IDC

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.

AI Governance: An Imperative for Harmonizing Innovation and Responsibility

August 2024

Written by: Ritu Jyoti, GVP/GM, AI, Automation, Data, and Analytics Research

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. AI is enabling 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), which is just beginning to unfold, can revolutionize customer experiences, boost employee productivity, enhance creativity and content creation, and accelerate process optimization. AI is a powerful driver of economic growth. IDC research estimates that generative AI's global economic impact will be close to $10 trillion by the end of 2033. This impact will include increased revenue, lower expenses, and improved productivity.

AT A GLANCE

KEY STATS

- IDC estimates GenAI's economic impact by the end of 2033 to be close to $10 trillion.
- The lack of AI governance and risk management tools and the lack of AI regulations are 2 of the top 5 inhibitors to AI adoption.

WHAT'S IMPORTANT

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

KEY TAKEAWAY

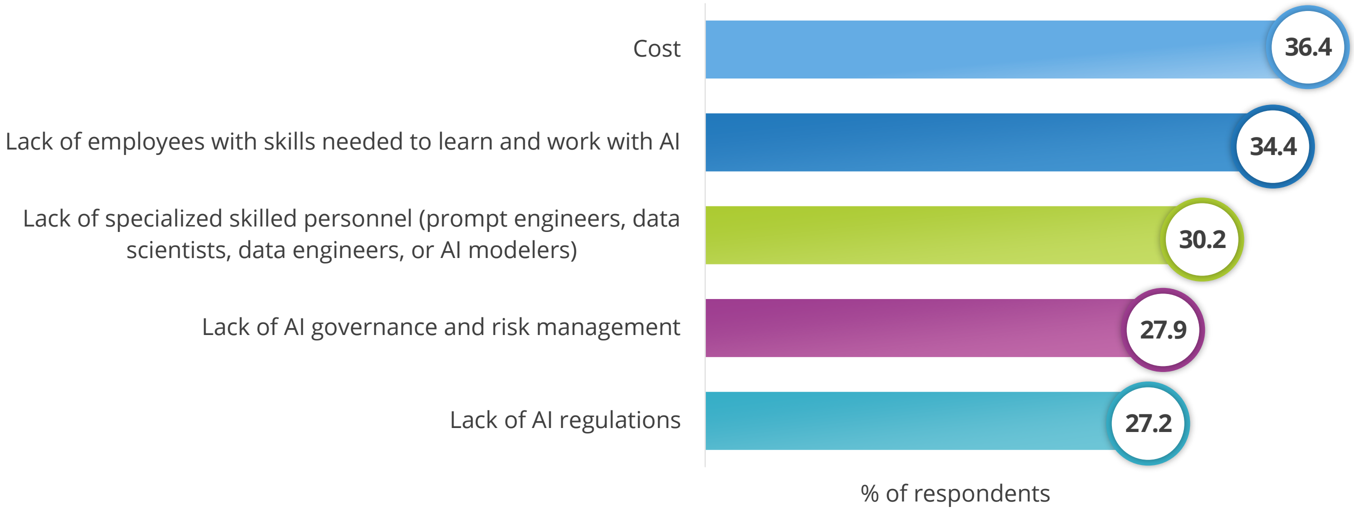
AI governance ensures ethical AI development and deployment.

However, AI also creates real risks and unintended consequences. 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. Chatbots can cheat on tests. 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 October 2023 *Global AI (Including GenAI) Buyer Sentiment, Adoption, and Business Value Survey* (n = 607 worldwide), one-third of the respondents noted the lack of AI governance and risk management tools and the lack of regulations as 2 of the top 5 inhibitors to adopting AI (see Figure 1).

FIGURE 1: ***Inhibitors to AI Adoption***

Q ***What challenges have you experienced/expect to experience when implementing AI technology at your organization?***



n = 607 worldwide

Source: IDC's Global AI (Including GenAI) Buyer Sentiment, Adoption, and Business Value Survey, October 2023

Legislators, regulators, and standard setters are starting to develop frameworks to maximize AI's benefits to society while mitigating AI's risks. These frameworks must be resilient, transparent, and equitable.

Powering innovation requires several key governance components. To properly launch AI governance, clear strategy and goals are necessary. Further:

- Once goals have been set, there is a need for a body of work to ensure these goals are being met in addition to defining and enforcing guidelines and tracking progress.
- Companies need to ensure that AI best practices can be carried out by collating and sharing accessible training resources for everyone who is part of the AI strategy.
- It is necessary to 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, processes must be in place to clarify anything that the general guidance does not cover.

- Systems and tools need to be in place to roll out initiatives, provision infrastructure and data, measure progress, and ensure that due process is followed throughout. Because of how AI learns over time and therefore 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 a platform designed to help organizations manage and monitor their AI activities on cloud or on premises. Key capabilities of the platform include:

- The capability includes the ability to govern GenAI and ML models from any vendor, including IBM watsonx.ai, Amazon SageMaker and Bedrock, Google Vertex, and Microsoft Azure. With the ability of IBM watsonx.governance to now monitor both development time and runtime metrics, the software can monitor metrics from quality to faithfulness to drift, regardless of the AI platform in use.
- The solution can evaluate and monitor model health, accuracy, drift, bias, and GenAI quality.
- The platform provides access to powerful governance, risk, and compliance capabilities featuring workflows with approvals, customizable dashboards, risk scorecards, and reports.
- IBM watsonx.governance provides factsheet capabilities to collect and document model metadata automatically across the AI model life cycle.

IBM and AWS have recently announced the integration of IBM watsonx.governance and 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 manage model risk and support their compliance obligations 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. These new metrics include:

- Faithfulness metrics that measure how faithful the model output is to the provided reference data
- Relevance metrics that measure how relevant the large language model output/response was to the user query
- Unsuccessful request metrics that measure the ratio of unsuccessfully answered questions out of the total number of questions

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

In addition, according to IBM, the IBM watsonx.governance enables the EY.ai Confidence Index, which helps end users drive confidence in the data, technology, and processes that form the infrastructure of the AI ecosystem. 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 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 generative AI, 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. IBM should continue to innovate 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. Fortunately, there is wide agreement among countries on the foundational principles to govern AI use. At this unique moment of possibility and peril, now is the time to cooperate to turn those principles into practice.

About the Analyst

Ritu Jyoti, *GVP/GM, AI, Automation, Data, and Analytics Research*

Ritu Jyoti is group vice president/general manager of Worldwide Artificial Intelligence (AI), Automation, Data, and Analytics Research with IDC's software market research and advisory practice. Ms. Jyoti is responsible for leading the development of IDC's thought leadership for AI research and managing the Worldwide AI, Automation, Data, and Analytics Software Research team. Her research focuses on the state of enterprise AI efforts and global market trends for rapidly evolving AI and ML, including generative AI, innovations, and ecosystems.

MESSAGE FROM THE SPONSOR

IBM watsonx.governance accelerates responsible, transparent and explainable AI for both Generative AI (Gen AI) and machine learning (ML) models developed in IBM watsonx.ai or from any third-party AI model vendor. Deployed on cloud or on-premises, 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
Twitter @IDC
blogs.idc.com
www.idc.com

This publication was produced by IDC Custom Solutions. The opinion, analysis, and research results presented herein are drawn from more detailed research and analysis independently conducted and published by IDC, unless specific vendor sponsorship is noted. IDC Custom Solutions makes IDC content available in a wide range of formats for distribution by various companies. A license to distribute IDC content does not imply endorsement of or opinion about the licensee.

External Publication of IDC Information and Data — Any IDC information that is to be used in advertising, press releases, or promotional materials requires prior written approval from the appropriate IDC Vice President or Country Manager. A draft of the proposed document should accompany any such request. IDC reserves the right to deny approval of external usage for any reason.