As Seen in
Risk and Regulation (*an FT publication)*
March 2024

Four Ways Banks Are Harnessing AI to Manage Model Risk

By Ben Peterson

Humans must always own model risk management, but artificial intelligence can empower senior risk managers to make the necessary judgement calls.

Banks are embracing automation to combat the growing challenges of model risk management, both in their in-house processes and via open platforms such as MLFlow or TRAC.

This attempt has been largely successful, with most banks' model risk data now far more extensive and consistent than even five years ago. Humans are still core to the MRM process though and, as models and data proliferate, this dependence on human observation is becoming a significant cost and risk to banks.

There is no guarantee that staff will have the bandwidth, model skills, or understanding of legacy data flows to fully interpret and correctly document all the required risk parameters. They may also just have a bad day.

But over the past two years, banks have been discovering that many MRM operations are surprisingly amenable to "AI enablement", and that deploying AI to streamline MRM could leave senior staff better equipped to manage risk. While MRM will always depend on human judgement, it may be among the next processes to benefit from AI.

This is still at an early stage, but here are some examples I am seeing of work that's going on right now.

Bucketing models into risk categories

At the very first step in the MRM pipeline, one bank is employing AI not to analyse models themselves but to assign them to risk categories such as customer welfare, compliance, and credit risk. The AI model takes a holistic view of a model's inputs and context and allocates the model's risk across risk types and priority classes before human analysis begins. This streamlines the classification step and helps MRM staff to use their time efficiently.

Model risk: GenAI doing the talking

Another firm is using generative AI to identify similar or duplicate models. Generative AI is used to produce a verbal description of a model, and these descriptions are then compared to find models that may be fundamentally similar. In practice, they may be two versions of the same code, or even two copies of the same spreadsheet.

This clustering technique can identify groups of functionally similar models even when the model's actual code is different; staff can then rationalise and consolidate the models, reducing risk. Caveat: a critical input to the AI is the model's data lineage and the set of key data sources, so strong data lineage and governance is paramount.

Several banks and software vendors are looking to exploit GenAI to analyse the nuts and bolts of models. For relatively simple models, GenAI offers an effective way to convert the parameters of a model into a verbal description – for example, to produce a text description based on the weights of a regression model.

But why?

Explainability, the process of evaluating why a model produced a specific output, can also be boosted by AI. Neural network models in particular are well known for the difficulty of attributing an output decision to a specific input, and Explainable AI – or XAI – techniques such as saliency mapping tend to produce complex technical outputs.

To achieve business-meaningful explanation, one system vendor is focusing on relating model output to specific inputs (for example a document used in training) so that even when the model's internals are complex, the firm can at least attribute a decision to a business artefact. Work on explainable AI continues to advance rapidly, and over the coming years we may see GenAI used to convert complex explainability outputs such as saliency maps into readable statements.

Model risk - making lending fairer

Some families of models, such as loan approval models, present particularly significant regulatory and reputational risk. Lenders are looking at AI tools to enhance the current layers of manual assurance to which these models are subject.

Traditionally, lending models are tested via counterfactual analysis — comparing pairs of inputs that lead to different decisions, and asking "What could have been different about borrower A, to achieve the same decision that borrower B got?"

This process is difficult to automate and fraught with problems. A naive approach to counterfactual analysis might conclude that borrower A would have been approved had they been 15 years younger, failing to realise that they would then have been a minor.

GenAI can improve this process by explaining the difference between two loan applicants in business terms and generating realistic counterfactuals. At least one bank is exploiting GenAI to generate realistic borrower profiles across even extreme market conditions such as a pandemic, thus validating Fair Lending compliance under stress scenarios.

The dawn of AI-driven model risk management

Model risk management by its very nature is about accountable, informed humans seeing beyond the models under their management to glean the bigger picture. Banks and regulators are rightly wary of the idea of "AI-driven MRM".

Yet it is becoming clear that the power of AI to summarise, compare, and group complex entities is of huge value in the MRM process. It empowers staff to apprehend the nature of models, their sensitivity to inputs and their probable future behaviour more clearly than simple documentation alone.

AI enablement in MRM is just starting: no bank can afford to ignore it.

About the Author:

Ben Peterson, Treliant's Data Lead for EMEA, is a technology leader with more than 20 years' experience in Financial Services and fintech. He understands the role that strong data management plays in increasing revenue and reducing risk and believes that data management can have a compelling RoI at both program and organization level. He is a tireless advocate of the power of data governance to create organizational transparency, leading to better compliance and decision making.