Minutes

Artificial Intelligence Public-Private Forum – Third meeting

15 June 2021

Attendees

Co-Chair	Organisation
Ramsden, Dave	Bank of England
Mills, Sheldon	Financial Conduct Authority

Moderator	Organisation
Saporta, Victoria	Bank of England

Member	Organisation
Baldwin, Michael	Google Cloud
Barto, Jason	Amazon Web Services
Browne, Fiona	Datactics
Campos-Zabala, Javier	Experian
Christensen, Hugh	Amazon Web Services
Dewar, Michael	Mastercard
Dorobantu, Cosmina	Alan Turing Institute
Gadd, Sarah	Credit Suisse
Kellett, Dan	Capital One UK
Kundu, Shameek	Truera
Lennard, Jessica	Visa
Moniz, Andy	Acadian Asset Management
Morrison, Gwilym	Royal London
Rees, Harriet	Starling Bank
Rosenshine, Kate	Microsoft UK
Sandhu, Jas	Royal Bank of Canada
Shi-Nash, Amy	National Australian Bank
Tetlow, Phil	IBM UK
Treleaven, Philip	University College London

Observer	Organisation
Beswick, Jacob	Office for Artificial Intelligence
Durkee, Mark	Centre for Data Ethics and Innovation
Yallop, Mark	FICC Markets Standards Board

Apologies	Organisation
Dipple-Johnstone, James	Information Commissioner's Office
Kirkham, Rachel	Mindbridge AI
Mountford, Laura	HM Treasury

As stated in the AIPPF Terms of Reference, the views expressed by the AIPPF members in these minutes and all other outputs do not reflect the views of their institutions, the Bank or FCA. Also, the activities, discussions, and outputs of the AIPPF should not be taken as an indication of future policy by the Bank or FCA.

Item 1. Opening remarks by co-chairs

Co-chairs Dave Ramsden and Sheldon Mills welcomed the members and observers to the third meeting of the Artificial Intelligence Public-Private Forum (AIPPF), which focused on model risk and model risk management (MRM).

Dave Ramsden

Dave noted that mathematical and statistical modelling are not new to financial services, and neither are the risks involved. Model risk can arise in several ways and have implications for customers, firms and the financial system. So managing and mitigating model risk is important not just for individual firms but for the wider stability of the financial system.

Dave highlighted how the use of artificial intelligence (AI) and machine learning (ML) represent a step change for at least three reasons: complexity, speed, and scale. All of which may amplify existing risks and introduce new ones. That is why MRM is becoming ever more important as a primary framework to manage risks related to AI.

Much of the existing guidance on MRM, such as the PRA's guidance for stress-testing models, is focussed on documentation and validation of a model to deliver a particular result. Dave discussed how this approach to MRM is designed for static models and may not be particularly well-suited to AI models, which are more dynamic. Future MRM frameworks for AI models may instead focus on the behaviour or outputs of models once they are deployed.

Finally, Dave acknowledged that there are different approaches to defining, identifying, and managing model risks for different types of financial firms. He explained that the purpose of this meeting, and the subsequent workshops, is to explore the key risks arising from AI models, how best to manage these risks, and what that may mean for the current regulatory framework.

Sheldon Mills

Sheldon underscored how the AIPPF has built a strong ethos of open, critical and honest debate and exchange amongst the members, observers and regulators. Sheldon emphasised this is an important achievement as it builds the foundations for a shared understanding of what the future regulation framework may need to be.

Sheldon noted in particular that MRM may seem a very technical area but could provide a basis from which to develop a broader regulatory approach to AI. With that in mind, Sheldon said it would be helpful to consider current MRM frameworks and how effective they are in relation to the use of AI, including if they need adjusting; how MRM practice differs between banking and other areas of financial services; if there are general MRM principles that can be applied to AI, both from financial services and other sectors; and how we can strike the right balance between providing a framework that allows for certainty, regulatory effectiveness and transparency, as well as beneficial innovation?

Lastly, Sheldon introduced Jessica Rusu as the new FCA representative on the AIPPF as of next quarter when Sheldon will step down from his role. Jessica is the FCA's new Chief Data Information and Intelligence Officer. Sheldon noted that, given her past experience and the remit of her role at the FCA, Jessica is uniquely placed to make a substantial contribution to the AIPPF.

Item 2. Roundtable discussion

The aim of the roundtable discussion was to identify and discuss the key issues and challenges for each of three topic areas, all within the context of MRM:

1. Risks arising from AI models
2. Management of risks arising from AI models
3. Regulatory framework

1. Risks arising from AI models

Key challenges

1.1. Members identified and considered a number of risks arising from AI models categorised within three broad areas: risks to the consumer, risks to the firm, and systemic risks. Some of the key risks included: deterioration of model performance due to incorrect training data, operational risk exposures and change management problems, tacit collusion, and amplification of herd behaviour.

Discussion

1.2. Members agreed that most of the risks related to the use of AI models in financial services exist in other frameworks, processes and domains within society, and are not necessarily new. What *is* new is the scale at which AI is beginning to be deployed and the opacity or complexity of the models.

1.3. Members also said there was an overarching theme emerging around shifting power relationships between individuals, groups and institutions. In some cases, these shifts involve the creation of new power relationships and in others they can widen existing misalignments.

1.4. For example, AI has given firms the capacity to influence, profile and target consumers in a way that hasn't happened (and was not technically possible) in the past. In extremes, this shift in power could significantly disadvantage customers. One member questioned if this could have implications for life insurance underwriting and pooling of risk, for example, since the insurers could potentially know everything about an individual, including aspects that couldn't be analysed in the past. However, another member said that the use of AI models in life insurance underwriting provides a more concise but not totally holistic picture of customer behaviours.

1.5. Within firms, the relationship between the technology teams developing AI models and departments involved in assurance of those models (such as audit and compliance teams responsible for MRM) has shifted. This is because AI is much more complex and opaque than traditional techniques, which means an assurance team's ability to monitor, validate and truly understand AI models is diminished relative to static models.

1.6. Another member highlighted the systemic risks and the potential for networks or clusters of AI models to have a significant and unpredictable impact on wholesale market structure, which may in turn have implications for consumers, firms and the system as a whole. Several members agreed that inadvertent risks can emerge

because there are many unknowns with AI, especially when multiple models interact within a network.

1.7. A key challenge is in identifying when model outputs shift or degrade, especially with reinforcement learning models that can change their behaviour over time. This challenge is often amplified because models are trained separately and in isolation, so it can be very difficult to understand how they will interact and what emergent behaviour may look like. In addition to this, one of the members said there could be an increase in cybersecurity risks to and from AI models which could pose systemic risks.

1.8. Building on the conversation around systemic risk, one member noted there are two distinct categories of risks: intentional and unintentional. Intentional risks arise because people are trying to exploit technological advantages for commercial gain. Unintentional risks occur because of the dynamic nature of model interactions, and can have potential implications for systemic risk. The members debated the merit of using this categorisation for responding to risks, with intended risks warranting stricter actions compared to unintended risks.

1.9. In terms of addressing unintended and systemic risks, several members agreed that the key question is how much emphasis should be placed on a robust monitoring framework both prior to deployment and on an ongoing basis. Often there can be a degree of urgency to using new techniques that firms employ very few metrics to monitor model performance. However, there is such a wide array of effects after putting AI models into production that a singular view or metric isn't appropriate any longer. Instead, firms need to think about the wider impact of models outside the initial field or business area and take a holistic view. This could, for example, involve the first line of defence monitoring each individual model and the second line having a centralised view across all models.

2. Management of risks arising from AI models

Key challenges

2.1. Members considered the key challenges related to MRM frameworks and the general management of risks arising from AI models under three broad topics: mapping and assessing model risk (including dependencies), risk management and controls (including validation), and governance of AI models and accountability.

Discussion

2.2. As the first discussion highlighted, model risks existed before the use of AI but the speed and complexity of models have increased. This means that MRM requirements are higher for AI models to be trustworthy, reliable and secure compared to traditional methods. For example, MRM for AI requires an understanding of hyperparameters; issues like explainability and reproducibility may be harder to address; risks related to data privacy and bias can also flow through into the models and algorithms; and the number of inputs has increased exponentially; all of which makes it more difficult to have a holistic understanding of the model over time and to manage the risks accordingly. The members agreed that all of these issues become even more challenging when using third party models, since there are questions about assurance, control and where the responsibility lies for validation and monitoring of

those models.

2.3. There was strong agreement that the complexity of AI models in financial services is increasing. This is because complexity often corresponds to improved performance and use of ever greater volumes of data. As firms begin to use more complex methods and data sets, and apply AI to more complex use cases, they will become more comfortable with the technology and adoption will further increase. Members also agreed that smaller firms can sometimes use more complex AI compared to larger firms, since they do not have the constraints of legacy systems. In some cases, they may also be incentivised to use more complex AI models to compensate for a lack of data.

2.4. Similarly, the members agreed that complexity is the key challenge for MRM when dealing with AI models. The increase in complexity of both the inputs (some models have multiple input layers and dimensions) and the models themselves (especially deep learning models) means that traditional MRM becomes harder and less effective. Several of the members discussed how monitoring outputs and performance may make more sense for AI MRM, rather than the more traditional MRM focus on assessment of inputs. Members recognised that there are challenges with this approach since the output labels are also becoming more complex and moving from binary to multidimensional (e.g. from yes/no to descriptions).

2.5. Another member pointed out that the relationship between variables can become so complex with AI that the human brain would struggle to understand them. This can also impact attempts to reproduce and audit AI models as part of the MRM process. Even when second and third line functions have the underlying data library and source code, the latter is sometimes altered by the first line to make it less complex for the second line, thereby impacting the reproducibility of the model.

2.6. Several members noted that reproducibility is an important consideration, especially if customers ask a firm about a decision at a later date. However, the scale of AI and data being used by firms can also pose a challenge, since it is not clear what data, models and other metrics should be logged (for example: test data, training data, live business data, source code, explainability metrics like SHAP values, etc.), and for how long (weeks, months, years perhaps), all of which come at a cost to the firms.

2.7. Several of the members noted that although some models can be highly complex, they can also be deployed in low risk applications. Similarly, high-risk applications in financial services tend to use less complex models because explainability for the relevant stakeholders (consumers, internal compliance, regulators, etc.) is very important. The members discussed the merits of a tiered, risk-based approach to MRM that would account for an implicit ratio between the risk AI models pose to an organisation and the complexity they're willing to bear.

2.8. In terms of addressing the challenges associated with complexity, the members agreed that part of the model validation and MRM process should involve challenge around the complexity. For example, the second line function could ask about the business case, the trade-off in using more complex models versus the value added, and whether a simpler model would suffice. On the last point, several members said that this could be addressed by having a challenger model that uses less complex techniques and that this should be part of MRM.

2.9. Members also agreed that differentiating between validation and ongoing monitoring is important for managing AI model risks. Especially because the latter may need to be real-time for certain AI models and conducted by first or second line functions. Traditionally this was done periodically by third line and audit functions – an approach that may no longer be effective.

2.10. Lastly on complexity, one of the members observed that it was likely that firms may use more complex AI models for material and high-risk use cases in the near future as they become more comfortable and proficient with the use of AI. The challenge for MRM is what to do if those models become non-performing or the outputs of those models deteriorate beyond the acceptable risk tolerance. If the firms switch to a different model there may be operational risks and implications for business continuity. However, fixing the models whilst they continue to operate could take time and the models would continue to generate poor outputs. Several members thought that these questions are not being given sufficient attention and that they will become more important as AI adoption increases. One of the members said that failover procedures should be a key part of any MRM regardless of model complexity.

2.11. Another key challenge is around ensuring that accountable executives have an understanding of the algorithms, interactions and risks to fully consider the trade-offs. Leadership is important but there are still uncertainties. For example, what are the right questions executives need to ask without being experts, how can firms ensure appropriate training, and do executive committees need operational layers to ensure compliance and to deliver the right outcomes? These would expand that aspect of a firm's capability for MRM oversight.

2.12. In terms of addressing the challenges to MRM, several members discussed the merits of introducing guidelines specific to AI. These could cover aspects like the appropriate level of validation that is applied to AI models during the development and deployment stages, moving from batch to real-time monitoring, and addressing the need for more detective controls and greater preventative controls. In order to manage the risks, one of the members suggested that a strong culture of experimentation could be built alongside the right guardrails. This could include up-front testing in the development stage, gradual roll-out of new models (testing with 1% of traffic, then 5%, and so on), having a human-in-the-loop where appropriate, and developing automated processes where appropriate.

2.13. With all of the above, the members agreed that documentation is key and there are some emerging techniques that can automatically generate documentation (including the feature engineering aspects) as part of the model development process.

3. Regulatory Framework

Key challenges

3.1. Finally, members considered the regulatory framework and its application to AI models and MRM under three questions: what does a regulatory environment conducive to innovation and AI adoption look like? Is existing guidance sufficient to capture risks associated with AI models, and if not, where are the main gaps? And what is the role of model standards and model auditing?

Discussion

3.2. Members acknowledged that the topic of regulatory responses to AI is highly complex and that they wanted to explore some specific examples to encourage further debate. One example is the recent European Commission (EC) proposal for AI regulation. Members also highlighted that some jurisdictions already have explicit guidance that applies to AI models, like Singapore. Therefore, international regulatory fragmentation could pose a significant challenge.

3.3. Members explained that clarity of regulatory expectations is a key component of fostering innovation. There is some concern within technology and compliance departments that definitions of AI may be ambiguous. For example, the definition provided in the draft EC regulation is so broad that it could include any statistical model, including ordinary least squares regression. The members discussed how the definition could be better distilled by focusing on aspects of the complexity of AI, such as hyperparameters.

3.4. Another area of uncertainty and challenge for compliance departments and potential regulation of AI is the use of alternative data provided by a third party, such as geospatial data or satellite images. It's not clear if this information may be considered insider information when the data are only sold to a small number of firms. Also, could it be considered outsourcing if a firm uses third-party cloud computing storage services to analyse that data and should that activity be subject to the relevant outsourcing regulations? Similarly, what level of documentation is required if a firm processes that data and builds a model in the third-party cloud?

3.5. Members agreed with the documentation requirements of the proposed EC regulation and, in contrast to the wide definition of AI, the narrow definition of high-risk use-cases. On the latter point, one of the members said that financial services regulators could provide more clarity by focusing on specific AI use-cases or activities. These would most likely be the activities, such as credit, insurance underwriting, investment advice, that pose the highest risk to consumers, firms and financial stability. This could also provide a template for other use-cases over time and a more proportionate approach, with lower risk use-cases having lower regulatory burdens.

3.6. Members debated the merits of MRM guidance specifically for AI and whether this would require something entirely new or if existing frameworks can be amended. They acknowledged there are many views from different jurisdictions and the dominant view tends to be that existing frameworks can be enhanced. Several of the members said that explainability and fairness were likely to be two of the areas where existing MRM guidance may need further enhancement. There was also some debate about the right frameworks and mechanisms to allow firms to innovate safely, which could be in updating MRM or even in a sandbox approach.

Item 3. Closing remarks and next steps from the moderator and co-chairs

As moderator, Vicky Saporta concluded by saying that further work mapping AI model risks to AI MRM practices and frameworks would be useful. She also noted that any regulatory framework should aim to address the key challenge of complexity and associated risks.

The moderator and co-chairs thanked the members and observers for the engaging

conversation as well as their continued support. They outlined the next steps, which include preparations for the forthcoming workshops on MRM in Q2 and the Q3 meeting on governance.