

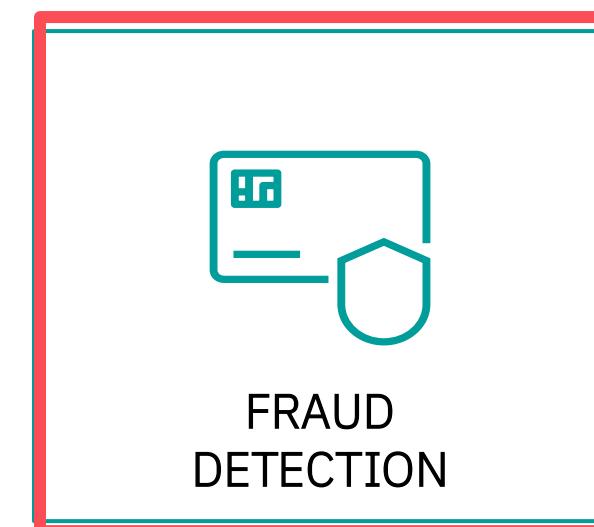
Monitoring AI models in IBM z/OS

Galina Ford
Machine Learning for IBM z/OS Technical Lead
and Architect
gorelik@us.ibm.com



AI on IBM Z:
make more
valuable
outcomes
possible
for every
industry

Financial Services



FRAUD
DETECTION



ANTI-MONEY
LAUNDERING

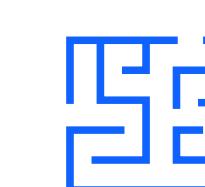


RISK SCORING



CREDIT
DECISIONING

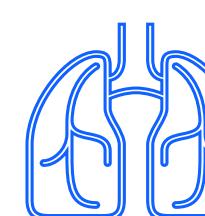
Insurance



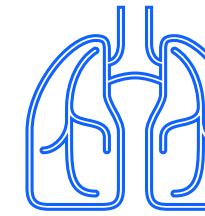
CLAIMS FRAUD
DETECTION



INTELLIGENT
UNDERWRITING

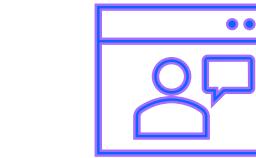


TEXT/IMAGE
PROCESSING

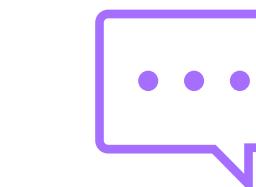


PRODUCT
RECOMMENDATIONS

Government



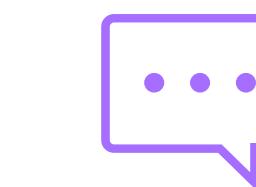
GEOSPATIAL IMAGE
ANALYSIS



FRAUD & SENTIMENT
ANALYSIS



AUDIT &
COMPLIANCE



CHAT SERVICE

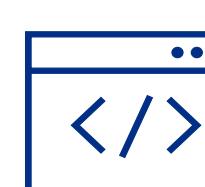
Others



RETAIL
INVENTORY/DEMAND
FORECASTING



SYSTEM ADMIN
ASSISTANT



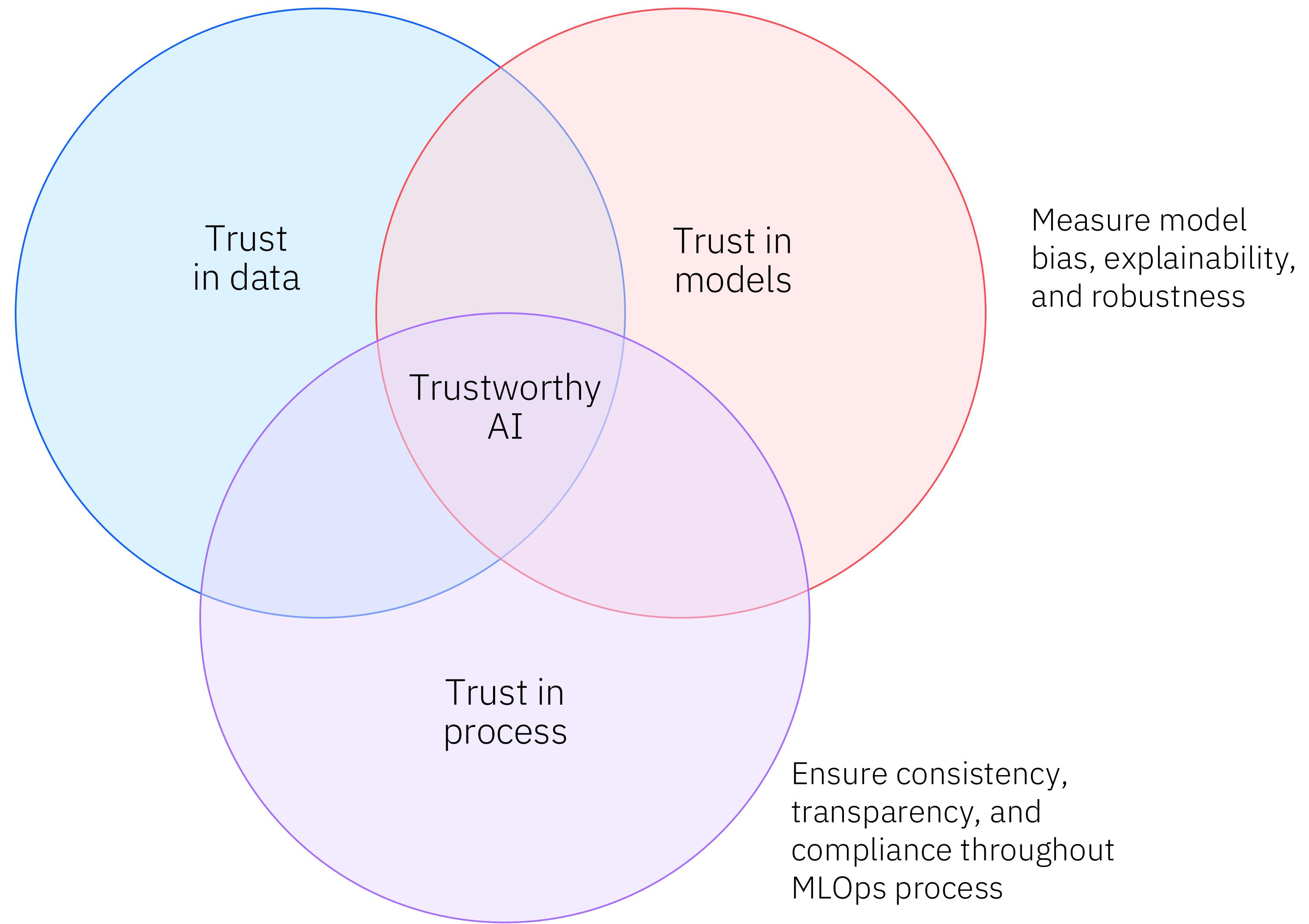
CODE
ASSISTANT



TRANSPORTATIO
N LOGISTICS

Trustworthy AI

Govern data to track lineage, understand quality, and control access to sensitive fields - while being available for self-service usage

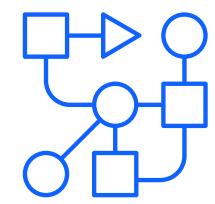


Machine Learning for IBM z/OS

Accelerated inferencing at scale for Transactional AI workloads on z/OS

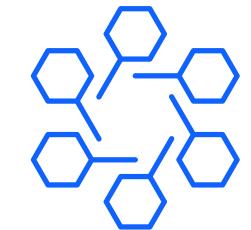
The full-featured machine learning platform on z/OS designed to infuse AI into mission critical IBM Z enterprise applications by leveraging AI models trained anywhere or on IBM Z and deploying them on z/OS including a multi-model architecture, co-located with enterprise applications, transaction data, and business logic for high throughput and extremely low latency to modernize applications and drive business insights at scale

Benefits



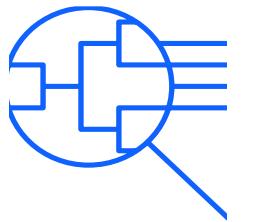
End-end platform

GUI and script driven features for end-end AI model lifecycle management



Scalable AI

Leverage the latest accelerator to score transactions natively in CICS, IMS and batch applications at scale



Easy import and deploy

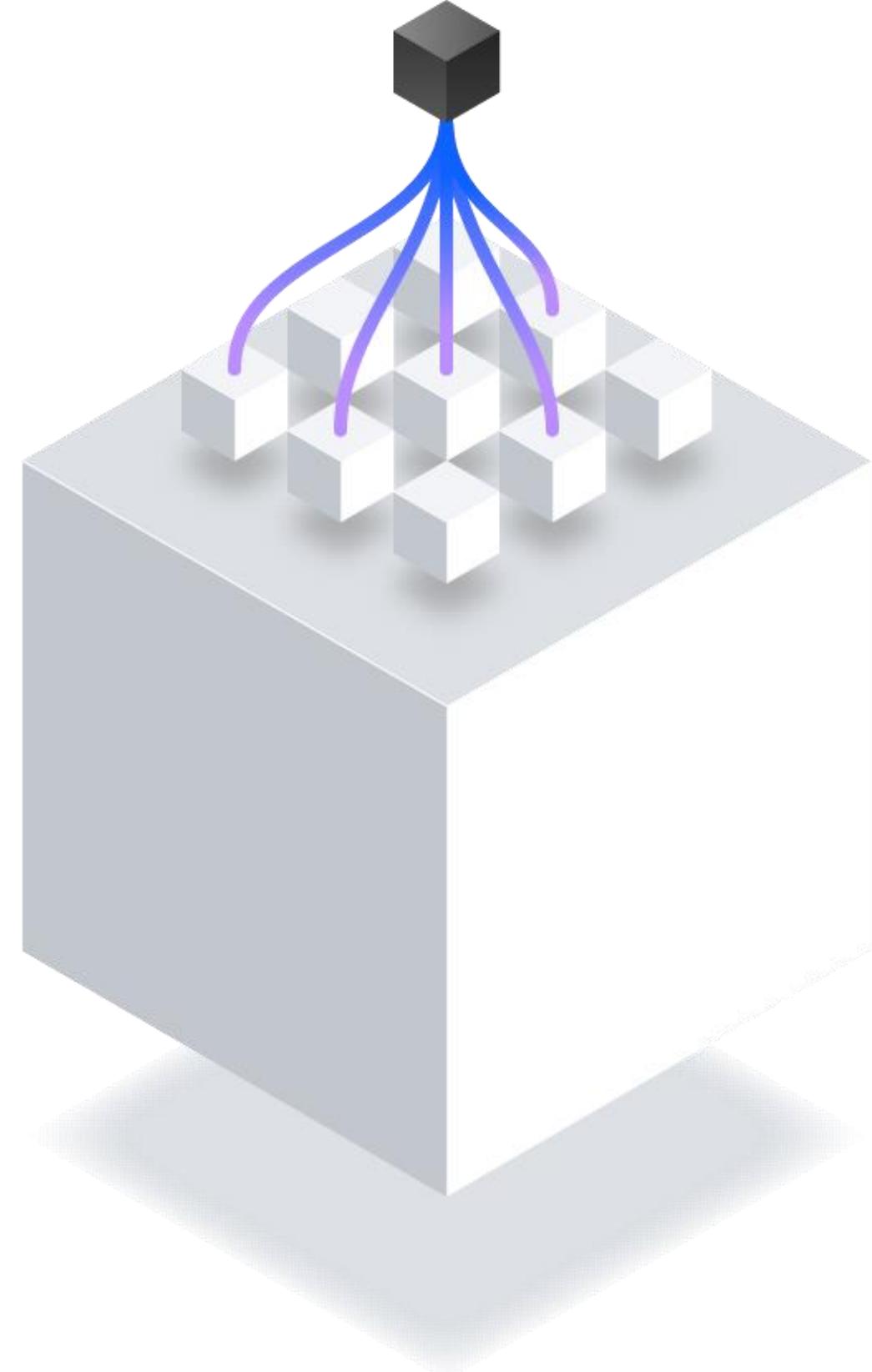
Leverage predictive models and Encoder LLMs trained anywhere or with [watsonx.ai](#) and deploy with MLz



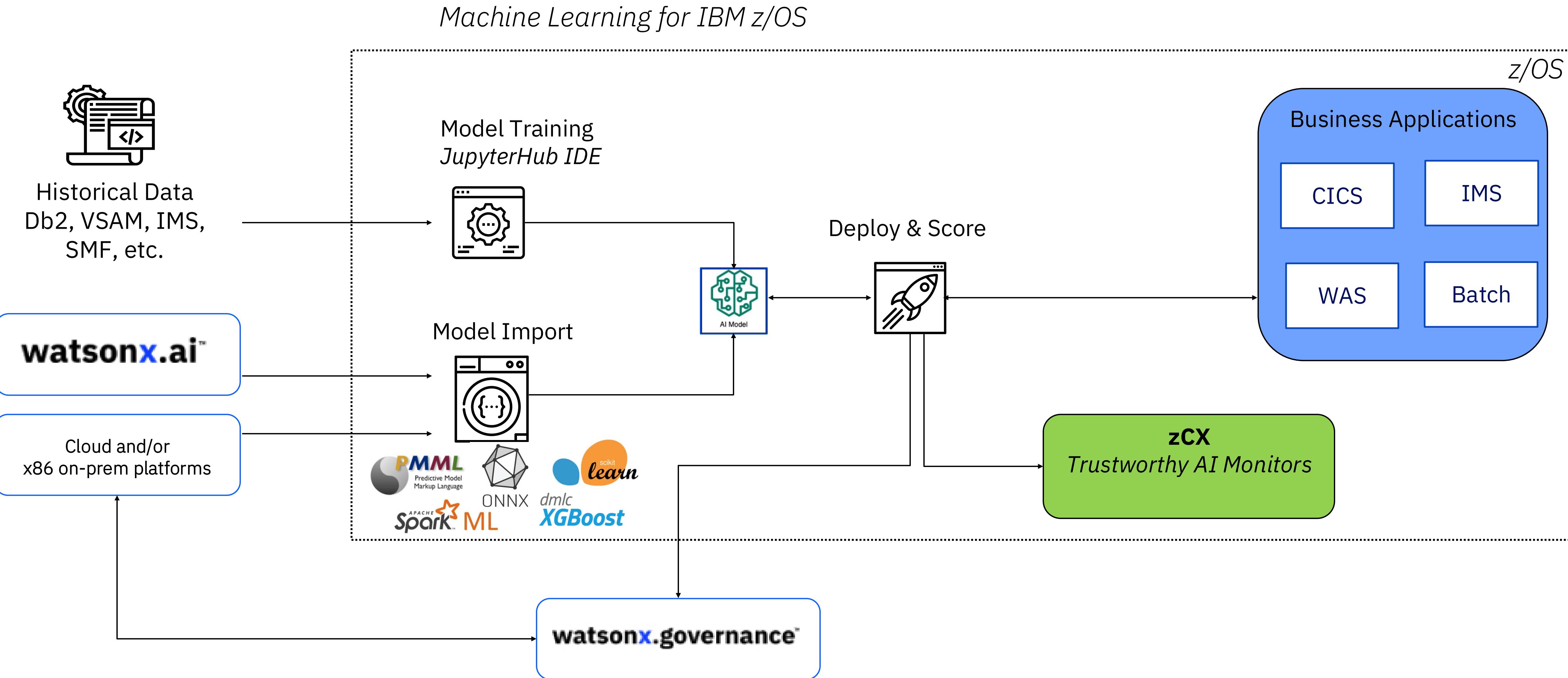
Trustworthy features

Leverage built in explainability and drift detection features or enterprise AI governance with [watsonx.governance](#)

Transactional AI
Infuse AI into every transaction



Train anywhere – Deploy on IBM Z



Key Features of Machine Learning for IBM z/OS v3.2 – Enterprise Edition

3

Scoring Engines

Online scoring for various types of machine learning models and deep learning and encoder LLMs

Leverage on-chip AI accelerator for scoring acceleration

4

Integrated Scoring

In-transaction scoring through native CICS and WOLA interface for CICS, IMS and BATCH COBOL applications

1

GUI Configuration

[Web-based Configuration Tool](#) for single instance and HA configuration

2

Model training tool

Integrated Jupyter server for model training on Z

Leverage IBM Z Spark 3.5 and Python AI Toolkit for training and scoring

5

Model management

[Web-based UI and REST APIs](#) for end-to-end model lifecycle management

6

Trustworthy AI

Deploy your models with the utmost trust leveraging the integrated trustworthy AI capabilities

Trustworthy AI Capabilities

*Supports multiple scoring interfaces:
REST, CICS, and WOLA (e.g., from IMS
transactions and batch jobs)*

*Supports multiple model types:
SparkML, Python based scikit-learn
and XGBoost, PMML, ONNX models*

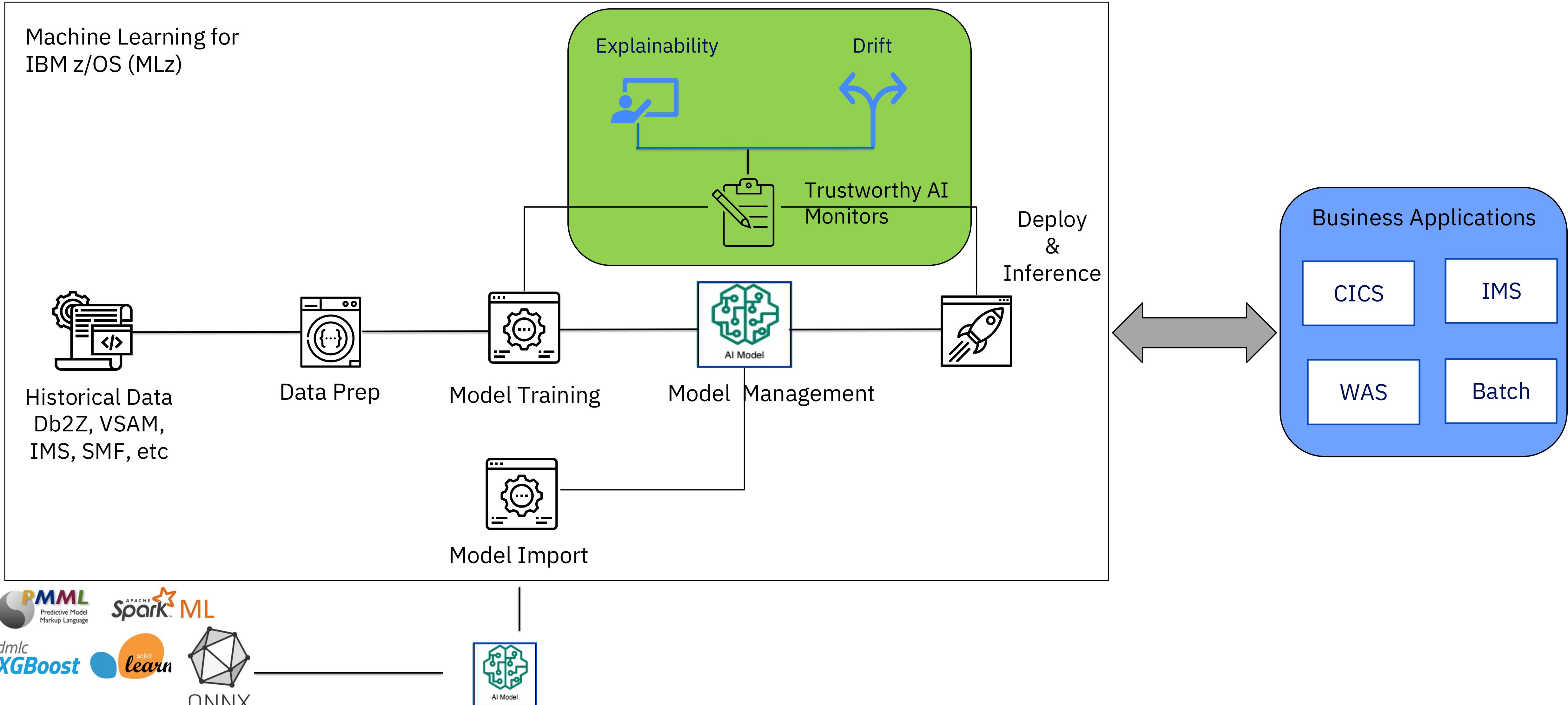
Explainability

*Enables continuous monitoring
of z/OS transactions that are
scored by AI models and
provides explanations on the
model's output at the individual
transaction level*

Drift Detection

*Enables monitoring and
evaluation of both output drift
and feature drift, providing
data scientists with insights
from different perspectives*

Trustworthy AI – in Machine Learning for IBM z/OS



Model deployment and management

- Intuitive web user interface for AI model life cycle management
- REST APIs available for automation
- Dual control available for deployment related activities

Machine Learning for IBM z/OS

Models

Name	Published by	Published on	Type	Latest version	Status	Evaluator
spark01	mlzuse1	Aug 3, 2025 10:11 AM	IBM SparkML	v1	✓	—
xgboost11	mlzuse1	Aug 3, 2025 10:11 AM	XGBoost	v1	✓	—
snapMLTest1	mlzuse1	Aug 3, 2025 10:10 AM	PMML	v1	✓	—
pmmltest1	mlzuse1	Aug 3, 2025 10:09 AM	PMML	v1	✓	—
onnxModel1	mlzuse1	Jul 8, 2025 12:20 PM	ONNX	v1	✓	—

Models per page: 10 ▾ 1–5 of 5 items

Import model

View details

Deploy

Export

Delete

Machine Learning for IBM z/OS

Deployments

Deployment name	Deployed by	Model name (version)	Scoring service	Type	Engine	Deployed on	Next evaluation
pmmlTest1D1	mlzuse1	pmmltest1 (v1)	scoringServer1	online	PMML	Pending approval	—
spark01D1	mlzuse1	spark01 (v1)	scoringServer1	online	Spark	Aug 3, 2025 10:17 AM	—
onnxModelServ1	mlzuse1	onnxModel1 (v1)	scoringServer1	online	ONNX	Jul 8, 2025 1:21 PM	—

Deployments per page: 10 ▾ 1–3 of 3 items

1 ▾ of 1 page

Model Monitoring

- Requires a [z/OS Container Extensions instance](#)
- [Intuitive web user interface](#) for AI model life cycle management
- [REST APIs](#) available for automation

Machine Learning for IBM z/OS

Dashboard Models Deployments Data sources Runtimes

Search deployments

Deployment name	Deployed by	Model name (version)	Scoring service	Type	Engine	Deployed on	Next evaluation
PMML19	mlzuse1	fraudPMML (v1)	scoringServ1	online	PMML	Nov 4, 2025 3:39 PM	—

Deployments per page: 10 1–1 of 1 items

View details

Schedule evaluation...

Test API call

Update

watsonx integration...

Monitor

Remove

Create monitor

Create a new monitor by filling out deployment and data information, including feature and output details. All monitor details must be completed and the monitor successfully created prior to enabling trustworthy AI capabilities. All monitor details apply to all enabled trustworthy AI capabilities.

Deployment

Monitor name ⓘ
Enter monitor name

Deployment scoring endpoint ⓘ
https://127.0.0.1:12345

Features

Algorithm type
Binary classification

Data collection ⓘ
 Store samples

Upload training data
Trustworthy AI capabilities require access to training data to train the deployed model. Max file size types are .csv and .json.
Drag and drop files here or click

training.csv ×

Edit monitor

Update the existing monitor by filling out deployment and data information, including feature and output details. All monitor details must be completed and the monitor successfully created prior to enabling trustworthy AI capabilities. All monitor details apply to all enabled trustworthy AI capabilities.

Deployment

Features

Feature columns
Shown are the features and label derived from the deployed model. Select the Categorical checkbox for any features that should be identified as categorical data. The Type column is automatically set based on your model schema, as well as the Label column, which represents the ground truth data your model is trying to predict. You cannot change the input for Type and Label from this table.

Feature	Type	Categorical	Label
fraud	integer	<input type="checkbox"/>	<input checked="" type="checkbox"/>
transaction_amount	real	<input type="checkbox"/>	<input type="checkbox"/>
merchant_category	real	<input type="checkbox"/>	<input type="checkbox"/>
customer_age	real	<input type="checkbox"/>	<input type="checkbox"/>
transaction_location	real	<input type="checkbox"/>	<input type="checkbox"/>
customer_type	real	<input type="checkbox"/>	<input type="checkbox"/>
transaction_day_of_year	real	<input type="checkbox"/>	<input type="checkbox"/>
transaction_day_of_month	real	<input type="checkbox"/>	<input type="checkbox"/>
transaction_day_of_week	real	<input type="checkbox"/>	<input type="checkbox"/>
transaction_hour_of_day	real	<input type="checkbox"/>	<input type="checkbox"/>

Monitor settings
monitorFraud1

Explainability allows you to explain specific transactions and understand what influenced the model to come to a decision for that transaction. Enabling explainability for this monitor requires selecting explainer type(s).

Enable explainability for this monitor

Explainer type ⓘ
2 × LIME, SHAP
 LIME
 SHAP

Monitor details

Evaluations ⓘ

Items per page: 10 1–10 of 14 items

Explainability

Drift

Model output
Select the prediction column (the column containing the in the prediction it provides) from the training data.

Feature	Type
probability(0)	number
probability(1)	number

Monitoring: real time or add-hoc?

- Collected transactions
Drift evaluation will run at configured time intervals if the minimum number of transactions have occurred

Explainability will support selection from stored transactions

- Add-hoc input
Drift evaluation and **Explainability** require data to be imported

The image displays two screenshots of a monitoring interface. The left screenshot shows the 'Monitor settings' page for a monitor named 'monitorFraud1'. It includes sections for 'Monitor details', 'Evaluations', 'Explainability', and 'Drift'. The 'Drift' section is active, showing settings for 'Time interval' (1 hour), 'Minimum sample size' (10), 'Important features' (selected: 'customer', 'customer_age', 'customer_type'), and 'Thresholds' (Output drift threshold: 0.5, Feature drift threshold: 0.5). The right screenshot shows the 'Edit monitor' page for the same monitor. It includes tabs for 'Deployment' and 'Features'. Under 'Deployment', there is a 'Monitor name' field (monitorFraud1) and a 'Deployment name' field (PMML19). Under 'Features', there is a 'Data collection' section with a checked checkbox for 'Store samples'. A callout box highlights this checkbox with the text: 'Storing samples (inferences) allows the system to present a database of transactions to explain and continually monitor for drift. If unchecked, you will not be able to schedule automatic evaluations and will need to import the necessary data to explain or monitor for drift at the time of evaluation.' Below this is a 'Upload training data' section with a file upload area labeled 'Drag and drop files here or click to upload' containing 'training.csv'. A dashed blue arrow points from the 'Drift' section of the left screenshot to the 'Data collection' section of the right screenshot, indicating the connection between the two.

Explainability

LIME

Model-agnostic
local, interpretable
explanations -
explains single
predictions of a model

SHAP

Model-specific
global explanations -
the effect of a feature
on the target variable

demoMonitor Explained transactions

View the status and results for explained transactions, and add new transactions to be explained. You may close the page while explanations are loading and return once they are available for viewing.

Search explained transactions

Add transactions

Transaction ID	Transaction date & time	Explanation date & time	Status	
HZZU00IN_0	2025-02-23 09:00:16	2025-02-23 09:00:16		View results
9AyfWrBX_0	2025-02-23 09:00:16	2025-02-23 09:00:16		View results
9AyfWrBX_0	2025-02-22 10:51:57	2025-02-22 10:51:57		View results
HZZU00IN_0	2025-02-22 10:51:57	2025-02-22 10:51:57		View results
uhwE4pQL	2025-02-22 10:39:52	2025-02-22 10:39:52		View results

LIME SHAP

Prediction outcome label: probability(0)

Predicted outcome: 0.9998

The top 3 features with positive contribution:

- transaction_amount
- transaction_hour_of_day
- transaction_day_of_month

The top 3 features with negative contribution:

- transaction_day_of_week
- merchant_category
- customer_type

View transaction record →

Feature influence

This visualization depicts Shapley values, which quantify each feature's influence on a specific transaction's predicted outcome ($f(x)$). Positive (red) values indicate features pushing the prediction above the model's average predicted value ($E[f(x)]$), while negative (blue) values indicate the opposite.

Waterfall

f(x)=0.9998

transaction_amount = 3229.8247 +0.0033

transaction_hour_of_day = 4 +0.0019

transaction_day_of_week = 1 -0.0019

transaction_day_of_month = 7 +0.0017

transaction_amount_1m = 27403.8397 +0.0016

customer_age = 23 +0.0013

transaction_day_of_year = 273 +0.0011

transaction_type = 1 +0.0008

transaction_amount_last_day = 301.0995 +0.0007

merchant_category = 2 -0.0005

3 other features = ... +0.0006

0.989 E[f(x)]=0.989 0.9998

1 of 1 page

Drift

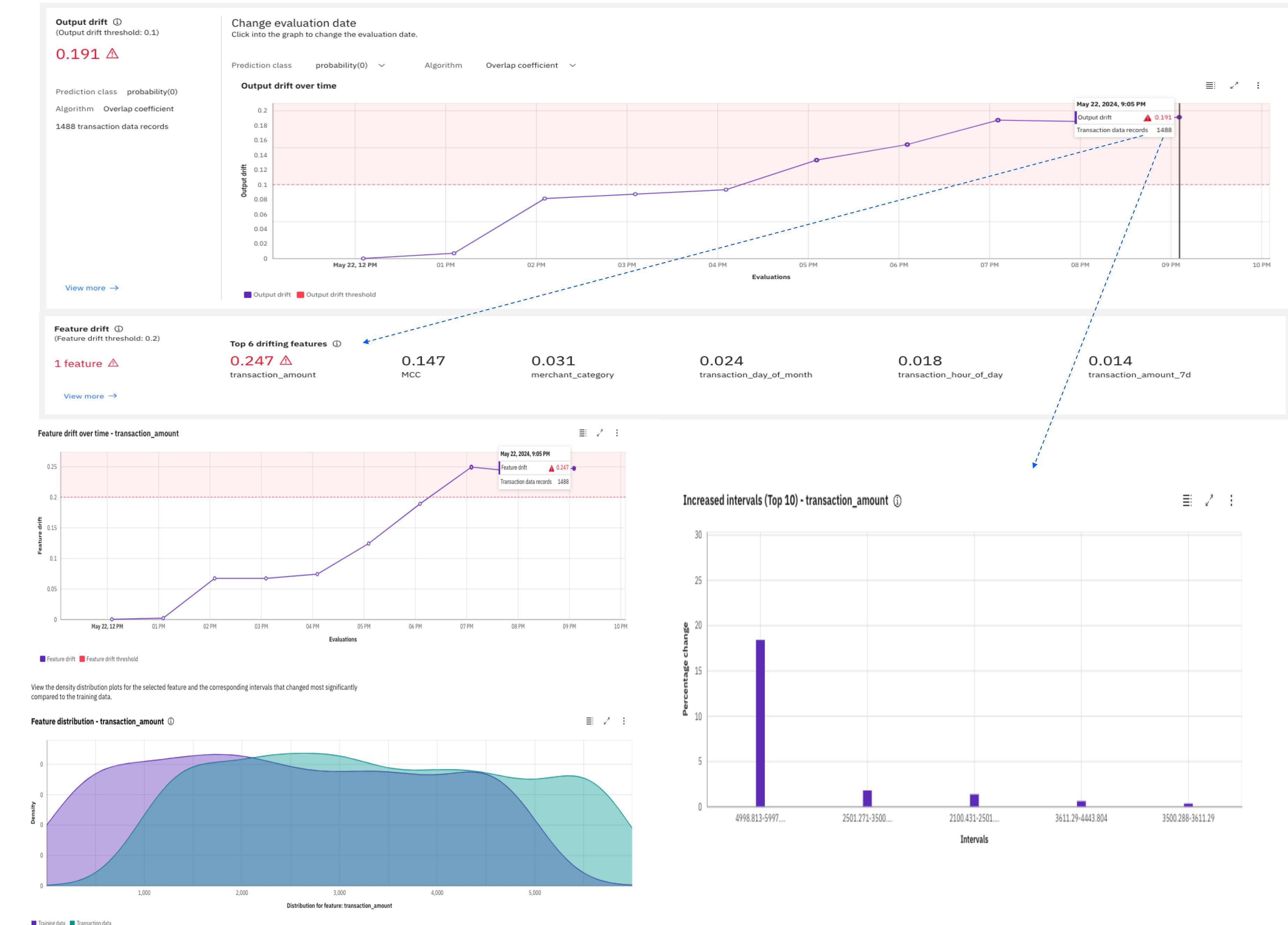
Output Drift

Measured by *Overlap coefficient* and *Total variation distance*

Feature Drift

Categorical features are measured by *Jensen-Shannon* and *Total variation distance*

Non-categorical features are measured by *Overlap coefficient* and *Total variation distance*



Thank you

© 2025 International Business Machines Corporation

IBM, the IBM logo, IBM Z, DataStage and Db2 are trademarks of IBM Corporation, registered in many jurisdictions worldwide. Other product and service names might be trademarks of IBM or other companies. A current list of IBM trademarks is available on <https://www.ibm.com/legal/copyright-trademark>.

This document is distributed “as is” without any warranty, either express or implied. In no event, shall IBM be liable for any damage arising from the use of this information, including but not limited to, loss of data, business interruption, loss of profit or loss of opportunity.

Client examples are presented as illustrations of how those clients have used IBM products and the results they may have achieved. Actual performance, cost, savings or other results in other operating environments may vary.

Not all offerings are available in every country in which IBM operates.

Any statements regarding IBM’s future direction, intent or product plans are subject to change or withdrawal without notice.

All names and references for organizations and other business institutions used in this deliverable’s scenarios are fictional. Any match with real organizations or institutions is coincidental.

All names and associated information for people in this deliverable’s scenarios are fictional. Any match with a real person is coincidental. Videos may show information about or belonging to third parties. Such videos do not suggest endorsement of third parties or their products or services.