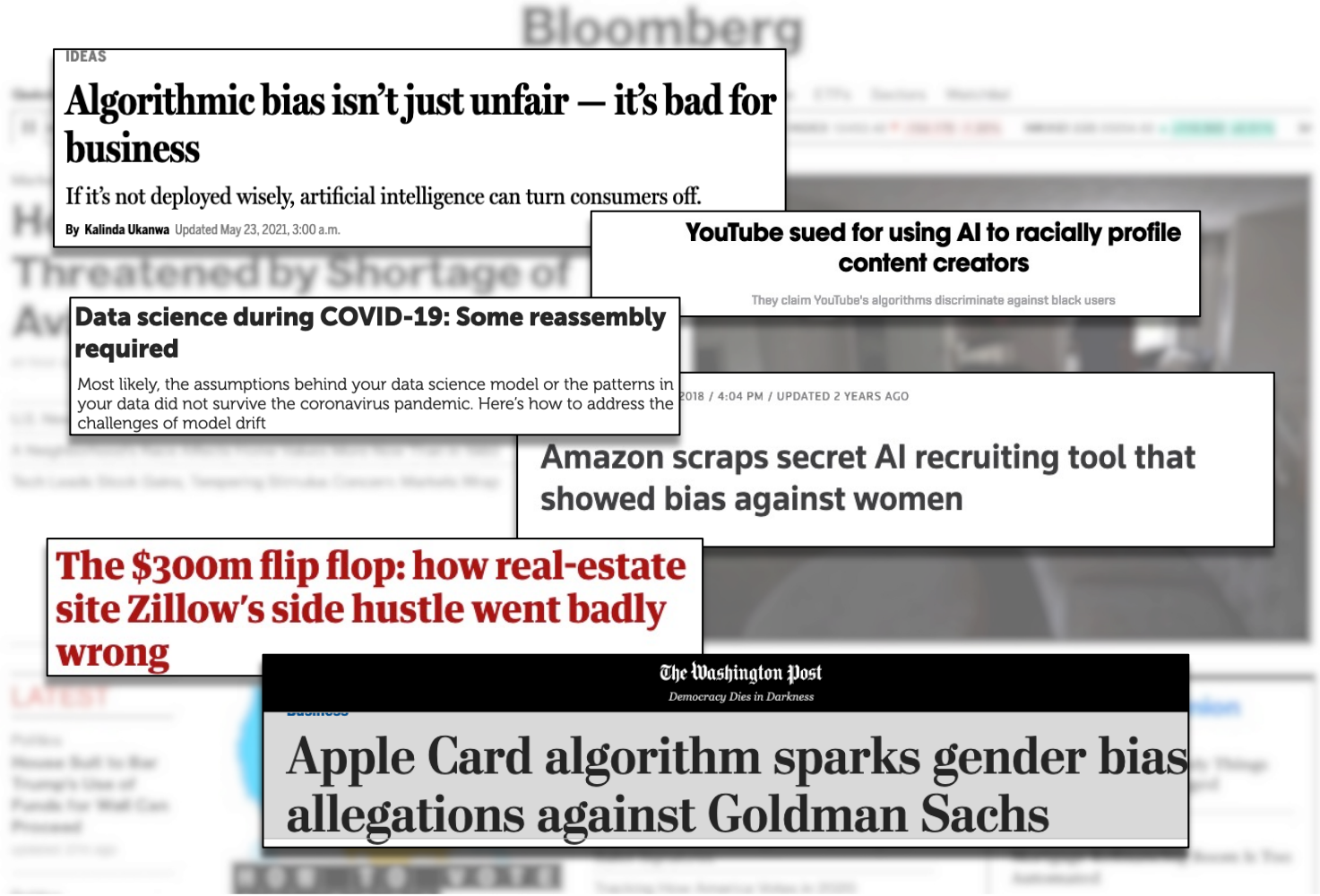
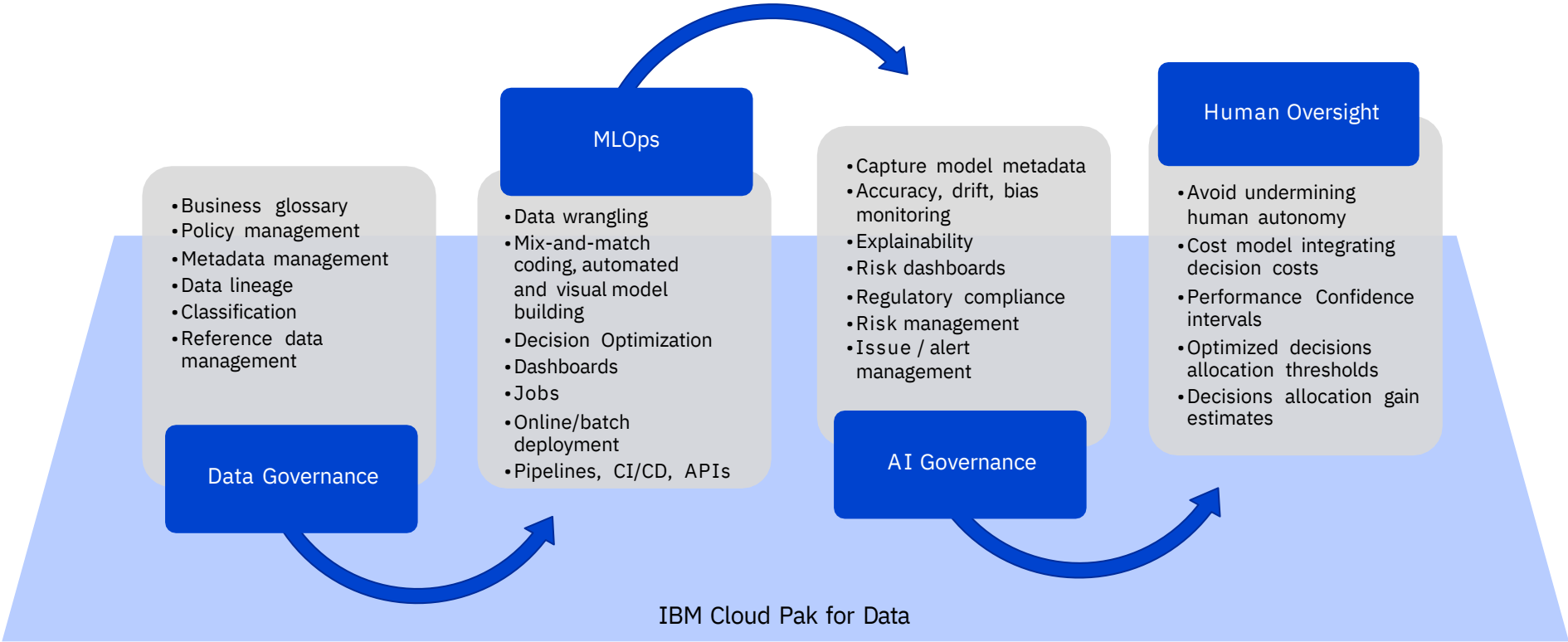
Árnyék a gépezetben: mit kezdjünk a mesterséges intelligencia fekete dobozával?

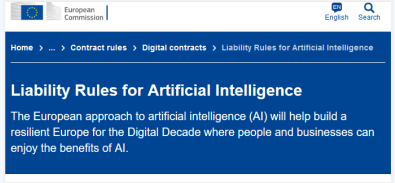
Huszti Dániel
daniel.huszti1@ibm.com
+36 20 823 5737



Putting AI Governance in perspective



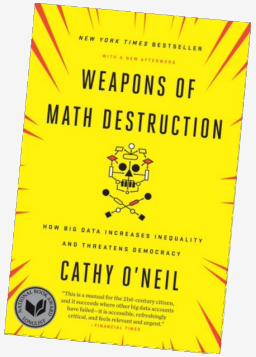
Regulation and reputation are the driving factors behind AI governance



Nearly all (97%) respondents believe that regulation will impact them to some extent and 95% believe that at least part of their business will be affected by the EU regulations specifically.

25% have yet to establish any meaningful Responsible AI capabilities.

Accenture - From AI compliance to competitive advantage, 2022



"Fewer than 20% of executives strongly agree that their organizations' practices and actions on AI ethics match (or exceed) their stated principles and values."

IBM and Oxford Economics – AI ethics in action, 2021

AI Governance solution

IBM solution is built on 3 pillars to meet clients on their maturity curve

Lifecycle governance	Monitor, catalog, and govern AI models from anywhere, throughout the AI lifecycle
Risk management	Manage risk and compliance to business standards, through automated facts and workflow management
Regulatory compliance	Ensure clients adhere to external AI regulations for audit and compliance

IBM's differentiation

Comprehensive: only vendor to support all three layers of AI governance

Automated: automated facts collection and lineage tracking within python notebooks

Open: AI vendor-neutral, supports governance of models built and deployed in third party tools

Unified: unified cataloging for models & data and unified notions of data quality for AI and reporting

Regulations-driven: support for AI regulations natively

Active policy enforcement: support for policies and rules to automate regulatory compliance

In practice, this means that...

Business stakeholders need to ...
(C-Suite, CPO/CRO, Business Owner)

- ✓ Define new responsibilities, policies, guidelines and processes based on AI principles & enterprise values
- ✓ Establish organizational structures
- ✓ Understand regulation and translate to business & technical requirements
- ✓ Raise awareness and train leaders and practitioners
- ✓ Assess maturity, risk level and conformity
- ✓ Oversee, manage and mitigate risks
- ✓ Design architecture, pipelines and governance rules for trust and scale

Technical stakeholders need to ...
(Data Scientist, Ops Lead, Risk Manager)

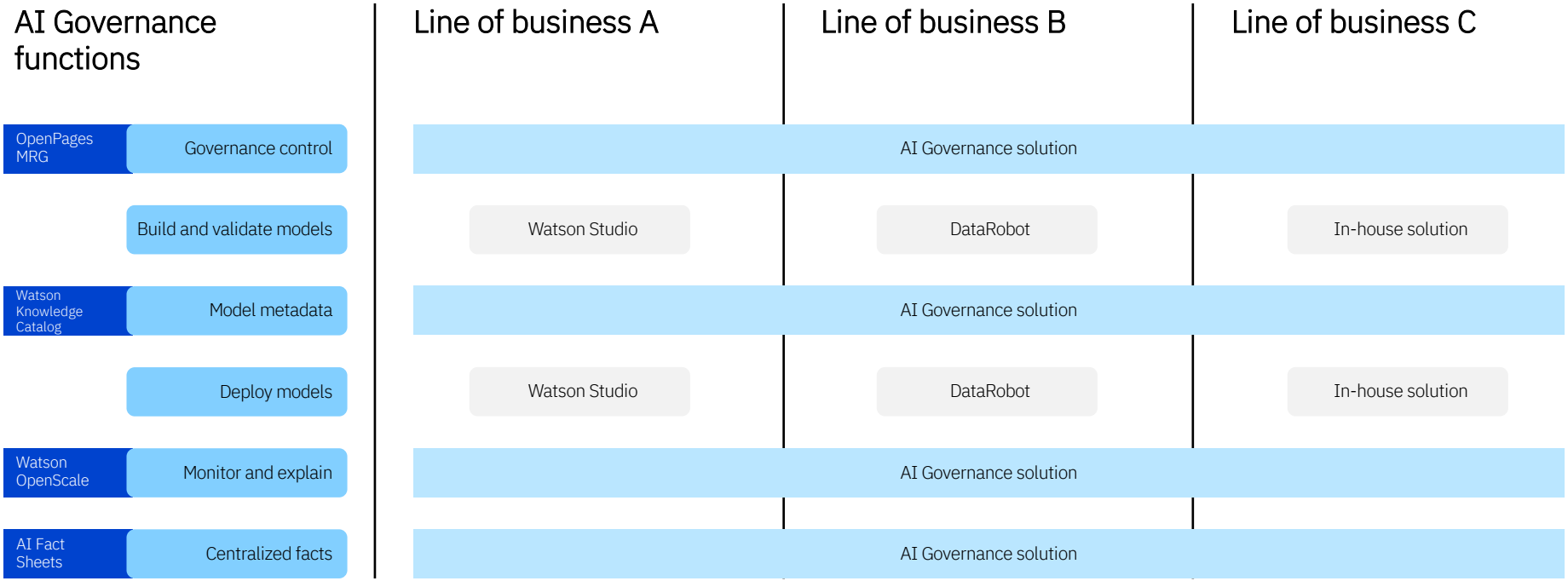
- ✓ Adhere to data governance rules
- ✓ Keep track of security risks
- ✓ Document every new and updated model
- ✓ Find and fix direct and indirect bias in each data set and model
- ✓ Find and fix model accuracy drift & data consistency drift in each model
- ✓ Explain model decisions as requested
- ✓ Find and fix adversarial attacks on your data and models
- ✓ Complete all model review process activities

Which is easier said than done

Common challenges	Optimization approach
AI governance often requires lots of manual work; amplified by changes in data and model versions.	Automate the governance activities as much as possible.
Companies have models in multiple tools, applications and platform, developed inside and outside the organization	Consolidate in one governance platform for consistency and efficiency.
It's not a one-size-fits-all approach.	Configure to your specific situation.
Current practices and tools not optimized for AI governance	Enhance or replace suboptimal tooling.

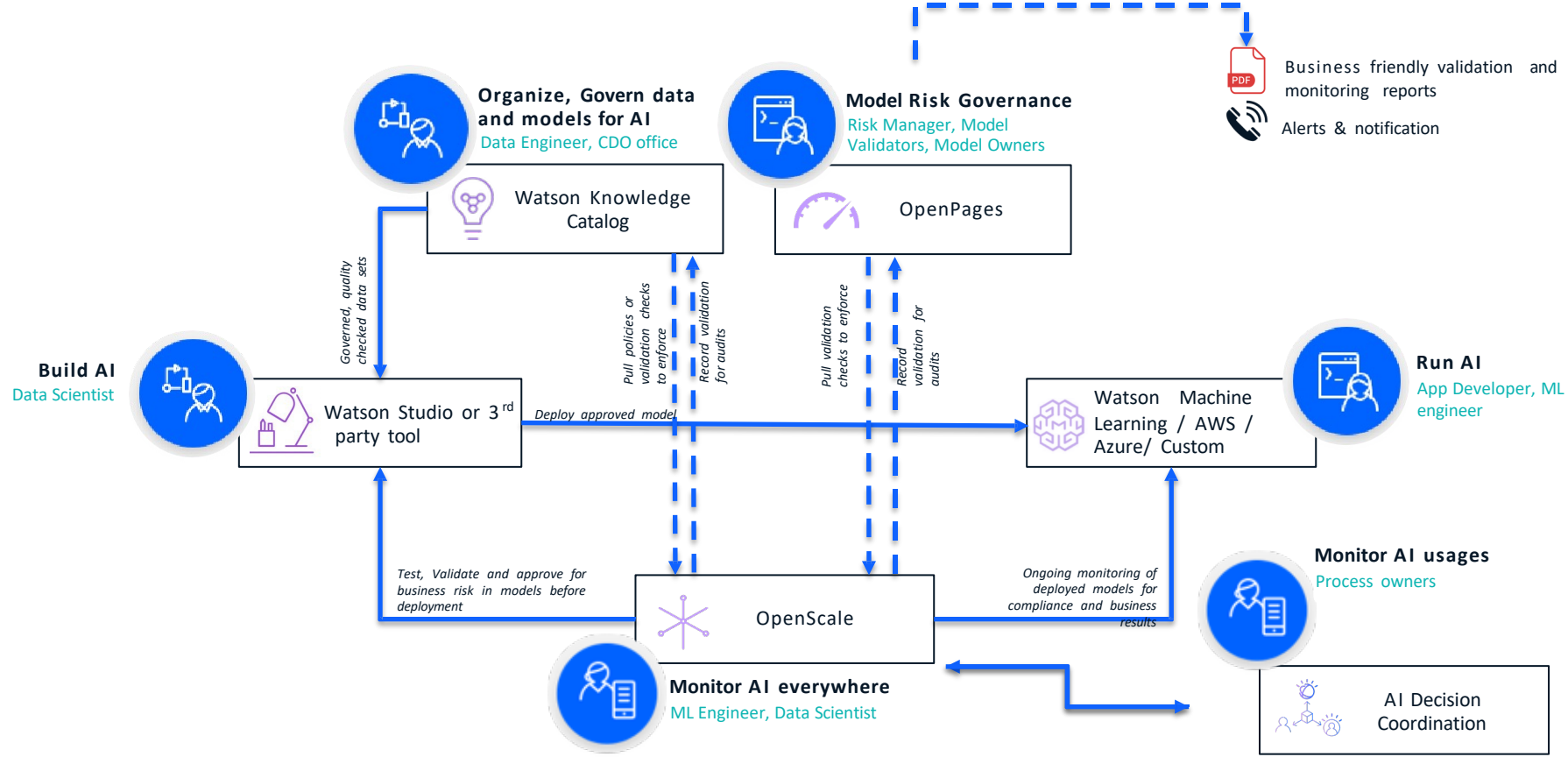
Putting AI Governance to work

A typical customer context



Technology View – AI Governance

Cloud Pak for Data components in scope for Trusted AI



Bias detection

Continuous calculation of model fairness

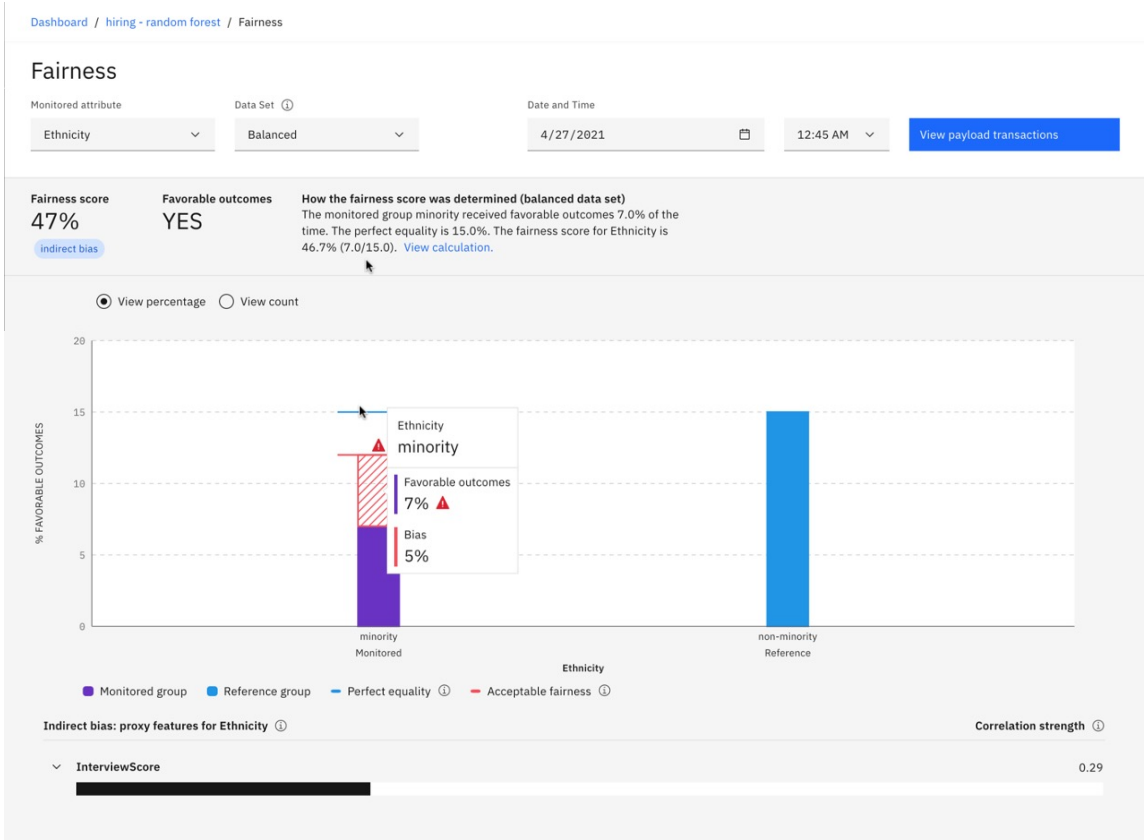
– Analyze deployed model predictions for bias

– Collect and aggregate bias data for dashboards and alerts

– Find non-feature data correlations

– Use a corrected model for “de-biased” predictions

Ensuring fairness in model scoring



Model Explainability

Explain model predictions

– Show the most influential features

– Explain in natural language

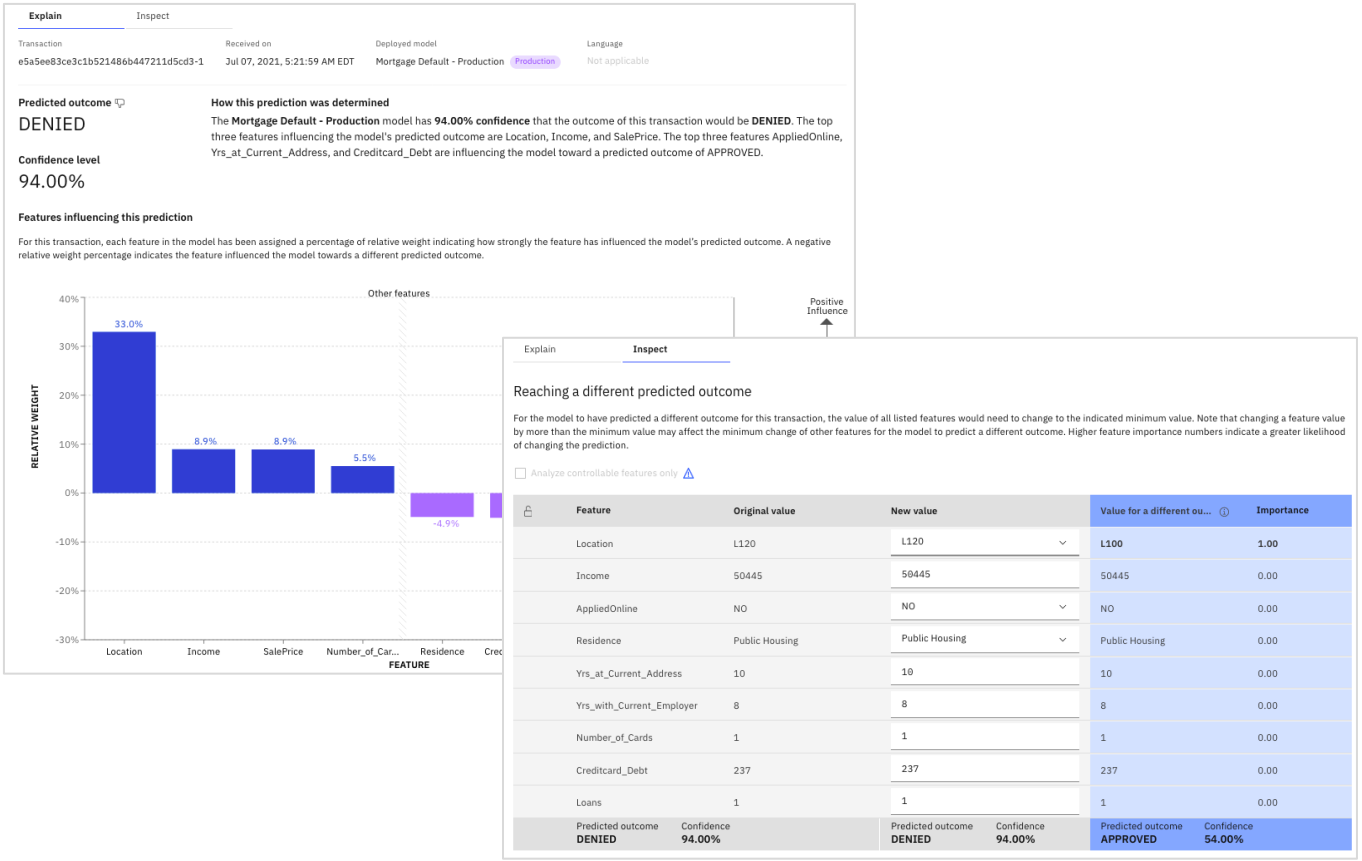
– Available API for prediction explanations

What-if analysis

– Experiment with values

– Assess effects of changes to features

Understand model outcomes



Drift detection

Measure the degree to which a model has moved away from reality

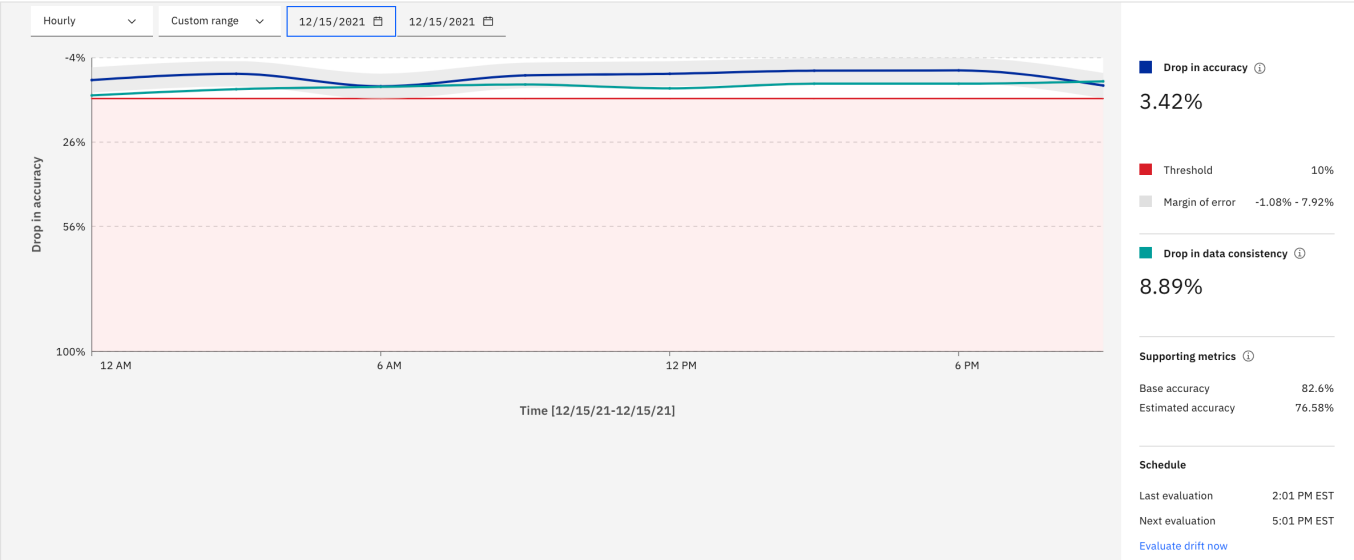
– Drop in accuracy – reality has changed, as shown by the scoring data

– Drop in consistency – reality is the same, the events vary

Drift monitoring and alerts

– Degradation of model performance can trigger retraining and redeployment

Handle changing scenarios

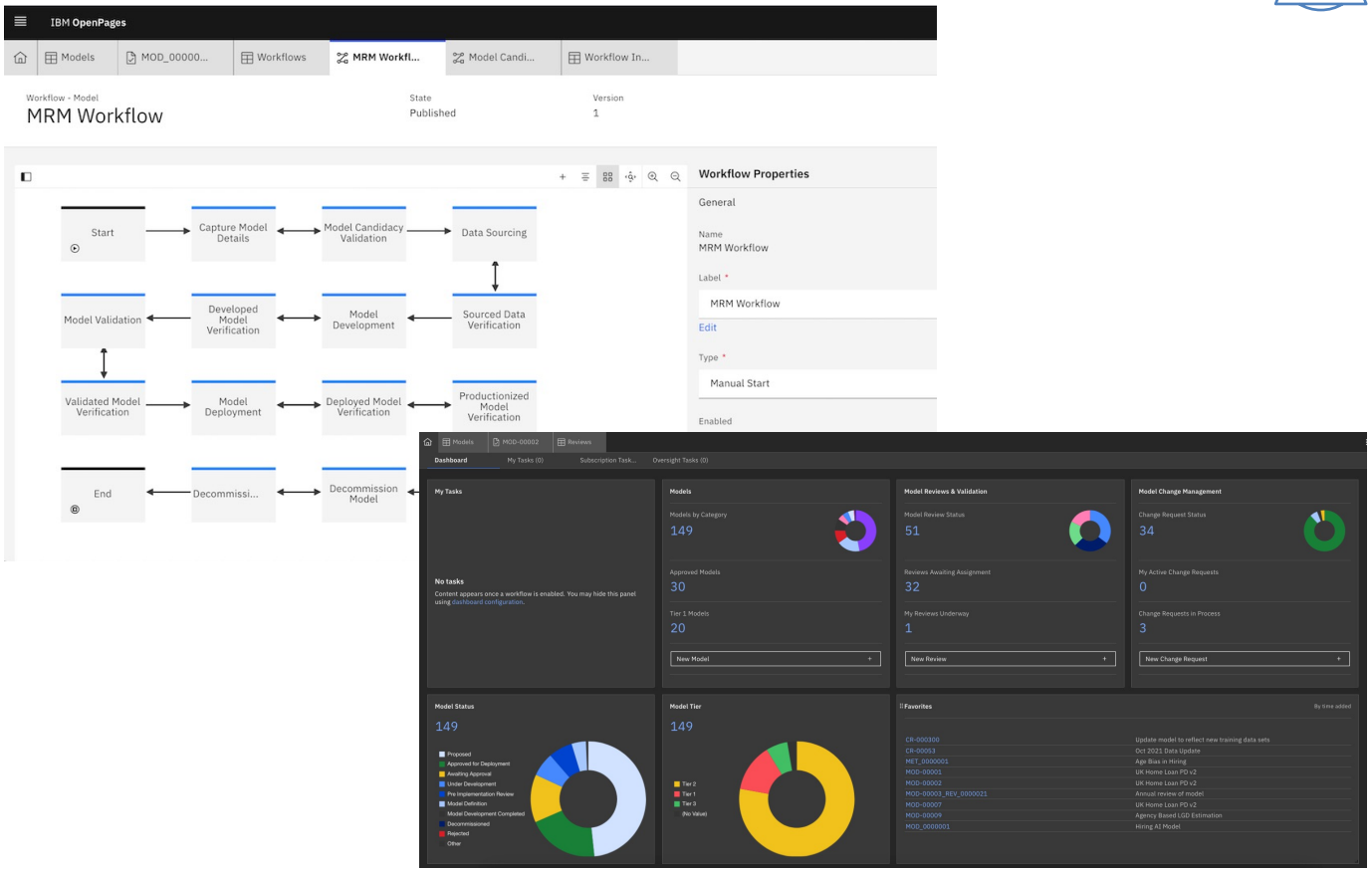


Manage risk across the enterprise – Open Pages

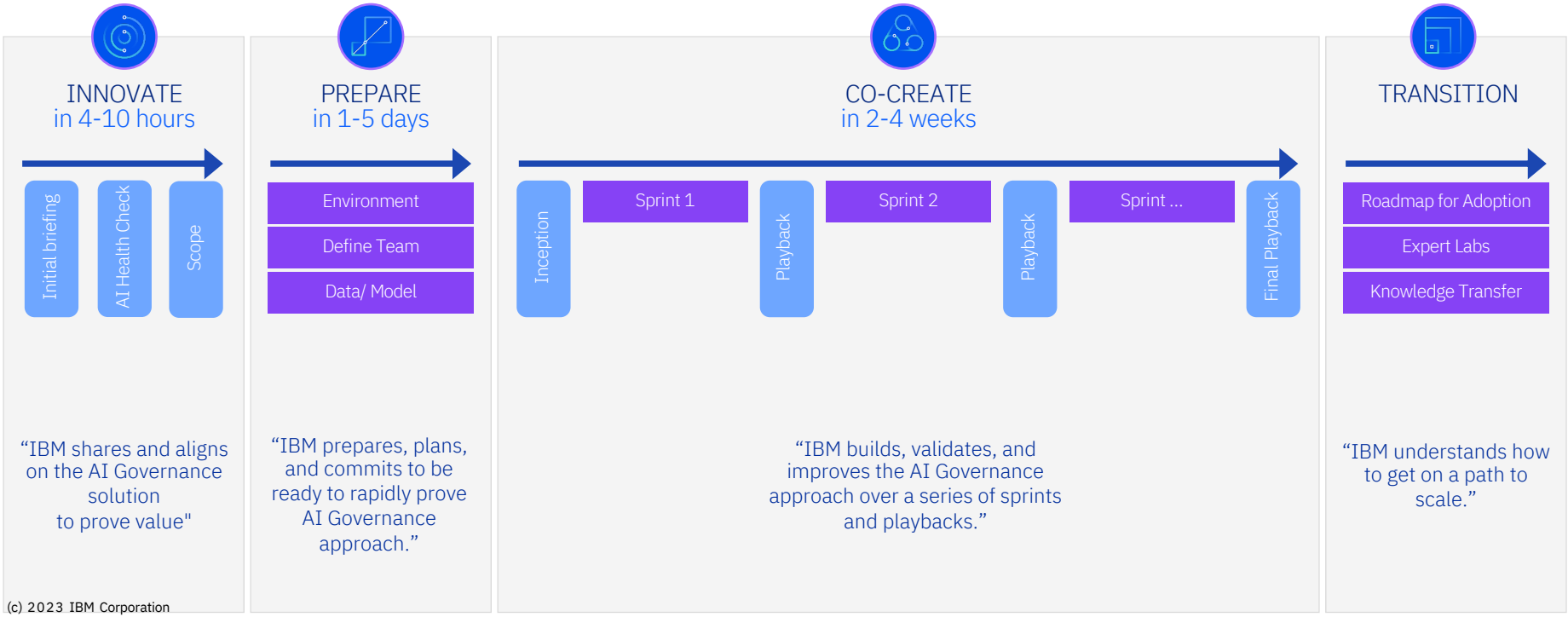
– Consistent holistic views of risk and compliance

– Drive GRC adoption

– Embedded self-service reporting, analytics, and dashboarding



How to get started?
Launch a pilot project on AI Governance



Recently announced at IBM Think 2023

Put AI to work with **watsonx**

Scale and accelerate the impact of AI with trusted data.