

Red Hat
Summit

Connect

AI generativa:
una rivoluzione per le imprese, con rischi e opportunità.
IBM watsonx in azione

Daniele Pietropaoli, Senior Data & AI Technical Sales
IBM Technology, Italy

Rome, 07 - 11 - 2024



Agenda

- Evolution AI : risks and opportunities
- Pillars of AI governance
- Elements of AI Risk
- Watsonx Platform
- watsonx.governance

Rapid evolution in AI drives evolution in the governance of AI

Generative AI

65/35

Split in software spend on non-generative AI (65%) and generative AI (35%). ⁽¹⁾

36

Amplified and new risk with generative AI. ⁽²⁾



Govern both forms of AI in a consistent manner.
Extend your governance to account for the new aspects of generative AI.

Innovation in models

~5000

New foundation models posted on Hugging Face every week.

Continuous innovation in open source and commercial offerings gives you an increasing range of options and trade-offs.



Govern the onboarding of new models.
Govern the trade-offs in use cases.

Consumption models

60/40

Split in software spend on AI platforms (60%) and AI embedded in enterprise applications (40%). ⁽¹⁾

70%

Of independent software vendors will have embedded generative AI capabilities. ⁽¹⁾



Govern all AI, regardless of how and it's created or consumed.

Legislation

200+

Pages in the EU AI Act.

10

Standardization requests in the EU AI Act.

Consequences

Non-compliance with the prohibition of the AI practices, referred to in Article 5, shall be subject to administrative fines of up to 35M EUR or, if the offender is a company, up to 7% of its total worldwide annual turnover.

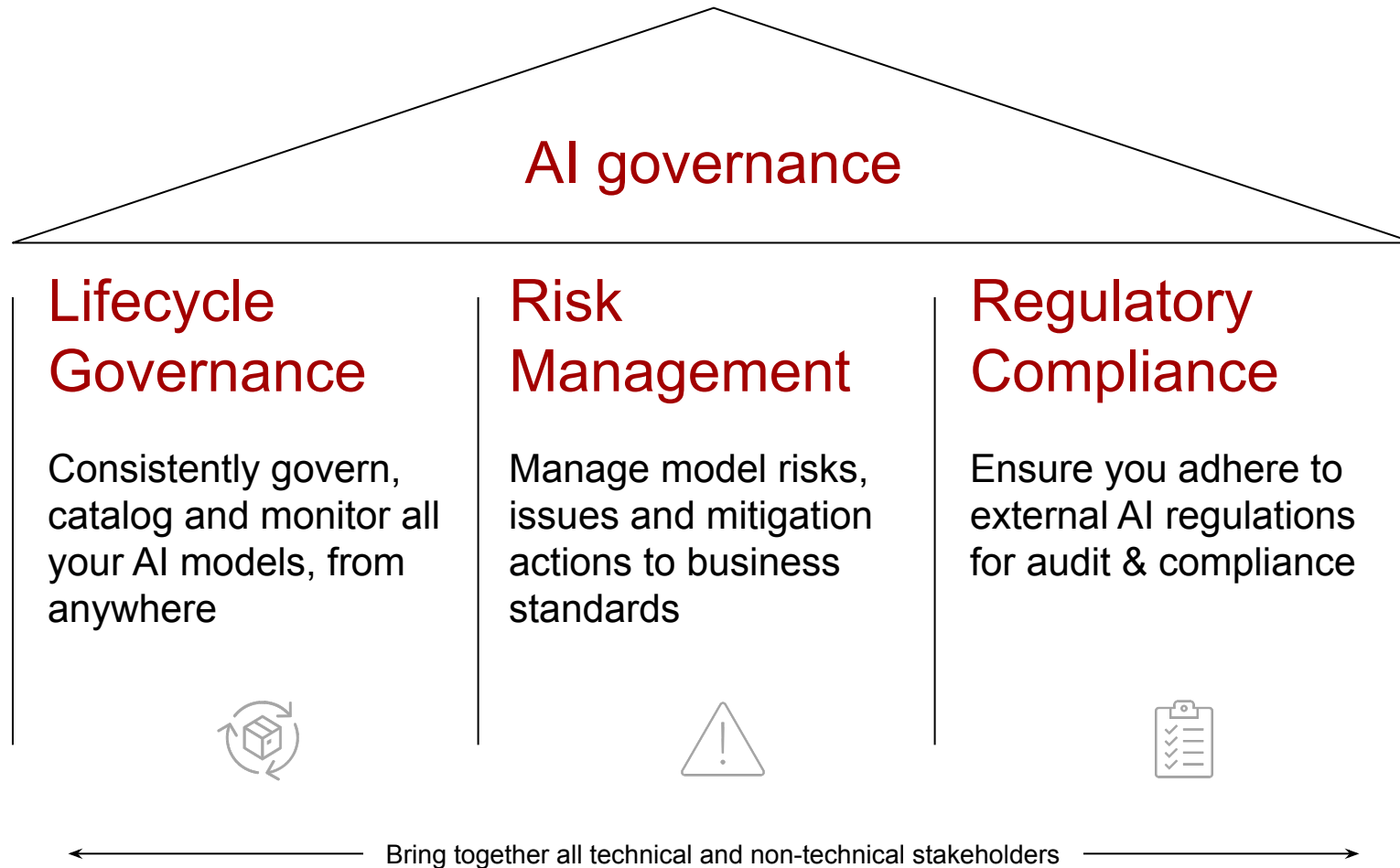


Translate your responsibilities into controls and workflows.
Risk-assess your use cases.
Adopt technical standards.

⁽¹⁾ Gartner – Forecast Analysis: AI Software. 2023-2027, Worldwide

⁽²⁾ IBM AI Risk Atlas

Three pillars of AI governance



Which is easier said than done

Common challenges



AI governance collaboration requires lots of **manual work**; amplified by changes in data and model versions.



Automate the governance activities as much as possible.



Companies have AI in **multiple tools, applications and platform**, developed inside and outside the organization



Consolidate as much as possible in one governance platform.



Governance is **not a one-size-fits-all** approach.



Configure to your specific situation.



Constrain technical teams in their choice of technology.




Open architecture to wrap around tooling of choice.

Elements of AI Risk




Accountability


Accuracy


Fairness



Truthfulness


Transparency


Drift


Trusted data


Sustainability


Explainability


Adversarial
Robustness


IP/PII leakage

...


Regulatory
Risk

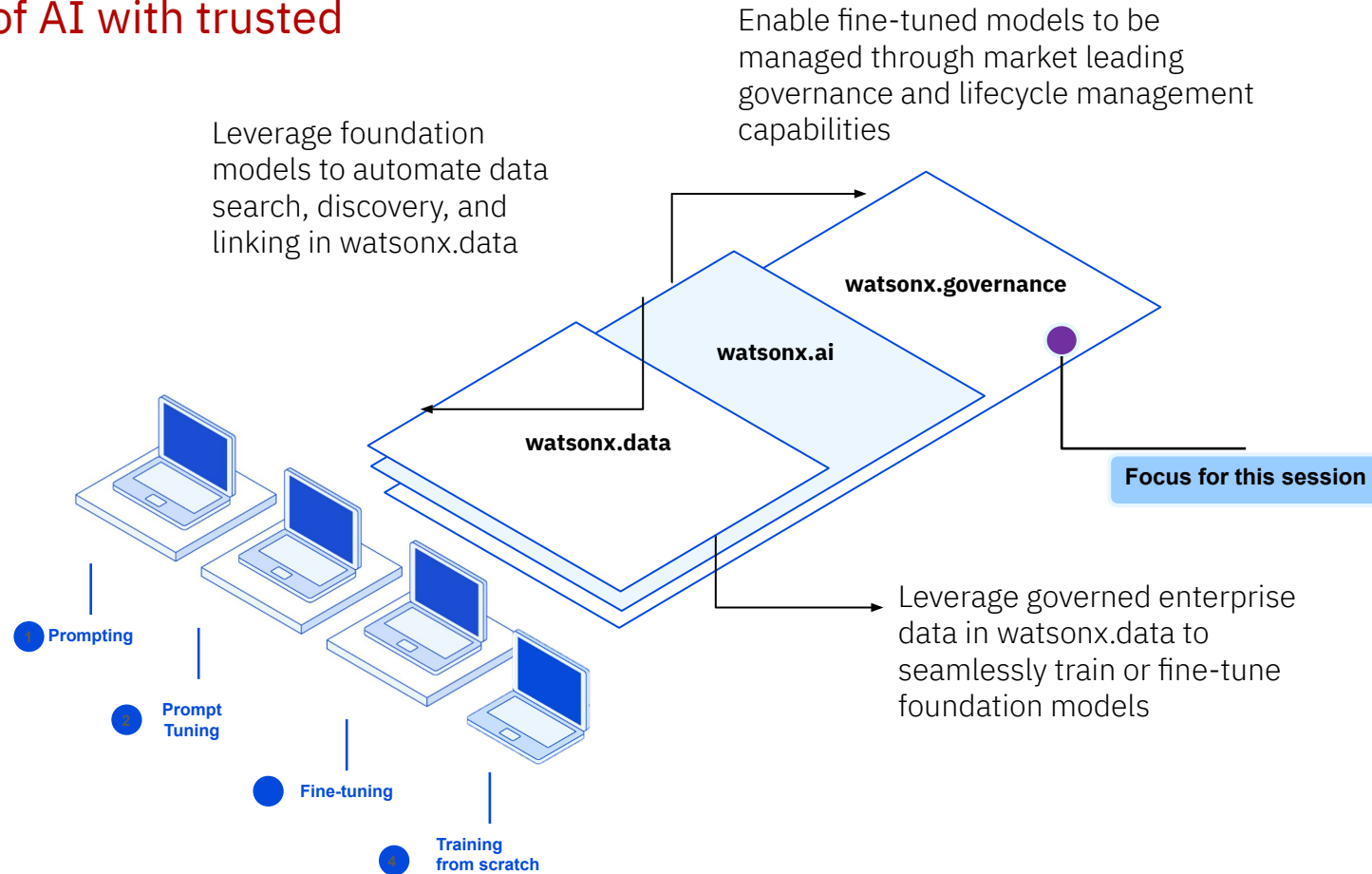

Reputational
Risk


Operational
Risk

watsonx

Main Software Offerings

Scale and accelerate the impact of AI with trusted data.



watsonx.data *Scale AI workloads, for all your data, anywhere*

Fit-for-purpose data store, built on an open lakehouse architecture, supported by querying, governance and open data formats to access and share data.

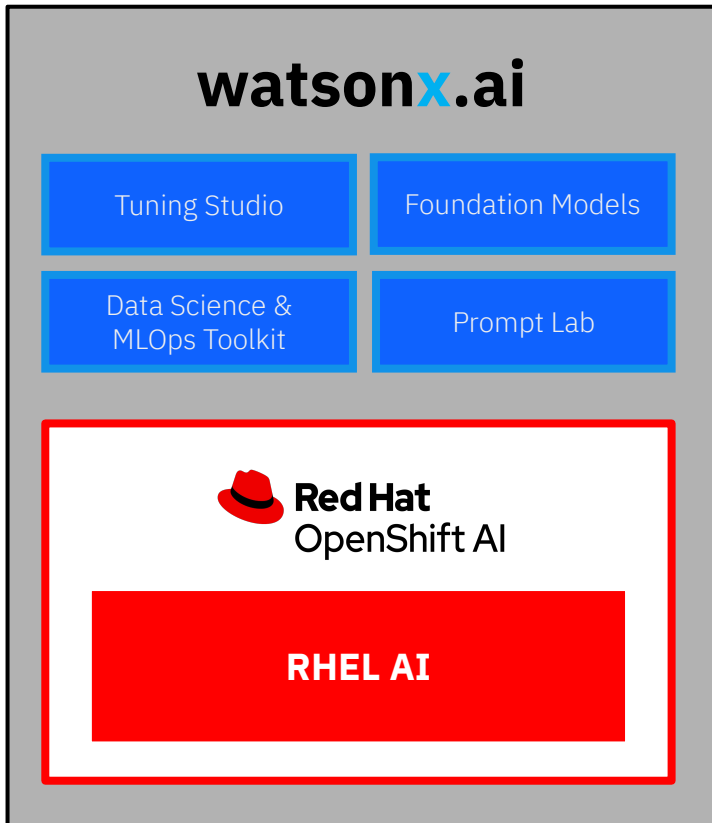
watsonx.ai *Train, validate, tune and deploy AI models*

A next generation enterprise studio for AI builders to train, validate, tune, and deploy both traditional machine learning and new generative AI capabilities powered by foundation models. It enables you to build AI applications in a fraction of the time with a fraction of the data.

watsonx.governance *Enable responsible, transparent and explainable AI workflows*

End-to-end toolkit encompassing both data and AI governance to enable responsible, transparent, and explainable AI workflows.

watsonx.ai runs on Red Hat OpenShift AI & RHEL AI



watsonx.ai

Multi-model flexibility

Access a variety of fit-for-purpose foundation models

Enterprise AI/ML lifecycle

Address LOB multi-user and organizational needs for scale and lifecycle governance

Red Hat OpenShift AI

Model serving & monitoring

Deploy models anywhere with consistent cloud-to-edge production deployment, and monitoring capabilities

Resource optimization and management

Create model pipelines for tuning and alignment, validation and delivery; Share and optimize resources across teams

RHEL AI

Foundation model runtime engine

Develop, deploy, and test open Granite language/code models, supported & indemnified from Red Hat and IBM

Developer-friendly model toolkit

Bring rapid LLM experimentation to developers via supported InstructLab CLI

Runs on any infrastructure



Physical



Virtual



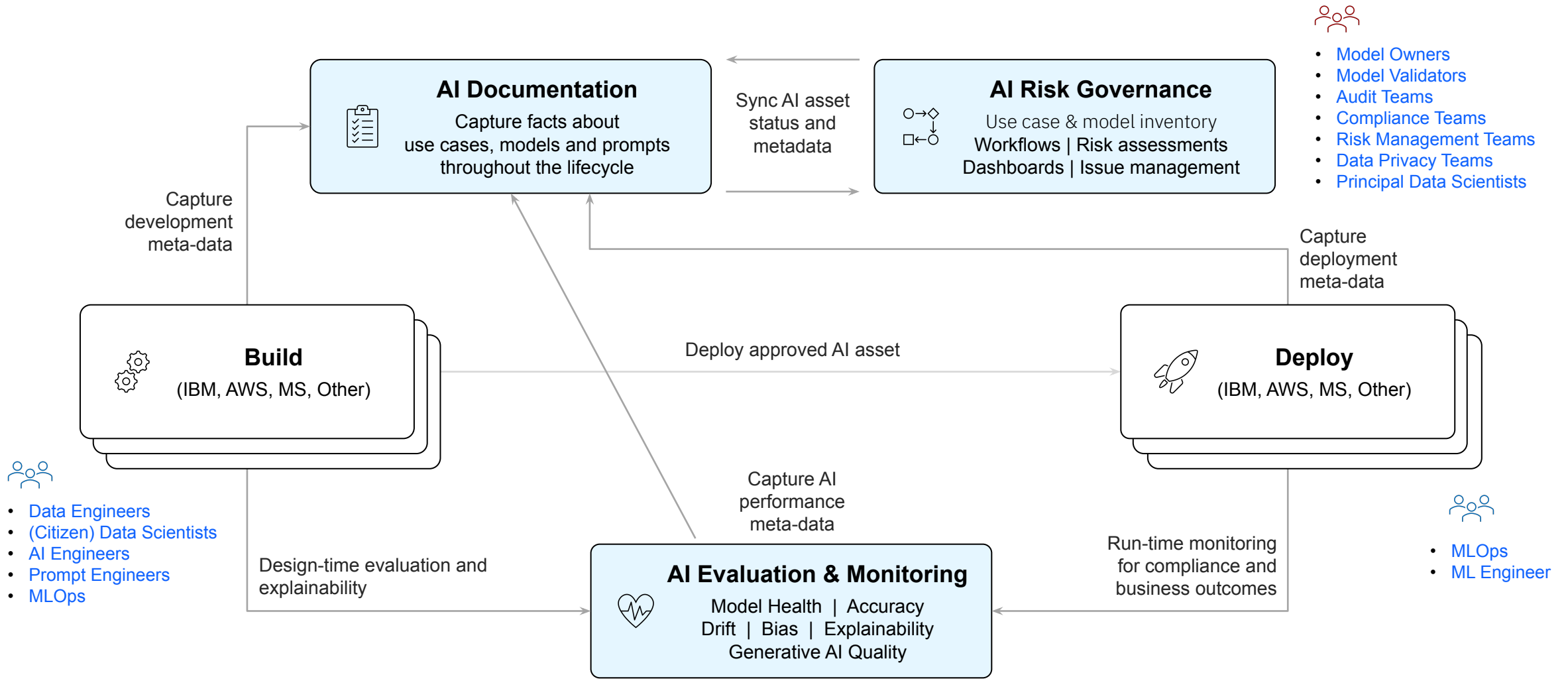
Private



Public

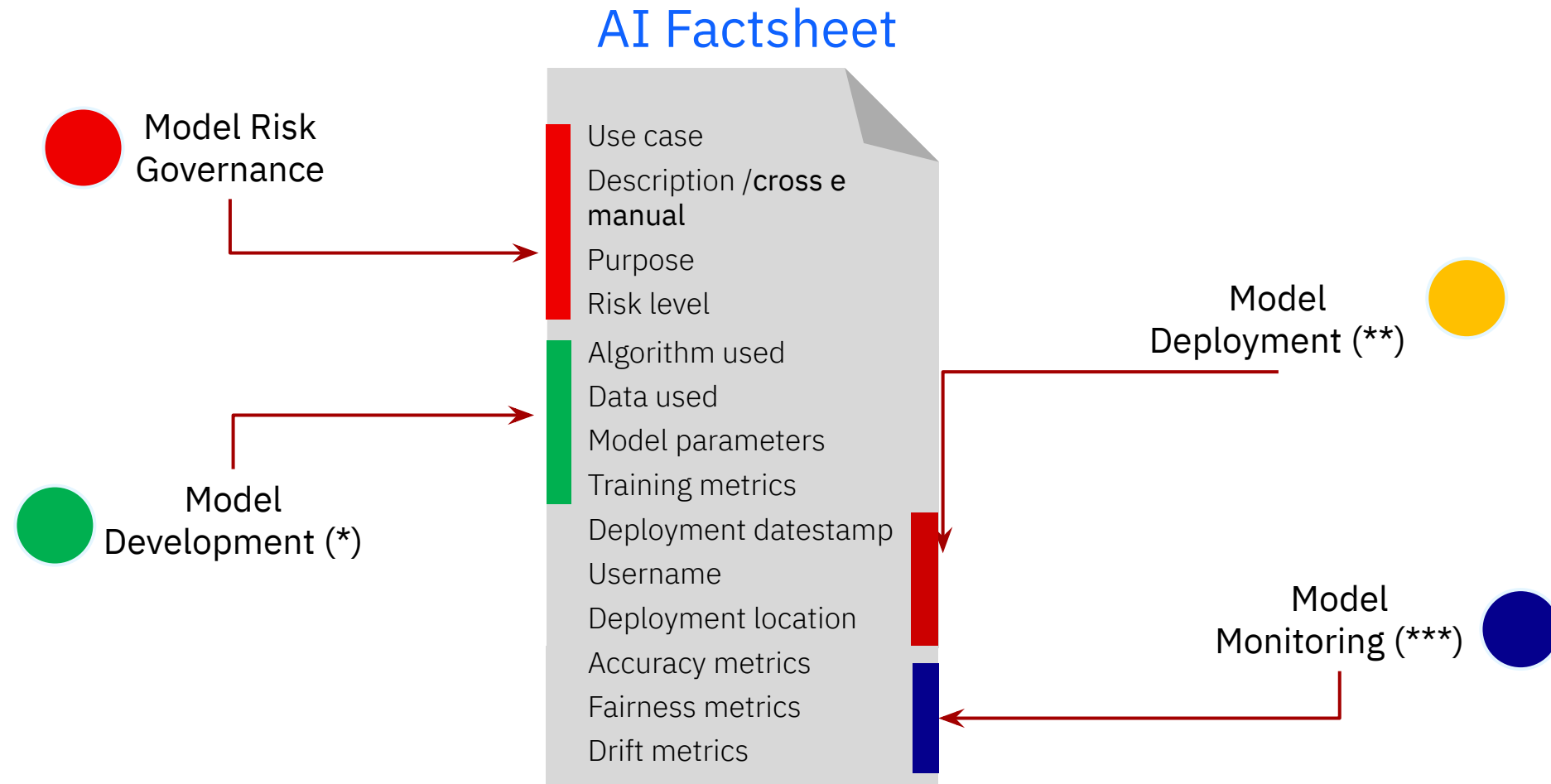


watsonx.governance



AI Documentation

Captures "facts" about an AI use case throughout its lifecycle, such as:



(*) Repeat for each model in the use case

(**) Repeat for each deployment of each model

(***) Repeat for each run for each deployment

AI Evaluation and Monitoring

Performance Monitoring

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

The image shows two screenshots of the IBM Watson OpenScale interface. The top screenshot is the 'Insights dashboard' which provides a high-level overview of model health. It features a summary table with the following data:

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

Below the summary table, there are filters for 'Tags', 'Alert type', and 'Machine learning provider'. A search bar asks 'Which deployment are you looking for?'. Three model cards are displayed, each showing a table of metrics:

Model	Quality	Fairness	Drift	Global explanation	Custom
GermanCreditRiskModelPrePr...	1	1	1	--	--
GermanCreditRiskModelChall...	1	1	--	--	--
GermanCreditRiskModel	1	--	1	--	--

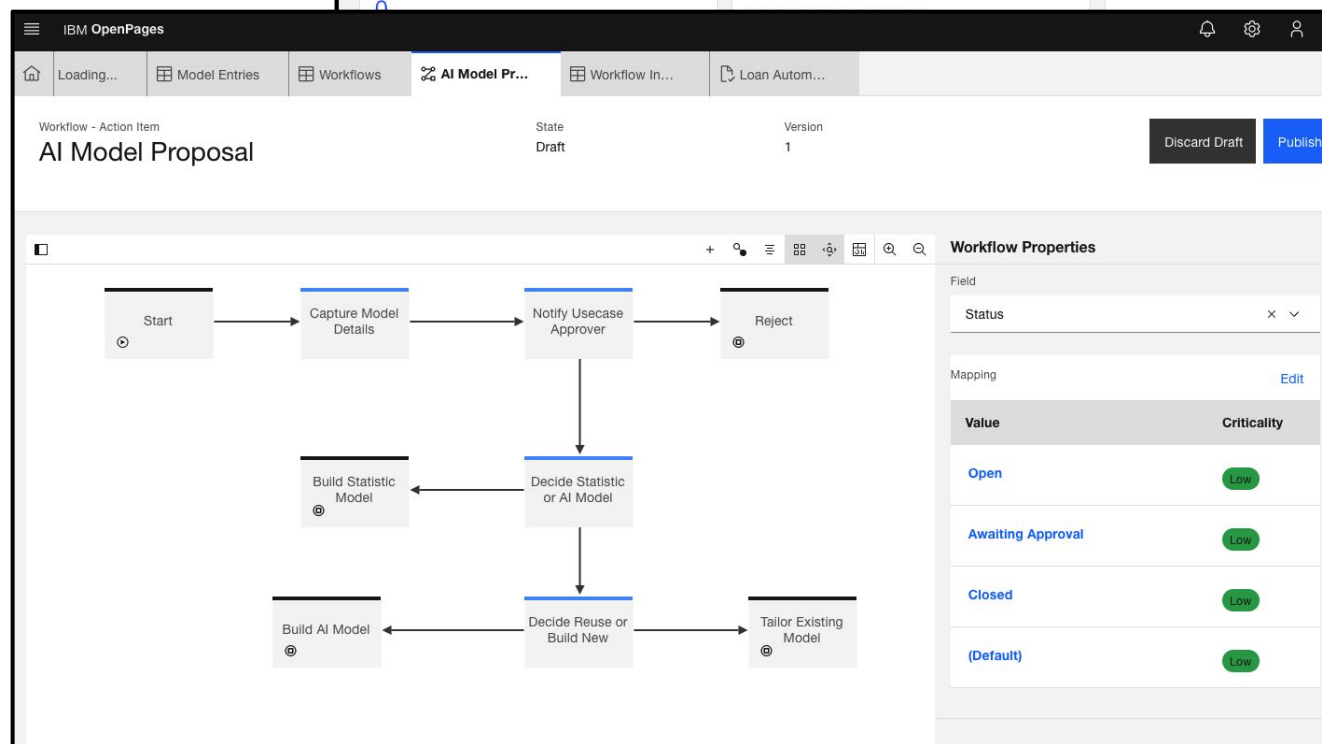
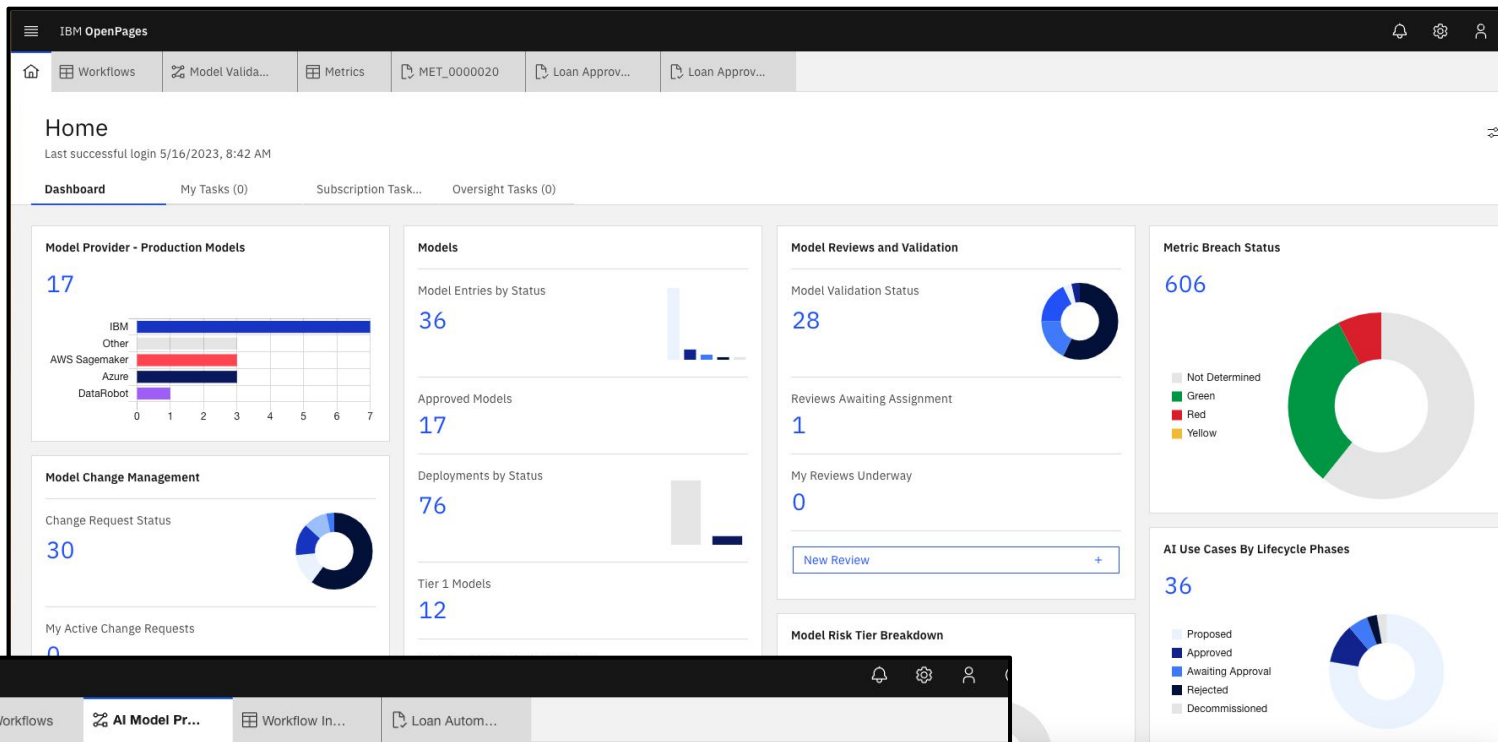
The bottom screenshot shows a detailed view of a 'Credit Risk Evaluation' for a 'Pre-production' model. It includes a circular gauge showing '3 tests run' and a table of performance metrics:

Metric	Value	Status
Fairness	90%	Green - within threshold
Quality	.99	Green - within threshold

Additional metrics include: Area under ROC (.99), Area under PR (.99), Accuracy (.95), True positive rate (TPR) (.89), False positive rate (FPR) (.15), and Recall (.89). A 'Compare model' panel on the right allows for comparison between 'Credit Risk' and 'Credit Risk V2' across various metrics, all of which are shown as '99'.

AI Risk Governance

- Consolidated view of models from multiple platforms
- View development status, model performance and alerts or emerging issues
- Monitor and trigger workflows for model validation, retraining and performance issues



Red Hat
Summit

Connect

Q&A



Red Hat
Summit

Connect

Thank you

