Microsoft Security

How Microsoft Secures Generative AI

EVERON
SWITCHED ON I.T

Introduction

We are at the start of an exciting new age: an age of limitless possibilities, an age of generative AI. Like nothing else in human history, generative AI has emerged as an engine for innovation, with vast applications and new use cases that continue to reveal themselves every day.

Because the technology is rapidly changing business and the world at large, and its unbridled potential has captured the popular imagination like nothing else in recent memory, we at Microsoft feel a responsibility to share what we have learned about using this technology safely.

We take this responsibility seriously. Microsoft is committed to security, privacy, and compliance across everything we do, and our approach to AI has been no different.

In 2016, we began our work on responsible AI. By 2018, we had identified our Responsible AI Principles, and by 2019, we became the first major cloud provider to create a permanent Office of Responsible AI to both govern our AI program and provide actionable guidance for engineering teams building AI systems.

During our decade-long focus on delivering AI to our customers, we have developed standards and best practices to address them—standards and best practices we share openly.

No matter which AI solutions you choose—one of the Microsoft Copilot offerings, your own AI application built on the Azure AI platform, or an AI system offered elsewhere—we want to help you use AI safely and responsibly across systems in a way that keeps your data secure and private.

How Microsoft ensures generative AI remains safe to use

We believe the possibilities for generative AI are limitless. We take a holistic approach to generative AI security that considers the technology, its users, and society at large across four areas of protection: data privacy and ownership, transparency and accountability, user guidance and policy, and secure by design.

Data privacy and ownership

Microsoft applies our privacy commitments to all Copilot experiences and generative AI software products. Through our transparent data protection and privacy policies, we ensure customer data remains private. We empower customers to retain control of their information, which will never be used to train foundational models or be shared with OpenAI or other Microsoft customers without authorized user permission.

Transparency and accountability

Generative AI sometimes gets it wrong. To make sure that the content created by generative AI is credible, it's essential for the AI to (1) use accurate, authoritative data sources; (2) showcase reasoning and sources to maintain transparency; and (3) encourage an open dialog with the provision for feedback—an avenue that permits users to contribute substantially to the enhancement of AI results.

User guidance and policy

To mitigate potential overreliance, we encourage users to think critically about the information served with generative AI by using carefully considered language and referencing cited sources. We also consider hostile misuse, where users try to engage the AI in harmful actions, like generating dangerous code or instructions to build a weapon. To shield against this kind of misuse, we layer deep safety protocols into the system, setting clear boundaries on what AI can and cannot do to maintain a safe and responsible usage environment.

Secure by design

To prepare for generative AI threat vectors, we've added new steps to our Security Development Lifecycle, including updating the Threat Modeling SDL requirement to account for AI and machine learning-specific threats and mandating that teams adhere to the Responsible AI Standard. We also continually monitor and log our large language model interactions for threats and implement strict input validation and sanitization of user-provided prompts. Finally, we put all of our generative AI products through multiple rounds of AI red teaming to look for vulnerabilities and ensure we have proper mitigation strategies in place.

AI Red Team

Instituted in 2018, Microsoft's AI Red Team mirrors the tactics and techniques of potential adversaries to find and fix vulnerabilities.

The team's charge extends beyond just securing against potential threats; it encompasses a critical examination of other system failures, including the generation of potentially harmful content, providing us with a comprehensive picture of the system's integrity, confidentiality, and availability. The world of AI is always in flux, and as such, our red teaming efforts are relentless and adaptive, embracing an ongoing cycle of testing both before and after product release.

Making AI safe for everyone

AI adoption rates are increasing fast—frequently without the knowledge or oversight of management—and demand for AI applications continues to rise exponentially. As a business decision-maker, embracing generative AI is a strategic move that can enhance your organization, and make your teams more efficient.

Is your organization ready? Here’s how to get started:

Step 1: Implement a Zero Trust security model

The Zero Trust security model uses rich intelligence and analytics to ensure every access request is fully authenticated, authorized, and ecrypted before granting access. Instead of assuming everything behind the corporate firewall is safe, the Zero Trust model assumes breach and verifies each request as though it originates from an open network.

Learn more about Zero Trust >

Step 2: Adopt cyber hygiene standards

The Microsoft Digital Defense Report 2023 shows that basic security hygiene still protects against 99% of attacks. Meeting the minimum standards for cyber hygiene is essential for protecting against cyberthreats, minimizing risk, and ensuring the ongoing viability of the business.

Learn more about cyber hygiene >

Step 3: Establish a data security and protection plan

For today's environment, a defense-in-depth approach offers the best protection to fortify your data security. There are five components to this strategy, all of which can be enacted in whatever order suits your organization's unique needs and possible regulatory requirements.

Learn more about data security and protection >

Step 4: Establish an AI governance structure

In order for organizations to get AI ready, it's critical that they implement processes, controls, and accountability frameworks that govern data privacy, security, and development of their AI systems, including the implementation of responsible AI standards.

Learn more about AI governance >

Microsoft Security

EVERON
SWITCHED ON I.T