June 30, 2021

Via Federal Register

Chief Counsel's Office
Attention: Comment Processing
Office of the Comptroller of the Currency
400 7th Street SW, Suite 3E-218
Washington, DC 20219

Ann E. Misback
Secretary
Board of Governors of the Federal Reserve System
20th Street and Constitution Ave NW
Washington, DC 20551

James P. Sheesley
Assistant Executive Secretary
Federal Deposit Insurance Corporation
550 15th Street NW
Washington, DC 20429

Comment Intake
Bureau of Consumer Financial Protection
1700 G Street NW
Washington, DC 20552

Melane Conyers-Ausbrooks
Secretary of the Board
National Credit Union Administration
177 Duke Street
Alexandria, VA 22314-3428

RE: Request for Information and Comment on Financial Institutions' Use of Artificial Intelligence, including Machine Learning

Ladies and Gentleman:

DataRobot, Inc., appreciates the opportunity to provide public comment on financial institutions' use of artificial intelligence (AI), including machine learning (ML). DataRobot is an enterprise AI platform and automated decision intelligence tool that pioneered the use of automated machine learning (AutoML). DataRobot has spent hundreds of millions of dollars on research and development. DataRobot currently works with over 800 public and private sector organizations and companies, including eight of the top ten financial institutions in the world.

The below responses reflect DataRobot's experience and expertise in working with DataRobot's financial institution customers.

#1: ***How do financial institutions identify and manage risks relating to AI explainability? What barriers or challenges for explainability exist for developing, adopting, and managing AI?***

Financial institutions identify and manage risks relating to AI explainability by setting standardized requirements for which analytics need to be provided for models that are validated

by model risk management functions. This process can sometimes take up to two years.

The biggest challenge for this is when these standard requirements meet non-standardized processes for developing AI models (i.e., where models are hand-coded ad hoc). Such processes can result in different implementations of the analytics required, and bear a high risk of inconsistencies and errors.

From DataRobot's perspective, model behavior post-hoc is a solved issue. Instead, there is ambiguity related to standard practices, which tool to use, and how to interrogate a model post-hoc. AI explainability is relevant to both the training and deployment states of machine learning models. In training, the data scientist or the AI model developer uses explanations to validate the model's learned pattern. However, after deployment, explainability is primarily used when users (e.g., a loan officer or a fraud detection analyst) can understand the reason behind the model decision. As a result, consistency in interpretability tools and metrics is important in addition to simply providing information that users may use to understand predictions.

Technology has advanced in such a way that institutions are now able to use built-in tools in enterprise AI software to manage risks associated with explainability and model performance. Such software (like DataRobot) provides robust model monitoring tools that allow data scientists and AI users to keep track of their models' performance. For example, when data drift is identified (a significant shift in the characteristics of the data used for predictions as compared to the data used to train the models), the software will alert users that the model needs to be retrained.

#2: *How do financial institutions use post-hoc methods to assist in evaluating conceptual soundness? How common are these methods? Are there limitations of these methods (whether to explain an AI approach's overall operation or to explain a specific prediction or categorization)? If so, please provide details on such limitations.*

These methods are common and generally follow Guidance on Model Risk Management (SR-11-7) requirements. General limitations to these methods are described in question #1.

One notable limitation of typical post-hoc methods in outcomes-based explainability is resilience to cases where the modelling data contains multiple series which are highly intercorrelated. Without proper treatment of such quasi-redundant series, while the model performance will not suffer, post-hoc explainability techniques such as permutation importance (for identification of important variables) or partial dependence analysis (for sensitivity analysis of predictions with regards to important input variables) are likely to give misleading information unless compensated for correctly. This can lead to the importance of certain variables being understated if they are highly correlated. At DataRobot, for instance, we address these issues by using feature reduction techniques such as Feature Importance Rank Ensembling or Feature Clustering, both of which DataRobot developed specifically to address these problems.

#3: *For which uses of AI is lack of explainability more of a challenge? Please describe those challenges in detail. How do financial institutions account for and manage the varied challenges and risks posed by different uses?*

Unsupervised machine learning and reinforcement learning algorithms are the classes of algorithms where lack of explainability is more of a challenge. This is because many post-hoc explainability techniques rely on understanding changes to algorithm performance metrics as a result of the application or removal of certain stimuli (data scenarios); for unsupervised and reinforcement learning methods, performance metrics tend to be less well defined, which makes a sensitivity analysis of changes difficult.

With supervised machine learning, which accounts for the large majority of use cases in financial institutions, interpretability (as opposed to explainability) techniques are well understood and straightforward. Outcomes-based interpretability is widely used. Various input data scenarios (stimuli) are applied to models, and the outcomes (predictions) are measured and quantified. Some circles distinguish between explainability — the ability to understand in detail how an algorithm reaches its decisions (like understanding the workings of an internal combustion engine in an automobile) and interpretability — the ability to understand what the sensitivities of an algorithm's decisions are in (knowing what happens when you turn the steering wheel, step on the gas, or press a button). Arguably, in financial institutions, interpretability is more important and of more practical value than explainability.

Platforms, like DataRobot, that provide model-agnostic frameworks to enable users to interpret results, make informed adjustments, and access easy-to-use and state-of-the-art interpretability techniques for all models, are now widely available. Using software like DataRobot helps financial institutions account for and manage explainability challenges by providing tools that allow the user to compare models and explore feature fit, accuracy over time, lift charts, feature impact, hot spots, and feature effects.

#4: *How do financial institutions using AI manage risks related to data quality and data processing? How, if at all, have control processes or automated data quality routines changed to address the data quality needs of AI? How does risk management for alternative data compare to that of traditional data? Are there any barriers or challenges that data quality and data processing pose for developing, adopting, and managing AI? If so, please provide details on those barriers or challenges.*

In DataRobot's experience, there has not been a significant change in control processes or automated data quality routines as a result of AI. DataRobot notes that in many cases, AI actually has less strict data quality requirements versus "traditional" modeling. However, it is important to take into account the risk that biases can creep into machine learning models if the sampling of data is skewed. This can be a particular challenge with alternative data.

Similar to explainability tools, software is available on the market that enables users to implement robust exploratory data analysis including bias, unbalanced data, and target leakage checks. Software like DataRobot enables organizations to standardize evaluations and not rely on bespoke analysis by a data scientist or group.

Both supervised and unsupervised machine learning techniques can be used effectively to improve data quality, both by making data quality teams more efficient (reducing the number of false positive alerts) and by more effectively finding issues (by using sophisticated unsupervised pattern recognition techniques to identify more complex data quality issues).

#5: *Are there specific uses of AI for which alternative data are particularly effective?*

AI has advanced to be able to accommodate multimodal datasets. For example, DataRobot can accommodate visual, time aware, location, text-aware, and traditional tabular data. As a result, data scientists and AI users can build models using all of the data available and the software will identify the most informative data used to train the model. In addition, feature effect tools will also highlight to the user what specific features were the most useful. This removes much of the guesswork out of determining which data should be used in which use case.

In many use cases (especially client services, marketing, and financial crime prevention), the addition of textual data (natural language) can improve model performance by 10-15 percent. In event prediction, alternative data sources in general are valuable.

#6: *How do financial institutions manage AI risks relating to overfitting? What barriers or challenges, if any, does overfitting pose for developing, adopting, and managing AI? How do financial institutions develop their AI so that it will adapt to new and potentially different populations (outside of the test and training data)?*

It is of vital importance to ensure that AI models are not overfit, otherwise they will not perform as needed once deployed, which in turn will create barriers to adoption (from a change management perspective). In addition, measuring model stability is important.
Risks of overfitting are managed through the building of challenger models in the context of SR-11-7 model validation processes, notably through the use of challenger models *that have been validated on out-of-sample data* (i.e., data where the outcomes are known but were not seen or used in the course of training the models). Using automated machine learning techniques makes this challenger model building process more efficient and scalable. It is crucial to ensure that models are validated on out-of-sample data before deployment; good automated machine learning approaches will enforce this requirement and make it very difficult for users to deploy any models that have not been validated thus.

Best practice for adapting AI models to new and potentially different populations outside of test and training data is to track data drift (the extent to which production data is distributed differently from training and test data) regularly and re-train and re-deploy models once data drift exceeds certain thresholds. This discipline should be autoMented by tracking ongoing model accuracy in cases where outcomes (actuals) are available soon after the time of predictions, which may not always be the case (e.g., predicting defaults on term loans). Automated machine learning should also be used to measure model stability ahead of deployment, by making simulations to understand the model deterioration over time.

#8: *How do financial institutions manage AI risks relating to dynamic updating? Describe any barriers or challenges that may impede the use of AI that involve dynamic updating. How do financial institutions gain an understanding of whether AI approaches producing different outputs over time based on the same inputs are operating as intended?*

Most financial institutions manage risks relating to dynamic updating ("online learning") by not permitting it, preferring instead to carry out "batch learning" and updating the batch-trained models at regular intervals or when indicated by data drift/accuracy tracking as described in question #6. To reduce the model validation burden in these cases, an update strategy should be specified at the time of training the original models. For example, implementing triggers/thresholds for retraining the models, and "circuit breakers" describing when models should be taken out of production for more detailed review. Regular refreshes of model validation for models in production, including periodic review of model performance, are also vital tools.

This can be done using MLOps tools built into machine learning software that trigger a retraining. Due to advances in machine learning software, lifecycle management of models is no longer a guess.

#10: *Please describe any particular challenges or impediments financial institutions face in using AI developed or provided by third parties and a description of how financial institutions manage the associated risks. Please provide detail on any challenges or impediments. How do those challenges or impediments vary by financial institution size and complexity?*

Third party solutions, like DataRobot, are a vital part of the ecosystem. Without third party partners, institutions (especially, but not limited to, community institutions) are at risk of inferior optimized results or open source deployments with limited understanding. Use of third party automated machine learning software can help mitigate these risks for small and large financial institutions. The worldwide shortage of skilled data scientists in the labor market further reinforces this need.

While financial institutions may not be able to achieve the same level of transparency with third-party AI solutions as is achieved through developing AI systems in-house, these risks can be managed by rigorously testing system and model performance prior to implementation, across multiple scenarios including edge cases.

Financial institutions can further mitigate such risks by using third-party software to build models in-house while providing the maximum amount of transparency. For example, DataRobot's automated machine learning (AutoML) and automated time series (Auto Time Series) capabilities allow users with little to no coding experience to build models while enforcing data science best practices. DataRobot's software mitigates risks by providing detailed analytics and transparency measures ranging from feature engineering to model selection and tuning to prediction explanations to model monitoring, and guardrails to ensure any model built is trusted and bias is mitigated.

#11: *What techniques are available to facilitate or evaluate the compliance of AI-based credit determination approaches with fair lending laws or mitigate risks of non-compliance? Please explain these techniques and their objectives, limitations of those techniques, and how those techniques relate to fair lending legal requirements.*

Multiple fairness metrics comply with Equal Credit Opportunity Act and evaluate the existence and the level of discrimination in AI-based credit determination systems. Additionally, there are mitigation techniques that can be implemented to reduce the level of bias detected in the AI-based credit determination systems.

There are three main approaches. The first approach, called pre-processing, focuses on the data provided to the AI system. If the data provided to the AI system contains historical bad practices, these techniques assist with reducing the level of bias present in the data. In the second approach, in-processing, during the learning process, the AI system is forced to learn a fairer pattern. These techniques may require modification of the AI system. Lastly, the third approach, post-processing, directly modifies the outcome of the AI-based credit determination to enforce fairness.

For all the approaches mentioned above, the protected attributes need to be present in the data in the AI system at some point in the model building life cycle. Thus, if protected attributes are only available before the modeling and building the AI system, pre-processing methods can be used. However, these techniques are blind to how the AI system infers credit determination, which can reduce its effectiveness in reducing the bias. If the protected attributes are available at the stage when the AI-based credit determinations are about to be deployed, the post-processing techniques can be applied. These techniques modify the final outcomes till some level of fairness is achieved. However, these techniques are changing AI systems' decisions which can result in reducing the systems accuracy. Lastly, if protected attributes can be available during the learning process of the AI system (they won't be used in the system), it is

advised to use in-processing techniques. Since these techniques enforce fairness during the learning process, the AI system learns patterns that are fairer while preserving its level of accuracy. All these approaches can use the fairness metric that complies with ECOA to ensure the fairness and mitigate AI systems that are non-compliant. However, in some cases the desired outcome may not be achievable due significant reduction in AI system's performance or lack of representation in the data available to the system.

#12: ***What are the risks that AI can be biased and/or result in discrimination on prohibited bases? Are there effective ways to reduce risk of discrimination, whether during development, validation, revision, and/or use? What are some of the barriers to or limitations of those methods?***

There are many methods that exist to reduce risk of discrimination. The issue is clearer direction on which is to be employed, barriers to proper monitoring, and the use of run-time assertions that can help restrict model behavior. The risks themselves, and the approaches taken to mitigate them, are broadly comparable to those set out in question #11. As with the above answer, it is of paramount importance that best practices are enforced and that the prohibited classes of data are taken into account when understanding a model's behavior, prior to its deployment.

#13: ***To what extent do model risk management principles and practices aid or inhibit evaluations of AI-based credit determination approaches for compliance with fair lending laws?***

Model risk management principles and practices aid the development of the model and can be expanded using stakeholder theory to include an impact assessment in addition to model mechanics. They provide a structured forum in which to evaluate the compliance of such approaches and guard against the deployment of hastily developed, poorly considered or evaluated models. That said, such processes need to be efficient and agile; if not, the replacement of credit determination approaches which have inbuilt biases with those that do not can be delayed substantially. At present, they typically are not.

#16: ***To the extent not already discussed, please identify any additional uses of AI by financial institutions and any risk management challenges or other factors that may impede adoption and use of AI.***

DataRobot has seen banks take years to build and deploy new AI and machine learning models. The bulk of that time is due to regulatory compliance. Regulations on AI are necessary and ultimately helpful, but the amount of time spent on compliance could be reduced to weeks or days if the regulations were updated to be AI-friendly.

The risk management and compliance officers need one to two years to inspect models for regulatory compliance because the regulations are not model-agnostic. To comply, the bank

needs to inspect a new model down to its finest level of detail. This method of model inspection is far riskier than a model-agnostic approach, in which the government would set expectations with respect to model behavior.

In a model-agnostic regulatory framework, the government would publish a model-agnostic inspection method (such as the partial dependency algorithm), and how the model must behave.

A model-agnostic framework would also require a data specification or inspection method and data protection rules. These too should come with automation that would enable any entity to verify their compliance with the regulation. The issue is that a model may behave correctly with respect to protected classes, such as gender, but that would not be adequate, because inspection of the data could show that gender is actually random, not a true measure of something meaningful.

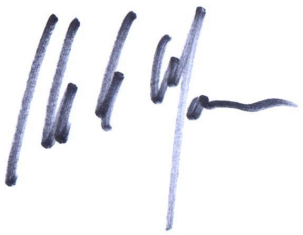
#17: *To the extent not already discussed, please identify any benefits or risks to financial institutions' customers or prospective customers from the use of AI by those financial institutions. Please provide any suggestions on how to maximize benefits or address any identified risks.*

AI is able to decrease bias when compared to human loan officers if the appropriate mitigation steps are taken with the data and features used. However, it is easier to fix a biased algorithm than a biased human. When using AI, it is easier to identify bias, and correct it within an algorithm. The issue is the amplification of missteps if the AI does not have model assurance, so the risk is in governance and monitoring model behavior.

The financial sector is the most sophisticated with respect to the organization required for risk management and regulatory compliance. If the government can provide a model-agnostic regulatory framework and automation to implement it, the financial services industry should gain rapid regulatory relief without a loss of regulatory compliance. The financial institutions have optimized their efforts to the limits of what is possible now, but allowing for automation, the government can set a new optimum that will be substantially less expensive.

If there are any questions, please contact the undersigned at steven.moore@datarobot.com or at (630) 200-6485.

Very truly yours,



Steven Moore
Vice President, Global Government Affairs
DataRobot, Inc.