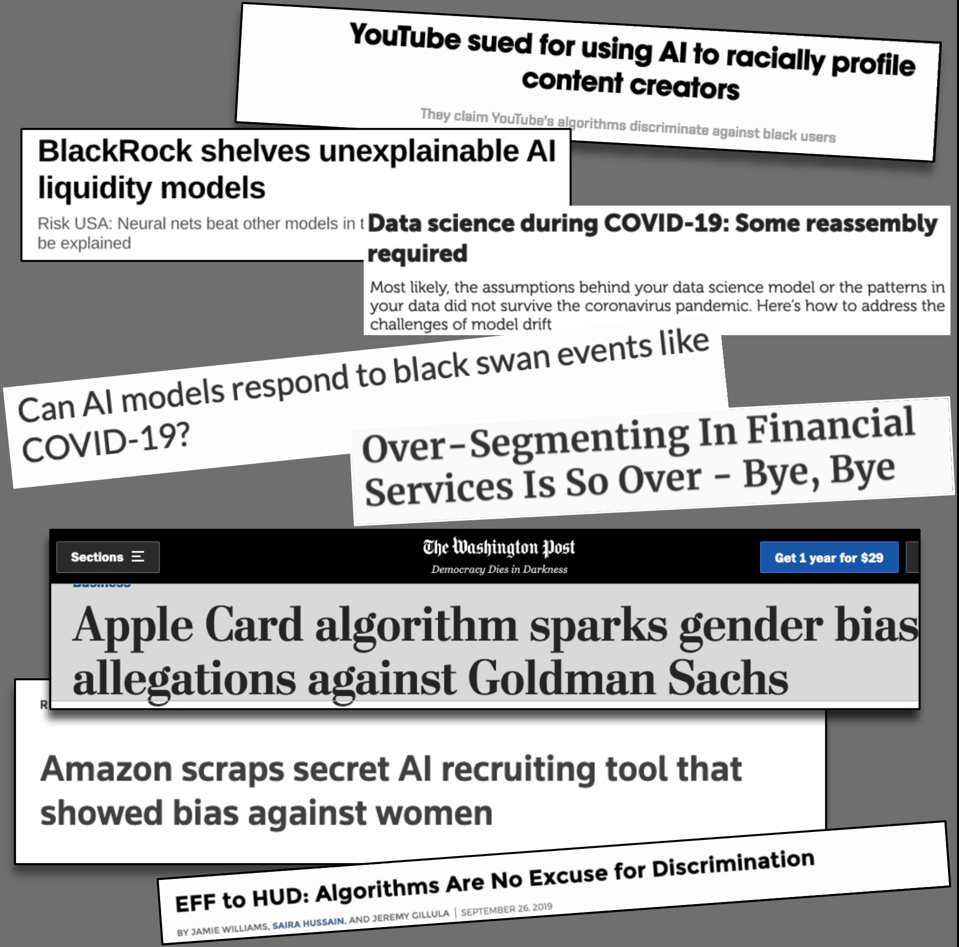
Trustworthy AI with IBM

Brian Snyder
Sr Data Science Technical Specialist
bsnyder@us.ibm.com

Rising concerns on trust in AI decisions



YouTube sued for using AI to racially profile content creators

They claim YouTube's algorithms discriminate against black users

BlackRock shelves unexplainable AI liquidity models

Risk USA: Neural nets beat other models in t
be explained

Data science during COVID-19: Some reassembly required

Most likely, the assumptions behind your data science model or the patterns in your data did not survive the coronavirus pandemic. Here's how to address the challenges of model drift

Can AI models respond to black swan events like COVID-19?

Over-Segmenting In Financial Services Is So Over - Bye, Bye

Sections

The Washington Post

Democracy Dies in Darkness

Get 1 year for $29

Apple Card algorithm sparks gender bias allegations against Goldman Sachs

R

Amazon scraps secret AI recruiting tool that showed bias against women

EFF to HUD: Algorithms Are No Excuse for Discrimination

BY JAMIE WILLIAMS, SAIRA HUSSAIN, AND JEREMY GILLULA | SEPTEMBER 26, 2019

…and regulators are catching up

USA

SR 11–7 requires model risk management for all models in financial services

2019—Proposal for Algorithmic Accountability Act

2021 – US Govt. National AI Initiative Act

Canada

2017—National AI Strategy launched. Impact Analysis

2020—All public agencies must do an impact analysis for AI models

European Union

2021 – Draft regulation for trust in AI development

2019—Guidelines for AI development

Mexico

2018—General principles for AI development in the government

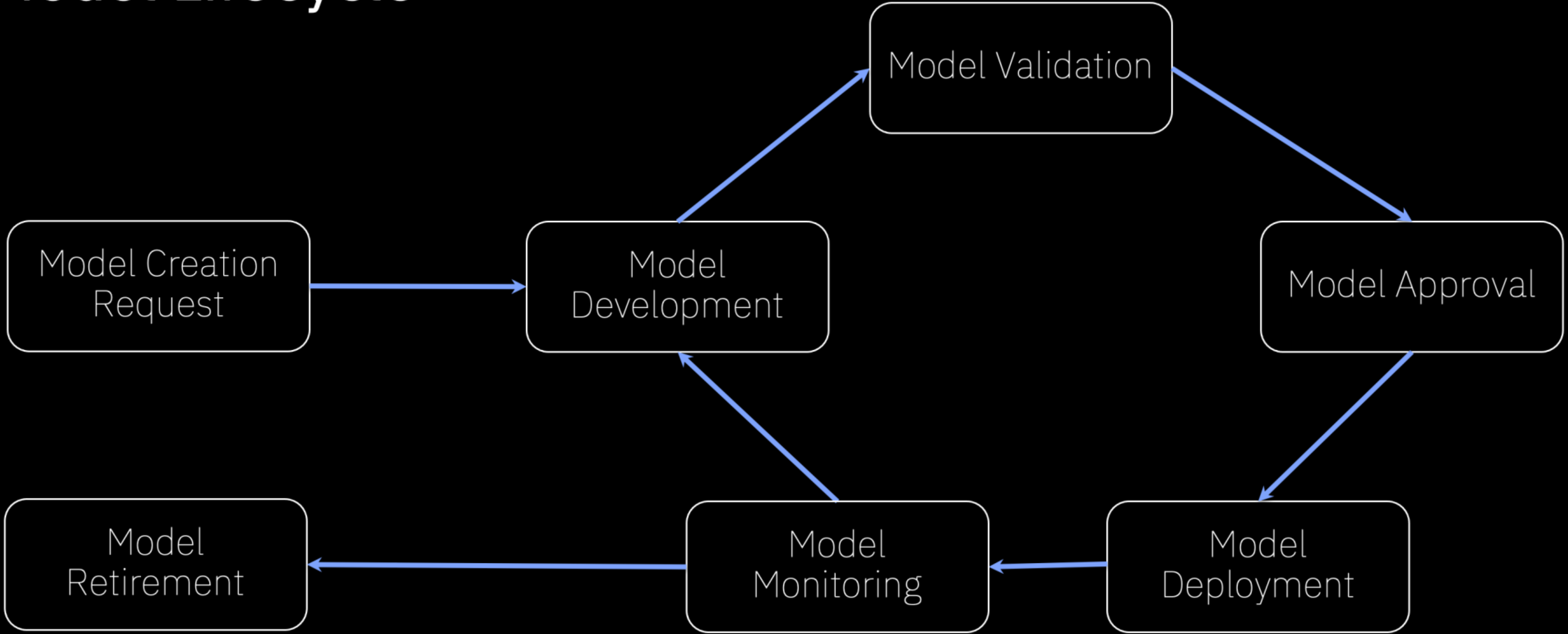
Partnerships on AI

Partnership between tech companies to study best practices and impact of AI

AI Now Institute

NYU research center focused on social implications of AI

Model Lifecycle



Model Creation Request

Model Development

Model Validation

Model Approval

Model Deployment

Model Monitoring

Model Retirement

Aspects of Trustworthy AI

Fairness

Impartial and addressing bias

*Are **privileged groups** at a systematic advantage compared to other groups?*

Robustness

Models need to perform well across the lifecycle, handle exceptions effectively, enable confidence in systems outcomes

*Are relevant performance **metrics** monitored over time?*

Drift

Changes in input data cause model to make inaccurate decisions.

Do anomalies exist between training data and data ranges or combinations seen in real life?

Explainability

Easy to understand outcomes/decisions

Why did the AI arrive at an outcome? At what point would the outcome have been different?

Transparency

Open to inspecting facts and details

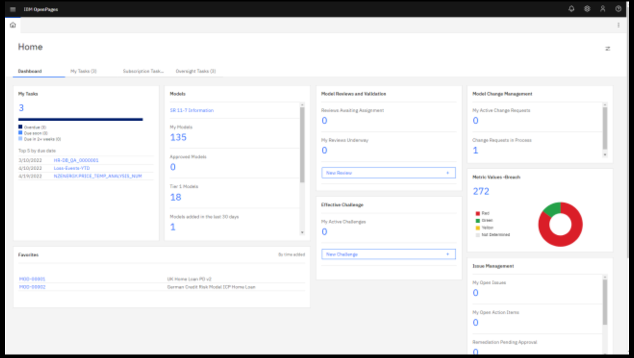
Can we increase understanding of why and how AI was created?

Common AI Use Cases

- Lending: Loan Origination, Loan Default
- Collections
- Claims Processing
- Underwriting

- Targeted Marketing Campaigns
- Segmentation
- Customer Management
- HR

Trustworthy AI Demo



My Tasks
3
Overdue (3)
Due soon (0)
Due in 2+ weeks (0)
Top 5 by due date
3/10/2022 HR-DB_QA_0000001
4/10/2022 Loss-Events-YTD
4/19/2022 NZENERGY.PRICE_TEMP_ANALYSIS_NUM

Models
SR 11-7 Information
My Models
135
Approved Models
0
Tier 1 Models
18
Models added in the last 30 days
1

Model Reviews and Validation
Reviews Awaiting Assignment
0
My Reviews Underway
0

Effective Challenge
My Active Challenges
0

Model Change Management
My Active Change Requests
0
Change Requests in Process
1

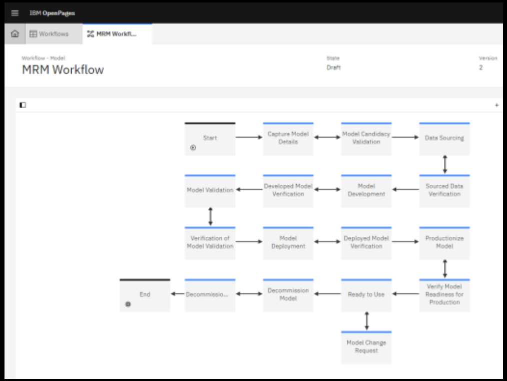
Metric Values -Breach
272
Red
Green
Yellow
Not Determined

Issue Management
My Open Issues
0
My Open Action Items
0
Remediation Pending Approval
0

Favorites
By time added

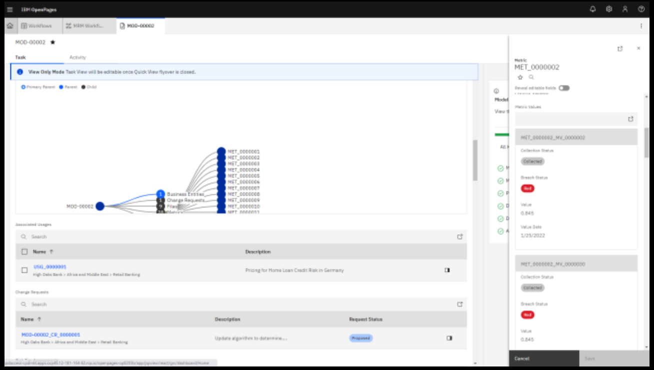
MOD-00001	UK Home Loan PD v2
MOD-00002	German Credit Risk Model ICP Home Loan

Enterprise Inventory Dashboard



Workflow - Model
MRM Workflow
State
Draft
Version
2

Enterprise Workflow



MOD-00002
Task
Activity
View Only Mode Task View will be editable once Quick View flyover is closed.

Associated Usages

Name	Description
USG_0000001 High Oaks Bank > Africa and Middle East > Retail Banking	Pricing for Home Loan Credit Risk in Germany

Change Requests

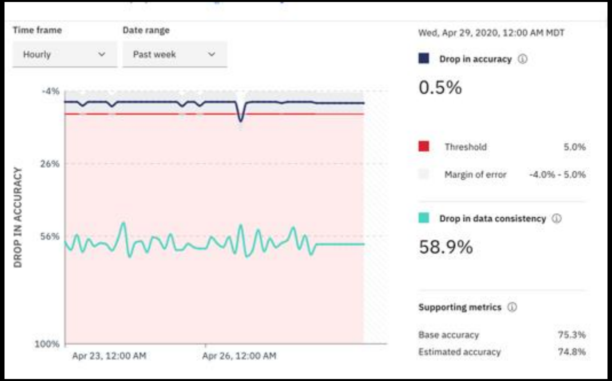
Name	Description	Request Status
MOD-00002_CR_0000001 High Oaks Bank > Africa and Middle East > Retail Banking	Update algorithm to determine...	Proposed

Metric
MET_0000002
Reveal editable fields
Metric Values

MET_0000002_MV_0000002
Collection Status
Collected
Breach Status
Red
Value
0.845
Value Date
1/25/2022

MET_0000002_MV_0000030
Collection Status
Collected
Breach Status
Red
Value
0.845

Variety of Risk Metrics Captured For Reporting



Time frame
Hourly
Date range
Past week

Wed, Apr 29, 2020, 12:00 AM MDT
Drop in accuracy
0.5%

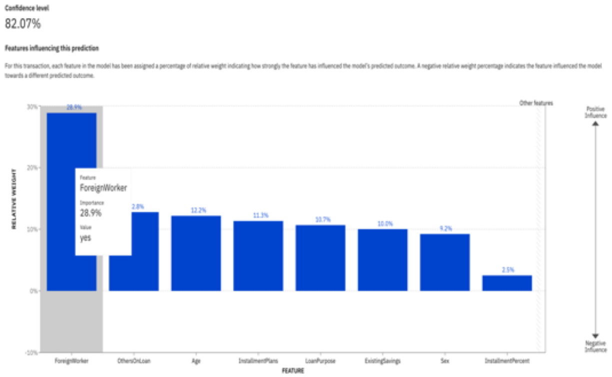
Threshold	5.0%
Margin of error	-4.0% - 5.0%

Drop in data consistency
58.9%

Supporting metrics

Base accuracy	75.3%
Estimated accuracy	74.8%

Drift in data consistency, Drift in accuracy, Bias Detection



Confidence level
82.07%

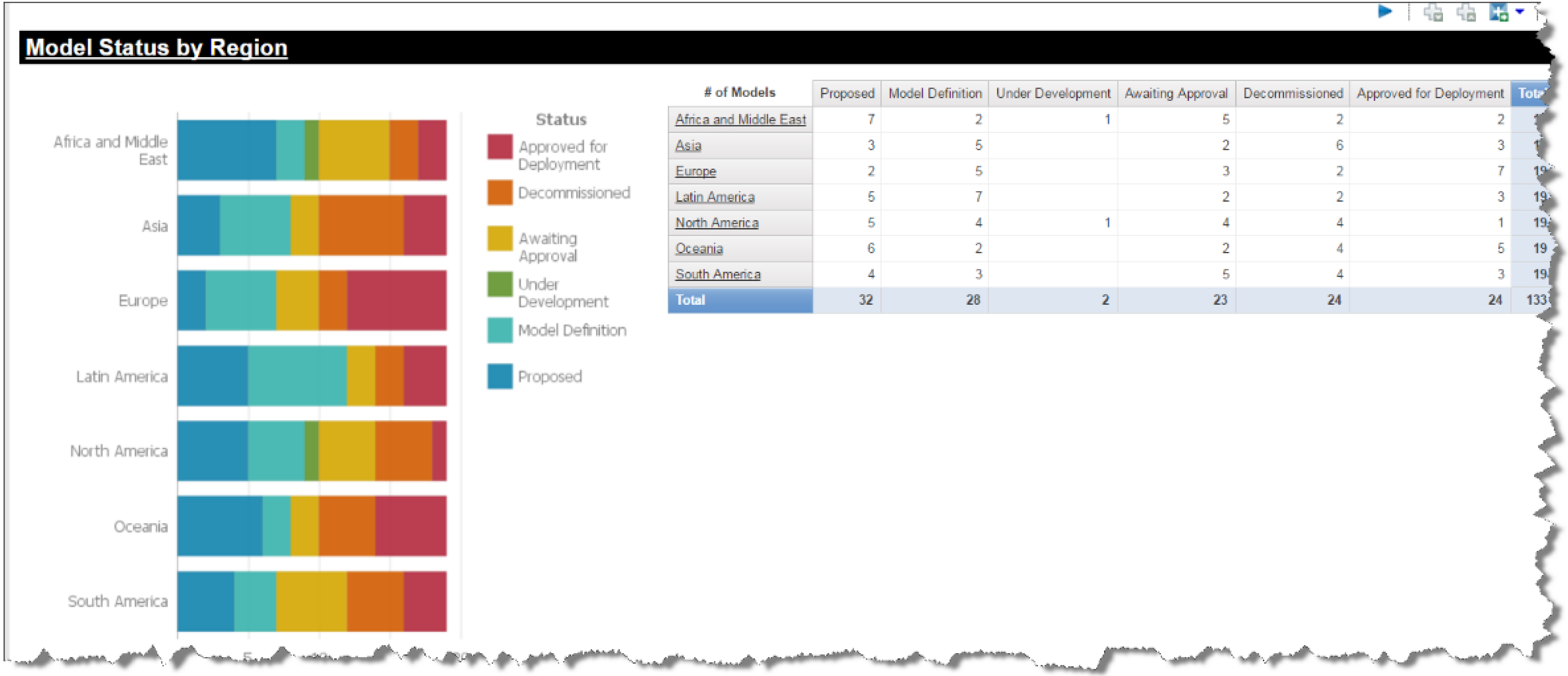
Features influencing this prediction

For this transaction, each feature in the model has been assigned a percentage of relative weight indicating how strongly the feature has influenced the model's predicted outcome. A negative relative weight percentage indicates the feature influenced the model towards a different predicted outcome.

Feature
ForeignWorker
Importance
28.9%
Value
yes

Local and contrastive explanations

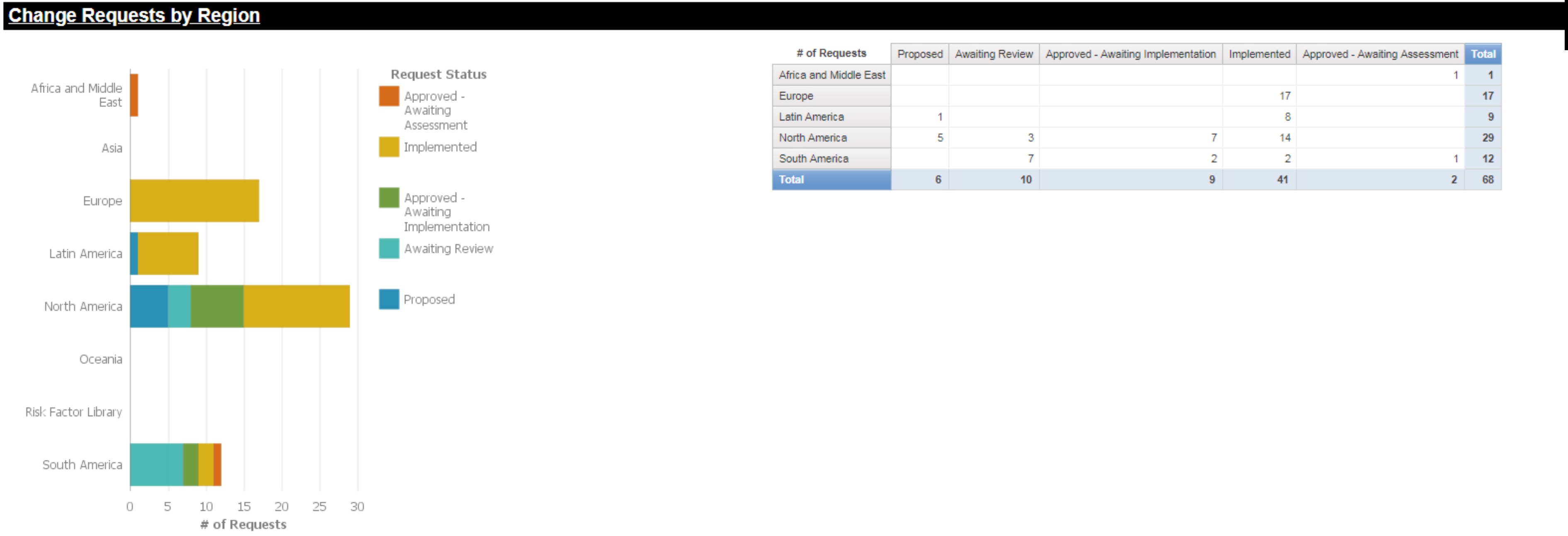
Model Status by Region



Model Status by Region

# of Models	Proposed	Model Definition	Under Development	Awaiting Approval	Decommissioned	Approved for Deployment	Tot
Africa and Middle East	7	2	1	5	2	2	1
Asia	3	5		2	6	3	1
Europe	2	5		3	2	7	19
Latin America	5	7		2	2	3	19
North America	5	4	1	4	4	1	19
Oceania	6	2		2	4	5	19
South America	4	3		5	4	3	19
Total	32	28	2	23	24	24	133

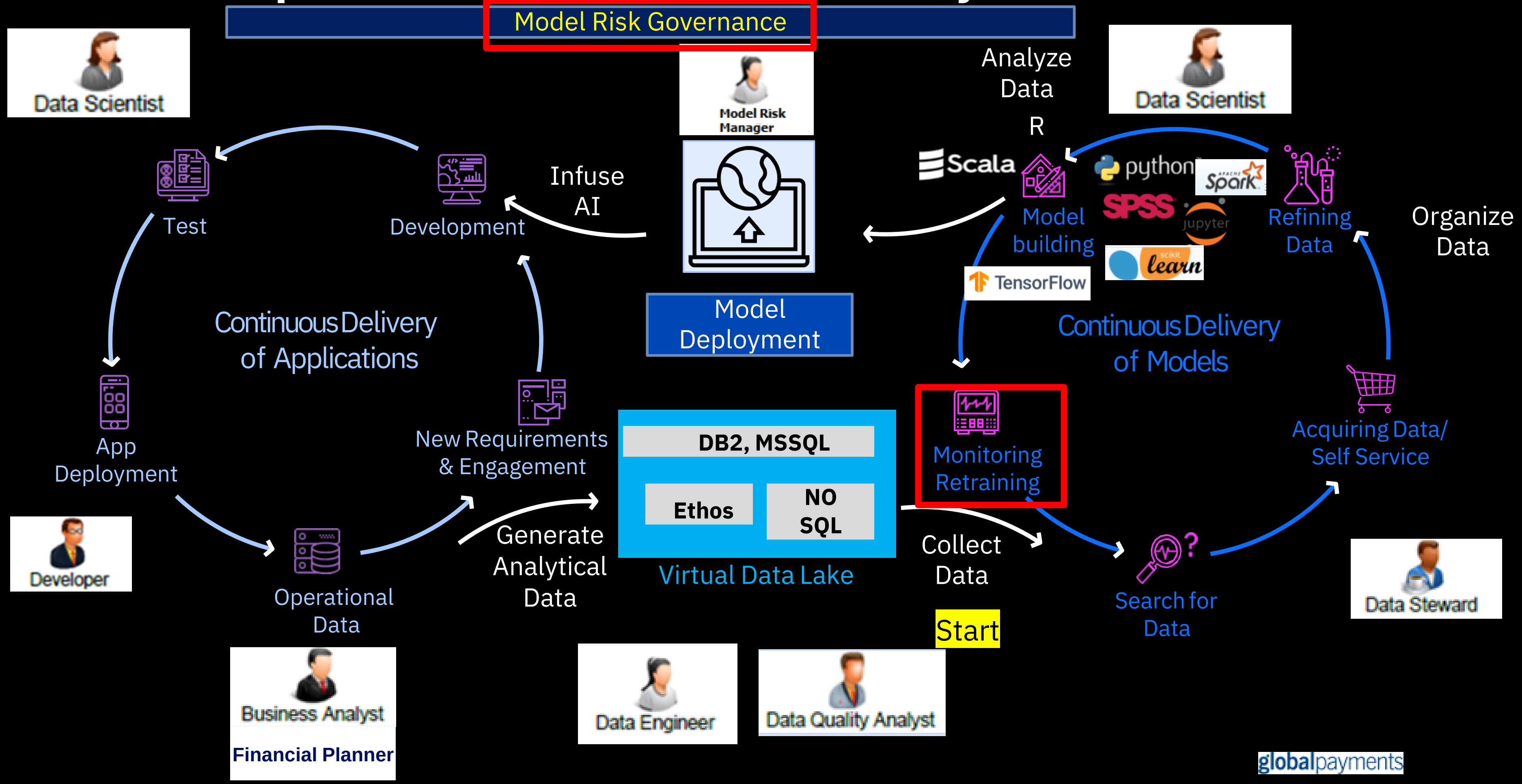
Change Request by Region



Change Requests by Region

# of Requests	Proposed	Awaiting Review	Approved - Awaiting Implementation	Implemented	Approved - Awaiting Assessment	Total
Africa and Middle East					1	**1**
Europe				17		**17**
Latin America	1			8		**9**
North America	5	3	7	14		**29**
South America		7	2	2	1	**12**
Total	**6**	**10**	**9**	**41**	**2**	**68**

ModelOps Data Science and Trustworthy AI as a Team - DEMO



US-based Multinational Bank

Upon movement of all proof-of-concept projects into production, the **bank will have the ability to govern all AI projects** using their existing technology and skill investment so that existing business units do not need to change their current systems and reskill their employees.

*IBM Cloud Pak for Data provides organizations a transversal and centralized view of the AI lifecycle through **an integrated platform that covers the three key functions of model build, model deploy, and model management.**

Business problem

This bank uses many tools and systems for model development and deployment, making model governance challenging. They were unable to **ensure models comply with enterprise policies, identify inefficiencies, provide standardized regulatory reporting, learn and scale best practices**

Solution

IBM used its leading enterprise insights platform (Cloud Pak for Data) with its open architecture enabling a **smooth integration of models developed and deployed on other platforms** from a governance perspective

29 October 2021

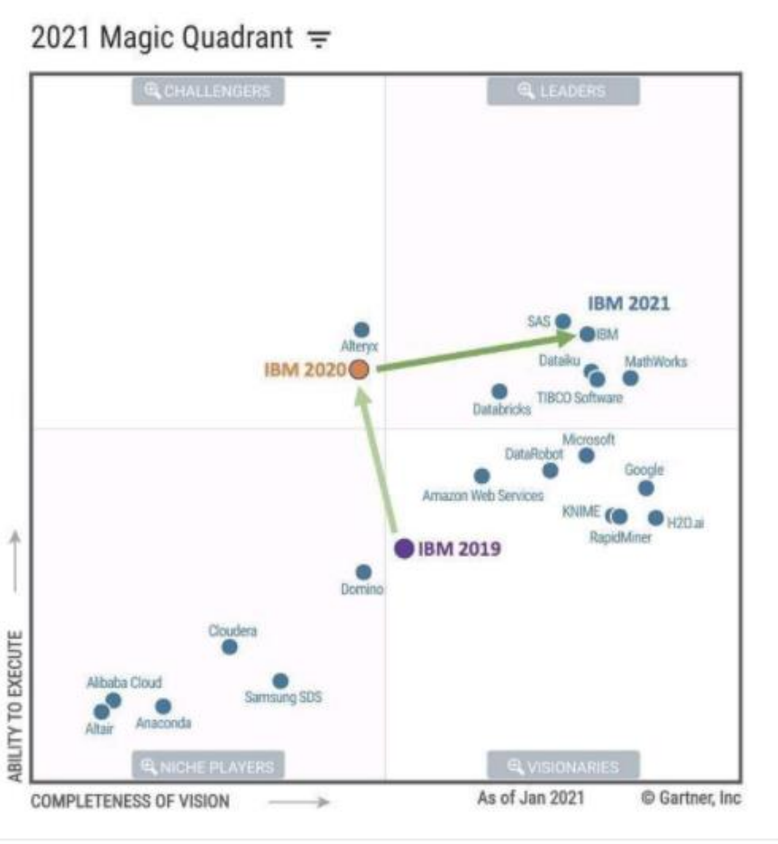
Trustworthy AI with IBM

Brian Snyder
Sr Data Science Technical Specialist
bsnyder@us.ibm.com

THANK YOU!

Demo Snapshots

Gartner Magic Quadrant for Data Science and Machine Learning Platforms

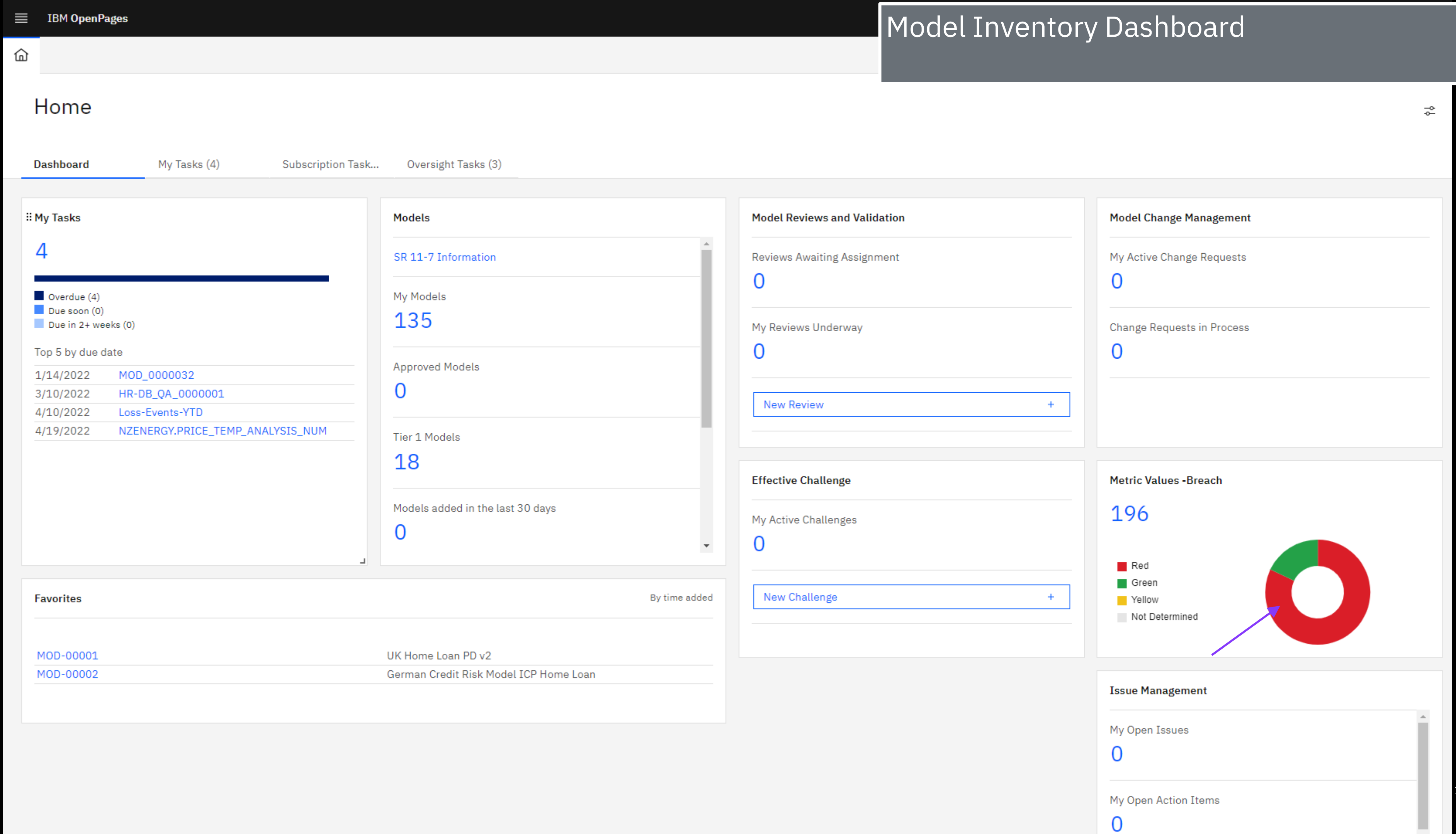


Supports multiple tasks across the data science life cycle, including:

- Problem and business context understanding
- Data ingestion
- Data preparation
- Data exploration
- Feature engineering
- Model creation and training
- Model testing
- Deployment
- Monitoring
- Maintenance
- Data and model governance
- Explainable artificial intelligence (XAI)
- Business value tracking
- Collaboration

https://www.gartner.com/doc/reprints?id=1-25DOZD29&ct=210304&st=sb

Model Inventory Dashboard



IBM OpenPages

Home

Dashboard | My Tasks (4) | Subscription Task... | Oversight Tasks (3)

My Tasks

4

- Overdue (4)
- Due soon (0)
- Due in 2+ weeks (0)

Top 5 by due date

1/14/2022	MOD_0000032
3/10/2022	HR-DB_QA_0000001
4/10/2022	Loss-Events-YTD
4/19/2022	NZENERGY.PRICE_TEMP_ANALYSIS_NUM

Models

SR 11-7 Information

My Models
135

Approved Models
0

Tier 1 Models
18

Models added in the last 30 days
0

Model Reviews and Validation

Reviews Awaiting Assignment
0

My Reviews Underway
0

New Review +

Model Change Management

My Active Change Requests
0

Change Requests in Process
0

Favorites

By time added

MOD-00001	UK Home Loan PD v2
MOD-00002	German Credit Risk Model ICP Home Loan

Effective Challenge

My Active Challenges
0

New Challenge +

Metric Values -Breach

196

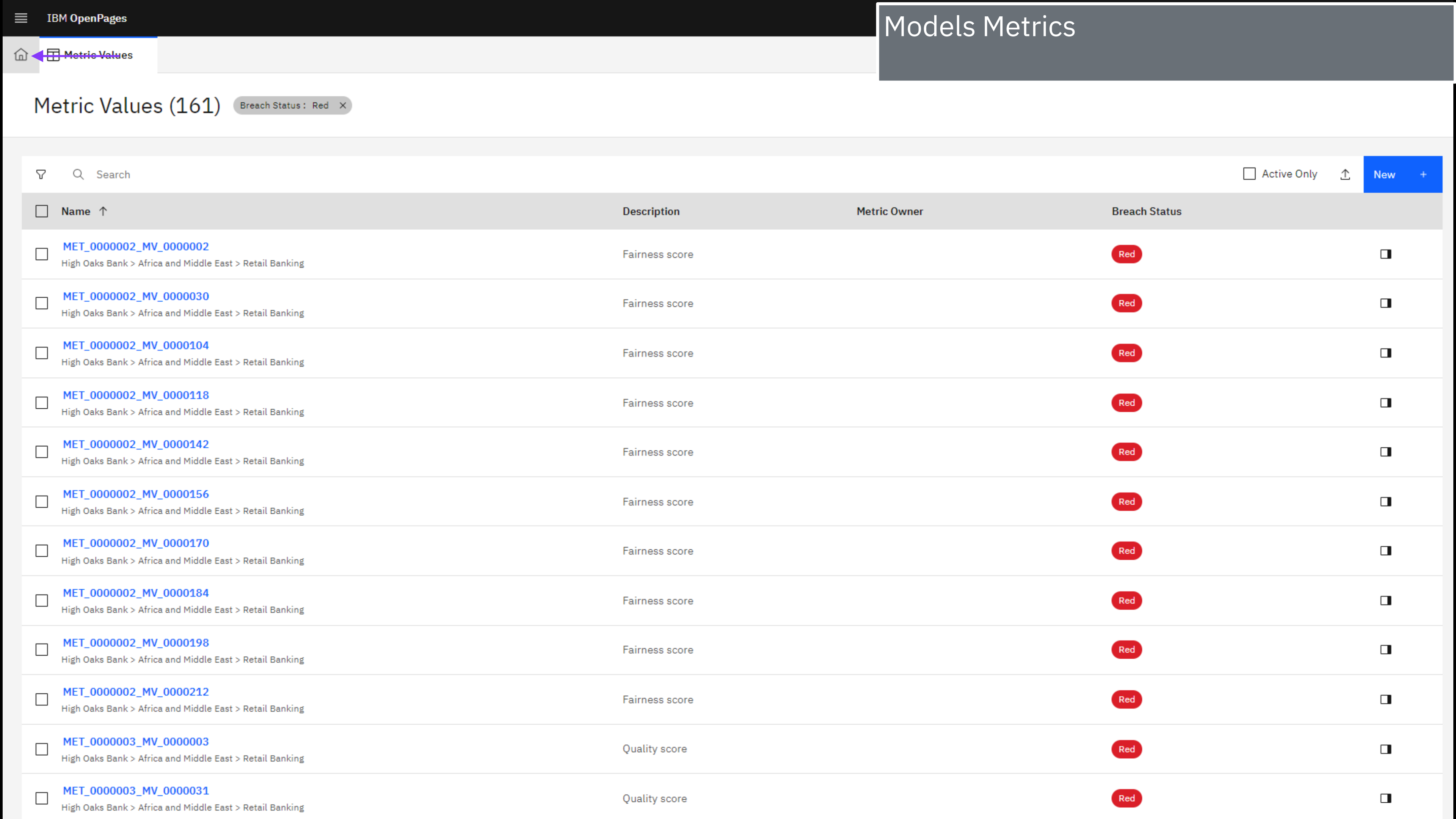
- 🟥 Red
- 🟩 Green
- 🟨 Yellow
- ⬜ Not Determined

Issue Management

My Open Issues
0

My Open Action Items
0

Models Metrics

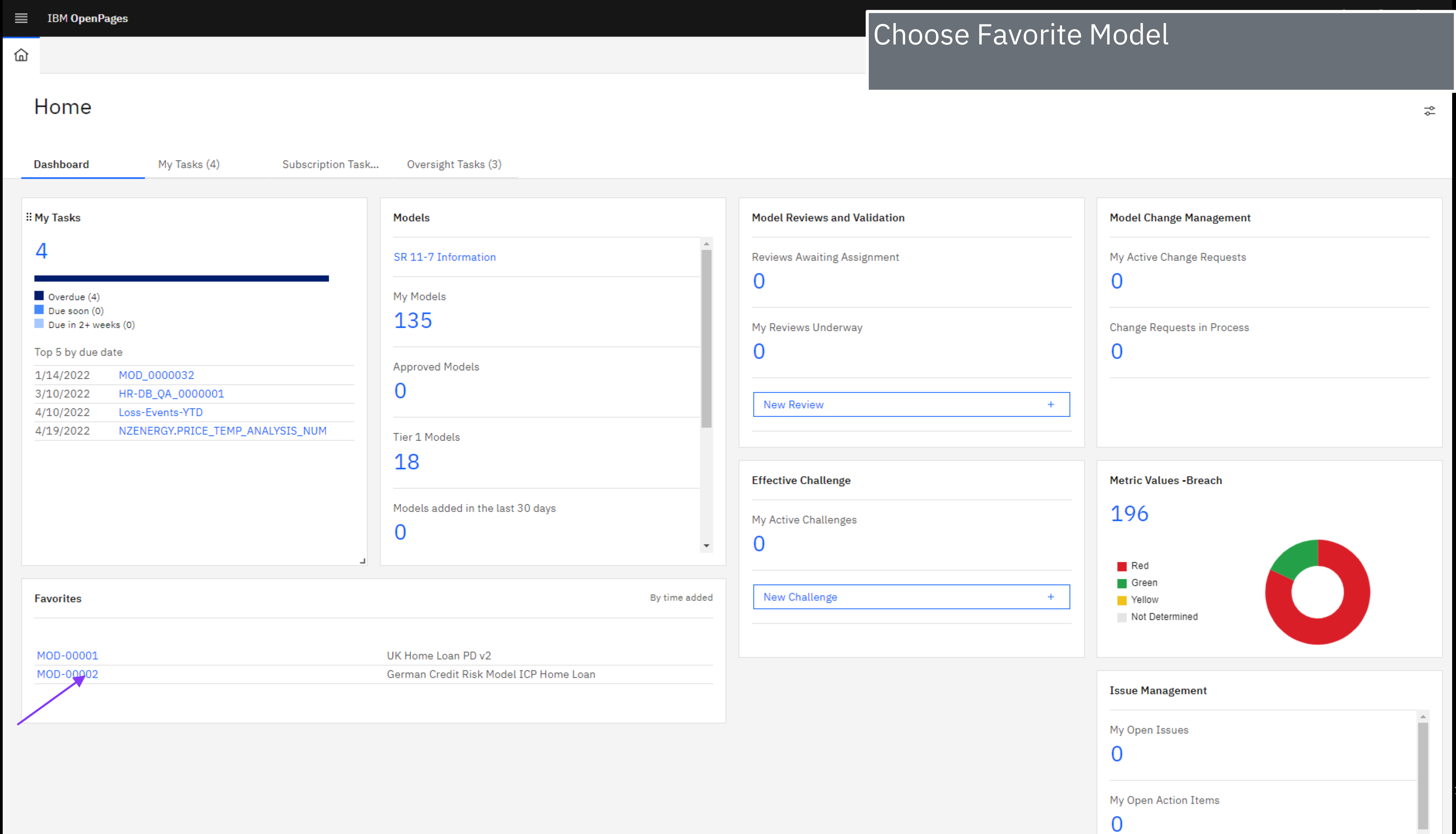


Metric Values (161)

Breach Status : Red ✕

□ Name ↑	Description	Metric Owner	Breach Status
□ MET_0000002_MV_0000002 High Oaks Bank > Africa and Middle East > Retail Banking	Fairness score		Red
□ MET_0000002_MV_0000030 High Oaks Bank > Africa and Middle East > Retail Banking	Fairness score		Red
□ MET_0000002_MV_0000104 High Oaks Bank > Africa and Middle East > Retail Banking	Fairness score		Red
□ MET_0000002_MV_0000118 High Oaks Bank > Africa and Middle East > Retail Banking	Fairness score		Red
□ MET_0000002_MV_0000142 High Oaks Bank > Africa and Middle East > Retail Banking	Fairness score		Red
□ MET_0000002_MV_0000156 High Oaks Bank > Africa and Middle East > Retail Banking	Fairness score		Red
□ MET_0000002_MV_0000170 High Oaks Bank > Africa and Middle East > Retail Banking	Fairness score		Red
□ MET_0000002_MV_0000184 High Oaks Bank > Africa and Middle East > Retail Banking	Fairness score		Red
□ MET_0000002_MV_0000198 High Oaks Bank > Africa and Middle East > Retail Banking	Fairness score		Red
□ MET_0000002_MV_0000212 High Oaks Bank > Africa and Middle East > Retail Banking	Fairness score		Red
□ MET_0000003_MV_0000003 High Oaks Bank > Africa and Middle East > Retail Banking	Quality score		Red
□ MET_0000003_MV_0000031 High Oaks Bank > Africa and Middle East > Retail Banking	Quality score		Red

Choose Favorite Model



IBM OpenPages

Home

Dashboard | My Tasks (4) | Subscription Task... | Oversight Tasks (3)

My Tasks

4

- 🟦 Overdue (4)
- 🟦 Due soon (0)
- 🟦 Due in 2+ weeks (0)

Top 5 by due date

1/14/2022	MOD_0000032
3/10/2022	HR-DB_QA_0000001
4/10/2022	Loss-Events-YTD
4/19/2022	NZENERGY.PRICE_TEMP_ANALYSIS_NUM

Models

SR 11-7 Information

My Models
135

Approved Models
0

Tier 1 Models
18

Models added in the last 30 days
0

Model Reviews and Validation

Reviews Awaiting Assignment
0

My Reviews Underway
0

New Review +

Model Change Management

My Active Change Requests
0

Change Requests in Process
0

Effective Challenge

My Active Challenges
0

New Challenge +

Metric Values -Breach

196

- 🟥 Red
- 🟩 Green
- 🟨 Yellow
- ⬜ Not Determined

Favorites

By time added

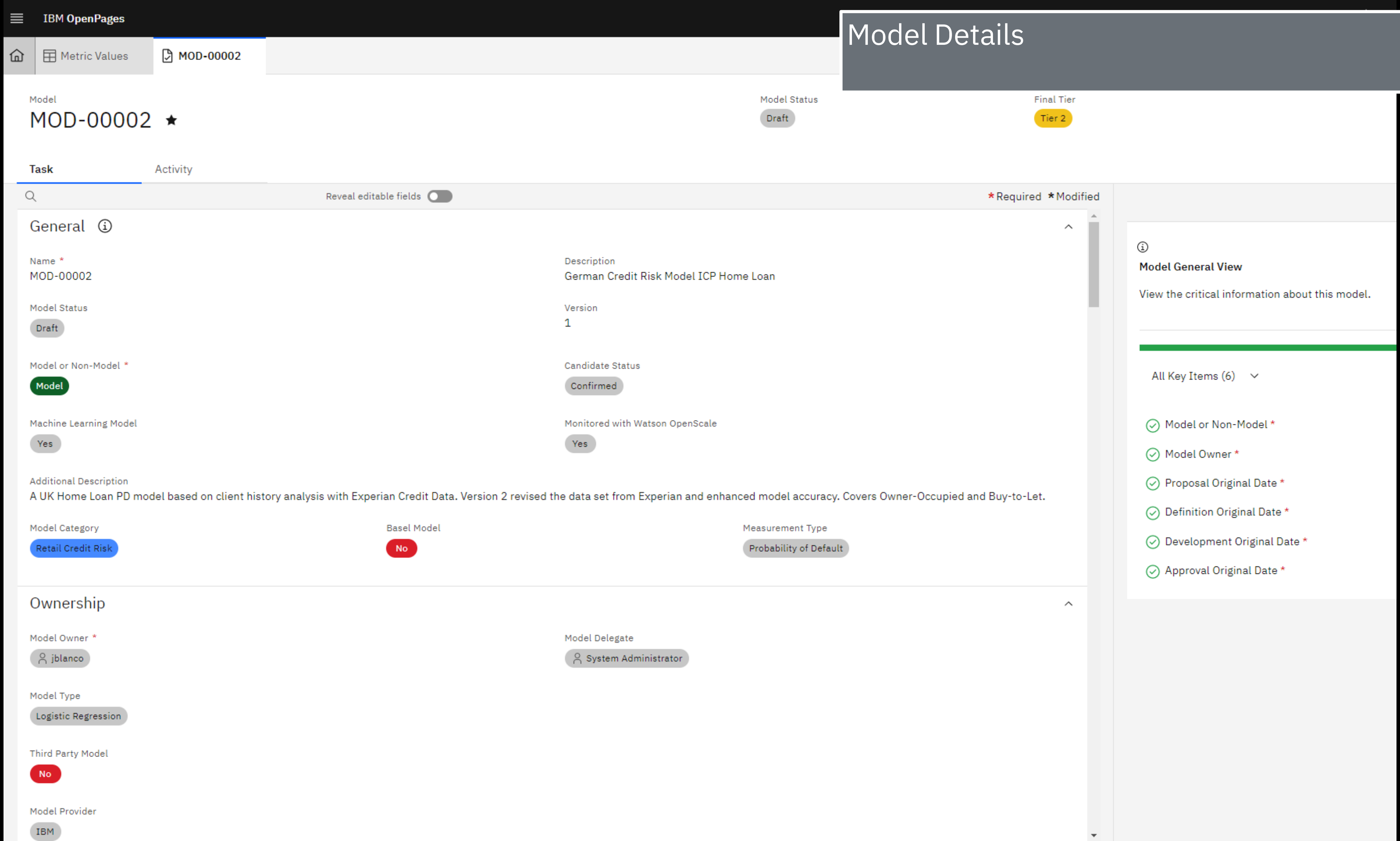
MOD-00001	UK Home Loan PD v2
MOD-00002	German Credit Risk Model ICP Home Loan

Issue Management

My Open Issues
0

My Open Action Items
0

Model Details



Model
MOD-00002 ★

Model Status
Draft

Final Tier
Tier 2

* Required * Modified

General

Name *
MOD-00002

Description
German Credit Risk Model ICP Home Loan

Model Status
Draft

Version
1

Model or Non-Model *
Model

Candidate Status
Confirmed

Machine Learning Model
Yes

Monitored with Watson OpenScale
Yes

Additional Description
A UK Home Loan PD model based on client history analysis with Experian Credit Data. Version 2 revised the data set from Experian and enhanced model accuracy. Covers Owner-Occupied and Buy-to-Let.

Model Category
Retail Credit Risk

Basel Model
No

Measurement Type
Probability of Default

Ownership

Model Owner *
jblanco

Model Delegate
System Administrator

Model Type
Logistic Regression

Third Party Model
No

Model Provider
IBM

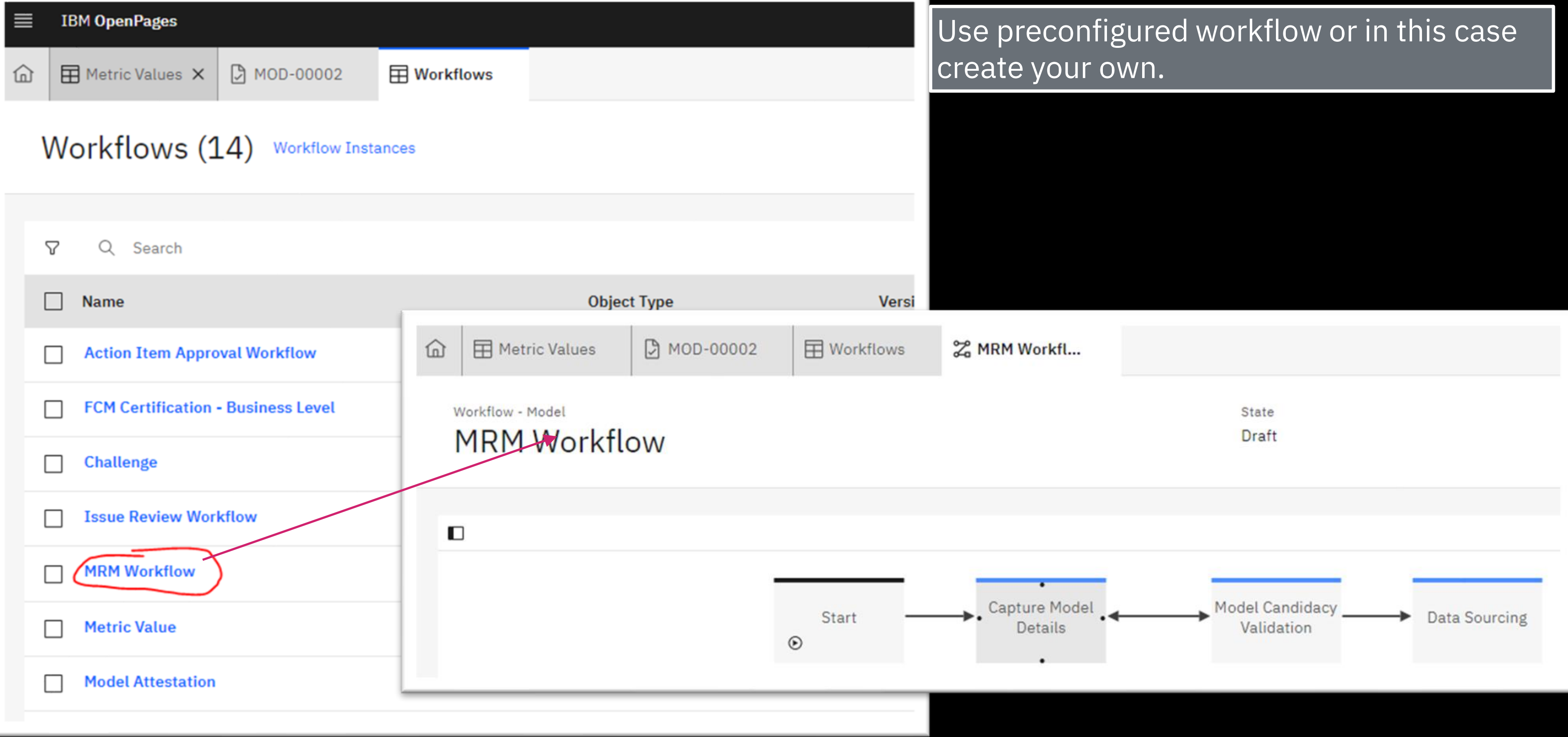
Model General View

View the critical information about this model.

All Key Items (6)

- ✓ Model or Non-Model *
- ✓ Model Owner *
- ✓ Proposal Original Date *
- ✓ Definition Original Date *
- ✓ Development Original Date *
- ✓ Approval Original Date *

Use preconfigured workflow or in this case create your own.



IBM OpenPages

Metric Values ✕ | MOD-00002 | Workflows

Workflows (14)

Workflow Instances

Search

□	Name	Object Type	Versi
□	Action Item Approval Workflow		
□	FCM Certification - Business Level		
□	Challenge		
□	Issue Review Workflow		
□	MRM Workflow		
□	Metric Value		
□	Model Attestation		

Metric Values | MOD-00002 | Workflows | MRM Workfl...

Workflow - Model

MRM Workflow

State

Draft

Fairness and Artificial Intelligence in Banking

Use of AI/ML creates many risks – bias, lack of explainability, operational, … – with implications for both compliance and risk management

Regulators issued guidance regarding these risks, but the regulatory landscape continues to evolve as new evidence of consumer impact emerges

Fairness and bias

Complexity

Explainability

Alternative Data

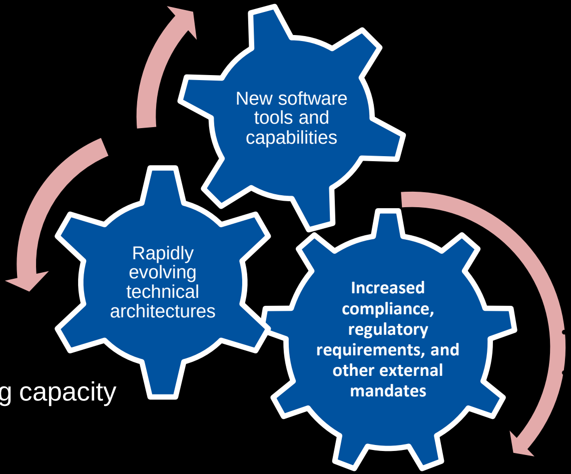
Very specific challenges relate to the nature of a data science model.

> The technology and regulations, and how they are applied, are constantly evolving.

Traditional (SAS, Matlab, IBM, etc.)

ML-centric (IBM Watson, Datarobot, RapidMiner, H2O, etc.)

New entrants (Algorithmia, ModelOp, Modzy, etc.)


New software tools and capabilities
Rapidly evolving technical architectures
Increased compliance, regulatory requirements, and other external mandates

Cloud

Unlimited computing capacity

Containerization

Security

DataOps, DevOps, MLOps

FRB SR 11-7

OSFI E-23

ECB TRIM Guideline

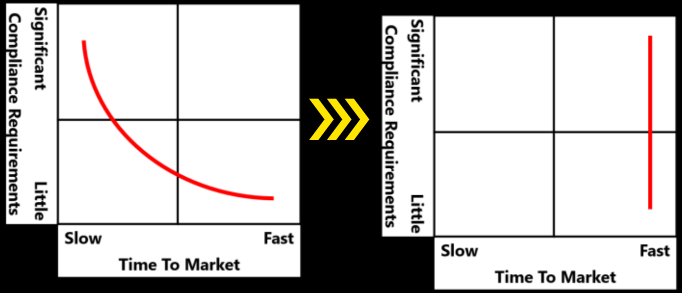
World Bank Credit Scoring Guidelines

OECD AI Principles

IOSCO Consultation report on AI

CA Consumer Privacy Act

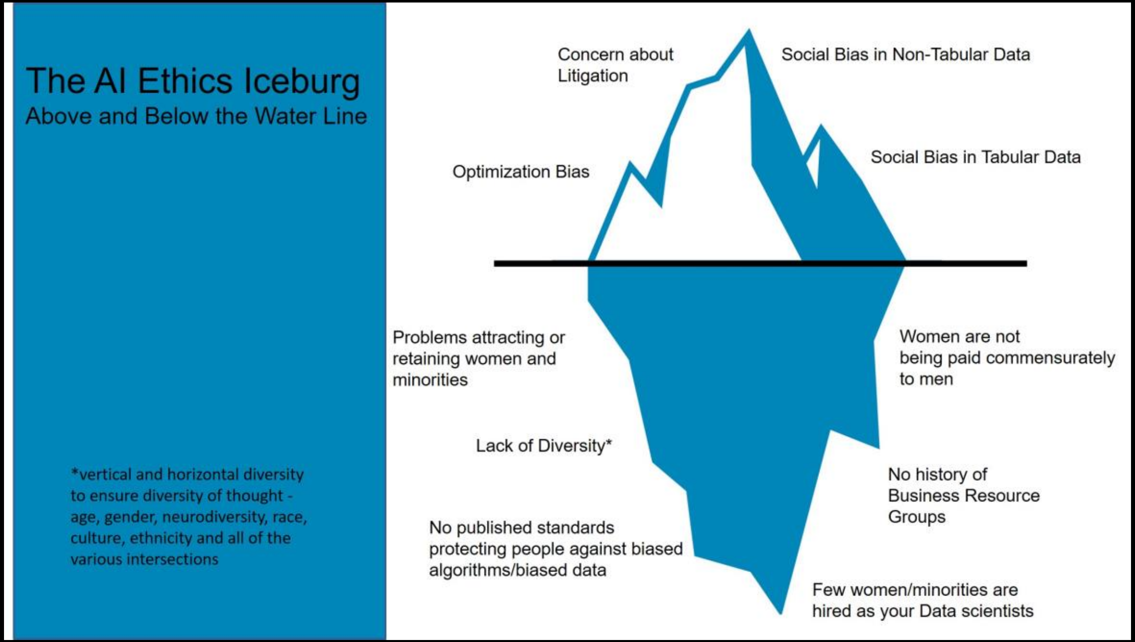
How to ensure the necessary *velocity* and *quality* in a changing, multi-system and multi-source environment?



Operationalizing AI requires organizations to demonstrate responsible, explainable use of AI

AI

Driven by data ➔ validity not guaranteed
Models ➔ not code, often unexplainable
Probabilistic ➔ non-deterministic, uncertain
Not just focus on version control, not an assembly line ➔ Model development MUST be part of lifecycle management


The AI Ethics Iceburg
Above and Below the Water Line

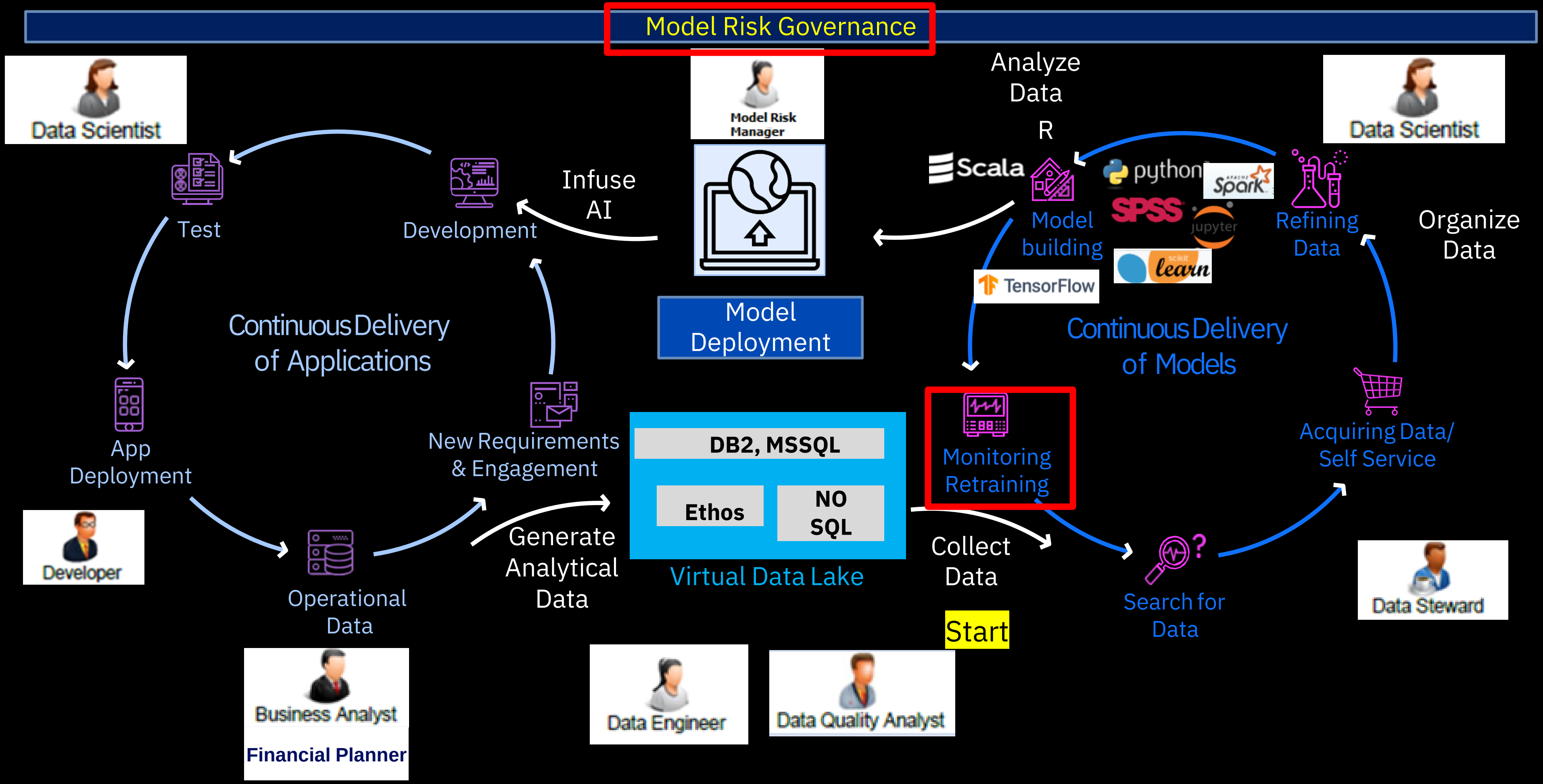
Concern about Litigation
Social Bias in Non-Tabular Data
Optimization Bias
Social Bias in Tabular Data
Problems attracting or retaining women and minorities
Women are not being paid commensurately to men
Lack of Diversity*
No history of Business Resource Groups
No published standards protecting people against biased algorithms/biased data
Few women/minorities are hired as your Data scientists

*vertical and horizontal diversity to ensure diversity of thought - age, gender, neurodiversity, race, culture, ethnicity and all of the various intersections

Model drift, bias and risk can pose significant liabilities and damage

AI fairness is a corporate social responsibility

ModelOps Data Science and Trustworthy AI as a Team



Model Risk Governance

Model Risk Manager

Model Deployment

Infuse AI

Development

Test

App Deployment

Operational Data

New Requirements & Engagement

Generate Analytical Data

Continuous Delivery of Applications

Data Scientist

Developer

Business Analyst

Financial Planner

Virtual Data Lake

DB2, MSSQL

Ethos

NO SQL

Data Engineer

Data Quality Analyst

Collect Data

Start

Monitoring Retraining

Search for Data

Acquiring Data/ Self Service

Data Steward

Refining Data

Organize Data

Data Scientist

Analyze Data

R

Model building

Continuous Delivery of Models