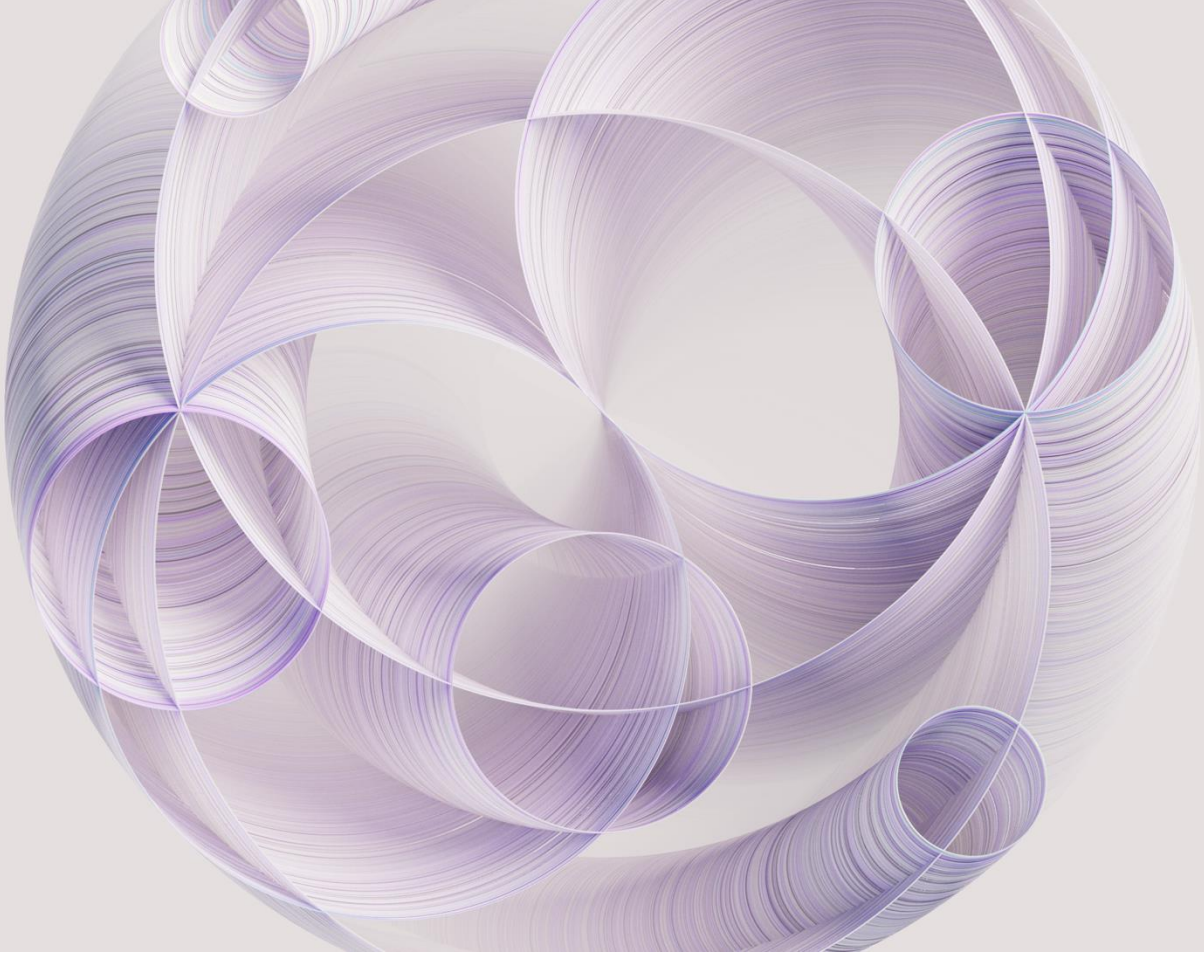
watsonx.governance
Governing AI Agents



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

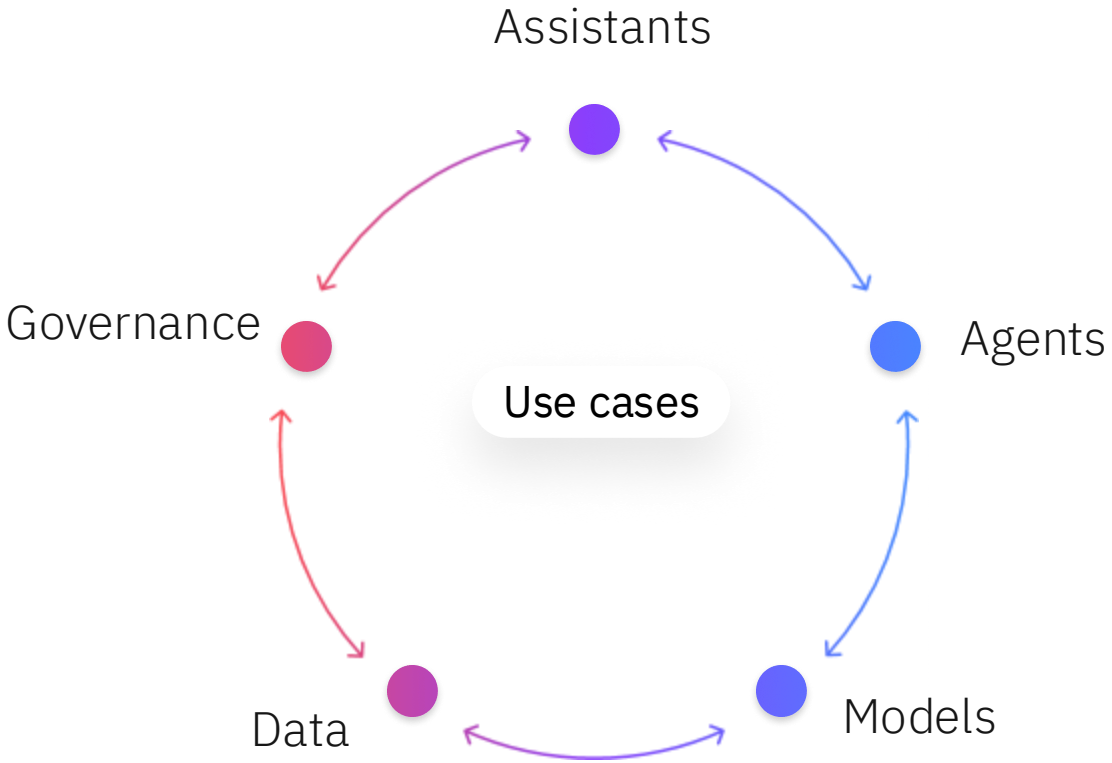
- Govern any AI, anywhere, without sacrificing speed and performance
- Automate time consuming audit and documentation processes

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



AI agents can significantly enhance how effectively humans perform tasks and achieve business outcomes.

Augmenting human intelligence

Automating routine or time-consuming tasks

Improving efficiency and productivity

Enhancing decision-making and quality of responses

Agentic AI brings with it immense promise

Opportunity

$4.4T

global profits annually[1]

Expansion

45%

projected 5-year CAGR* for the Agentic AI market [2]

Innovation

1/3

“By 2028, one-third of interactions with generative AI (Gen AI) services will use action models and autonomous agents for task competition, according to Gartner, Inc.” [3]

*compound annual growth rate

1 - AI in the workplace: A report for 2025 | McKinsey

2 - AI Agents: What They Are and Their Business Impact | BCG

3 – Gartner®, Gartner Predicts One-Third of Interactions with GenAI Services Will Use Action Models & Autonomous Agents for Task Completion by 2028., March 11, 2024

Agents will power the future

By 2028, one-third of interactions with gen AI services will use action models and autonomous agents for task competition.[1]

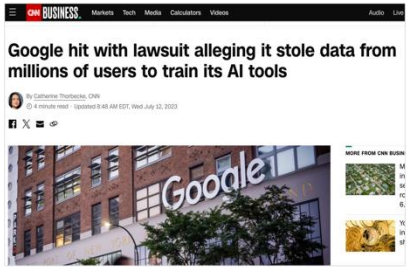
1. Gartner Predicts One-Third of Interactions with GenAI Services Will Use Action Models & Autonomous Agents for Task Completion by 2028, Gartner, March 11, 2024

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.***

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.***

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.***



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.***

Assess, evaluate and monitor across agentic AI

Agentic AI Governance

Evaluate, deploy and monitor all types of agents.

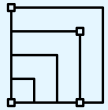
Manage compliance, security and risk.

Orchestrator for Tools and Agents

Multi-agent, multi-tool supervisor, router, and planner which facilitates complex task execution

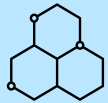
Prebuilt AI Agents

Accelerate AI Agents with pre-built utility agents and domain agents.
Provide a searchable catalog of AI agents and tools.



Build Your Own Agents

Build custom designed agents with no-code and pro-code tooling.
Integrate 3rd party agents built in any tool or framework



AI Agent Ops

Discover, manage, monitor and optimize autonomous AI agents

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

Changing regulations

Multiple stakeholders

Manual, error-prone documentation

Increased risk

Disparate tools and data

Vulnerable data

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

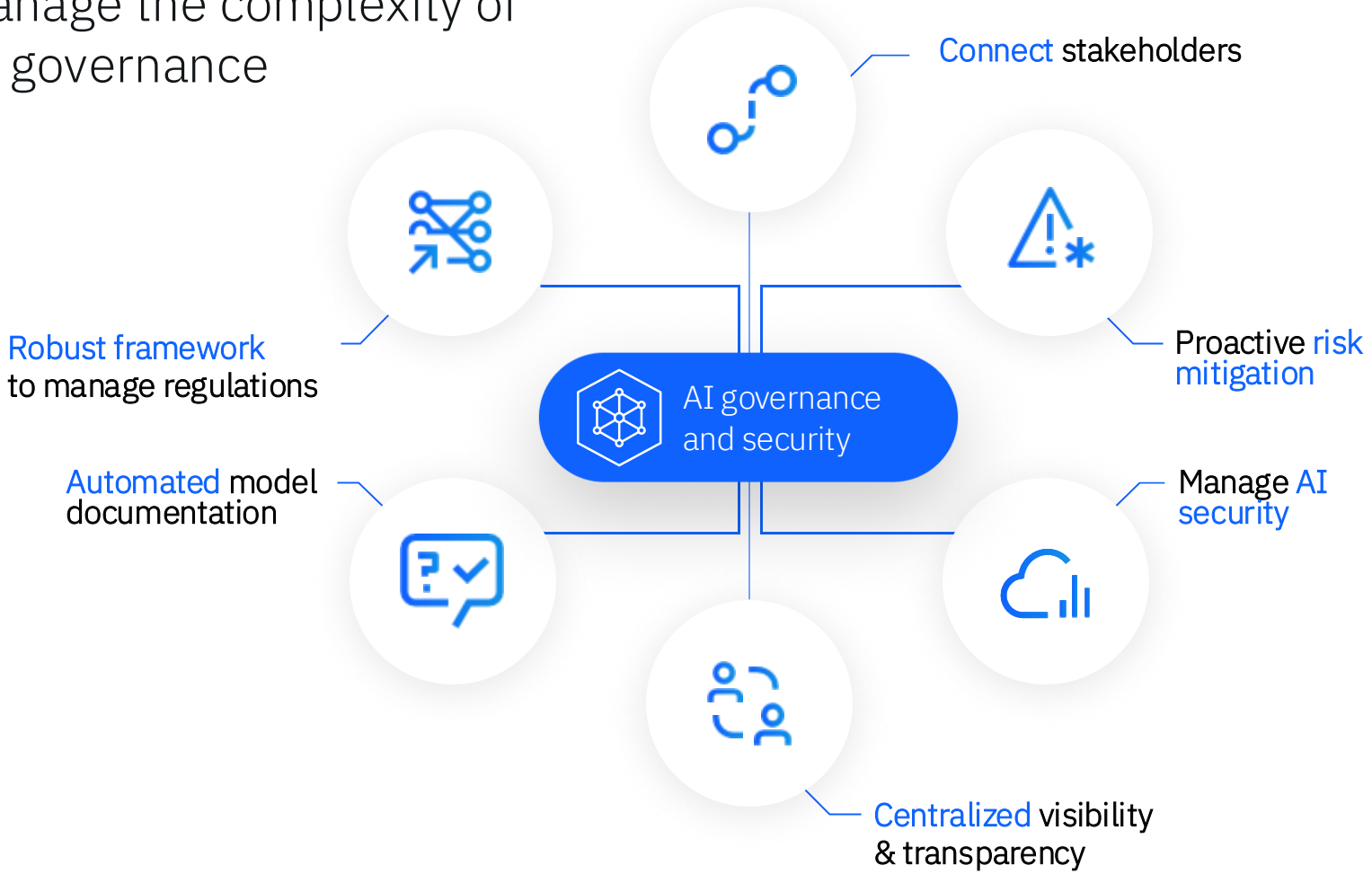
Manual and error prone documentation

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 manager **risk and compliance across the entire AI lifecycle**.

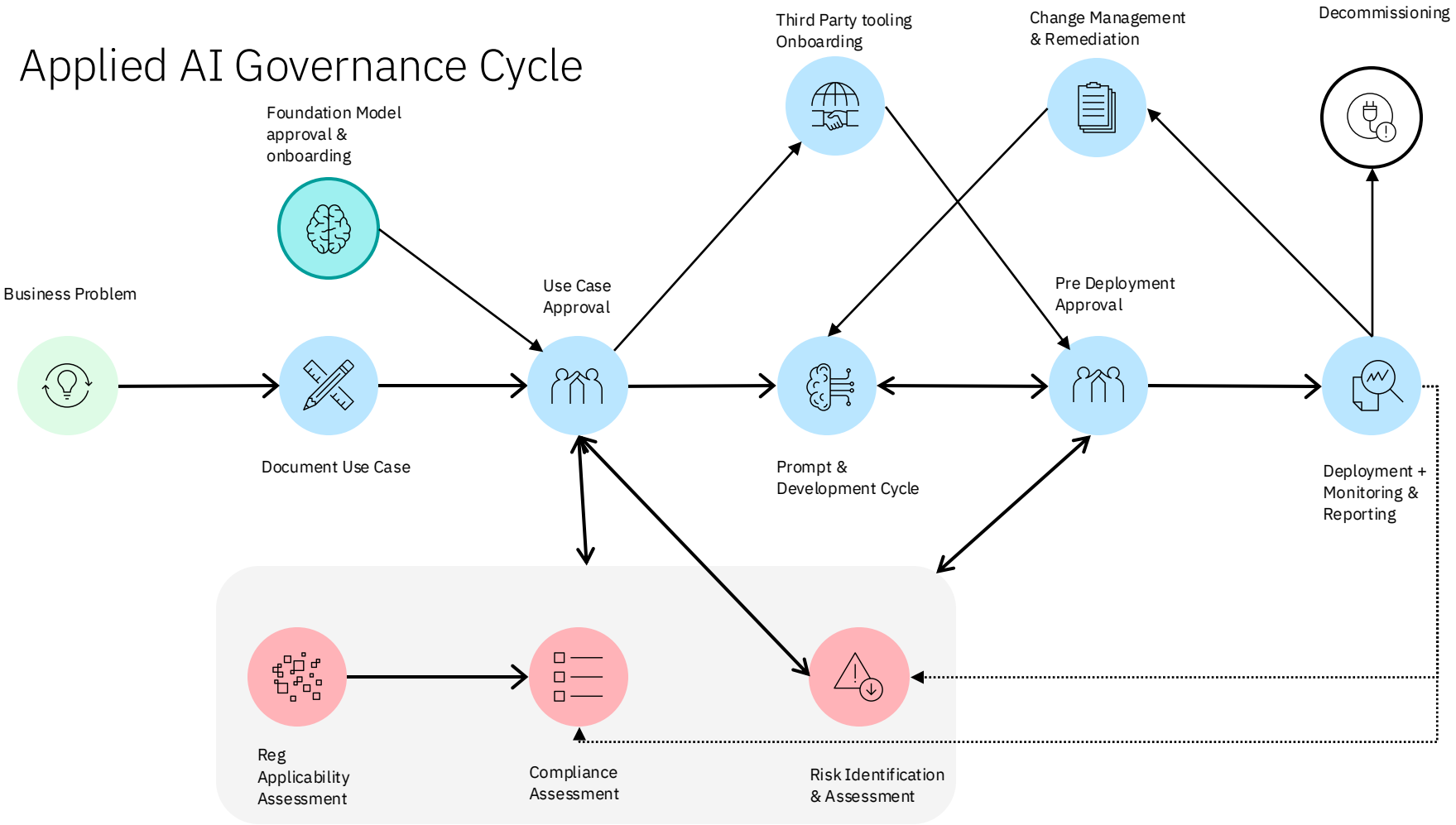
watsonx Orchestrate

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**

Applied AI Governance Cycle



IBM watsonx.governance

Accelerate responsible, transparent and explainable AI workflows

Centralized
AI lifecycle governance

Manage, monitor and govern any AI: model, app, agent or tool; across IBM and 3rd party like OpenAI, AWS, Azure, GCP, Meta, etc.

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 Agent, deployed anywhere

salesforce, SAP SuccessFactors, SAP Ariba, workday., zoominfo, Seismic, milvus, Marketo, servicenow, coupa, dun & bradstreet

watsonx.governance at a glance

- AI use case owners
- Data Scientist / AI Engineers
- Model validators
- Audit teams
- Compliance teams
- Risk management teams
- AI Security teams

- Data engineers
- AI Engineers
- (Citizen) data scientists
- MLOps
- ML engineers

Lifecycle governance | **Risk management** | **Regulatory compliance**

AI risk governance and security

Model inventory | Risk assessments | Workflows | Dashboards | Issue management

AI Factsheet

Capture model facts throughout the lifecycle

AI Observability and Guardrails

Model health | Harmful content detection | Accuracy | Drift | Bias | Explainability

Build and Deploy

IBM watsonx.ai | AWS | MS Azure | GenAI apps | SaaS solutions | Other

AI lifecycle governance

Key Highlights

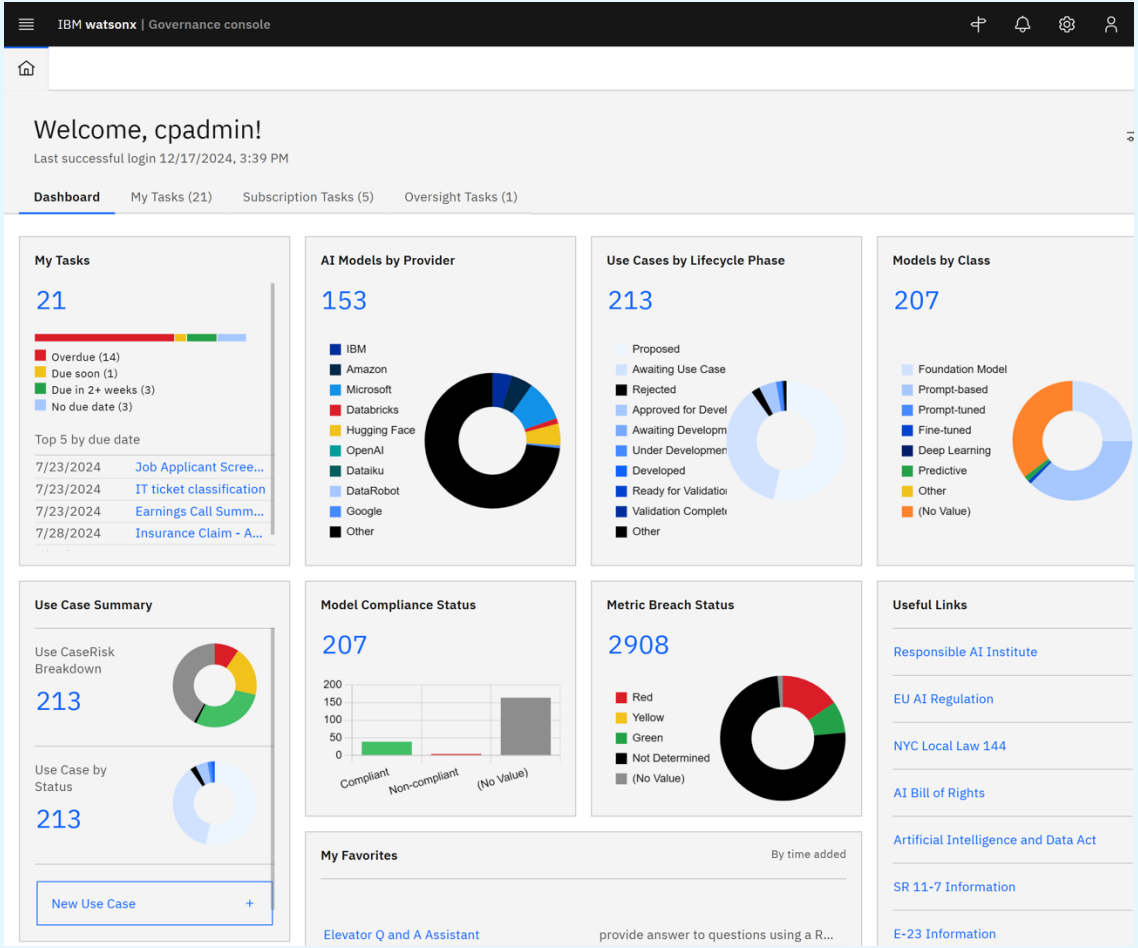
- Inventory and track ML models, GenAI apps and agents from concept/ideation through lifecycle
- Evaluate and assess your GenAI prompt templates, ML models, agents during build
- Automate the capture of the model, app and agent metadata to facilitate management and compliance

Solves for:

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

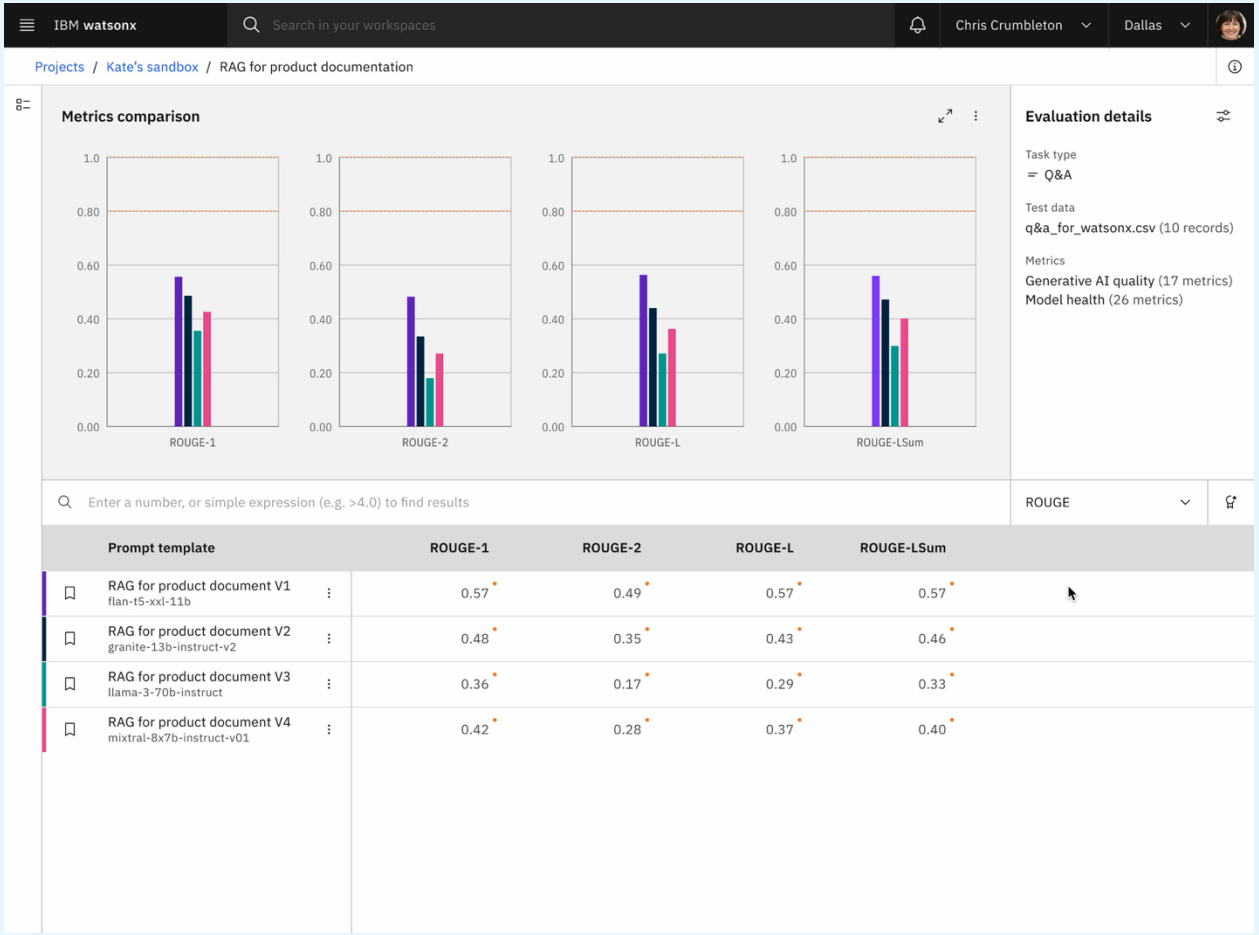
Scale responsible AI with watsonx.governance

Bring transparency and visibility into your AI use cases



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

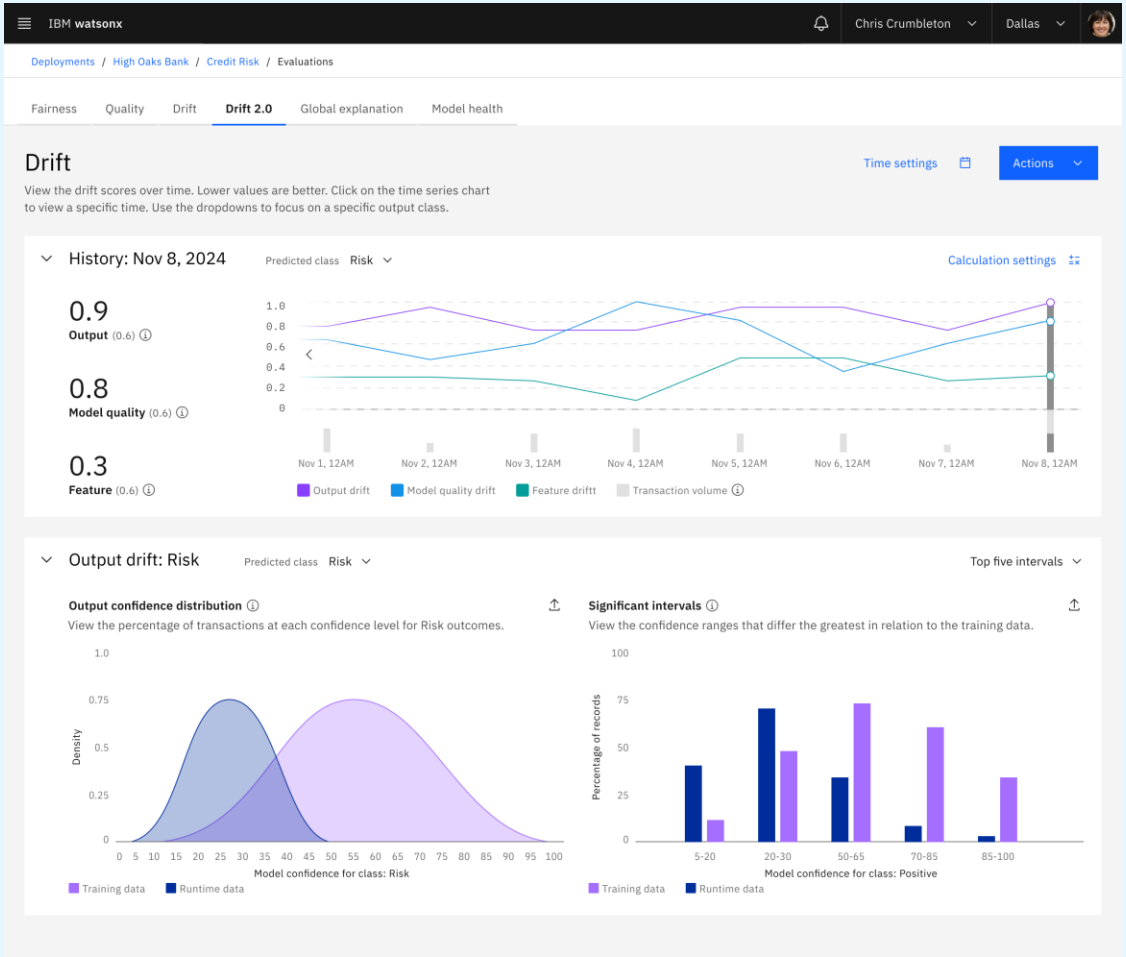
Key highlights

- 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

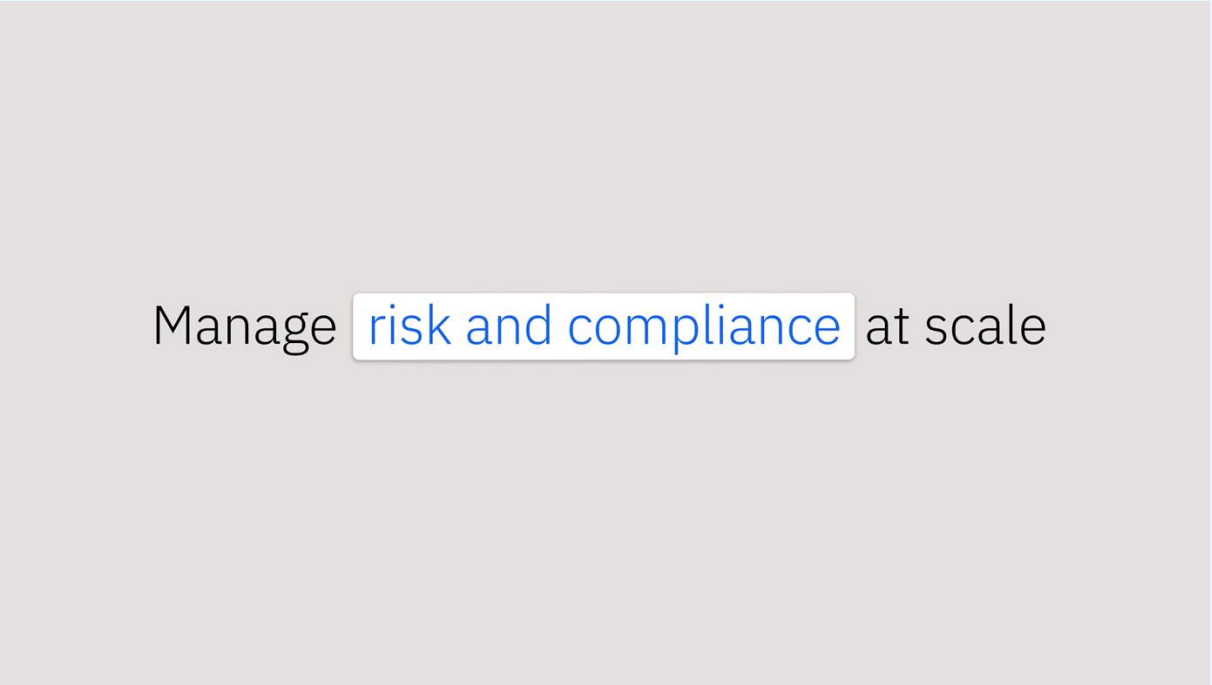
Assess, identify, and manage AI risk



Scale responsible AI with watsonx.governance

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

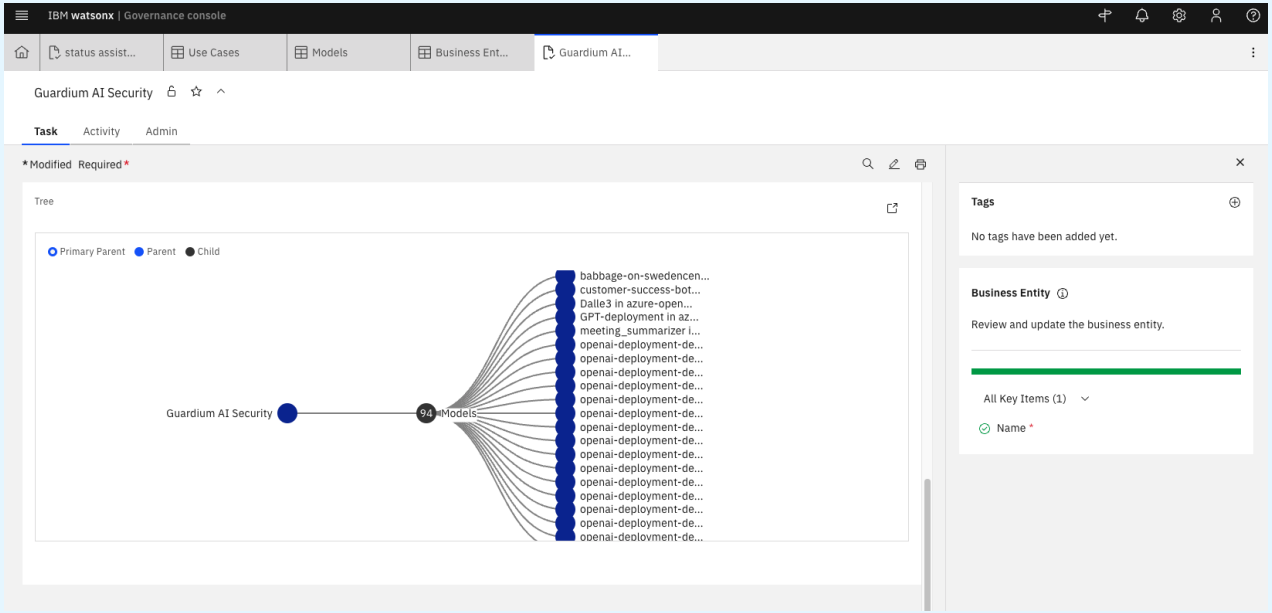


IBM Guardium AI Security integrates with watsonx.governance for Trusted AI

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



Discovered models grouped in same business entity

Dynamic and trustworthy compliance

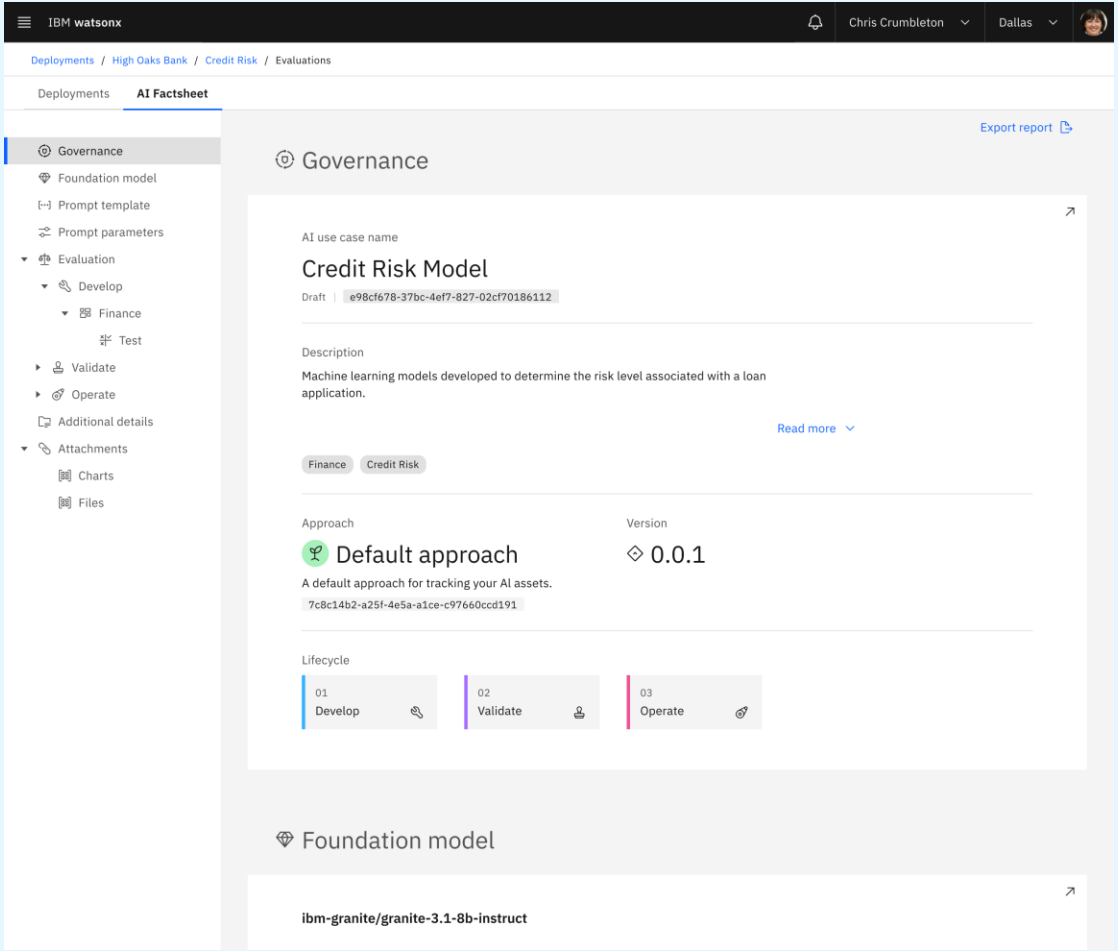
Key highlights

- 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:

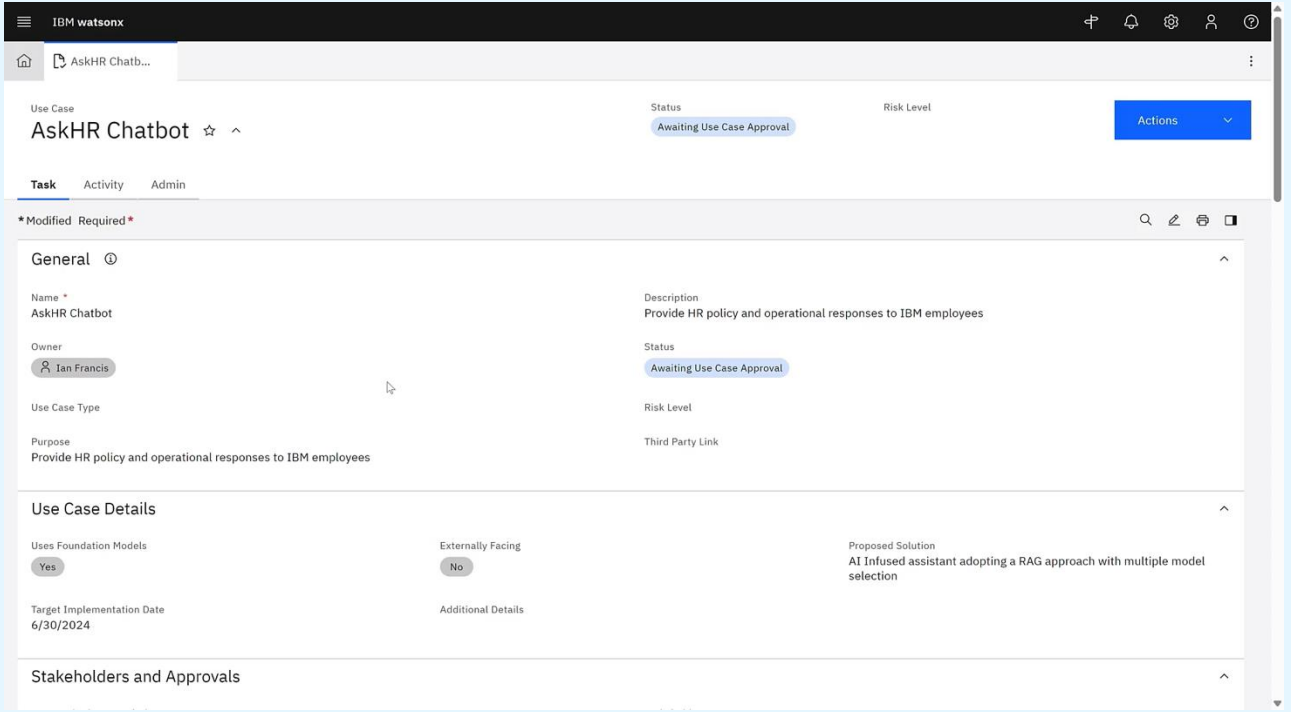
1. Changing regulations
2. Inaccurate documentation

Meet growing AI regulatory landscape



watsonx.governance for dynamic and trustworthy compliance

Example: assess regulatory applicability for your GenAI use cases



Agentic AI

Risks and Challenges

New

Emerging areas *intrinsic to* agentic AI

Risks

- Unsupervised autonomy
- Data bias
- Redundant actions
- Attack on AI agent's external resource
- Tool choice hallucination
- Sharing IP/PI/confidential information

Challenges

- Reproducibility
- Traceability
- Attack surface expansion
- Harmful and irreversible consequences

Amplified

Known areas *intensified by* agentic AI

Risks

- Misaligned actions
- Discriminatory actions
- Over- or under-reliance
- Unauthorized use
- Exploit trust mismatch
- Unexplainable or untraceable actions
- Lack of transparency

Challenges

- Evaluation
- Accountability
- Compliance
- Mitigation and maintenance
- Infinite feedback loops
- Shared model pitfalls

Key lifecycle governance activities
For agentic systems

Experimentation tracking

Track agentic app variants and compare results to inform which to push to production

Agentic system metrics, monitoring and alerts

Oversee elements such as hallucination, answer relevance, and system drift in production and development

Traceability

Help developers debug agentic app by tracing each step of the user interaction and agent processing

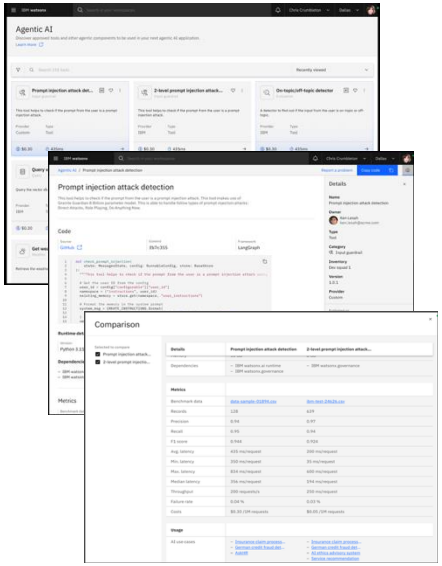
Cataloging of agentic AI applications

Single consolidated view of all in development and use

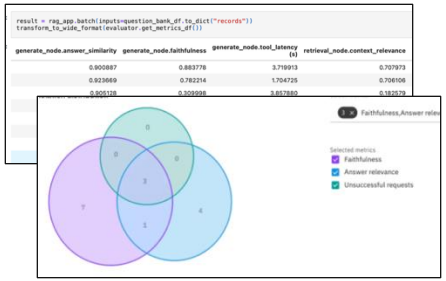
Agentic AI governance feature highlights*

Providing the tools and capabilities to develop, deploy, manage and govern AI agents

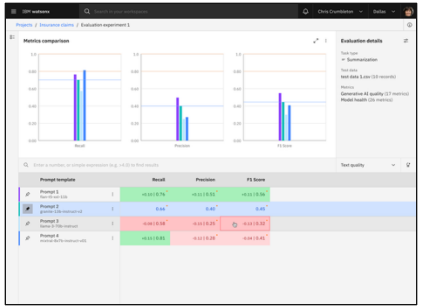
Agentic tool catalog



Agent evaluators



Agent evaluation studio



Agent production monitoring**

*Please note: roadmap items are subject to change
**Release planned for Q3 2025

watsonx.governance

Governance of agentic AI

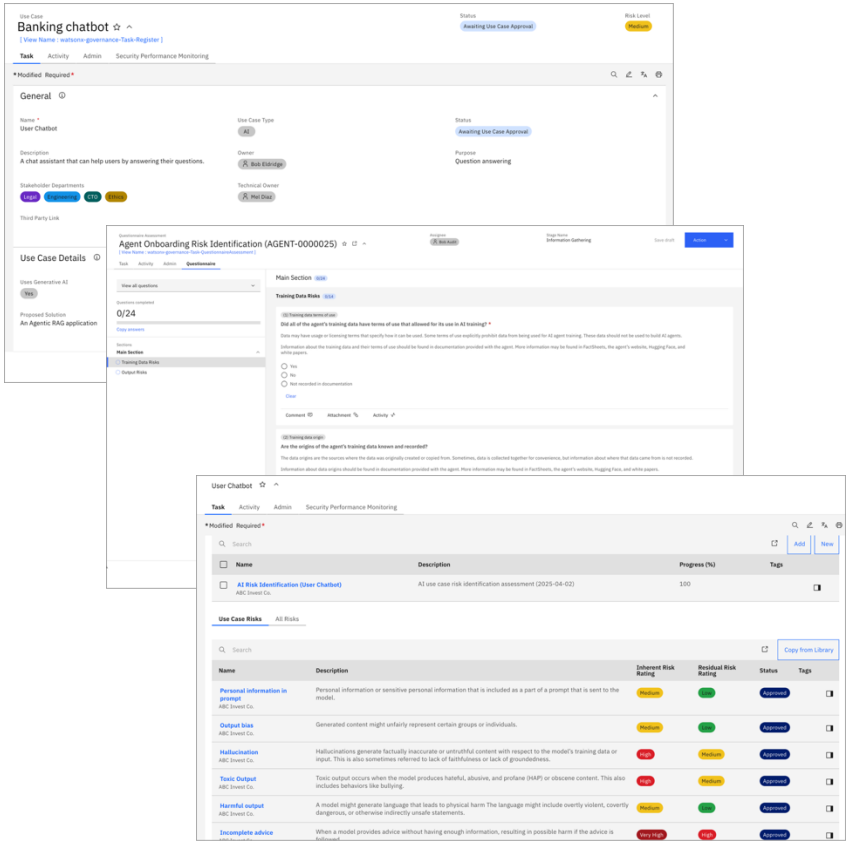
Agent Onboarding

- *Onboard Agents, assess risk, and capture metadata from build to deployment*
- **AI Risk Assessment** identifies AI agent risks during onboarding with automated workflows
- **Stakeholder management** fosters cross-functional collaboration to approve and onboard agents

- Identify and manage AI Risks by using automated workflows to onboard AI Agents
- Identify risk levels and applicable AI regulations for your AI use case

Key features:

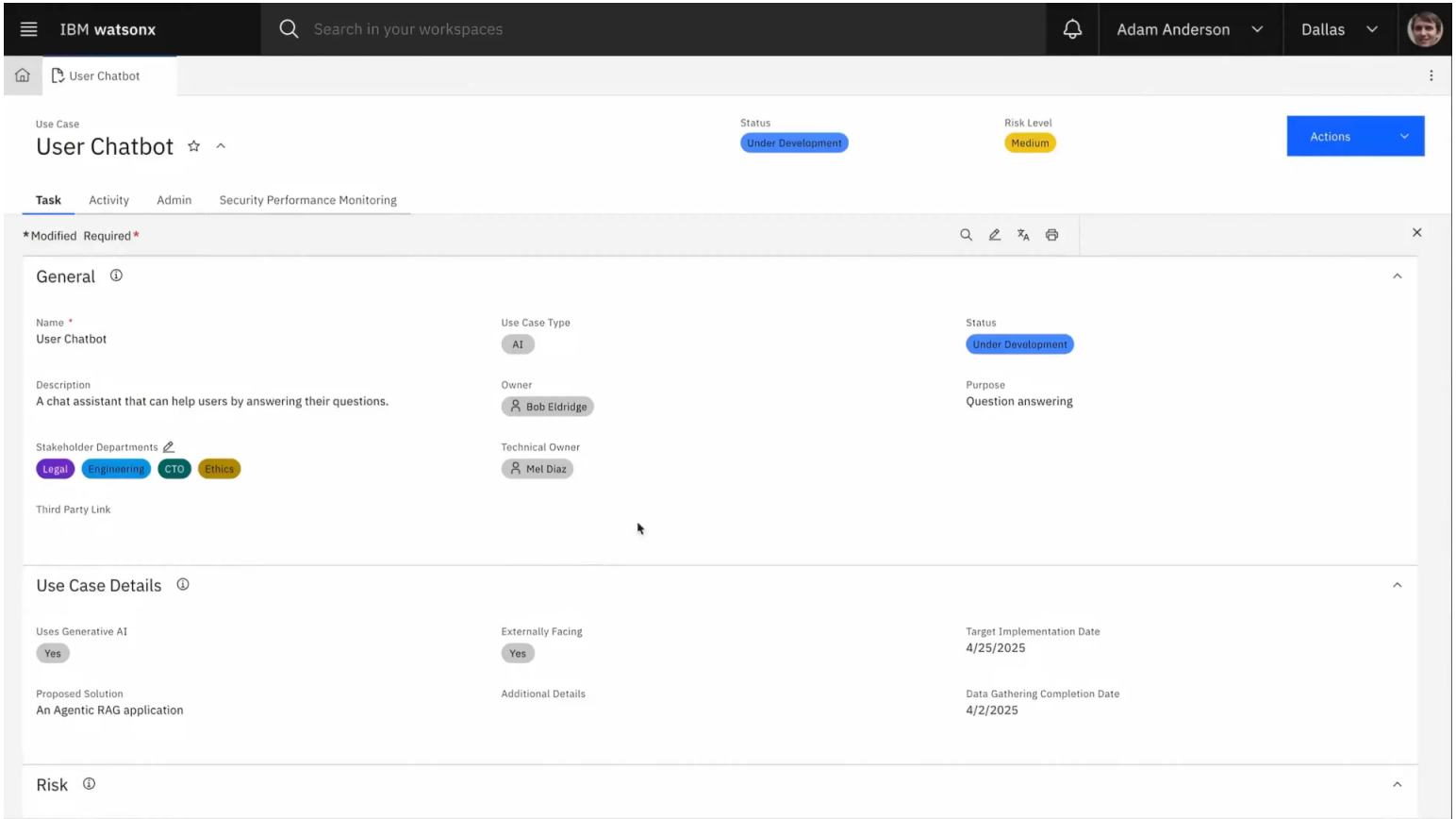
- Create an AI use case describing your business goal using the Governance Console
- Risk questionnaire to identify risk dimensions, compliance needs, and applicable AI regulations and generate an assessment for your use case



watsonx.governance

Governance of agentic AI

Agent onboarding demo (1 min)



watsonx.governance

Governance of agentic AI

Agentic Tool Catalog

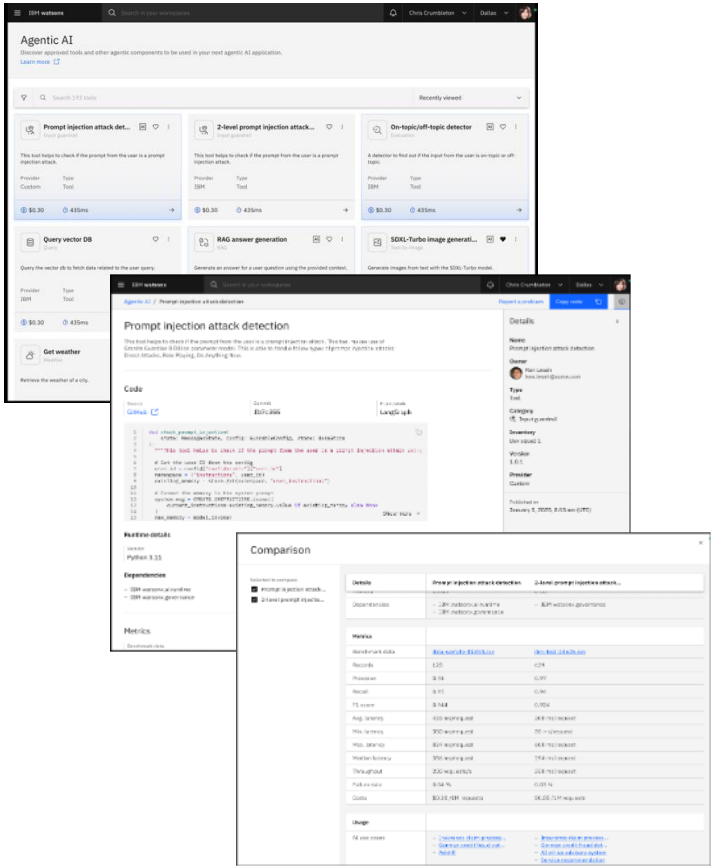
Consolidated list of agentic tools to manage and inform selection

- **Re-using agent tools** helps accelerate progress across users and use cases
- **Promote approved tools** to encourage utilization and proper use

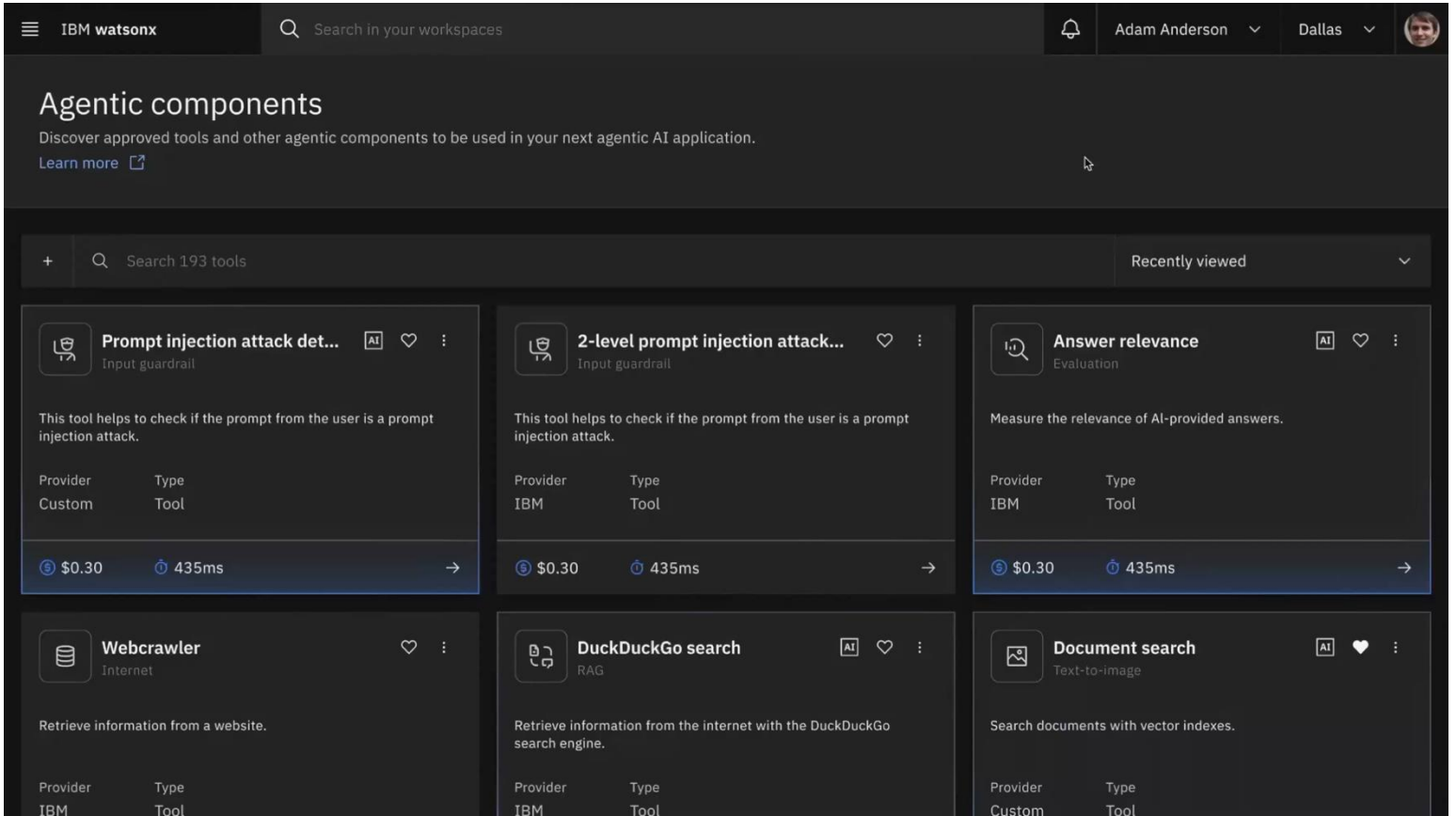
- Tools perform specific tasks and are key in designing and building agentic systems and span a broad range e.g. data retrieval, external connection tools, and more.
- Tool catalogs are a way to manage tool ‘sprawl’ and help bring consistency across teams/units

Key features:

- Tool lineage by mapping tools to use cases
- Search by use case type / domain to find tools and get started faster
- Filter tools available by type of tool
- Tool card that documents purpose of the tool, along with quality metrics
- Easy side-by-side comparison of various tools
- See tool ratings from other users



Agentic Tool Catalog Demo (1 min 15 secs)



watsonx.governance

Governance of agentic AI

Agent Evaluators and Evaluation Studio

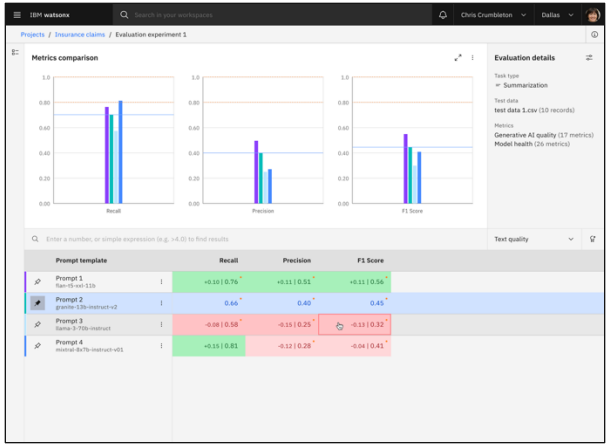
Help organizations understand and improve agents' performance to scale with confidence

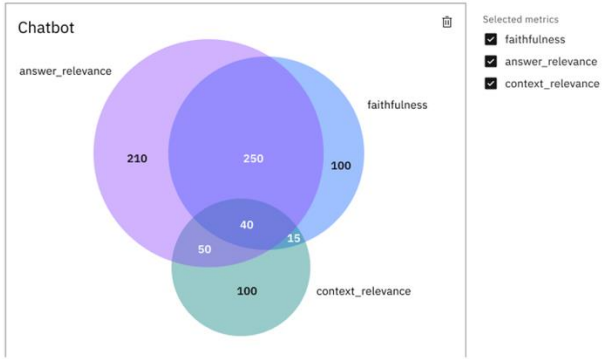
- **Evaluation metrics** assess agent competence for various tasks
- **Root cause analysis** identifies underlying reasons for poor performance to inform improvements

- Developers will be able to easily leverage the 50+ evaluation metrics from watsonx.governance to evaluate their tools and use the computed metrics in their agentic AI application
- Developers can run experiments to compare results and select the best performing version

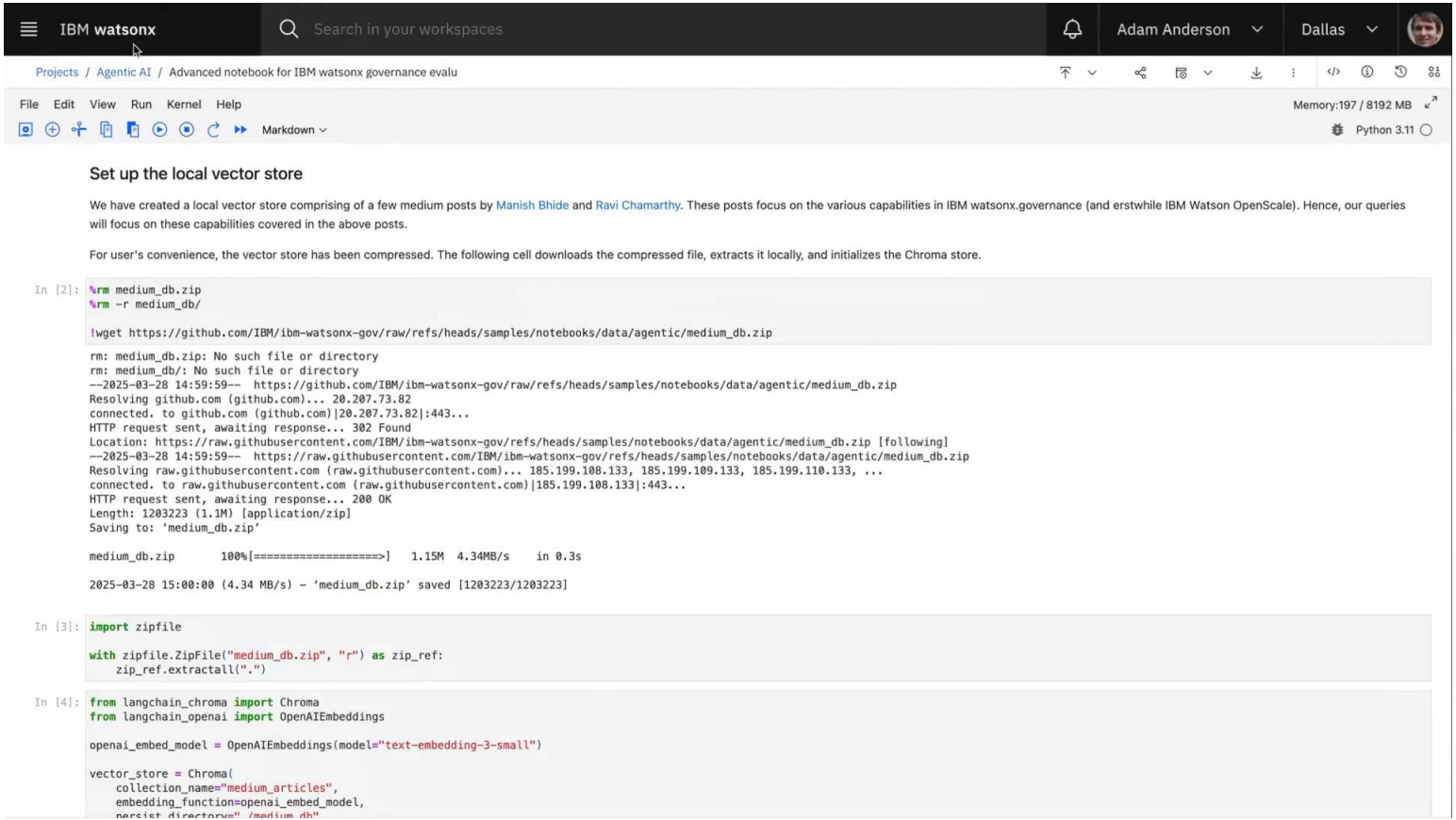
Key features:

- Compute metrics to evaluate the application by adding a one-line decorator
- Identify root cause for poor performance using Venn charts to understand interactions between metrics and make necessary improvements
- Use Evaluation Studio to track and compare experiments





Agent Evaluation Demo (2 mins)



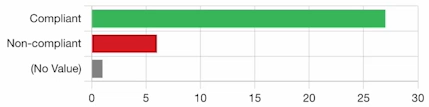
Welcome, Marc!

Last successful login 2/10/2025, 4:41 PM

Home | My Tasks (0) | Subscription Tasks (0) | Oversight Tasks (0)

Model Compliance Status

34



Model Inventory

Use Cases by Risk Level

32

Models by Risk Tier

34

Foundation Models by Hosting Location

0

No data available

Approved Models

18

Models added in the last 30 days

1

AI Agents

Agents by Level

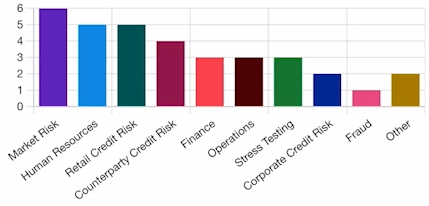
13

Agents by Risk Level

13

Models by Department

34



Change Management

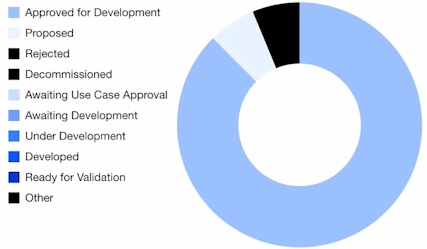
Change Requests by Status

41

My Active Change Requests

Use Cases by Lifecycle Phase

32



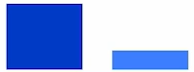
Model Performance

186

Deployments

Deployments by Status

40



Production Deployments

Model Validation

Model Reviews by Status

37

Reviews Awaiting Assignment

1

Reviews in Progress

13

Models by Provider

Foundation Models

0

No data available

Developed Models and Prompts

34

Useful Links

AI Risk Atlas

EU Artificial Intelligence Act

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).

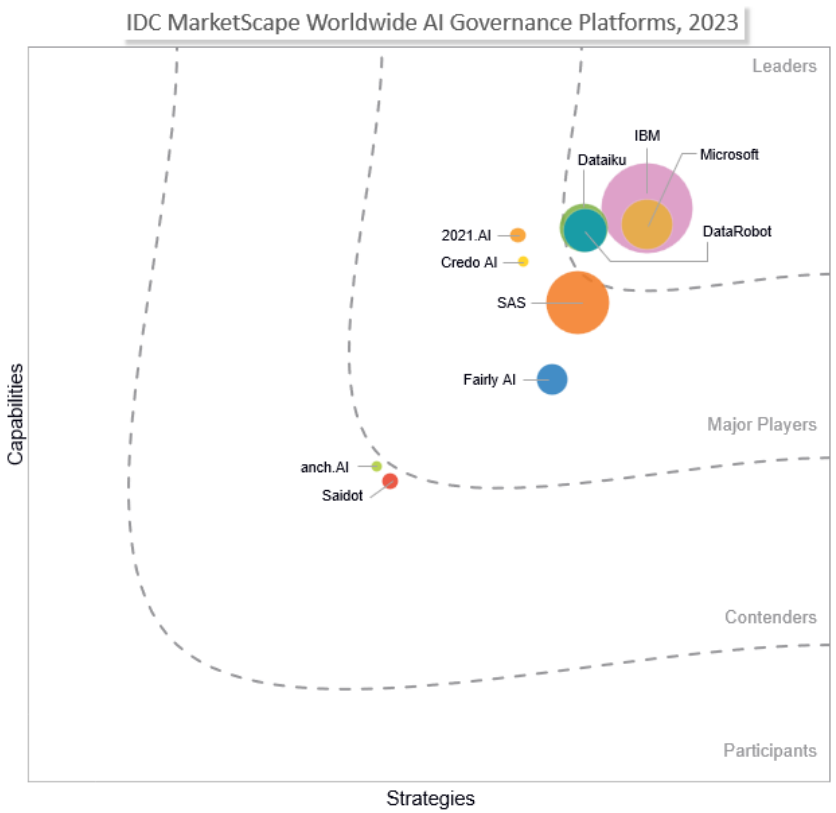
Assess and reduce AI risk at runtime

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

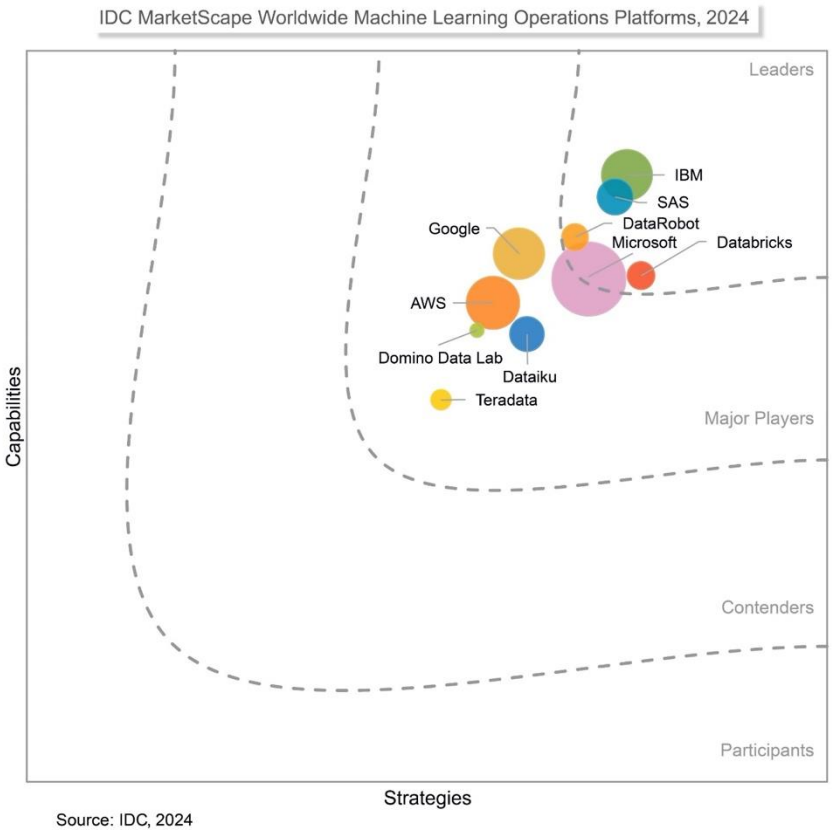
Worldwide compliance expertise

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

IBM is a leader in AI Governance and ML Ops



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



Source: IDC, 2024

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

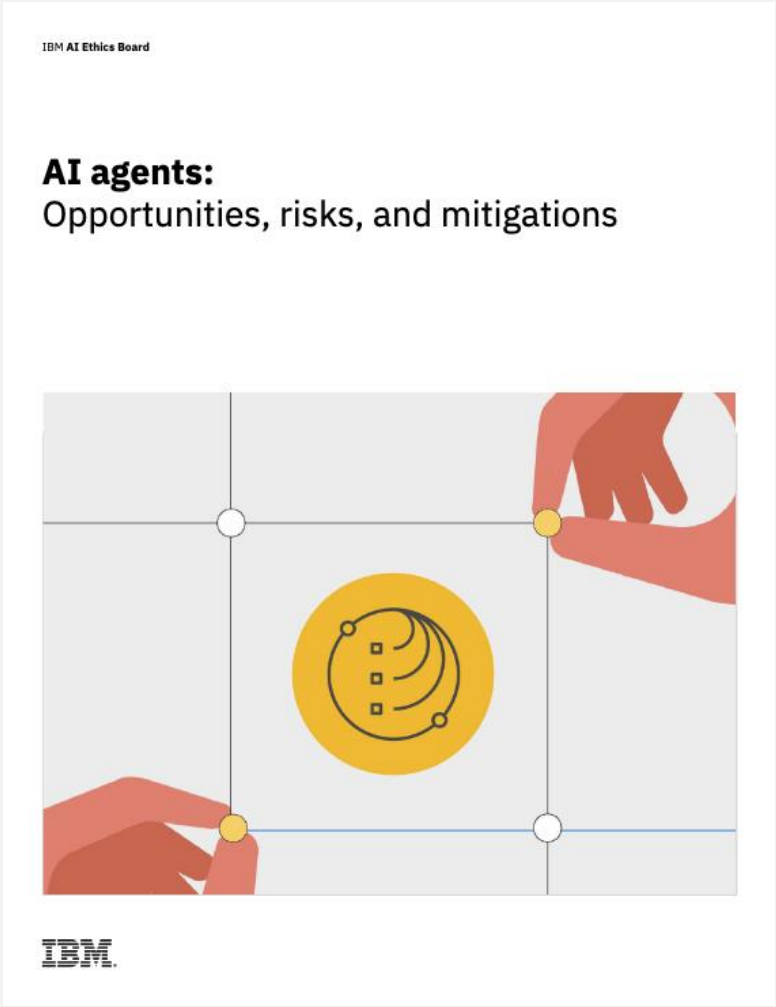
Key strengths: end-to-end solution and commitment to support

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

For more information
on IBM’s perspective

Read *AI agents: Opportunities, risks, and mitigations*, a deep-dive into the unique risks posed by AI agents and potential mitigations, written by the IBM AI Ethics Board.

Scan the QR code to
access the paper:



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