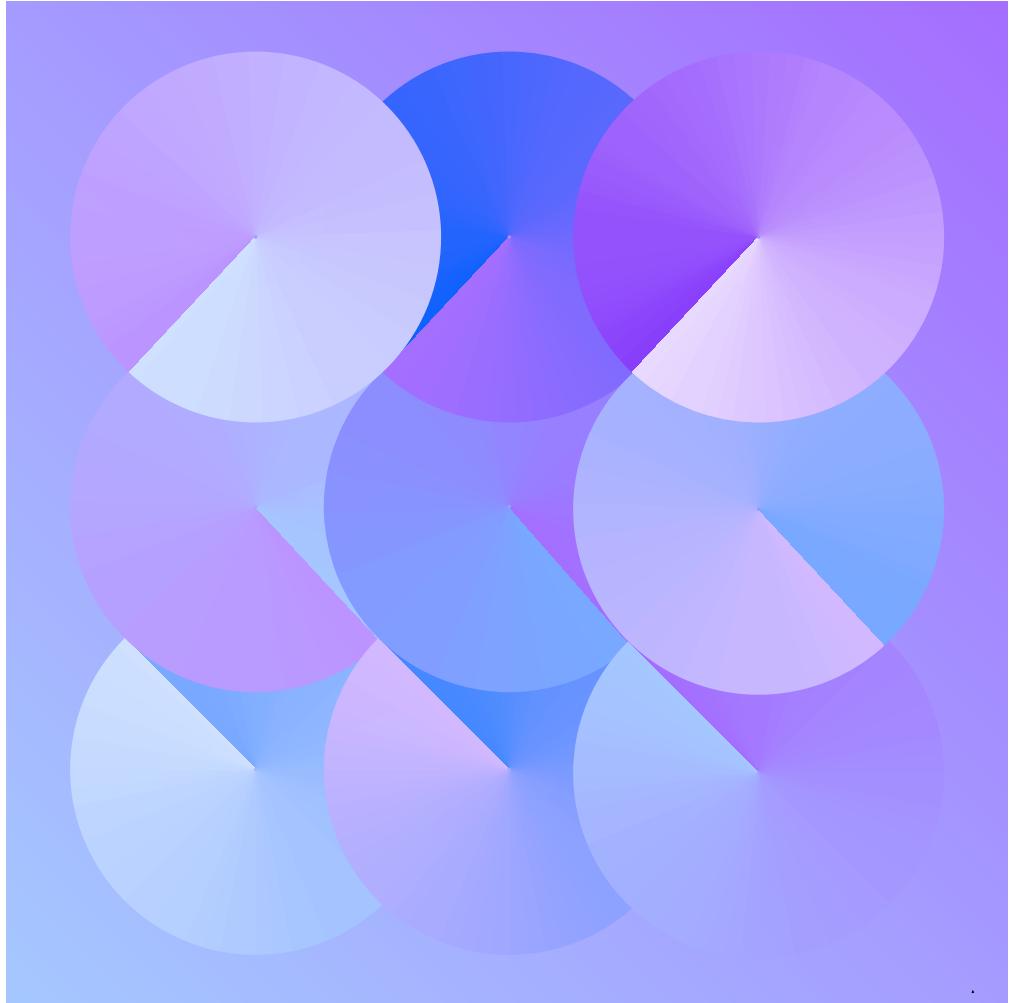


The enterprise guide to AI governance

Three trust factors that can't be ignored

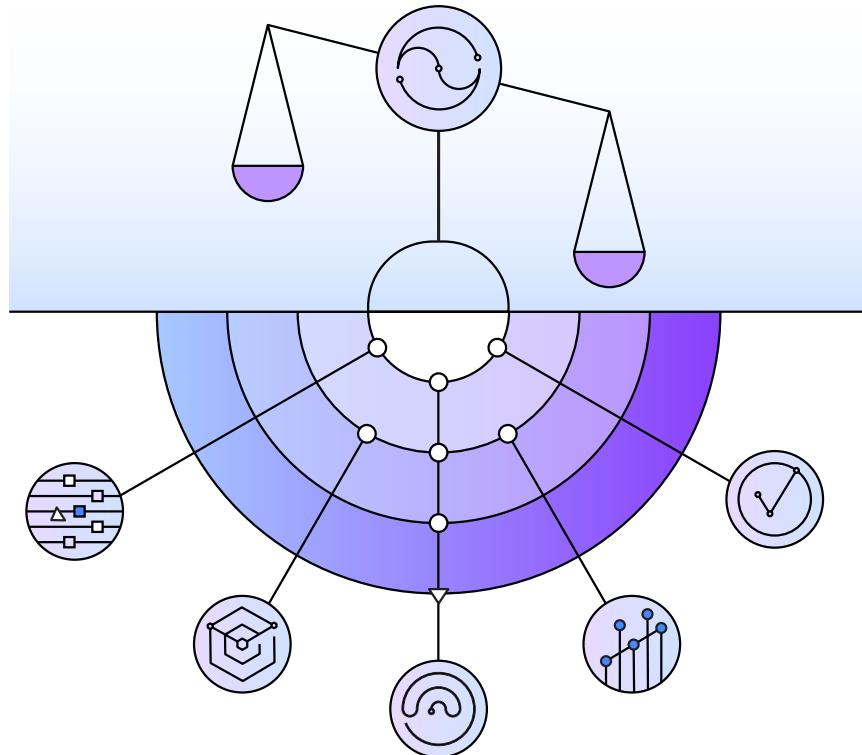


Foreword

Putting generative AI governance in context

This Research Brief is part of an ongoing series of reports published by the IBM Institute for Business Value (IBM IBV) about generative AI and the opportunities and challenges it presents to organizations worldwide.

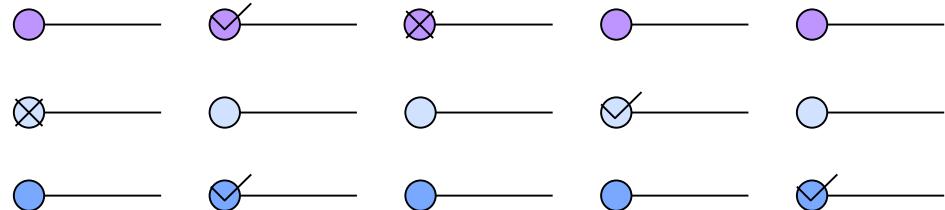
As business leaders adopt generative AI to boost competitiveness and increase productivity, leaders need information on the ever-shifting landscape. Other reports include: *5 trends for 2024: Deep tech requires deep trust*, *The CEO's guide to generative AI: Risk management*, and *The CEO's guide to generative AI: Responsible AI & Ethics*.



Introduction

Why AI governance matters more than ever

In less than two years, generative AI, the latest evolution of artificial intelligence, has moved from novelty to business necessity. With gen AI strategies shifting rapidly from exploring to focusing to expanding, 77% of business leaders in a recent IBM IBV study say they are convinced that gen AI is not only market ready, but that quick adoption is necessary to maintain competitiveness.¹



Gen AI is cutting coding time from days to minutes, boosting content creation, personalizing customer and employee interactions, automating cybersecurity operations, and optimizing processes. But at the same time, AI-related risks are also on the rise: compliance and regulation, data bias and reliability, and a loss of trust when users don't understand AI model operation and governance.

Governance refers to the principles, policies, and responsible development practices that align AI tools and systems with ethical and human values. It establishes the frameworks, rules, and standards that direct AI research, development, and application according to the principles that organizations deem worthy. Governance mitigates the potential risks associated with AI—bias, discrimination, and harm to individuals—through sound AI policy, data governance, and well-trained and maintained datasets.²

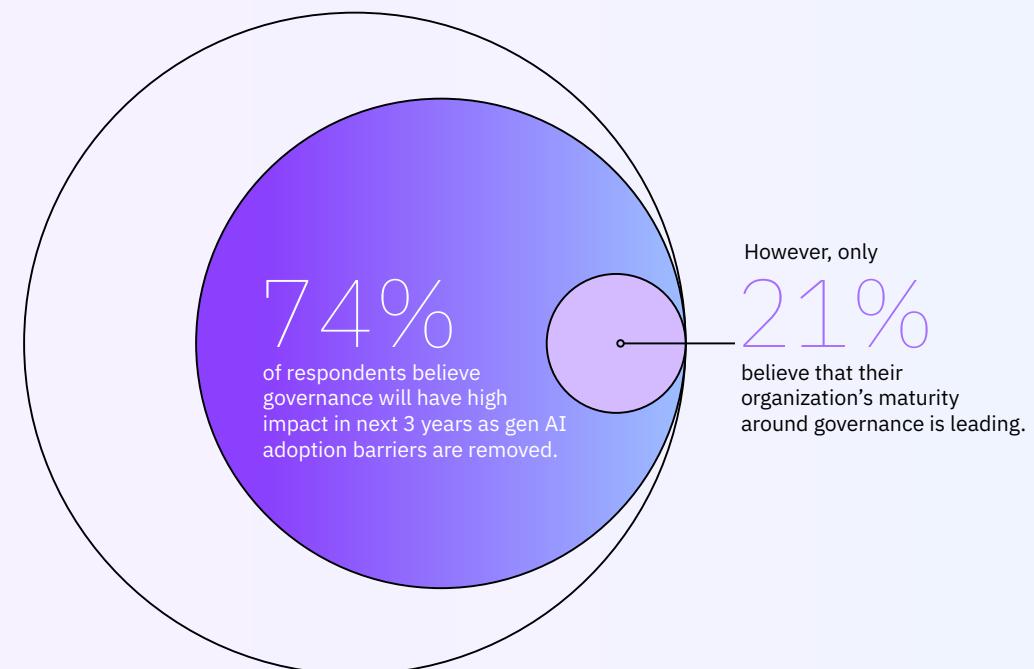
Effective governance is key to building a foundation of trust. Monitoring how AI models are trained and managed helps organizations not only build better models but reassures employees, customers, partners, and other stakeholders that the information and services they use are reliable.

Governance can also be a catalyst for growth—for instance, by facilitating more meaningful connections with customers. It's part of a management mindset that goes beyond risk management and compliance to unlock opportunity. What's troubling is that only 21% of executives in our research say their organization's maturity around governance is systemic or innovative (see Figure 1).

From training and tuning to inference and outputs, risks can crop up at every phase of AI development. In fact, MIT researchers recently compiled a list of over 750 AI risks to help identify gaps and uncertainties in how organizations perceive the AI risk landscape.³

Figure 1

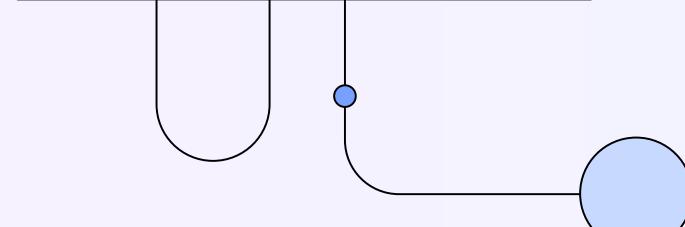
Governance can also be a catalyst for growth—for instance, by facilitating more meaningful connections with customers.



Over 65% of data leaders at a recent Gartner conference highlighted data governance as their top focus in 2024.⁴ Executives across the C-suite admit that their organizations need to do better: 60% of CEOs say they're looking into mandating additional AI policies to mitigate risk. While 63% of CROs and CFOs say they are focused on regulatory and compliance risks, only 29% say these risks have been sufficiently addressed⁵ (see Figure 2). Roughly 27% of public companies cited AI regulation as a risk in recent filings with the Security and Exchange Commission.⁶

So how do organizations move gen AI initiatives forward as fast as possible to capture business value—while also constructing governance guardrails to keep gen AI on track? In this report, we will explore this central leadership challenge, and provide specific recommendations for action, through the lens of three major trust factors.

Figure 2
Why AI governance matters more than ever.



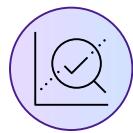
Trustworthy AI depends on effective governance. Incorporating these three trust factors is the starting point for building governance frameworks.



Trust factor 1

Accountability: Who is in charge of AI governance?

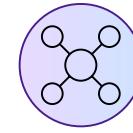
Effective AI governance must be a funded mandate from senior leadership. Organization-wide adoption requires flexible governance frameworks to mitigate risks and achieve business goals.



Trust factor 2

Transparency: How do you assess sources of data and what is shared about them?

Diverse and multidisciplinary teams should be deployed to assess data used to build models, matching the broad range of needs and expectations of AI users. This will answer questions about how models are audited and how they perform compared to humans.



Trust factor 3

Explainability: How do you explain the output of AI systems and models?

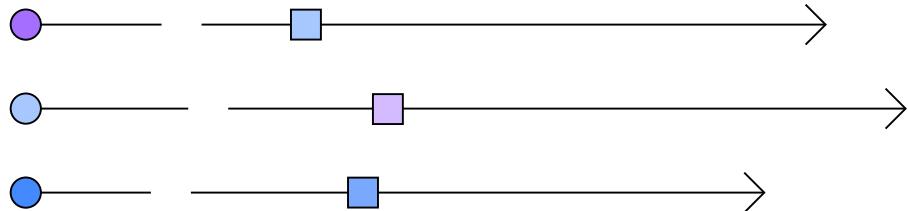
Deep collaboration between people and AI systems is built on transparency and explainability, adding the human touch to AI-informed decision-making. Sharing the provenance of models is fundamental to trust.

Clear AI governance practices and policies are at the core of addressing these trust factors. Without governance, the adoption of trustworthy and ethical AI systems can be inhibited. At the same time, gen AI itself can help improve governance—all across the enterprise.

Three key AI governance-related terms

Transparency, explainability, and provenance

Effective governance—delivered through corporate instructions, staff, processes, and systems—helps assure that AI systems operate as an organization intends, while meeting stakeholder expectations and regulatory requirements. To enable AI users to direct, evaluate, monitor, and take corrective action at all stages of the AI lifecycle, governance relies on transparency, explainability, and provenance.



Transparency is the ability to perceive how an AI system is designed and developed, typically supported by the sharing of appropriate details about the AI system.⁷ To build a trustworthy AI model, algorithms cannot be perceived as black boxes. AI developers, users, and stakeholders must understand the inner workings of AI to trust its results.

Explainability, in the context of AI, is a set of practices, tools and design principles that makes AI decisions more comprehensible to humans.⁸ The more explainable an AI system is, the greater its ability to provide insights that

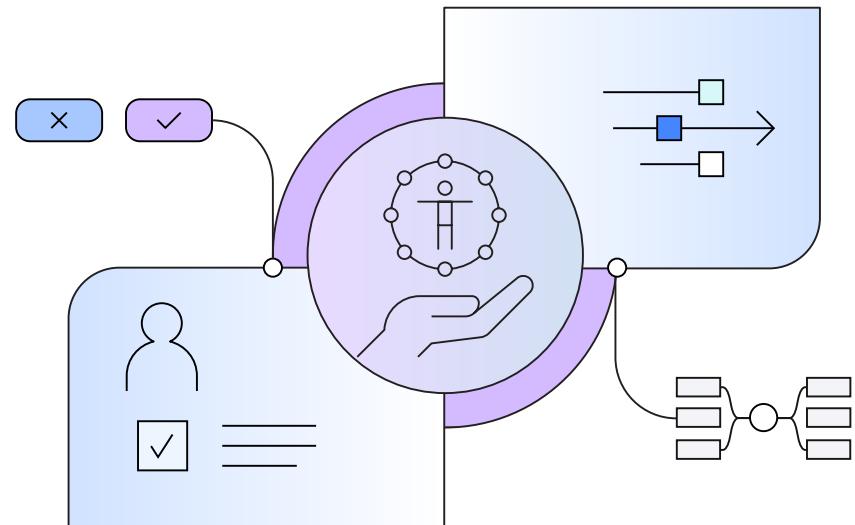
people can use and trust. Within its own governance framework, each organization adopts an explainability approach to meet its objectives.

Provenance refers to the ability to explain and verify the origins of the data that trains AI models throughout their lifecycles.⁹ It is vital for ensuring authentic data inputs and for enhancing trust in AI-generated insights and decisions. By recording metadata from the data's source, provenance provides historical context and supports data validation and auditing, leading to more accurate and trustworthy AI outputs.

Trust factor 1 | Accountability

Who is in charge of AI governance?

The first step is clear accountability. In our research, 60% of C-suite executives say they have placed clearly defined gen AI champions throughout their organization. And almost as many—59%—say they have a direct report responsible for organization-wide AI integration. What's more, 80% of C-suite executives say they have a separate risk function dedicated to using AI or gen AI. They want to be sure that in developing and deploying AI, they are mitigating the risks of unintended harm and unwanted biases.



Trust factor 1 | Accountability

Another IBM IBV survey of C-suite leaders, creative executives, creative managers, and designers revealed that 47% of respondents have established a generative AI ethics council to create and manage ethics policies and mitigate generative AI risks.¹⁰ The goal of these councils is to address the risk of “lawful but awful” AI.¹¹ The establishment of enterprise-wide governance frameworks helps streamline the process of detecting and managing technology ethics concerns in AI projects.¹²

Supporting governance requires a commitment of resources. Spending on AI ethics has steadily increased from 2.9% of AI spending in 2022 to 4.6% in 2024. This share is expected to increase to 5.4% in 2025.

Our research indicates that more technologically mature organizations tend to prioritize AI governance. For instance, 68% of CEOs in an IBM IBV survey say governance for gen AI must be integrated upfront in the design phase, rather than

retrofitted after deployment.¹³ Yet less sophisticated players and newer entrants to AI struggle with the complex choices that governance can raise. The solution is often flexible AI governance frameworks, which can help adapt to changing markets, mitigate risks, and encourage greater adoption to realize potential.¹⁴

Spending on AI ethics has steadily increased from

2.9% in 2022

to

4.6% in 2024

This share is expected to increase to

5.4% in 2025

Build robust AI governance frameworks under an executive mandate.

- 1. Empower a senior-level executive to lead AI and data governance initiatives.**

Send the message that governance is a senior management priority. Championing AI and data governance from an enterprise's highest levels minimizes the risk of failure due to fragmented ownership and fuzzy accountability.
- 2. Prioritize and build on responsible AI development and deployment.**

Give leaders who will be accountable for AI governance the authority to do the work and provide their teams with the necessary resources to support this mandate. Teams can also build on and evolve governance frameworks already in place.
- 3. Ensure that senior leadership aligns principles with practices.**

Align values related to the development and procurement of AI. Organizations achieve the outcomes they measure, and aligning principles with practices supports measurement of progress towards responsible AI adoption.
- 4. Develop a cultural foundation for governance structures.**

Without a strong cultural foundation, AI governance structures cannot gain traction. Healthy cultures have success measurements, incentives, messaging and communications, diversity and inclusiveness, psychological safety, proactive employee training, and a holistic approach to AI literacy.
- 5. Foster collaboration with stakeholders and ecosystem partners.**

Include stakeholders across the entire organization—all working toward the same goals. Collaboration with governments, trade and industry associations, and other groups helps establish AI governance guidelines, best practices, and regulations for responsible AI use. Ensure third-party software vendors and partners with embedded AI are subject to audits and other governance processes.

The Data & Trust Alliance

Enhancing AI business value and trust with data provenance standards¹⁵

The Data & Trust Alliance (D&TA) was established in 2020 by CEOs from leading companies, based on a shared conviction that the future of business will be powered by the responsible use of data and AI. The 27 members of the Alliance—including Deloitte, GM, IBM, Johnson & Johnson, Mastercard, Meta, Nike and UPS—represent 18 industries, employ over four million people and earn \$2 trillion in annual revenue.

“The value of AI depends on the quality of data. To realize and trust that value, we need to understand where our data comes from and if it can be used, legally. That’s why the members of the Data & Trust Alliance created a new business practice through cross-industry data provenance standards.”

Saira Jesani
Executive Director, Data & Trust Alliance

Case study | Data provenance standards



The Alliance creates tools and practices to enhance trust in data, models, and the processes around them. In 2023, the Alliance developed the first set of cross-industry data provenance standards, including 22 metadata fields that provide essential information about the origin of data and associated rights.

These standards were created with two objectives in mind: business value and implementation feasibility. By adopting D&TA data provenance standards, businesses can better understand datasets before purchase or use—and have a basis to decline data or request changes from third parties. To encourage adoption, only the

most essential metadata—required to understand more about a dataset's origin, its method of creation and whether it can be legally used—were selected.

In early 2024, IBM tested the D&TA data provenance standards as part of a clearance process for datasets used to train foundational models. IBM's data governance program already included a data clearance process that applied relevant controls, documented lineage, and defined guidelines for use and re-use. The challenge was a need to respond to an increasing volume of data clearance requests. The organization tested the standards to optimize the process for greater efficiency and accuracy.

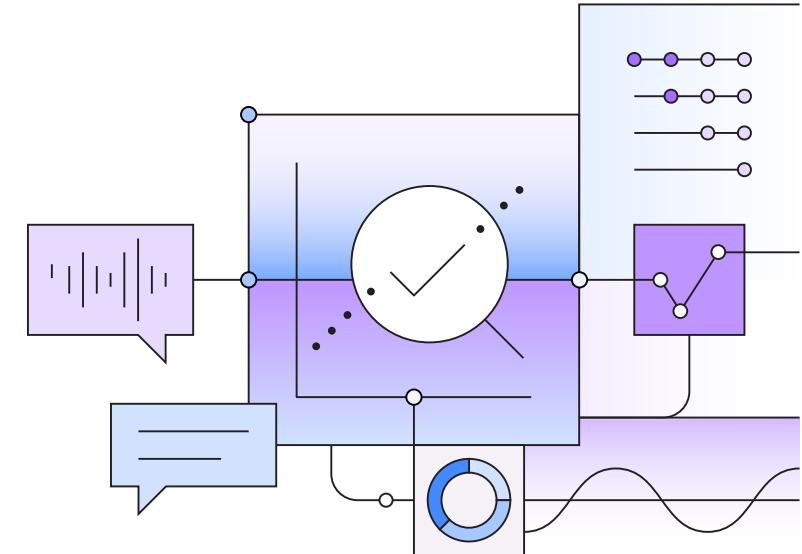
IBM saw increases in both efficiency—time for clearance—and overall data quality, with a 58% reduction in data clearance processing time for third-party data and a 62% reduction in data clearance processing time for IBM owned or generated data. The D&TA standards were a meaningful contributing factor to these improvements. IBM is now adopting the D&TA data provenance standards into its business data standards, where appropriate, to further optimize enterprise data governance.

Trust factor 2 | Transparency

How do you assess AI data sources and what is shared about them?

90% of the data available in the world was generated in the last two years—just as gen AI went from curiosity to ubiquity. Approximately 400 million terabytes of data are created every day, with 150 zettabytes estimated to be generated in 2024.¹⁶ But managing such huge amounts of data presents huge challenges. Almost half of surveyed CEOs say they are concerned about accuracy and bias—an issue that could create as many problems as generative AI promises to solve.

Before people can use AI, trust must be earned, and the most effective way to earn user trust is through transparency. With respect to personal data, transparency is a key privacy principle. It requires organizations to be open and forthcoming about their data processing practices. This enables people to determine how they want their data used and shared.



For transparency to be effective, organizations must provide explainability—the ability of an AI system to provide insights that people can use to understand the causes of the system's predictions. Clear explanations must be provided about accountability, data, models, algorithms, performance, audits, and related factors (see Figure 3). Otherwise, organizations take on a tremendous exposure to risk.

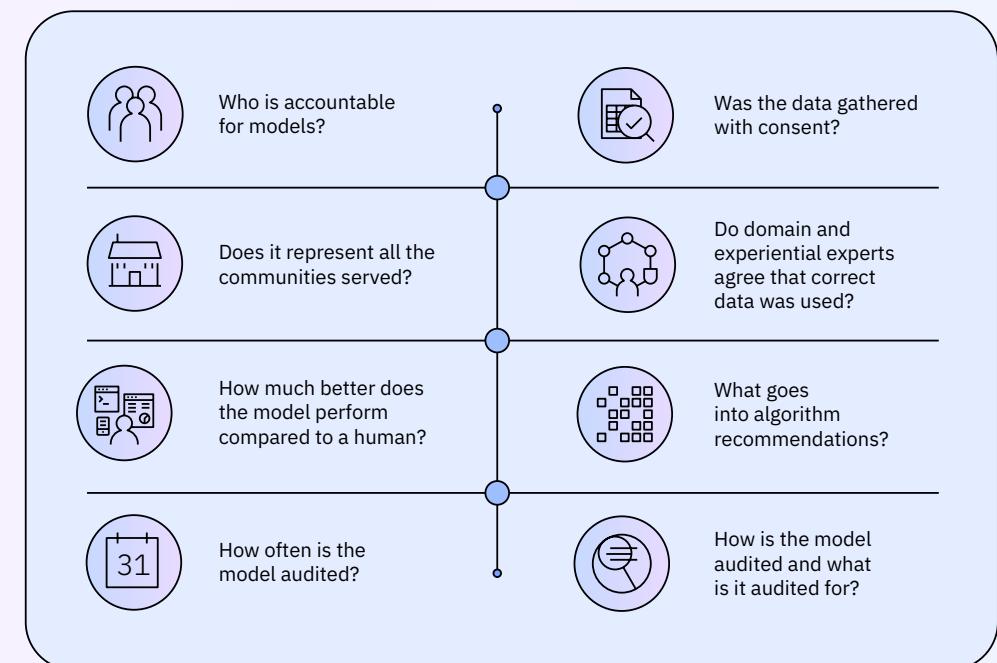
Trust factor 2 | Transparency

Transparency eliminates black box opacity and supports accurate and fair decision-making. As artifacts of the human experience, virtually all data is biased. AI mirrors our biases. The question is: which biases do not reflect our values? If bias aligns with an organization's values, there must be transparency about why that dataset and approach were chosen over others. If they don't align, a different approach is needed.

To help ensure assumptions are not overlooked, AI governance should include diverse, multidisciplinary teams to both build and govern these models. Outside experts in psychology, anthropology, law, philosophy, linguistics, and other disciplines can also help ensure that AI is used to augment human intelligence in ways that align with human values. Governance teams also need a psychological safety net when having challenging conversations about potential disparate impacts of an AI model.

Figure 3

Here are important questions for which organizations must provide clear explanations.



Assemble a Dream Team to build effective AI governance.

1. Establish a multidisciplinary AI governance team.

To avoid blind spots, include a broad range of expertise from technical, ethical, and social domains. This diversity can help identify gaps faster, help leverage existing governance mechanisms and enable an organization to proactively head off unintended impacts.

2. Train everyone in transparency.

Give employees at all levels opportunities to receive the training they need to build or procure AI models responsibly in their own domain, as well as awareness of when to seek help when working outside their domain, such as audits. Engage the workforce by creating a culture that celebrates openness and inclusion.

3. Ask questions and think beyond regulatory compliance.

Best practices for governance go deeper than rules and regulations and can open doors to innovation. Compliance starts with policies, procedures, and industry standards for AI. Building a framework for compliance enables efficient incorporation of new rules and regulations.

4. Embrace ideas from outside the organization.

Learn, follow, and look for opportunities to participate in the development of intergovernmental and international standards. AI principles promulgated by the OECD provide a useful starting-off point.¹⁷ Additional national and global standards are expected as gen AI adoption grows.

Australia Post

Delivering a more
efficient future¹⁸

With annual revenues over \$5.8 billion, Australia Post provides postal services from 4,310 locations. As a government-owned corporation, the Post must use transparency and rules to maintain customer trust. Much of the data handled is personal and sensitive.

Generative AI is a key part of the Post's mission to boost customer service and efficiency. After testing and reviewing thousands of customer calls and employee keystrokes, generative AI is now routing customer queries and answering business-as-usual questions. The Post is working toward a goal where gen AI could handle between 40% and 60% of calls, delivering a better customer experience while significantly reducing costs.

But mindful that while the public has concerns about AI, the Post is embracing transparency, because gen AI is well underway. The Post has already conducted a review of all its data and is now creating strict procedures around data governance. It is committed to not just aligning with regulatory frameworks but hardening its privacy and security protocols.

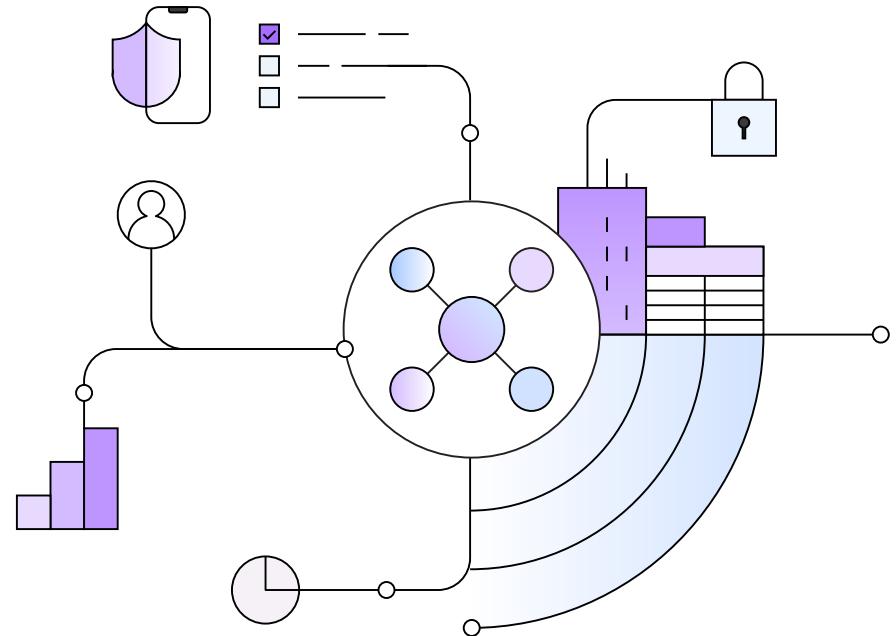
Australia Post is working toward a goal where gen AI could handle between 40% and 60% of calls to improve customer experience and reduce costs.

How do you explain the output of AI systems and models?

As more organizations adopt AI, AI acceptance is at a crossroads.

While 35% of respondents to the 2024 Edelman Trust Barometer survey say they accept this innovation, almost as many—30%—reject it.¹⁹

Demonstrating the trustworthiness of AI will be key to optimizing AI's impact. Trustworthy AI can also contribute to new ideas that separate innovators from those doing the bare minimum.



Trust factor 3 | Explainability

A key element of trustworthy AI is provenance—the ability to explain and verify the origins and history of data throughout its lifecycle. When training AI models, provenance is essential for ensuring that the data is authentic and trustworthy. Authentic data inputs into AI models enhances the trustworthiness of AI-generated insights and decisions.

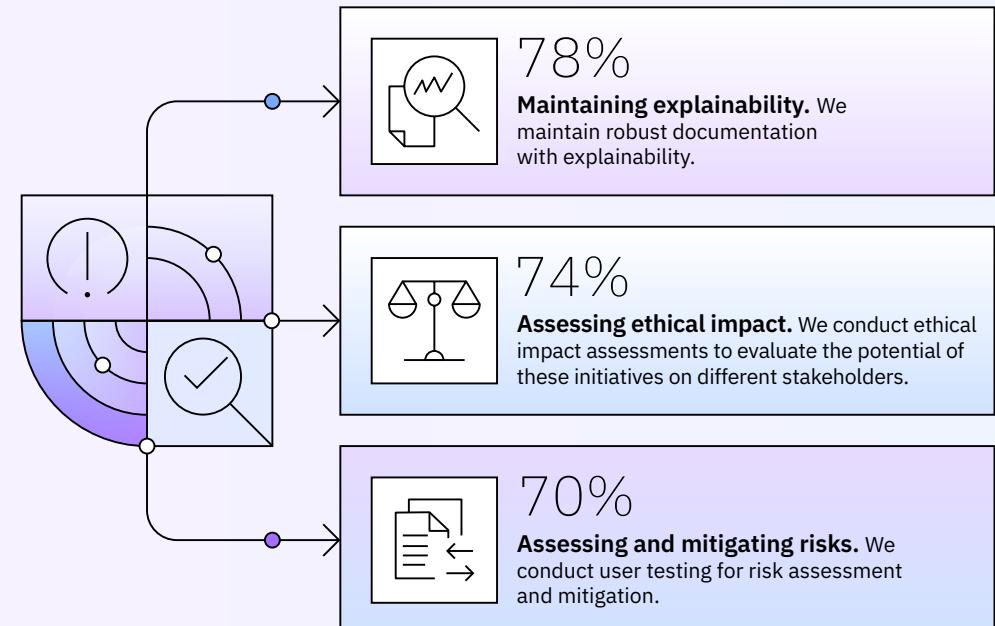
For people to trust what goes into and what comes out of AI models, explainability—the ability to understand and trust AI outputs—is informed by provenance. Explainability is

not just limited to explaining how a gen AI model renders outputs. In higher risk use cases, it is appropriate to have every output provide an explanation of its data lineage informed by provenance, along with evidence.

Most executives in our research say they recognize the importance of explainability; 78% maintain robust documentation; 74% conduct ethical impact assessments; and 70% conduct user testing for risk assessment and mitigation (see Figure 4).

Figure 4

Most executives in our research say they recognize the importance of explainability.



Keep humans in the loop.

1. Design AI systems that facilitate human-AI collaboration and oversight.

Create and scale repeatable patterns to ensure AI systems and their transparent metadata are accessible to the people using them, no matter what their level of technical understanding.

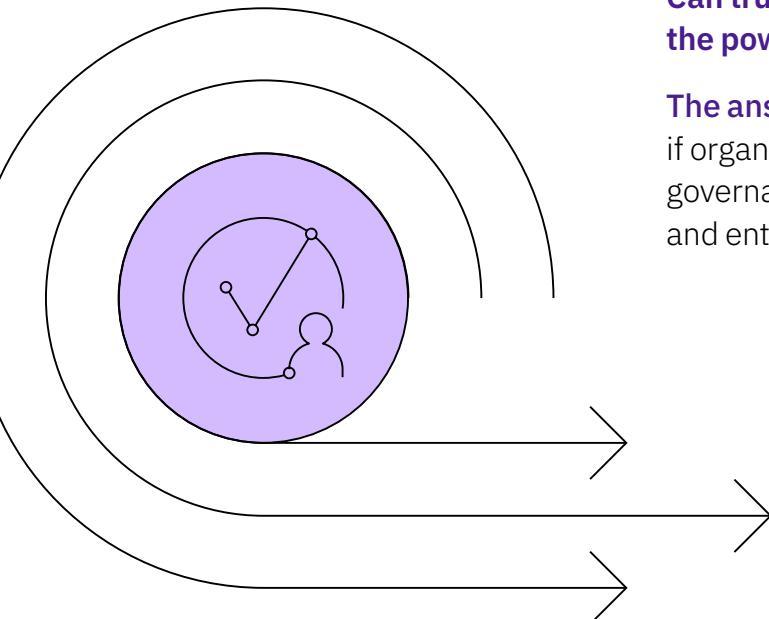
2. Prioritize AI output that is explainable and auditable.

Invest in applied training that improves AI literacy and provides clear guidance on designing and developing human-centric systems.

3. Incentivize employees to speak up and speak out if AI output is confusing.

Make sure those who build, design, and procure AI adopt a human-centric instead of a data-centric approach and consider how outputs can be evaluated after the fact. Provide appropriate communication so employees feel empowered and competent to ask about potential disparate impacts related to the AI models they work with.

Governance and building trust



Can generative AI be trusted? Can trust guardrails balance the power of gen AI?

The answer is yes—but only if organizations approach AI governance with commitment and enthusiasm.

Understanding the data used to train, tune, and make inferences from AI models is essential. What a company does with AI is defined, in large part, by how it selects, governs, analyzes, and applies data across the enterprise. Communicating that process transparently is how trust is built and maintained over time.

Governance needs to be embedded at every phase of the generative AI lifecycle—not in functional silos but across the enterprise. It must be championed by top leadership that provides strategic guidance, recognition, and feedback. According to the 2024 Edelman Trust Barometer,

79% of global respondents say it is important for their CEOs to speak out about the ethical use of technology.²⁰

Ultimately, AI governance is about much more than rules, restrictions, regulations, and requirements. It's about a shared understanding of practices for effective collaboration that can reduce uncertainty and increase predictability—practices which may actually accelerate development. When sponsored and promoted at the leadership level, AI governance will no longer be seen as just another IT issue, but as a core strategy for value creation, growth, innovation, and developing the potential of human-AI collaboration.

Authors

Phaedra Boinodiris

Global Leader for Trustworthy AI

IBM Consulting

pboinodi@us.ibm.com

<https://www.linkedin.com/in/phaedra/>

Brian Goehring

Associate Partner and AI Research Lead

Institute for Business Value

IBM Consulting

goehring@us.ibm.com

<https://www.linkedin.com/in/brian-c-goehring-9b5a453/>

Milena Pribic

Design Principal, Ethical AI Practices

IBM Software

mpribic@us.ibm.com

<https://www.linkedin.com/in/milenapribic/>

Catherine Quinlan

Vice President, AI Ethics

IBM Chief Privacy Office

cquinlan@us.ibm.com

<https://www.linkedin.com/in/catherinemquinlan/>

Contributors

Sara Aboulhosn, Lee Cox, Rachna Handa, Christina Montgomery, Shyam Nagarajan, Dasha Simons, Michael Tucker, and Kush Varshney.

The right partner for a changing world

At IBM, we collaborate with our clients, bringing together business insight, advanced research, and technology to give them a distinct advantage in today's rapidly changing environment.

IBM Institute for Business Value

For two decades, the IBM Institute for Business Value has served as the thought leadership think tank for IBM. What inspires us is producing research-backed, technology-informed strategic insights that help leaders make smarter business decisions. From our unique position at the intersection of business, technology, and society, we survey, interview, and engage with thousands of executives, consumers, and experts each year, synthesizing their perspectives into credible, inspiring, and actionable insights. To stay connected and informed, sign up to receive IBV's email newsletter at ibm.com/ibv. You can also find us on LinkedIn at <https://ibm.co/ibv-linkedin>.

Related reports

The CEO's Guide to Generative AI: Manage unpredictable risks

<https://www.ibm.com/thought-leadership/institute-business-value/en-us/report/ceo-generative-ai/ceo-ai-risk-management>

The CEO's Guide to Generative AI: Responsible AI & ethics

<https://www.ibm.com/thought-leadership/institute-business-value/en-us/report/ceo-generative-ai/responsible-ai-ethics>

The ingenuity of generative AI

<https://www.ibm.com/thought-leadership/institute-business-value/en-us/report/scale-generative-ai>

Notes and sources

1. Goehring, Brian, Manish Goyal, Ritika Gunnar, Anthony Marshall, and Aya Soffer. *The ingenuity of generative AI*. IBM Institute of Business Value. June 2024. <https://www.ibm.com/thought-leadership/institute-business-value/en-us/report/scale-generative-ai>
2. Mucci, Tim and Stryker, Cole. "What is AI governance?" IBM Blog. November 28, 2023. <https://www.ibm.com/topics/ai-governance>
3. Constantino, Tor. "AI's Risky Business, MIT Researchers Catalogue Over 750 AI Risks." *Forbes*. September 11, 2024. <https://www.forbes.com/sites/torconstantino/2024/09/11/ais-risky-business-mit-researchers-catalogue-over-750-ai-risks/>
4. "Data Governance is a Top Priority for 65% of Data Leaders-Insights From 600+ Data Leaders For 2024." *Humans of data*. March 28, 2024. <https://humansofdata.atlan.com/2024/03/future-of-data-analytics-2024/>
5. *The CEO's guide to generative AI: Risk management*. IBM Institute for Business Value. August 2024. <https://www.ibm.com/thought-leadership/institute-business-value/en-us/report/ceo-generative-ai/ceo-ai-risk-management>
6. Lin, Belle. "AI Regulation Is Coming. Fortune 500 Companies Are Bracing for Impact." *The Wall Street Journal*. August 27, 2024. <https://www.wsj.com/articles/ai-regulation-is-coming-fortune-500-companies-are-bracing-for-impact-94bba201>
7. IBM Design for AI guidelines and definitions.
8. What is explainable AI? IBM. <https://www.ibm.com/topics/explainable-ai>
9. What is data provenance? IBM. <https://www.ibm.com/think/topics/data-provenance>
10. *Disruption by design: Evolving experiences in the age of generative AI*. IBM Institute for Business Value. June 2024. <https://www.ibm.com/thought-leadership/institute-business-value/en-us/report/generative-ai-experience-design>
11. Foody, Kathleen. "Explainer: Questioning blurs meaning of 'lawful but awful'". AP. April 7, 2021. <https://apnews.com/article/death-of-george-floyd-george-floyd-cba9d3991675231122e2b68fd5b4b00>
12. Montgomery, Christina and Francesca Rossi. "A look into IBM's AI ethics governance framework." IBM Blog. December 4, 2023. <https://www.ibm.com/blog/a-look-into-ibms-ai-ethics-governance-framework/>
13. *6 hard truths CEOs must face*. IBM Institute for Business Value. May 2024. <https://www.ibm.com/thought-leadership/institute-business-value/en-us/c-suite-study/ceo>
14. A Flexible Maturity Model for AI Governance Based on the NIST Risk Management Framework. *IEEE USA*. July 2024. <https://ieeusa.org/product/a-flexible-maturity-model-for-ai-governance/>
15. The Data & Trust Alliance. Accessed September 28, 2024. <https://dataandtrustalliance.org/about>
16. Duarte, Fabio. "Amount of data created daily (2024)." June 13, 2024. *Exploding Topics*. <https://explodingtopics.com/blog/data-generated-per-day>
17. *OECD AI Principles overview*. OECD.AI Policy Observatory. May 2019. <https://oecd.ai/en/ai-principles>
18. Internal IBM case study.
19. Edelman, Margot. "Why the human touch is needed to harness AI tools for communications." World Economic Forum. June 18, 2024. <https://www.weforum.org/agenda/2024/06/human-touch-harness-ai-tools-communications/>
20. Ibid.



© Copyright IBM Corporation 2024

IBM Corporation
New Orchard Road
Armonk, NY 10504

Produced in the United States of America | October 2024

IBM, the IBM logo, ibm.com and Watson are trademarks of International Business Machines Corp., registered in many jurisdictions worldwide. Other product and service names might be trademarks of IBM or other companies. A current list of IBM trademarks is available on the web at "Copyright and trademark information" at: ibm.com/legal/copytrade.shtml.

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.

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.

This report is intended for general guidance only. It is not intended to be a substitute for detailed research or the exercise of professional judgment. IBM shall not be responsible for any loss whatsoever sustained by any organization or person who relies on this publication.

The data used in this report may be derived from third-party sources and IBM does not independently verify, validate or audit such data. The results from the use of such data are provided on an "as is" basis and IBM makes no representations or warranties, express or implied.

This document is printed on chlorine-free 100% post-consumer paper meeting Forest Stewardship Council (FSC) responsible forestry certification. The energy used to manufacture this paper was generated through renewable green energy. Please recycle.





ibm.co/consulting-ai