SIMAI : DEVELOPING RELIABLE AI IN INDUSTRY

Cyril Lemaire | Nicolas Girard | Grégoire Martinon
Anh Khoa Ngo Ho | Geoffray Brelurut
Philippe Neveux | Ryad Belhakem

1 MOTIVATION

AI capabilities grow exponentially as the associated risks. A deficit in reliability prevent AI use cases to deliver the intended value:

- Black box AIs are no longer accepted by business experts
- Lack of technical and scientific robustness can cause delays and additional costs for production run.
- The upcoming AI Act regulation prohibits non-certified high-risk classified AIs.

In that context, there is a growing need for the development of frameworks and good practices to implement reliable and responsible AI solutions. This project is a collaboration between Quantmetry, Michelin and the Ecole Normale Supérieure Paris Saclay, with financial support of Région Ile-de-France (AI for Industry grant).

2 OBJECTIVES

The objective of the SimAI project is to study scientific and technical aspects of AI reliability applied to a Michelin use case: the prediction of tire rolling resistance. We identified and implemented technical solutions by developing 2 complementary workflows:

- **Validity domain:** A concept that brings together a series of elementary checks and precautionary indicators that can guide the decision making of a human operator.
- **Explainability**: For an AI system to be fully explainable, it needs to answer the questions of all actors interacting with it. The objective is to propose and implement and explainable AI workflow.

3 VALIDITY DOMAIN

The validity domain comes into play in any machine learning lifecycle and is driven by two main questions:

How to ensure that the input data in production is valid? [1]
How can I trust the prediction? [2]

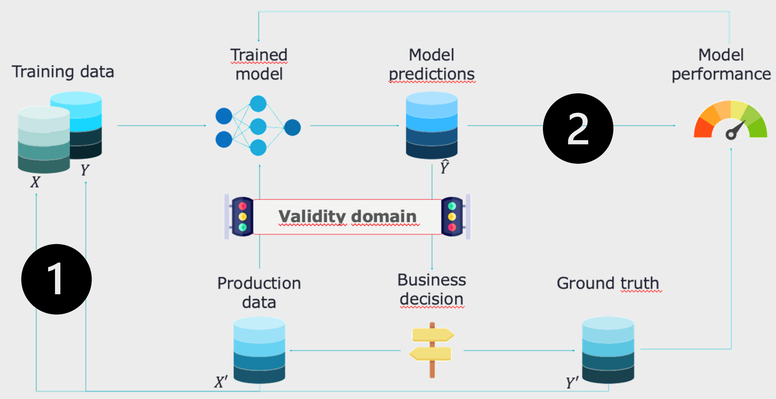


Figure 1: Validity domain within AI's lifecycle.

- **Auditability**: describe your training data and your predictions with words, understand and reproduce all experiments, model training and predictions.
- **Business validation:** set business acceptance criteria to data points and models' predictions.
- **Technical validation:** provide anomaly scores to data points and mathematical guarantees on the prediction accuracy.

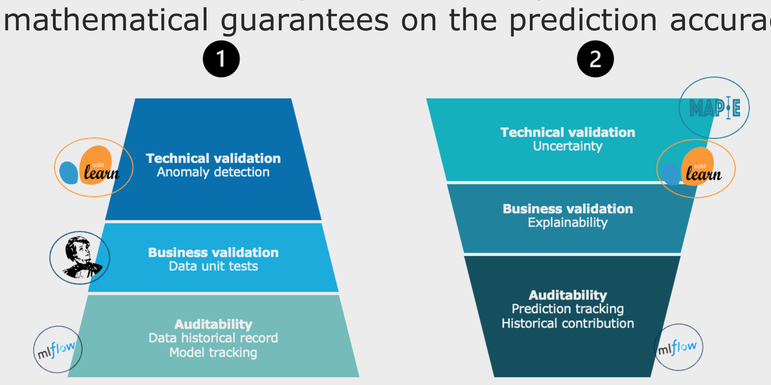
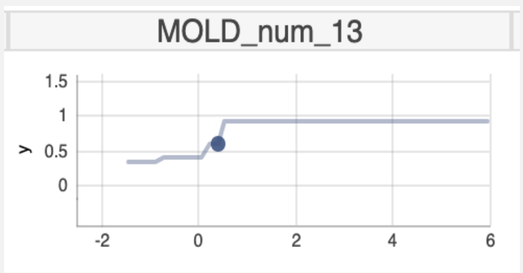


Figure 2: Validity domain pipeline's main concepts and tools for Michelin use case. This pipeline has been developed in an interactive app dedicated for business and data scientists.

4 EXPLAINABLE AI

End User (Material Designer):

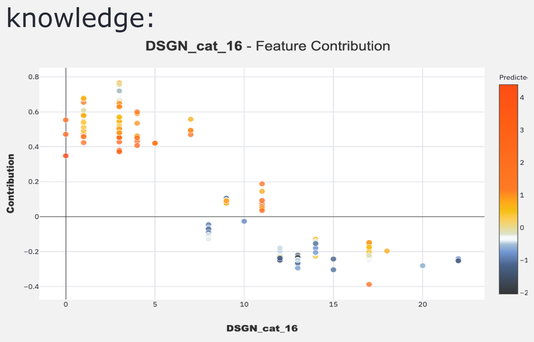
He designs the tire based on model prediction and engineering requirements. He requires local and actionable explanations:



Q: "What happens if I change this?"
A: Ceteris paribus profiles describe the model behavior around a prediction for a given variable.

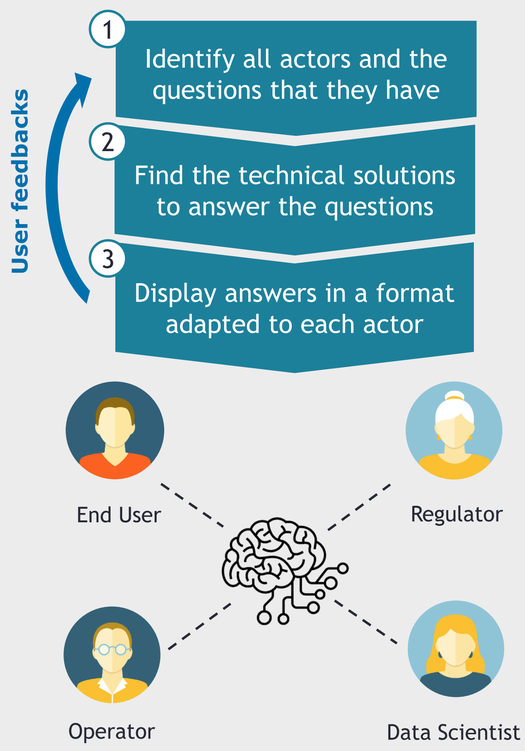
Operator (Performance engineer):

He is responsible for the conception and production of new tires. He needs to make sure the model is coherent with his knowledge:



***Q**: "How does this feature impact my model?"*
***A**: Feature importance plots using Shapash allows to understand the global influence of oneor several variablesof the model.*

The **Explainable** AI workflow developed in the SimAI project has been used to win a challenge organized by **ACPR** to explain a credit risk model in the banking sector. The actors are different between the 2 use cases, but we can still identify 4 common personas.



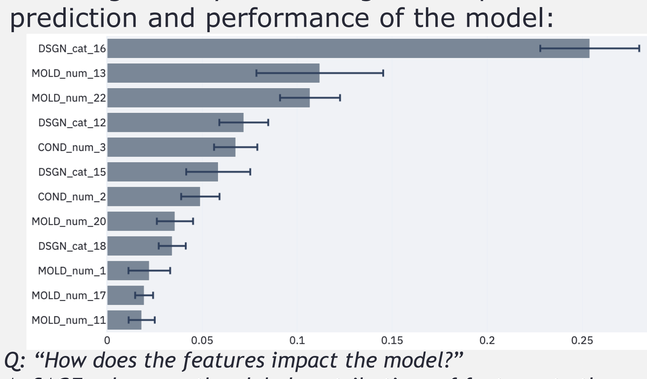
Regulator:

The regulator can either be an internal auditor for quality control or an external regulatory body if the use case falls under AI Act certification requirements.

The typical explanations to provide to the regulator for the audit are likely to be a summary of the explanations for the other actors.

Data Scientist:

The data scientist can use explainable AI to understand, debug, improve and monitor the model. He will generally focus on global explanations for the prediction and performance of the model:



Q: "How does the features impact the model?"
A: SAGE values are the global contributions of features to the model performance. It is a great tool for feature selection or model comparison.

5 CONCLUSION

PRINCIPLED METHODOLOGIES

Validity domain and Explainability are two central concepts to build a trustworthy model and help the business to be more confident with their predictions. Rigorous frameworks including auditability, business validation and technical validation is a way to highly improve AI acceptability and it is a first step to comply with the future AI Act regulation. These framework can be successfully implemented with existing open-source libraries.

NEXT STEPS

We are already working on the next steps of validity domain to comply with the upcoming AI Act regulation by implementing a digital end-to-end process to certify the validity domain of any high-risk AI use cases. Both workflows need to be applied on different tasks and use cases to improve generalization.

UNE SOLUTION TESTÉE