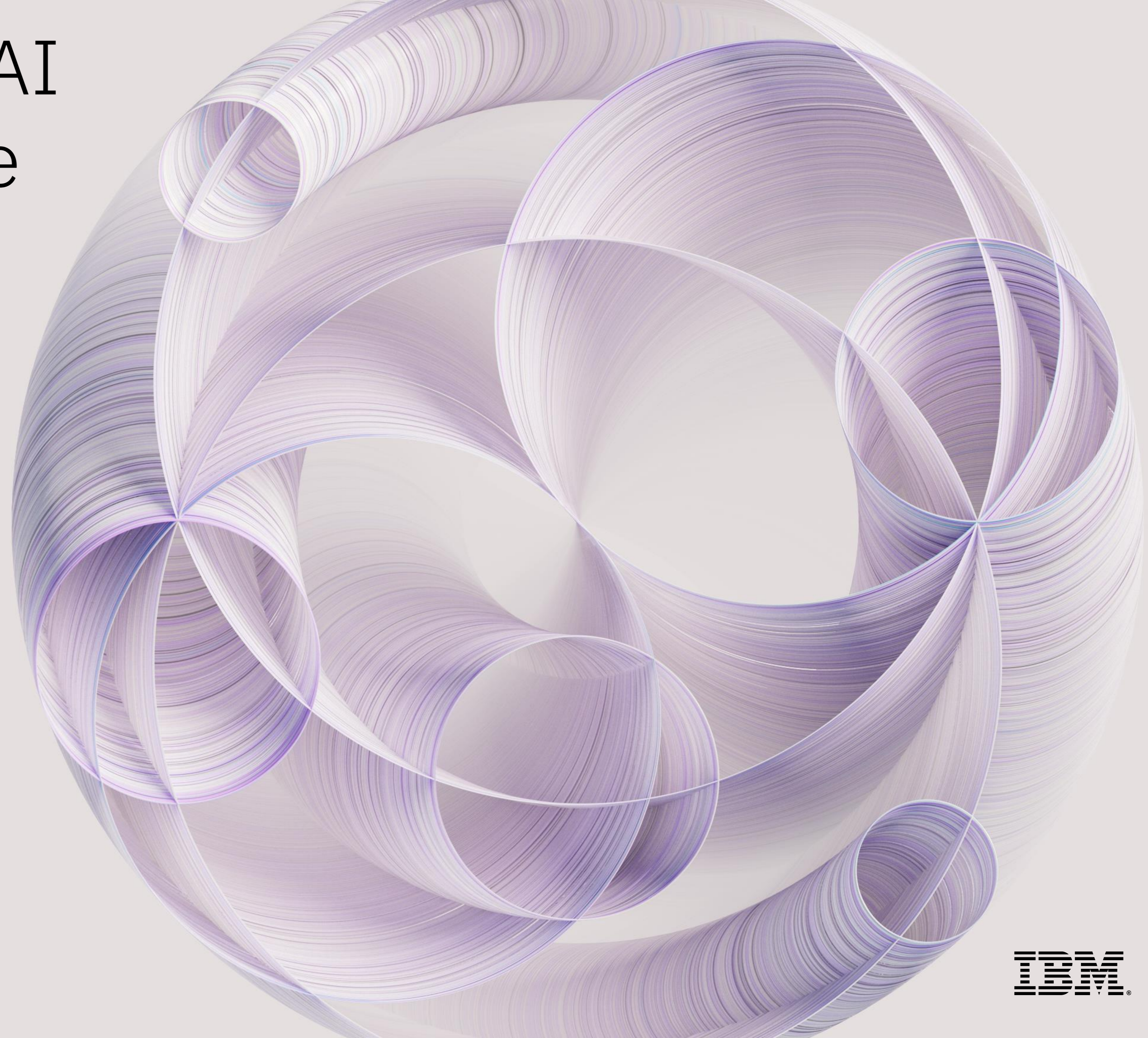


# Scale responsible AI with AI governance



# Today's talk

Scale responsible AI with  
watsonx.governance

# 1

Governance is needed to scale AI with trust, compliance and security  
Reduce risk, bias, model drift, profanity and hate speech

---

# 2

Watsonx.governance is the leading AI governance solution

- Govern any AI, anywhere, without sacrificing speed and performance
- Automate time consuming audit and documentation processes

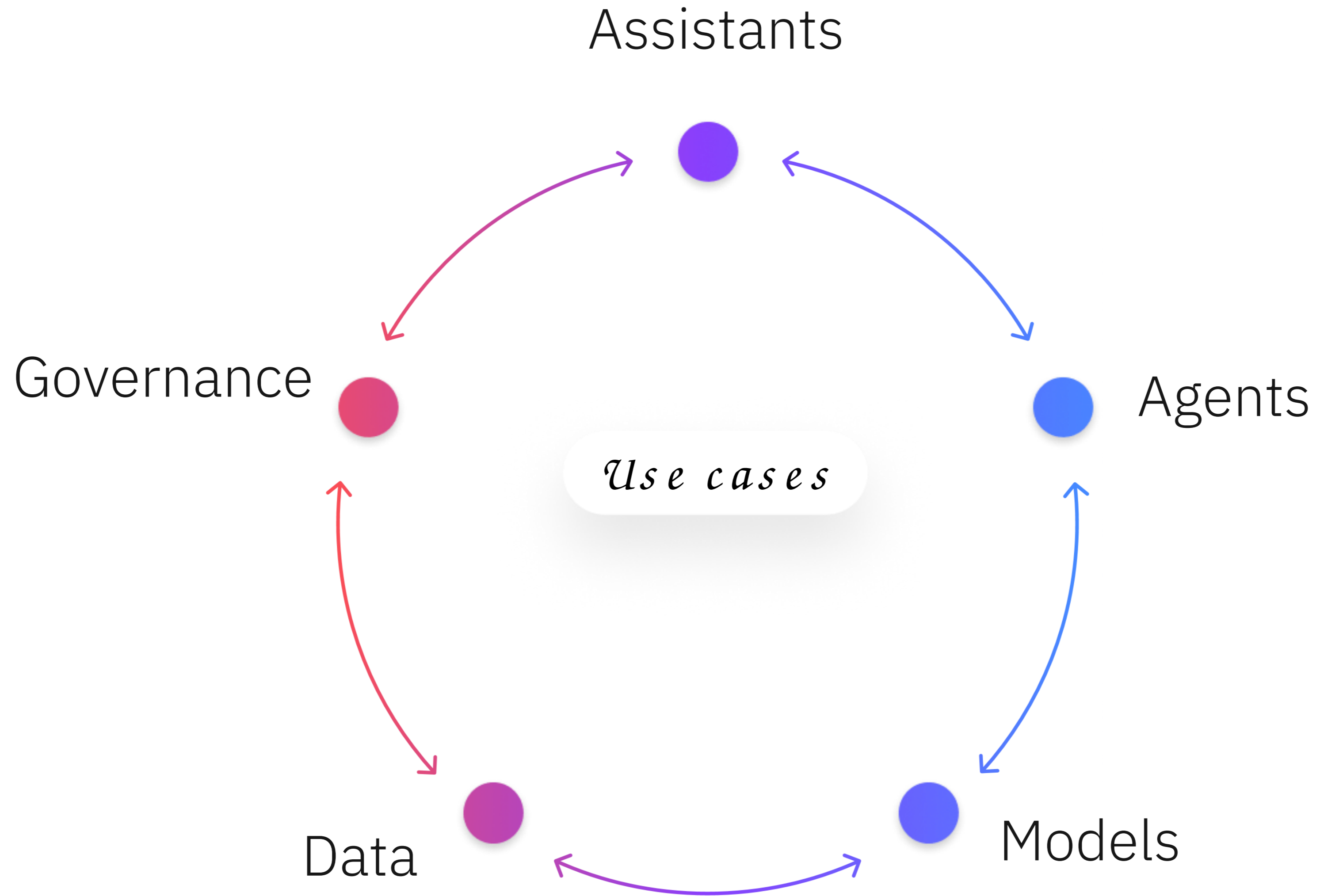
---

# 3

IBM's approach to data and AI

- Open, customized, multimodal and multi-model
- Both commercial and open-source innovation
- Customized to each business and use case to drive ROI

# AI building blocks to the future



The impact of generative AI is massive, yet so are the compliance, security and business risks

## Growing regulations

€35M

fine for noncompliance with EU AI Act, or 7% annual turnover, whichever is higher<sup>4</sup>

## New and amplified risks

73%

of IT leaders are concerned with biased outcomes, and 79% are concerned about security risks<sup>5</sup>

## Stakeholder complexity

<30%

of CROs and CFOs say regulatory and compliance risks are being sufficiently addressed in their organizations<sup>6</sup>

<sup>4</sup> [EU Artificial Intelligence Act](#), European Parliament and Council of the European Union, 13 June 2024.

<sup>5</sup> [Generative AI could raise global GDP by 7%](#), Goldman Sachs, 4 April 2023.

<sup>6</sup> [Risk Management – The CEO’s Guide to Generative AI](#), IBM Institute for Business Value, 12 August 2024.

# AI needs governance



The process of directing,  
monitoring and managing the  
AI activities of an organization  
through automation

# Your AI for business strategy can't succeed without AI governance



Changing regulations



Multiple stakeholders



Manual and error prone documentation



Increased risk

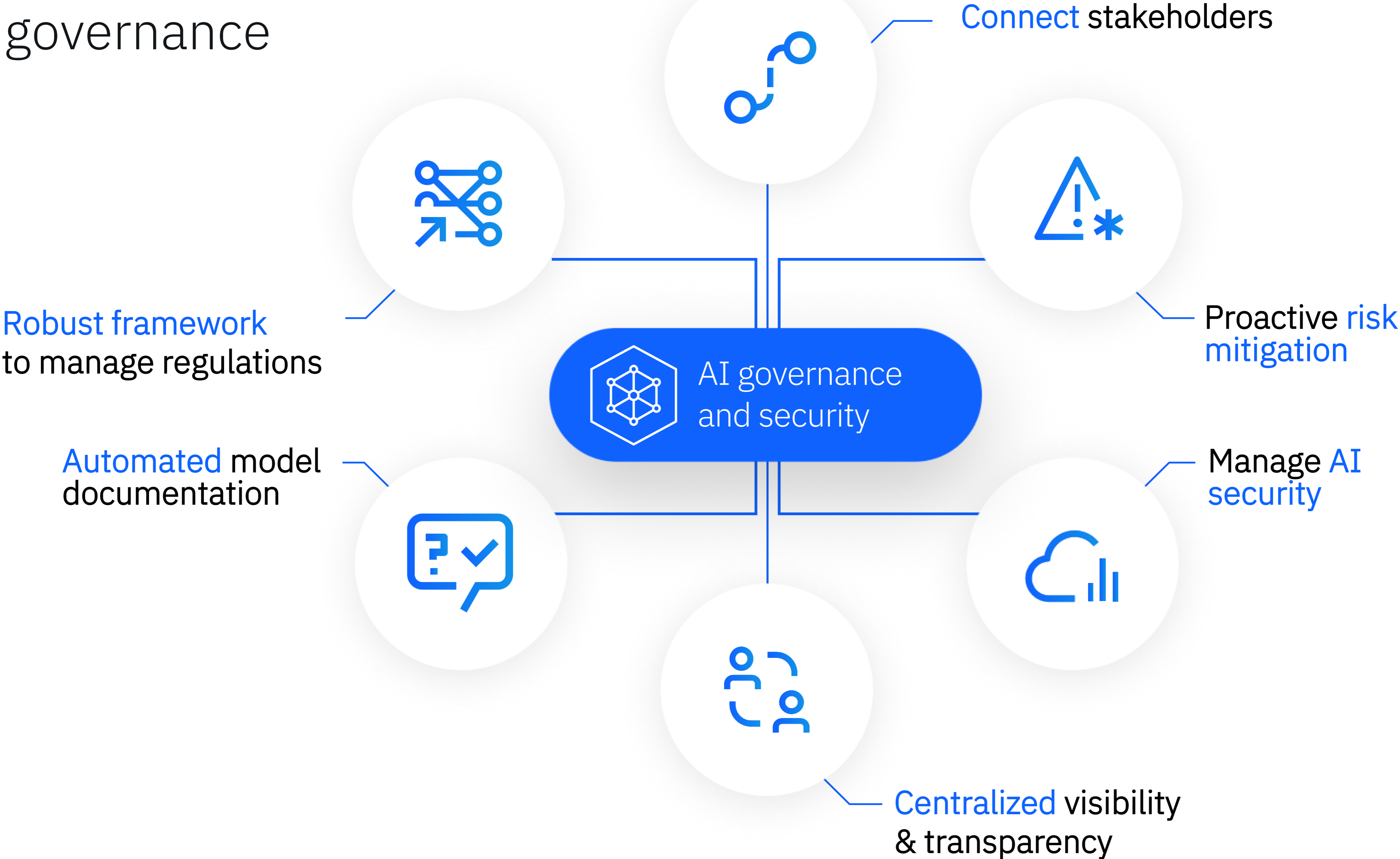


Disparate tools and data



Vulnerable data

# Manage the complexity of AI governance



Put AI to work with watsonx.

IBM watsonx is a portfolio of AI products that accelerates the impact of generative AI in core workflows to drive productivity.

# watsonx

A portfolio of AI products that accelerates the impact of generative AI in core workflows to drive productivity.

## watsonx.ai

Enterprise-grade AI studio that helps AI builders innovate with all the APIs, tools, models, and runtimes to build AI solutions

Featuring **IBM Granite**, and popular third-party models including **Mixtral**, **Llama** series

## watsonx.data

The **hybrid, open data lakehouse** to power AI and analytics with all your data, anywhere

## watsonx.governance

End-to-end toolkit for AI governance to manage **risk and compliance** across the entire AI lifecycle.

## watsonx Orchestrate

An enterprise-ready solution that helps create, deploy, and manage AI assistants and agents to automate processes and workflows.

## watsonx Code Assistant

Accelerate development, **application modernization**, and **assist with IT Operations**

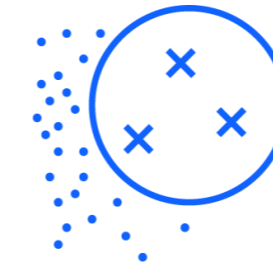
# IBM watsonx.governance

Accelerate responsible,  
transparent and explainable  
AI workflows



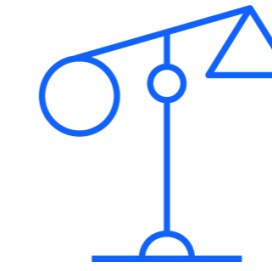
## Centralized AI lifecycle governance

Manage, monitor and govern  
any AI: model, app or agent;  
across IBM and 3<sup>rd</sup> party like  
OpenAI, AWS, and Meta



## Proactive AI risk and security management

Proactively detect and  
mitigate AI risks, evaluate  
AI assets, and secure AI  
deployments with Guardium  
AI security




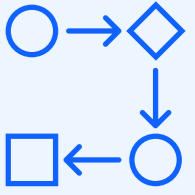



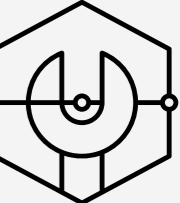
## Trustworthy and dynamic compliance

Manage AI for safety  
and transparency with  
our regulatory library,  
automation and  
industry standards

Platform agnostic: Govern any AI, deployed anywhere



# watsonx.governance at a glance

	Lifecycle governance	Risk management	Regulatory compliance
 <ul style="list-style-type: none"> <li>- AI use case owners</li> <li>- Data Scientist / AI Engineers</li> <li>- Model validators</li> <li>- Audit teams</li> <li>- Compliance teams</li> <li>- Risk management teams</li> <li>- AI Security teams</li> </ul>	<p><a href="#">AI risk governance and security</a></p> <p>Model inventory   Risk assessments   Workflows   Dashboards   Issue management</p>		
	<p><a href="#">AI Factsheet</a></p> <p>Capture model facts throughout the lifecycle</p>		
	<p><a href="#">AI Observability and Guardrails</a></p> <p>Model health   Harmful content detection   Accuracy   Drift   Bias   Explainability</p>		
 <ul style="list-style-type: none"> <li>- Data engineers</li> <li>- AI Engineers</li> <li>- (Citizen) data scientists</li> <li>- MLOps</li> <li>- ML engineers</li> </ul>	<p>Build and Deploy</p> <p>IBM watsonx.ai   AWS   MS Azure   GenAI apps   SaaS solutions   Other</p>		

# AI lifecycle governance



## Key Highlights

- Inventory and track ML models, GenAI apps and agents from concept/ideation through lifecycle
- Evaluate and assess your GenAI prompt templates, ML models, agents during build
- Automate the capture of the model, app and agent metadata to facilitate management and compliance

## Solves for:

1. Time-consuming documentation
2. Manual or non-systematic approach to evaluate prompts
3. Lack of transparency through lifecycle

# Bring transparency and visibility into your AI use cases

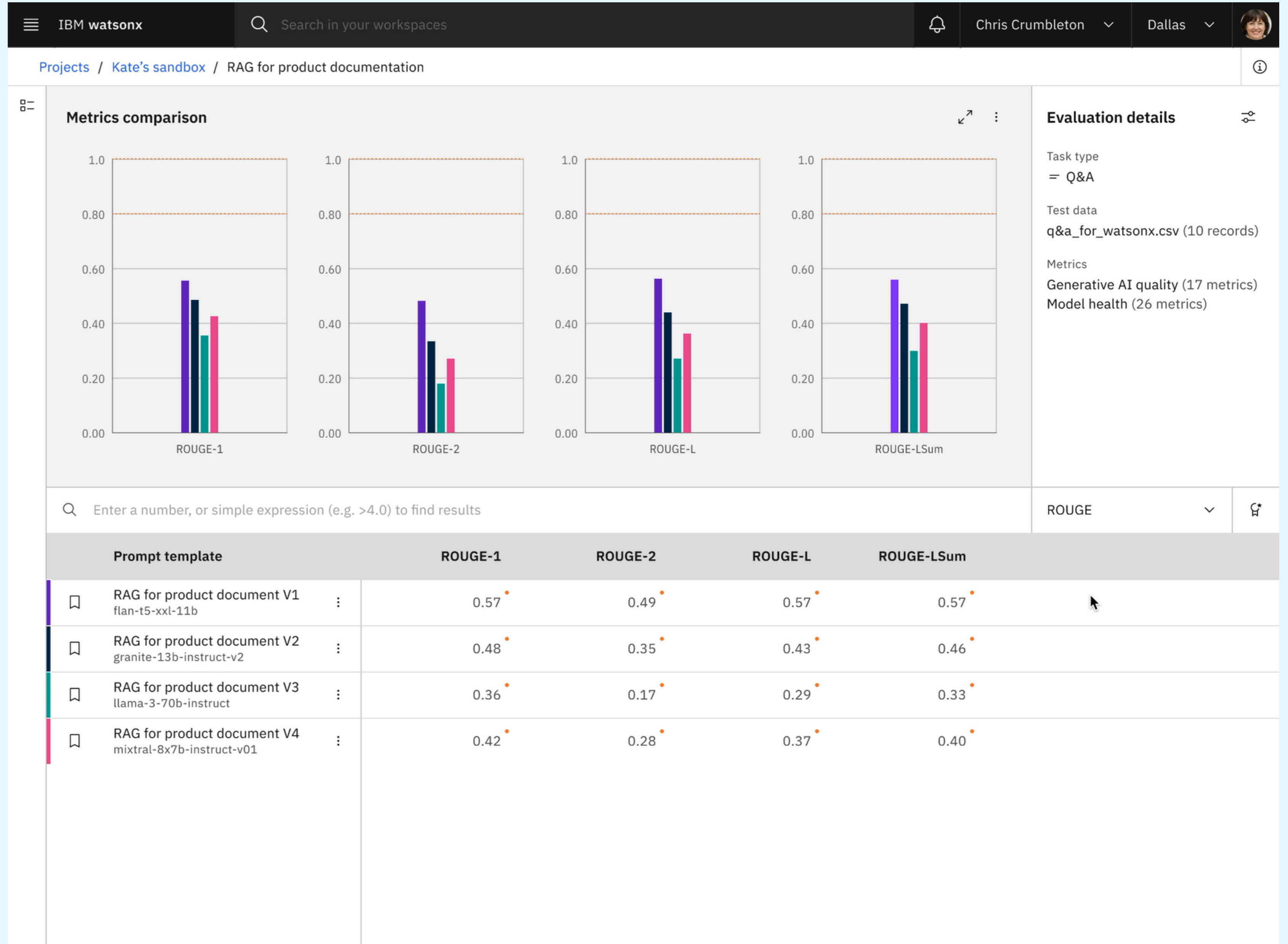
The screenshot shows the IBM Watsonx Governance console interface. At the top, it says 'Welcome, cpadmin!' with the last successful login on 12/17/2024 at 3:39 PM. The dashboard is divided into several sections:

- My Tasks:** Shows 21 tasks. A progress bar indicates 14 overdue, 1 due soon, 3 due in 2+ weeks, and 3 with no due date. A table lists the top 5 tasks by due date, including 'Job Applicant Scree...', 'IT ticket classification', 'Earnings Call Summ...', and 'Insurance Claim - A...'.
- AI Models by Provider:** A donut chart showing 153 models across providers like IBM, Amazon, Microsoft, Databricks, Hugging Face, OpenAI, Dataiku, DataRobot, Google, and Other.
- Use Cases by Lifecycle Phase:** A donut chart showing 213 use cases in various phases from Proposed to Validation Complete.
- Models by Class:** A donut chart showing 207 models categorized by class such as Foundation Model, Prompt-based, and Fine-tuned.
- Use Case Summary:** Includes a 'Use CaseRisk Breakdown' donut chart and a 'Use Case by Status' donut chart, both showing 213 items.
- Model Compliance Status:** A bar chart showing 207 models, with counts for Compliant, Non-compliant, and (No Value).
- Metric Breach Status:** A donut chart showing 2908 metrics across Red, Yellow, Green, Not Determined, and (No Value) categories.
- Useful Links:** A list of external links including Responsible AI Institute, EU AI Regulation, NYC Local Law 144, AI Bill of Rights, Artificial Intelligence and Data Act, SR 11-7 Information, and E-23 Information.
- My Favorites:** A section for user favorites, currently showing 'Elevator Q and A Assistant'.

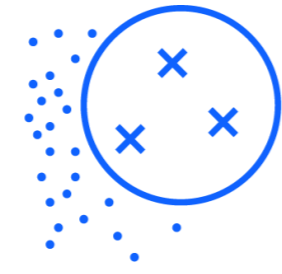
# watsonx.governance for centralized AI lifecycle governance



Example: Accelerate AI asset selection as you build with easy comparative evaluation on multiple quality metrics in Evaluation Studio



# Proactive and efficient risk management



## Key highlights

- Identify and manage AI risks early on with risk assessments at use case and model-level
- Mitigate risks and moderate content with Guardrails
- Observe and understand with continuous monitoring with alerts and explainability

## Solves for:

1. Limited visibility across AI use cases
2. Lack of risk management for AI
3. Identifying potentially harmful content in prompts

# Assess, identify, and manage AI risk

The screenshot shows the IBM watsonx interface for monitoring AI risk. The breadcrumb trail is: Deployments / High Oaks Bank / Credit Risk / Evaluations. The main navigation includes Fairness, Quality, Drift, **Drift 2.0**, Global explanation, and Model health. The 'Drift' section is active, displaying a time series chart for 'History: Nov 8, 2024' with the predicted class set to 'Risk'. The chart tracks four metrics: Output drift (purple), Model quality drift (blue), Feature drift (green), and Transaction volume (grey). On the left, three key metrics are highlighted: Output (0.6) at 0.9, Model quality (0.6) at 0.8, and Feature (0.6) at 0.3. Below the chart, two detailed views are provided: 'Output confidence distribution' showing a density plot of confidence levels for 'Risk' outcomes, comparing training data (purple) and runtime data (blue); and 'Significant intervals' showing a bar chart of confidence ranges (5-20, 20-30, 50-65, 70-85, 85-100) for 'Positive' outcomes, comparing training data (purple) and runtime data (blue).

watsonx.governance for  
proactive and efficient  
risk management 

Example: identify potential risks and mitigation for AI use cases early on with out-of-the-box AI risk assessments and applied AI risk taxonomy

Manage **risk and compliance** at scale

How IBM Guardium can help

## IBM Guardium AI Security integrates with watsonx.governance for Trusted AI

### Key highlights

- Discover AI models, data, and applications used by an organization across AI Services and enrich watsonx.governance use case inventory
- Automatically identify unregistered AI deployments and trigger the appropriate governance workflow
- Operationalize across product, risk, compliance, and security stakeholders

Govern, secure and monitor AI in one unified experience with Guardium AI Security and watsonx.governance integration

The screenshot displays the IBM Watsonx Governance console interface. At the top, the navigation bar includes 'Guardium AI Security' and 'Models'. The main content area shows a tree view with a legend for 'Primary Parent', 'Parent', and 'Child'. A node labeled 'Guardium AI Security' is connected to a '94 Models' node, which in turn connects to a list of various AI models. On the right side, there are panels for 'Tags' (indicating no tags have been added yet) and 'Business Entity' (with a prompt to review and update the business entity).

Discovered models grouped in same business entity

# Dynamic and trustworthy compliance



## Key highlights

- Ingest and represent internal and external AI regulations to present to use case owners and compliance officers
- Provide ability to record and assess AI Use case evidence for audit and compliance
- Use Factsheets for transparent model processes

## Solves for:

1. Changing regulations
2. Inaccurate documentation

## Meet growing AI regulatory landscape

The screenshot displays the IBM watsonx Governance interface. At the top, the navigation bar includes the IBM watsonx logo, a notification bell, and user information for Chris Crumbleton in Dallas. The breadcrumb trail shows the path: Deployments / High Oaks Bank / Credit Risk / Evaluations. The main content area is titled 'AI Factsheet' and features a sidebar with a navigation menu. The menu items include Governance, Foundation model, Prompt template, Prompt parameters, Evaluation (expanded), Develop (expanded), Finance (expanded), Test, Validate, Operate, Additional details, Attachments (expanded), Charts, and Files. The main panel shows the 'Governance' section for the 'Credit Risk Model' use case. It is currently in 'Draft' status with ID 'e98cf678-37bc-4ef7-827-02cf70186112'. The description states: 'Machine learning models developed to determine the risk level associated with a loan application.' There are tabs for 'Finance' and 'Credit Risk'. The 'Approach' is 'Default approach' (ID: 7c8c14b2-a25f-4e5a-a1ce-c97660ccd191) and the 'Version' is '0.0.1'. A 'Lifecycle' section shows three stages: '01 Develop' (active), '02 Validate', and '03 Operate'. Below this, the 'Foundation model' section lists 'ibm-granite/granite-3.1-8b-instruct'.

watsonx.governance for dynamic and trustworthy compliance



Example: assess regulatory applicability for your GenAI use cases

The screenshot displays the IBM Watsonx Governance console interface. At the top, the header shows 'IBM watsonx' and navigation icons. Below the header, a breadcrumb trail indicates the current page is 'AskHR Chatbot'. The main content area is divided into sections: 'General', 'Use Case Details', and 'Stakeholders and Approvals'. The 'General' section includes fields for Name, Owner, Description, Status, Use Case Type, Purpose, Risk Level, and Third Party Link. The 'Use Case Details' section includes 'Uses Foundation Models', 'Externally Facing', 'Proposed Solution', and 'Target Implementation Date'. The 'Stakeholders and Approvals' section is currently empty.

Use Case	Status	Risk Level
AskHR Chatbot	Awaiting Use Case Approval	

*Modified Required*		
<b>General</b>		
Name *	Description	
AskHR Chatbot	Provide HR policy and operational responses to IBM employees	
Owner	Status	
Ian Francis	Awaiting Use Case Approval	
Use Case Type	Risk Level	
Purpose	Third Party Link	
Provide HR policy and operational responses to IBM employees		
<b>Use Case Details</b>		
Uses Foundation Models	Externally Facing	Proposed Solution
Yes	No	AI Infused assistant adopting a RAG approach with multiple model selection
Target Implementation Date	Additional Details	
6/30/2024		
<b>Stakeholders and Approvals</b>		

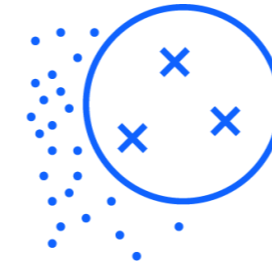
## Differentiation:

Accelerate responsible, transparent and explainable AI for both gen AI and ML models across any public or private cloud.



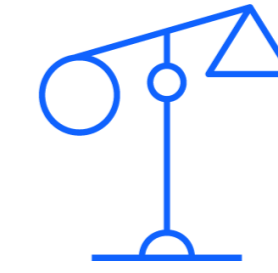
Govern any model, agent or AI app anywhere

Apply governance to ML and gen AI—open or closed—IBM and third parties (like OpenAI, AWS, Meta).



Assess and reduce AI risk at runtime

Continuous monitoring and recommendations with model risk assessment and real-time alerts.

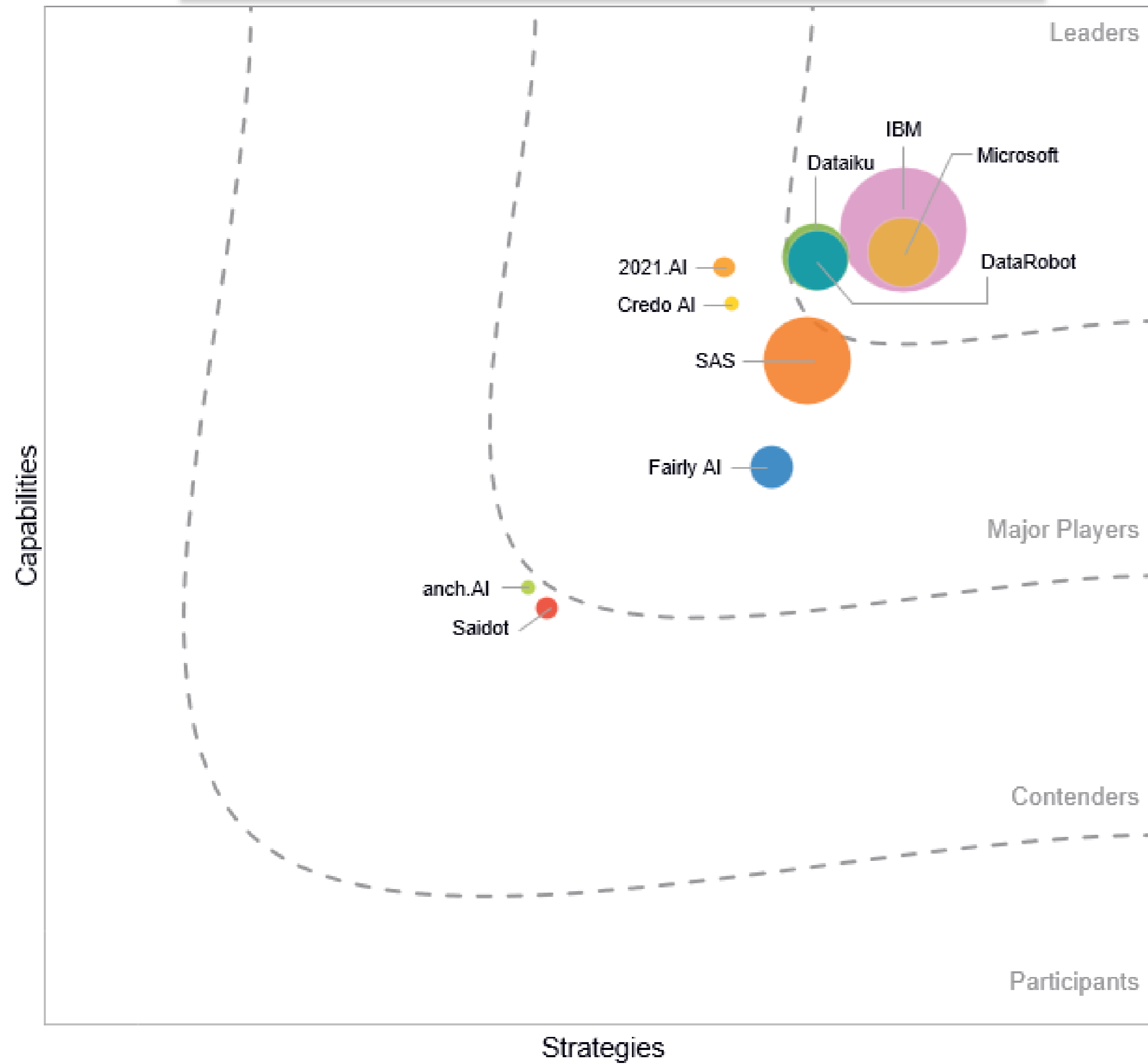


Worldwide compliance expertise

Compliance with internal policies, industry standards, and AI regulation, with automated audit processes.

# IBM is a leader in AI Governance and ML Ops

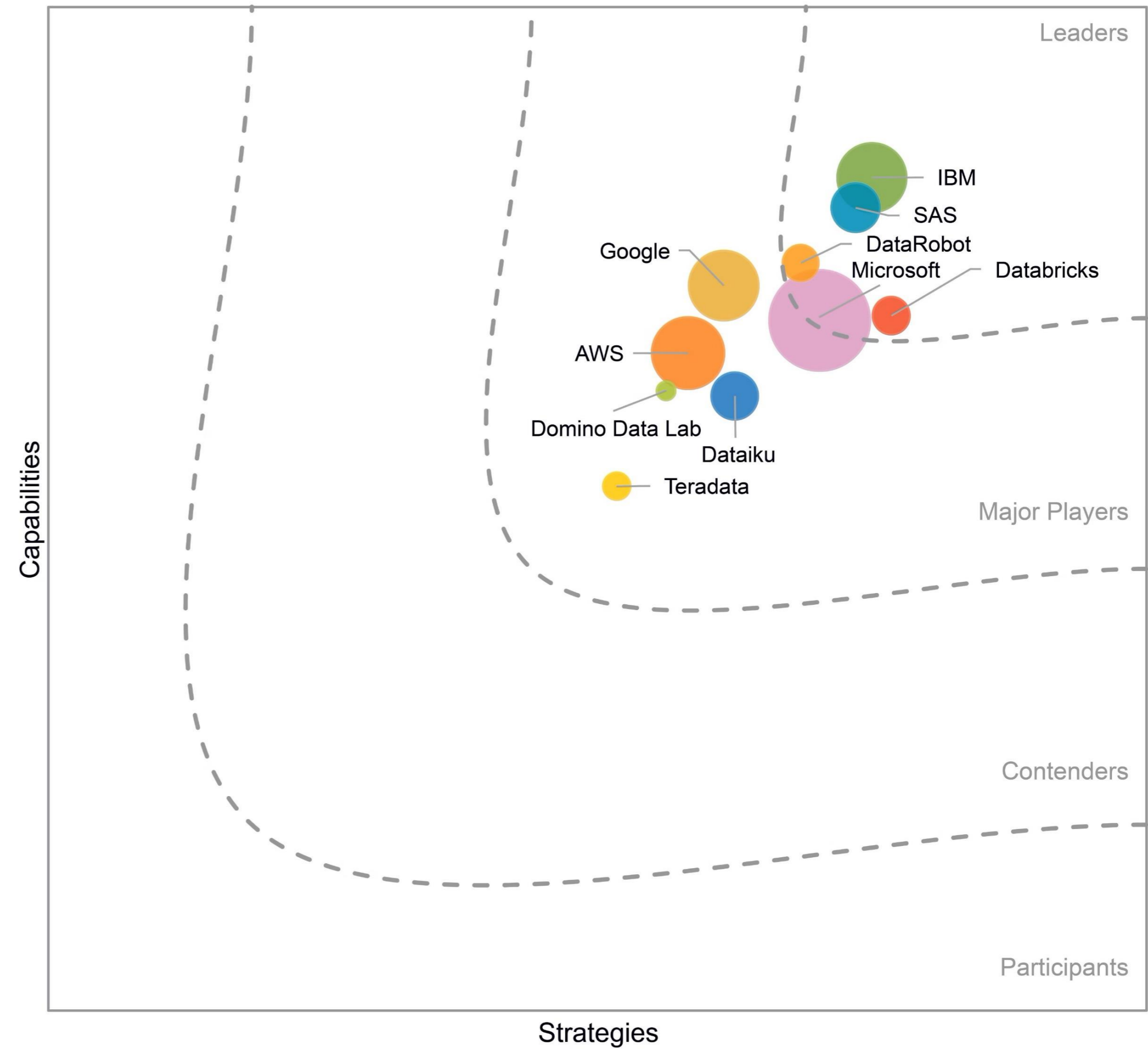
IDC MarketScape Worldwide AI Governance Platforms, 2023



Source: "IDC MarketScape: Worldwide AI Governance Platforms 2023 Vendor Assessment"

**Key strengths:** end-to-end solution and commitment to support

IDC MarketScape Worldwide Machine Learning Operations Platforms, 2024



Source: IDC, 2024

Source: "IDC MarketScape: Worldwide Machine Learning Operations Platforms 2024 Vendor Assessment" November 2024, IDC # US51573824

**Key strengths:** AI governance, no-code/low-code options and flexible deployment

# Client stories

## Scale generative AI ethically



Introduced watsonx.governance to apply debiasing to **increase court fairness from 71% to 82%**.



Embedded watsonx.governance to PepsiCo's advantage peeling solution **for end-to-end model lifecycle management**.



Integrated watsonx.governance with **Infosys Topaz, an AI-first offering**, to build responsible AI workflows.



Leveraging watsonx.governance, Deloitte is actively engaging customers to adopt a **responsible AI** approach to their generative AI initiatives.



Meta data-capturing capabilities from watsonx.governance modernized critical banking functions, saving **months of manual audit process work**.



Using AI governance tools from watsonx.governance streamlined collaboration, **reducing manual effort**.

Use case: IT modernization and governance

## IBM CIO: Building AI solutions with trust, transparency and speed

### Challenges

- AI development is accelerating
- Delivering quality, trustworthy and explainable AI solutions requires validation of accuracy, testing for undesired behaviors and generating fact sheets to provide transparency
- These cause increased work for development teams and extends the timeline to deliver solutions

Operating AI safely requires ongoing monitoring of model health, data safety and integration with enterprise IT operations.

### Solution

By incorporating watsonx.governance into the internal AI platform, we've enabled AI developers to take advantage of AI prompt evaluation, fact sheet generation and operational monitoring capabilities.

When solutions are delivered to the business, ongoing automated monitoring provides confidence to stakeholders and lays a foundation for robust testing and more trusted and secure AI deployments.

# Days to minutes

Projected reduction in time to evaluate AI solutions during the build stage

---

# 2x

Increase in gen AI metrics evaluated during solution validation

---

# 14

Model health, drift and safety metrics continuously monitored in production

---

# Compliance readiness

Aligned with AI governance requirements from the IBM Office of Privacy and Responsible Technology

# IBM's AI Governance

Vision and solution  
summary

1

AI needs governance to ensure trust and business success

---

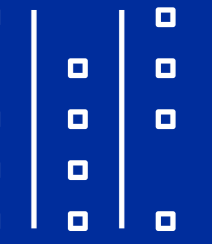
2

Governance is needed to scale responsibly

---

3

The leading AI governance solution is [watsonx.governance](#)



## AI Productivity

Reinvent how work is done with  
AI agents/assistants

- AI Assistants

## AI/ML Ops

Work with AI models, tools and  
governance that's built for  
business—engineered to ensure  
trust and scalability in  
applications

- AI Models
- AI Tools
- AI Governance

## Data Fabric

Bring all your business data  
together and optimize how it  
moves through your systems to  
scale analytics and AI in your  
applications while protecting it.

- Databases
- Data Intelligence
- Data Integration
- Data Security

## Data Storage

Store data across edge,  
core and clouds

- Software-defined Storage

# IBM Software

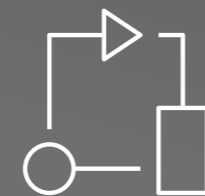
## Hybrid Cloud

Unify on-prem, public, private clouds and edge to scale virtualization and AI across environments



## Transaction Processing

Deliver unmatched transactional performance, security and reliability



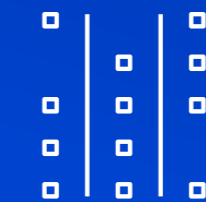
## Automation

Automate technology lifecycle management with AI for productivity, resiliency and spend optimization



## Data

Access trusted and secure data to drive AI productivity

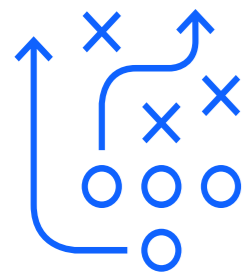


AI

*Open and Trusted*

# Three ways to get started with watsonx today

Looking forward to today's discussions



## Free trial

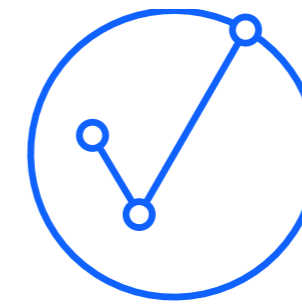
Core watsonx features to start building AI models and accessing data across your organization.



## Request a client briefing or demo

Discussion and custom demonstration of IBM's generative AI watsonx point of view and capabilities. Understand where generative AI can be leveraged now for impact in your business.

2-4 hours onsite or virtual



## 5-year business value assessment

Engagement with an IBM multi-disciplinary team to jointly innovate and rapidly prove the business value of generative AI solutions using watsonx.

1-4 weeks

**IBM**

# Why IBM watsonx for scaling enterprise AI to drive productivity

## Open

---

→ Offers choice to train the right foundation models, including open-source models, and the choice of data, tools, and frameworks to achieve desired business outcomes.

→ Run AI wherever the business needs to, across any cloud, at scale.

## Trusted

---

→ Built with open and transparent technology to give enterprises confidence in their AI and meet regulatory compliance demands.

→ Responsible AI and protected data backed by enterprise governance and security controls.

## Integrated

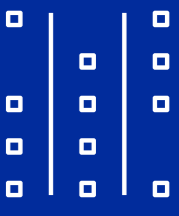
---

→ Integrates technology seamlessly into existing infrastructures, systems, and processes with choice of cloud to transform the enterprise and drive productivity from within.

→ Embedded AI for targeted use cases that drives enterprise scale productivity.

# Data

Access trusted and secure data to drive AI productivity



## AI Productivity

*Reinvent how work is done with AI agents/assistants*

*AI Assistants*



watsonx Code Assistant™



watsonx Orchestrate™



Planning Analytics

## AI/ML Ops

*Work with AI models, tools and governance that's built for business—engineered to ensure trust and scalability in applications*

*AI Models*



Granite™



Meta Llama



Mistral

*AI Tools*



watsonx.ai™

*AI Governance*



watsonx.governance™

## Data Fabric

*Bring all your business data together and optimize how it moves through your systems to scale analytics and AI in your applications while protecting it*

*Databases*



watsonx.data™

*Data Intelligence*



Data Product Hub



Knowledge Catalog



Manta Data Lineage

*Data Integration*



DataStage®



Databand®



Streamsets

*Data Security*



Guardium® Data Security Center

## Data Storage

*Store data across edge, core and clouds*

*Software-defined Storage*



Storage Ceph®

# Appendix

IBM's Point of View:  
[Governing DeepSeek models with IBM watsonx.governance](#)

### Platform and model agnostic

DeepSeek's open-source model aligns with IBM's vision.

IBM watsonx.governance helps enterprises govern open-source models including DeepSeek-R1, deployed anywhere.

### Risk-based decision making

DeepSeek-R1 like any other open-source model carry potential risks.

IBM watsonx.governance provides a systemic approach toward decision making of model onboarding and its enterprise-wide use based on risk identification and assessment across various risk dimensions.

### Security and Data Privacy

DeepSeek's Chinese origin introduces complexities.

IBM helps with organizational alignment and enforcement of established enterprise security and privacy policies for DeepSeek models.

31

### Compliance: EU AI Act

DeepSeek faces EU AI Office regulatory scrutiny across range of risks<sup>1</sup>

IBM watsonx.governance helps organization prepare to meet requirements of the EU AI Act, while also preparing for growing worldwide AI regulations and industry standards.

### The AI Alliance

IBM is a founding member of the AI Alliance<sup>2</sup> to advance Open, Safe, Responsible AI.

IBM is dedicated to client success in AI, providing open, governed, secure, and scalable solutions that meet their evolving needs.

### For more information

[Scale trusted AI with IBM watsonx.governance](#)

[Manage AI Security: IBM Guardium](#)

[IBM watsonx.governance in Action: Evaluating DeepSeek R1-Powered RAG Apps](#)

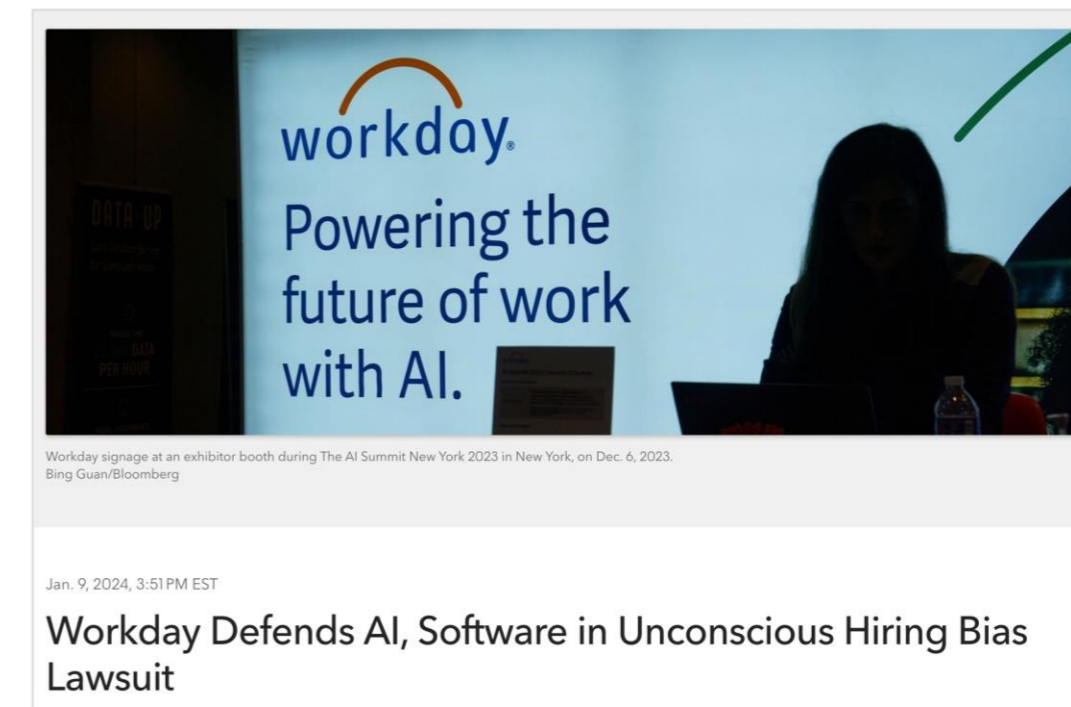
As leaders look to scale generative AI, *trust* will be critical.



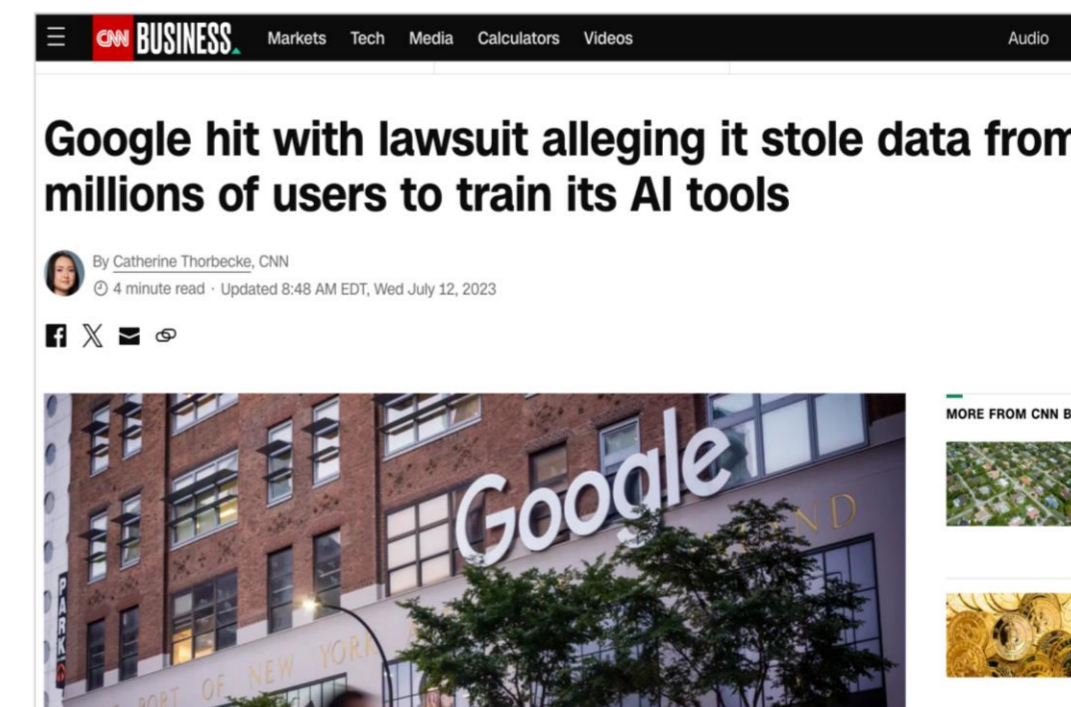
*New York Times sued OpenAI for the use of their copyrighted content. It is essential that data that drives GenAI is owned and safe to use from a legal standpoint.*



*DPD, a UK-based parcel delivery service discontinued its AI Chatbot after a frustrated user coaxed the system into speaking bad about DPD's customer service. GenAI needs to be consistent and avoid coercion from users to change.*



*Workday has faced many claims their AI tool used in hiring process is discriminatory. GenAI tools must be ethical and audited against racism, sexism and other prejudices.*



*Google was sued for creating GenAI tools based on the data collected by its users without the knowledge that their data would be used in this way. Transparency of data is needed for both the collection and distribution of GenAI data.*

Our approach to establish scalable and sustainable AI governance and AI model governance across the organization

### Organizational AI governance

#### Strategy

Who?  
Businesses, AI ethics boards, data/AI leaders, internal policy and regulations, CPOs, CISOs

#### Planning

Who?  
Businesses, AI ethics boards, data/AI leaders, internal policy and regulations, CPOs, ecosystems

- Trustworthy AI principles
- AI governance policies, processes and metrics
- Operating model
- Governance structures
- Regulatory and risk assessment

### Automated AI model lifecycle governance

#### Development and deployment

Who?  
Dev teams, IT leaders, CDAOs, software and digital science leaders

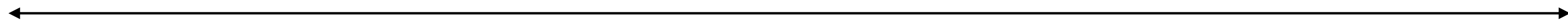
#### Operations

Who?  
IT leaders, MLOps teams

#### Monitoring and portfolio management

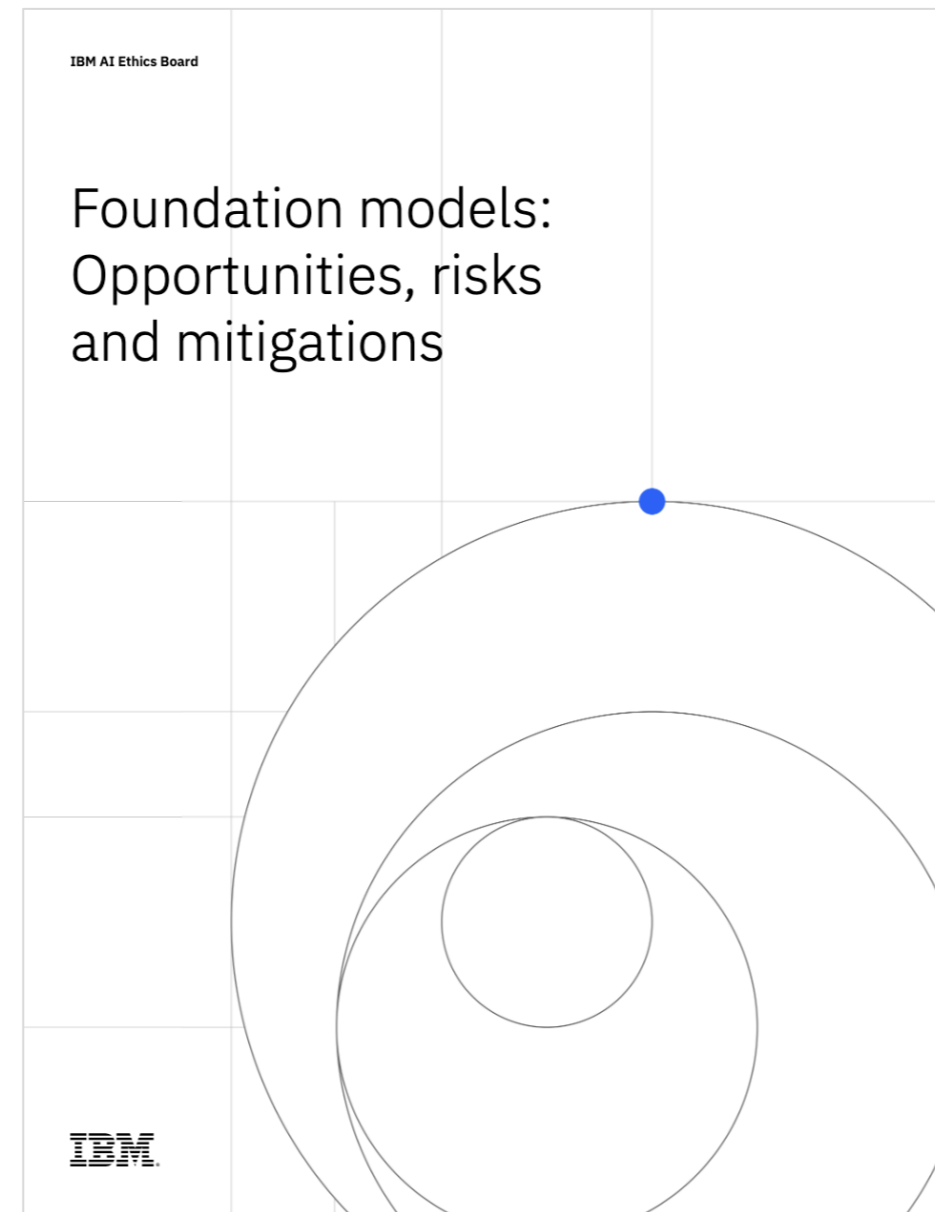
Business Outcomes and Model Governance  
Business leaders, MLOps teams

- Centralized platform for model lifecycle monitoring
- Trustworthy AI model lifecycle
- Defined model OKRs and KPIs
- Model onboarding and sustaining processes



Supported by watsonx.governance to accelerate responsible, transparent, and explainable AI workflows

Based on the AI Ethics Board's whitepaper *Foundation models: Opportunities, risks, and mitigations*, IBM's [AI Risk Atlas](#) analyzes unique risks of working with generative AI, foundation models, and machine learning models across all phases of the AI lifecycle.



## AI risk atlas

Last Updated: 2024-03-21

Explore this atlas to understand some of the risks of working with generative AI, foundation models, and machine learning models.

Risks are categorized with one of these tags:

- Traditional AI risks (applies to traditional models as well as generative AI)
- Risks amplified by generative AI (might also apply to traditional models)
- New risks specifically associated with generative AI

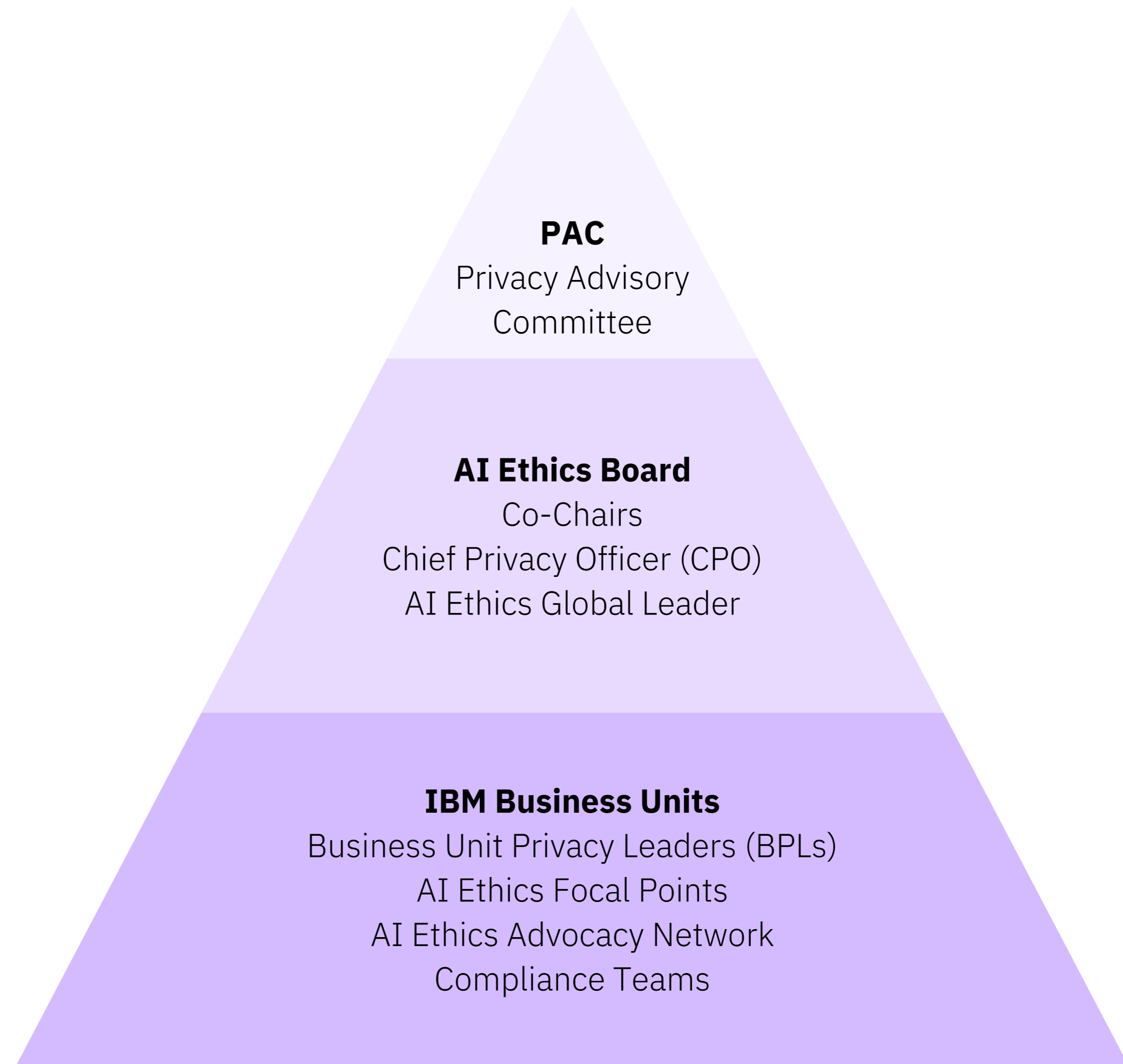
**Risks associated with input**

Training and tuning phase

 <b>Fairness</b> Data bias <b>Amplified</b>	 <b>Robustness</b> Data poisoning <b>Traditional</b>	 <b>Value alignment</b> Data curation <b>Amplified</b> Downstream retraining <b>New</b>
 <b>Data laws</b> Data transfer <b>Traditional</b> Data usage <b>Traditional</b> Data acquisition <b>Amplified</b>	 <b>Intellectual property</b> Data usage rights <b>Amplified</b> Confidential information in data <b>Traditional</b>	 <b>Transparency</b> Data transparency <b>Amplified</b> Data provenance <b>Amplified</b>

# IBM as Client Zero for organizational AI governance: AI Ethics Board & Office of Privacy and Responsible Technology

- Use case review process.
- Robust workflow using IBM tools to collect, consolidate, display, and monitor the lifecycle.
- Automate the capture and integration of facts from the AI lifecycle
- Automate and consolidate multiple tools, applications, and platforms while documenting the origin of datasets, models, associated metadata, and pipelines.
- Conduct compliance assessments, identify gaps and develop remediation campaigns based on regulatory requirements.



[Read the Integrated Governance Program Case Study to simplify and automate global privacy and AI compliance tasks for machine learning models managed by IBM](#)

Global Chief Data Office

Chief Information Security Office (CISO)

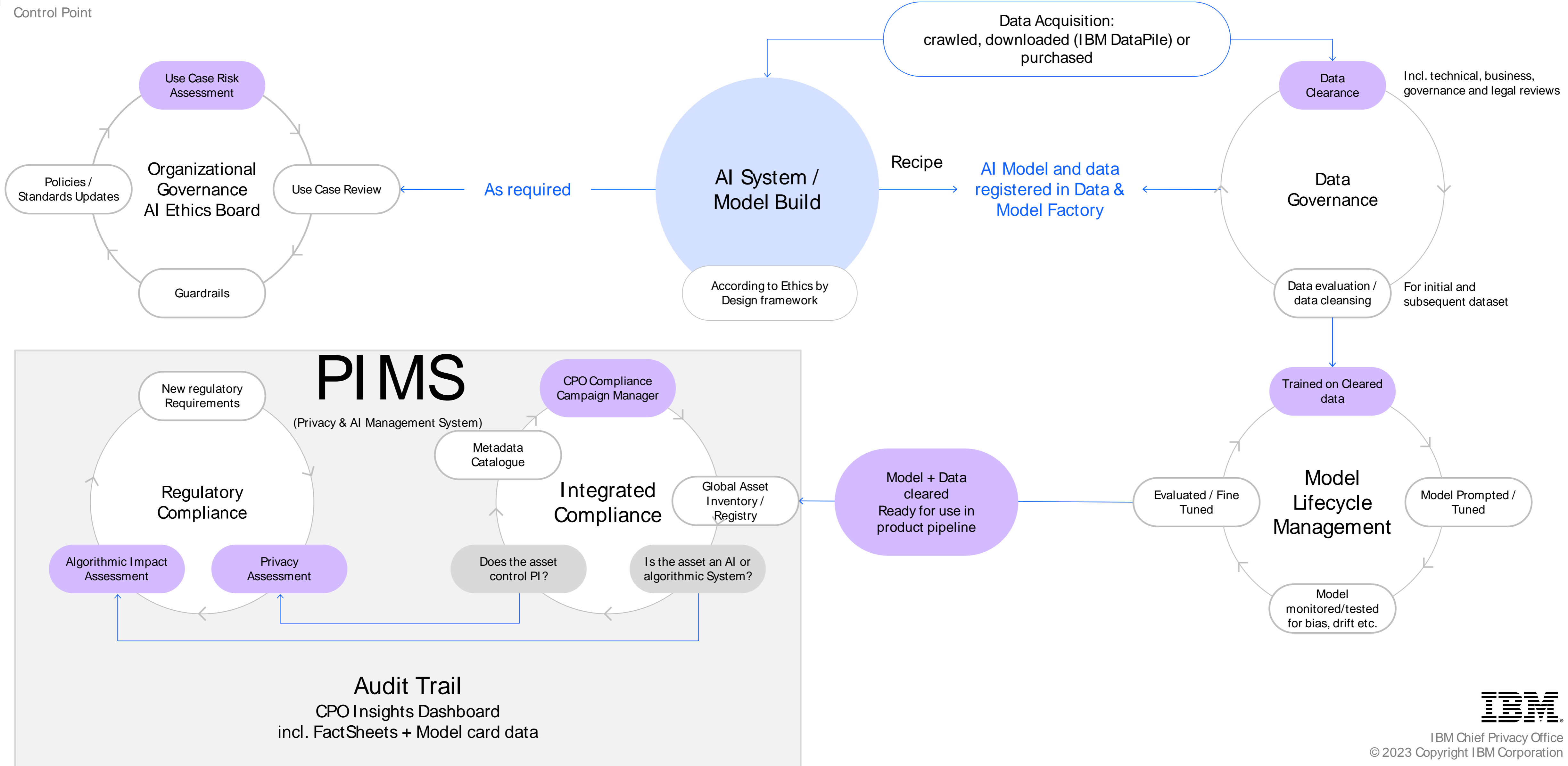
Government & Regulatory Affairs

Legal





Control Point



# Applied AI Governance Cycle

