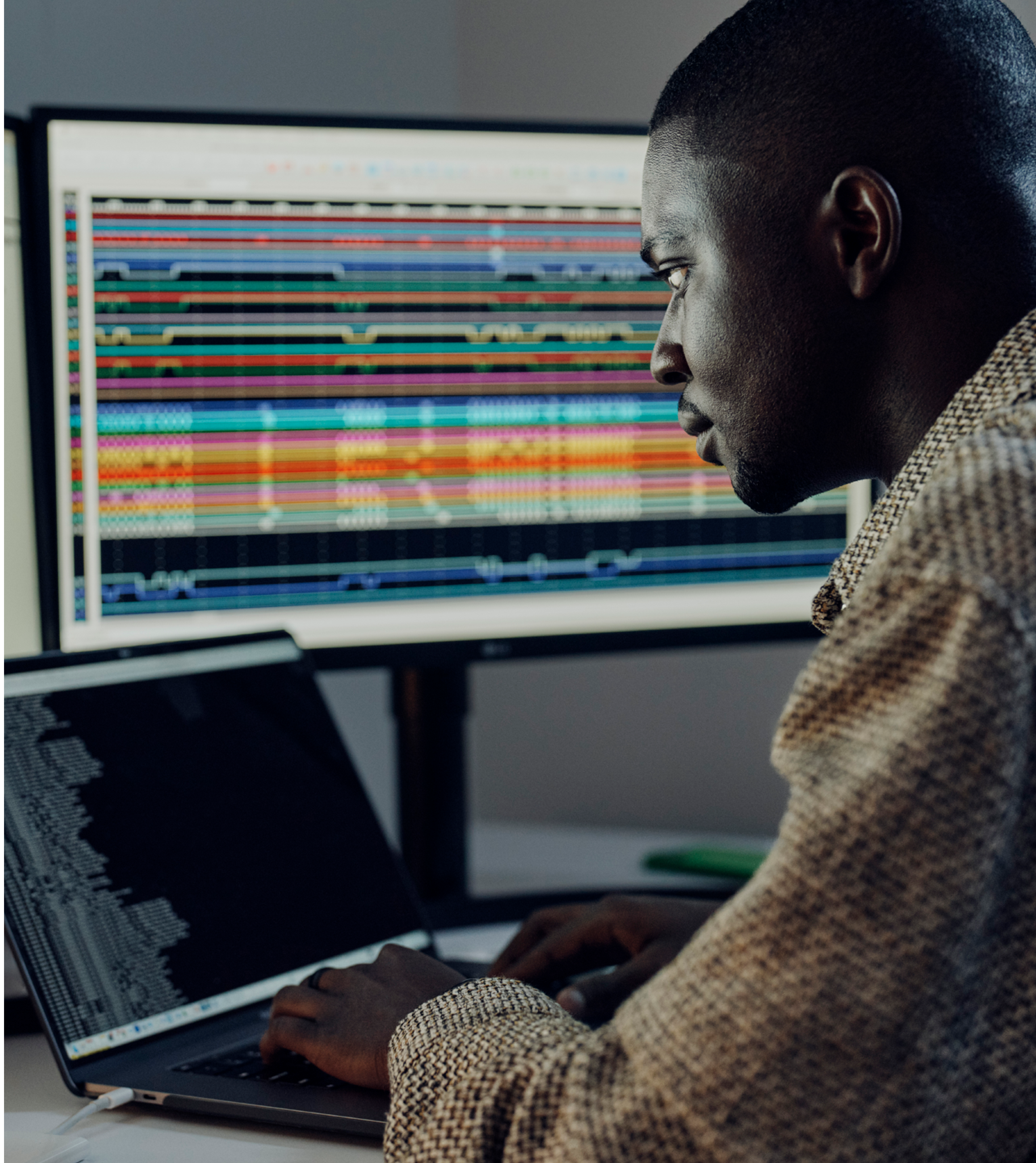
watsonx.governance™

# AI for business:

## Maximize AI ROI through smarter governance



# Contents

# 01 Introduction

## AI governance is critical for scalability

As generative AI (gen AI) becomes the new reality, businesses are racing ahead with AI-powered innovations. But the question remains: Is your AI being adequately governed?

Answering that requires embedding security and resilience by design into an organization's very DNA, not just stating it in policy. This means providing continuous, demonstrable evidence that controls are operating as intended, rather than relying solely on annual compliance checks. Such continuous assurance is essential to meeting today's regulatory demands and navigate an ever-shifting risk landscape.

Governance is the key to ensuring that all innovative AI-backed ideas for your business are on the right path and comply with ethical and regulatory standards across the globe.

With governance as your safety net, there's no reason to hold back from embracing the full potential of AI.

This ebook will help you learn more about the principles of AI governance and set your business on the fast track.

Read the full story or try watsonx.governance at no charge →

AI-related risks are on the rise—compliance and regulatory issues, data bias and reliability concerns, and a growing loss of trust when users don’t understand how AI models operate or are governed.

**AI agents are the future**
As the pace of digital transformation accelerates, enterprises are turning to AI agents as the next evolution of intelligent automation. According to a Gartner survey, 83%[1] of respondents expect AI agents to improve process efficiency and output by 2026. Whereas, an IBM survey states that 71%[2] believe AI agents will autonomously adapt to changing workflows.

Gartner predicts at least 15%[3] of day-to-day work decisions will be made autonomously by agentic AI by 2028, up from 0% in 2024.

Executives across the C-suite admit that their organizations need to do better.

60%

60% of CEOs say they’re mandating additional AI policies to mitigate risk.[4]

63%

While 63% of CROs and CFOs say they focus on regulatory and compliance risks, only 29% believe these risks have been sufficiently addressed.[4]

27%

Roughly 27% of public companies cited AI regulation as a risk in recent filings with the Security and Exchange Commission (SEC).[5]

# 02 Challenges of scaling AI

What's holding organizations back? In a word: trust. Executives cite cybersecurity, privacy and accuracy as the top barriers to implementing gen AI. As the landscape evolves, they expect to increase investments in AI ethics by at least 40% over[6] the next three years.

**Navigating AI governance: The current hurdles**
The AI governance landscape is equipped with a range of tools, yet many models grapple with transparency, consistent monitoring and accurate cataloging during development. The absence of a comprehensive, automated, end-to-end lifecycle management system often hampers scalability and introduces operational opacity. The pursuit of explainable AI outcomes remains elusive, especially with the rise of black box models. These models, while widely deployed, often obscure the logic behind their outputs—even from the developers who built them.

This lack of governance can lead to several inefficiencies. For example, it may contribute to scope creep, hinder timely model deployment, yield inconsistent model quality and lead to unidentified risks. Given the complexities of AI development and deployment, implementing robust, transparent and automated governance frameworks is essential to mitigate these potential issues.

Get IDC's perspective on the main barriers to scaling AI

Read now →

Navigating the complex landscape of risk and reputation management in the realm of AI is daunting. Headlines continue to highlight the dangers of opaque AI systems, which can yield unjust, inexplicable or biased outcomes when deployed in real-world scenarios. These flawed results—often influenced by hidden biases related to race, gender or age—can have far-reaching consequences, impacting both customers and the integrity of your brand.

For instance, consider the high stakes in sectors such as healthcare. AI systems that influence patient diagnoses or treatment plans must be transparent and fair. An incorrect or biased AI recommendation could lead to misdiagnosis or inappropriate treatment, with potentially life-threatening outcomes.

To minimize AI-related risks, organizations must focus on building systems that are transparent, fair and inclusive. Explainable AI plays a key role in detecting and preventing biased decision-making, while also reinforcing privacy, security and customer trust. Building AI that is trustworthy and unbiased is critical not only to enhancing operations but also to avoiding controversy and reputational damage.

Learn more about AI risk management →

**Adapting to the evolving AI regulatory landscape**
Successful AI adoption requires adherence to laws and regulations—local, regional and national—which are evolving at a rapid pace. Non-compliance could cost your organization tens of millions of dollars in fines,[7] as demonstrated by some of the most stringent AI regulations currently debated globally. The current draft of the EU AI Act, for example, contemplates penalties of up to EUR 35 million or 7% of a company's global revenue.[8]

Model documentation is critical, yet it's an area that's often overlooked by data scientists under time pressure, especially in organizations that lack clear governance requirements.

Organizations must not overlook this step; new regulations will require comprehensive model documentation, including metadata and lineage.

Learn how to streamline AI compliance

Read our blog →

## New

Emerging areas intrinsic to agentic AI

**Risks**

- Unsupervised autonomy
- Data bias
- Redundant actions
- Attack on AI agent's external resources
- 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

Discover how to unlock the power of agentic AI and mitigate risks

Learn more →

# All AI needs governance

Governance is indispensable for all AI, including unsupervised agents. Despite the absence of labeled data, these agents require oversight to ensure ethical, unbiased behavior, fostering trust and reliability in AI applications.

Consider an unsupervised AI agent tasked with customer segmentation for targeted marketing. Without proper governance, the agent could inadvertently cluster customers based on sensitive attributes—such as race or income—leading to potentially discriminatory practices.

Governance measures could include:

1. **Algorithmic auditing:** Regularly reviewing the agent's clustering process to ensure it doesn't rely on protected attributes
2. **Fairness metrics:** Implementing metrics to assess the agent's output for any signs of bias or discrimination
3. **Human-in-the-loop:** Including human oversight to validate and, when necessary, adjust the agent's decisions

By incorporating these governance strategies, you can mitigate the risk of unintentional bias and ensure the fair and effective operation of the unsupervised AI agent.

Explore how IBM helps govern agentic AI

Learn more →

**Generative models**
Gen AI models include both foundation models (FMs) and large language models (LLMs). These models have the potential to unlock trillions in economic value,[9] because they boost productivity with their remarkable performance and can adapt to a wide range of tasks.

Such models are highly customizable, scalable and cost-effective. They can query massive volumes of data—and they continue learning in real time. Off-the-shelf generative applications require minimal expertise and can help eliminate many tedious, time-consuming tasks.

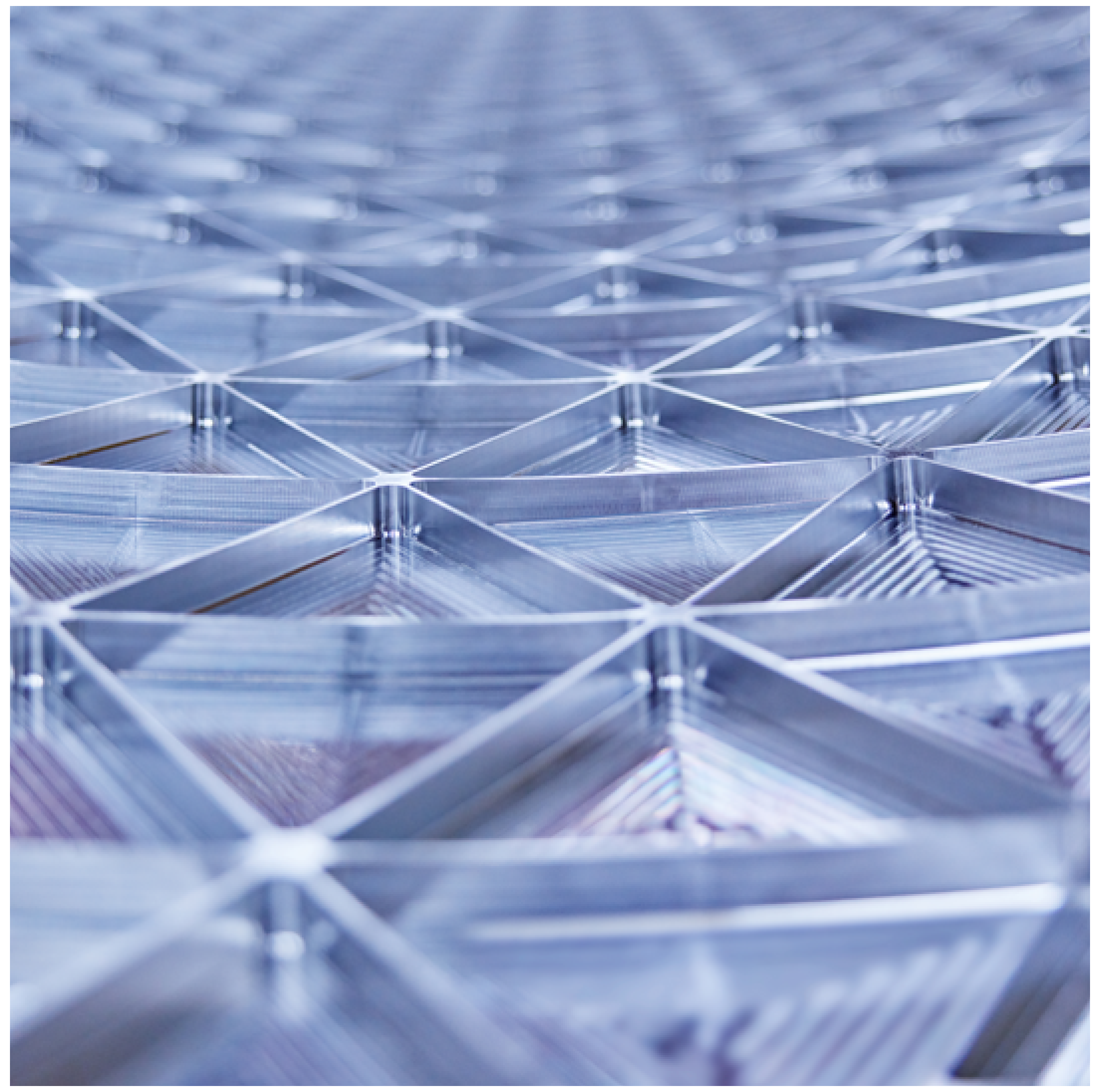
In statistics, generative models have long been used to analyze numerical data.[10] But with the rise of deep learning, their capabilities have expanded to include the generation of images, music, speech, video, text and even code. Use cases now span industries—from marketing and customer service to retail and education.

While generative models have pushed AI to the top of many business agendas, their capabilities introduce new complexities that can pose risks for organizations and for society alike.

Learn how to scale AI responsibly

Read the blog →

# 04 Holistic AI governance



Like any other initiative, successful AI governance depends on the intersection of people, process and technology.

To implement AI properly, you need a strong cross-functional team. AI is increasingly recognized as a strategic priority for many leaders, and the number of team members involved in adopting AI appears to be growing by the day. Some of these individuals may be new to the concept of the AI lifecycle, while others are finding renewed motivation to engage in AI initiatives. It's important to meet the needs of all these groups without overburdening your data scientists, who often lack the time to route or manage approvals and respond to requests for information.

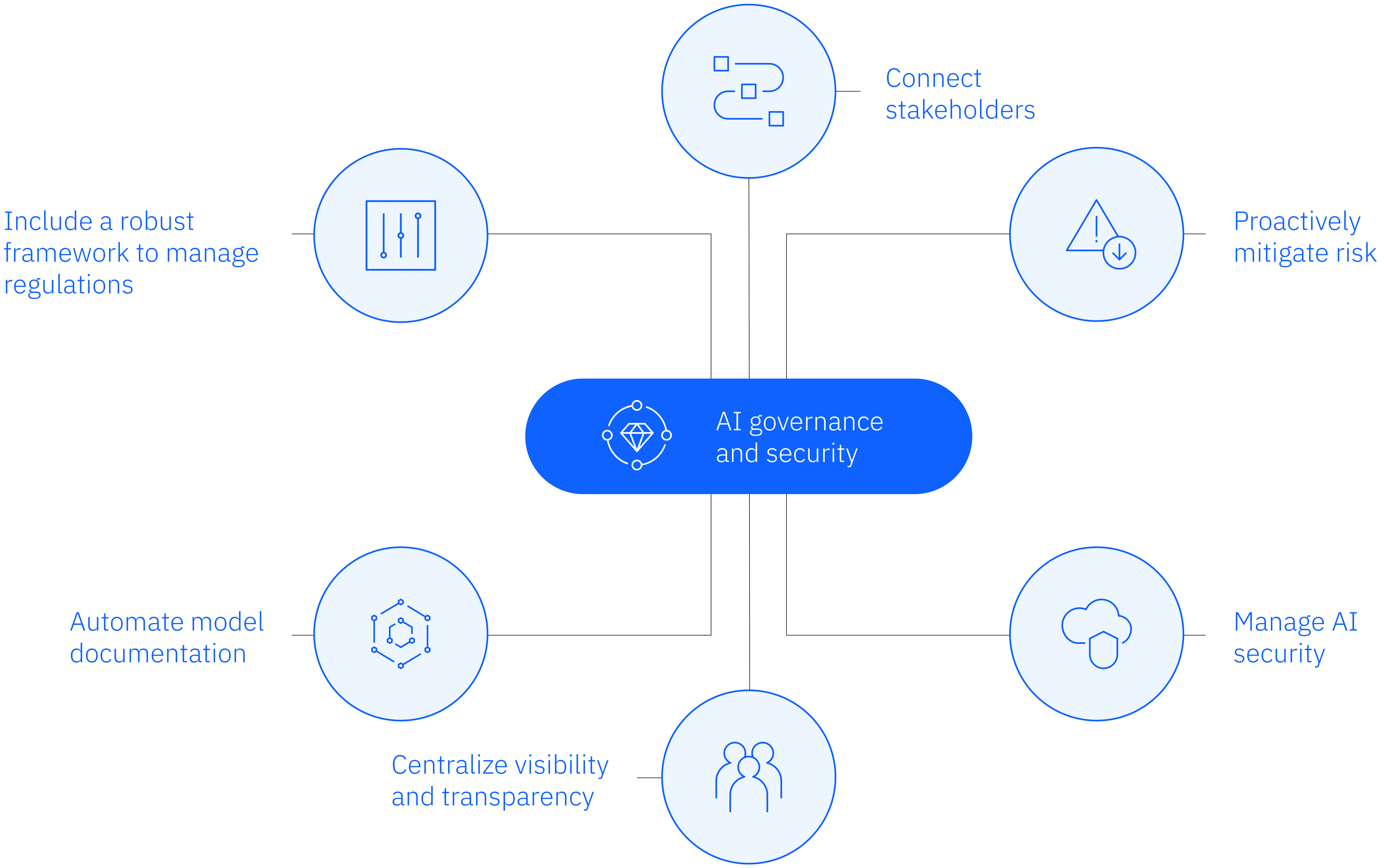
Start by aligning your stakeholders and securing buy-in from key parties. Then, involve them in ideation and build consensus around outcomes and the adoption of responsible AI.

Take steps to ensure that the correct set of metrics, KPIs and objectives are defined in accordance with your company's existing business controls and regulatory frameworks. Finally, monitor the specific metrics identified for your AI models.

Learn how to build a holistic approach to AI governance

Read the blog →

# Manage the complexity of AI governance



AI governance and security

Connect stakeholders

Include a robust framework to manage regulations

Proactively mitigate risk

Automate model documentation

Manage AI security

Centralize visibility and transparency

The principles are supported by the Pillars of Trust, our foundational properties for AI ethics.

**Explainability**
Good design does not sacrifice transparency in creating a seamless experience.

**Fairness**
Properly calibrated, AI can assist humans in making choices more fairly.

**Robustness**
As systems are employed to make crucial decisions, AI must be secure and robust.

**Transparency**
Transparency reinforces trust, and the best way to promote transparency is through disclosure.

**Fairness**
AI systems must prioritize and safeguard consumers' privacy and data rights.

**Process**
AI governance involves tracing and documenting the origin of data, associated models and metadata, and the overall data pipelines for audit. Your documentation should include the techniques used to train each model, the hyperparameters used and the metrics gathered during testing phases. This level of detail increases transparency and gives stakeholders visibility into the model's behavior throughout its lifecycle, including the data that influenced its development and the potential risk it may pose.

Start by benchmarking and evaluating your organization's current AI technologies and processes. Some processes and stakeholders may already be aligned and can be extended, while others may need to be replaced. Next, create a set of automated governance workflows that align with your compliance requirements.

New and existing AI models can adopt these workflows, preventing the process delays mentioned above. Finally, set up a monitoring framework to alert owners and users when a model's metrics exceed acceptable thresholds.

**How to start?**

- Streamline AI management, monitoring and governance across models, apps and agents. Enhance predictions by proactively identifying bias, drift and retraining needs. Improve asset quality, transparency and explainability, and reduce risk.
- Boost AI speed to help businesses scale operations and automation, while ensuring transparent, explainable results that are unbiased and free from drift.
- Leverage risk management for scalable risk identification, control, tracking and reporting. Proactively detect and fix bias, drift and behavioral shifts through model retraining or reconstruction.
- Evaluate multiple AI assets simultaneously to accelerate time to production and reduce the manual workload for developers and data scientists.
- Gain visibility into security vulnerabilities, misconfigurations and risk metrics, with preset alerts that flag unauthorized shadow AI deployments before they escalate.
- Simplify compliance by using a preconfigured list of AI regulations, reducing the time spent identifying obligations and managing noncompliance risks. Translate external regulations into automated enforcement and enhance audit and reporting compliance with fact sheet documentation.
- Document dataset origins, model metadata and pipelines using automated fact sheets. This automation collects model facts, which can free up the data science team for other important tasks while offering audit and litigation support.

Start with watsonx.governance →

## A framework for responsible, governed AI

| | Transparency and visibility | AI risk and security management | Regulatory compliance | Automation |
|---|---|---|---|---|
| Plan | Define measurable performance metrics for AI usage across your organization | Review existing processes that monitor fairness and explainability | Conduct a gap analysis against current and potential AI regulations | Review existing skills and demand for responsible AI, and align them with business objectives |
| Build | Establish traceability and auditability of current processes | Operationalize updated processes and checkpoints throughout the AI lifecycle | Ensure model documentation is accessible | Specify the new roles, skills and learning agendas required to implement responsible AI |
| Create | Create automatic documentation of model lineage and metadata. Track agent behavior during production, identify anomalies and assess performance metrics | Implement automated trigger alerts to detect key AI risk indicators—such as model drift, hallucinations and shadow AI—to proactively mitigate potential risks and security threats | Strengthen regulatory compliance for data science teams without additional overhead | Establish a repeatable, end-to-end workflow with built-in stakeholder approvals to reduce risk and increase scale |

05

# watsonx.governance for responsible, transparent and explainable AI

Meet IBM watsonx.governance®—a powerful AI governance toolkit designed to direct, manage and monitor your AI initiatives, helping you reduce risk, address compliance obligations and maximize ROI from your AI investments.

Built on IBM® watsonx®, this toolkit uses software automation to bolster regulatory compliance and support ethical AI practices. It offers comprehensive governance without requiring costly platform migrations. In the preproduction phase, IBM watsonx.governance® validates business risks. Post-deployment, it continuously monitors fairness, quality and model drift to ensure compliance. Auditors gain access to model behavior insights and prediction explanations, while teams benefit from visibility into model function and training details. Spanning the entire AI lifecycle, watsonx.governance assists teams across design, development, deployment and monitoring, with centralized documentation of AI facts. It simplifies audits with traceability across data, models, metadata and pipelines, and includes documentation of training techniques, hyperparameters and test metrics. Expect enhanced transparency into model behavior, greater insight into influential data and proactive identification of risks.

IBM was named a Leader in the 2024 IDC MarketScape for worldwide machine learning operations (ML Ops).

Read the report →

## Consider these components:

**Regulatory compliance**
Streamline your AI regulatory compliance process with automation and intelligence:

- Build transparent model processes
- Access a single repository of regulatory content
- Map AI use cases and projects to global regulations
- Accelerate documentation of regulatory obligations
- Improve the compliance assessment cycle
- Enhance collaboration and reduce manual processing time

Learn more

**Risk and security management**
Proactively detect and mitigate risks and monitor fairness, bias drift and new LLM metrics:

- Automatically identify unregistered AI deployments and trigger appropriate actions
- Gain visibility into security vulnerabilities, misconfigurations and risk metrics
- Unify security policy creation across risk, compliance and security stakeholders

Learn more

**Lifecycle governance**
watsonx.governance is open and platform-agnostic. You can govern any AI models, applications or agents that are built and deployed using IBM or third-party platforms such as OpenAI, Amazon and others.

- Evaluate multiple AI assets (models, apps or agents) in a single instance
- Streamline agent tool selection and monitor agent performance using advanced RAG metrics
- Track AI assets across their lifecycle in real time

Learn more

Regardless of where you intend to deploy AI in your business, effective governance drives ROI across use cases by aligning AI initiatives with enterprise goals.

**Improve customer experience (CX) chatbots**
Monitor chats for red flags such as toxicity, disclosure of personal information or off-topic responses using AI guardrails

**Enhance BPO processes**
Monitor natural language understanding (NLU) text models for drift, relevancy and more, through hands-off monitoring

**Avoid unwanted compliance costs**
Automate the compliance process using a pre-built library of global regulations, including the EU AI Act, ISO 42001, NIST AI RMF and more

**Build efficient RAG-based bots**
Analyze prompt template evaluation results for RAG tasks with built-in root cause analysis

**Automate the audit process**
Streamline the audit process to meet regulatory requirements by automatically capturing detailed, contextual information about AI risks relevant to your business

**Safeguard HR processes from potential business risks**
Detect and mitigate bias in ML models used for hiring decisions about AI risks relevant to your business

Vice President, Integrated Governance and Market Readiness, IBM

**Accelerating innovation through centralized AI governance**
Building on IBM's foundational principles for Trust and Transparency, the Office of Privacy and Responsible Technology (OPRT) launched the Privacy and AI Management System (PIMS) to help manage machine learning models reliably, comply with privacy and AI regulations, and promote transparency and accountability.

To advance their AI governance journey, OPRT developed the Integrated Governance Program (IGP), a unified approach to responsibility and compliance that integrates technologies such as watsonx.governance, IBM Cloud Pak® for Data, IBM Knowledge Catalog and IBM OpenPages®. This holistic view of IBM's data and models enables proactive risk management, regulatory compliance alignment, scaled governance workflows and uniform internal data standards for transparency and trustworthy AI, resulting in:

- 58% reduction in data clearance request processing time for third-party data
- 62% reduction in data clearance request processing time for IBM-proprietary data
- More than 1,000 datasets and models approved for potential re-use.[11]

07

# Next steps

See how quickly you can create responsible, transparent and explainable AI workflows with the watsonx.governance toolkit—without the costs of switching from your current data science platform. With watsonx.governance, you can:

- Manage risk and secure AI deployments
- Stay ahead of the evolving AI regulatory landscape
- Increase transparency and visibility into your AI use cases to drive greater ROI and faster time to market on AI initiatives

## Get started

Meet watsonx.governance →

Request a demo →

Try for free →

1. Gartner Predicts Over 40% of Agentic AI Projects Will Be Canceled by End of 2027, Gartner, 25 June 2025.

2. IBM Study: Businesses View AI Agents as Essential, Not Just Experimental, IBM Newsroom, 10 June 2025.

3. Top Strategic Technology Trends for 2025: Agentic AI, Gartner, 21 October 2024.

4. The CEO’s guide to generative AI: Risk management, IBM Institute for Business Value, 12 August 2024.

5. AI Regulation Is Coming. Fortune 500 Companies Are Bracing for Impact. The Wall Street Journal, 27 August 2024.

6. Generative AI: The state of the market, IBM Institute for Business Value, 25 May 2023.

7. EU's AI Act Takes Effect: Non-Compliance Could Cost Tens of Millions of Euros, aiexpoeurope.com, August 2024.

8. What is the Artificial Intelligence Act of the European Union (EU AI Act)? IBM, 20 September 2024.

9. The economic potential of generative AI: The next productivity frontier, McKinsey, 14 June 2023.

10. What is generative AI? IBM Research, 20 April 2023.

11. Accelerating innovation through centralized AI governance.

This document is current as of the initial date of publication and may be changed by IBM at any time.

Not all offerings are available in every country in which IBM operates.

Examples presented as illustrative only. Actual results will vary based on client configurations and conditions and, therefore, generally expected results cannot be provided.

It is the user’s responsibility to verify the operation of any non-IBM products or programs with IBM products and programs. IBM is not responsible for non-IBM products and programs.

THE INFORMATION IN THIS DOCUMENT IS PROVIDED “AS IS” WITHOUT ANY WARRANTY, EXPRESS OR IMPLIED, INCLUDING WITHOUT ANY WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND ANY WARRANTY OR CONDITION OF NON-INFRINGEMENT. IBM products are warranted according to the terms and conditions of the agreements under which they are provided.

No IT system or product should be considered completely secure, and no single product, service or security measure can be completely effective in preventing improper use or access. IBM does not warrant that any systems, products or services are immune from, or will make your enterprise immune from, the malicious or illegal conduct of any party.

The client is responsible for ensuring compliance with all applicable laws and regulations. IBM does not provide legal advice nor represent or warrant that its services or products will ensure that the client is in compliance with any law or regulation.