



Getting Started with AI Security:

AI Risks, How to Prevent Them, and AI for Defenders

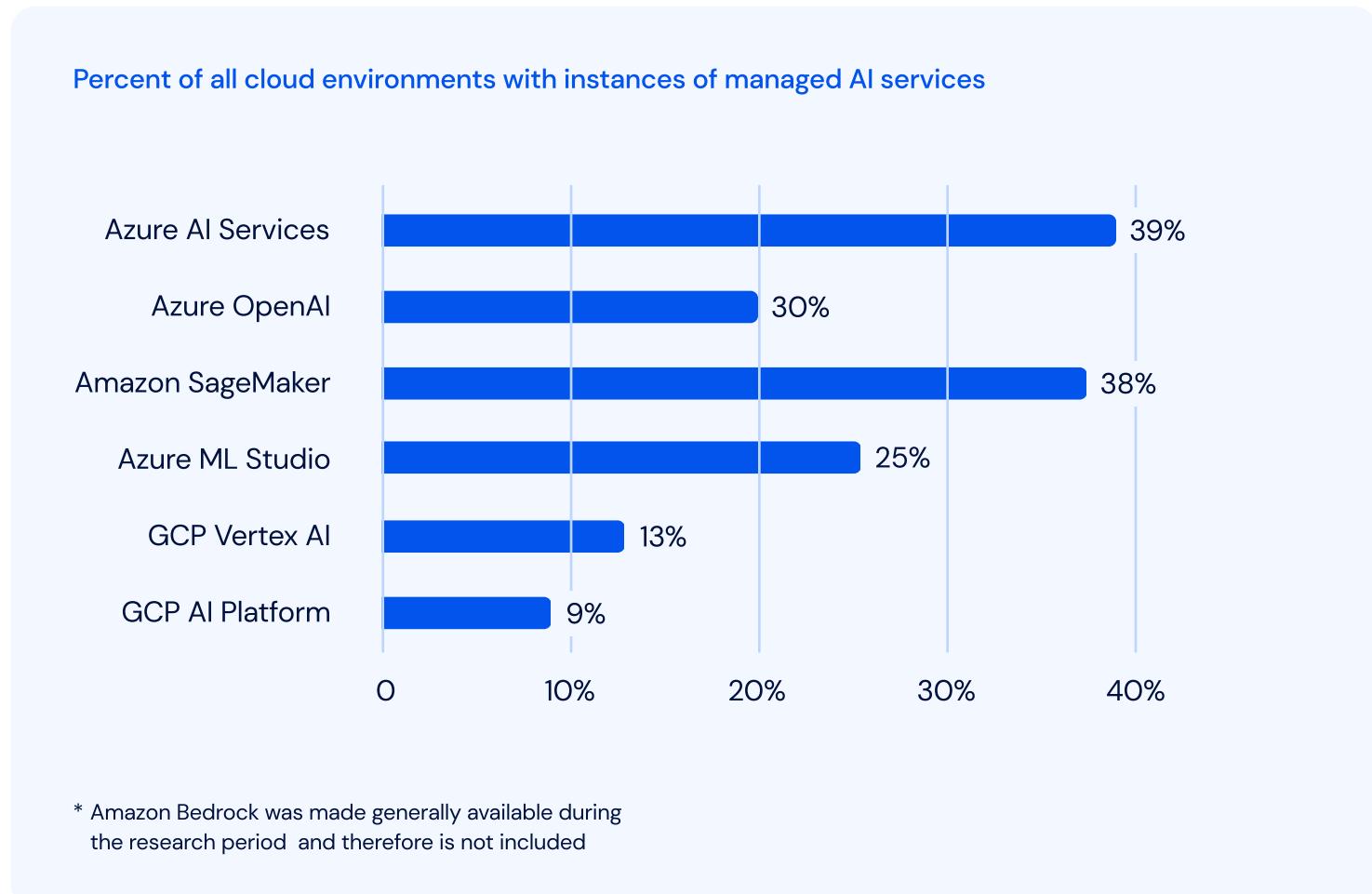


Table of Contents

Introduction	3
Top AI Security Risks	4
5 Tips to Fight AI Security Risk	5
AI-SPM: Core Features & Mini RFP	6
AI for Defenders	7
Conclusion	8

Introduction

With AI poised to reshape entire industries, every organization is now seeking to use it for business gain. AI is everywhere: Wiz researchers observe AI present in more than 70% of cloud environments¹. Azure, unsurprisingly, leads the pack:



Sanctioned or not, AI is here.

Wiz customers include 40% of the Fortune 100, as well as many of the world's fastest growing born-in-the-cloud companies. Regardless of size, industry, or maturity level, many of them ask us the same question: What's the best, safest way to use this powerful new technology?

In this ebook, we'll delve into the intersection of AI and cybersecurity, offering insights, outlining common challenges, and sharing success strategies for security teams. Key topics we'll cover:

- AI risks and best practices for mitigation of critical risks
- How to safeguard your AI development pipeline with AI-SPM
- Using AI to power security, including operations and incident response

¹ <https://www.wiz.io/blog/key-findings-from-the-state-of-ai-in-the-cloud-report-2024>

Top AI Security Risks

In this section, we will summarize several critical AI security risks. Our focus will be on attack types, versus regulatory implications or the risks inherent in Shadow AI, i.e. unknown instances of AI usage that lie outside the reach of corporate governance policies. (Shadow AI does, of course, pose significant risk, largely because it can be exploited by malicious actors in the ways we outline below.)

Most AI security risks fall into 4 primary categories:

Adversarial attacks, in which a malicious actor manipulates AI systems by feeding them carefully crafted inputs to deceive or compromise their functionality. These attacks can lead to bad decisions or breaches in security protocols.

Model inversion attacks, where attackers attempt to reverse-engineer AI models by exploiting outputs to infer sensitive information about the training data or the model itself. This poses a serious threat to data privacy and confidentiality.

Data poisoning, a technique used to corrupt training data with malicious inputs. Data poisoning can compromise the integrity of AI models, leading to biased or inaccurate outcomes during decision-making processes.

Model theft, where adversaries steal trained AI models to replicate their functionality or gain insights into proprietary algorithms. This intellectual property theft can undermine competitive advantage and compromise business interests.

Later on we'll explore the security measures and best practices to effectively mitigate these risks and safeguard AI-driven initiatives from potential threats and vulnerabilities.

Security Researchers Investigate the Risk Inherent to AI Systems

Wiz Research conducts ongoing investigations into AI infrastructure to examine the security implications of this powerful technology. Recent discoveries include:

- “Problama” (CVE-2024-37032), an easy-to-exploit Remote Code Execution vulnerability in the open-source AI Infrastructure project Ollama. [Read more](#).
- Hugging Face addressed security risks in AI infrastructure, including vulnerabilities that could allow malicious actors to execute code, escalate privileges, and compromise cross-tenant environments. [Read more](#).
- A critical vulnerability in the Replicate AI platform allowed remote code execution and potential access to private AI models and user data. [Read more](#).

In aggregate, the researchers' work underscores the reality that these tools are often at an early stage of development and lack standardized security features, such as authentication. Additionally, due to their young code base, it is relatively easier to find critical software vulnerabilities, making them perfect targets for potential threat actors.

5 Tips to Fight AI Security Risk

Given the risks present in AI systems, what's the best way to secure them? We see 5 primary strategies:



1. Strong Tenant Isolation:

Unsurprisingly, given that we deem this a critical risk, tenant isolation tops our list. Effective tenant isolation ensures that different users' data and workloads are securely separated, preventing unauthorized access. Implement rigorous privilege management, encryption, and authentication measures to maintain distinct and secure boundaries between tenants.



2. Conduct Regular Audits:

Continuous and thorough security audits can help identify and mitigate potential vulnerabilities in AI infrastructure. These audits should include penetration testing, vulnerability assessments, and code reviews to uncover and address security weaknesses.



3. Comprehensive Threat Modeling:

Threat modeling helps anticipate potential security threats so defenders can design systems to withstand them. By understanding the possible attack vectors and scenarios, you can implement targeted defenses to safeguard AI systems against specific threats.



4. Secure Development:

Secure development practices, such as using secure coding standards, incorporating security into the development lifecycle (DevSecOps), and conducting regular security training for developers, can significantly reduce the risk of introducing vulnerabilities into AI applications.



5. Technology Solutions and Frameworks:

Leverage advanced security solutions, tools, and frameworks that enhance visibility and control over your AI environment. These can help you monitor, detect, and respond to security incidents promptly, ensuring continuous protection of AI infrastructure.

AI security solutions abound, and sifting through the noise could be a full-time job in and of itself. In our next section, we'll unpack #5 on this list a bit more by talking about how to evaluate AI Security Posture Management (AI-SPM) solutions.

AI-SPM: Core Features & Mini RFP

AI-SPM solutions protect AI systems from vulnerabilities, offering comprehensive visibility, continuous monitoring, and threat detection across AI environments. With the right AI-SPM solution, you should be able to answer “yes” to the following questions:

Does my organization know what AI services and technologies are running in my environment?

Do I know the AI risks in my environment? Can I prioritize what’s critical?

Can I detect misuse in my AI pipeline?

Key features include robust access controls, data encryption, and compliance management. A strong AI-SPM solution will emphasize proactive risk management, providing real-time insights and automated responses to potential threats and ensuring the integrity and security of AI operations.

The following 10 questions can act as a “mini RFP” for any would-be buyer evaluating any of the AI-SPM offerings on market:

1. Is the offering a standalone, or is it a feature of a larger platform with additional cloud security capabilities?
 - a. Will this be additive to my current tech stack?
 - b. Will both security and development teams benefit?
2. How does the offering provide visibility into AI environments and detect potential vulnerabilities, misconfigurations, exposed data and other risks?
 - a. Can it continuously monitor AI systems in real-time?
 - b. What specific threats can it detect?
3. What are the access control and authentication mechanisms that are supported?
4. How does it ensure data encryption and protect sensitive information?
5. What compliance frameworks are supported?
6. How does the technology handle incident response?
 - a. What automated response features are available?
7. Can it integrate with my existing security and IT infrastructure?
8. What reporting and analytics capabilities does it offer?
9. How does the tool support proactive risk management?
 - a. What preventive measures does it provide?
10. What kind of support and training resources are available?
11. Can my AI engineers easily leverage the solution to understand risk?
 - a. What type of remediation information does the solution provide?

[More on Wiz AI-SPM.](#)

AI for Defenders

Given AI's relative newness and omnipresence, it presents an attractive target for cybercriminals. Until this point, we've largely focused on the risks inherent to AI systems, and how to mitigate them.

One stone remains unturned: how AI can benefit defenders.

There are many wonderful examples of how security teams can harness the power of traditional and generative AI to cope with growing demands and maximize impact. These gains fit nicely into the "people / process / technology" framework.

People: Bridging the Skills Gap

Use AI to upskill existing staff, enabling non-security experts to handle security tasks by converting natural language to queries and generating remediation steps quickly.

Process: Increase Efficiency and Velocity

Implement AI to benchmark against industry standards, detect control gaps, and streamline processes by analyzing large data sets swiftly.

Technology: Enhance Threat Detection

Employ AI-powered tools for real-time anomaly detection and data classification, which can accelerate incident response and improve overall security posture.

AI can enhance cybersecurity by automating threat detection, analyzing large datasets to identify patterns, and reducing response times. It aids in vulnerability management efforts by quickly identifying and addressing potential security gaps, and minimizes human error in routine tasks, freeing up cybersecurity professionals to focus on complex and adaptive threats.

Overall, AI complements – not replaces – human expertise by handling repetitive tasks and providing advanced data analysis, improving the efficiency and effectiveness of cybersecurity measures.

Conclusion

AI has cemented its status as a business technology mainstay. It will no doubt bring more challenges, opportunities, and new questions we haven't even considered yet.

Today, security teams are grappling with the best practices for securely implementing AI technologies. Very real risks exist, highlighting the critical need for robust security measures and frameworks tailored to AI environments.

To mitigate these risks effectively, organizations are advised to adopt comprehensive security strategies. These include implementing strong tenant isolation and leveraging AI-SPM solutions that empower organizations to proactively manage AI-related risks and ensure the integrity of their AI operations.

Though AI is unlikely to entirely replace human experts, it excels at processing vast amounts of data, identifying patterns, and automating repetitive tasks. It enhances threat detection, incident response, and vulnerability management by quickly analyzing and correlating data, which would be labor-intensive for humans.

Ultimately, while AI presents unprecedented opportunities for innovation and efficiency, its widespread adoption necessitates crystal-clear visibility and proactive security measures to ensure organizations are harnessing AI's potential while safeguarding against emerging threats, ensuring a secure future.