DECEMBER 2020

IIF MACHINE LEARNING GOVERNANCE

SUMMARY REPORT

INSTITUTE OF
INTERNATIONAL
FINANCE

INSTITUTE OF INTERNATIONAL FINANCE

INTRODUCTION

The adoption of machine learning (ML) in the financial sector continues to develop rapidly; its use influences and touches many aspects of the financial sector. The technology has become an increasingly important tool for financial institutions, and its capabilities offer substantial benefits for industry, individuals, and society.

For the last three years, the Institute of International Finance (IIF) has been analyzing financial institutions' applications of ML, in particular its use in credit risk and anti-money laundering (AML), through various surveys and research papers.

Our new report, *Machine Learning Governance*, covers the end-to-end governance of the machine learning development and implementation process.[1] The latest in our series of machine learning studies,[2] it explores foundational aspects, data and inputs to machine learning, governance mechanism, model validation, model implementation, and model monitoring. It also explores considerations around bias, ethics, and explainability/interpretability in machine learning, and the need for strong governance to ensure that models are built, and data management is performed, with the customer in mind.

This study covers six topics:[3]

- Foundational Aspects
- Data and Inputs to Machine Learning
- Governance Mechanism
- Model Validation
- Model Implementation
- Model Monitoring

Our study finds that there is no "one-size-fits-all" approach to ML governance, and there are interesting regional differences, many of which can be attributable to existing non-discrimination and data protection laws.

IIF staff surveyed[4] 66 financial institutions, representing a diversity of scales[5], business models, and geographies[6].

[1] This paper presents an abbreviated public summary of the key themes of the IIF's *Machine Learning Governance Detailed Survey Repor*t published on December 3, 2020. Distribution of that Detailed Survey Report is limited to the official sector (supervisory community) and the 66 financial institutions that participated in the survey.

[2] Previous IIF machine learning papers include Recommendations for Policymakers (2019), Machine Learning in Credit Risk, 2nd Edition Summary (2019), Bias and Ethical Implications in Machine Learning (2019), Explainability in Predictive Modeling (2018), Machine Learning in Anti-Money Laundering (2018), and Machine Learning in Credit Risk (2018).

[3] All figures and tables contained in this report are from our 2020 survey results, unless otherwise stated.

[4] IIF staff surveyed participant firms during the period of January to August 2020. While our survey and interviews were framed to be representative on a firm-wide basis, it is acknowledged that there may be limitations in some responses, given the scale of some of the participating firms, and the visibility of some individual interviewees.

[5] By way of firms' scale, 17 out of the 66 FIs have total assets greater than $1 trillion, 17 are in the range of $500 billion to $1 trillion, 15 are in the range of $150 billion to $500 billion, while another 17 have less than $150 billion.

[6] FIs are categorized by region according to where they are headquartered, while acknowledging that many have operations across multiple jurisdictions There are nine regions in the study: Asia-Pacific (8 firms); Canada (5); China (5); Euro Area (12); Japan (5); Latin America (4); Middle East and Africa (8); "Other Europe" (10); and the U.S. (9). The Euro Area region consists of firms that are headquartered in countries that use the euro as a currency. The "Other Europe" region consists of firms that are headquartered in the Nordics, Switzerland, and the UK.

The majority of firms (68%) in our sample are using ML in production, and over a quarter of respondents (26%) have active pilot projects. The technology is increasingly employed in areas such as credit risk, compliance, market risk assessment, and insurance underwriting.

Noting that there are some differing views in defining "machine learning," a broad, inclusive scope was applied for the purpose of this study, including approaches that conform to at least some of the distinctive machine learning features.[7]

FOUNDATIONAL ASPECTS

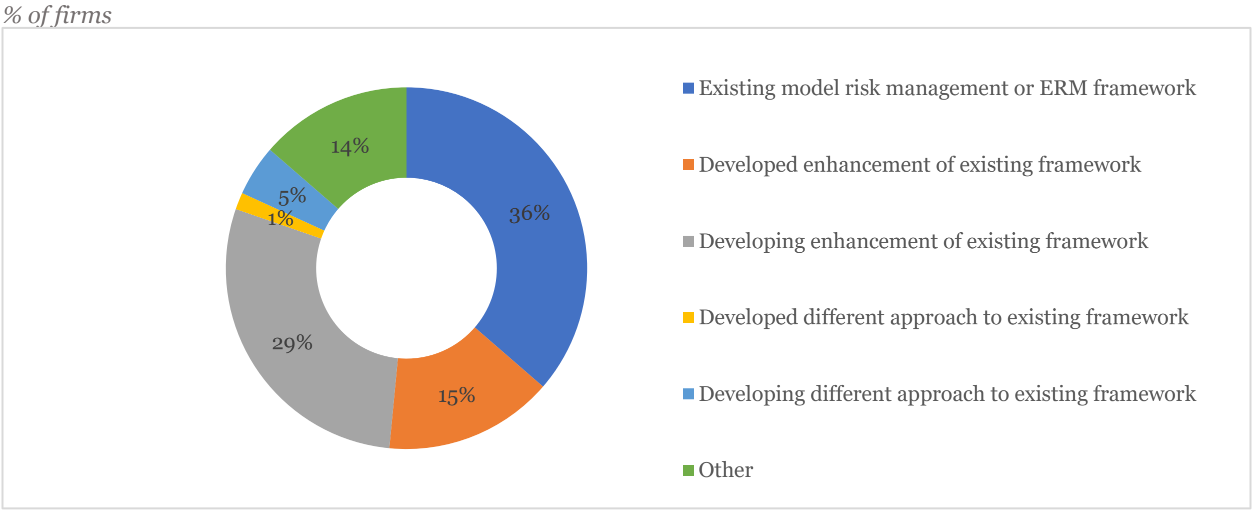
We acknowledge that levels of supervisory scrutiny for models differ across jurisdictions; in this new Report we consider the range of approaches across our diverse sample of firms. Most supervisors expect all firms using models to have in place certain common model risk management (MRM) components and implement these appropriately for their models.

While there are many ways to define these frameworks, to ensure consistency in responses, we use terminology set out in the *Supervisory Guidance on Model Risk Management* (SR 11-7) issued by the Board of Governors of the Federal Reserve System and Office of the Comptroller of the Currency.[8] This terminology is consistent with European Central Bank (ECB) expectations.

Model Governance Process in Place for Machine Learning

Most respondents are applying their existing model risk management or enterprise risk management (ERM) framework to ML applications, and several others are developing enhancements to account for new risks arising from ML techniques, such as control processes to mitigate against bias and discrimination of models. Those that are in the process of developing an enhancement to their existing MRM are looking to incorporate new associated risks that come from using more complex ML techniques, while taking particular care not to "over-govern" ML applications (see Figure 1).

Figure 1: What is the process of model governance currently in place for ML?



[7] See Annex for the ML definition used in IIF reports.

[8] SR 11-7 Guidance on Model Risk Management, accessed at: https://www.federalreserve.gov/supervisionreg/srletters/sr1107a1.pdf

At the regional level, we found that European and American firms were using their existing MRM framework for ML at a noticeably higher level than their counterparts in other parts of the world. Meanwhile, firms in Canada and the Asia Pacific regions are placing a heavy emphasis on developing enhancements to existing frameworks.

Defining Machine Learning Models for Governance

As touched upon above, consensus is lacking on a clear definition for ML—not only globally or industry-wide, but also *within* half of the survey respondents: only one in two respondents have a clear, internal definition of what constitutes a ML model.

In terms of what criteria qualifies entry in the governance process, many firms highlighted materiality as the key driver in deciding the extent of monitoring, validation and governance. Materiality is assessed typically across multiple dimensions such as financial impact, the use case objective, and considering any impacted party. However, in cases where all models are considered in the governance process, one consideration may be the complexity of the methodology.

Machine Learning Ethics Framework or Specialist Committees on ML

Roughly one-third of firms have established a specialist committee to advise the respective governance bodies and risk management functions on ML-specific questions, and nearly another one-third have established a ML ethics framework to address ethical issues raised by ML models and the use of new data source. Among these, several have established both.

Controls Against Bias and Discrimination

Firms rely on several control processes to mitigate against bias and discrimination in ML models. Firms were able to select more than one option, and their answers indicate that the controls are very much dependent on the use case. We define "bias" as an unfair inclination for or prejudice against a person, group, object, or position. Discrimination, whether intentional or unintentional, may occur when one group of people is more adversely affected by a decision or process than another group without a legitimate and neutral justification.[9] This area continues to evolve, along with the supporting literature.

The most common selections made by firms regarding how to mitigate bias and discrimination in ML models were "auditing, testing and controls", "code of ethics defined at the institution level", and "excluding sensitive attributes[10] from the beginning and not including these as part of the feature analysis / selection / engineering process" (see Figure 2).

At the regional level, we see that there were vast differences in the way firms responded. Firms from the U.S. and "Other Europe"[11] regions selected several options at an equal or higher rate than the overall sample average, while firms from China selected several options at an equal or lower rate than the overall sample average.

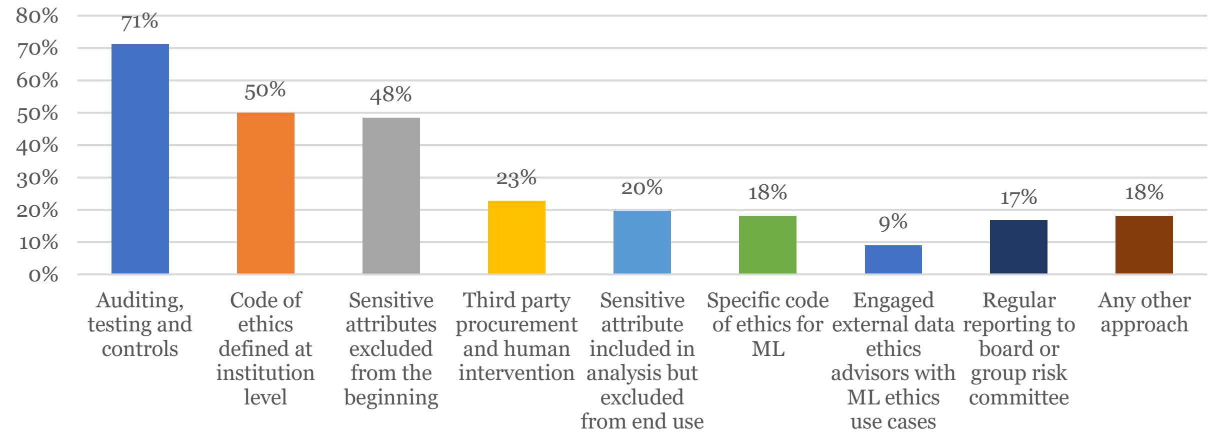
[9] Laws typically evaluate the discrimination using two distinct notions: disparate treatment, and disparate impact. Disparate treatment includes overt discrimination, as well as more subtle unjustified differences in outcome on a prohibited basis. Disparate impact occurs when a neutral policy or practice results in a disproportional exclusion or burden on certain group of people, whether or not the policy was created with the intent to discriminate.

[10] "Sensitive attributes" related to protected/sensitive features that could create moral, ethical and legal problems. Although, this view of whether to include or exclude sensitive attributes is very much linked to whether it is lawful or unlawful to do so. For some executives, withholding sensitive demographic information from an algorithm does not solve the problem of "redundant encodings," where membership in a protected/sensitive class is encoded in other data.

[11] The "Other Europe" region consists of firms that are headquartered in the Nordics, Switzerland, and the UK.

*Figure 2: What controls are in place to mitigate against ML models producing unfairly biased or discriminatory outcomes? **

% of firms



* Firms were able to select as many options as appropriate.

DATA AND INPUTS TO MACHINE LEARNING

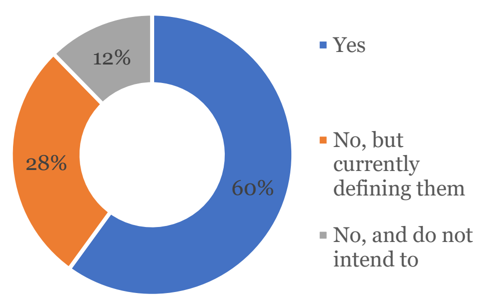
Data and inputs used to develop a ML model are critical to model development and implementation, and as such firms were asked questions on their assessment of data quality. Key findings touch on data governance, accountability for quality of training data sets, and the existence of data governance committees as they relate to ML. Important questions around the use of external data and third-parties were also covered in the survey, and key findings are presented below.

Externally Sourced Data and Third Parties

The majority of firms have principles in place for the use of externally sourced data and aggregate scores provided by third parties, and over a quarter are in the process of defining them (see Figure 3).

*Figure 3: Are there any principles in place for the use of externally sourced data? **

% of firms



* Externally sourced data can include sources such as social media data and aggregated scores provided by third parties. Sample size of 65 firms.

The interlinkage between advanced analytics (including AI/ML) and broader data privacy principles was reflected in the answers provided by firms in the Asia Pacific, Canada, China, Euro Area, "Other Europe", and the U.S. regions.

Firm-Wide Data Governance Committees

The vast majority of respondents have established a firm-wide data governance committee as it relates to ML applications, with nearly all of them indicating that they have ML models in production.[12]

Accountability for Quality of Training Data Sets

The three most common responses among the 59 firms that answered the question on the accountability for the quality of training data sets were "model owner", "data owner", and a shared responsibility between the model owner and data owner.

Compliance with Relevant Data Privacy Regimes

Overwhelmingly, participants comply with relevant data privacy regimes. Once again, answers to this question show digital regulation and data strategy working together around ML implementation.

In many instances, firms indicated that privacy matters and regimes are handled by their legal and compliance teams. Across all regions, firms highlighted that bank operations and applications have to comply with existing requirements from government and regulatory bodies, including for data privacy.

GOVERNANCE MECHANISM

Our findings indicate that participating firms have in place most of the different control and governance mechanisms such as board and senior management oversight, policies and procedures, controls, and organizational structure. However, the practical implementation varies and is closely linked to the use case – i.e., a firm's business activities, the complexity, and extent of its model use.

Governance Centralization

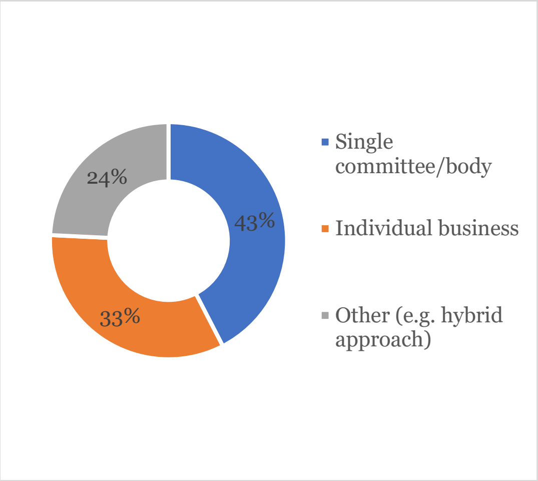
Figure 4 shows that the most common approach is to have ML model governance centralized to a single body within the institution, while a third of respondents revealed that it was the responsibility of individual businesses within the firm. Of the near quarter of firms that selected "other", several indicated that they had something of a hybrid model in place.

An interesting remark we heard from a couple of respondents is how they are striving to move away from what they currently have in place.

[12] Figures in the "Firm-Wide Data Governance Committees" sub-section reflect a sample size of 60 firms.

Figure 4: Is the governance of ML models centralized to a single committee/body in the firm or is it the responsibility of the individual businesses (with associated committees/boards)?

% of firms



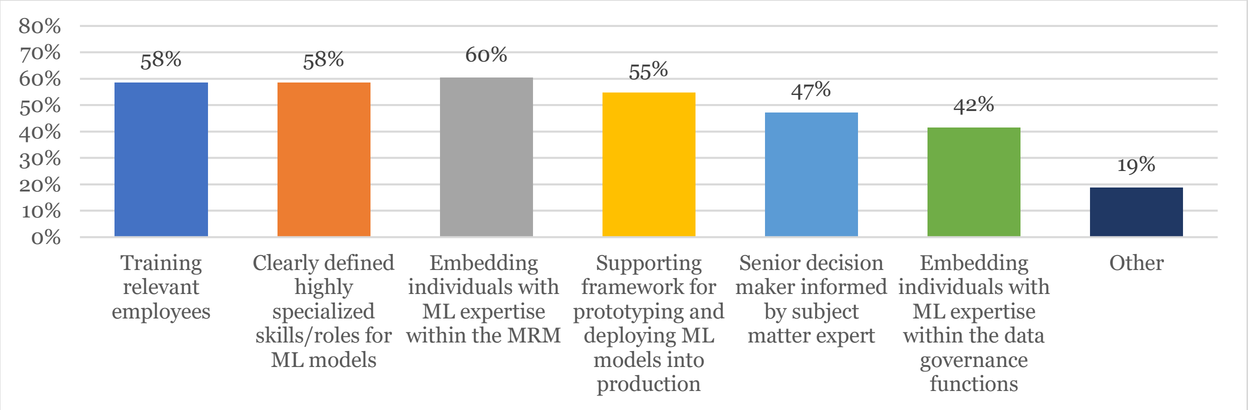
Articulation of Roles and Responsibilities

When asked whether the roles and responsibilities of the parties responsible for the models throughout the model lifecycle (from inception to retirement) were clearly articulated in the governance for ML models, the vast majority of firms responded “yes.” One of the firms that responded “no” elaborated by saying that the roles and responsibilities are clearly articulated in their model risk policy and associated standards, but that they apply to all models within the firm and are not ML-specific.

Among the firms that answered “yes,” a follow-up question asking for the specific steps taken was posed to the group. And as Figure 5 illustrates, “embedding individuals with ML expertise within the model risk management” was the top answer, followed closely by “providing training for relevant employees” and “clearly defining the highly specialized skills/roles required for ML models.”

*Figure 5: What steps have been taken to establish roles and responsibilities for parties responsible for the governance of ML models? **

% of firms



*Firms were able to select multiple choices. Sample of 53 FIs that answered "yes" when asked whether the roles and responsibilities are clearly articulated in the governance for ML models.

MODEL VALIDATION

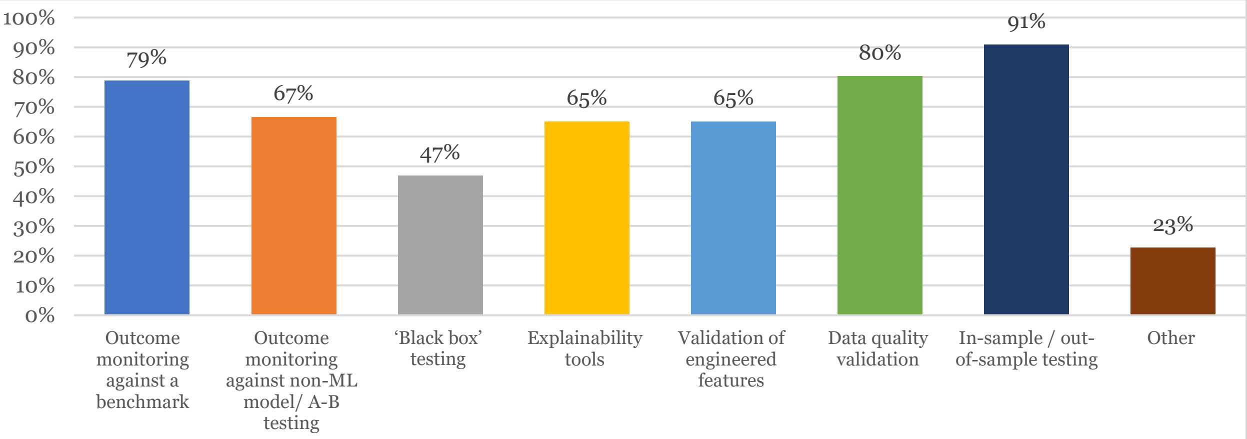
The sophistication of validation and the choice of techniques employed to assess the robustness of ML models vary depending on a number of factors, including the use of models, complexity and/or materiality.

Firms validate ML applications before and after deployment. The most common validation methods are in-sample / out-of-sample testing followed closely by data quality validation and outcome monitoring against a benchmark (see Figure 6).

A common point that study participants raised when answering this question was that each technique has its own limitations, and its usefulness really depends on the business application, the team using it, and the complexity of the model.

*Figure 6: What model validation techniques are used to assess machine learning model robustness? **

% of firms



* Firms were able to select as many options as appropriate.

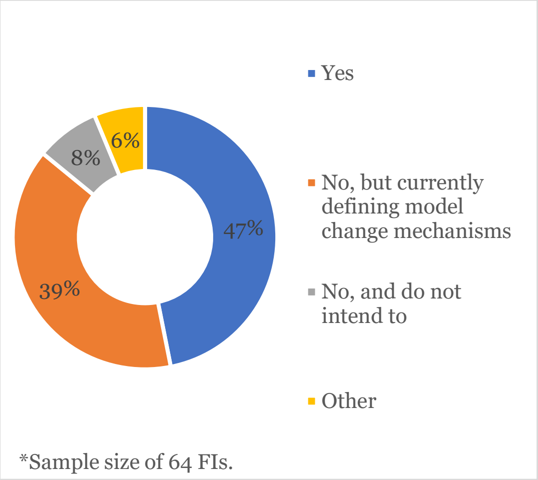
Although most firms use ML applications that are developed in-house, firms also reported relying on ML vendor models for a myriad of applications, including to help with natural language processing, image recognition used for onboarding, etc. Over a third validate ML vendor models in the same manner as internal models.

MODEL IMPLEMENTATION

Our findings indicate that most firms either have implementation platforms that cater for the need to frequently update/change model parameters or are in the process of establishing them (see Figure 7). A reoccurring remark centered around how updating model parameters was highly situational and implementation specific as some models may be designed to be updated on a more frequent basis depending on the use case. Several firms explained that models may be reviewed on a more frequent basis depending on the complexity/materiality of the model, which could result in the identification of limitations/overlays, where appropriate.

*Figure 7: Does your implementation platform cater for the need to frequently update/change model parameters based on refreshed data collected? **

% of firms



*Sample size of 64 FIs.

MODEL MONITORING

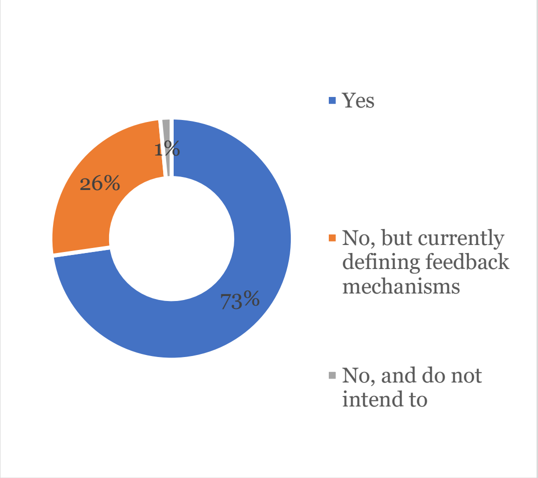
Similarly, in terms of model monitoring, there is a variety of feedback mechanisms and controls, and of safeguards to mitigate the risks of ML models, and which are used is very much dependent on the individual model in question.

Feedback Mechanisms and Controls

As Figure 8 illustrates, nearly three-quarters of firms have feedback mechanisms or controls in place to ensure expected outcomes and to prevent the distribution of input data and features from drifting over a period of time.

Figure 8: Are there any feedback mechanisms or controls in place for correcting the ML model (ensuring outcomes are as expected) and to ensure that the distribution of input data/features does not drift over a period of time?

% of firms



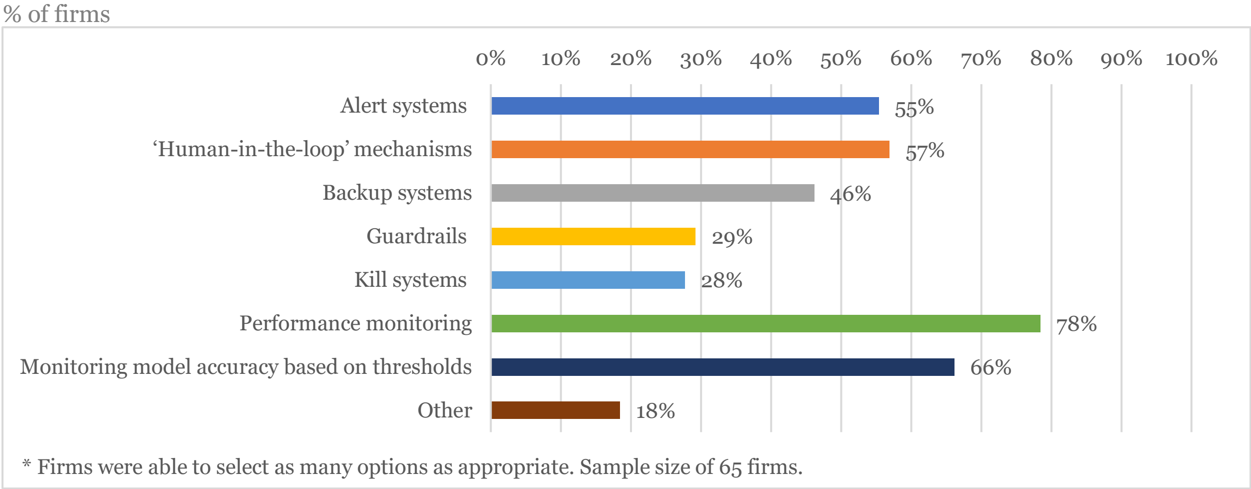
At the regional level, European, American, and Canadian firms indicated they had feedback mechanisms or controls in place at a significantly higher rate than their peers in the Middle East and Africa, China, Japan, and Asia Pacific. The U.S. was the only region where all firms reported having feedback mechanisms and controls already in place.

Safeguards

In order to manage the risks associated with ML, firms use a variety of safeguards. The most common safeguards are performance monitoring, followed closely by monitoring model accuracy based on thresholds (see Figure 9). Many remarked that performance monitoring is important as models will degrade overtime as you use them (i.e., model drift or the degradation of the model's predictive power).

"Human-in-the-loop" (HITL)[13] mechanisms ranked third, but when selected, firms referred to it as critically important in the development of ML models.

*Figure 9: What safeguards are built into the software? **



* Firms were able to select as many options as appropriate. Sample size of 65 firms.

[13] Human-in-the-loop (HITL) is the process where decisions made by the ML application are only executed after review or approval from a human. This process starts by involving human intervention in training stages when building an algorithm, creating a continuous feedback loop that allows the algorithm to give better results, and for testing and validating the model.

LOOKING AHEAD

This survey and report are a first step to understanding the governance around machine learning in financial institutions represented by our membership. Given its power and impact, ML requires a collaborative effort between the industry and the supervisory community to ensure that it protects customers without stifling its adoption or stalling innovation in the financial sector. Further work is needed to gain a deeper understanding on the state of deployment for particular business areas such as credit risk, compliance (AML, fraud prevention/detection, anti-financial crime), predictive marketing, trading, portfolio management, customer acquisition, etc., given the choice of safeguards and controls are very much dependent on the use case.

With this in mind, the IIF is considering repeating the survey in the future to track the development and deployment of governance aspects by FIs. Concurrently, we plan on continuing our existing dialogue with policymakers, financial institutions, and subject matter experts on how to identify common good examples of practice, and help support the safe, ethical development of ML models.

At the IIF, we will continue to monitor its application and identify challenges and areas in which firms can share knowledge to support a safer deployment of ML solutions. Future surveys on specific business areas will include questions around data validation, governance, and potential ethical issues that arise from the use of ML.

ANNEX: DEFINING MACHINE LEARNING

Given the lack of consensus on a clear definition for ML, in previous IIF reports and for the purposes for this Report, we use a wider definition for ML. Rather than providing a constraining definition, participating financial institutions were asked to consider four key attributes that most ML approaches conform to. These attributes are:

1. A primary goal of optimizing out-of-sample predictive performance facilitated by well-tuned regularization.[14]
2. A significant degree of automation in the model development process.
3. The use of cross-validation to model relationships in the data, i.e., divide data into random separate sets for the purpose(s) of training, testing, and validation. [15]
4. Applicable to very large volumes of data (although some techniques also work well on small data sets), including, in some cases, unstructured data sources. [16]

The main component of ML is that it provides systems with the ability to automatically learn over time, generally from large quantities of data. The learning process is based on observations or data, such as examples, in order to identify patterns in data and make better predictions. An ML algorithm can therefore be seen as an algorithm that, from data, generates another algorithm, usually referred to as a model.

[14] *Regularization* refers to optimizing a model's ability to predict data points out-of-sample. Requires finding an optimal fit of the modeled relationship that neither underfits or overfits the data. In other words, this technique constrains or shrinks the coefficient estimates towards zero. In overfitting the model describes random error or noise in the data (i.e., data points that don't represent the true properties of the data), rather than the underlying relationship.

[15] *Cross-validation* entails fitting a model by running it on different randomly sampled datasets. While cross-validation can be used to examine any model, ML techniques employ it to create the model and smoothen or regularize the modeled relationship. For more detail, see Leo Breiman, "*Statistical modeling: the two cultures*," Statistical Science vol. 16, no. 3 2001, 199-231.

[16] *Unstructured data* refers to data that does not have a structure that makes it readily accessible for analysis. Examples are mobile and sensor data, social media streams, images, and videos. *Structured data* refers to data that is well-organized and clarifies and standardizes the relationships in the data. Conventional statistical methods have been well able to analyze this type of information.

Authors

Natalia Bailey
Policy Advisor, Digital Finance
nbailey@iif.com

Dennis Ferenzy
Associate Economist, Digital Finance
dferenzy@iif.com

Contributor

Brad Carr
Managing Director, Digital Finance
bcarr@iif.com