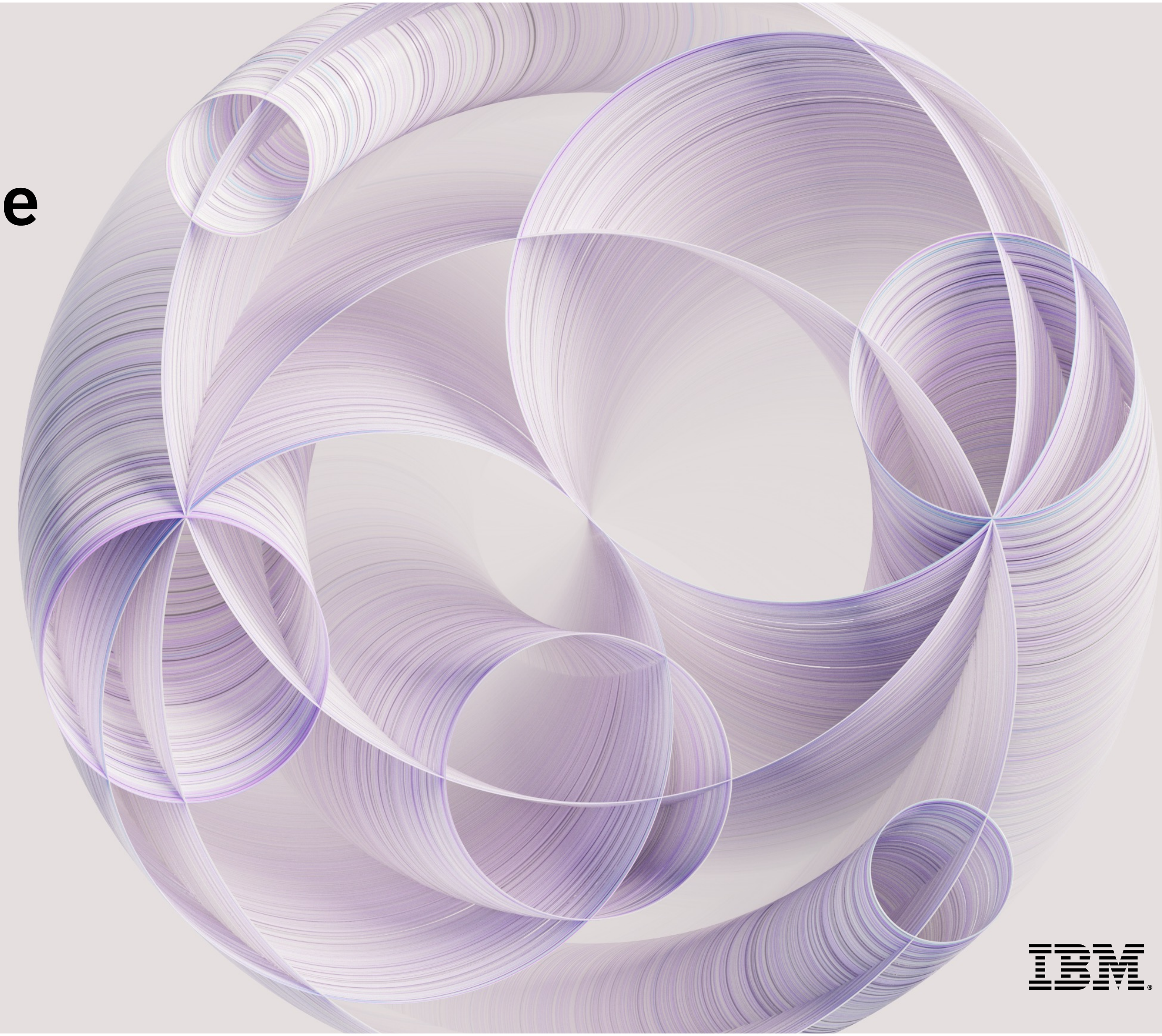
Introducing watsonx.governance

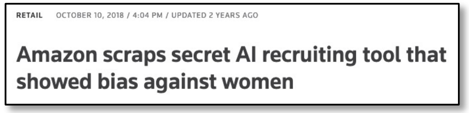


The promise of AI is clear, but implementations come with questions

How do we operationalize AI with confidence?

“**Protected/sensitive data like ethnicity and gender** should not be used by AI models for hiring. How can we ensure fairness ?”

How do I enable responsible use of AI to manage risk?



RETAIL OCTOBER 10, 2018 / 4:04 PM / UPDATED 2 YEARS AGO

Amazon scraps secret AI recruiting tool that showed bias against women

How do I protect and scale against regulations?

European Commission

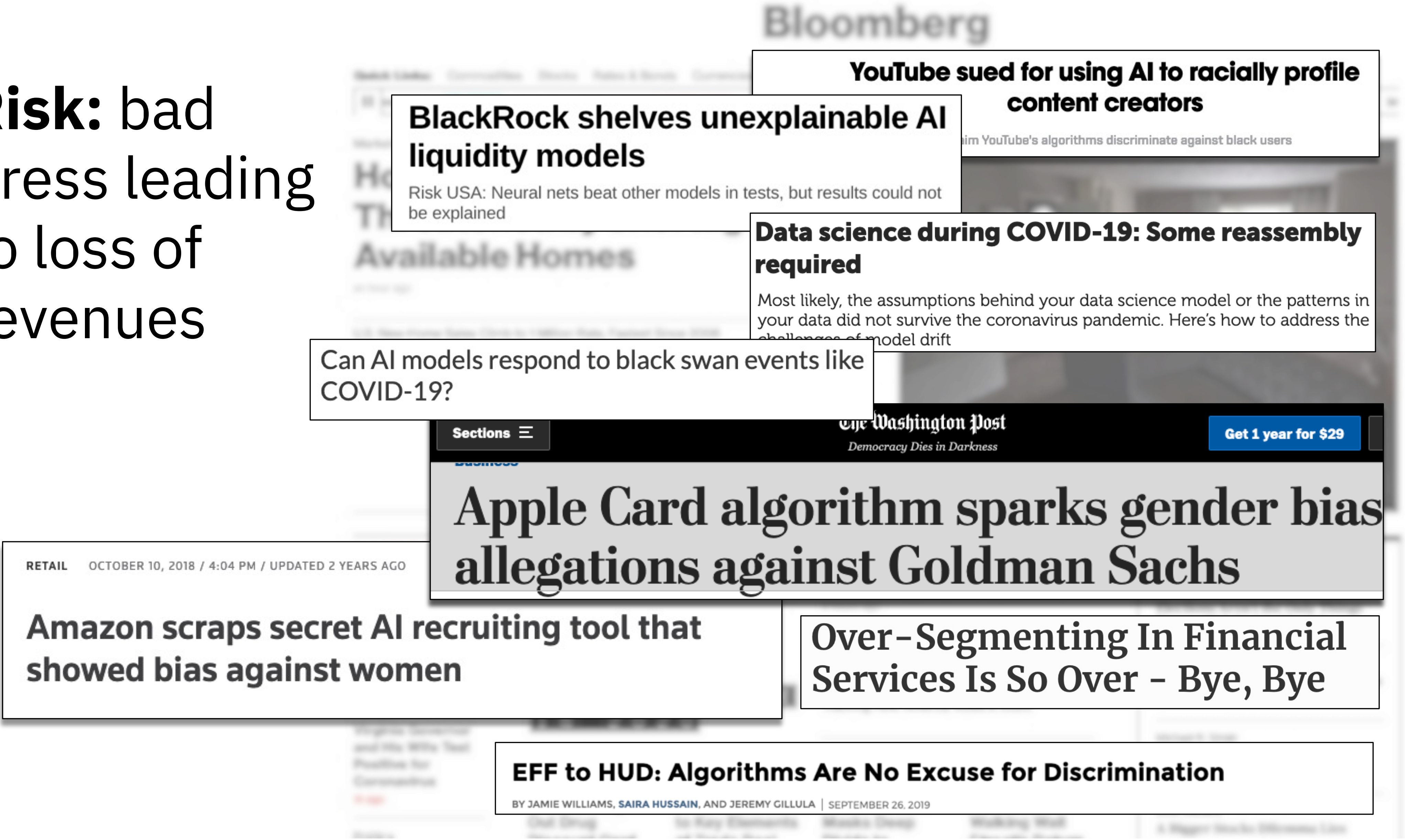
IDC predicts AI lifecycle opportunity is 6.7B in 2022 with AI Governance opportunity growing at 55%

Reputation: poor governance can damage consumer trust

Organizations need to:

- Only use personal information when it is needed and with the user's consent
- Ensure results are free from bias (ex. sex, age, etc) and guided by ethical and transparent principles
- Build consumer confidence, providing explanations for business decisions (ex. credit, membership denials)
- Proactively detect fraud and risk to consumer’s accounts

Risk: bad press leading to loss of revenues



YouTube sued for using AI to racially profile content creators

im YouTube's algorithms discriminate against black users

BlackRock shelves unexplainable AI liquidity models

Risk USA: Neural nets beat other models in tests, but results could not 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?

Apple Card algorithm sparks gender bias allegations against Goldman Sachs

RETAIL OCTOBER 10, 2018 / 4:04 PM / UPDATED 2 YEARS AGO

Amazon scraps secret AI recruiting tool that showed bias against women

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

EFF to HUD: Algorithms Are No Excuse for Discrimination

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

What does it take to trust a generative AI platform?

How was it trained?

- Garbage in -> garbage out
- An enterprise cannot use a foundation model trained with a Wikipedia crawl
- The training material needs to be huge and comprehensive but must also be curated

Can it detect & minimize bias & hallucination?

- How does the platform detect and correct bias?
- How can it prevent hallucination (providing random and untrue answers with absolute aplomb and convictions)?

Is it transparent?

- Open vs black-box
- How to **audit**, and explain the model and the answers it generates?
- Does the model track **drift and bias**? And how does it address them?

Does it support regulatory compliance?

- How do foundation models and their usage comply with privacy and government regulations?
- What are the guardrails?
- Who is responsible for an inadvertently exposed PII or a “wrong answer”?

Is it safe?

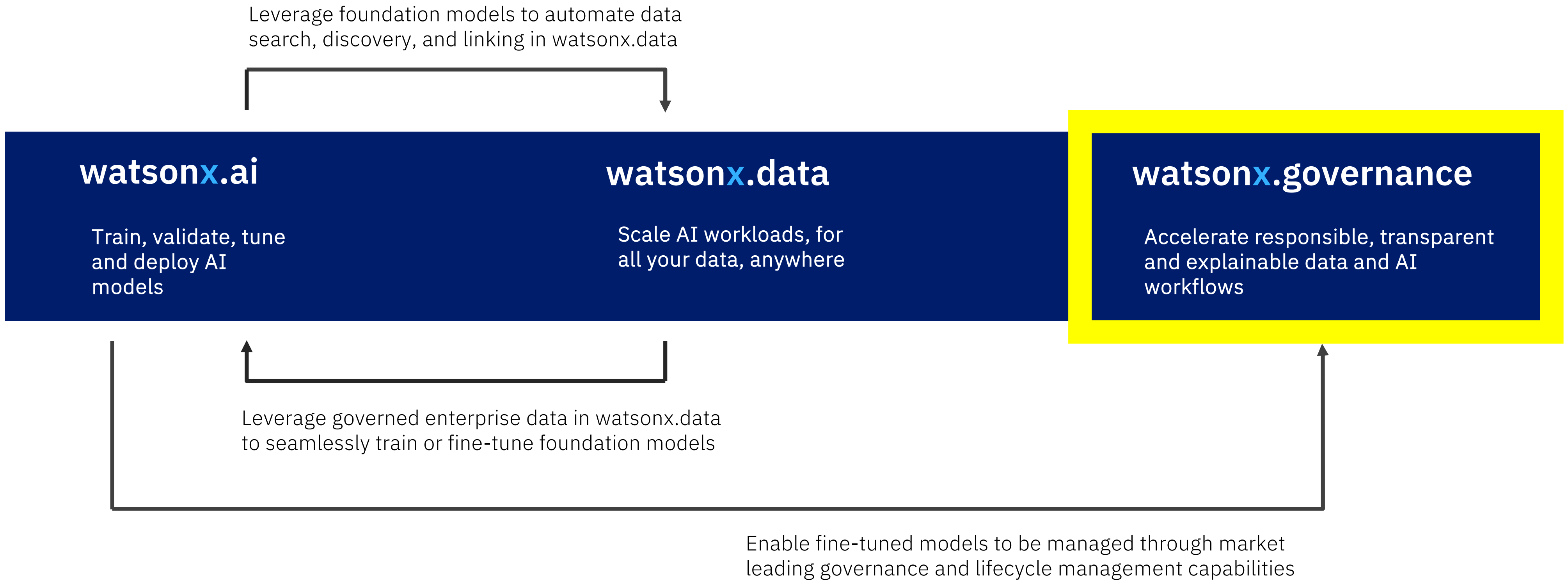
- Who has control over the model, input data, and output data?
- How to ensure that confidential information is not given out?
- How is it monitored?
- What safety features and guardrails are in place?

Can it be customized?

- Hybrid and multicloud?
- Can the model be fine-tuned with clients’ data?
- How can clients update, and extend the model to make it more suitable for their use cases?
- How to integrate with applications? What APIs are in place?

Put AI to work with watsonx

Scale and accelerate the impact of AI with trusted data



Leverage foundation models to automate data search, discovery, and linking in watsonx.data

watsonx.ai

Train, validate, tune and deploy AI models

watsonx.data

Scale AI workloads, for all your data, anywhere

watsonx.governance

Accelerate responsible, transparent and explainable data and AI workflows

Leverage governed enterprise data in watsonx.data to seamlessly train or fine-tune foundation models

Enable fine-tuned models to be managed through market leading governance and lifecycle management capabilities

IBM watsonx.governance
↓
end-to-end automated AI lifecycle governance toolkit built to mitigate risk and improve compliance

watsonx.governance

Enable responsible, transparent and explainable AI workflows

An end-to-end AI lifecycle governance toolkit encompassing both data and AI governance to mitigate risk and improve compliance

Govern across the AI lifecycle by automating and consolidating tools, applications and platforms.

Manage risk and protect reputation by automating workflows to better detect fairness, bias and drift.

Adhere to regulatory compliance by translating growing regulations into enforceable policies.

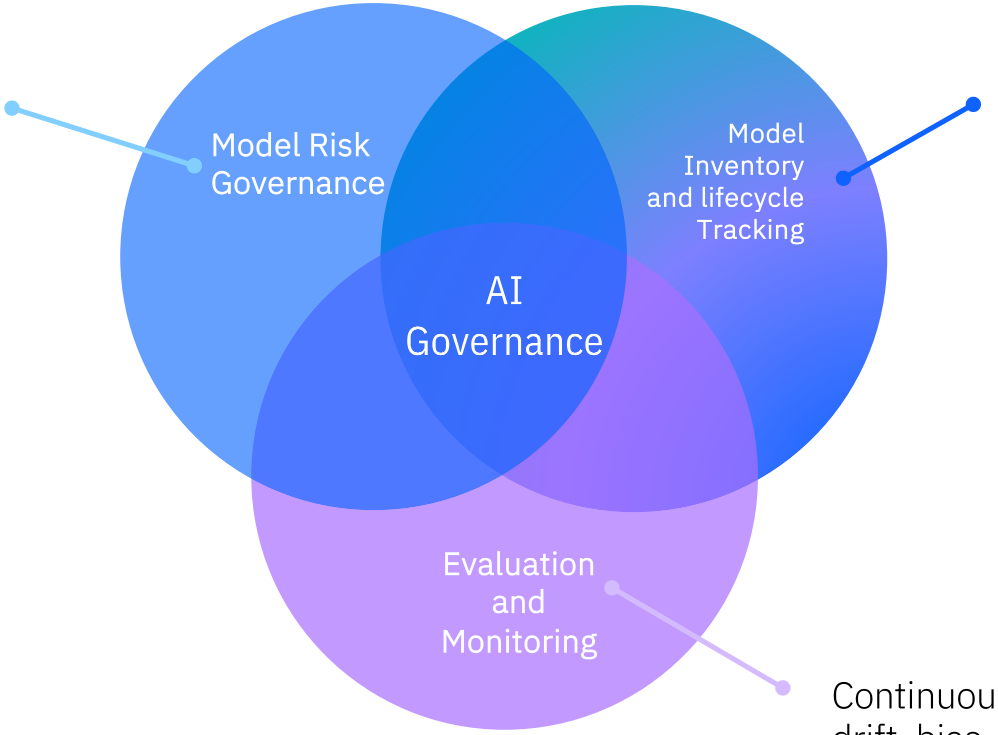
Comprehensive
Govern the end-to-end AI lifecycle with metadata capture at each stage

Open
Support governance of models built and deployed in 3rd party tools.

Automatic metadata and data transformation/lineage capture though Python notebooks.

A Governed, Trusted AI Lifecycle

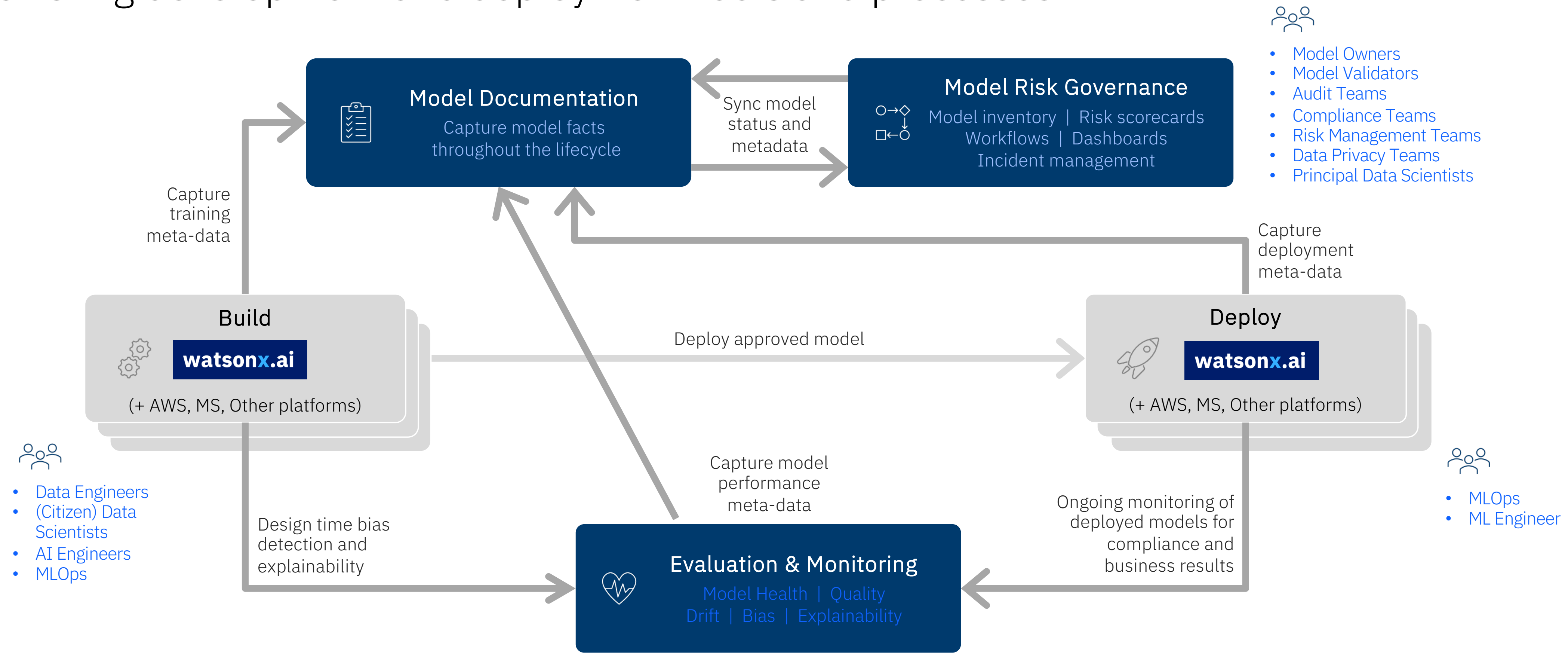
Workflows, dashboards for risk & incident management


Model Risk Governance
Model Inventory and lifecycle Tracking
AI Governance
Evaluation and Monitoring

Capture model facts throughout the entire lifecycle

Continuous monitoring for accuracy, drift, bias, explainability, and performance

AI Governance based on watsonx.governance integrates and augments your existing development and deployment tools and processes



Model Documentation

Capture model facts throughout the lifecycle

Model Risk Governance

Model inventory | Risk scorecards

Workflows | Dashboards

Incident management

Sync model status and metadata

- Model Owners
- Model Validators
- Audit Teams
- Compliance Teams
- Risk Management Teams
- Data Privacy Teams
- Principal Data Scientists

Capture training meta-data

Capture deployment meta-data

Build

watsonx.ai

(+ AWS, MS, Other platforms)

Deploy approved model

Deploy

watsonx.ai

(+ AWS, MS, Other platforms)

- Data Engineers
- (Citizen) Data Scientists
- AI Engineers
- MLOps

Capture model performance meta-data

Design time bias detection and explainability

Ongoing monitoring of deployed models for compliance and business results

- MLOps
- ML Engineer

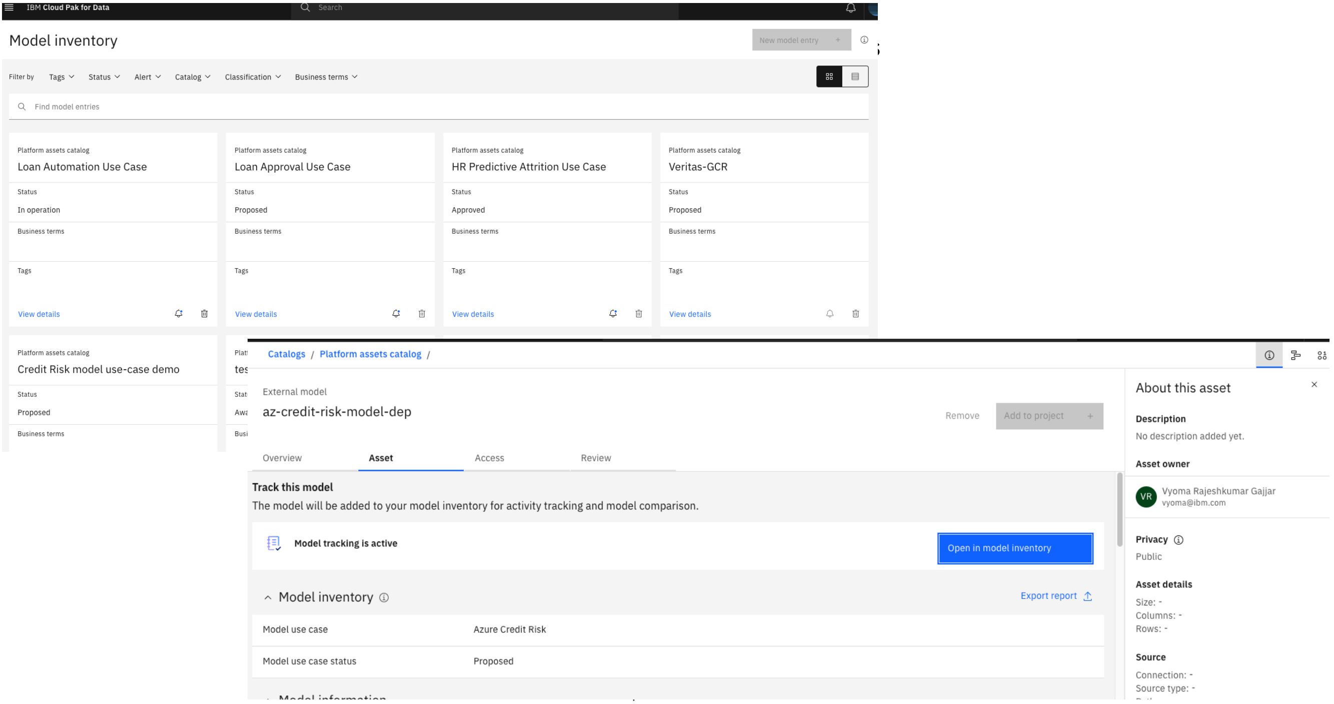
Evaluation & Monitoring

Model Health | Quality

Drift | Bias | Explainability

Model Inventory & Lifecycle Management

- Captures Model facts throughout the lifecycle.
- Provide a singular view of facts across the model lifecycle
- Facilitate subsequent enterprise validation, understand how the model will behave in different business situations
- Support audits, and requests for model facts from auditors, management, stakeholders, and customers



Model inventory

Platform assets catalog
Loan Automation Use Case
Status
In operation
Business terms
Tags
View details

Platform assets catalog
Loan Approval Use Case
Status
Proposed
Business terms
Tags
View details

Platform assets catalog
HR Predictive Attrition Use Case
Status
Approved
Business terms
Tags
View details

Platform assets catalog
Veritas-GCR
Status
Proposed
Business terms
Tags
View details

Platform assets catalog
Credit Risk model use-case demo
Status
Proposed
Business terms

Plat
tes
Stat
Awa
Busi

Catalogs / Platform assets catalog /

External model
az-credit-risk-model-dep

Track this model
The model will be added to your model inventory for activity tracking and model comparison.

Model tracking is active

Model inventory

Model use case	Azure Credit Risk
Model use case status	Proposed

Model information

About this asset

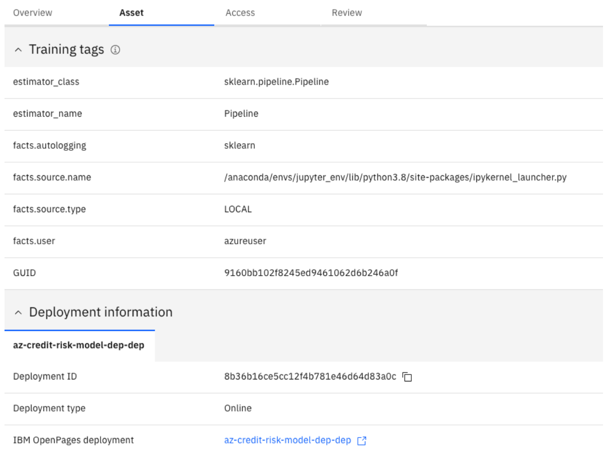
Description
No description added yet.

Asset owner
VR Vyoma Rajeshkumar Gajjar
vyoma@ibm.com

Privacy
Public

Asset details
Size: -
Columns: -
Rows: -

Source
Connection: -
Source type: -



Training tags

estimator_class	sklearn.pipeline.Pipeline
estimator_name	Pipeline
facts.autologging	sklearn
facts.source.name	/anaconda/envs/jupyter_env/lib/python3.8/site-packages/ipykernel_launcher.py
facts.source.type	LOCAL
facts.user	azureuser
GUID	9160bb102f8245ed9461062d6b246a0f

Deployment information

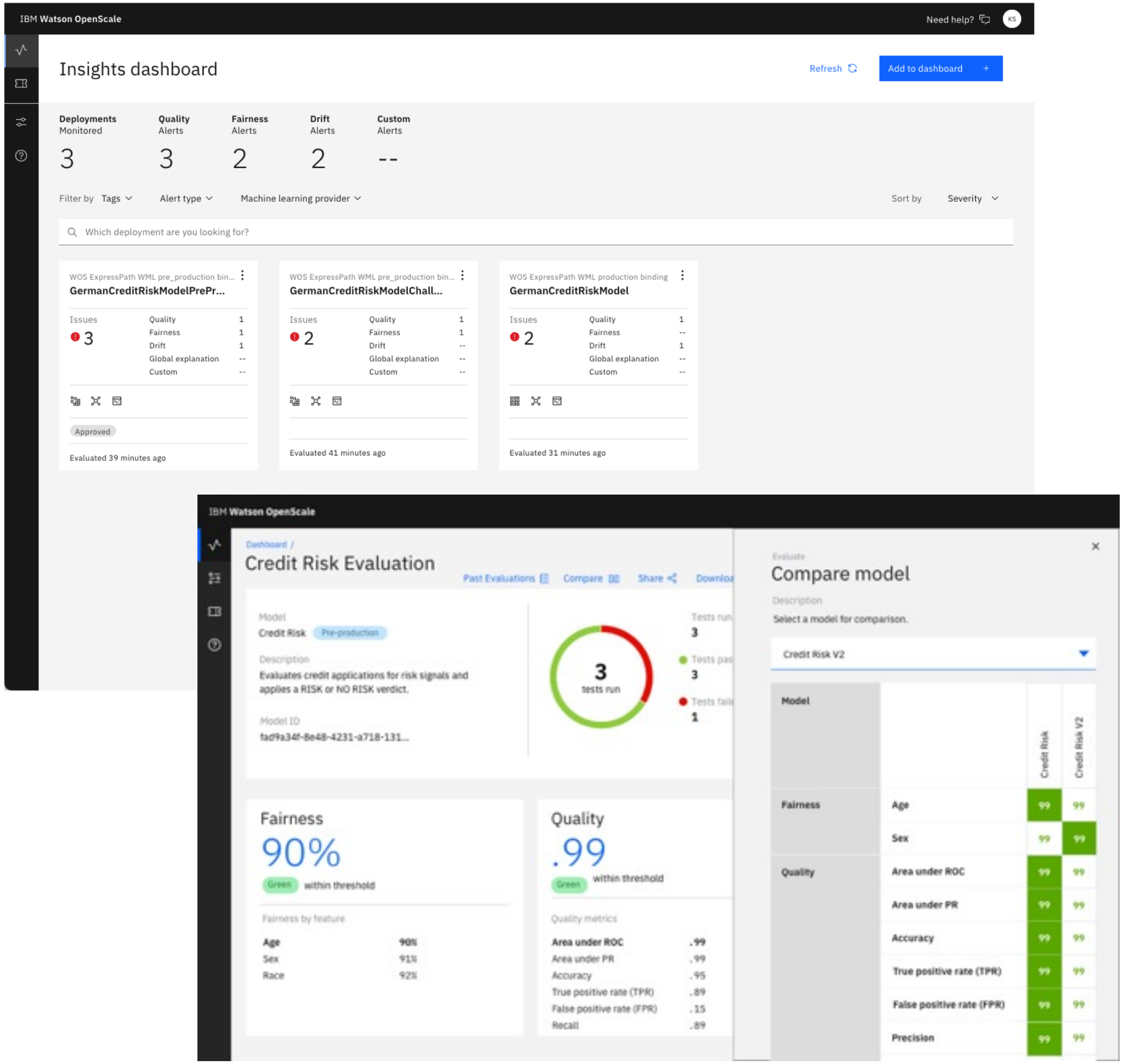
az-credit-risk-model-dep-dep

Deployment ID	8b36b16ce5cc12f4b781e46d64d83a0c
Deployment type	Online
IBM OpenPages deployment	az-credit-risk-model-dep-dep

Evaluation and Monitoring

- Ongoing health monitoring of Models
- Trace and explain Model predictions
- Document metrics and track metric values over time
- Bias detection and mitigation
- Notification of issues when quality thresholds or business KPIs are violated

OpenScale



Insights dashboard

Deployments Monitored	Quality Alerts	Fairness Alerts	Drift Alerts	Custom Alerts
3	3	2	2	--

Filter by Tags ⌄ Alert type ⌄ Machine learning provider ⌄ Sort by Severity ⌄

Which deployment are you looking for?

WOS ExpressPath WML pre_production bin...
GermanCreditRiskModelPrePr...

Issues		
● 3	Quality	1
	Fairness	1
	Drift	1
	Global explanation	--
	Custom	--

Approved

Evaluated 39 minutes ago

WOS ExpressPath WML pre_production bin...
GermanCreditRiskModelChall...

Issues		
● 2	Quality	1
	Fairness	1
	Drift	--
	Global explanation	--
	Custom	--

Evaluated 41 minutes ago

WOS ExpressPath WML production binding
GermanCreditRiskModel

Issues		
● 2	Quality	1
	Fairness	--
	Drift	1
	Global explanation	--
	Custom	--

Evaluated 31 minutes ago

Credit Risk Evaluation

Model
Credit Risk Pre-production

Description
Evaluates credit applications for risk signals and applies a RISK or NO RISK verdict.

Model ID
fad9a34f-8e48-4231-a718-131...

3 tests run

Tests run
3
● Tests pas
3
● Tests faile
1

Fairness
90%
Green within threshold

Fairness by feature

Age	90%
Sex	91%
Race	92%

Quality
.99
Green within threshold

Quality metrics

Area under ROC	.99
Area under PR	.99
Accuracy	.95
True positive rate (TPR)	.89
False positive rate (FPR)	.15
Recall	.89

Evaluate
Compare model ×

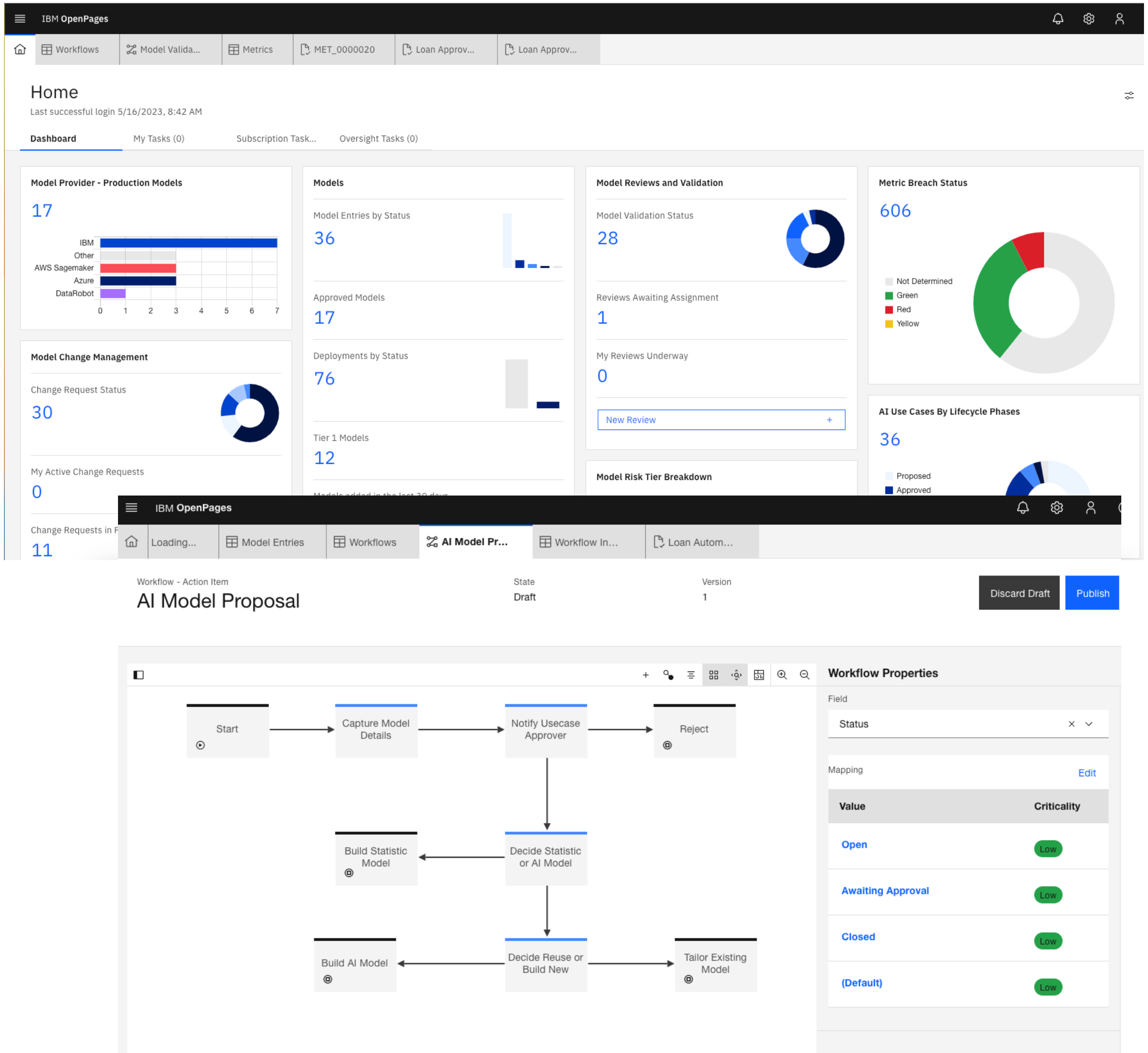
Description
Select a model for comparison.

Credit Risk V2 ▼

Model		Credit Risk	Credit Risk V2
Fairness	Age	99	99
	Sex	99	99
Quality	Area under ROC	99	99
	Area under PR	99	99
	Accuracy	99	99
	True positive rate (TPR)	99	99
	False positive rate (FPR)	99	99
	Precision	99	99

Model Risk Governance

1. Consolidated view of models from multiple platforms
2. View development status, model performance and alerts or emerging issues
3. Monitor and trigger workflows for model validation, retraining and performance issues
4. Track issues and incidents related to models in OpenPages included Issue Management Solution
5. Workflow to document and approve changes to models



Home

Last successful login 5/16/2023, 8:42 AM

Model Provider - Production Models

17

Model Change Management

Change Request Status 30

My Active Change Requests 0

Change Requests in F 11

Models

Model Entries by Status 36

Approved Models 17

Deployments by Status 76

Tier 1 Models 12

Models added in the last 30 days

Model Reviews and Validation

Model Validation Status 28

Reviews Awaiting Assignment 1

My Reviews Underway 0

Model Risk Tier Breakdown

Metric Breach Status

606

AI Use Cases By Lifecycle Phases

36

Workflow - Action Item

AI Model Proposal

State: Draft

Version: 1

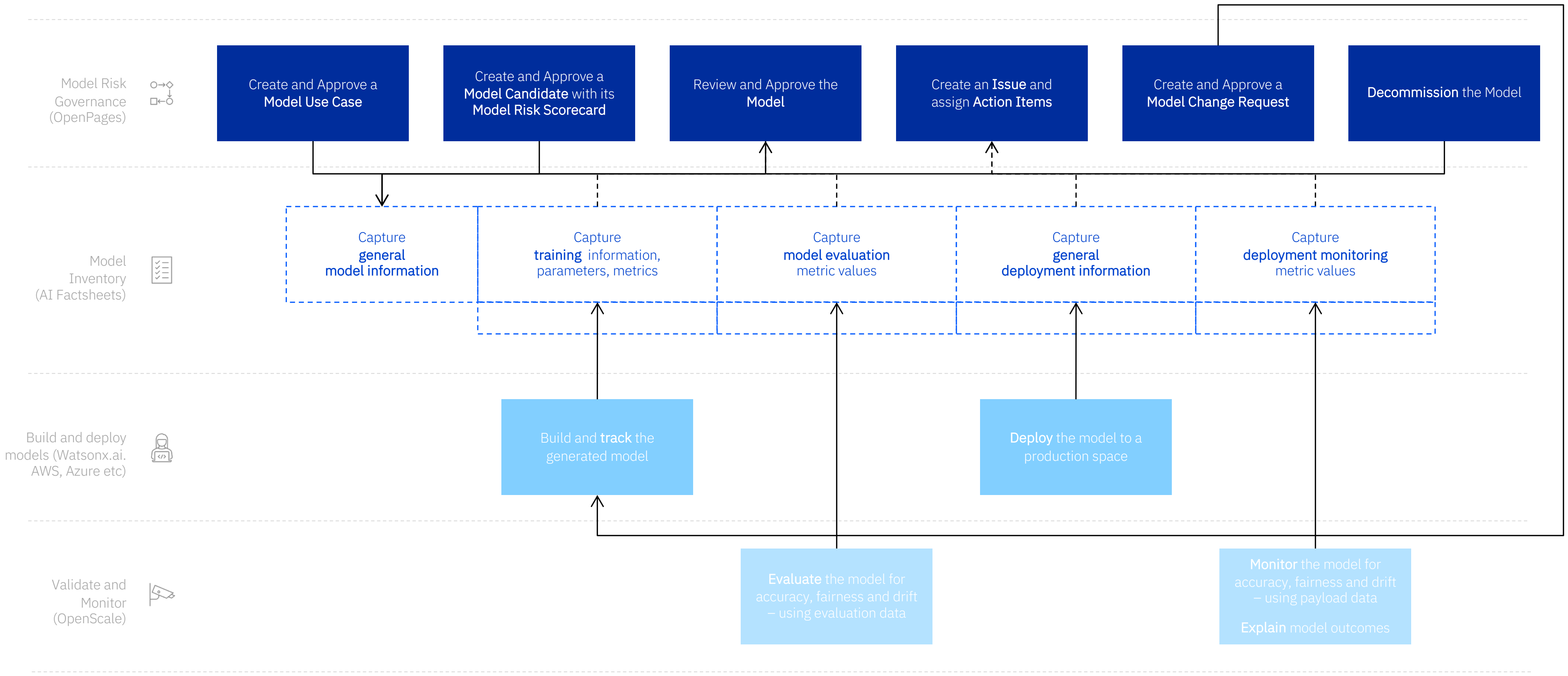
Workflow Properties

Field: Status

Mapping Edit

Value	Criticality
Open	Low
Awaiting Approval	Low
Closed	Low
(Default)	Low

AI Lifecycle Governance – Process View



Model Risk Governance (OpenPages)

Create and Approve a **Model Use Case**

Create and Approve a **Model Candidate** with its **Model Risk Scorecard**

Review and Approve the **Model**

Create an **Issue** and assign **Action Items**

Create and Approve a **Model Change Request**

Decommission the Model

Model Inventory (AI Factsheets)

Capture **general model information**

Capture **training** information, parameters, metrics

Capture **model evaluation** metric values

Capture **general deployment information**

Capture **deployment monitoring** metric values

Build and deploy models (Watsonx.ai. AWS, Azure etc)

Build and **track** the generated model

Deploy the model to a production space

Validate and Monitor (OpenScale)

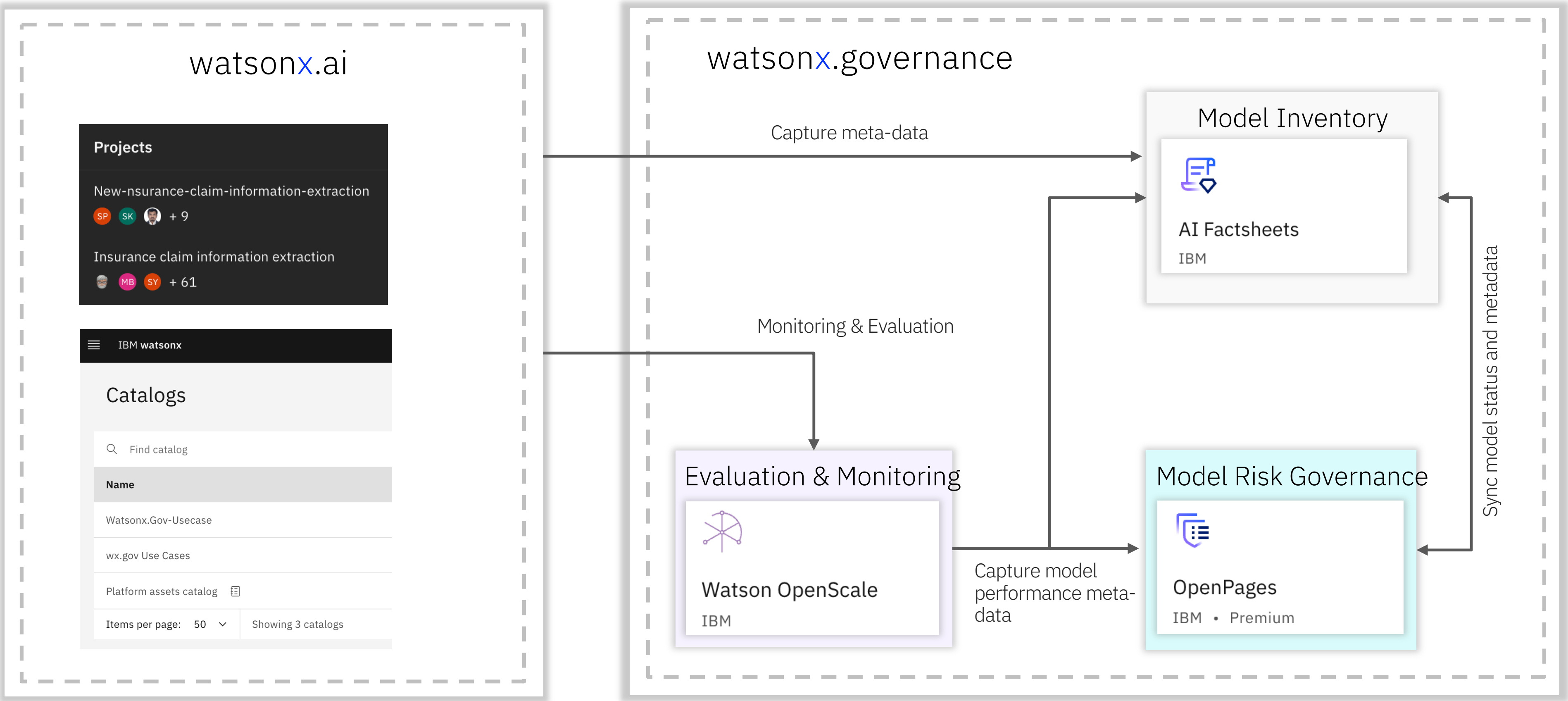
Evaluate the model for accuracy, fairness and drift – using evaluation data

Monitor the model for accuracy, fairness and drift – using payload data

Explain model outcomes

(boxes with solid fill colors are user actions, boxes with dashed outlines are what the integrated tooling does automatically)

Reference Architecture



watsonx.ai

Projects

New-nsurance-claim-information-extraction

SP SK + 9

Insurance claim information extraction

MB SY + 61

IBM watsonx

Catalogs

Find catalog

Name
Watsonx.Gov-Usecase
wx.gov Use Cases
Platform assets catalog

Items per page: 50 Showing 3 catalogs

watsonx.governance

Capture meta-data

Model Inventory

AI Factsheets

IBM

Monitoring & Evaluation

Evaluation & Monitoring

Watson OpenScale

IBM

Capture model performance meta-data

Model Risk Governance

OpenPages

IBM • Premium

Sync model status and metadata

Stakeholder Pain Points

Head of Model Risk (Alex)

I can’t keep track of the **thousands AI models and how we are tracking them and making sure model does not Risk on day 2.**

Enterprise Risk Director (Mary)

We have difficulty in effectively **managing AI models** introduced by evolving new business requirements

Head of Model Validation (Doe)

We have **fragmented and disparate practices** for managing models and metrics to measure their effectiveness across the organization

Prompt Engineer (Ann)

I need a single view to track all the crucial model parameters for multiple prompts on multiple models.

Business owner (Lisa)

I need the ability to see organizations 360 degree view of all models

Chief Compliance Officer (Jane)

I need the ability to define **multiple risk and compliance frameworks like EU AI Ethics act with harmonized controls around AI models**

Demo

IBM