09:45– 10:20 on 15.12.2022

EIOPA - AI Governance 2022

Artificial Intelligence Explainability in Insurance

Alexander Thamm GmbH

Andreas Gillhuber & Dr. Johannes Nagele

European Insurance and Occupational Pensions Authority

Welcome

Your speakers for today.

Andreas Gillhuber
Co-CEO

Andreas has around **30 years of experience** in executive and management positions with a focus on IT and AI, including at BMW, RWE, Nokia Siemens Networks, Siemens and IBM.

At [at], he leads Delivery and is thus **responsible for more than 300 ongoing projects** at our customers. Andreas himself works primarily in data engineering, data ops and data strategy. He is also a book author and speaker at conferences, and serves as CFO for the German Data Science Society e.V.

Dr. Johannes Nagele
Principal AI Researcher & Consultant

With a science background in biophysics and brain research, Johannes has over **10 years of experience in statistics, data science, machine learning, and artificial intelligence**. He combines his many years of hands-on experience with conceptual approaches to the analysis of complex systems.

At [at] he leads the Excellence Cluster on Explainable AI and supports his delivery team in the implementation of numerous cross-industry projects as an expert and team lead.

(Animated presentation of a AI-art comic describing a typical use case of AI for claim processing is being presented. The importance of Explainable AI in Insurance Is being highlighted)

“Black-box” ML solutions replace traditional statistical models

Typical AI cases along the insurance value chain ordered by technology.

	Marketing/Sales	Operations	Claims Management	Risk planning & strategy
Image Recognition & AI Art	**Creating personalized content for ads & brochures**	**Disaster response** Providing real-time information & analysis to assess damage and make decisions about payouts	**Camera inspection** for damage detection	
Anomaly detection		**Cybersecurity**	**Chat Bot** in customer service	
Classification	**Underwriting** Checking eligibility for coverage	**Fraud detection** **Fraud prevention**	„**Google**“ for claims **Workflow optimization**	**Automated immediate regulation** of small claims “Dark processing” **Supply chain management**
Prediction/ Recommender	**Fraud prevention** already during application **Reduction of risk issues** in pricing	**Risk prevention** **Personalized coaching**	**Prediction of high claim load**	**Personalized employee training**

European AI Act & AI Liability Directive

Explainable AI (XAI) for adequate risk management in insurance applications (typically in the high-risk category)

Unacceptable Risk

AI systems considered a clear threat to the safety, livelihoods and rights of people.

- Manipulative, subliminal or exploitative techniques
- Classification of people based on their social behavior

High-Risk

AI systems targeting sensitive GDPR-related topics

- Recruiting-/Employee management
- Safety-critical systems that endanger health in case of failure
- Administration & justice
- Evaluation individuals' credit scores, creditworthiness, insurance premiums

Limited- and Minimal-Risk

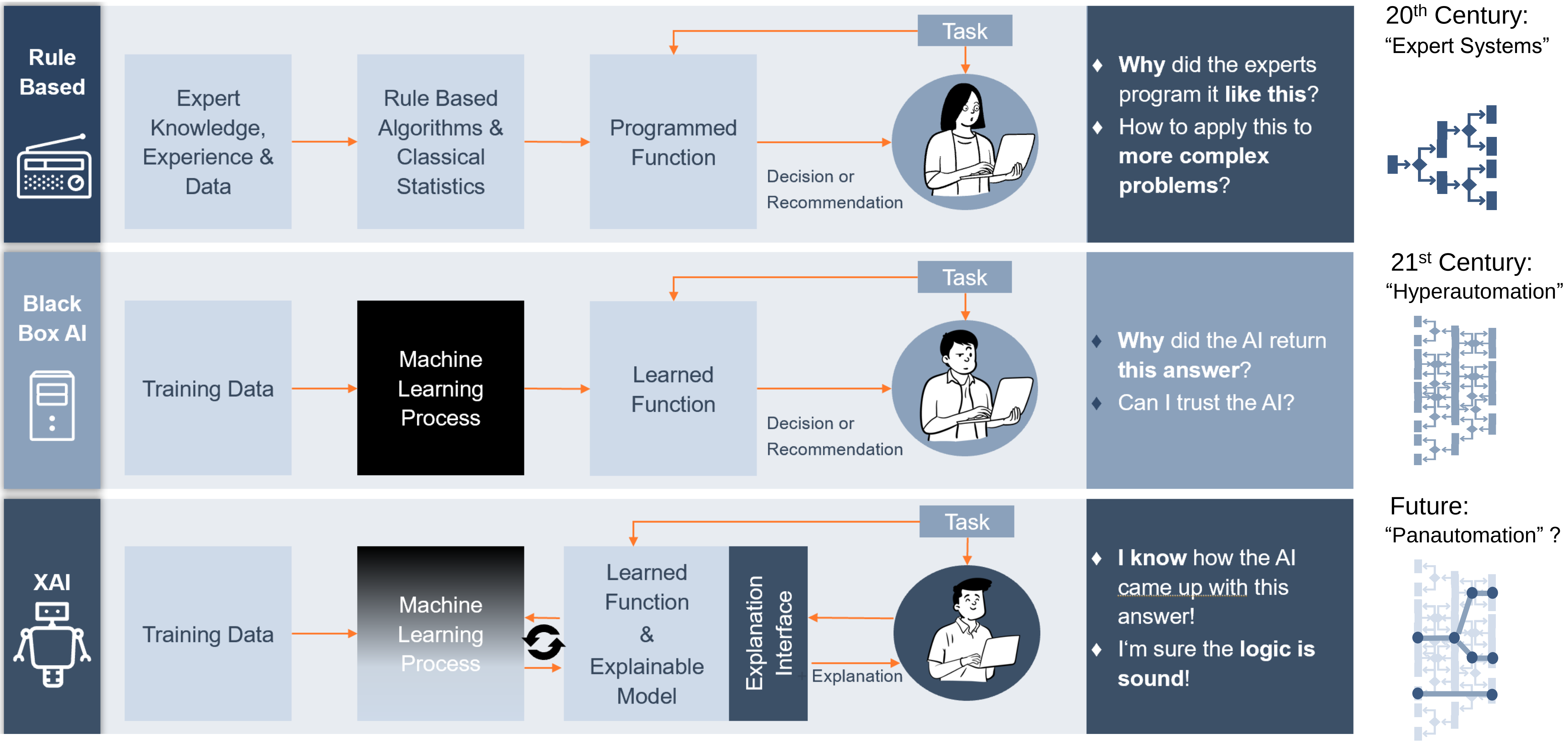
AI systems in day-to-day use not containing critical data

- AI-Chatbots
- Spam filters
- Inventory management
- Market segmentation

[at] alexanderthamm

From rule based data processing to explainable AI

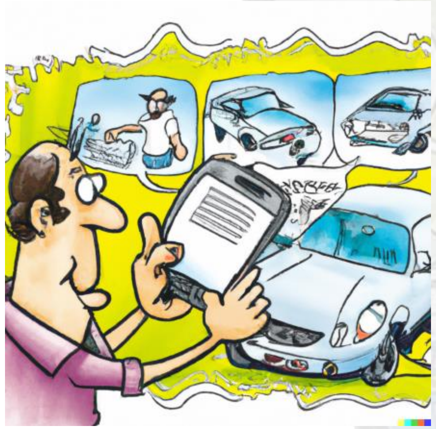
Rules Based vs. Machine Learning vs. XAI



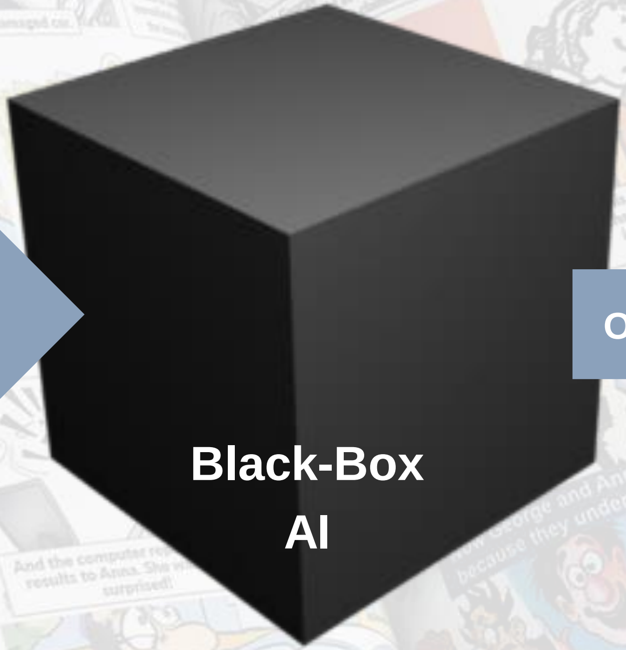
[at] alexanderthamm

Black-Box AI Models

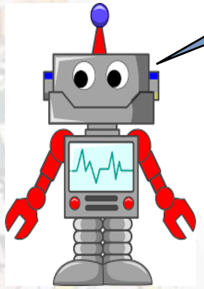
The need for Explainable Artificial Intelligence (XAI)



INPUT



OUTPUT



Trustworthy?

Reliable?

Convincing?

Really?

I mean… sounds good!

Biased??

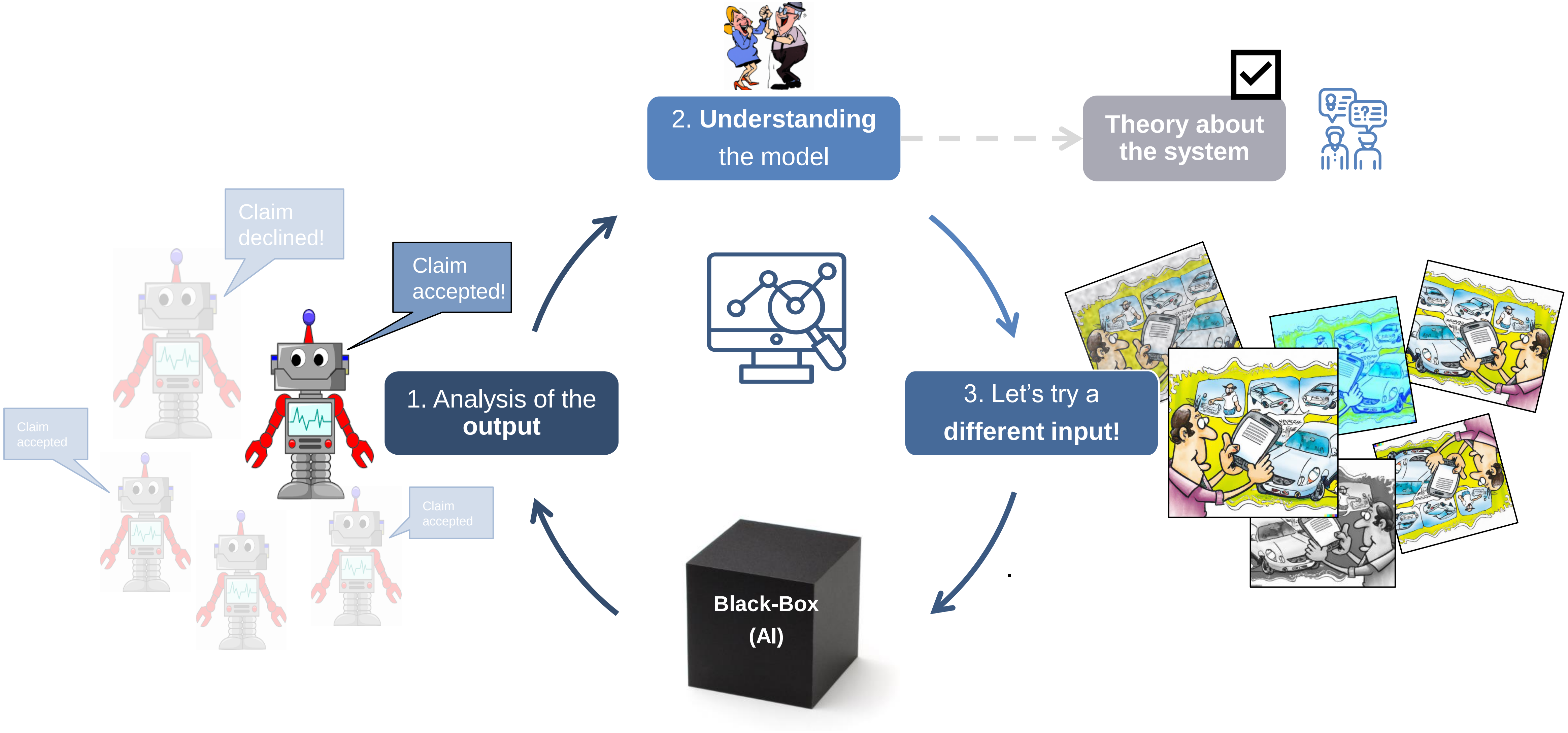
Compliant?

What? No, I do it how I always did it!

→ **Explainable AI (XAI)** for analysis of black-box AI models → **Transparency** and **Interpretability**

How do we achieve explainability?

Natural approach: Systematic exploration of model responses to varying inputs



Example 1: Image Processing & Class Activation Maps (CAM)

Finding explanations in images: Example for damage recognition in car images

Input Data

Image of the damaged vehicle:
Which part of the image influences the ML-model the most?

Pixel attribution

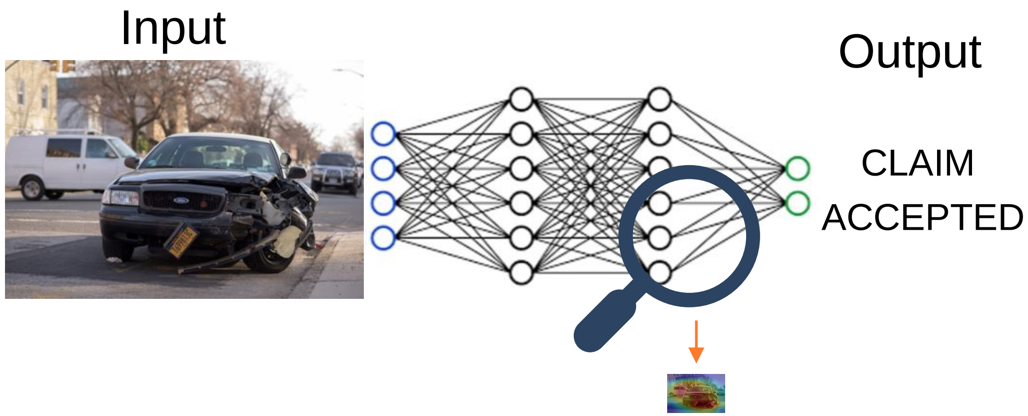


Image explanations:
- Highlights image locations contributing to an explanation
- Traces back a decision to single pixels by reversing the model's analysis of the image
- Alternatively, permutation methods like SHAP can be used

What did the AI look at?

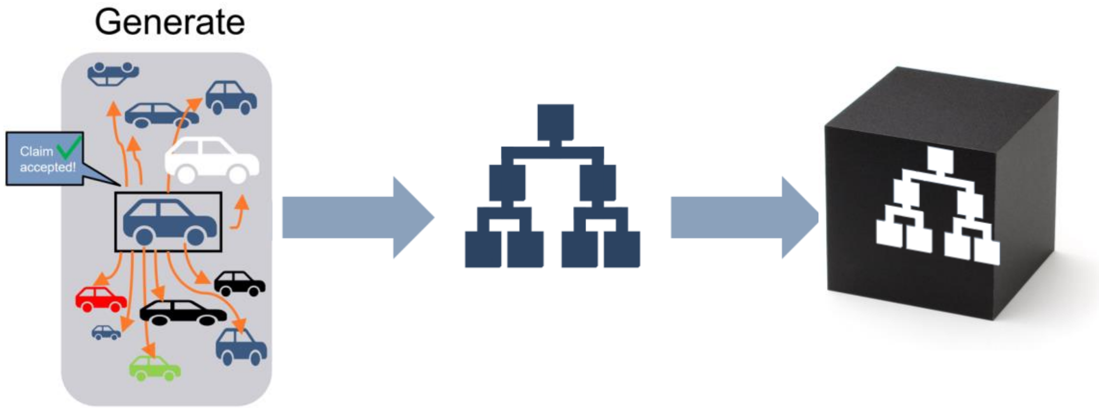
Pros & Cons:
- Easy to interpret
- Easy to implement and use
- Many methods need detailed knowledge of the ML-model
- Doesn't explain general decision logic

Example 2: Local Surrogate Models (LIME)

Transparency by mimicking the complex model by transparent models

Input Data

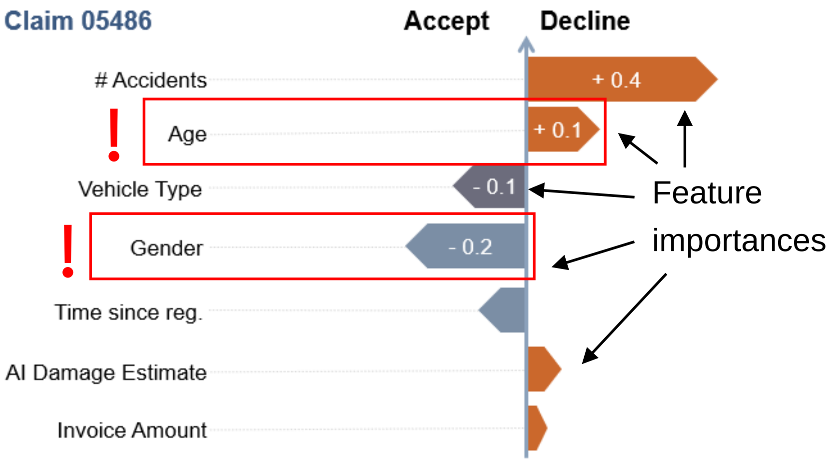
Training simple model on the black-box output



LIME:

- Creates novel samples close to the original
- Trains an interpretable model to emulate the behavior of the black-box model
- The interpretable model is used to explain the decision characteristics of the AI-Model locally

Why was this particular claim accepted?



Pros & Cons:

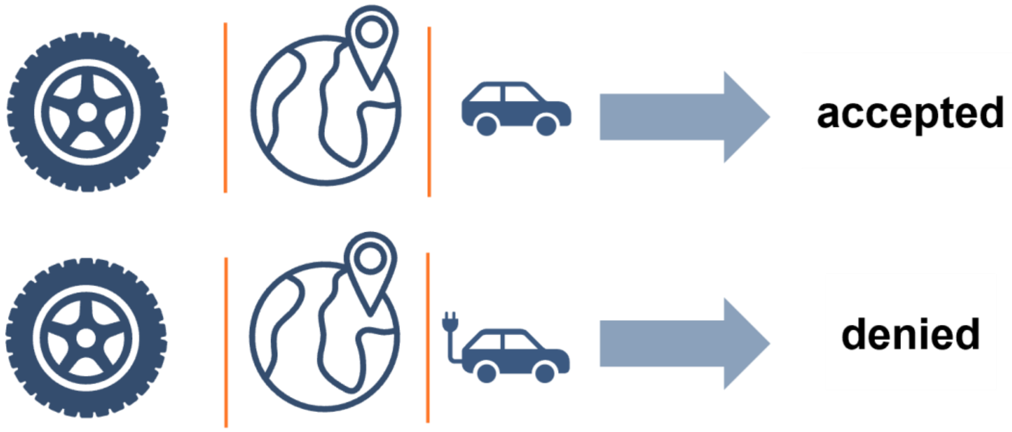
- Versatile when it comes to data (+)
- Easy to implement and use (+)
- Creation of correct novel samples is an unsolved problem (−)
- Ignores correlation between features (−)

Example 3: Shapley Values and SHAP

With “Feature Importances” to a possible explanation

Input Data

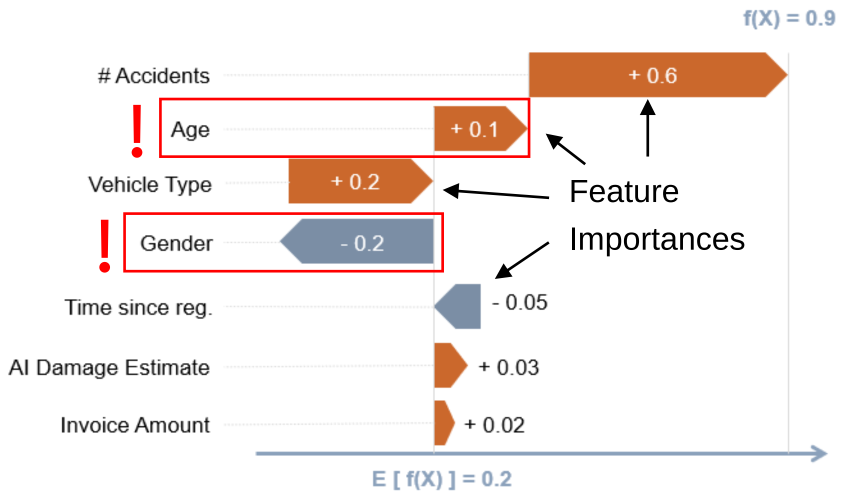
Feature Importance Analysis



SHAP:

- Explains the contribution of each feature of an insurance claim on the prediction
- Based on the theory of collaborative **game theory**
- Data is **permutated and repeatedly tested** to determine the contribution of the features

Why are claims typically accepted?



Pros & Cons:

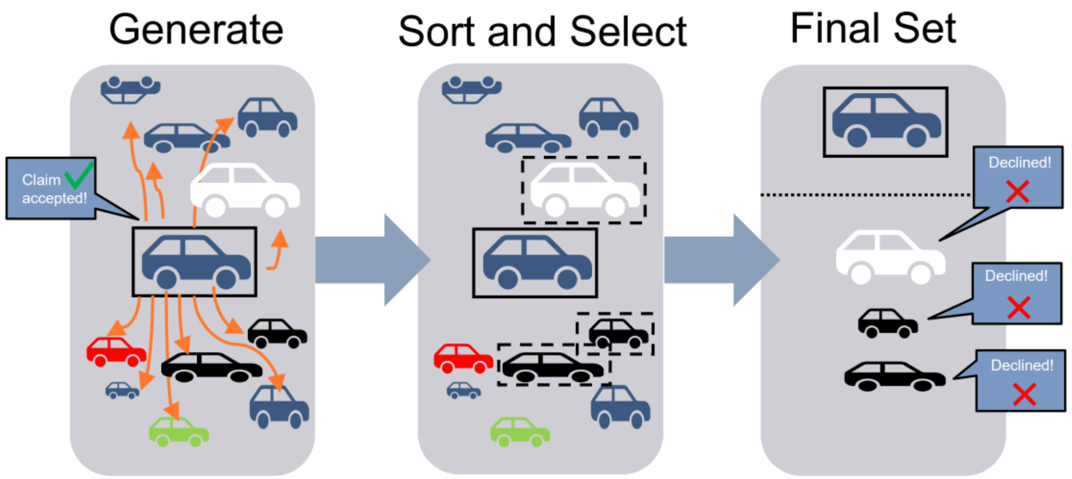
- Returns case specific and general explanations (local & global) ⊕
- Solid theoretical foundation ⊕
- Computationally very expensive ⊖
- Correlated features are problematic ⊖

Example 4: Counterfactual Explanations

Understanding through counter-examples

Insurance Data

Counterfactual Explanations



Counterfactuals:

- Describes the smallest change to the feature values that changes the prediction to a predefined output.
- “**If George was a woman**, the claim would have not been accepted by the algorithm.” → Bias!!!

Explanation

Features	Real Case	Scenario 1	Scenario 2	Scenario 3
Age	55	Age = 55	Age = 55	Age = 70
Gender	Male	Female	Female	Female
Vehicle Type	VW Golf	VW Golf	VW Golf	VW Golf
#Accidents	3	0	3	3
Time since registration	3	3	2	3
Claim accepted	Yes	No	No	No

Explanation:

- Very clear explanations ⊕
- No access to data or model needed ⊕
- Many possible counterfactuals can be correct ⊖

European AI Act & AI Liability Directive

Explainable AI (XAI) for adequate risk management in insurance applications (typically in the high-risk category)

Unacceptable Risk

AI systems considered a clear threat to the safety, livelihoods and rights of people.

- Manipulative, subliminal or exploitative techniques
- Classification of people based on their social behavior

High-Risk

AI systems targeting sensitive GDPR-related topics

- Recruiting-/Employee management
- Safety-critical systems that endanger health in case of failure
- Administration & justice
- Evaluation individuals' credit scores, creditworthiness, insurance premiums

Limited- and Minimal-Risk

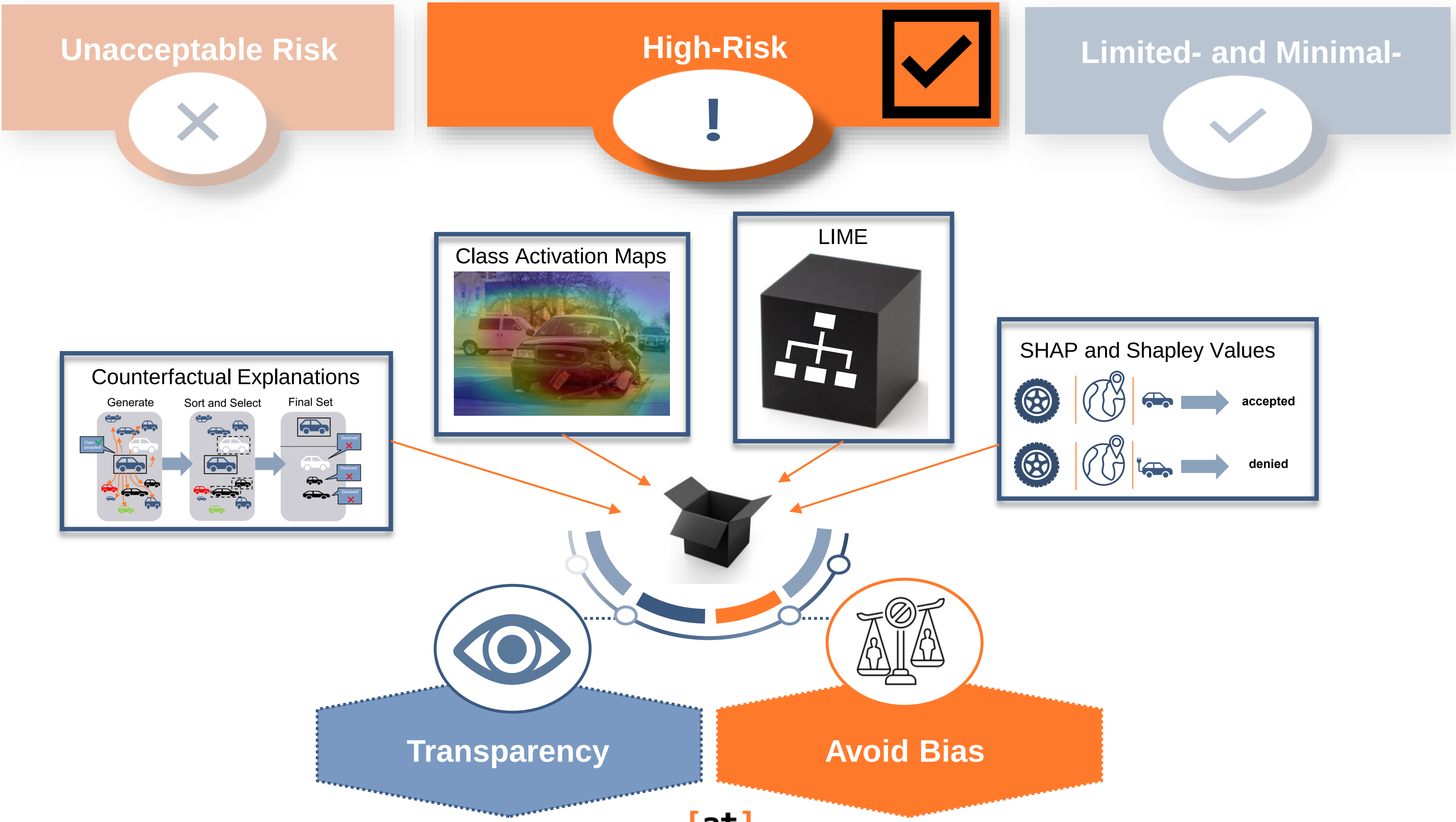
AI systems in day-to-day use not containing critical data

- AI-Chatbots
- Spam filters
- Inventory management
- Market segmentation

[at] alexanderthamm

XAI to comply with European AI Act & AI Liability Directive

The provided XAI-methods open up the black-box and increase transparency for high-risk applications in insurance



Thank you!

Start your
Data Journey

Contact:

Andreas Gillhuber

CO-CEO & Head of Delivery
andreas.gillhuber@alexanderthamm.com

Dr. Johannes Nagele

Principal AI researcher & consultant
johannes.nagele@alexanderthamm.com

Alexander Thamm

CEO
at@alexanderthamm.com

Follow us on:

fb.com/alexanderthammgmbh

Alexander Thamm GmbH

Alexander Thamm GmbH
Sapporobogen 6-8, 80637 München
T +49 89 307 60 880