

Getting AI Model Governance Right – Banking Domain

Venkat Srinivasan
Consultant

SYNOPSIS

Artificial intelligence (AI) is here to stay and transform the banking landscape at an exponential pace. Banks across the world are marching ahead with the absorption of AI through a wide range of use-cases with a firm anticipation that deployments would reap significant benefits in efficiency, profitability, customer service and consequently the brand value. However, coupled with this rapid move towards absorption there is a parallel element of risk in the inability of the financial institution to keep pace with the regulatory frameworks and have robust model risk management frameworks. This paper looks at banks specifically and the ways and means for these financial institutions to develop model frameworks which mitigate the risk of model failures.

I. INTRODUCTION

The scope of AI includes models in Machine learning, Deep learning and Generative AI. Each has its own flavours of risk which move through the entire life cycle from design to implementation and beyond. Artificial intelligence (AI) is technology that enables computers and machines to simulate human learning, comprehension, problem solving, decision making, creativity and autonomy¹. Associated with this capability is the model which drives this intelligence. Underlying this near human behaviour are models or software applications which include machine learning such as linear regression, logistic regression, decision trees, random forest, support vector machines (SVMs), k-nearest neighbour (KNN), clustering and Artificial Neural Networks which are modelled on the human brain structure and function. Each of these approaches is suited to different kinds of problems and data.

¹ <https://www.ibm.com/think/topics/artificial-intelligence>

All of these models and associated technologies are evolving and are therefore prone to risk and fear of the unknown. This brings us to the central question on two aspects (a) What are the risks and (b) What are the controls and procedures to mitigate these risks.

II. BACKGROUND

To start with let's look at model governance. A governance framework for models brings structure to what is often a sprawling ecosystem of algorithms, datasets, stakeholders and workflows. AI governance framework can be looked upon as a set of guidelines which oversee and managed AI development and application. It provides a clear, repeatable process to ensure that AI applications and products are sustainable and reliable over a period of time².

What are the risks which a bank should control as part of the AI Model Governance Framework:

(a) Algorithmic bias-> Models primarily learn from historical data. There may be patterns and aberrations in the historical data which cause a bias to leak into the model. This can cause an increase in the False Positives or/and False negatives for example when you use a model to automate the approval and extensions of credit. Consequences can be in the form of legal exposure to the bank and losing genuine customers who otherwise would have added value to the portfolio.

(b) Data privacy -> A good AI model can only be as good as the data on which it is trained. Where there is customer data there is risk and there are inadvertent occasions when the data can be exposed and misused. This results in privacy risks which can rapidly translate into legal and compliance exposure.

² <https://www.colibra.com/blog/ai-model-governance-what-it-is-and-why-its-important>

(c) Explainability gaps -> The more accurate or complex a model is, the more difficult it is to interpret and consequently explain in simple terms as to why it took decisions and behaved in a certain way. Deep learning models are black box systems, so if there are for example queries from regulators or compliance agencies the ability to interpret and explain is limited and subject to conjectural conclusions which are unacceptable.

(d) Third-party risk -> compounds when you buy models from external vendors. By doing so the risks the vendors carry in terms of transparency, documentation, explainability become your (the bank's) risk. You're accountable for their failures.

(e) Operational failures -> Most machine learning and AI algorithms work at high speed. The speed of response comes with risks. Before you can say Jack Robinson, your bank would have probably lost millions of Dollars in the bargain. There are also challenges of integrating these systems with legacy applications (core banking) and come up with a robust architecture.

III. MITIGATION MEASURES

The best way to understand the build of a governance model is to take a live use case and map it to proposed steps in the governance process. Let's take the case of fraud detection which a bank wishes to use an AI model for its credit card business.

The first step in the governance model is *assessment and planning*. Here one needs to understand where the business is and where it wants to be.

Current State -> Large number of False Positives and False Negatives and reviews are more ex-post with time delay in identifying genuine cases. Huge manual intervention and fair number of compliance issues from compliance and regulators on timely reporting are visible.

There have been frequent changes in the software which has resulted in patches here and there with limited capability to explain and interpret behaviour. Further there have been instances of absorption of fraud losses since there was a lag in detection of fraud which breached the time threshold by which a chargeback needs to be raised.

Desired Goto -> Decisions need to be accurate (Low levels of FP and FN), near online real-time with clear explanation on what drove the decision and reduction in fraud losses.

The next stage is *framework design*. Here you need to work out the departments who are going to drive the model. First step is that there has to be a model owner who should be identified who takes the overall ownership. Then you need to have a core group which drives the framework and development. This could draw on key people from Credit, Finance, compliance, IT, Information security and Operations who are familiar with AI concepts. Probably engage a third party as well who has rich expertise in AI model, design, risk and implementation. So, from a team perspective you have a rounded 360-degree skill set.

The next stage could be defining how the *development and implementation* would be done. What kind of model is best suitable (logistic, random forest, Clustering. Etc). The framework should be able to define how the business model would work. Which attributes would play a role in identification of the outlier transactions. If you are engaging a third party, they should be able to outline the design flow on the model based on the domain inputs. Coding could be in-house, outsourced or buying an off the shelf product. The model framework should outline the guidelines on taking a decision on this front. The bank needs to document this clearly because this would drive the explainability and interpretability of the model and risks on third party engagement if any.

The next stage is *validation*. This is the most important phase where the bank has to test the model. Now if you go back, you would note that the bank had already implemented a transaction monitoring system albeit with large limitations. So, the bank could readily split this data into training and testing and run the model. Post running you could generate a confusion matrix and check if the errors of FP and FN in the older model have been remediated and new ones do not emerge. Further the new model has FPs and FNs which are within the acceptable confidence levels which should be stated in the model governance document as an exit criterion for the model.

Having developed and tested the model, the governance framework should define as to how it needs to be *implemented* in production. AI models are black boxes, known to lack explainability and

interpretability. A bank needs to be sure with close to 99% confidence that what it is going to be ported to production environment works right, scalable, responsive and accurate. In this specific use case, the bank already has a Fraud Detection System albeit with some limitations. A good way to start with would be to have a six-month parallel run where you run the old and the new model in production and compare to see if there are any significant diversions other than those remediated by the new model. This should satisfy a central bank auditor that what the bank is going to implement has been verified thoroughly not only on test data but also actual production data through a parallel run.

The last stage would be of *audit and feedback loop* where one needs to review the aspects of bias, privacy, interpretability, explainability and operational failures. The model governance document should specify who will do this, how would it be done and what would be the form of reporting exceptions with associated change management guidelines if necessary.

IV. CONCLUSION

Model governance is critical and non-negotiable to get AI implementation done the right way. While there is no 'one size fits all' framework, a bank may need to tweak its core guideline to come up with a good framework on the use case being implemented. A good place to start with to frame the model governance structure would be the EU governance tasks which covers practically almost all of what one needs to tick off before implementing an AI model³.

³ <https://ai-governance.eu/ai-governance-framework/task-list/>