Microsoft

Grow Your Business with AI You Can Trust

Considerations for business leaders to plan their Frontier Transformation with confidence

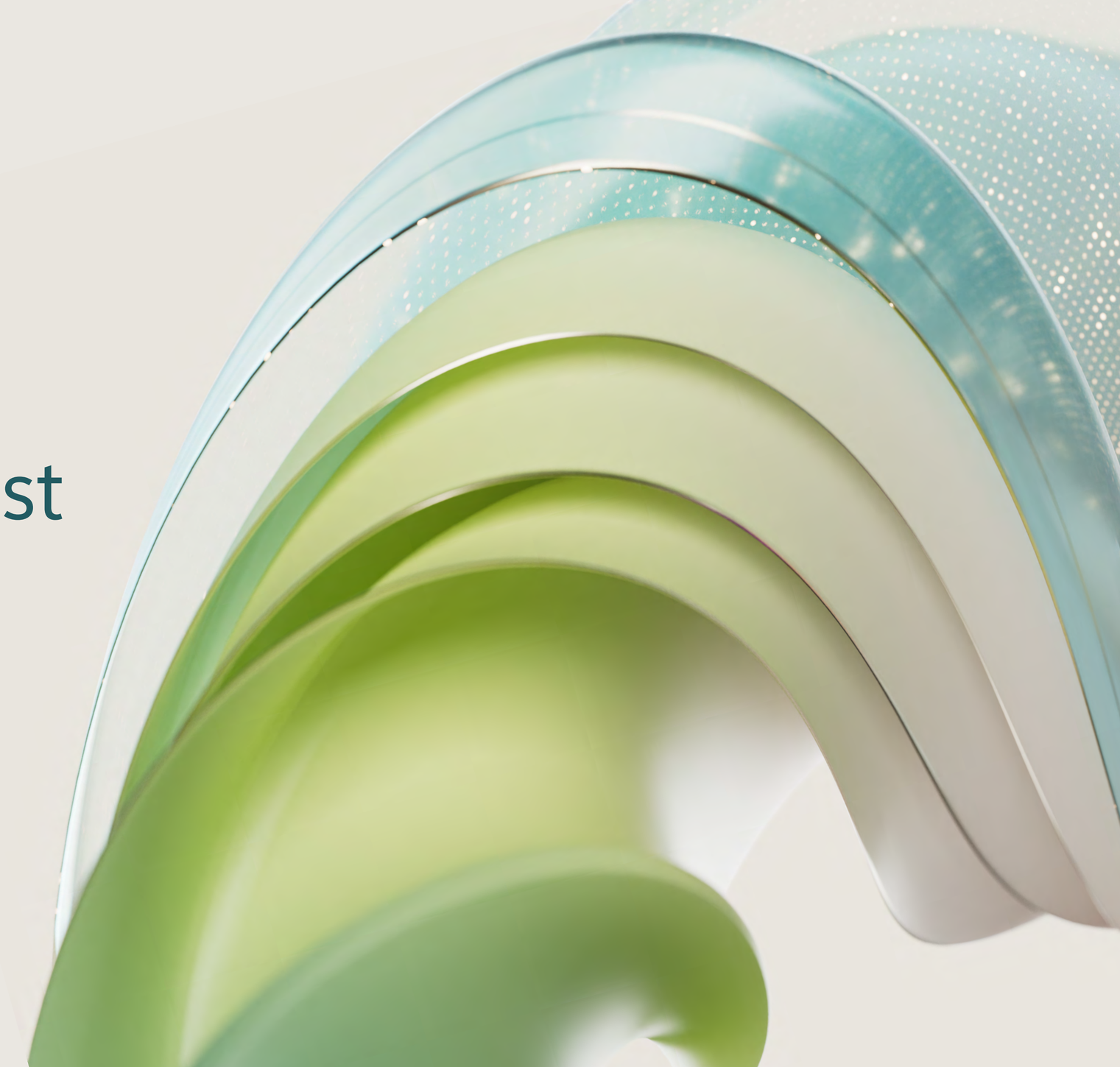


Table of contents

Introduction

AI solutions present a profound opportunity for organizations of every size and in every industry to grow revenue, reduce costs, enhance employee wellbeing, and operate more efficiently. It's no surprise, then, that business leaders are under pressure to adopt AI solutions as quickly as possible and avoid falling behind.

For all the hype around AI, there's as much concern about negative side effects of the technology. In the following sections, we outline several considerations business leaders can plan for to help unlock the promise of this new technology and avoid unintended consequences.

Establishing responsible and secure AI practices for your company helps you safely implement AI tools. And as global AI regulation increases, investing in responsible AI now better equips you to meet new regulatory requirements as they come. Taking a thoughtful approach to the implementation of responsible and secure AI in your business can help leaders embrace AI and innovation.

of business leaders said they were "**equally concerned and excited**" about generative AI.[1]

We're committed to Trustworthy AI and building industry-leading supporting technology. With capabilities that improve security, safety, and privacy, we continue to enable customers to use and build trustworthy AI solutions.

Learn more

Practice responsible AI

Consider responsible AI principles

Public policy and industry best practices are still catching up to the latest developments in AI technology, which leaves business leaders searching for reliable guidance on how to implement AI systems that have a positive impact on businesses, individuals, and society.

Taking the time to think through a responsible AI approach for your organization can help you and your teams move forward with confidence and protect against unintended risks. We've outlined six responsible AI principles for you to consider as you plan and build your own approach. These principles include:

- **Privacy and security**
- **Reliability and safety**
- **Accountability**
- **Inclusiveness**
- **Transparency**
- **Fairness**

Privacy and security

AI systems should adhere to the same privacy and security standards that businesses apply to their most sensitive data.

Prioritize the security of your infrastructure and the privacy of your data.

Know where data is located, how it's used, and confirm that it's secure at rest and in transit. Check that AI tools adhere to your company's values for privacy and security.

When you implement stringent data permissions and assign user permissions based on roles and group memberships, only authorized individuals will have access to sensitive information—reducing the risk of internal breaches.

Additionally, you can maintain security and meet compliance requirements with a governance solution, which includes retaining and logging interactions with AI apps, helping detect any regulatory or organizational policy violations when using those apps, and investigating incidents once they arise.

AI implementation must also comply with all local laws and regulations about data use and privacy. When it comes to data security, it's best to be overcautious and to work with security tool providers with a track record of reliability.

An example of privacy and security in practice:

Use of an AI tool to analyze customer communication to resolve a support ticket could involve access to sensitive or identifiable customer data. Understanding local laws and regulations, adhering to your organization's own high standards for security, and enforcing adequate controls can help keep that sensitive data private.

Complying with AI regulations

Microsoft is committed to building products and solutions that comply with regulations like the EU AI Act to help our customers use AI compliantly.

Learn more

Reliability and safety

Reliability and safety mean AI systems perform as expected, without errors or interruptions. AI tool developers are responsible for making sure their product provides accurate outputs through testing and documentation, but a system of oversight helps verify if the tool is delivering on its intended use. Systems should also undergo regular monitoring, maintenance, feedback, and evaluation processes to identify new uses, troubleshoot and resolve issues quickly, and improve the AI system over time.

Regularly conduct stress testing.

Stress testing prepares an AI system to handle the types of uses and volume of use it's intended to handle without producing errors or becoming vulnerable to risks.

Red teaming is a type of stress testing that involves simulating real-world attacks and using the techniques hackers commonly use to gain access to secure systems. In 2018, Microsoft established our dedicated AI Red Team, and we've expanded the team's mission to map risks outside of traditional security risks, including risks from non-adversarial users compromising responsible AI standards. For example, red teaming a generative AI tool may involve testing whether a user can generate content that stereotypes a marginalized group using the tool. An AI model can also be red teamed to identify potential misuses, scope its capabilities, and understand its limitations. The insights can then be applied to future versions of the model to ensure it will operate reliably and safely.[2]

Conduct due diligence on reliability and safety measures of an AI system at purchase and conduct regular stress testing to identify risks afterward.

Careful review of documentation helps organizations understand what steps the AI system provider has taken to facilitate the reliable and safe use of their system and helps organizations comply with all requirements to operate the system safely.

An example of reliability and safety in practice:

An AI tool is used to model financial outcomes and report on performance. Testing is conducted regularly to ensure the AI tool reliably produces accurate results, avoiding adverse impact on the organization's financial health.

Accountability

In many instances, responsible AI is human-centered. Establishing a clear system of oversight helps your people control the AI tools you implement and stay accountable for the outcomes those tools produce.

Establish a system of oversight that clearly defines roles and responsibilities at every stage of the AI journey.

Implementing a system of oversight that conducts impact testing and responds to impact results, keeps people at the center. This helps protect against potential adverse impacts, and helps ensure adequate controls are implemented at every stage.

Ensure AI tools are fit for purpose.

Regularly assess that AI tools provide the right solutions for the problems they were intended to solve—and determine how your organization will respond if a tool fails to serve its intended purpose.

An example of accountability in practice:

Use of an AI tool to review legal contracts involves oversight by an individual with adequate context and expertise to verify compliance with applicable laws and regulations, and to sign off on the final outcome of the AI-supported review.

Inclusiveness

At its core, inclusivity requires that AI tools be accessible to people of all abilities. That means the tools that business leaders create or procure should follow accessible design principles and comply with the European accessibility standard, EN 301 549; Section 508 of the U.S. Rehabilitation Act; and the Web Content Accessibility Guidelines (WCAG).

Identify opportunities to build inclusivity in your organization with AI.

For example, recipients of Microsoft grants are creating a hiring platform for neurodiverse applicants, building better and more affordable braille displays for students with visual impairment, and creating a web app to help individuals with speech disabilities communicate more effectively.

Transparency

Transparency is a building block of trust. To achieve and maintain transparency, always be clear about how and when AI is being used, as well as its capabilities and limitations.

Be open about how AI is being implemented and used across your organization.

Stakeholders and employees may feel more confident using AI tools when they understand how the tool arrives at its conclusion, but also are clear regarding its limitations. This transparency helps build their capability for using AI-supported tools and knowing when to supplement its outputs with information outside the tool's scope.

Customers want to know when they're interacting with an AI tool, when AI is being used in decision-making, or when an audio or visual asset has been generated or manipulated with AI. Business leaders should consider how that information will be disclosed to customers.

An example of transparency in practice:

Generative AI is used to create content for a marketing campaign, and the organization identifies which elements are created by AI. A specialist reviews the AI-created content to verify its accuracy so it doesn't mislead customers about the features or capabilities of the advertised product.

Fairness

Fair AI implementation allocates opportunities, resources, and information equitably among the people who use and are impacted by AI.

Ensure that your AI systems provide a similar quality of service and delivery of resources and opportunities to all who use it or are affected by it, across demographic groups.

When using AI tools that describe, depict, or otherwise represent people, minimize the potential for stereotyping or demeaning people—especially those of marginalized groups—to promote fairness.

Include members of different backgrounds, experiences, education levels, and perspectives on the team that manages AI implementation, and identify statistical bias in datasets to help drive fairness in an AI system. Human review by subject matter experts in decisions that use AI can also help protect against biased outcomes.

An example of fairness in practice:

An AI tool is used to review applications and identify priority candidates in the hiring process with oversight from a human resources representative. This person confirms that the tool assesses information accurately, without statistical bias, and has the final review in decision-making.

Make informed decisions about AI and safety

Implementing AI responsibly minimizes risks while allowing your business to benefit from the potential of its various uses. Use the following questions as conversation starters with your team as you begin to think through your AI implementation.

Privacy and security

- [ ] Is the data accessed by AI systems secured according to your organization's policies for handling sensitive data?
- [ ] Are you in control of your data, including where it's stored and how it's used?
- [ ] Is your data secured at all times, including when it is in transit from one system to another?
- [ ] Do you have quality security tools in place to defend against third-party access or cyberattacks?
- [ ] Do you have threat identification and response tools in place in case of cyberattacks?

Inclusiveness

- [ ] Have you confirmed that the tools you intend to use meet accessible design principles?
- [ ] Do the tools comply with the European accessibility standard, EN 301 549?
- [ ] Do the tools comply with Section 508 of the U.S. Rehabilitation Act?
- [ ] Do the tools comply with the Web Content Accessibility Guidelines (WCAG)?

Reliability and safety

- [ ] Have the tools you intend to use been adequately tested to minimize errors?
- [ ] Do you have a plan in place to remediate any failures that occur?
- [ ] Will the tools be regularly monitored for reliability issues?
- [ ] Are you prepared to comply with all requirements to operate the tools safely?

Transparency

- [ ] Have you educated stakeholders about how this implementation will work, including its capabilities and limitations, or do you have a plan to do so before they start to use AI tools?
- [ ] Do you have a plan to communicate with employees about how your organization is going to be using AI and how its outputs should be interpreted?
- [ ] Have you determined how and when you'll notify customers that they are interacting with AI or viewing AI-generated content?

Accountability

- [ ] Have you assessed the impact this implementation would have on your employees, organization, and customers?
- [ ] Have you established a system of oversight and response in case of potential negative impacts?
- [ ] Have you implemented data governance and management best practices?
- [ ] Have you determined who will have oversight of AI tools and ensured they have adequate training and control?
- [ ] Have you made sure this solution is fit for its intended purpose?

Fairness

- [ ] Have you ensured this implementation will provide the same quality of service to all affected by it?
- [ ] Have you tested the system's outputs to make sure it will fairly allocate resources and opportunities across demographic groups?
- [ ] Are outputs of the system free of stereotypes and negative portrayals of marginalized groups?

Protect your data and AI systems

Secure your AI tools now and in the future

Implemented incorrectly, any technology that has access to sensitive data can present a security risk for businesses. Because AI systems require a large amount of proprietary data, it's critical to prioritize security from the start when considering AI implementation or procuring AI solutions.

At Microsoft, we launched our Secure Future Initiative, bringing together our learnings to address and prepare for the increasing scale and heightening risks of cyberattacks in the age of AI. The Secure Future Initiative identifies three principles that Microsoft upholds to help secure the entire digital ecosystem:

1. **Secure by design**
   Security comes first when designing any product or service.
2. **Secure by default**
   Security protections are enabled and enforced by default, require no extra effort, and aren't optional.
3. **Secure in operations**
   Security controls and monitoring will be continuously improved to meet current and future threats.

Security is the essential underpinning of any AI implementation. When you ensure basic security hygiene practices, you protect your data, your people, and your devices from more than 98% of cyberattacks.[3]

Effective practices for security hygiene include:

- Enabling multifactor authentication to protect against compromised user passwords and help provide extra resilience for identities.
- Applying Zero Trust principles, which involves explicit verification, use of least privileged access, and assumption of breach, to limit the impact of an attack.
- Using extended detection and response and antimalware to automatically block attacks and gain insight into the security operations software for faster response.
- Ensuring systems are up to date with the latest versions of firmware, operating systems, and applications.
- Implementing the right protections for critical data, which requires knowledge of which data is most important and where it's located.

Manage your AI systems with governance

A strong governance model helps build the solid foundation for responsible AI implementation. It's the role of governments and regulatory bodies to maintain baseline requirements to minimize adverse effects of AI use on society. Commercial organizations also have an ethical responsibility to create a governance structure to manage their own development or use of AI systems according to their organizational values, local laws and regulations, and the greater good.

Establishing your own governance

When creating your organization's system of governance, remember that the purpose of governance is to adhere AI solutions to company policy and responsible AI principles through a series of policies and procedures. This includes applying policies for assessing and implementing third-party AI solutions, coordinating stakeholder involvement and education, and producing documentation to inform employees, customers, and other users about AI tools.

Risks

Map

Mapping risks is the first stage of governing AI and should inform decisions about a tool's safety, reliability, and fitness for purpose. Mapping risks involves running AI impact assessments and privacy and security reviews—including red teaming and stress testing.

Measure

Measuring risks involves developing metrics by which to assess any identified risks and testing planned mitigation methods to determine how effective they will be.

Manage

Managing risks requires organizations to consistently monitor performance. At this stage, you should identify opportunities for user agency and educate stakeholders about responsible use. Human review and oversight should be included in the management process, as well as best practices for transparency according to the responsible AI principles.

Build observability at every layer

Putting innovation in the hands of every employee unlocks speed and creativity. Today, anyone in the organization can build agents, automate workflows, and solve problems in real time. As companies rapidly adopt AI, agents are emerging across teams, functions, and regions. When everyone can create them, you need a new level of observability to ensure what's built is safe, compliant, and aligned with business goals.

Although organizations want agentic AI in the heart of the business, governance pressures, regulatory scrutiny, and internal risk controls keep them grounded. Without full visibility into agents, interactions, and data flows, AI remains untrusted.

To deploy agentic AI, Frontier Firms need understanding:

- What agents exist?
- Who is using them?
- What systems and data do they access?
- What workloads do they drive and what outcomes do they produce?

According to Cyber Pulse (2026), an AI security report, 29% of employees have turned to unsanctioned AI agents.[4] Without clear governance and answers to the questions outlined above, organizations risk the rise of shadow AI and double agents operating outside established controls. These unsanctioned tools can introduce security vulnerabilities, create entry points for malicious actors, disrupt operations, and increase the likelihood of sensitive data leakage.

Getting the most out of your AI agents

Effective observability is built on a set of foundational capabilities that allow you to act quickly on performance, behavior, and risk signals before they impact the business:

Registry

Get a complete view of all agents in your organization, including agents with agent ID, agents you register yourself, and shadow agents.

Agent analytics

Track key agent metrics, including adoption, performance, speed, quality, cost, and ROI to make more informed decisions.

Agent maps

Visualize agent usage, behavior, and trends in one comprehensive agent map to simplify monitoring and accelerate issue identification.

Role-specific oversight

Build extended visibility across your organization through role-based reporting for security leaders to manage agent risks and business leaders to monitor business metrics and ROI.

Observability is an organization-wide responsibility

Observability is critical for deploying and managing trusted AI agents. This introduces unique responsibilities for every layer of organizations to deliver unified visibility and coordination.

IT teams enable and govern every agent running in the environment. This includes operating the AI registry, managing identities and access controls, and enforcing consistent policies. IT teams are also responsible for risk factors like over-permissioned agents.

Developers ensure that security, compliance, and safety guardrails are implemented. They must also continuously evaluate models for vulnerabilities, monitor runtime performance and cost, observe production behavior, and apply guardrails consistently across all agents.

Security teams protect the entire AI ecosystem. They're responsible for detecting agent sprawl and inappropriate access, validating regulatory and compliance obligations, identifying data oversharing and leakage risks, and evaluating AI-specific threats and vulnerabilities.

Read more about Agent 365

Address digital sovereignty requirements

Digital sovereignty has become urgent because nations see cloud, data, and AI as the backbone of economic competitiveness and digital resilience. For business leaders, it's a board-level issue too because of fragmented regulations, rising cyber threats, and geopolitical volatility.

Across the globe, regulations for AI, cybersecurity, and privacy are evolving quickly, reshaping how organizations use AI, manage data, and secure digital services. In fact, there are over 1,000 global policy initiatives across 69 countries, and over 100 nations enforcing privacy laws.[5] The result is pressure to innovate responsibly, meeting regulatory expectations without slowing progress.

Organizations encountering one or more of the following scenarios often evaluate how digital sovereignty fits into their cloud and AI strategy.

1. **You operate in markets with evolving regulatory requirements.** Data and operations must align with local and national compliance standards at scale.
2. **You are scaling AI across regions and need clear governance over data processing.** AI workloads require defined storage and processing boundaries, trustworthy AI controls, and flexibility in model choice across a broad ecosystem.
3. **You must ensure that no external operator can access your sensitive data.** Your data and workloads need enforceable controls against unauthorized access even from the cloud provider.
4. **You are required to keep data within specific geographic boundaries.** You need the ability to choose where data is stored and processed to meet local laws and privacy expectations.
5. **You depend on uninterrupted digital services.** Mission-critical systems must continue operating securely, even during natural or geopolitical disruptions.

Digital sovereignty doesn't have to slow innovation. It's about managing risk so you can scale AI using the tools and environments your business depends on. The following three principles provide a foundation for building cloud environments that are compliant, resilient, and innovation ready.

1. **Sovereignty built into the cloud with no migration required.** Data and operations must align with local and national compliance standards at scale.
2. **Trust, transparency, and accountability.** AI workloads require defined storage and processing boundaries, trustworthy AI controls, and flexibility in model choice across a broad ecosystem.
3. **Flexible deployment models tailored to risk and priorities.** Different workloads carry different operational and regulatory considerations, so deployment models should be adjustable to fit local requirements and business priorities.

Learn more about Microsoft Sovereign Cloud

Realize the potential of AI

How companies have transformed with safety in mind

Used responsibly and securely, AI can improve business operations, help your organization address your most pressing challenges, and establish trust and confidence with your customers. Learn from real-word examples and impact to understand how.

Wipro

Objective

Wipro has long prioritized AI, training 200,000 employees on GenAI principles. The goal was to embed AI across the business to deliver faster value and better outcomes while ensuring adoption is responsible and aligned with business needs.

Solution

To guide responsible adoption, Wipro established an AI council. In bi-weekly sessions, leaders created a responsible, persona-driven model to implement Microsoft Copilot across the organization. By defining distinct roles such as sales, developers, and CTO teams, Wipro ensured each group used the Copilot tools best suited to their work.

Impact

Wipro successfully scaled AI with strong governance and measurable outcomes. Most Copilot use cases were approved due to clear business value, and employees reported increased productivity and efficiency. By treating AI as a strategic business tool not a novelty Wipro demonstrated how Responsible AI can drive innovation, trust, and real impact.

Learn more

Accenture

Objective

Accenture (800,000+ employees) saw AI innovation stall in proof-of-concept. Clients wanted production-grade AI—safety, and accuracy at scale—but delivery was slowed by disconnected tools and manual handoffs.

Solution

They built a centralized enterprise AI platform on Azure AI Foundry with observability baked in from evaluation through deployment. Using a single SDK, they unified safety, accuracy, and performance tracking, eliminating tool sprawl and stitching across fragmented systems.

Impact

Accenture enabled clients to move beyond generative AI demos to production-grade, scalable, and compliant applications. The teams significantly accelerated time to market, reducing AI application build time by up to 50%. As a result, Accenture reports up to a 30% increase in overall efficiency while scaling trusted AI solutions.

Learn more

Objective

POST Luxembourg sought to modernize its core banking system to improve agility, meet new EU instant payment regulations, and expand financial inclusion while ensuring compliance with European requirements and maintaining control within its national regulatory environment.

Solution

The organization began migrating its core banking system to Microsoft Azure, becoming the first public sector institution in Luxembourg to do so. By moving to the cloud, POST Luxembourg was able to deploy new payment modules quickly, meet Single Euro Payments Area (SEPA) instant payment obligations, and modernize development environments.

Impact

With its core banking platform in the cloud, POST Luxembourg gained the flexibility to add, test, and scale services without downtime, respond rapidly to regulatory change, and strengthen its ability to deliver inclusive banking.

The transformation supports digital sovereignty by aligning infrastructure, compliance, and innovation within the European regulatory landscape.

Learn more

Raiffeisen Bank International

Objective

Raiffeisen Bank International (RBI) aimed to automate repetitive tasks, accelerate document summarization across legal and regulatory content, and enhance customer service—while ensuring security, privacy, and compliance with European regulations as part of its broader digital transformation.

Solution

RBI built its own internal ChatGPT using Azure OpenAI in Foundry Models and Azure AI Search within Microsoft Foundry. The solution enables employees to summarize financial and regulatory documents, draft communications, extract insights from reports, and automate common inquiries, all within a secure and compliant cloud environment.

Impact

RBI deployed the solution first to 2,000 users and expanded to more than 20,000 active users. Employees report higher productivity, faster resolution of customer issues, and greater collaboration—establishing a scalable foundation for generative AI adoption across the organization.

Learn more

Advance the sustainability of AI

Just as security is an essential foundation of responsible AI, sustainability is crucial to the conscientious use of it. A powerful tool in understanding and reducing environmental impact, AI can help businesses advance their sustainability goals and help private businesses and public institutions better understand and pursue environmental conservation, resource management, and climate change mitigation.

With AI data management tools and reporting, organizations can bring visibility into their sustainability activities so they can record, report, and reduce their environmental impact. AI can provide the insight necessary to make more informed decisions and stay on track toward your sustainability goals.

At the same time, we recognize the resource intensity of these applications and the need to address environmental impact from every angle.

Business leaders can make their own datacenters more sustainable or work with providers that are already taking these actions to reduce the environmental impact of the datacenters that fuel their AI solutions.

At Microsoft, we're deeply invested and are increasing our focus in three main areas in line with our sustainability commitments: optimizing datacenter energy and water efficiency, advancing low-carbon materials, and improving the energy efficiency of AI and cloud services—all with the goal of empowering our customers and partners with tools for collective progress.

Begin your Frontier Transformation

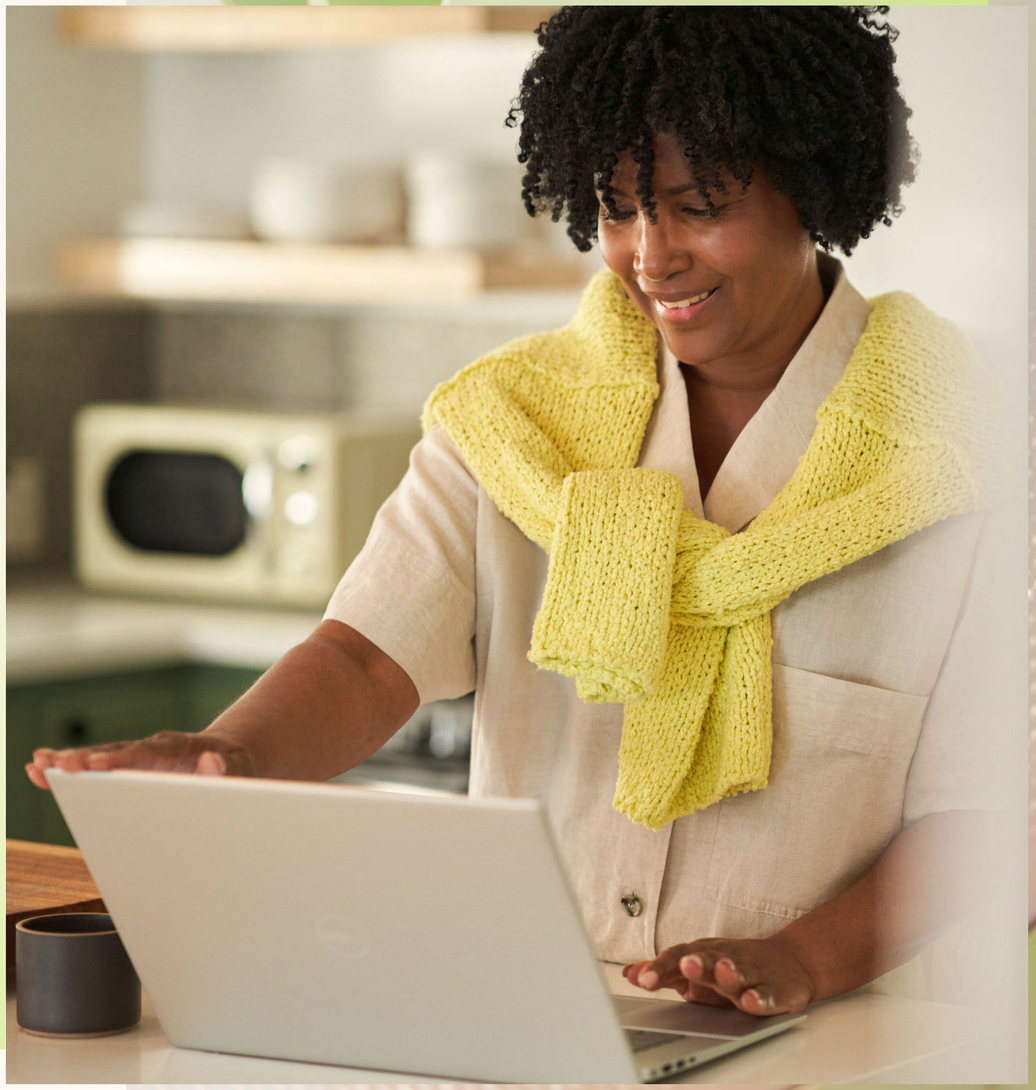
By carefully considering responsible AI practices and prioritizing security throughout your organization, you can explore the potential of AI to help grow your business.

Our commitments and capabilities make it possible for you to accelerate Frontier Transformation with confidence, and you can trust Microsoft to put your AI security, privacy, and safety first.

Learn more about Microsoft AI to begin your journey.

Discover how Microsoft is committed to responsibly developing and advancing AI.

Explore how Microsoft helps you safeguard AI with comprehensive security and governance solutions.



Sources

[1] "What Business Leaders Really Think About Generative AI," INSEAD, April 11, 2024
https://knowledge.insead.edu/leadership-organisations/what-business-leaders-really-think-about-generative-ai

[2] "2024 Microsoft Responsible AI Transparency Report," Microsoft, accessed July 10, 2024
https://www.microsoft.com/corporate-responsibility/responsible-ai-transparency-report

[3] Quy Nguyen, "Basic cyber hygiene prevents 98% of attacks," Microsoft Tech Community, September 18, 2023
https://techcommunity.microsoft.com/t5/security-compliance-and-identity/basic-cyber-hygiene-prevents-98-of-attacks/ba-p/3926856

[4] "Cyber Pulse: An AI Security Report," Microsoft Security Experts, February 10, 2026
https://www.microsoft.com/en-us/security/security-insider/emerging-trends/cyber-pulse-ai-security-report

[5] "The AI regulations that aren't being talked about," Deloitte Insights, Deloitte, November 10, 2023
https://www.deloitte.com/us/en/insights/industry/government-public-sector-services/ai-regulations-around-the-world.html

OECD.AI: The OECD's Hub for AI Policy, Organisation for Economic Co-operation and Development, https://oecd.ai/en/

"Building a Foundation for AI Success: Governance," Microsoft Cloud Blog, Microsoft, March 28, 2024
https://www.microsoft.com/en-us/microsoft-cloud/blog/2024/03/28/building-a-foundation-for-ai-success-governance/

"2025 AI Index Report," Stanford Institute for Human-Centered Artificial Intelligence
https://hai.stanford.edu/ai-index/2025-ai-index-report.

"AI Regulations Around the World," Mind Foundry Blog, Mind Foundry, January 13, 2026
https://www.mindfoundry.ai/blog/ai-regulations-around-the-world

"Identifying Global Privacy Laws Relevant to DPAs," IAPP News, International Association of Privacy Professionals, March 19, 2024
https://iapp.org/news/a/identifying-global-privacy-laws-relevant-dpas

"Data Protection and Privacy Legislation Worldwide," UNCTAD, United Nations Conference on Trade and Development, February 17, 2026
https://unctad.org/page/data-protection-and-privacy-legislation-worldwide