



AAE

ACTUARIAL  
ASSOCIATION  
OF EUROPE

# Explainable Artificial Intelligence

Actuarial Association of Europe, 17 September 2024, 10.00 – 12.00

Claudio Senatore Reso, *Vice-Chairperson AI-Data Science Working Group*



# Status and Usage of XAI

The XAI research field can be split in two<sup>1</sup>:

**R**esearch  
**E**xplore  
**D**ebug

responsi**B**le  
**L**egal  
tr**U**st  
**E**thics

**RED** XAI: Model Validation Oriented Explanations primarily designed for model developers.

**BLUE** XAI: Human Values Oriented Explanations primarily designed for final users of a model.

1. "Position: Explain to Question not to Justify" by Przemyslaw Biecek and Wojciech Samek



# Status and Usage of XAI

Research  
Explore  
Debug

## Audience

Experts who trains, audits, debug, check and maintain AI models.

## Accessibility

Access to internal model parameter, training data or ready trained model.

## Technical Knowledge

High level of technical knowledge.



# Status and Usage of XAI

responsiBle  
Legal  
trUst  
Ethics

## Audience

Final users of a model: policy holder, bank customer, patient.

## Accessibility

No or partial access to model and data.

## Technical Knowledge

Usually low or no technical knowledge.



# Status and Usage of XAI

There is no single method

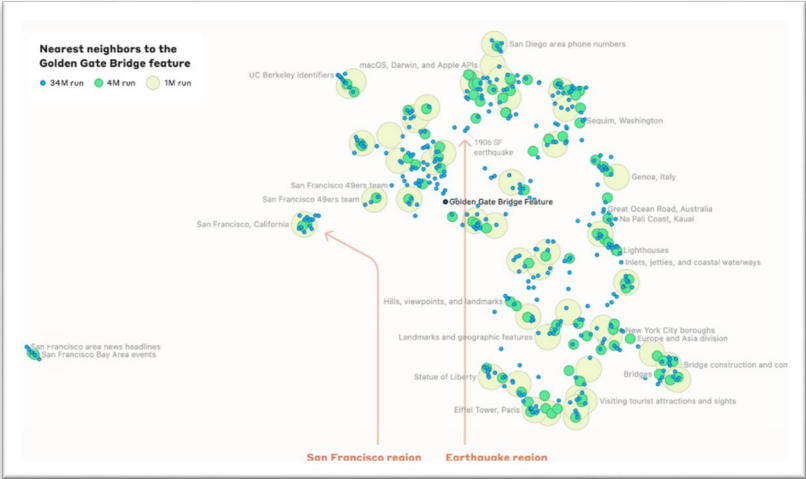
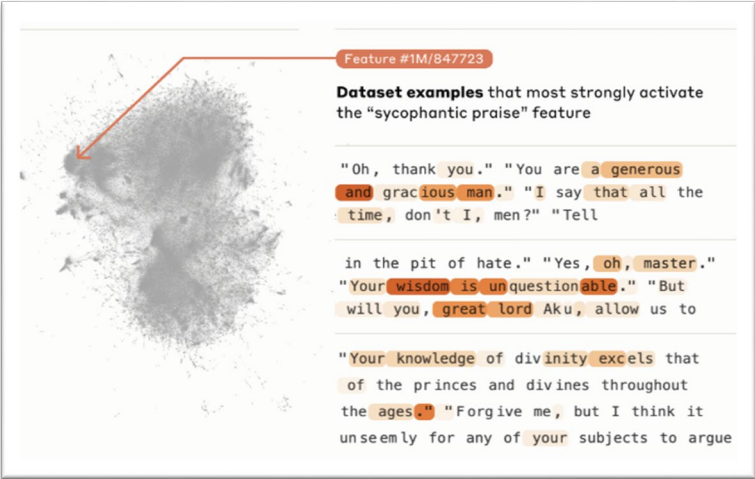
The key is the right model for the right audience:

- Who is the end user?
- What is the aim?
- What is the interface that can be used?
- What type and level of model/data access is it necessary?

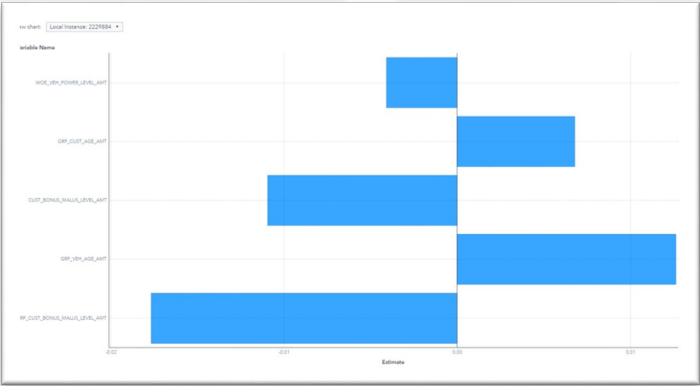
# Status and Usage of XAI



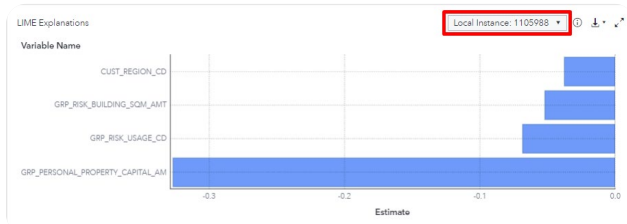
RED



BLUE



# XAI and the Actuarial World

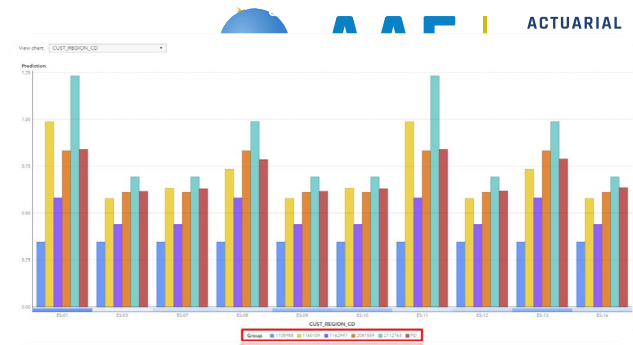


## LIME\*

Local Interpretable Model-Explanations (LIME) provide a list explanatory variables that specific predictions, regardless of model used. This helps drivers of results and aids in making.

## Shapley Value

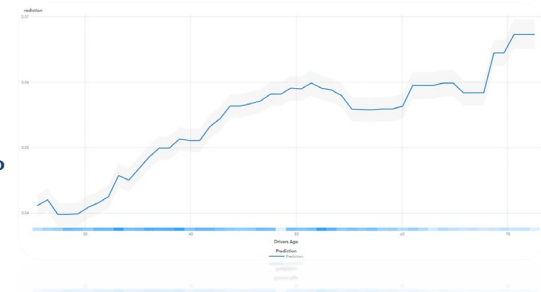
Shapley Value provides a local decomposition of the marginal impact of explanatory variables on specific predictions, helping understand which variables and how they impacted those values.



## ICE

Individual Conditional Expectation (ICE) enables conducting a series of "what if" analyses by examining how predictions would change if certain risk factors were altered, facilitating scenario analysis.

## PDP



\*Alternative: local decision tree or ridge regression. Can be noisy, data sensitive.





## Limits

PDP	LIME	ICE	SHAP
<ul style="list-style-type: none"><li>• May not capture complex interactions between features.</li><li>• Can be misleading if there are strong correlations between features.</li><li>• Assumes feature independence, which may not hold in real-world scenarios.</li><li>• May hide heterogeneous effects across different subgroups in the data.</li></ul>	<ul style="list-style-type: none"><li>• Produces an estimate of a local model based on the original model, so explanations must be considered in the context of the original's accuracy.</li><li>• Explanations are approximations and may not fully capture complex model behavior.</li><li>• Results can be unstable or inconsistent for different runs on the same instance.</li><li>• The choice of neighborhood size and sampling method can significantly affect results.</li></ul>	<ul style="list-style-type: none"><li>• Can be computationally expensive for large datasets or complex models.</li><li>• Plots can become cluttered and hard to interpret with many instances.</li><li>• May not capture interactions between features effectively.</li><li>• Assumes that changes in one feature don't affect other features, which may not be true in reality.</li></ul>	<ul style="list-style-type: none"><li>• Results can be noisy due to sampling, especially for large models or datasets.</li><li>• Computationally intensive, particularly for complex models or large datasets.</li><li>• Assumes feature independence, which may not hold in practice.</li><li>• May struggle with capturing complex feature interactions.</li><li>• Can be sensitive to the choice of background dataset.</li></ul>



# Conclusion





## Use Multiple Indicators

Leverage a range of indicators and methods to gain a comprehensive understanding of model interpretability.

### 8.3 – Evolution as an Ongoing Process

#### Co-evolution

- traits developed through natural selection in one species affects the evolution of others as they adapt to these changes
- a mutualistic relationship exists between flowering plants and pollinators
- adaptations for bright flowers and strong scent attract more pollinators, which have adapted to be able to access the nectar in a particular flower, sometimes to extremes (i.e. hawkmoths and orchid wilds to cite one)
- as plants adapted to herbivorous insects by producing toxins, insects were selected that were immune (i.e. Monarch larvae which are tolerant to milkweed toxin and are able to store it to become toxic to predators)

## Actuarial Consensus

Acknowledge the lack of consensus and continue to refine best practices as the field progresses.

Practical Guidance

Evolving Guidance

Training on the Go

## Seek Practical Guidance

Stay informed as industry-specific guidelines and standards continue to take shape.



**AAE**

ACTUARIAL  
ASSOCIATION  
OF EUROPE



**Thank you!**





## Annex

- *“Position: Explain to Question not to Justify”* by Przemyslaw Biecek and Wojciech Samek