
AWS
re:Invent
NOV. 28 – DEC. 2, 2022 | LAS VEGAS, NV

AIM328-R

Explain your ML models with Amazon SageMaker Clarify

Pinar Yilmaz (she/her)

Principal Engineer
AWS

Chris Knoch

Principal Project Manager
AWS

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

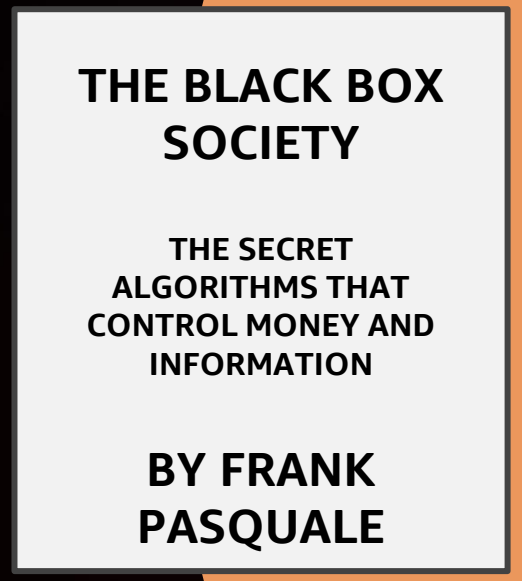
Question – How big is the responsible AI opportunity?

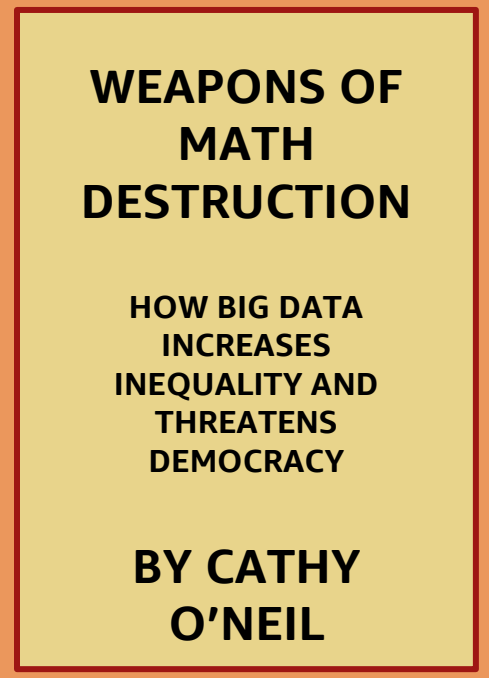
Where will the responsible use of AI rank in public discussions in the coming years?

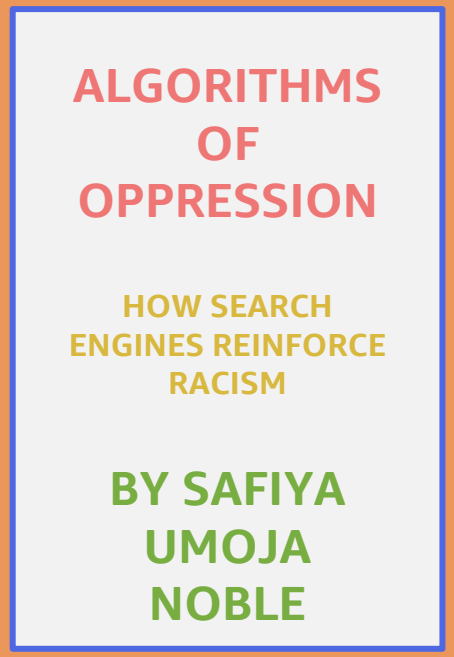
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Algorithmic bias

- Ethical challenges posed by AI systems
- Inherent biases present in society
- Reflected in training data
- AI/ML models prone to amplifying biases



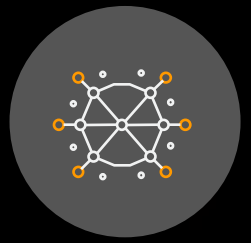






aws
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Responsible AI challenges today



Complexity increasing



Low transparency with black box models



Cost of missteps increasing



Focus on “why?” and “how?” over “what?”

aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Responsible AI motivation and business opportunities

Regulatory. Regulations may require companies to understand why the ML model made a given prediction and also whether the prediction it made was free from bias, both in training and at inference (e.g., **financial services, healthcare, and public sector**)

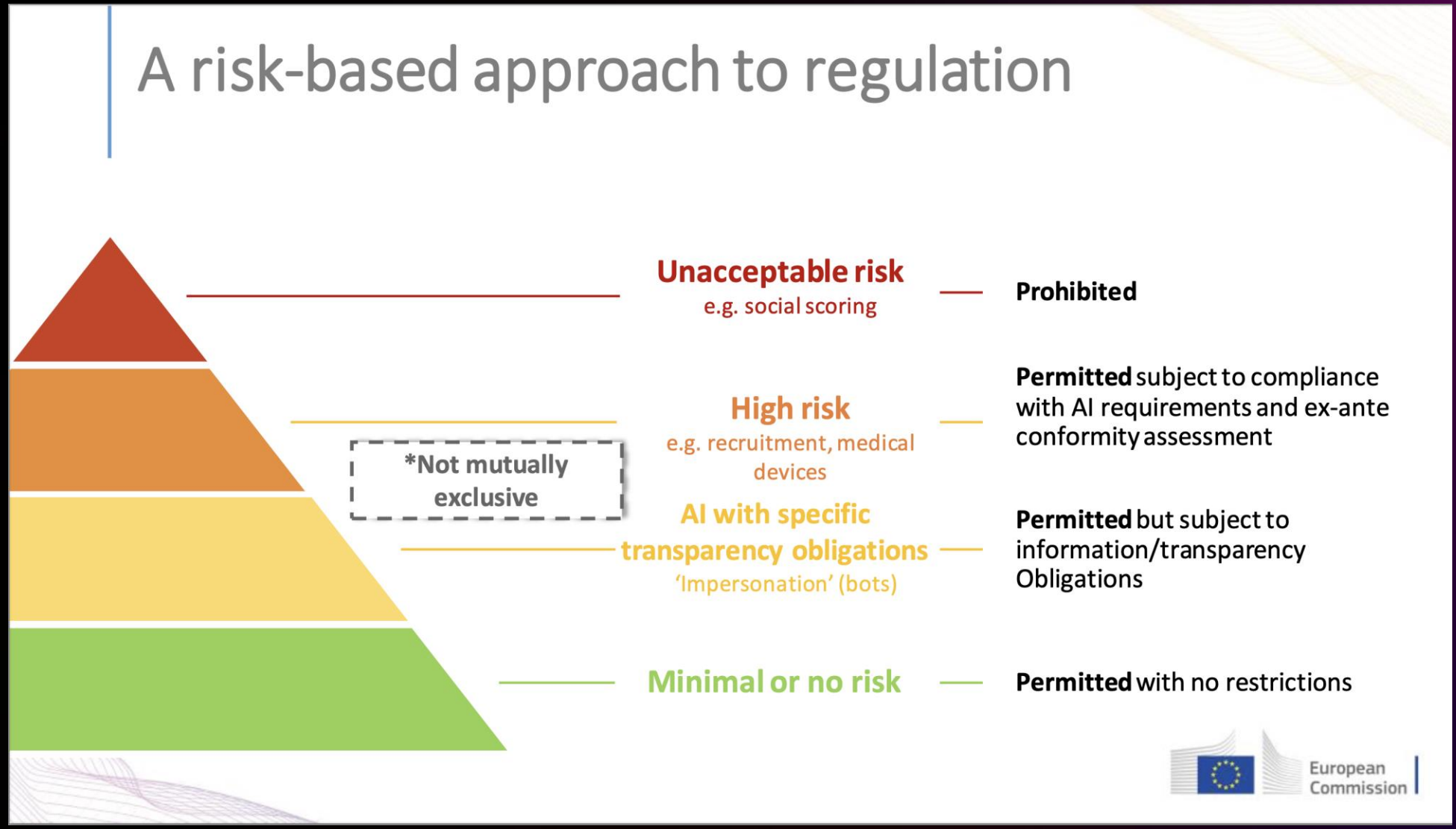
Business. Providing explanations to internal teams (loan officers, customer service reps, compliance teams) and end users/customers to drive adoption and improved outcomes

Data science. Improving models through better feature engineering and training data generation, understanding failure modes of the model, debugging model predictions, etc.

aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Regulation example – EU AI Act



Source: A European Strategy for Artificial Intelligence, CEPS webinar, Lucilla Sioli

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

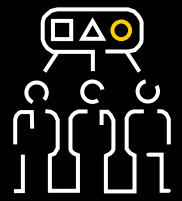
Question – Are there frameworks to help chart a long-term plan?

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

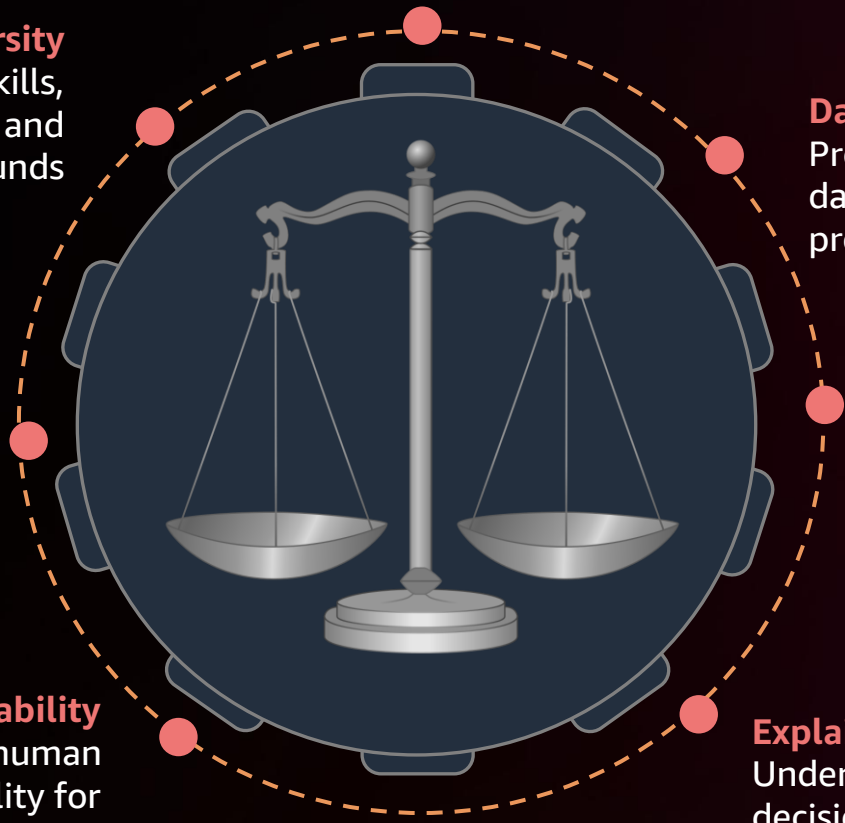
Pillars of responsible AI

Value alignment
Systems should be designed and used in ways that align with company mission, social norms, and legal compliance





Diversity
Inclusion of unique skills, experiences, perspectives, and cultural backgrounds



Data privacy & protection
Protects the quality and integrity of data used, its relevance, access, and processing





Training & education
Appropriate knowledge sharing and education to understand purpose, use, and impact

Fairness
Systems must be designed to minimize bias and promote inclusive representation





Accountability
Structures maintaining human involvement and responsibility for design, development, decision processes, and outcomes

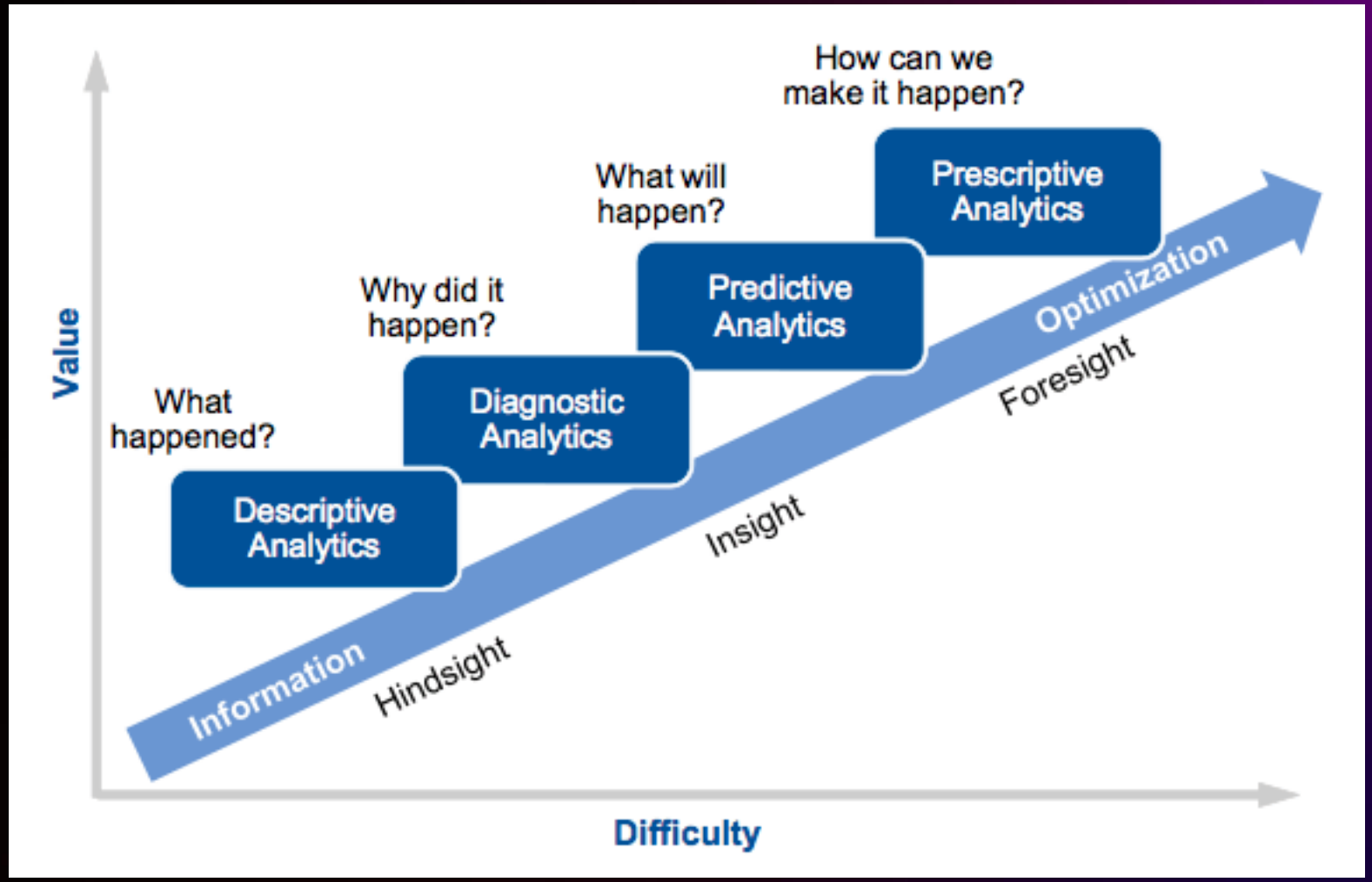
Explainability & transparency
Understanding how data is used, how decisions and outcomes are made in an human-understandable way



aws
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Predictive analytics maturity cycle

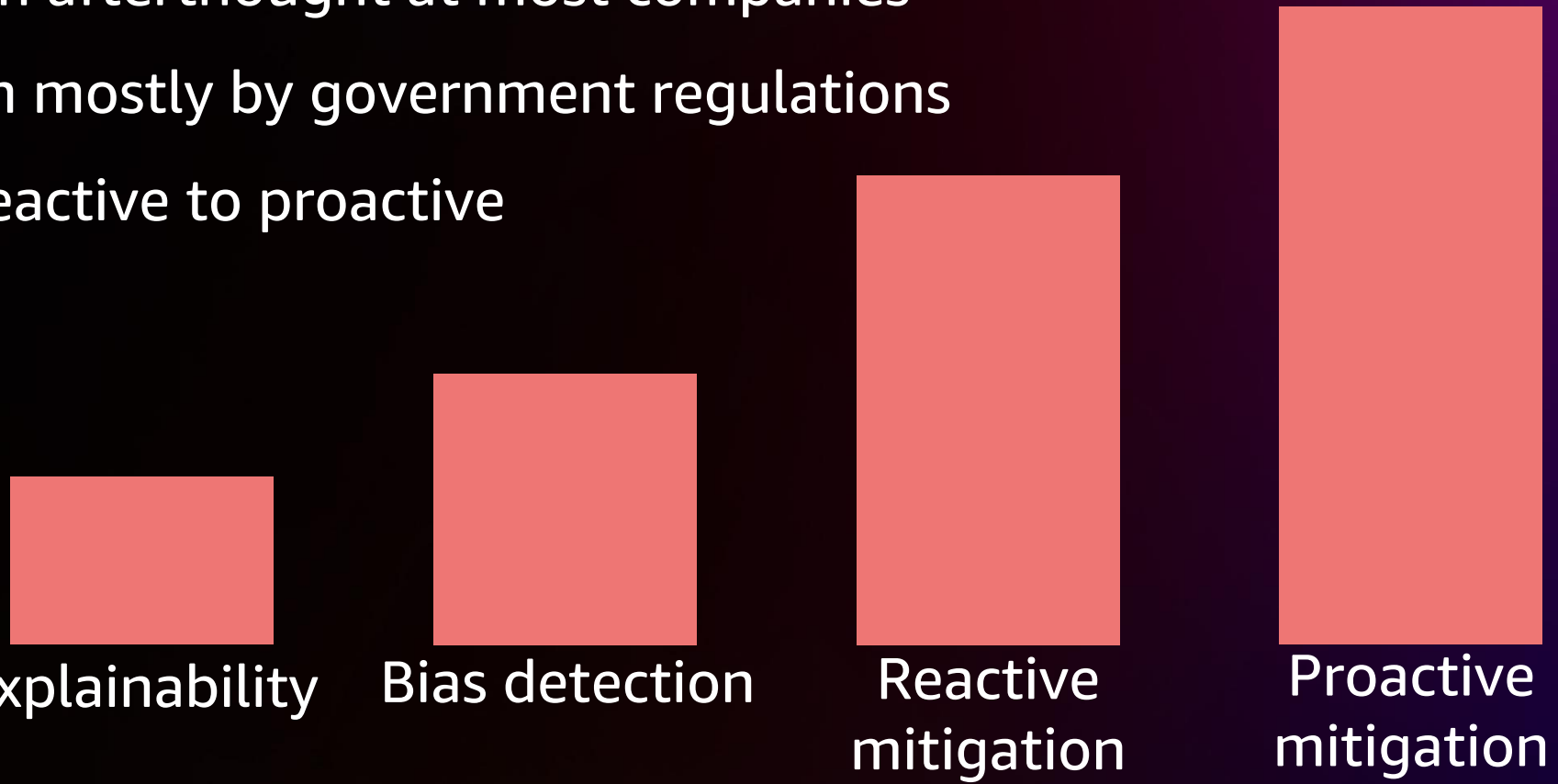
- Current solutions address “what?” questions in descriptive and predictive analytics
- **Explainability** is needed to address the “how?” and “why?” questions in diagnostic and prescriptive analytics



© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Bias detection and mitigation maturity cycle

1. Explainability (aka model understanding) drives the most immediate value
2. Bias is currently an afterthought at most companies
3. Adoption is driven mostly by government regulations
4. Transition from reactive to proactive



aws
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Question – How is Amazon SageMaker Clarify approaching these challenges?

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Amazon SageMaker Clarify

Explain machine learning models and assess your data and models for bias



Detect bias during data preparation

Detect bias across attributes in your training data utilizing nine different pre-training detection methods



Detect bias in your trained model

Evaluate the degree to which various types of bias are present in your model utilizing nine different post-training detection methods



Explain both global and local predictions

Understand relative importance of each feature to your model's overall behavior and individual predictions



Explain decisions made by CV and NLP models

Understand how CV and NLP model predictions are being made in unstructured data models that are typically "black box" in their approach



Detect drift in bias and model behavior over time

Provide alerts and detect drift over time in SageMaker Model Monitor to understand changes due to dynamic real-world conditions



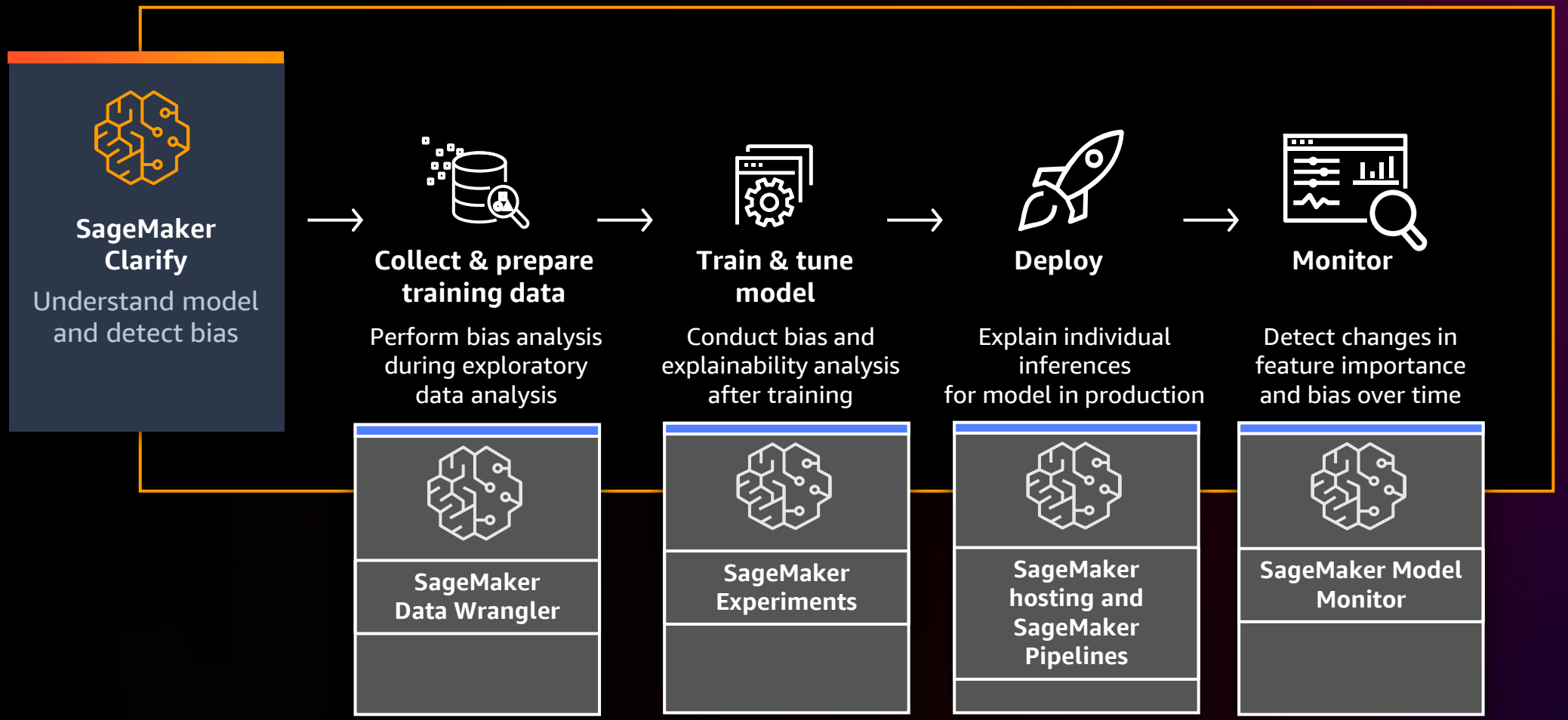
Provide model reports to your internal teams and auditors

Offer internal risk teams and external regulatory bodies the visibility they need

aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

SageMaker Clarify – How it works



© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Question – How do I drive the most immediate responsible AI value with SageMaker Clarify?

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Immediate value with SageMaker Clarify

- Prediction level explainability
  - Batch
  - Online
- Monitoring feature and bias drift
- NLP explainability

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Customer spotlight – Local prediction explainability

PROBLEM

Capcom is a Japanese game company famous for game titles such as the *Monster Hunter* series and *Street Fighter*. In order to keep users satisfied, CAPCOM needed to ensure game quality and identify possible churners and their causes.

SOLUTION

Created a churn prediction model using the access log data of their mobile game, which could better inform the customer service team in order to drive customer satisfaction and retention.

IMPACT

The customer churn model based on AutoGluon Tabular was able to predict customer churn one week out 94% of the time. **SageMaker Clarify provides global and local model explainability via local SHAP values while decreasing 50% of the computation cost compared to open source solutions.** By checking it, CAPCOM is understanding and improving user satisfaction on both global and customer-by-customer levels.

NEXT

Use online explainability feature to drive real-time prediction explainability

CAPCOM®

aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Feature attribution solutions

Examples –

- End customer facing example – Why am I seeing this ad?
- Attribute a lending model's prediction to its features
- Attribute an object recognition model's prediction to its pixels
- Attribute a text sentiment model's prediction to individual words

What if updated data and training causes changes over time?

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Explainability SageMaker Model Monitor & SageMaker Clarify

- Monitor drift in relative feature importance via model quality monitoring jobs
- SageMaker Model Monitor will automatically
  - Create a baseline and compute feature statistics
  - Inspect merged data and generate quality metrics and violations (use KernelShap to compute relative importance of features)
  - Push generated reports to Amazon Simple Storage Service (Amazon S3) for further analysis in a notebook
  - Emit Amazon CloudWatch metrics to set up alerts and triggers
- Cost – An average of $20 per month to monitor hosted production models
- Getting Started
  - Example Notebooks: https://github.com/aws/amazon-sagemaker-examples/blob/main/sagemaker_model_monitor/fairness_and_explainability/SageMaker-Model-Monitor-Fairness-and-Explainability.ipynb
  - Video: https://youtu.be/-tGFB6njOJM?t=780 (Minute 13:00)
  - Service team is here to help

aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Thank you!

Pinar Yilmaz
yilmazp@amazon.com

Chris Knoch
knocc@amazon.com



Please complete the session survey in the **mobile app**

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.