust.

As an example, automation in a delivery project helps introduce standards of working and seamless collaboration practices; ensure product quality; eliminate tedious, repetitive manual efforts; reduce time to production. MLOps allows the production of the necessary documentation for model review, which can be automated and updated automatically with every change made in the model.

When it comes to versioning, historisation and reproducibility, MLOps records and tracks all test results and performance measures. This task can even be reperformed on demand, creating complete and searchable lineage tracking and audit trails for all production models.

As data or business evolves, different versions of the models need to be created. MLOps solutions provide capabilities that can operationalise different versions of models, notify users of version changes and generally allow an efficient management of model and data version history.

There is also improved quality and reliability as more robust and reliable processes for model development, testing, and deployment are defined and normalised across the organisation.

The imperative of expanding the traditional MRM function

Pre-requisites to deploying MLOps

In recent years, various tools have emerged that can help with ML pipeline automation. That said, the choice of tools for MLOps should be driven by the specific context for the ML solution and the way the financial institution's operations are set up.

Multiple tools might need to be selected as different tools automate different phases of the ML workflow, such as data preparation, model training, deployment, review and monitoring. These tools should be added to banks' DevOps toolkit, including CI/CD, audit logs, repo integration and alerts.

In assembling the right skills and processes, highly specialised resources are scarce, especially when combining ML engineer and DevOps profiles in a high turnover environment.

MLOps needs to be integrated into the financial institution MRM framework to better address ML-specific risks. Still, the oversight of those models should be consistent with the processes used for traditional models.

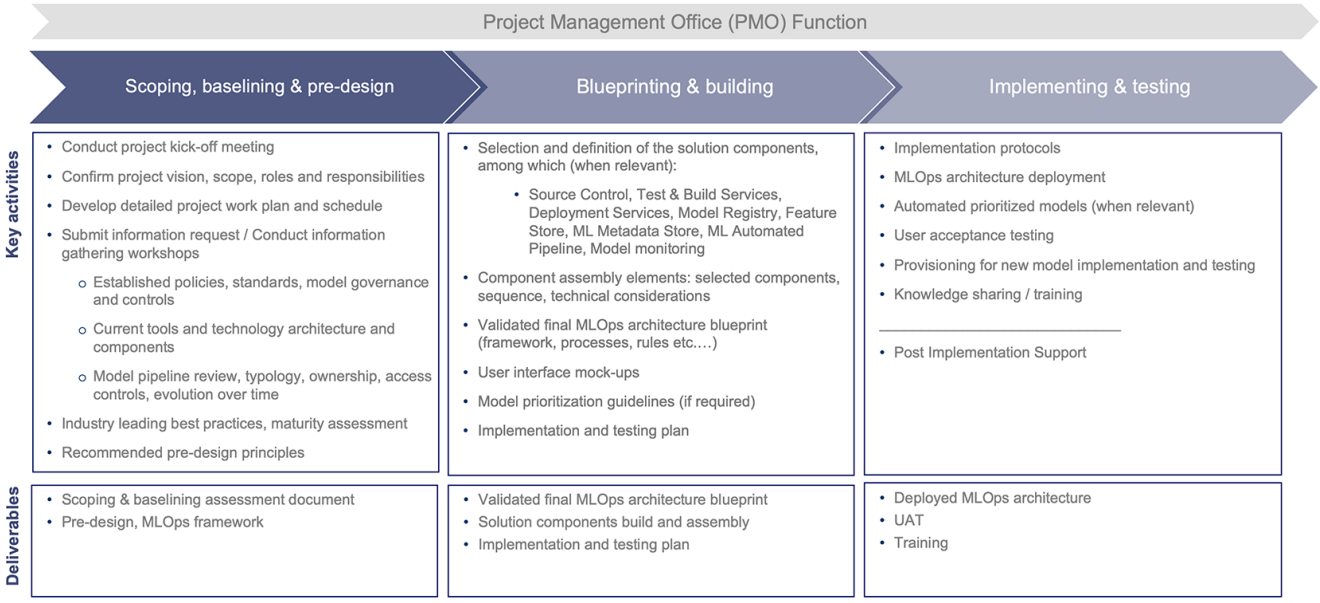
Typically, formal processes for the ML lifecycle are not easy to define accurately. Understanding the complete process is usually spread across several teams involved, often with no one person having a detailed end-to-end understanding of it.

Getting started on implementing MLOps

Mazars has designed a customised approach to help our clients implement an MLOps process to streamline and optimise their ML model pipeline.

Conceptually, a blueprinting and building module should follow after the joint scoping, baselining, and pre-design steps. At this stage, the complete solution should be mocked-up for validation and then built before the final deployment and testing phase.

Components of a potential MLOps solution (Illustrative only)

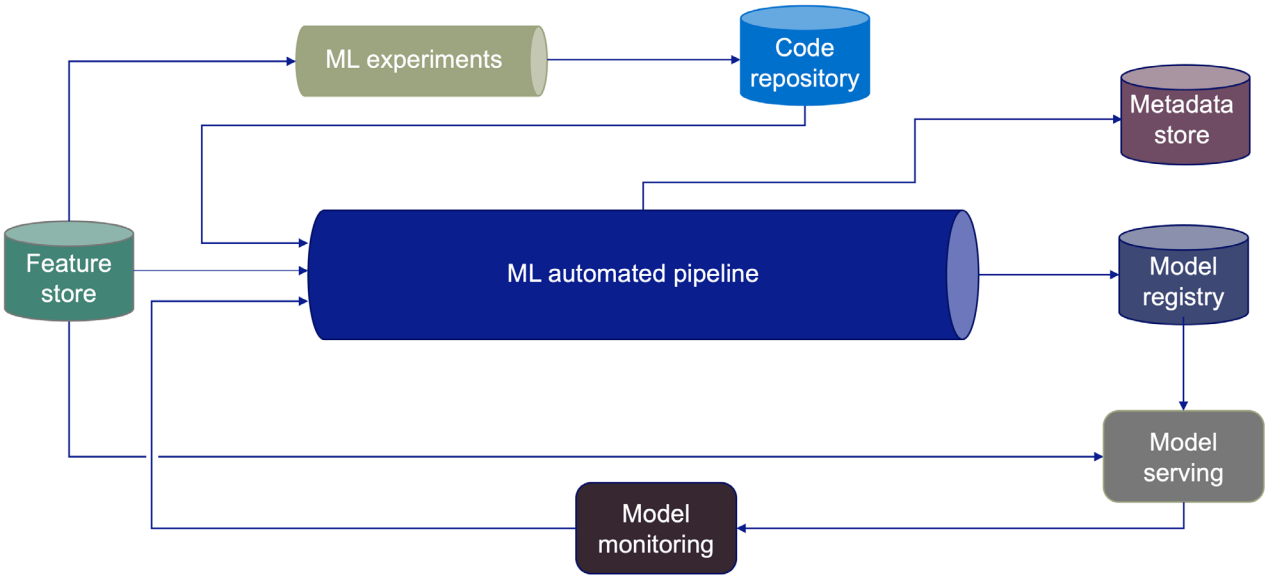


Components of a potential MLOps solution (Illustrative only)

The following components have to be customised to account for the financial institution's specific baseline infrastructure, model pipeline, needs, internal capabilities and objectives:

- Code repository and model registry: versioning the code, data, and ML model artifacts.
- Test & build services: using CI tools for quality assurance for all ML artifacts and building packages and executables for pipelines.
- Deployment services: using CD tools for deploying pipelines to the target environment.
- Model registry: a registry for storing already trained ML models.
- Feature store: pre-processing input data as features to be consumed in the model training pipeline and during the model serving.
- Metadata store: tracking metadata of model training, for example model name, parameters, training data, test data, and metric results.
- ML automated pipeline: automating the steps of the ML experiments.

Model monitoring: Monitor data input drift, model computational performance and results quality



Conclusion

Faced with the challenge of managing future complex model environments, financial institutions need to implement MLOps to limit their risk exposure vis-à-vis their business and regulators alike. The MLOps solution blueprint recommended by Mazars provides a path towards achieving these objectives.

Contacts

Emmanuel Dooseman
Partner, Global Head of Banking, Mazars
emmanuel.dooseman@mazarsusa.com

Patrick Zerbib
Partner, Mazars
patrick.zerbib@mazarsusa.com

Luis Belmar Letelier
Partner, Mazars
luis.belmar-letelier@mazars.fr

Gilles Cuyaubère
Manager, Mazars
gilles.cuyaubere@mazarsusa.com