MIT Technology Review
Insights

Produced in partnership with

HCLTech

To deliver value from generative AI, businesses must take concrete steps to ensure responsible AI becomes part of the organization's operating model.

Implementing responsible AI in the generative age

Key takeaways

1. **Responsible AI can provide competitive advantage. Most executives say their business will increase investments in building responsible AI in the next 12 months.**
2. **Although 87% of executives believe responsible AI principles are critical to adopt, only 15% of respondents feel very prepared to implement them.**
3. **Half of respondents have a handle on managing operational risks, but less than a quarter feel they can address user adoption, change management, and bias-related risks.**

Many organizations have experimented with AI, but they haven't always gotten the full value from their investments. A host of issues standing in the way center on the accuracy, fairness, and security of AI systems. In response, organizations are actively exploring the principles of responsible AI: the idea that AI systems must be fair, transparent, and beneficial to society for it to be widely adopted.

When responsible AI is done right, it unlocks trust and therefore customer adoption of enterprise AI. According to the US National Institute of Standards and Technology the **essential building blocks** of AI trustworthiness include:

- Validity and reliability
- Safety
- Security and resiliency
- Accountability and transparency
- Explainability and interpretability
- Privacy
- Fairness with mitigation of harmful bias

To investigate the current landscape of responsible AI across the enterprise, MIT Technology Review Insights surveyed 250 business leaders about how they're implementing principles that ensure AI trustworthiness. The poll found that responsible AI is important to executives, with 87% of respondents rating it a high or medium priority for their organization (see Figure 1).

A majority of respondents (76%) also say that responsible AI is a high or medium priority specifically for creating a competitive advantage (see Figure 2). But relatively few have figured out how to turn these ideas into reality. We found that only 15% of those surveyed felt highly prepared to adopt effective responsible AI practices, despite the importance they placed on them (see Figure 3).

Putting responsible AI into practice in the age of generative AI requires a series of best practices that leading companies are adopting. These practices can include cataloging AI models and data and implementing governance controls. Companies may benefit from conducting rigorous assessments, testing, and audits for risk, security, and regulatory compliance. At the same time, they should also empower employees with training at scale and ultimately make responsible AI a leadership priority to ensure their change efforts stick.

"We all know AI is the most influential change in technology that we've seen, but there's a huge disconnect," says Steven Hall, chief AI officer and president of EMEA at ISG, a global technology research and IT advisory firm. "Everybody understands how transformative AI is going to be and wants strong governance, but the operating model and the funding allocated to responsible AI are well below where they need to be given its criticality to the organization."

Methodology

MIT Technology Review Insights conducted a poll on responsible AI in November 2024. The 250 respondents are global executives and senior leaders and represent a broad range of industries.

"It is important for organizations to institutionalize the right AI governance policies and implement tech guardrails to reduce incident risks."

Apoorv Iyer, Executive Vice President and Global Head, Generative AI, HCLTech

The need for responsible AI

Generative AI – and enterprise AI at large – has moved from proof of concept experiments to much wider adoption. Applying the technology to areas such as customer service and software development has the potential to transform entire processes and functions. For example, **ISG Research estimates** that generative AI delivers productivity gains in software development of around 30% to 42% as it automates insights, facilitates robust and error-resistant coding practices, and enhances software quality and security.

Respondents in our survey recognize the importance of responsible AI to unlock the full value of AI in their organizations. Some 87% of respondents said that implementing responsible AI practices is a high or medium priority for their business (see Figure 1). And looking ahead, the vast majority of respondents said that implementing responsible AI across a number of critical business areas will also be a high or medium level priority for them in the next 18 months (see Figure 4).

For AI-fueled transformation to succeed, however, organizations must ensure that they use AI responsibly. "AI has always been one of the techniques we use to create differentiation, but generative AI is now so all encompassing," says Apoorv Iyer, executive vice president and global head of the generative AI practice at HCLTech. "It gets into your products, how you deliver your services, and how you interact with your customers or consumers. Because of the wide usage of generative AI, it becomes very important to adopt responsible AI principles across the enterprise."

The risks of failing to implement responsible AI

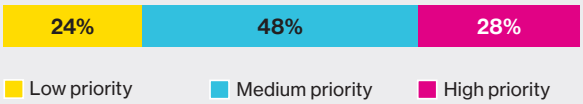
Executives polled lack confidence in their organization's abilities to implement responsible AI principles to address a number of AI-related challenges they face (see Figure 5). Respondents rated their capabilities in addressing the following risks as mostly less than stellar:

Figure 1: Responsible AI is important to business leaders

How would you rate the priority level of implementing responsible AI practices in your organization?

Figure 2: Responsible AI can create a competitive advantage

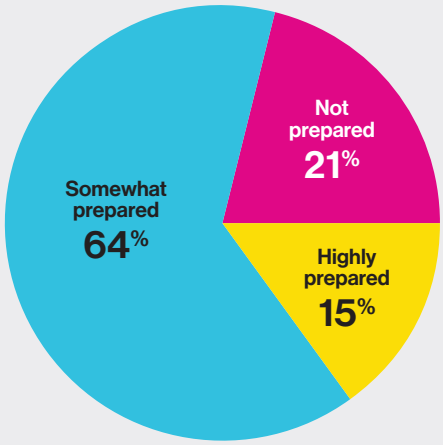
How would you rate the importance of responsible AI to the creation of competitive advantage for your business?



Source: MIT Technology Review Insights poll, 2025

Figure 3: Preparedness for adopting responsible AI

Rate your organization's level of preparedness in adopting effective responsible AI practices.



Source: MIT Technology Review Insights poll, 2025

User adoption and change management. Some AI features can feel intrusive, leading to resistance from users and some people may not feel comfortable using AI products widely in the market and across the organization. "If you lose trust, it will take a long time to rebuild," says Vijay Guntur, chief technology officer and head of ecosystems at HCLTech. Only 23% of respondents felt highly capable of addressing this risk (see Figure 5).

Bias and lack of fairness. This risk could occur if the data used to train AI models is skewed or unrepresentative of certain populations, leading to outputs like biased hiring decisions or AI-generated photos that are unrepresentative of minority populations. Only 26% of respondents felt highly capable of addressing this risk (see Figure 5).

Ability to comply with regulatory frameworks. Regulations at the regional, national, and local levels include the European Union Artificial Intelligence Act, Canada's Artificial Intelligence and Data Act, and California's AI data transparency acts. Non-compliance with regulatory frameworks can result in large fines, legal challenges, and damage to an organization's reputation. Only about a third of respondents (34%) felt highly capable of addressing this risk (see Figure 5).

Operational disruptions. Different parts of the organization adopt "shadow AI" without oversight, leading to inconsistent or irresponsible uses. Flaws in the algorithm or its training data can generate inaccurate or malicious results that undermine confidence in AI and lead to financial and reputational damage. Fifty percent of respondents felt highly capable of addressing this risk (see Figure 5).

Data privacy and security breaches. The news has been filled with examples of major large language models leaking sensitive data, with some companies pulling back from generative AI experiments. Data breaches from security vulnerabilities can potentially lead to theft of sensitive data, financial loss, and damage to an organization's reputation. **In the case of Samsung**, employees inadvertently released proprietary source code while using ChatGPT to help with coding. Despite these fears, a slight majority of respondents (58%) said they feel highly capable of addressing the risk of data privacy and security breaches (see Figure 5).

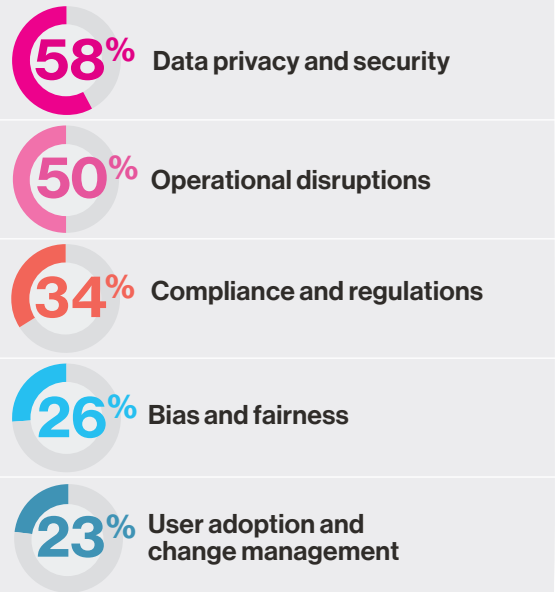
Figure 4: Responsible AI will be a priority across the business into the future

Most respondents say that responsible AI is a high or medium level priority for critical business areas in the next 18 months.

Source: MIT Technology Review Insights poll, 2025

Figure 5: Companies lack confidence in their ability to address AI-related challenges

The percentages of respondents who rate their organization as highly capable of addressing various risks associated with AI.



Source: MIT Technology Review Insights poll, 2025

"Certain industries are going to have to adopt responsible AI much quicker."

Vijay Guntur, Chief Technology Officer and Head of Ecosystems, HCLTech

When companies fail to establish effective responsible AI practices, they ultimately run the risk of creating risks to reputation and stalled implementations. "We're seeing an increased risk of AI- and gen AI-related incidents as AI becomes more prevalent," says Iyer. "It is important for organizations to institutionalize the right AI governance policies and implement the tech guardrails – for example, algorithmic audits, bias detection tools, explainability frameworks (such as SHAP and LIME), and human in the loop systems – to reduce incident risks."

The risks are even greater in highly regulated industries such as financial services and health care. Industries such as these will need to pay much greater attention to responsible AI standards than in the past. "Certain industries are going to have to adopt responsible AI much quicker," says Guntur. "Once you move beyond horizontal adoption across functions within individual enterprises and take an industry view of AI, your tolerance to risk changes."

Best practices for implementing responsible AI

Companies must move beyond experiments and pilot programs to fully integrate responsible AI into their operations. A number of best practices have proven beneficial at leading organizations.

Catalog AI models and data. Many companies have a range of decentralized and uncoordinated AI implementations across the organization. "There's a proliferation of models, some of them are good for certain things, and some of them are better than others," says Iyer, who recommends cataloging all active models and AI systems within the organization as a first step in establishing transparency and consistency. Companies must also accurately catalog and protect their most critical and sensitive data – such as proprietary source code and customer personal data, for example – from unauthorized access or misuse, ensuring that only

The five stages of responsible AI

According to Apoorv Iyer, executive vice president and global head of the generative AI practice at HCLTech, organizations typically follow five stages when integrating responsible AI effectively and sustainably:

STAGE 1

Establish foundational policies.
Every organization must start with a well-defined policy and set of guiding principles for responsible AI. These foundational policies should align with the organization's ethical standards and operational needs.

STAGE 2

Identify key stakeholders.
Successful responsible AI implementation requires involvement across departments, including technology, IT, business, legal, and risk and compliance. These stakeholders should be represented in the responsible AI office and serve as champions of responsible AI within their organizations.

STAGE 3

Develop a responsible AI architecture.
Effective IT architectures encompass foundational principles and tools, and can leverage capabilities from trusted partnerships. This architecture serves as the backbone for implementing and scaling responsible AI across the organization.

STAGE 4

Encourage user adoption and acceptable use.
User training, acceptable use policies, and change management processes are crucial for ensuring adoption of responsible AI at scale. Proper training helps ensure that employees, users, and partners understand and adhere to AI usage guidelines.

STAGE 5

Look to refinement and continuous improvement.
AI systems require ongoing refinement based on user feedback, incidents, and evolving AI advancements. Establishing a feedback loop helps organizations respond to emerging challenges and improve their responsible AI practices.

"If it's not repeatable, users lose trust in AI. Actively monitoring the outputs for potential bias or inaccuracies is also critical to being able to determine how you derived an answer."

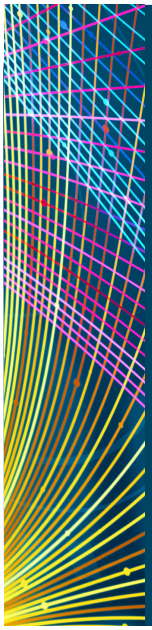
Steven Hall, Chief AI Officer and President of EMEA, ISG

suitable data gets integrated into AI applications. Once created, it is important to do regular algorithmic and data audits on these catalogs.

Create a centralized governance structure. Leading organizations are creating a dedicated AI governance office that oversees all AI models used within the enterprise. The office has responsibility for access control, model approval, and risk assessment, and also helps select and deploy models for specific use cases in compliance with responsible AI principles. In addition, the AI governance office maintains version control for upgrading models, particularly open source models. Robust software development practices (e.g., DevOps and MLOps) can help ensure that models remain up to date and align with responsible AI standards.

Test for explainability, repeatability, and observability. AI models should be able to explain how they derived an answer. But even more important, they should be able to provide consistent results based on the same inputs and allow visibility into how the model behaves and makes decisions. If the answer a model gives changes when you provide the same prompt, it's not repeatable. Repeatable results enable organizations to explain AI decision-making processes clearly, which is particularly important for regulatory compliance. "If it's not repeatable, users lose trust in AI," says Hall. "Actively monitoring the outputs for potential bias or inaccuracies is also critical to being able to determine how you derived an answer."

Conduct thorough risk assessments. AI systems carry different levels of risk. Responsible AI implementation involves assessing models based on the potential impact on individual users and society in general, especially in high-stakes applications like finance or health care. "Categorizing uses by levels of risk and then properly validating and testing for risks is the way to think about compliance for responsible AI," says Guntur. **Under the EU AI Act**, unacceptable risks include systems that are harmful, manipulative, or exploitative. High-risk applications are categorized as those that have the



The future of agentic AI

Agentic AI can perform tasks autonomously with little or no human involvement. These systems are capable of making decisions independently and learning from experience.

With AI agents increasingly in charge, companies need predefined frameworks and approaches for agents to be able to responsibly handle high-stakes decisions and avoid the greatest risks. "It is early days, but companies can mitigate the risk through a human-in-the-loop approach," says Vijay Guntur, chief technology officer and head of ecosystems at HCLTech.

Companies are adopting agentic AI first in areas with lower risk profiles where it can work alongside humans, such as in IT operations. Beyond IT, agentic AI is likely to be used to streamline operations in customer service, marketing, and other functions, working semi-independently while still under human oversight. For high-stakes areas like software testing or code generation, fully autonomous agentic AI will take time to mature and adopt.

potential to put people's health, safety, or fundamental rights on the line, such as critical infrastructure or essential public or private services, for example. Limited risk systems include chatbots or those used to generate or manipulate images.

Test for security and ethical compliance. Many organizations conduct regular audits that evaluate AI models on specific metrics to ensure that AI models meet fairness, security, and other responsible principles consistently. AI security goes beyond traditional cybersecurity, emphasizing authentication, identity management, and access controls to prevent internal and external threats. "With AI security, the threat can also be internal, so that's something executives have to think about," says Guntur. Regular, holistic code testing throughout the AI lifecycle, often within the firewall, can ensure that both AI-generated and human-coded software gets rigorously vetted before deployment. Structuring information as data products also ensures secure, controlled access to specific data sets for different AI applications without exposing the full data set.

Hardwire compliance with regulatory standards. With AI regulations evolving rapidly, companies must ensure that AI systems and processes align with a wide range of regional, national, local, and industry-specific standards. "Every regulation has a set of standards about how the data is supposed to be treated," says Hall. Responsible AI requires companies to perform regular audits to harmonize and confirm regulatory compliance, using off-the-shelf and customized tools that assess AI models for potential regulatory violations.

Empower employees with training at scale. Training employees about AI principles and responsible practices is essential to successful adoption. AI champions within the organization can also facilitate responsible AI practices by fostering a culture of accountability and awareness. Creating a channel for employee feedback and ideas about AI tools and practices can also measurably improve responsible AI implementation and adoption.

Responsible AI starts at the top

Ultimately, leaders need to articulate clear expectations and provide ethical guidelines for AI projects to ensure responsible use across functions, departments, and business units. Implementing responsible AI well pays dividends. It gives leaders the confidence that they are striking the right balance in safeguarding the organization's most valuable asset: trust.

The goal is to balance productivity goals with regulatory adherence. "Leaders want to gain a competitive advantage with AI, whether that's productivity improvement or the discovery of new business processes and services," says Hall. "But at the same time, they want to use AI to make sure that they're managing things in regulated or unregulated environments. So there's a balance."

"Implementing responsible AI in the generative age" is an executive briefing paper by MIT Technology Review Insights. We would like to thank all participants as well as the sponsor, HCLTech. MIT Technology Review Insights has collected and reported on all findings contained in this paper independently, regardless of participation or sponsorship. Laurel Ruma was the editor of this report and Nicola Crepaldi was the publisher.

About MIT Technology Review Insights

MIT Technology Review Insights is the custom publishing division of MIT Technology Review, the world's longest-running technology magazine, backed by the world's foremost technology institution – producing live events and research on the leading technology and business challenges of the day. Insights conducts qualitative and quantitative research and analysis in the US and abroad and publishes a wide variety of content, including articles, reports, infographics, videos, and podcasts. And through its growing MIT Technology Review Global Insights Panel, Insights has unparalleled access to senior-level executives, innovators, and entrepreneurs worldwide for surveys and in-depth interviews.

From the sponsor

HCLTech's full-stack AI/GenAI capabilities enable it to assist clients in unlocking the full potential of AI through four key offerings: AI Force, a platform that accelerates software and IT operations lifecycle; AI Foundry, providing consultancy and managed SI services for data, AI, and infrastructure; AI & Cloud Native Labs, dedicated spaces for catalyzing AI/GenAI and Cloud Native ideas; and AI Engineering, which includes semiconductor and hardware engineering services. The company's approach is grounded in the "art of possible," empowering clients to explore new possibilities and deploy practical solutions that deliver business impact and ROI.

HCLTech

Illustrations
Cover art and spot illustrations created with Adobe Stock.

MIT Technology Review Insights

www.technologyreview.com

insights@technologyreview.com