The hidden cyber risks of deploying generative AI - Against Invaders - Noti

Data: 2025-09-26 14:36:18

Autor: Inteligência Against Invaders

Organizations increasingly think it’s a great idea, even an absolute necessity, to integrate artificial intelligence into their operations. And it can be both. But many organizations don’t understand the cybersecurity risks involved with AI, and they don’t realize how unprepared they are to secure their AI deployments.

Whether for internal productivity or customer-facing innovation, AI — especially generative AI — can revolutionize a business.

But if they’re not secure, AI deployments can lead to more problems than benefits. Without proper safeguards, AI can introduce vulnerabilities that open the door to cybercriminals rather than strengthen defenses.

AI adoption outpaces security readiness

The appetite for AI is undeniable. According to EY, 92% of technology leaders expected to increase AI spending in 2025, a 10% increase over 2024. Agentic AI is emerging as a particularly transformative frontier, with 69% of technology leaders saying their organizations need it to staycompetitive.

Unfortunately, organizations aren’t thinking enough about security. The World Economic Forum (WEF) reports that 66% of organizations believe AI will significantly affect cybersecurity in the next 12 months, but only 37% have processes in place to assess AI security before deployment. Smaller businesses are even more exposed—69% lack safeguards for secure AI deployment, such as monitoring training data or inventorying AI assets.

Accenture finds similar gaps: 77% of organizations lack foundational data and AI security practices, and only 20% express confidence in their ability to secure generative AI models.

In practice, that means most enterprises are embracing AI with little assurance that their systems and data are truly protected.

Why insecure AI deployments are dangerous

Deploying AI without security can be a major compliance risk. Beyond that, it actively empowers cyberattackers, who can exploit generative AI in several ways:

- **AI-driven phishing and fraud.** WEF notes that 47% of organizations view AI-enabled

cyberattacks as their top concern. And for good reason: 42% of organizations experienced social engineering attacks last year.
- **Model manipulation.** Accenture highlights how AI worms such as Morris II can embed malicious prompts into models, hijacking AI assistants to exfiltrate data or spread spam.
- **Deepfake-enabled scams.** Criminals increasingly use AI-generated voices, images and videos to commit fraud. One attack impersonated Italy's defense minister with convincing voice deepfakes, defrauding prominent business figures into wiring funds abroad.

AI lowers the barrier to entry for attackers, making scams faster, cheaper and harder to detect.

Building security Into AI from the start

If organizations want to realize the full benefits of AI safely, they need to adopt a security-first mindset. Instead of retrofitting defenses after incidents or cobbling together multiple disparate tools, companies should seek natively integrated cybersecurity solutions from the outset. With solutions that are easy to manage from a central console and work together without manual integrations, organizations can:

- **Embed security into AI development pipelines.** Secure coding, data encryption and adversarial testing should be standard at every stage.
- **Continuously monitor and validate models.** Organizations need to test AI systems for manipulation, data poisoning and other emerging risks.
- **Unify cyber resilience strategies.** Security cannot be siloed. Defenses should be natively integrated across endpoints, networks, cloud environments and AI workloads. This strategy reduces complexity and ensures attackers cannot exploit weak links.

Both WEF and Accenture emphasize that the organizations best prepared for the AI era are those with integrated strategies and strong cybersecurity capabilities.

Accenture's research shows that only 10% of companies have reached what it calls the "Reinvention-Ready Zone," which combines mature cyber strategies with integrated monitoring, detection and response capabilities. Firms in that category are 69% less likely to experience AI-powered cyberattacks than less prepared organizations.

The role of MSPs and enterprises

For managed service providers (MSPs), the AI wave presents both a challenge and an opportunity. Clients will increasingly demand AI-powered tools, but they will also rely on their MSPs to keep them secure.

According to the Acronis Cyberthreats Report H1 2025, cyberattackers have ramped up their AI-enabled attacks on MSPs. More than half of all attacks on MSPs in H1 2025 were phishing attempts, largely driven by AI capabilities.

So, MSPs have to provide integrated protection that spans cloud, endpoint and AI environments, ensuring they can protect themselves and their clients.

For enterprises, the path forward is about balancing ambition with caution. AI can boost efficiency, creativity and competitiveness, but only if deployed responsibly.

Organizations should make AI security a board-level priority, establish clear governance frameworks, and ensure their cybersecurity teams are trained to address emerging AI-driven threats.

The future of AI deployments is tied to security

Generative AI is here to stay, and it will only become more deeply embedded in business operations. But rushing ahead without securing these systems is like building a skyscraper on sand: The foundation is too weak to support the structure.

By adopting integrated, proactive security measures and solutions, organizations can harness AI's potential without amplifying their exposure to ransomware, fraud and other evolving threats.

About TRU

TheAcronis Threat Research Unit (TRU)is a team of cybersecurity experts specializing in threat intelligence, AI and risk management. The TRU team researches emerging threats, provides security insights, and supports IT teams with guidelines, incident response and educational workshops.

See the latest TRU research

Sponsored and written by Acronis.