

2026 Responsible AI Transparency Report

Executive Summary

[Read the full report](#)



AI can benefit communities worldwide—if we earn and sustain trust that it’s developed and deployed responsibly. Responsible AI is not a one-time exercise but a continuous discipline. Our AI governance program is how we translate this discipline into practice. Through a combination of people, principles, policies, and tools, we empower teams across Microsoft—and the customers who build on our platforms—to develop and deploy AI responsibly.

Three key themes emerge from this year’s Responsible AI Transparency Report:

We are designing and implementing more adaptive governance as AI rapidly advances. As AI capabilities advance, adoption grows, and societal expectations evolve, governance cannot remain static. This year, we re-engineered our Responsible AI Standard to adapt more readily to new capabilities, uses, and risks while applying it consistently across the models, platforms, and applications we develop and deploy.

We are strengthening safeguards across the AI lifecycle, from anticipating emerging risks to responding to real-world use. Over the past year, we updated our approach to frontier and sensitive-interaction risks; expanded evaluation, red teaming, controls, and monitoring of agentic systems; and strengthened defenses and incident response readiness for fraud, cyber threats, and other forms of misuse. These efforts reflect a broader shift from point-in-time assessments toward continuous governance.

We are making responsible AI more practical, transparent, and accessible across the AI ecosystem. We are expanding the tools, training, and technical guidance that customers and developers need to govern AI in practice, while investing in more reliable AI performance across languages and cultural contexts. We are also working with governments, researchers, civil society, and industry on shared approaches to evaluation, transparency, and assurance.

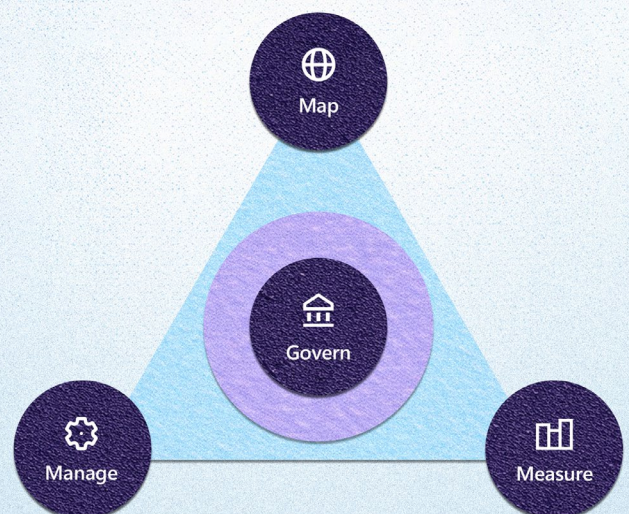
Looking back: What is changing and how we are evolving

The report highlights major shifts shaping governance over the past year:

- Accelerating AI innovation and diffusion even as gaps widen across geographies
- Expanding generative and agentic AI capabilities
- Increasing use of conversational AI in more sensitive contexts
- Scaling AI misuse, increasing people’s exposure to fraud, scams, and cyber exploits
- Evolving AI regulations while trust remains a key adoption enabler

Across these trends, our core commitment is to implement responsible AI with continuous operational discipline that scales with real-world deployment and stays responsive as risk profiles change.

How we develop, release, and operate AI responsibly



This year marks a major step in our governance evolution. We refreshed our Responsible AI Standard to improve agility and clarity across a complex AI lifecycle and value chain, strengthening our alignment with the NIST AI Risk Management Framework (Govern, Map, Measure, Manage) as a continuous loop spanning pre- and post-release practices.

Our updated Standard is intended to help teams meet evolving societal expectations and regulatory requirements more efficiently while continuing to innovate.

This report highlights our expanded investments in evaluations, safeguards, transparency, and post-release learnings to make our governance more scalable and evidence-driven:

- Scaled pre-release checks for generative AI products
- Increased adversarial testing and systematic measurement
- Strengthened defenses against prompt injection and other attacks
- More comprehensive transparency documentation
- More robust post-release monitoring, incident response, and learning processes

When AI systems fall short or are misused, Microsoft is prepared with an incident response program designed to support learning and



drive improvement. The report also describes work to build consistency and readiness amid regulatory uncertainty, including reviewing over 100 enacted and pending regulations to stress-test the Standard's durability.

How we use AI responsibly

This is the first time our Responsible AI Transparency report includes a dedicated view into Microsoft as an AI deployer. It shows how our governance approach goes beyond development to how AI is deployed across internal operations and customer-facing experiences. We outline a practical deployment approach:

- Assess the underlying AI application.
- Deploy safeguards and human oversight.
- Detect and respond to issues over time.
- Inform users to support appropriate use and reliance.

We also use the deployment experience as a feedback loop so that those real-world lessons can strengthen our broader governance and engineering practices to better support customers and partners with their AI deployment efforts.

How we contribute to and learn from the ecosystem

Our work in responsible AI is part of a broader ecosystem effort: trustworthy AI at scale requires shared research foundations, interoperable standards, and practical tools that help organizations implement governance without slowing innovation. The report describes investments in advancing research, deepening partnerships, building stronger communities, contributing to best practices and benchmarks, and equipping customers through compliance leadership, tools, and training.

It also highlights ecosystem-level challenges that require shared progress, such as improving evaluation rigor and comparability, expanding multilingual performance and safety coverage, and strengthening societal resilience to AI-enabled misuse. Together, these efforts aim to support more coherent governance outcomes across jurisdictions and sectors while enabling faster and safer adoption.

Looking ahead

As we look to the year ahead and beyond, we are optimistic about AI's opportunity. The potential for AI to drive many discoveries and address global challenges across all industries is unfolding before us, even as risks continue to

emerge and evolve. That is why our optimistic vision is inseparable from our commitment to responsibility. AI's benefits depend on trust, and our focus is on sustaining it through disciplined action, continuous improvement, and openness about our progress and priorities. As AI scales in capability and reach, we are working to ensure our governance evolves in ways that remain clear, consistent, and grounded in real-world experience—so that innovation can move forward responsibly.

Read the full report

