

Establishing AI Governance to Meet Compliance Challenges

Thursday, June 13 | 12:00 - 1:00 PM ET



Kay Firth-Butterfield
CEO, [Good Tech Advisory LLC](#)



Mrinal Manohar
CEO, [Casper Labs](#)

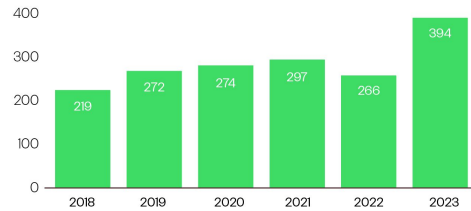


Shyam Nagarajan
Executive Partner, [IBM Consulting](#)

The AI era is here

Number of Fortune 500 earnings calls mentioning AI, 2018-2023

Number of earnings calls



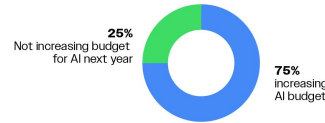
Rising demand

AI has emerged as a strategic imperative to drive new efficiencies and enhance productivity to create massive new value.

Source: IEEE Spectrum (April 2024)

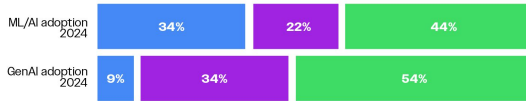
Adoption rates for AI & GenAI - Budget plans for AI

Budget



Adoption

■ Adopted already ■ Plan to in the next year ■ Enterprises intending to adopt in more than one year (or never)

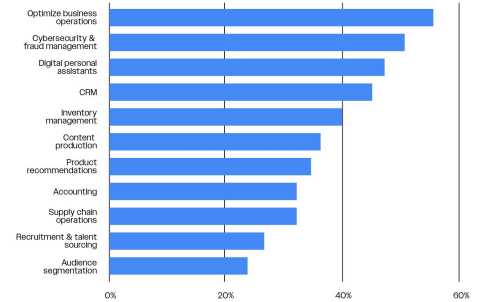


Steadily climbing AI budgets

A vast majority of IT departments are allocating consistently more resources to support adoption initiatives.

Source: Gartner CEO Insight Report (March 2024)

Most popular business use cases

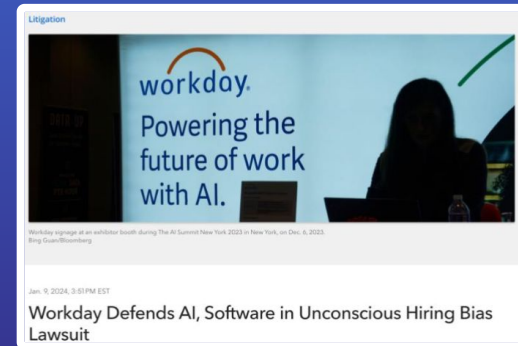
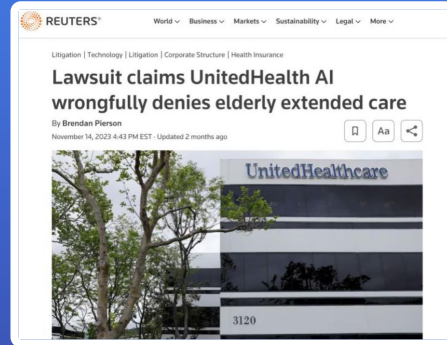
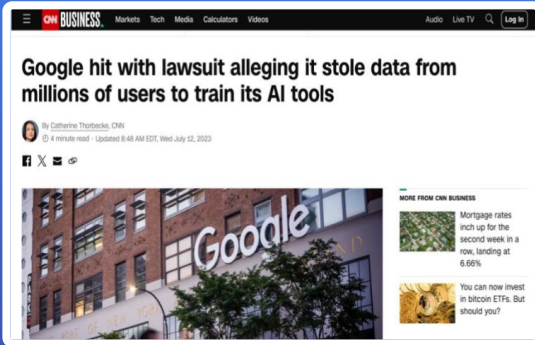


Established use cases

AI is driving meaningful results across a growing number of proven enterprise use cases.

Source: CompTIA 2024 IT Industry Outlook


... And so is unbounded risk



Regulators are stepping in.

MEPs approve world's first comprehensive AI law

8 days ago
By Shiona McCallum, Liv McMahon & Tom Singleton, Technology reporters

A graphic featuring the European Union flag (a circle of twelve gold stars on a blue background) with a large black smartphone icon in the center. The letters 'AI' are written in white on the phone's screen.

The European Parliament has approved the world's first comprehensive framework for constraining the risks of artificial intelligence (AI).

AI Regulation is Coming- What is the Likely Outcome?



A digital illustration of a glowing blue and purple circuit board or data network structure against a dark background.

Photo: gutflores_daniel/istock Stock

Blog Post by Bill Whyman
Published October 10, 2023

Biden wants to move fast on AI safeguards and signs an executive order to address his concerns

President Joe Biden is seated at a desk in the East Room of the White House, signing an executive order. Vice President Kamala Harris stands behind him. A large blue screen in the background displays the text 'ARTIFICIAL INTELLIGENCE SAFETY, SECURITY, AND TRUST'.

U of AI President Joe Biden signs an executive on artificial intelligence in the East Room of the White House, Monday, Oct. 30, 2023. (AP Photo/Evan Vostz)


BY JOSH BONE AND MATT O'BRIEN
Updated 12:07 PM EDT, October 30, 2023

China Brief: Beijing Pushes for AI Regulation

A weekly digest of the news you should be following in China, plus exclusive analysis. Delivered Tuesday.

A campaign to control generative AI raises questions about the future of the industry in China.

By James Frazee, a deputy editor at Foreign Policy

A woman is sitting in a white chair, wearing a VR headset and holding a controller. She appears to be in a virtual reality environment. Other people are visible in the background, also wearing VR headsets.

A virtual tour is VR product during the World Artificial Intelligence Conference in Shanghai on July 7, 2023. (www.hk.hk.com/newsroom)

The state of AI regulation

- AI is **entering a new phase of maturity**, underpinned by new regulations
- **Historical parallel**: regulation spurred automotive innovation
- **EU AI Act as a global template**: AI's “GDPR moment” has arrived



Definition, law and regulation

“An AI system is a machine-based system that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments.”

- OECD / EU AI Act

Regulation is not an abstract; numerous bodies are already enforcing existing regulations around AI misuse/abuse



Consumer Financial
Protection Bureau



Federal
Communications
Commission



Global AI regulation



United Nations



COUNCIL OF EUROPE

CONSEIL DE L'EUROPE

Key actors influencing global regulations & frameworks



OECD

WORLD
ECONOMIC
FORUM



GPAI



Addressing AI's data gap

Good decisions need good data

- Current LLM challenges
 - 2M GPT users; disproportionate global representation
 - 3B are unconnected
 - Connected communities who don't produce much data
 - Global North Women and persons of colour

What is Responsible AI?

The world agrees on what is responsibility
but not on how we get there

- ✔ Bias
- ✔ Fairness
- ✔ Security reliability & robustness
- ✔ Privacy
- ✔ Explainability / transparency
- ✔ Human agency
- ✔ Accountability
- ✔ Lawfulness
- ✔ Sustainability
- ✔ Safety



South America

- Argentina
- Brazil
- Chile
- Colombia
- Peru



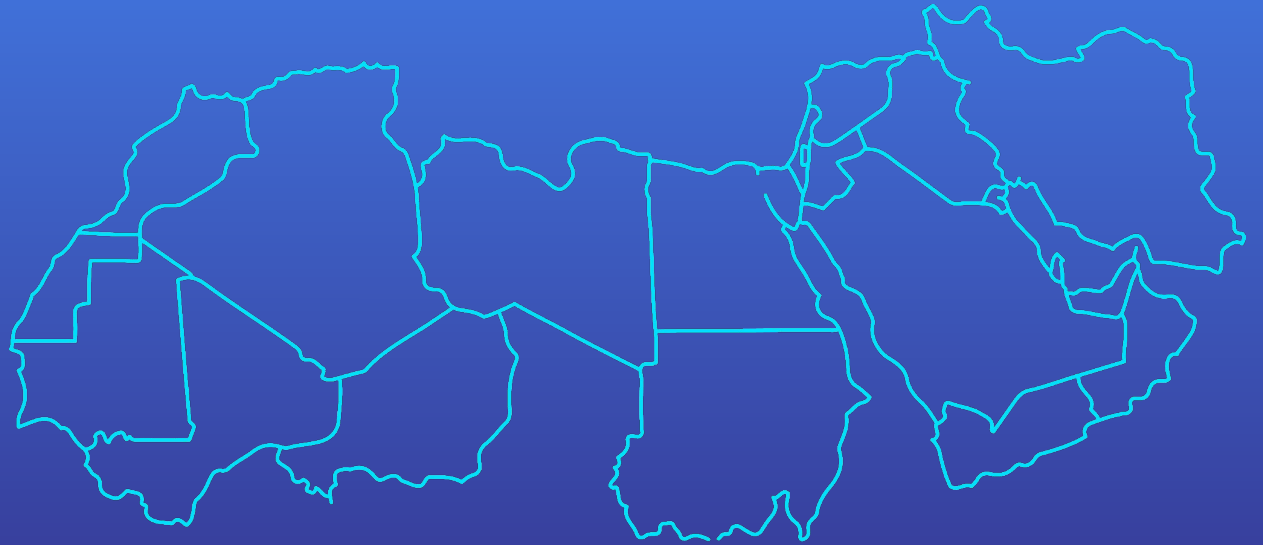
Asia and Oceania

- Australia
- Bangladesh
- China
- India
- Indonesia
- Japan
- Mauritius
- New Zealand
- Singapore
- South Korea
- Taiwan



MENA

- Egypt
- Israel
- Saudi Arabia
- UAE



North America

- Air Canada/ Getty/Open AI
- Indemnity statements from big tech firms
- Biden Executive Order
- Consumer Protections for Interactions with Artificial Intelligence (CO)
 - *Goes into effect 2/1/26*
- NIST Framework and arising from the EO
- AIDA of Canada



EU and UK

- UK – CMA and existing laws
- EU AI ACT





Risk levels

- Unacceptable
- High
- Limited
- Minimal (no risk)



Product liability

- AI Liability Directive
- Product Liability Directive

Timeline for implementation starts 10 June 2024



JUNE 2024

6 months – prohibitions on unacceptable risk AI will apply (Chapters 1 & 2)



JUNE 2026

24 months – remainder of the AI Act will apply, except;



DECEMBER 2024

12 months - notifying authorities, general purpose AI models, governance confidentiality and penalties will apply (Chapters 3 s4 and 7 and 12 & Article 78). Exception of Article 101 for GPAI providers



JUNE 2027

36 months – Article 6 (1) and the corresponding obligations in this regulation will apply

Codes of practice must be ready 9 months after entry into force according to Article 56

Some examples





HR

EEOC

Colorado

EU

Sometimes bias is
necessary (Hardoon et al)



Healthcare

Chatbots

FDA

Radiology

And more...



Banking / Financial Services

Chatbots

Loans

Financial analysis

And more...

Deepfakes

- C2PA unifies the efforts of the Adobe-led [Content Authenticity Initiative \(CAI\)](#) which focuses on systems to provide context and history for digital media, and [Project Origin](#), a Microsoft- and BBC-led initiative that tackles disinformation in the digital news ecosystem
- What does it mean for business and how to spot it (HK and CEO WPP)
- Australia – Revenge porn, deepfakes and child abuse. UK Online Harms Act





Establishing trust with vulnerable users

- Children – smart toys
- Older adults

3 key takeaways on regulation

- 1 AI regulations: not an abstract possibility; a concrete reality
- 2 Every org. needs an AI governance strategy
- 3 When thinking about governance, don't overlook the need for a) **multi-party access**; and b) **independent certification**



A New Framework for AI Risk Management



To trust AI, you need

Explainability

Easy to understand, to be able to **interpret** results and the training sets.



Transparency

How the AI models were built, the datasets, and **accountability**.



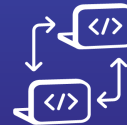
Adversarial Robustness

Was something **tampered** with and how did it affect results.



Data Privacy

Safeguard data security and **integrity**. **Prove** model's dataset lineage.



Fairness

Equitably, **context of the results**, equity, and societal safeguards.



Companies need clarity, regulators demand action

Key AI risk vectors



Training data opacity



Data privacy / confidentiality concerns



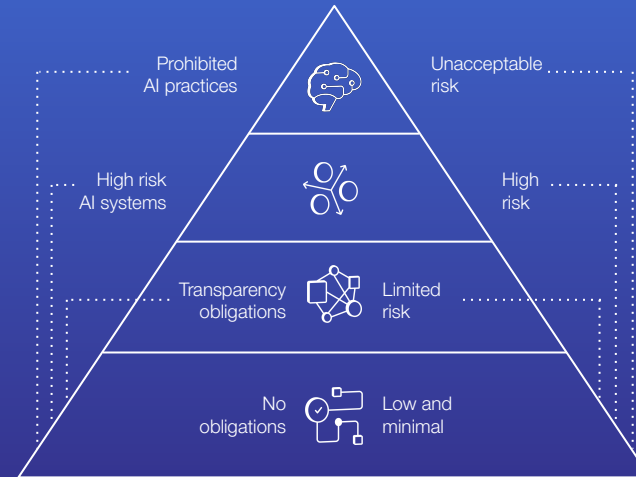
Unintended consequences of model output



Lack of auditability of model data

Failure to comply with the EU AI Act can result in significant fines of **€35m+**. It comes into effect **this year**

The EU AI Act employs a risk-based approach to regulate AI systems based on their level of risk:



Risk Is Everyone's Business

In today's turbulent environment, the need to **take on risk with confidence** is greater than ever before.

90%

of compliance leaders expect evolving business, regulatory, and customer demands to increase compliance-related operating costs by up to 30%.¹

79%

of organizations report that keeping up with the speed of digital and other transformations is a significant risk management challenge.²

77%

of organizations recognize the need to upgrade their Third-Party Risk Management operating model.³

Casper Labs & IBM are advancing AI governance



Problem:

Enterprises lack effective governance tools that can reign in uncontrollable AI and increase transparency and auditability of AI systems.

Solution:

Casper Labs is developing a comprehensive governance solution that delivers a highly secure datastore for managing, monitoring and sharing AI data.

Tamper-Proof



Hybrid Functionality



Fully Auditable



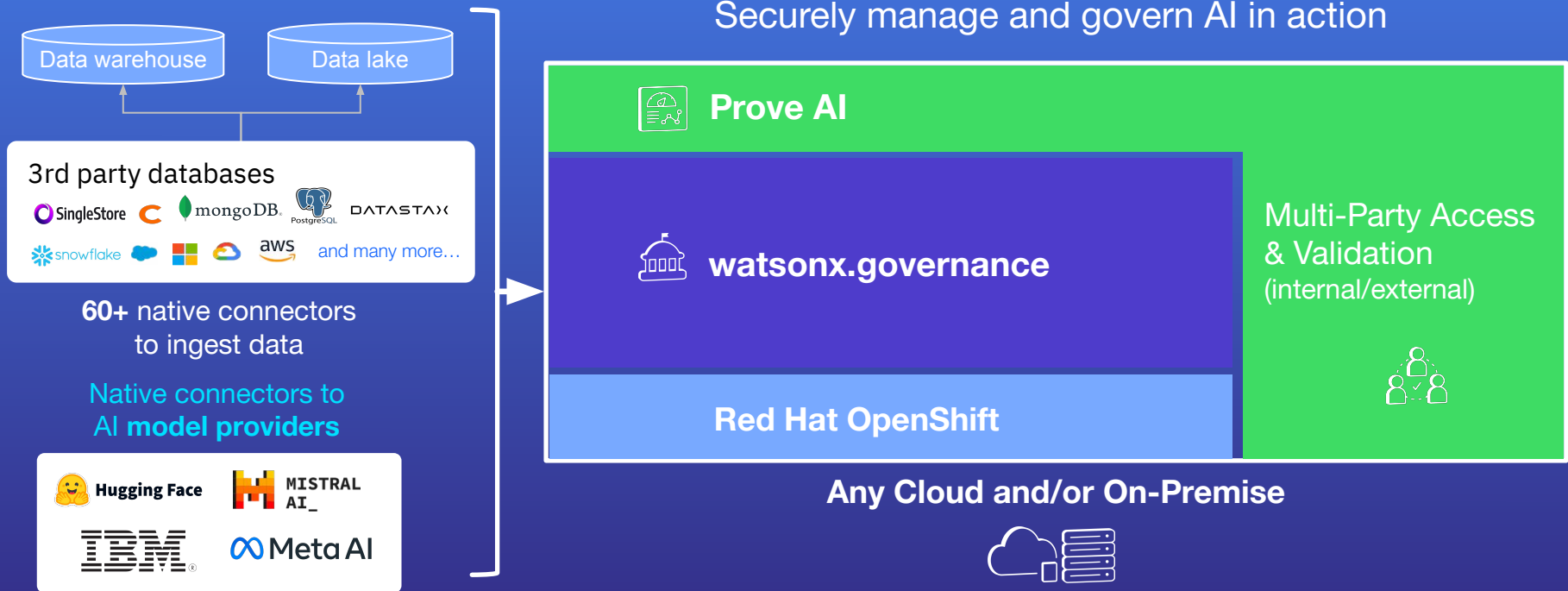
Version Control



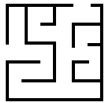
Multi-Party Attestation



Better together: an End-to-End AI + Data Governance Platform



AI governance is complicated



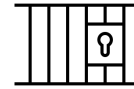
AI governance collaboration requires lots of **manual work**; amplified by changes in data and model versions.



Companies have models in **multiple tools, applications and platforms**, developed inside and outside the organization



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



The **lack of tools for collaboration and communication** impacts stakeholder management

Given the rapidly-increasing regulation of AI, it's critically important that any governance solution can manage risk across the enterprise

Risks and governance requirements differ by:

- Use Case
- Industry
- Geography
- Company
- Technology used



Your governance solution needs adjust to your specific situation:

- Risk assessment
- Governance workflows
- Dashboards
- Model metadata
- Monitoring metrics

watsonx.governance

Accelerate responsible, transparent, and explainable AI workflows

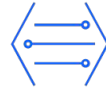
One unified, integrated AI governance platform to govern generative AI and predictive machine learning (ML)



Compliance

Manage AI to meet upcoming safety and transparency regulations and policies worldwide—a “nutrition label” for AI

31



Risk

Proactively detect and mitigate risk, monitoring for fairness, bias, drift, and new LLM metrics



Lifecycle Governance

Manage, monitor and govern AI models from IBM, open-source communities and other model providers

Comprehensive

Govern the end-to-end AI lifecycle with metadata capture at each stage

Open Platform, Open Capabilities

Support an OpenShift platform with governance of models built and deployed in 3rd party tools.

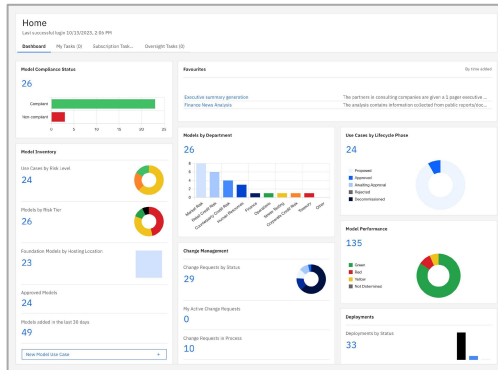
Automated

with automatic documentation and model evaluation

Respond to the growing and changing AI regulation

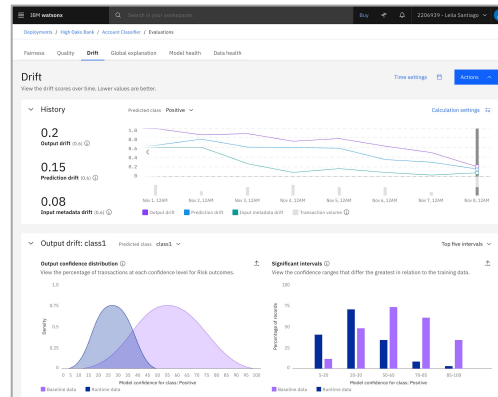
Compliance: satisfy AI regulations

- Translate external AI regulations into enforceable policies for automated enforcement.
- Provide core services to help adhere to external AI regulations for audit and compliance
- Use factsheets for transparent model processes



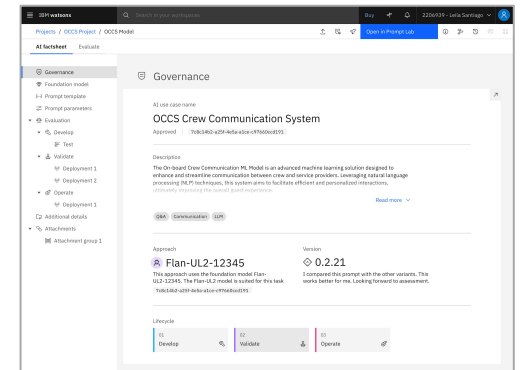
Trust: manage risk and protect reputation

- Manage by exception when key metrics are breached
- Identify, manage and report on risk and compliance at scale
- Provide explainable model results in support of audits and to avoid fines



Lifecycle governance: operationalize AI with confidence

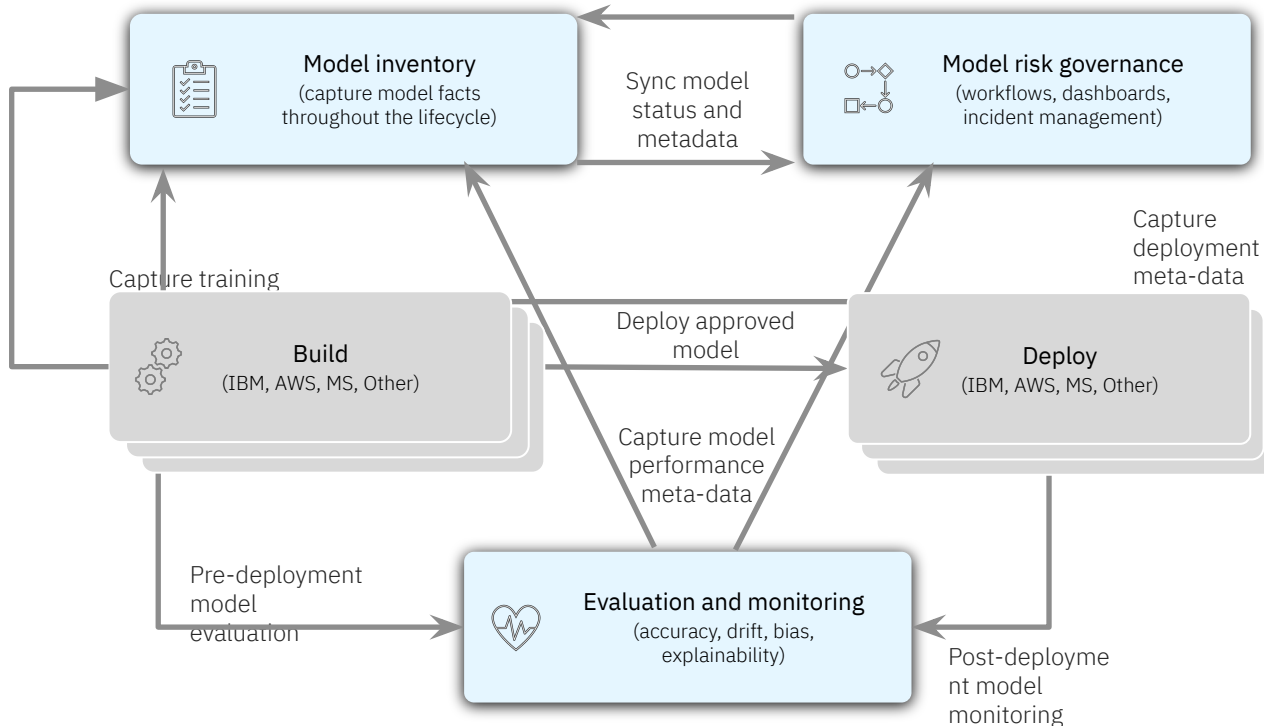
- Monitor, catalog, and govern models across the AI lifecycle
- Automate the capture of model metadata for to facilitate management and compliance
- Oversee model performance across the entire organization with dynamic dashboards and dimensional reporting



What IBM offers

watsonx.governance

Trusted: Accelerate responsible, transparent, and explainable AI workflows



A toolkit for AI governance

- Govern generative AI and traditional ML models across the entire AI lifecycle
- Automate and consolidate multiple tools, applications & platforms while documenting the origin of data sets, models meta data, and pipelines
- Manage risk and protect reputation by automating workflows to better detect fairness, bias, and drift
- Improve adherence to AI regulations, such as the proposed EU AI Act, and internal compliance standards



Manage lifecycle and risk

Simplify data governance, risk management and regulatory compliance

Fully-customizable dashboards for model status across the entire enterprise

A highly scalable, AI-powered, and unified GRC platform

Risk and Compliance

The following regulations may be applicable:

- EU AI Act
- Canada Directive on Automated Decision-Making

The following risk categories may be applicable:

- Fairness
- Drift
- Explainability

Risk and Compliance Assessments

Name	Description
<input type="checkbox"/> Regulatory Applicability Assessment	2023-09-18 - Use case regulato High Oaks Bank
<input type="checkbox"/> Risk Metric Elicitation	2023-09-18 - Use case risk me High Oaks Bank

Models

Deployments Model Links

Search

Name	Description	Model Owner	Model Class	Model Status	Third Party Link	Tags
Executive Summary Generation (LLM) (MOD-00034) High Oaks Bank	Model is a fine-tuned version of google/mt5-small foundational LLM optimized for summarizing more	Mark Owens	Fine-tuned	Pre Implementation Review		

Associated Foundation Models


Search

Name	Description	License	Model Status	Hosting Location	Model Provider	Tags
mt5-small (FM-00015) Library > MRG > Foundation Models	mt5 is pretrained on the mC4 corpus, covering 101 languages	Apache 2.0	Approved for Deployment	External	Hugging Face	

Models by Provider


Foundation Models

19



Non Foundation Models

34



Manage lifecycle and risk

Fully customize approval workflows and notifications to foster collaboration between data science teams, AI developers, and business stakeholders

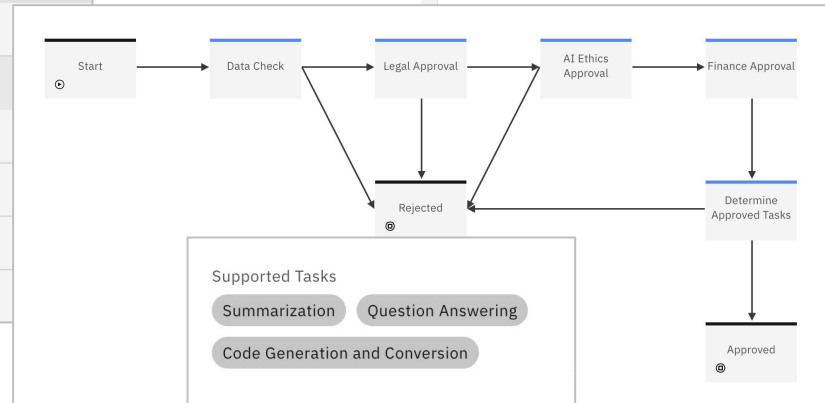
Automatically updated with data from Factsheets and OpenScale, reducing manual labor for regulatory compliance and improving time-to-value for AI projects

FM Usage Cost by Department

41



<input type="checkbox"/>	Name	Description	Value	Breach Status	Value Date	Tags
<input type="checkbox"/>	MET_0000001 Library > MRG > Foundation Models	ROUGE Score	82.14	Green	9/18/2023	
<input type="checkbox"/>	MET_0000002 Library > MRG > Foundation Models	BLEU Score	0.6523	Green	9/18/2023	
<input type="checkbox"/>	MET_0000003 Library > MRG > Foundation Models	Toxicity Score	0.87	Green	9/18/2023	
<input type="checkbox"/>	MET_0000004 Library > MRG > Foundation Models	CorefE	0.67	Yellow	9/18/2023	
<input type="checkbox"/>	MET_0000005 Library > MRG > Foundation Models	LinkE	0.4298	Red	9/18/2023	
<input type="checkbox"/>	MET_0000006 Library > MRG > Foundation Models	Coherence	0.7852	Green	9/18/2023	



Prove AI: a new standard for AI governance

A next generation secure synchronous data + AI governance platform

Key differentiators beyond IBM watsonx and other platforms



**Multi-Party Access and
Accountability**



**Tamper-proof
Auditability**



**Version Control
Automation**

De-risking AI; enabling companies to build with confidence

Open

Designed with interoperability at its core. Any cloud, any model, any data.

Auditable

Delivers real time, tamperproof audit trail of data inputs and model contexts.

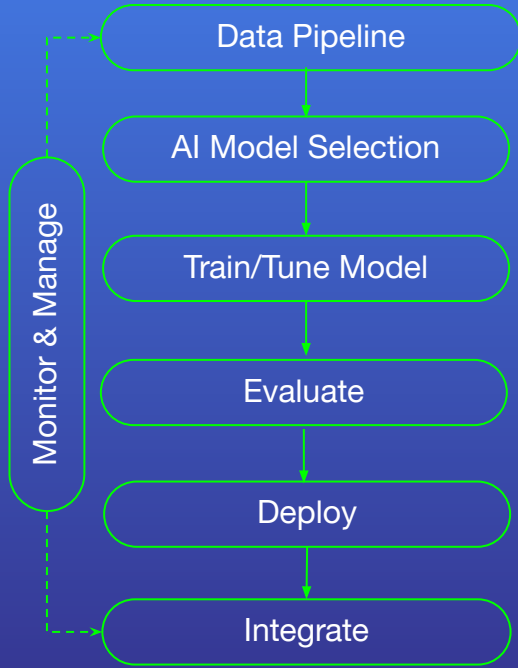
Private

Impenetrable version management with enhanced data management.

Secure

Built on a highly secure and controlled environment.

Prove AI at a glance



Prove AI Frontend GUI

Data Registration (automated + IBM ...)

Model Registration (automated + IBM ...)

Audit History

Issue Resolution

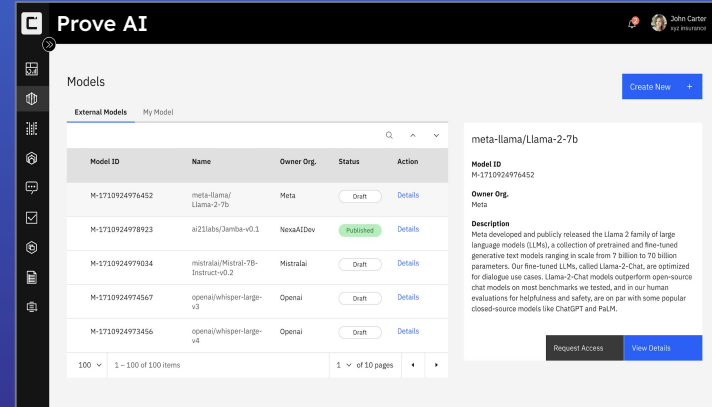
Version Control & Rollback

Prompt Playground

Third Party Attestation

Hybrid Blockchain Integration

A tamper-proof, serialized, and automated immutable source of truth for AI systems with built-in secure 3rd party access, assessment, and attestation.



Case Study: AI-powered food inspection

Trusted assurance for AI-enabled packaged food safety inspection with verified results

Challenges

- Complex, highly variable food safety quality across production
- Trust of computer-vision powered detection of metal, plastic, glass, etc. in packaged foods
- Variable results due to calibrations
- Certified inspection of results that is secure, auditable, & timestamped

Why Prove AI

- Tamper-proof, serialized, & automated secure datastore
- Multi party AI model 360-view & validation of results
- Enhanced AI data security & data provenance insights
- Dynamic hybrid enterprise blockchain backend

Impact

- End customer food distributors' assurance of inspection & results
- More accurate detection
- Certifiable trust of AI model performance & detection
- Drastically lowered packaged goods detection overhead costs



Q&A





Become a prove ai
design partner

casperlabs.io



Thank you!

casperlabs.io