SPOTLIGHT

Sponsored by: IBM

AI has huge potential to unlock value — but it can also create unexpected and undesirable outcomes. AI governance helps ensure ethical AI development, fostering customer trust and loyalty.

Evolving Regulations and Emergence of Agentic AI Fuel AI Governance Imperative

September 2025

Written by: Nancy Gohring, Senior Research Director, AI

Introduction

AI is transforming a diverse range of industries, from finance and manufacturing to agriculture and healthcare, by enhancing operations and reshaping the nature of work. It enables smarter fleet management and logistics, optimizing energy forecasting, creating more efficient use of hospital beds, improving quality control in advanced manufacturing, and creating personalized consumer experiences. Governments are also adopting AI because of its ability to deliver better service to citizens at a lower cost to taxpayers. Enterprises' application of generative AI (GenAI) can revolutionize customer experiences, boost employee productivity, enhance creativity and content creation, and accelerate process optimization.

What a difference a year makes. The emergence of agentic AI, which determines how to accomplish multistep work and selects the right tools to execute actions, promises a new level of autonomy poised to deliver even greater impact to organizations. AI agents can expand the utility of automation technologies and broaden the aperture of use cases, handling nearly entire workflows such as processing mortgage applications, onboarding employees, and developing sophisticated software.

AT A GLANCE

KEY STATS

» IDC estimates that by 2030, every $1 spent on AI solutions and services will generate $4.9 into the worldwide economy.

» Security breaches, customer data exposure, and regulatory risks are top AI concerns for the enterprise.

WHAT'S IMPORTANT

A clear strategy and set of goals are necessary to properly launch and support evolving AI governance requirements.

KEY TAKEAWAY

AI governance ensures ethical AI development and deployment.

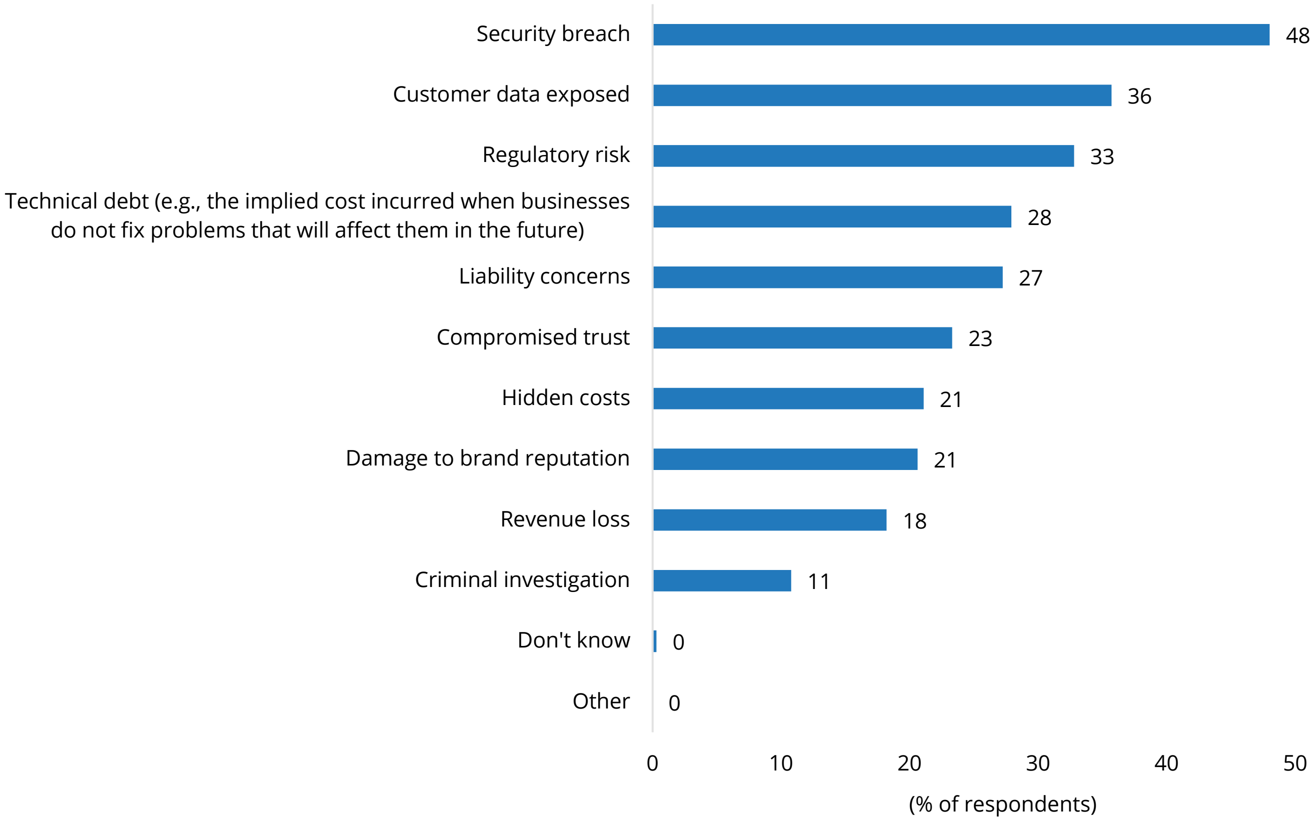
IDC anticipates that the advent of agents will boost the potential of AI to drive economic growth. IDC research estimates that by 2030, for every new $1 spent on AI solutions and services will generate $4.9 into the worldwide economy.

However, AI also creates real risks and unintended consequences. An agentic AI–enabled workflow that is permitted to change customer files, approve a loan, assess applications, and rate candidates can make mistakes, with consequences that could expose organizations to reputational, financial, regulatory, and legal risk. A text generation engine that can convincingly imitate a range of publications is open to misuse; voice imitation software can mimic an individual's speech patterns well enough to convince a bank, workplace, or friend. AI platforms can reinforce and perpetuate historical human biases (e.g., based on gender, race, or sexual orientation), undermine personal rights, compromise data security, produce misinformation and disinformation, destabilize the financial system, and cause other forms of disruption globally.

The stakes are high. According to IDC's June 2025 *AI Tech Buyer Survey,* the main concerns that responsible AI policies protect against include security breaches (48%), customer data exposure (36%), and regulatory risk (33%) (see Figure 1). Responsible AI and AI governance are closely linked, where AI governance describes a broader framework that encompasses and enables responsible AI, which is typically oriented around ethical principles.

FIGURE 1: ***Responsible AI Policy Topics***

Q ***What are the main concerns your responsible AI policy is protecting against?***



n = 2,276 for worldwide

Base = respondents indicated clear principles/rules, policies, and processes and a governance body are currently in place at their organization

Notes:

The data is managed by IDC's Global Primary Research Group.

The data is weighted by IT spending by country.

Multiple responses were allowed.

Use caution when interpreting small sample sizes.

Source: IDC's AI Tech Buyer Survey, June 2025

Legislators and regulators worldwide have developed some frameworks to maximize AI's benefits to society while mitigating its risks. Some regulations are already in effect, most notably the EU AI Act, prohibiting a set of AI applications and related activities such as harmful AI-based manipulation, deception, and social scoring. For other types of AI systems, application providers are required to implement risk assessment and mitigation systems and use high-quality data sets that minimize discriminatory outcomes.

Other regions, such as the United States with its AI Action Plan, have taken less prescriptive approaches, leaving organizations to carefully interpret and apply any guidance, which in some cases may counter state-level regulations. For organizations operating AI applications in the United States or in multiple geographies, navigating layers of regulation is a complex undertaking. To mitigate complexity and drive innovation, enterprises require several key governance components that include:

- Build a governance strategy developed in collaboration with experts in AI, law, compliance, technology, and business priorities. It should include responsible AI principles that define approaches to fairness, inclusiveness, transparence, and accountability.
- Set goals and put in place processes to make sure that the goals are met, in addition to defining and enforcing guidelines and tracking progress.
- Establish a set of clear rules that make it easy for anyone launching new AI initiatives to understand what is acceptable.
- Beyond the guidelines and best practices, put in place processes to clarify anything that the general guidance does not cover.
- Implement tools and systems to roll out initiatives, provision infrastructure and data, measure progress, and ensure that due process is followed throughout. Given how AI learns over time and depends on the data it is fed, clear retention policies to enable auditing past decisions are crucial.

AI governance helps ensure ethical AI development, fostering customer trust and loyalty. Transparent processes improve decision-making, boosting operational efficiency and innovation. Compliance with regulations prevents legal issues and safeguards a company's reputation.

Definition

AI governance refers to the set of policies, frameworks, practices, and tools that help the development, deployment, and use of AI technologies. It involves establishing rules, standards, and ethical guidelines to help organizations implement AI in a responsible and accountable way. AI governance ensures that the quality of the AI implemented is upheld and that all stakeholders can communicate clearly regarding AI best practices.

Benefits

AI governance addresses the potential risks, challenges, and ethical considerations of this technology. It aims to ensure that the development and deployment of AI systems align with societal values, protect user rights, and minimize potential harm. By using AI governance, businesses can continue to adapt to and be prepared for future changes in AI technology and regulation.

AI governance is a combination of leadership, training, communications, guidance, policies, processes, and tools that address the three following areas:

- **Law:** The rules that legal systems enforce

- **Ethics:** The rules that culture and society enforce
- **Regulation:** The rules that governments enforce and the compliance with growing industry standards, especially in highly regulated industries, such as financial services and healthcare

Trends

The AI regulatory landscape is rapidly changing. The EU AI Act is the first comprehensive AI law by a major regulator globally. The law aims to ensure that AI systems are safe and respect the law and the EU's fundamental rights and values. It also assigns AI applications to three risk categories: unacceptable, high, and unregulated.

Organizations building or using AI systems in the EU market or whose system outputs are used within the EU will be responsible for complying with the EU AI Act.

Enterprise obligations depend on the level of risk an AI system poses to people's safety, security, or fundamental rights along the AI value chain. AI systems classified as "high risk," and general-purpose AI system providers determined to be of high impact or posing "systemic risks," will have the most stringent transparency and reporting requirements. Depending on the risk threshold of AI systems, enterprises have some level of responsibility — ranging from classification to risk management to technical documentation and human oversight.

Considering IBM

IBM watsonx.governance is designed to help organizations manage, monitor, and scale their AI with automation, regardless of whether it is on premises, in the cloud, or in a hybrid environment. Key capabilities of the product include:

- Governing AI models, apps, or agents from any vendor, including IBM watsonx.ai, Amazon SageMaker and Bedrock, Google Vertex, and Microsoft Azure (With IBM watsonx.governance now capable of monitoring both development time and runtime metrics, the software can track metrics from quality to faithfulness to drift, regardless of the AI platform in use.)
- Evaluating and monitoring model health, accuracy, drift, bias, and GenAI quality
- Using agentic AI governance to evaluate, monitor, and manage agents during development and in production, with enhanced metrics, risk management, tracing, and a catalog of approved agents and assets with which to deploy or build other agents
- Providing risk and regulatory compliance management capabilities featuring automated risk and compliance assessments and workflows, customizable dashboards to bring stakeholders together, risk scorecards, and reports
- Leveraging an integrated AI security and governance solution that combines watsonx.governance with IBM Guardium AI Security to provide comprehensive penetration testing, monitoring, mitigation, and detection of shadow AI, using continually evolving metrics
- Providing Factsheet capabilities to collect and document model metadata automatically across the AI model life cycle

- Enabling configurable AI guardrails that detect and remove harmful language and personal identifiable information (PII) in inputs and outputs

IBM works with partners like AWS to integrate IBM watsonx.governance with their solutions. With AWS, the product is integrated with Amazon SageMaker — a service for building, training, and deploying ML and GenAI models with fully managed infrastructure, tools, and workflows — to help Amazon SageMaker and IBM watsonx customers govern the entire life cycle, manage risk, monitor model health, and support their compliance obligations with global regulations such as the EU AI Act. This integration rounds out the availability of the watsonx platform in AWS Marketplace, which already includes IBM watsonx.ai and watsonx.data as customer-managed offerings.

One of the most significant challenges faced by large language models (LLMs) is hallucinations, which refer to the phenomenon of generating outputs that are factually incorrect or contextually inappropriate. Organizations need effective strategies and tools to detect hallucinations and mitigate the associated risks. IBM watsonx.governance now supports out-of-the-box evaluation of retrieval-augmented generation (RAG) metrics during development and runtime. The set of metrics to evaluate performance include HAP, PII, prompt injection, context relevance, faithfulness, answer similarity, answer relevance, hit rate, average precision, reciprocal rank, and unsuccessful requests designed to ensure a thorough assessment of a system's effectiveness.

All of these metrics provide a score from 0 to 1, and their combination will help developers and prompt engineers create more accurate and efficient AI use cases while worrying less about hallucinations.

IBM watsonx.governance also enables support of key regulatory requirements out of the box. For example, it supports:

- AI model risk assessments to help users understand which AI risks apply to their use case
- Assessment of AI systems' applicability to the EU AI Act

IBM collaborated with Credo AI to launch Compliance Accelerators to expedite the process of understanding which regulations to comply with for each AI use case and automating the risk and compliance management.

IBM watsonx.governance also continues to forge relationships with key partners such as EY, Deloitte, Infosys, and KPMG to offer various AI governance capabilities. For example, the EY.ai Confidence Index helps end users drive confidence in the data, technology, and processes that form the infrastructure of their AI ecosystems. It supports enhanced decision-making and more efficient operations through reliability and explainability, and it promotes responsible AI by improving transparency and privacy through measurable confidence levels.

Essentially, IBM watsonx.governance provides tools for transparency, explainability, and responsible AI use, making it easier for businesses to deploy and manage AI at scale.

Challenges

Ensuring effective AI governance can be challenging. Building the processes and frameworks to help ensure AI models and applications comply with various regulations, such as the EU AI Act and the evolving landscape, can be complex and time-consuming. Likewise, monitoring AI models for bias, drift, and other risks is crucial but challenging. It requires continuous oversight and sophisticated tools to detect and mitigate these issues.

With the advent of GenAI, organizations' awareness of the risks and rewards of embracing AI responsibly has improved. However, alongside the benefits the technology provides, the attacks and threats from bad actors are rapidly evolving. Agentic AI delivers a new set of challenges, including potentially higher risk levels related to the increasingly complex and extensive work that agents can conduct autonomously. Plus, the process of composing AI agents has become easier for nontechnical users, opening the door to shadow IT that organizations must figure out how to govern. IBM plans to continue to innovate to deliver granular governance of agents and agentic workflows, as well as to inhibit jailbreaks and prompt attacks. It also should continue to partner with innovative start-ups and ISVs to address industry-specific threats and risk management.

Conclusion

AI has huge potential to unlock value — but it can also create unexpected and undesirable outcomes. The additional risks that AI-powered decisions entail mean that AI governance should always accompany them to help drive innovative value through responsible AI adoption.

For company leaders, understanding the core principles underlying AI rules, even if those rules may not currently apply to their company, can instill confidence in customers and regulators in the use of AI, potentially providing a competitive advantage in the marketplace. It can also help companies anticipate the governance needs and compliance requirements that may apply to their development and use of AI, making them more agile.

Based on the identified trends, there are at least three actions businesses can take now to remain a step ahead of the rapidly evolving AI regulatory landscape:

- Understand AI regulations in effect in the markets in which the business operates, aligning internal AI policies with these regulations and any associated supervisory standards
- Establish robust, clear governance, and risk management structures and protocols, as well as accountability mechanisms, where appropriate, to enhance how the business manages AI technologies
- Engage in dialogue with public sector officials and others to better understand the evolving regulatory landscape and to provide information and insights that might be useful to policymakers

For governance approaches to strike the right balance between government oversight and innovation, companies, policymakers, and other stakeholders must engage in open conversations. All these parties are testing the waters and working to find new possibilities that AI is enabling. New rules will be necessary. At this unique moment of possibility and peril, now is the time to cooperate to turn those principles into practice.

About the Analyst

***Nancy Gohring,** Senior Research Director, AI*

Nancy Gohring is a senior research director, coleading IDC's GenAI and Agentic AI Strategies program. Gohring covers big picture trends related to enterprise adoption of AI, including GenAI and agentic AI. Key research themes include business, organizational, and technology architecture transformation in the context of AI and GenAI. As part of the Worldwide AI, Automation, Data, and Analytics Research practice, Gohring supports a range of clients across the technology stack including hyperscalers, developer tool providers, enterprise application vendors, professional services organizations, automation frameworks providers, and infrastructure suppliers.

MESSAGE FROM THE SPONSOR

IBM watsonx.governance accelerates responsible, transparent and explainable AI for both Generative AI (Gen AI) and machine learning (ML) models from any third-party AI model vendor or custom, purpose-built models, regardless of whether the AI is in a cloud, on-premises or hybrid environment. Watsonx.governance helps direct, manage and monitor AI across the end-to-end model lifecycle. Models are monitored to detect and mitigate risk based on pre-determined thresholds for bias, drift and the inputs and outputs for Gen AI models for toxic language, hate speech, or hallucinations. Gen AI models are also monitored for data size, latency and changes in throughput. GRC capabilities provide automated workflows with approvals, access to persona based customizable dashboards/reports, and risk scorecards in support of compliance with the growing and changing AI regulations, industry standards and internal policies. Factsheets automate the tracking and documentation of model metadata across the AI lifecycle in support of stakeholder requests, audits and fines.

Learn more about watsonx.governance at https://www.ibm.com/products/watsonx-governance.

IDC Custom Solutions

The content in this paper was adapted from existing IDC research published on www.idc.com.

IDC Research, Inc.
140 Kendrick Street
Building B
Needham, MA 02494, USA
T 508.872.8200
F 508.935.4015
blogs.idc.com
www.idc.com

IDC Custom Solutions produced this publication. The opinion, analysis, and research results presented herein are drawn from more detailed research and analysis that IDC independently conducted and published, unless specific vendor sponsorship is noted. IDC Custom Solutions makes IDC content available in a wide range of formats for distribution by various companies. This IDC material is licensed for external use, and in no way does the use or publication of IDC research indicate IDC's endorsement of the sponsor's or licensee's products or strategies.

International Data Corporation (IDC) is the premier global provider of market intelligence, advisory services, and events for the information technology, telecommunications, and consumer technology markets. With more than 1,300 analysts worldwide, IDC offers global, regional, and local expertise on technology and industry opportunities and trends in over 110 countries. IDC's analysis and insight helps IT professionals, business executives, and the investment community to make fact-based technology decisions and to achieve their key business objectives.