OakNorth
CREDIT INTELLIGENCE

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

AGENCY: Board of Governors of the Federal Reserve System, Bureau of Consumer Financial Protection, Federal Deposit Insurance Corporation, National Credit Union Administration, and Office of the Comptroller of the Currency (agencies).

ACTION: Request for information and comment

LINK: www.federalreserve.gov/newsevents/pressreleases/files/bcreg20210329a1.pdf

QUESTION 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?

While AI allows the development of models with great expressiveness and adaptability, the complexity of the result can mean that AI models often do not tend to yield to the human the sorts of intuition that other models can. The obvious counter-example would be a simple linear regression where the relationships between inputs and outputs is extremely clear but the model may not accurately capture all the important features of a complex problem space. This can lead to a lack of transparency. How can a user fully trust a model they find difficult to intuitively understand? It can also hide problems such as overfitting, where a model can perform superbly on a training set but fail to replicate this in real-world conditions. Explainable AI[1] is a field of active research which has yielded a multitude of technical means which financial institutions can use to manage the risk in their models such as concept testing, input attribution and others. Risk teams however still generally need to understand this new world so they are better able to exercise their judgement when managing AI adoption. One strategy that can help is to transition gradually and adopt "human-in-the-loop" processes where AI provides efficiency benefits by augmenting existing processes but experienced staff are still able to exercise discretion and thereby maintain control. In such a process, explaining decision-making does not change substantially with the adoption of AI. AI models simply provide an additional data point.

QUESTION 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.

One possible method is using Shapley values – a technique which ML has borrowed from game theory. These values provide a way to explain predictions of non-linear models in such a way as to provide a means of improving the score. This tends to give humans more intuition about the performance of the model and how inputs link to outputs. For example, in the ON Credit Intelligence Suite we have an AI model (Learning to Rank) that given a business will rank a list of other businesses based on how "comparable" they are for the purposes of credit analysis. This is a complex multidimensional ranking problem and the results are not necessarily always intuitive as they include non-structured data inputs such as text descriptions of a business. Shapley values can be used to highlight particular segments of text which explain the score. To a human, this helps to explain the model and can be used to check soundness. Another possible approach is LIME (Locally interpretable model-agnostic explanations).

[1]Doshi-Veles and Kim [2017] Towards a Rigorous Science of Interpretable Machine Learning may be a reasonable starting point for investigating the basic questions https://arxiv.org/pdf/1702.08608.pdf

Continued...

QUESTION 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?

Lack of explainability is a very significant challenge where institutions seek to fully automate decision-making around an AI-based model (e.g. automated lending decisions) over and above the normal challenges associated with conventional model risk management. In such a process, if decisions are challenged, the institution must have a way to respond which demonstrates that the process was fair and compliant with legal and regulatory obligations. This could potentially be difficult if model decisions cannot directly be traced back to specific inputs which led to the decision. Financial institutions can account for and manage these risks via standard model control processes with enhanced diligence around decisions in particular (where models are used for automated decision-making) or by using AI-enhanced human decision-making processes, where AI and other advanced analytical inputs are provided to a human, who has ultimate decision-making authority. In this approach, existing human processes (which already include checks and balances to manage risk in decision-making) can be enhanced with additional data and analytics without giving rise to risk with respect to model explainability. As the human is always in the loop, the onus remains on them to ensure their decision is explainable and model outputs they themselves do not understand can be rejected, leading to a process that should be at least no worse than before the addition of AI.

QUESTION 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.

AI is no different from other statistical approaches in that poor data inputs will lead to a bad model, however as models can learn extremely subtle features of the input data there may be an enhanced risk to poor input data quality. In our experience (as a provider of an AI-enabled credit intelligence solution to financial institutions), institutions deal with this risk by imposing an extremely high bar of acceptance. They do this through extensive human-led quality assurance and testing of outputs and this in turn leads to us as a vendor performing significant sampling and cross-checking of third-party data and rejection of data inputs that seem to be of poor quality. Risk management for alternative data benefits from the fact that very frequently there are additional "traditional" (non-alternative) data sources which can be used to verify correctness. Commonly alternative data sources are a proxy for traditional sources but have advantages such as being more timely, more frequently updated etc. In these cases the correctness of the alternative sources can easily be checked on a periodic basis against the slower, less timely traditional sources to ensure accuracy while also checking that they remain a valid proxy for these traditional sources. As further data sources are added, they can be triangulated against existing sources to look for errors and anomalies.

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

Alternative data, as in all its use cases, increases the power of AI exponentially by providing real-time, frequently-updating training and testing sets. Due to being less explored, it also allows for more creative experimentation to find new use cases and faster adopting models. This is especially relevant in the current environment as the changes that came with the pandemic broke all past correlations, so alternative data allows more timely models

Continued...

to be built quickly. For this reason, it is particularly effective for forward-looking scenario building in fast-changing environments/sectors.

QUESTION 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)?

We use two specific strategies. Firstly, we have a strict policy of model testing against a holdout set which is not used often for validation. So long as the holdout set is not overused, the model should not be able to learn features of this set end therefore it will provide a good final test without overfitting. Secondly, we perform backtesting with stratified sampling with particular focus on areas of the data where correlations are broken. In timeseries analysis for example we would focus on 2008, 2012 and 2020 to test for generalisability.

If the population is completely different you are at the information limit of the model so rather than adapting, look for when the data distribution has changed and decide how best to proceed. In this way we monitor continually for concept drift.

QUESTION 7: Have financial institutions identified particular cybersecurity risks or experienced such incidents with respect to AI? If so, what practices are financial institutions using to manage cybersecurity risks related to AI? Please describe any barriers or challenges to the use of AI associated with cybersecurity risks. Are there specific information security or cybersecurity controls that can be applied to AI?

In our experience, the key risks with respect to AI stem from the fact that AI/ML-based solutions need vast amounts of training and testing data sets to learn from and be effective. It implies that security, privacy and compliance matters are "job zero" for financial institutions and providers of AI and data-driven solutions like OakNorth.

From a cybersecurity perspective, AI systems should be safe and secure throughout their operational lifetime and verifiably so. It is no longer enough to simply follow good practices, but also be able to demonstrate them to customers, auditors and regulators on demand. The usual principles of defence in depth, minimising attack surface, reducing blast radius, least privilege and segregation of duties continue to apply. These principles translate into strong identity and access management, logging and monitoring, infrastructure protection, application and data security, and incident response. Fortunately, recent developments in DevSecOps, Cloud Computing and Artificial Intelligence act as an enabler, rather than a barrier, in implement such controls programmatically at scale. The OakNorth Credit Intelligence Suite leverages these technology developments to deliver reliable and demonstrable security for our customers.

From a privacy viewpoint, customers should have the right to access, manage and control the data they generate. It is also important that AI technologies respect local and industry-specific regulations around data sovereignty (especially international data transfers), data ownership (customers retain ownership and control over their data), data minimisation (processing only data that is strictly necessary), purpose limitation (using personal data for authorised purposes) and data security (technical and organisational measures to protect sensitive data against loss or disclosure).

In our experience, the specific information security controls that are most useful in an AI context are ubiquitous encryption, environment isolation and fine-grained identity and access management. The OakNorth Credit

Continued...

Intelligence Suite achieves these through fully isolated virtual environments for data processing; encrypting all data at rest and in transit; and utilising fine-grained roles and permissions.

QUESTION 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?

At the present we update with a relatively low frequency and completely retrain on update rather than dynamically updating, because reproducibility of results is extremely important for our use case. Maintenance of a validation set when dynamically updating is a significant challenge.

QUESTION 9: Do community institutions face particular challenges in developing, adopting, and using AI? If so, please provide detail about such challenges. What practices are employed to address those impediments or challenges?

Community financial institutions generally lack the resources and expertise internally to develop AI and ML models or model enabled solutions and are therefore almost entirely reliant on their vendor network. This lack of institutional knowledge means that they also struggle to evaluate AI/ML models and tools for both efficacy and soundness. As a result, they have been much quicker to adopt operational AI/ML tools (like OCR), whose results can be more easily verified by an end-user, than deeper analytical models. Where they have begun to adopt more analytical tools, they frequently rely on the brand strength of the vendor. This trust-based adoption applies both to efficacy (how the model is delivering results superior to, and not available from, more traditional methods of data analytics), and soundness (whether the model is consistent, reliable, and unbiased). As such, community banks are almost incapable of evaluating AI/ML technology without an independent framework for assessing models as discussed below and/or a significant investment in personnel resources.

QUESTION 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?

In many respects, the risks for a financial institution associated with AI developed by third parties are not materially different from other areas in which a financial institution may use a vendor service or software product, and the first defences from those fields (model risk management, good vendor due diligence and management, third party assessment where appropriate) still apply. A particular challenge of AI is that there isn't a reliable independent standard or set of best practise guidelines that a financial institution may insist a vendor provide evidence of compliance with. Whereas in cybersecurity for example, an FI may require a vendor to provide a SOC1 or SOC2 report or show how their processes map to ISO27001 or the NIST cybersecurity guidelines (and ask a 3rd-party auditor to assess) in the field of AI there is no such common standard nor might such a thing even be possible. So objective assessment by a neutral party may be difficult, and smaller institutions in particular may lack the internal skills to perform this validation or assessment themselves. For larger, more complex institutions, the skills may exist, but those staff would be in high demand and therefore AI validation may only be possible to prioritise in the most critical applications. That is a clear hindrance to adoption. Finally, the general obstacles to AI adoption described above (difficulty of explanation, lack of transparency, potential problems with overfitting or bias etc) can be more significant for third-party developed AI solutions given that the vendor may

Continued...

be reluctant to expose the inner workings of their models for validation by the financial institution (for fear the FI will simply take their methods and implement the solution themselves).

QUESTION 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.

The most important technique that we know of at the moment is for an institution to use extensive human checking of outputs and ensure there are no undesirable outcomes, especially groups who are systematically discriminated against by the decision-making framework. Human-in-the-loop decision-making processes (as are used by clients of the ON Credit Intelligence Suite) ensure this checking is done. On the face of it, it may seem as if this "passes the buck" by not solving the problem of fair lending in the AI decision-making process, however banks already are responsible for fair lending and therefore this is simply relying on procedures they already have in place given the complexity of the problem from a technical perspective.

QUESTION 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?

While fairness is a very important objective for institutions, it is extremely challenging from a technical perspective to ensure that AI-based decision making is fair given that there isn't a consensus about what fairness means and satisfying certain definitions of fairness breaks others[2]. The problem of fairness in ML particularly is that one of the desirable properties of ML models is their ability to learn features of the underlying data set that may not be intuitively obvious. As such, there have been numerous cases of ML algorithms unintentionally learning features of unstructured data sets that result in outputs that reinforce unfairness in subtle ways[3]. Our approach at OakNorth is to ensure that human decision-makers are kept in the loop. This avoids these problems by making use of the processes that banks already have in place to ensure that groups are not discriminated against on prohibited bases and simply extends them to cover decisions when AI algorithms are in use. In development, validation and revision, the team developing the algorithms may not in fact have the data needed to check for unfairness or discrimination. In particular, in our case (as a 3rd party provider) we don't know the client and in fact don't generally have any information about whether or not the client may be a member of a protected class.

QUESTION 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 help to ensure control over models and can form a natural control point to ensure fair lending considerations are built into AI model evaluation. That said, in the current form it is a matter for each individual institution to do this.

[2]See, for example, https://fairmlbook.org/tutorial2.html for a very good background on definitions of fairness as applied to fairness in machine learning and the tradeoffs between them. It also discusses Chouldechova's impossibility theorem proving this https://arxiv.org/pdf/1703.00056.pdf

[3]https://www.reuters.com/article/us-amazon-com-jobs-automation-insight-idUSKCN1MK08G is a famous example where Amazon scrapped an AI system used to screen candidates due to bias

Continued...

QUESTION 14: As part of their compliance management systems, financial institutions may conduct fair lending risk assessments by using models designed to evaluate fair lending risks ("fair lending risk assessment models"). What challenges, if any, do financial institutions face when applying internal model risk management principles and practices to the development, validation, or use of fair lending risk assessment models based on AI?

Fair lending models are particularly important in the consumer and small business retail markets, where lenders are more likely to automate decisioning. In these cases, transparency into the factors, weights, and biases of the models rendering the decision is especially important. The potential for bias in lending, and violation of fair lending regulation, is amplified by the presence of more personal and demographic data than is commonly considered in middle market commercial lending. As a result, the difficulties in developing, validating, and evaluating models for compliance with fair lending best practices are compounded by the lack of a clear and consistent regulatory framework for considering fair lending in the mid-market commercial space.

QUESTION 15: The Equal Credit Opportunity Act (ECOA), which is implemented by Regulation B, requires creditors to notify an applicant of the principal reasons for taking adverse action for credit or to provide an applicant a disclosure of the right to request those reasons. What approaches can be used to identify the reasons for taking adverse action on a credit application, when AI is employed? Does Regulation B provide sufficient clarity for the statement of reasons for adverse action when AI is used? If not, please describe in detail any opportunities for clarity.

Shapley values and similar techniques used in explainable AI can help to determine the reasons for adverse action on credit as the basis of a model output. These have the benefit of showing which inputs contribute most significantly to the outcome and what an applicate could do to improve the output. As we are not a US lender, and our models produce analytical inputs into a credit decision process (rather than being decision models per se) we have not been subject to a direct US bank examination (and the associated feedback) and don't have an informed opinion about regulation B or any requirement for additional clarity.

QUESTION 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.

The OakNorth Credit Intelligence Suite uses AI for (among others) understanding the structure of borrower documents, finding peer companies similar to a given prospective borrower, forecasting borrower-specific revenue and costs based on macroeconomic, sector-specific and alternative data sources. These use cases help banks to improve their risk management and credit underwriting. At present, a certain lack of clarity around model risk management requirements for AI models does sometimes cause sometimes impede adoption among financial institutions. Any possible increase in clarity around requirements would help to improve this.

QUESTION 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.

When effectively used, AI is a tool that can help a bank to embrace data-driven decision-making in areas that previously had exclusively been the realm of intuition and give confidence and insight, allowing banks to lend to credit-worthy businesses that previously had been underserved and overlooked. The ability of AI to process very

Continued...

large quantities of data can help with speed to decision and a more efficient process that again has the potential to unlock finance for these sectors and businesses. It brings the potential for lending which is tailored to the specific needs of a particular borrower, giving them the service that previously would only have been available to very large corporates. The ability of AI and granular forecasting to provide early warning indicators of potential future distress can help banks to better monitor their portfolios and provide more effective help to borrowers earlier before losses crystalize, leading to better outcomes for both banks and borrowers. Once these granular forecasts are in place, the bank has the ability to aggregate these up and see how a particular scenario would affect their book at the top level and therefore what risk mitigations they need to put in place for the borrowers most affected. All of these are potentially significant benefits to financial institutions' customers or prospective customers.

In commercial lending, the primary risks as discussed above stem from lack of transparency and the associated possibility of bias or unfairness in decision-making. As discussed above this can be mitigated in part by technical means (e.g. Shapley values or LIME) to help explain the findings of the AI models and in part by ensuring humans are still in the loop and able to make final determinations.