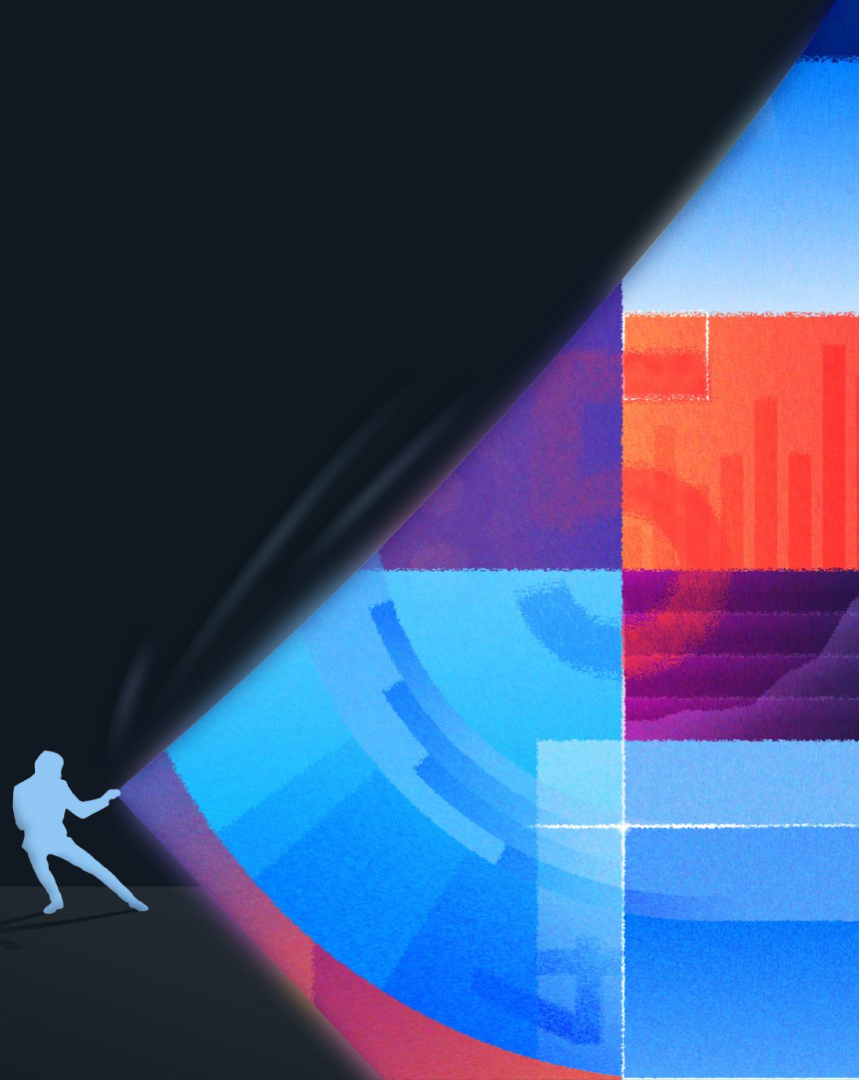
orca security

2024

State of AI Security Report

Unveiling the numbers and insights behind the prevalence of AI risks in the cloud



Inside This Report

Foreword

AI usage is exploding. Gartner predicts that the AI software market will grow 19.1% annually, reaching $298 billion by 2027. In many ways, AI is now in the stage reminiscent of where cloud computing was over a decade ago.

At that time, speed of innovation was the focus, and it came at the expense of security. One such example was where storage buckets were spun up at the speed of the cloud, but were being left exposed to the Internet - without considering the security implications.

Fast forward to today, we are now witnessing the signs that history may repeat itself. Many AI services are defaulting to wide access and full permissions, focusing on speed of delivery while sacrificing security measures.

Yet unlike a decade ago, we are now more prepared to secure emerging AI technologies and models. Awareness and education play a key role in achieving this goal, which is why we are releasing this inaugural report.

We hope the report will help developers, CISOs, and security professionals better understand how to secure their AI models, while not slowing down innovation.

Thank you for reading our research.

Gil Geron
CEO and Co-Founder of Orca Security

About the Orca Research Pod

The Orca Research Pod is a group of cloud security researchers that discover and analyze cloud risks and vulnerabilities to strengthen the Orca Cloud Security Platform and promote cloud security best practices.

Research Methodology

This report focuses on the security of deployed AI models in cloud services and environments. It was compiled by analyzing data captured from billions of cloud assets on AWS, Azure, Google Cloud, Oracle Cloud, and Alibaba Cloud scanned by the Orca Cloud Security Platform.

Report Data Set:

- Cloud workload and configuration data
- Billions of real-world production cloud assets
- Data referenced in this report was collected from January - August 2024
- AWS, Azure, Google Cloud, Oracle Cloud, and Alibaba Cloud environments

25+ vulnerabilities discovered on AWS, Azure, and Google Cloud

2024

- System:authenticated default Google Kubernetes Engine (GKE) group
- LeakyCLI in AWS and Google Cloud

2023

- Azure Digital Twins SSRF
- Azure Functions App SSRF
- Azure API Management SSRF
- Azure Machine Learning SSRF
- Azure Storage Account Keys Exploitation
- Azure Super FabriXss
- Two Azure PostMessage IFrame Vulnerabilities
- Bad.Build Supply Chain Risk in GCP
- 8 Cross-Site Scripting (XSS) vulnerabilities on Azure HDInsight
- Unauthenticated Access Risk to GCP Dataproc

2022

- AWS BreakingFormation
- AWS Superglue
- Databricks
- Azure AutoWarp
- Azure SynLapse
- Azure FabriXxs
- Azure CosMiss

This report harnesses unique insights from scans performed by the **Orca Cloud Security Platform** , and uncovers key AI security risks and considerations for CISOs, developers, and security professionals. The AI security risks discussed in this report are mapped to each of the OWASP Top 10 Machine Learning Risks.

Executive summary

Our three primary findings are as follows:

1. **More than half of organizations are deploying their own AI models**
   We found that **56%** of organizations have adopted AI to build custom applications. Azure OpenAI is currently the front runner among cloud provider AI services, with **39%** of organizations with Azure using it. Sckit-learn is the most used AI package (**43%**) and GPT-35 is the most popular AI model, with **79%** of organizations using GPT-35 in their cloud.

2. **Default AI settings are often accepted without regard for security**
   The default settings of AI services tend to favor development speed rather than security, which results in most organizations using insecure default settings. For example, **45%** of Amazon SageMaker buckets are using non randomized default bucket names, and **98%** of organizations have not disabled the default root access for Amazon SageMaker notebook instances.

3. **Most vulnerabilities in AI models are low to medium risk - for now**
   62% of organizations have deployed an AI package with at least one CVE. Most of these vulnerabilities are low to medium risk with an average CVSS score of **6.9**, and only **0.2%** of the vulnerabilities have a public exploit (compared to the **2.5%** average).

OWASP Top 10 Machine Learning Risks

OWASP
https://owasp.org/

01

Input Manipulation

Adversarial attacks, in which threat actors intentionally modify input data to deceive the model.

02

Data Poisoning

Manipulation of the training data to induce the model to act in an unintended and undesirable way.

03

Model Inversion Attack

Attackers reverse-engineer the model to obtain information from it.

04

Membership Inference Attack

Manipulation of the model's training data to induce behavior that reveals sensitive information.

05

Model Theft

Unauthorized, malicious users access the model's parameters.

06

AI Supply Chain Attacks

Alteration or substitution of a machine learning library or model employed by a system.

07

Transfer Learning Attack

Training a model on a particular task before fine-tuning it on another task to induce it to behave undesirably.

08

Model Skewing

Altering the distribution of training data to induce the model to behave undesirably.

09

Output Integrity Attack

Altering the output of a model to induce unintended or harmful behavior directed at the system using it.

10

Model Poisoning

Manipulation of the model's parameters to induce undesirable behavior.

Key findings

56%

of organizations have adopted AI services for custom applications

Many organizations are using AI models to create custom solutions and integrations specific to their environment(s).

27%

of organizations have not configured Azure OpenAI accounts with private endpoints.

This increases the risk that attackers can access, intercept, or manipulate data transmitted between cloud resources and AI services.

77%

of organizations using Amazon SageMaker have not configured metadata session authentication (IMDSv2) for their notebook instances

Not having IMDSv2 enabled leaves notebook instances and their sensitive data potentially exposed to high-risk vulnerabilities.

20%

of organizations using OpenAI have at least one access key saved in an unsecure location

A single leaked key can lead to a breach and risk the integrity of the OpenAI account.

45%

of Amazon SageMaker buckets are using the default bucket naming convention

Even though AWS fixed the default naming structure, adding randomized characters to the default bucket name, nearly half of organizations are still using the easily discoverable non randomized default name.

98%

of organizations using Amazon SageMaker have a notebook instance with root access enabled

The default setting enables local root access, allowing a potential attacker to access sensitive information, exfiltrate data, poison data, and more.

62%

of organizations have deployed an AI package with at least one CVE

AI packages enable developers to create, train, and deploy AI models without developing brand new routines. These packages are often susceptible to known vulnerabilities.

98%

of organizations using Google Vertex AI have not enabled encryption at rest for their self-managed encryption keys

This leaves sensitive data exposed to attackers, increasing the chances that a bad actor can exfiltrate, delete, or alter the AI model.

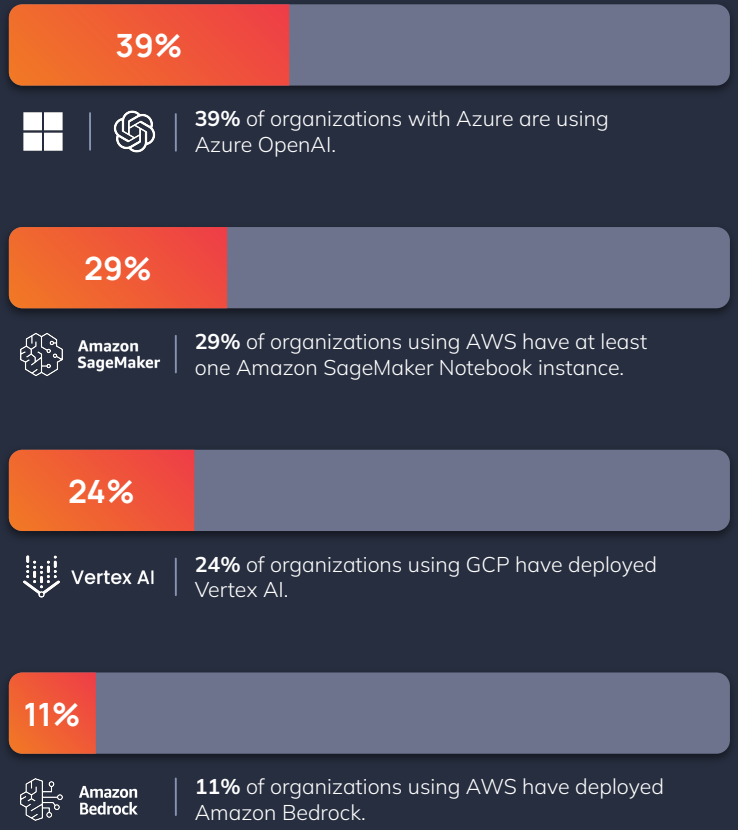
01 AI usage

1. AI usage

1.1 General AI usage

Our research indicates that more than half of organizations have adopted AI models for custom applications (company-specific AI solutions addressing a specific use case or set of use cases). This represents a significant percentage, meaning that many companies are not only using and testing AI models, but also creating their own custom solutions and integrations specific to their environment.

56% of organizations have adopted AI models for custom applications.



1.2 Usage by AI service

Among organizations in our study, Azure OpenAI is the most frequently used AI service.

The figures illustrate significant AI adoption among organizations, considering that Google Vertex AI became generally available in May 2021, Azure OpenAI in November 2021, and Amazon Bedrock in September 2023. We expect these figures to grow significantly over time, as more organizations leverage these AI services.

May 2021
Google Vertex AI becomes generally available

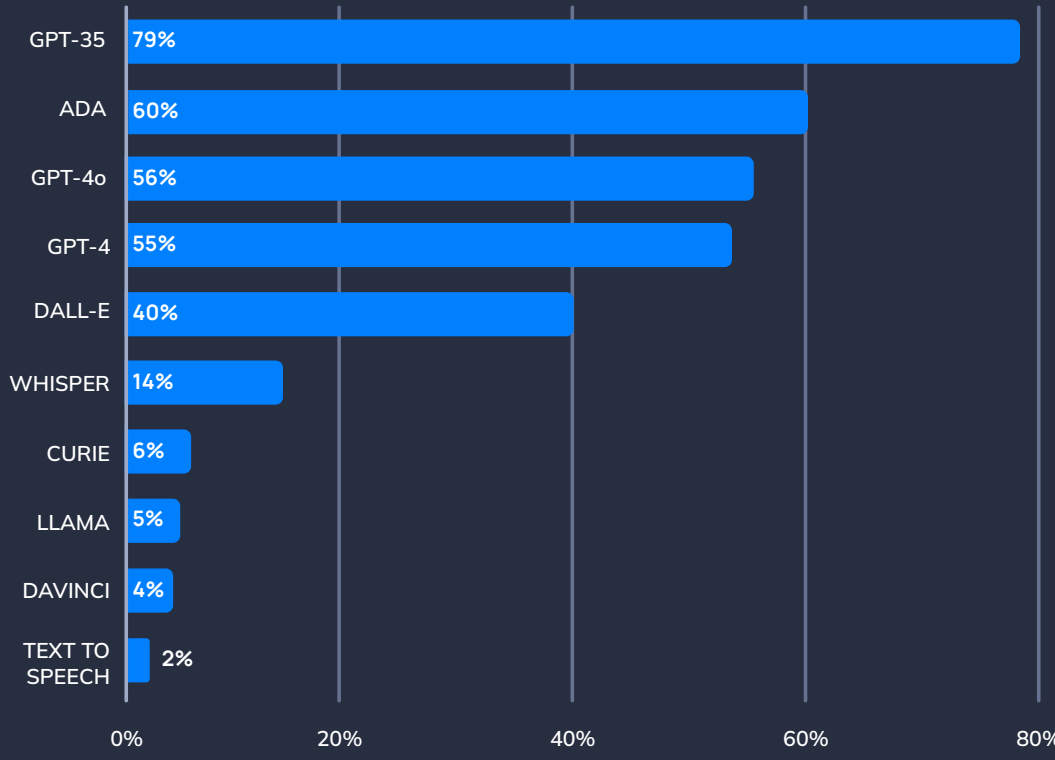
Nov 2021
Azure OpenAI becomes generally available

Sept 2023
Amazon Bedrock becomes generally available

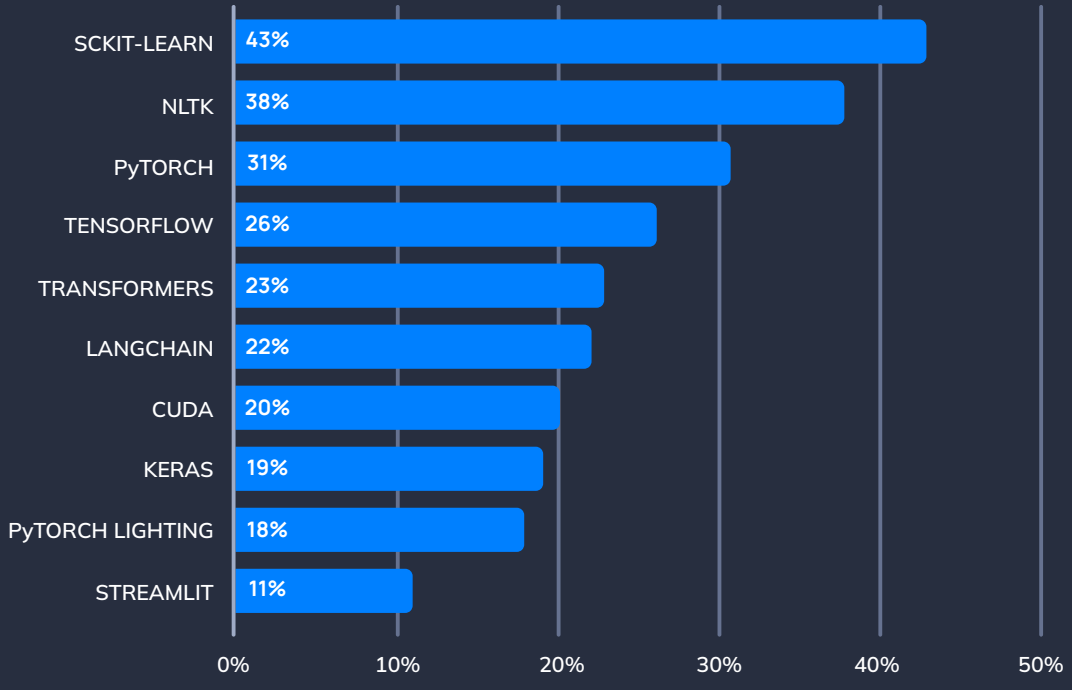
1.3 Usage by AI model

AI models are systems or algorithms trained to perform specific tasks according to learned patterns and relationships. Organizations can choose which AI model(s) to integrate with the AI services referenced previously.

Top 10 most popular AI models



Top 10 packages for building custom AI models



These figures clearly indicate that most organizations are not only consuming AI services from cloud providers, but also actively developing their own AI solutions.

1.4 Usage by AI package

AI packages are designed to automate the development of AI and ML applications. These packages contain a collection of pre-developed modules and tools that enable developers to create, train, and deploy AI models without developing brand new routines.

02

Vulnerabilities

2. Vulnerabilities in AI applications

A significant share of deployed AI packages contain at least one CVE. While alarming, most of them present low to medium risk, with an average CVSSv3 score of **6.9**. Additionally, **0.2%** of these vulnerabilities have a public exploit, far less than the general average of cloud assets (**2.5%**).

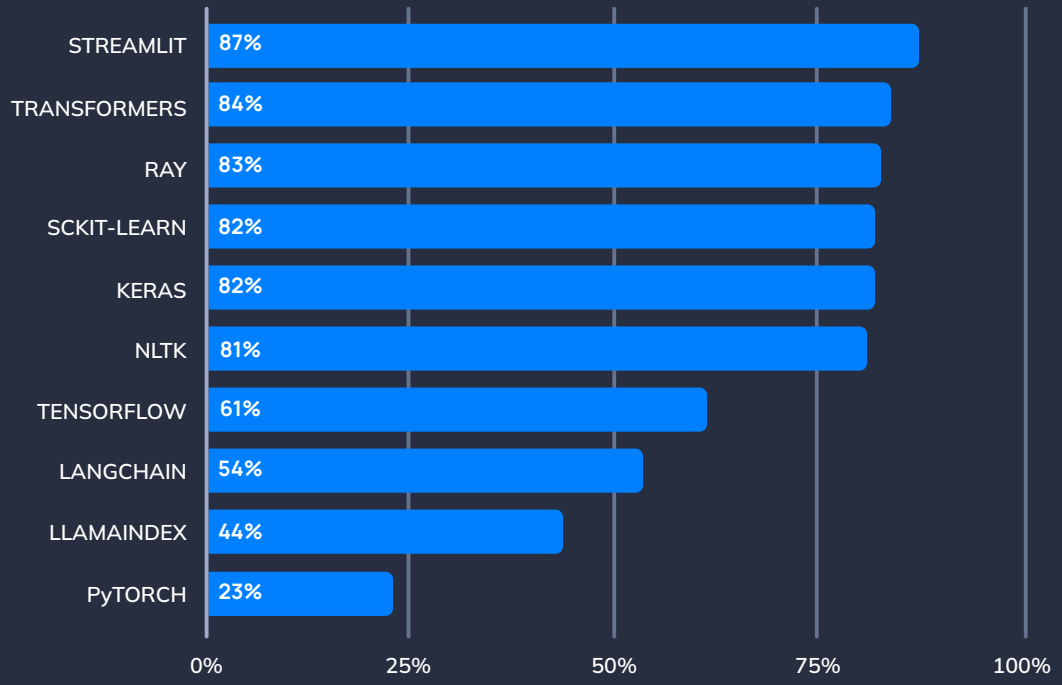
This may explain why most of these packages remain unpatched: security teams are prioritizing more critical risks. Also, upgrading some AI packages can be complicated, especially when they have dependencies (e.g., Numpy and Pytorch), which make it difficult to understand which version is compatible.

OWASP Risks:
data poisoning attack, model inversion attack, membership inference attack, model theft, AI supply chain attacks, transfer learning attacks, model skewing, output integrity attack, and model poisoning

62% of organizations have deployed an AI package with at least one CVE.

Deployed AI packages with at least one CVE

Percent of organizations with deployed package in at least one of their virtual machines



Our analysis illustrates that vulnerabilities vary by package. All CVEs associated with NLTK maintain a CVSS score deemed medium risk. Transformers packages contain four CVEs and most organizations have yet to upgrade those packages. Streamlit and Keras both contain at least one recently reported CVE, which most organizations have yet to patch.

The significant percentage of affected packages suggests that these applications are generally susceptible to known vulnerabilities.

It's important to note that even low- or medium-risk CVEs can constitute a critical risk if they are part of a high-severity attack path—a collection of interconnected risks that attackers can exploit to endanger high-value assets.

03

Exposed AI models

3. Exposed AI models

3.1 Introduction

AI model code that is exposed to the Internet gives attackers the opportunity to target and exploit that code. Exposed code presents risks such as code and model theft, the installation of a cryptominer, and more. This explains why attackers continuously search for exposed AI model assets.

For example, the Qubitstrike campaign targeted publicly exposed Jupyter notebooks to implement cryptominers. The attack aimed to breach the notebook instances, gain access to cloud environments, and illicitly mine cryptocurrency. To prevent exposure of AI model code, organizations must configure AI assets to limit network and Internet access.

> OWASP Risks:
> model inversion attack, model theft, and model poisoning.

3.2 Default Amazon SageMaker bucket names

Aqua Security recently discovered that when a user creates an Amazon SageMaker Canvas, the service automatically creates an S3 bucket to store files utilized by the service, using the following default naming convention: `sagemaker-{Region}-{Account-ID}`.

The main risk here is that an attacker can create the bucket before the victim, make it public, and have the victim use this bucket for their SageMaker needs, risking data leakage and data manipulation. Additionally, if your bucket is public, an attacker just needs to know, or guess, an AWS account ID and region to access and manipulate the contents of the bucket.

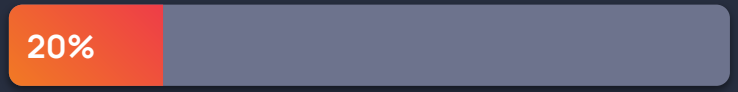
Aqua notified AWS back in February, who promptly changed the default naming to add a randomized number. However, if buckets had already been created, users would have to manually rename them. We found that nearly half (45%) of all SageMaker buckets are still using the default bucket name.

> Make sure to rename your Amazon SageMaker buckets if they are in the format sagemaker-{Region}-{Account-ID}.

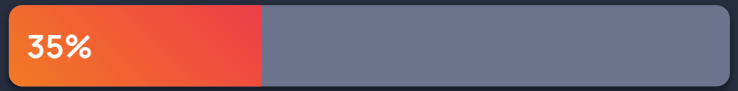
45% of Amazon SageMaker buckets are using the non randomized default bucket name

04

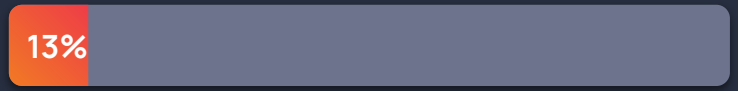
Insecure access



20% of organizations using OpenAI have an exposed OpenAI access key.



35% of organizations using Hugging Face have an exposed Hugging Face access key.



13% of organizations using Anthropic have an exposed Anthropic access key.

Before introducing project-level API keys (“Project keys”) earlier this year, OpenAI used API access keys (“User keys”) that provided admin permissions to all actions. These keys, which are now labeled “Legacy keys” in the OpenAI user interface, allowed full permissions for the secret key by default. OpenAI still allows developers to create Legacy keys, which if found, attackers can exploit to facilitate data theft, account theft, and more.

4. Insecure access

4.1 Exposed access keys

API keys provide access to AI services in code repositories. This enables developers to access AI models through the API and build applications that can generate code, images, and text. Because some of the repositories are public, unencrypted access keys can allow unwanted access to the model and its code—increasing the chances of a significant security incident.

OWASP risks:
model inversion attack, model theft, model poisoning, input manipulation attack, and output integrity attack

4.2 Exposed keys in commit history

The security risk of exposed keys also applies to a repository's commit history—not just its master branch. The Orca 2023 Honeypotting in the Cloud Report confirmed that attackers frequently scan for secrets in old commits and retrieve them.

Keep your API keys safe by following best practices, such as securely storing them, rotating keys regularly, deleting unused keys, avoiding hard coding keys, and using a secrets manager to manage their usage.

Note that attackers search the Git History for keys so it's also important to remove them from the commit history.

81% of OpenAI exposed access keys are stored in a repository's commit history.

77% of Hugging Face exposed access keys are stored in a repository's commit history.

4% of organizations using Amazon SageMaker have a notebook instance assigned with an administrative privileges IAM role.

4.3 Roles and permissions

Our findings reveal that a relatively low number of organizations are assigning over-privileged roles to Amazon Sagemaker, and instead are applying the principle of least privilege (PoLP) to their notebook instances, which is good to see. This helps prevent attackers from exploiting notebook instances to move laterally, access sensitive information, exfiltrate data, and more.

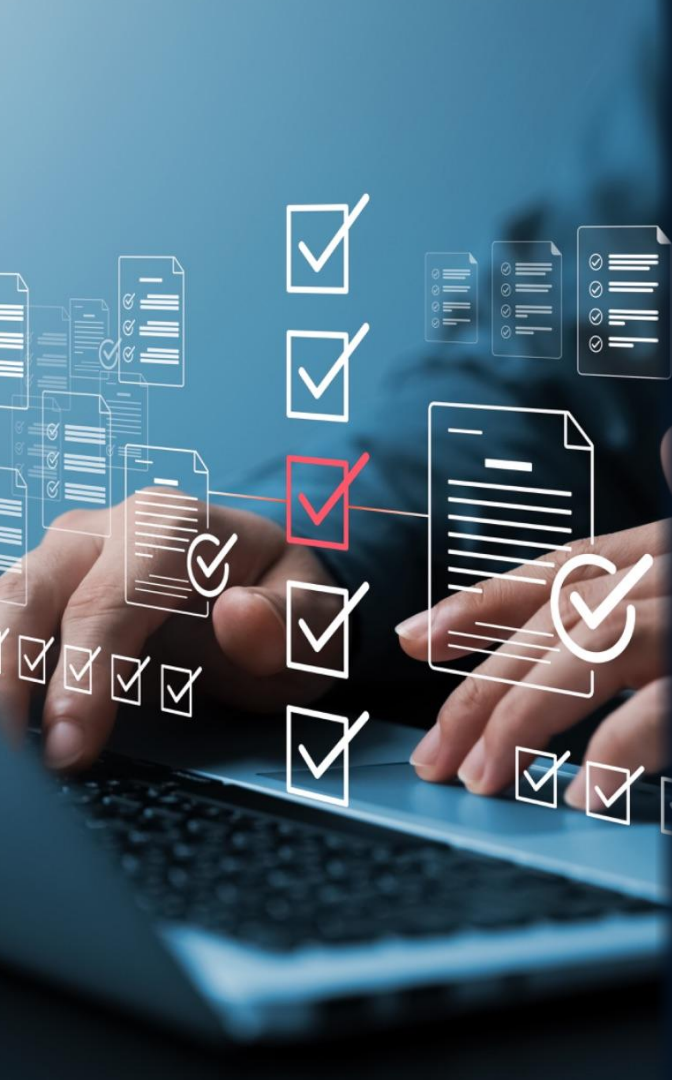
Because the field of AI is relatively new—especially its broad use through AI services and chatbots—numerous vulnerabilities and loopholes are continually being discovered, some of which may grant access to the resources these services utilize.

That's why it's especially important to properly assign roles and apply the PoLP when setting the permissions for AI users, resources, and services.

> To ensure cloud environments remain as secure as possible, ensure you manually configure roles for your users and AI services and embrace the PoLP.

05

Misconfigurations



5. Misconfigurations

Misconfigurations are a top security concern and leading cause of breaches in cloud environments. This especially applies to AI. The nascency of AI services, together with the general tendency for AI technologies to favor wider access and privileges, make AI misconfigurations particularly important for organizations to address.

This may explain why all risks in the OWASP Machine Learning Security Top Ten list apply to misconfigurations.

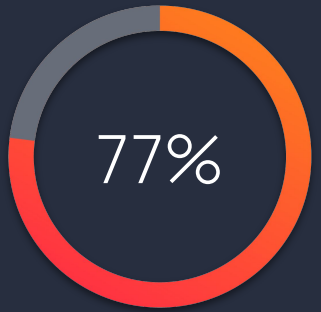
OWASP risks (all):
Input manipulation attack, data poisoning attack, model inversion attack, membership inference attack, model theft, AI supply chain attacks, transfer learning attack, model skewing, output integrity attack, and model poisoning.

5.1 Session-Based Metadata Access (IMDSv2)

Minimum Instance Metadata Service (IMDS) is an AWS security measure for cloud assets and helps organizations build secure and reliable AI applications. The feature provides access to temporary and frequently rotated credentials, eliminating the need to hardcode secrets to instances. Unlike the original version of IMDS, the latest version (IMDSv2) leverages session-based authentication that uses secret tokens and offers expanded protection against vulnerabilities. To use IMDSv2, organizations must manually disable v1 and enable v2.

In our study, most organizations using Amazon SageMaker (**77%**) have yet to configure IMDSv2 for their notebook instances. This leaves notebook instances and their sensitive data potentially exposed to high-risk vulnerabilities.

Still, the problem isn't limited to SageMaker. Analyzing general EC2 instances on AWS, we found that **95%** of organizations have yet to configure IMDSv2 for at least one EC2 instance. We also found that **35%** of all AWS EC2 instances are not configured with IMDSv2.



77% of organizations using Amazon SageMaker have yet to configure IMDSv2 for their notebook instances.

98% of organizations have not disabled root access for Amazon SageMaker notebook instances.

5.2 Root access

According to our findings, nearly all organizations using Amazon SageMaker have yet to disable root access for their notebook instances. This gives attackers free reign and full control over SageMaker's Jupyter notebooks and the AI models and services that run in them.

Over-privileged users give attackers a convenient platform to perform any action on the asset. Assuming they gain unauthorized access to an asset, excessive privileges can give them the freedom and ability to escalate privileges and carry out more malicious actions, such as data exfiltration and poisoning.

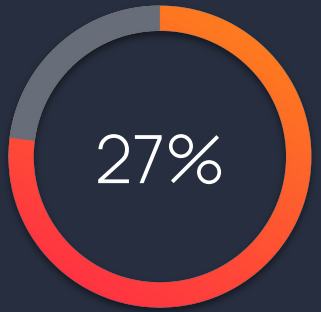
Unfortunately, AI services like Amazon SageMaker assign root privileges to assets by default - the highest level of permissions in a cloud environment.

5.3 Private endpoints

Approximately one out of every four organizations using Azure OpenAI have not configured their accounts with private endpoints. This increases the risk that attackers can access, intercept, or manipulate data transmitted between cloud resources and AI services.

Endpoints determine the connection between cloud resources and AI services. As the name suggests, private endpoints ensure that organizations can protect data transmission from exposure to the Internet. These endpoints establish a secure and efficient channel for transmitting data between infrastructure and the AI service, thereby reducing the potential risks associated with conventional public endpoints, which use public IP addresses.

> To ensure AI services remain as secure as possible, ensure you manually configure private endpoints to take advantage of this important security feature.

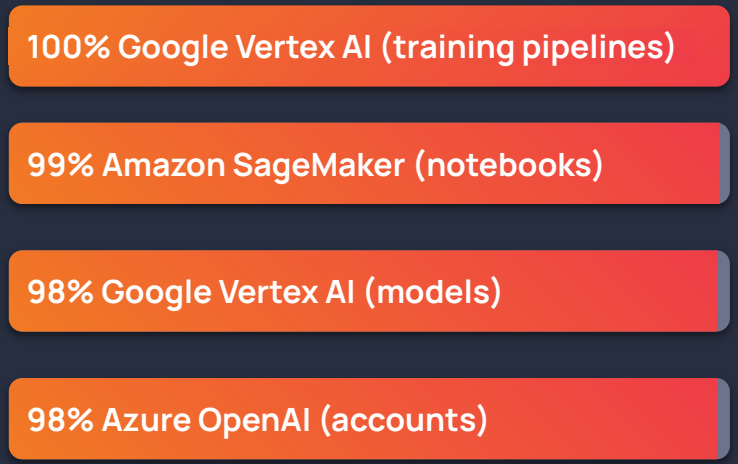


27% of organizations have not configured Azure OpenAI accounts with private endpoints.

06

Encryption

No encryption at rest
with self-managed keys:



6. Encryption

Nearly all organizations have yet to configure encryption at rest for their self-managed keys. This applies to all cloud providers (AWS and its Key Management Service (KMS) keys; Google and its customer-managed encryption keys (CMEK); and Azure and its customer-managed keys (CMK)) and across multiple locations where AI data is housed.

While our analysis didn't confirm whether organizational data was encrypted via other methods, choosing not to encrypt with self-managed keys raises the potential attackers can exploit exposed data.

OWASP risks (all):
data poisoning attack, membership inference attack, model theft, model skewing, and model poisoning.

Encrypting data at rest is an important security measure that enables organizations to enhance the security and integrity of their AI model code and training data. This includes the data stored in training pipelines (used to build custom AI models), the AI models themselves, as well as compute instances (which act as AI development environments).

For AI services, CSPs offer two options for encrypting data:

CSP-managed: Requires organizations to use encryption keys controlled and managed by the CSP

Self-managed encryption: Gives organizations complete control over their keys for enhanced security and flexibility, but it also calls for greater awareness and vigilance.

When self-managing keys, organizations must configure the encryption of data at rest.

> When choosing to self-manage encryption keys, ensure you actively enable encryption at rest across your AI pipelines, models, assets, and services. Best practice is to make this a standard practice and raise awareness through ongoing training and communication.

07

Conclusion

Challenges in AI security

01

Pace of innovation:

The speed of AI development continues to accelerate, with AI innovations introducing features that promote ease of use over security considerations. Maintaining pace with these advancements is challenging, requiring ongoing research, development, and cutting-edge security protocols.

02

Shadow AI:

Security teams lack complete visibility into AI activity. These blind spots prevent the enforcement of best practices and security policies, which in turn increases an organization's attack surface and risk profile.

03

Nascent technology:

Due to its nascent stage, AI security lacks comprehensive resources and seasoned experts. Organizations must often develop their own solutions to protect AI services without external guidance or examples.

04

Regulatory compliance:

Navigating evolving compliance requirements requires a delicate balance between fostering innovation, ensuring security, and adhering to emerging legal standards. Businesses and policymakers must be agile and adapt to new regulations governing AI technologies.

05

Resource control:

Resource misconfigurations often accompany the rollout of a new service. Users overlook properly configuring settings related to roles, buckets, users, and other assets, which introduce significant risks to the environment.

Key recommendations

The following 6 best practices can help you strengthen your AI security posture and minimize risks:

#1 Beware of default settings

Cloud provider AI services cater to the needs of developers, providing them with features and settings that enhance efficiency. This often translates into default settings that can produce security risks in a live environment. To limit risk, ensure you change these default settings during the early stages of development.

#2 Manage vulnerabilities

While the field of AI security is relatively new, most vulnerabilities are not. Often, AI services rely on existing solutions with known vulnerabilities. Detecting and mapping those vulnerabilities in your environments is still essential to managing and remediating them appropriately.

#3 Isolate networks

It's best practice to always limit network access to your assets. This means opening assets to network activity only when necessary, and precisely defining what type of network to allow in and out. This is especially relevant to AI services, since they are relatively new and untested, and possess significant capabilities.

#4 Limit privileges

Excessive privileges give attackers freedom of movement and a platform to launch multi-phased attacks—should bad actors successfully gain initial access. To protect against lateral movement and other threats, eliminate redundant privileges and remove unnecessary access between services, roles, and instances.

#5 Secure data

Securing data calls for combining several best practices. This includes opting for self-managed encryption keys while also ensuring that you enable encryption at rest. Additionally, favor more restrictive settings for data protection and offer awareness training that instructs users on data security best practices.

#6 Follow best practices

When designing and integrating AI services in your environments, always consult best practices recommended by the service provider and apply their most restrictive settings. This enables you to properly use the AI service and secure your environments.

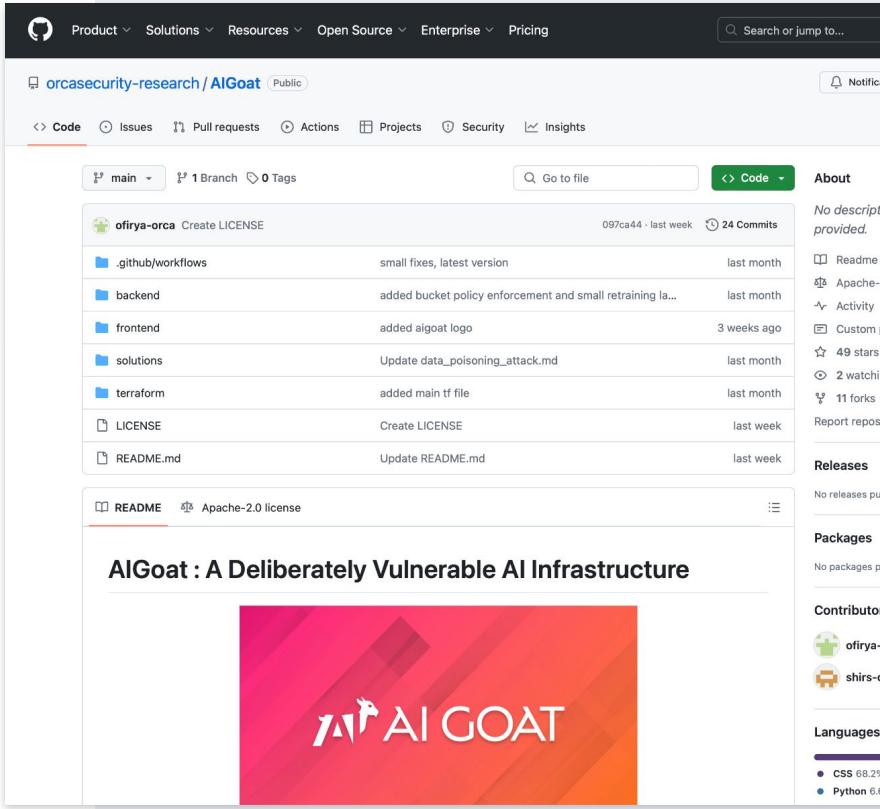
Get Hands On: AI Goat

Orca's AI Goat is the first open source AI security hands-on learning environment based on the OWASP top 10 ML risks. Provided as an open source tool on the Orca Research GitHub repository, Orca's AI Goat is an intentionally vulnerable AI environment that includes numerous threats and vulnerabilities for testing and learning purposes.

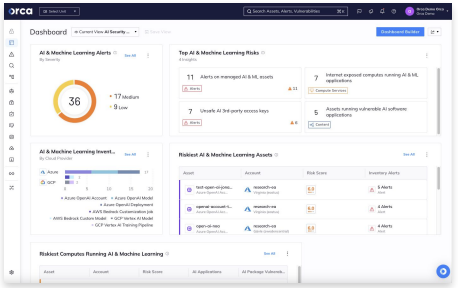
The learning environment was created to help security professionals and pentesters understand how AI-specific vulnerabilities—based on the OWASP Machine Learning Security Top Ten risks—can be exploited, and how organizations can best defend against these types of attacks.

Deploying AI Goat is straightforward and fully automated using Terraform on AWS infrastructure. This approach ensures that you can quickly set up the environment and start exploring the vulnerabilities and how they can be leveraged.

Learn more about AI Goat by visiting orca.security

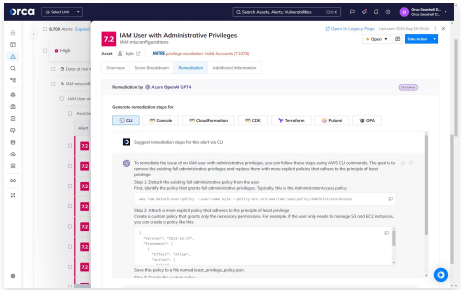


How can Orca help?



AI Security Posture Management (AI-SPM)

Orca's AI Security Posture Management (AI-SPM) solution provides unmatched visibility and deep risk analysis for AI services, models, resources, and use cases. Orca's AI-SPM solution covers 50+ AI models and software packages, allowing you to confidently adopt AI tools while fortifying security across your entire tech stack—no point solutions needed. Orca's AI-SPM provides a complete inventory of all the AI models in your environment (including any shadow AI) as well as end-to-end AI security to protect your AI models.



AI-driven cloud security

The Orca Platform itself widely leverages built-in Generative AI to simplify investigations and accelerate remediation. For example, Orca's AI-powered search supports plain language queries in more than 50 languages, eliminating specialized knowledge of cloud-provider terminology. Meanwhile, Orca's AI-powered remediation generates detailed remediation code and instructions tailored to your unique process. Orca also offers AI-powered features for IAM policy management and generating alert and asset descriptions.

About Orca Security

Orca's agentless-first Cloud Security Platform connects to your environment in minutes and provides 100% visibility of all your assets on AWS, Azure, Google Cloud, Kubernetes, and more.

Orca detects, prioritizes, and helps remediate cloud risks across every layer of your cloud estate, including vulnerabilities, malware, misconfigurations, lateral movement risk, API risks, sensitive data at risk, AI risks, and overly permissive identities.

To find out more, schedule a personalized demo of the Orca platform.

©2024 Orca Security. All rights reserved.