Responsible AI – (k)ein Widerspruch

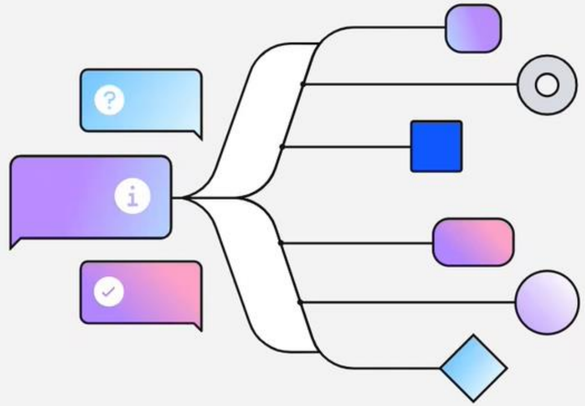
Ani Harreither
+ 43 664 6000 31779
ani.harreither@at.ey.com

Thomas Jirku
+ 43 664 618 7324
jirku@ibm.com


watsonx™

Der Titel heute ist Responsible AI - kannst du vielleicht kurz erklären was Responsible AI ist und warum wir Responsible AI brauchen bzw. für wen Responsible AI relevant ist ?

Before you start chatting, you can update the current settings and ground the chat with documents. To upload documents or an image, click next to the input field.



Sample questions

What are more efficient alternatives to a 'for loop' in Python?

What is the Transformers architecture?

Create a chart of the top NLP use-cases for foundation models.

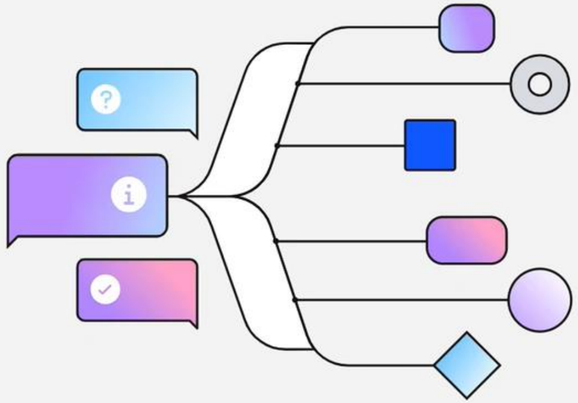
Describe generative AI using emojis.

Type something...

Wie kann man also AI responsible machen ?

Das ist sicherlich sehr kompliziert ?

Before you start chatting, you can update the current settings and ground the chat with documents. To upload documents or an image, click next to the input field.



Sample questions

What are more efficient alternatives to a 'for loop' in Python? →

What is the Transformers architecture? →

Create a chart of the top NLP use-cases for foundation models. →

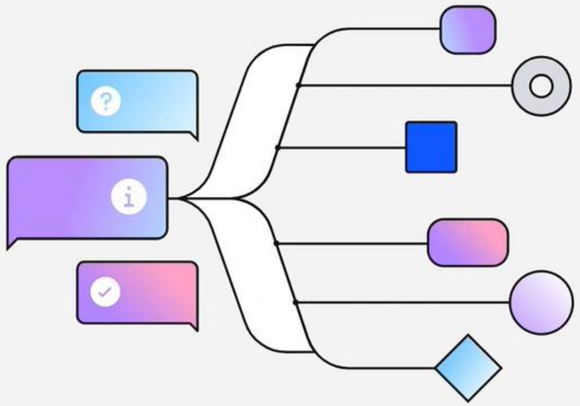
Describe generative AI using emojis. →

Type something...

Hilft der EU AI Act dabei AI Systeme noch vertrauenswürdiger zu machen?

Halte die Antwort bitte kurz.

Before you start chatting, you can update the current settings and ground the chat with documents. To upload documents or an image, click next to the input field.



Sample questions

What are more efficient alternatives to a 'for loop' in Python? →

What is the Transformers architecture? →

Create a chart of the top NLP use-cases for foundation models. →

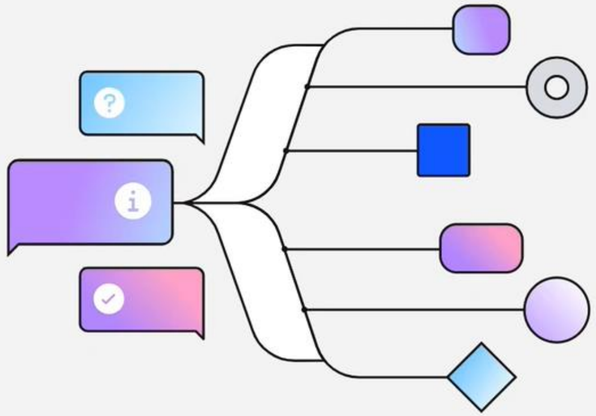
Describe generative AI using emojis. →

Type something...

Was sollten Juristen in Bezug auf den EU AI Act und AI Systemen besonders beachten?

Halte deine Antwort bitte kurz.

documents or an image, click ↑ next to the input field.



Sample questions

What are more efficient alternatives to a 'for loop' in Python? →

What is the Transformers architecture? →

Create a chart of the top NLP use-cases for foundation models. →

Describe generative AI using emojis. →

Type something...

EU calls for responsible development, and EY brings in the journey and frameworks to ensure AI systems are efficient, transparent and compliant

1 Risk-Classification

To-Dos

- Identify AI tools within your company based on AI Acts AI definition
- Classifying risk according to the AI Act (low, medium, high-risk, prohibited)
- Classifying AI models regarding other risk factors like business impact

Accelerator

Use EY's Online Questionnaire For Risk categorization

2 AS-IS Analysis

- Analysis of Roles & Responsibilities regarding efficiency & compliance
- Analysis of AI Procedures regarding efficiency & compliance
- Technical Analysis of AI models:
  - Data Quality & Reprocessing
  - Validation & Testing Mechanism
  - Model Explainability & Interpretability

Using EY's AI Maturity model for assessing current status

3 To-Be Definition

- Introduce specific requirements for high-risk models
- EY's North Star:
  - Roles & Responsibilities (RACI-Matrix)
  - Procedures for the AI Lifecycle (AI Procedure)
  - Technical solution (Confidence-Index)

EY's RACI Matrix

EY's AI procedure

EY's Confidence-Index

4 Roadmap

- Developing a Roadmap for closing identified gaps, including timelines and milestones
- Defining short-term and long-term actions to address immediate needs
- Resource planning to determine the required FTEs for both short-term and long-term efforts
- Training for ongoing improvement of the whole AI Lifecycle

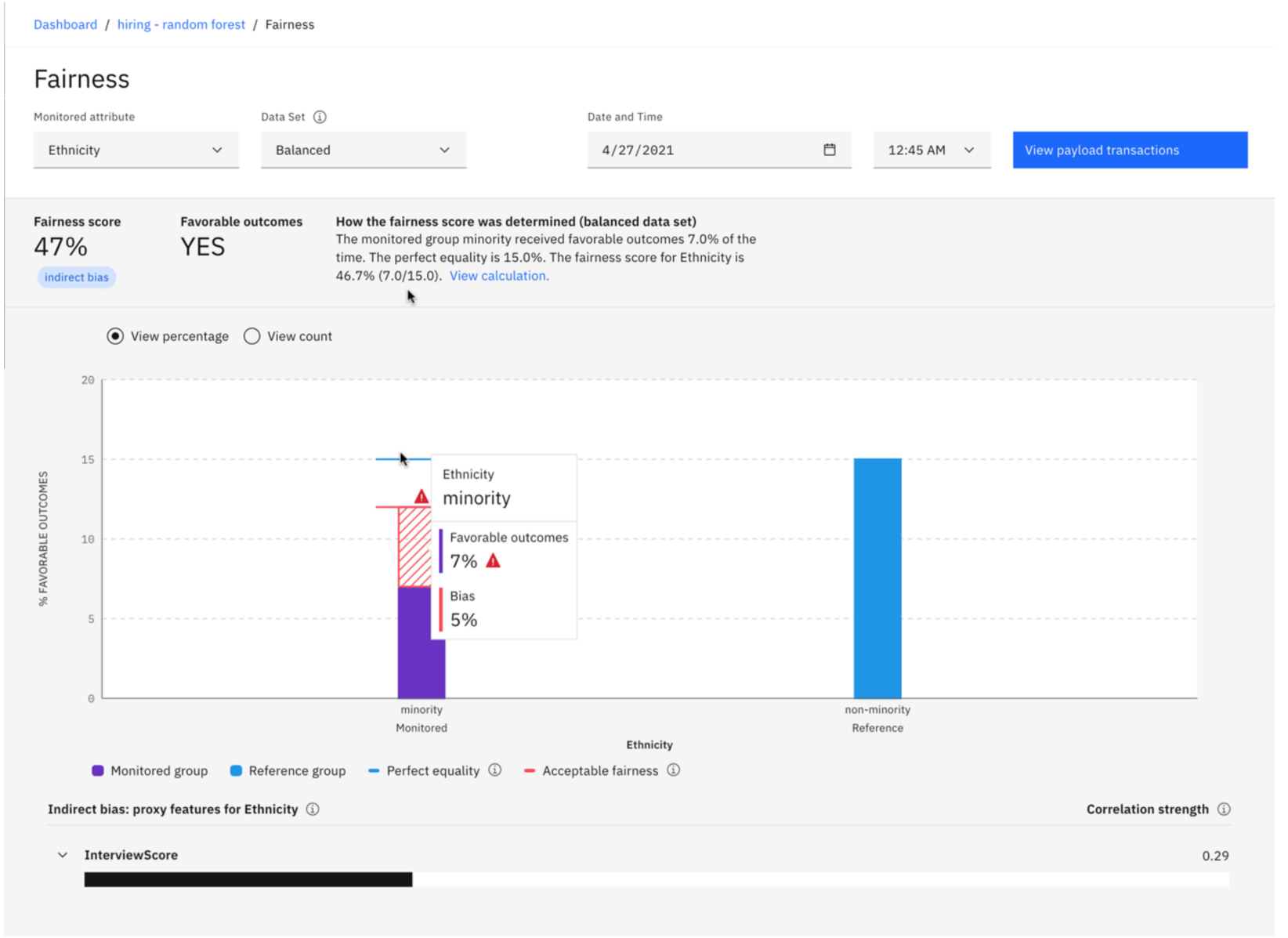
EY's Action & Resource Plan Framework includes Short-Term & Long-Term Actions & Resource Planning

Bias detection

Ensuring fairness in model scoring

Continuous calculation of model fairness

- Analyze deployed model predictions for bias
- Collect and aggregate bias data for dashboards and alerts
- Find non-feature data correlations
- Use a corrected model for “de-biased” predictions



Model Explainability

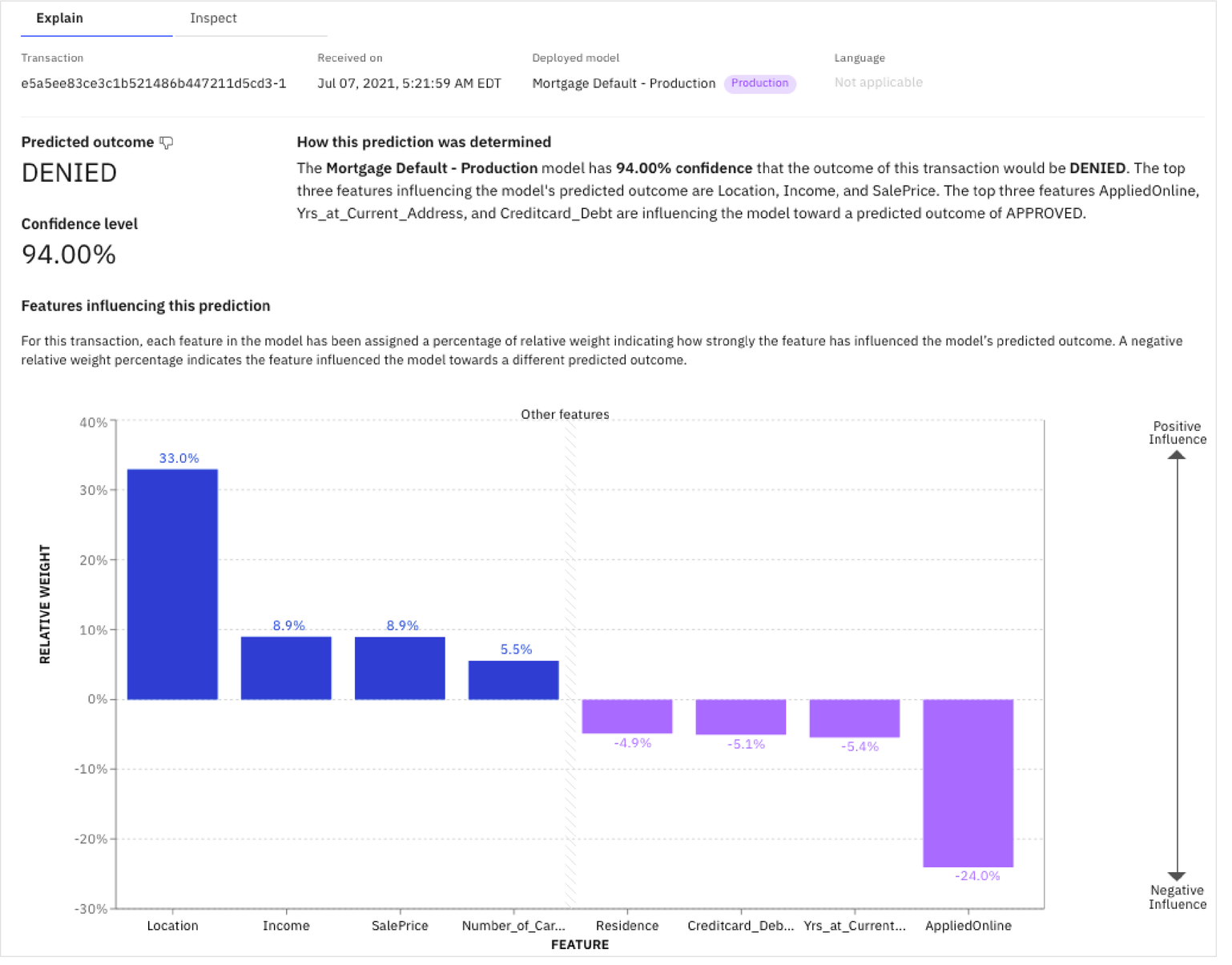
Explain model predictions

- Show the most influential features
- Explain in natural language
- Available API for prediction explanations

What-if analysis

- Experiment with values
- Assess effects of changes to features

Understand model outcomes



Drift detection

Handle changing scenarios

Measure the degree to which a model has moved away from reality

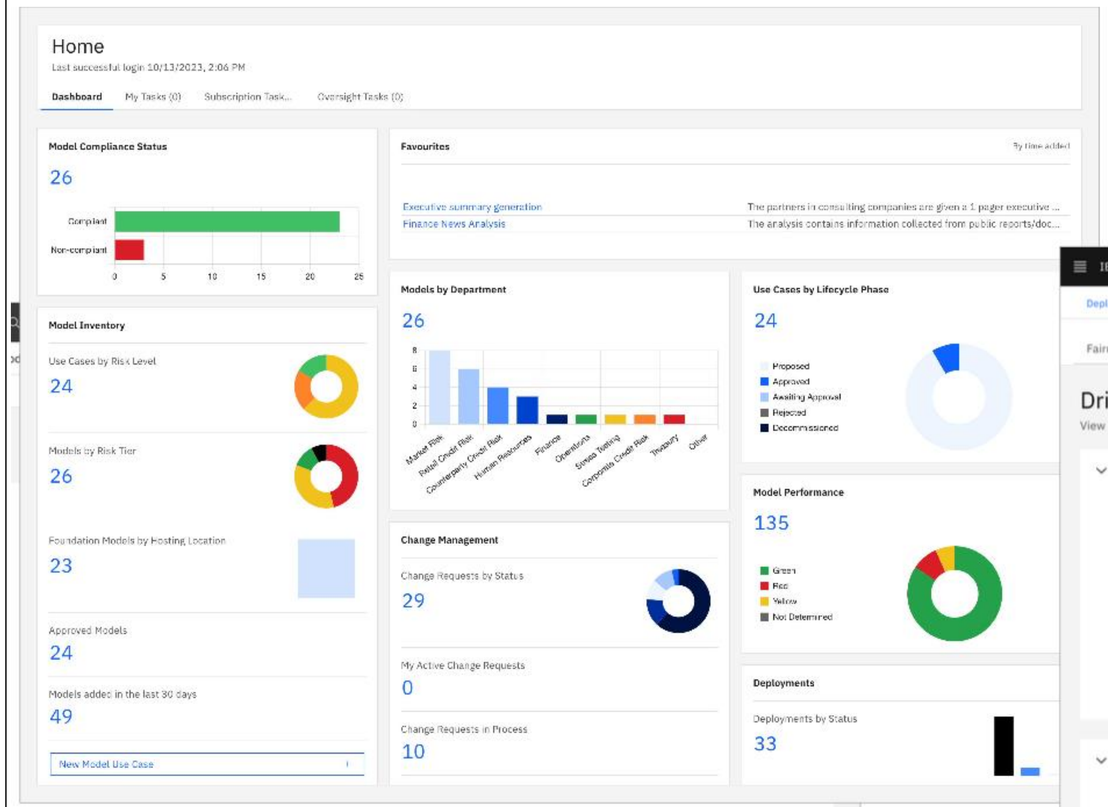
- Drop in accuracy – reality has changed, as shown by the scoring data
- Drop in consistency – reality is the same, the events vary

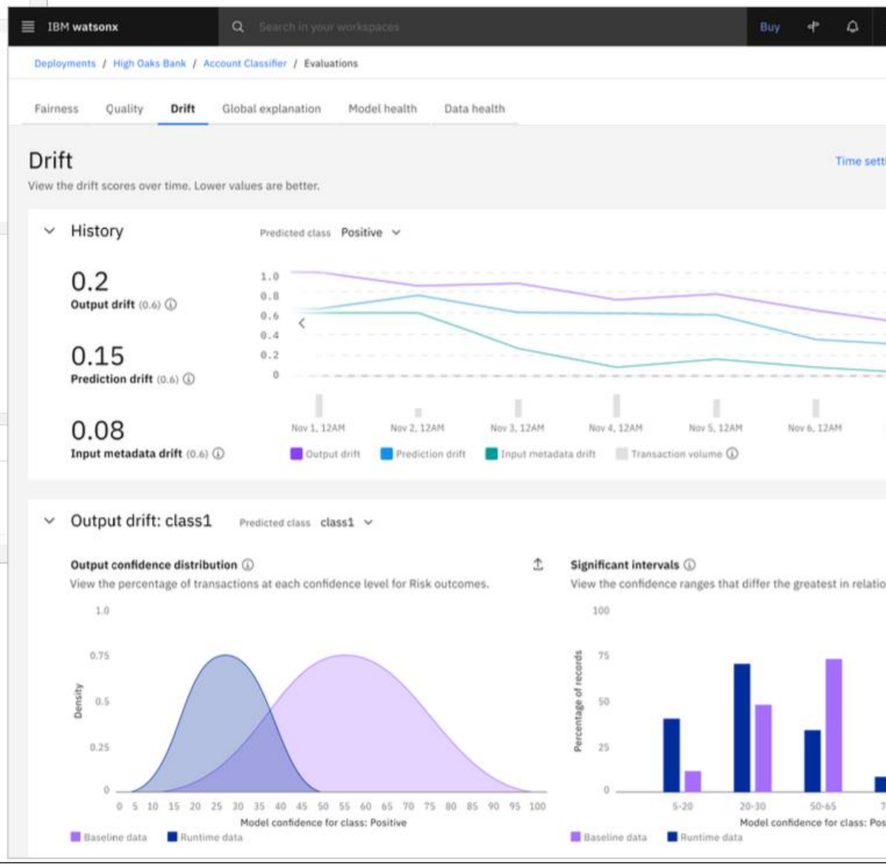
Drift monitoring and alerts

- Degradation of model performance can trigger retraining and redeployment

Manage risk across the enterprise with IBM

- Consistent holistic views of risk and compliance
- Drive GRC adoption
- Embedded self-service reporting, analytics, and dashboarding





Thank You

Ani Harreither, EY
+ 43 664 6000 31779
ani.harreither@at.ey.com

Thomas Jirku
+ 43 664 6187324
jirku@ibm.com