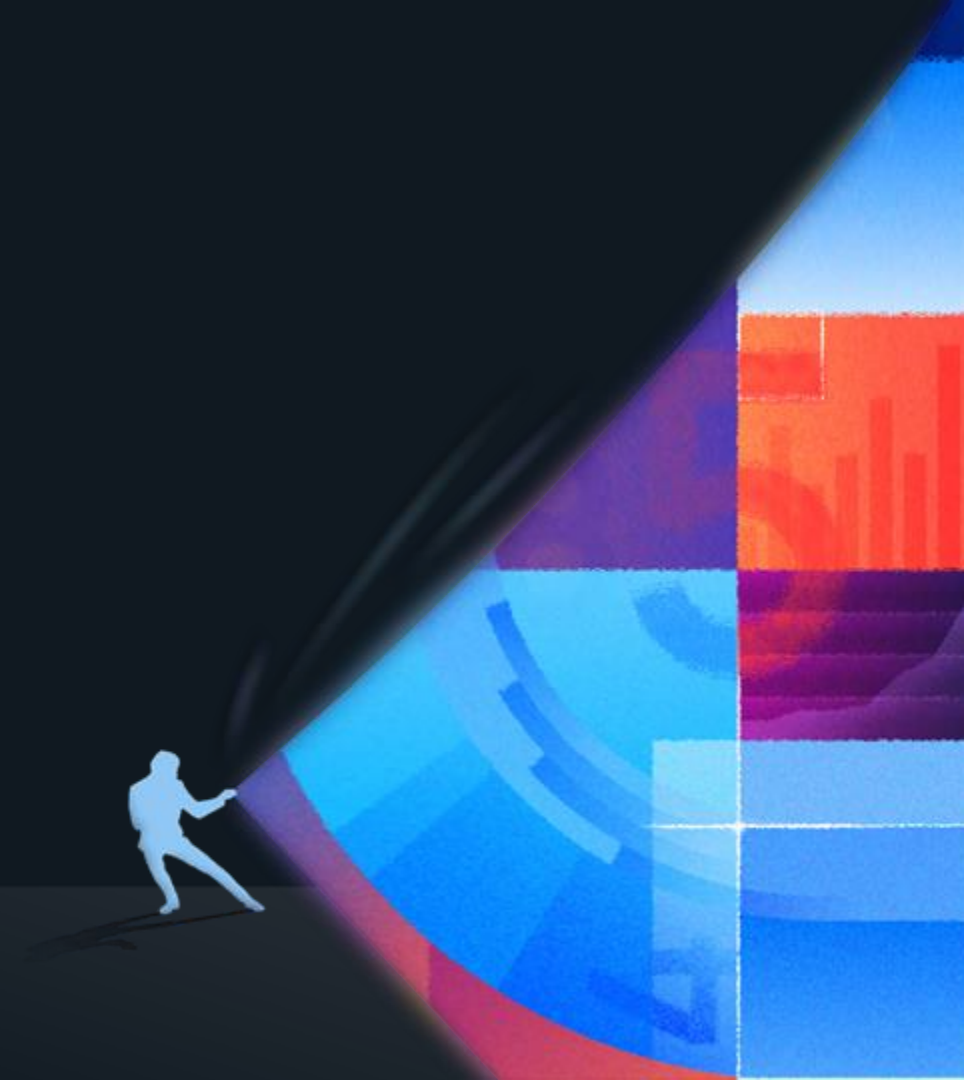
2024

State of AI Security Report

Revelando os números e os insights por trás da prevalência de riscos de IA na nuvem



Introdução

O uso da IA está em alta. O Gartner prevê que o mercado de software de IA crescerá 19,1% ao ano, atingindo US$ 298 bilhões até 2027. De muitas maneiras, a IA está agora em um estágio que lembra o da computação em nuvem há mais de uma década.

Naquela época, a velocidade da inovação era o foco, e isso ocorreu às custas da segurança. Um exemplo disso foi quando os buckets de armazenamento foram criados na velocidade da nuvem, mas estavam sendo deixados expostos à Internet - sem considerar as implicações de segurança.

Hoje, estamos testemunhando os sinais de que a história pode se repetir. Muitos serviços de IA estão adotando como padrão o acesso amplo e as permissões totais, concentrando-se na velocidade de entrega e sacrificando as medidas de segurança.

No entanto, diferentemente de uma década atrás, agora estamos mais preparados para proteger as tecnologias e os modelos emergentes de IA. A conscientização e a educação desempenham um papel fundamental para atingir esse objetivo, e é por isso que estamos lançando este relatório inaugural.

Esperamos que o relatório ajude os desenvolvedores, os CISOs e os profissionais de segurança a entender melhor como proteger seus modelos de IA, sem retardar a inovação.

Obrigado por ler nossa pesquisa.

Gil Geron
CEO e cofundador da Orca Security

Sobre o Orca Research Pod

O Orca Research Pod é um grupo de pesquisadores de segurança na nuvem que descobrem e analisam riscos e vulnerabilidades na nuvem para fortalecer a Orca Cloud Security Platform e promover as melhores práticas de segurança na nuvem.

Metodologia de pesquisa

Este relatório se concentra na segurança de modelos de IA implantados em serviços e ambientes de nuvem. Ele foi compilado por meio da análise de dados capturados de bilhões de ativos de nuvem no AWS, Azure, Google Cloud, Oracle Cloud e Alibaba Cloud verificados pela Orca Cloud Security Platform.

Conjunto de dados do relatório:

- Carga de trabalho na nuvem e dados de configuração
- Bilhões de ativos de nuvem de produção do mundo real
- Os dados mencionados neste relatório foram coletados de janeiro a agosto de 2024
- Ambientes AWS, Azure, Google Cloud, Oracle Cloud e Alibaba Cloud

Mais de 25 vulnerabilidades descobertas no AWS, Azure e 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

Este relatório aproveita insights exclusivos de monitoramentos realizados pela Orca Cloud Security Platform e revela os principais riscos e considerações de segurança de IA para CISOs, desenvolvedores e profissionais de segurança. Os riscos de segurança de IA discutidos neste relatório são mapeados para cada um dos 10 principais riscos de aprendizado de máquina da OWASP.

Sumário executivo

Nossas três principais descobertas são as seguintes:

1. **Mais da metade das organizações está implantando seus próprios modelos de IA**
   Descobrimos que **56%** das organizações adotaram a IA para criar aplicativos personalizados. O Azure OpenAI é atualmente o líder entre os serviços de IA de provedores de nuvem, com **39%** das organizações que usam o Azure. O Sckit-learn é o pacote de IA mais usado (**43%**) e o GPT-35 é o modelo de IA mais popular, com **79%** das organizações usando o GPT-35 em sua nuvem.

2. **As configurações padrão de IA geralmente são aceitas sem levar em conta a segurança**
   As configurações padrão dos serviços de IA tendem a favorecer a velocidade de desenvolvimento em vez da segurança, o que faz com que a maioria das organizações use configurações padrão inseguras. Por exemplo, **45%** dos buckets do Amazon SageMaker estão usando nomes de buckets padrão não aleatórios, e **98%** das organizações não desativaram o acesso à raiz padrão para instâncias de notebook do Amazon SageMaker.

3. **A maioria das vulnerabilidades nos modelos de IA é de risco baixo a médio - por enquanto, 62%** das organizações implantaram um pacote de IA com pelo menos um CVE. A maioria dessas vulnerabilidades é de risco baixo a médio, com uma pontuação CVSS média de **6,9,** e apenas **0,2%** das vulnerabilidades têm uma exploração pública (em comparação com a média de **2,5%**).

Os 10 principais riscos de aprendizado de máquina da OWASP

https://owasp.org/

01

Manipulação de entrada
Ataques adversários, nos quais os agentes da ameaça modificam intencionalmente os dados de entrada para enganar o modelo.

02

Envenenamento de dados
Manipulação dos dados de treinamento para induzir o modelo a agir de maneira não intencional e indesejável.

03

Ataque de inversão de modelo
Os atacantes fazem engenharia reversa do modelo para obter informações dele.

04

Ataque de inferência de associação
Manipulação dos dados de treinamento do modelo para induzir um comportamento que revele informações confidenciais.

05

Roubo de modelo
Usuários mal-intencionados e não autorizados acessam os parâmetros do modelo.

06

Ataques à cadeia de suprimentos de IA
Alteração ou substituição de uma biblioteca ou modelo de aprendizado de máquina empregado por um sistema.

07

Ataque ao aprendizado por transferência
Treinar um modelo em uma tarefa específica antes de ajustá-lo em outra tarefa para induzi-lo a se comportar de forma indesejável.

08

Inclinação do modelo
Alteração da distribuição dos dados de treinamento para induzir o modelo a se comportar de forma indesejada.

09

Ataque à integridade da saída
Alteração da saída de um modelo para induzir um comportamento não intencional ou prejudicial direcionado ao sistema que o utiliza.

10

Envenenamento do modelo
Manipulação dos parâmetros do modelo para induzir um comportamento indesejável.

Principais descobertas

56%

das organizações adotaram serviços de IA para aplicações personalizadas.

Muitas organizações estão usando modelos de IA para criar soluções e integrações específicas para seus ambientes.

27%

das organizações não configuraram contas do Azure OpenAI com endpoints privados.

Isso aumenta o risco de que atacantes possam acessar, interceptar ou manipular dados transmitidos entre recursos de nuvem e serviços de IA.

77%

das organizações que usam Amazon SageMaker não configuraram autenticação de sessão de metadados (IMDSv2) para suas instâncias de notebook.

A ausência de IMDSv2 deixa as instâncias de notebook e seus dados sensíveis potencialmente expostos a vulnerabilidades de alto risco.

20%

das organizações que usam OpenAI têm pelo menos uma chave de acesso salva em uma localização insegura.

Uma única chave vazada pode levar a uma violação e comprometer a integridade da conta OpenAI.

45%

dos buckets do Amazon SageMaker estão usando a convenção de nomenclatura de bucket padrão.

Embora a AWS tenha corrigido a estrutura de nomenclatura padrão, adicionando caracteres aleatórios ao nome padrão do bucket, quase metade das organizações ainda utiliza o nome padrão não randomizado, facilmente descoberto.

98%

das organizações que usam Amazon SageMaker têm uma instância de notebook com acesso root habilitado.

O acesso root permite que desenvolvedores criem, treinem e implantem modelos de IA sem novas rotinas de segurança. Esses pacotes são frequentemente vulneráveis a problemas de segurança conhecidos.

62%

das organizações implantaram um pacote de IA com pelo menos uma CVE.

Pacotes de IA permitem que desenvolvedores criem, treinem e implantem modelos de IA sem novas rotinas. Esses pacotes são frequentemente vulneráveis a problemas de segurança conhecidos.

98%

das organizações que usam Google Vertex AI não habilitaram a criptografia em repouso para suas chaves de criptografia gerenciadas por eles mesmos.

Isso deixa os dados sensíveis expostos a atacantes, aumentando as chances de que um agente malicioso possa exfiltrar, excluir ou alterar o modelo de IA.

Conclusão

Desafios na segurança de IA

01

Ritmo da inovação:
A velocidade do desenvolvimento de IA continua a acelerar, com inovações que promovem a facilidade de uso em detrimento das considerações de segurança. Manter o ritmo com esses avanços é desafiador, exigindo pesquisa contínua, desenvolvimento e protocolos de segurança de ponta.

02

Shadow AI (IA não controlada):
As equipes de segurança não têm visibilidade completa sobre a atividade de IA. Esses pontos cegos impedem a aplicação das melhores práticas e políticas de segurança, o que, por sua vez, aumenta a superfície de ataque e o perfil de risco da organização.

03

Tecnologia nascente:
Devido ao seu estágio inicial, a segurança de IA carece de recursos abrangentes e de especialistas experientes. As organizações frequentemente precisam desenvolver suas próprias soluções para proteger os serviços de IA sem orientação ou exemplos externos.

04

Conformidade regulatória:
Navegar pelos requisitos regulatórios em constante evolução exige um equilíbrio delicado entre fomentar a inovação, garantir a segurança e aderir aos padrões legais emergentes. Empresas e legisladores devem ser ágeis e se adaptar às novas regulamentações que regem as tecnologias de IA.

05

Controle de recursos:
Erros de configuração de recursos costumam acompanhar a implementação de um novo serviço. Os usuários negligenciam a configuração adequada de definições relacionadas a funções, buckets, usuários e outros ativos, o que introduz riscos significativos ao ambiente.

Principais recomendações

As 6 melhores práticas a seguir podem ajudá-lo a fortalecer sua postura de segurança de IA e minimizar riscos:

#1 Atenção às configurações padrão

Os serviços de IA de provedores de nuvem atendem às necessidades dos desenvolvedores, oferecendo recursos e configurações que aumentam a eficiência. Isso geralmente se traduz em configurações padrão que podem gerar riscos de segurança em um ambiente ao vivo. Para reduzir o risco, certifique-se de alterar essas configurações padrão nas primeiras etapas do desenvolvimento.

#2 Gerencie vulnerabilidades

Embora o campo da segurança de IA seja relativamente novo, a maioria das vulnerabilidades não é. Frequentemente, os serviços de IA dependem de soluções existentes com vulnerabilidades conhecidas. Detectar e mapear essas vulnerabilidades em seus ambientes ainda é essencial para gerenciá-las e corrigi-las adequadamente.

#3 Isole redes

É uma boa prática sempre limitar o acesso à rede para seus ativos. Isso significa abrir os ativos para a atividade de rede apenas quando necessário e definir com precisão o tipo de rede que deve ser permitida para entrada e saída. Isso é especialmente relevante para serviços de IA, pois são relativamente novos e não testados, e possuem capacidades significativas.

#4 Limite privilégios

Privilégios excessivos dão aos atacantes liberdade de movimento e uma plataforma para lançar ataques multifásicos, caso consigam obter acesso inicial. Para se proteger contra movimentos laterais e outras ameaças, elimine privilégios redundantes e remova acessos desnecessários entre serviços, funções e instâncias.

#5 Proteja os dados

Proteger os dados exige a combinação de várias práticas recomendadas. Isso inclui optar por chaves de criptografia gerenciadas por você mesmo e garantir que você habilite a criptografia em repouso. Além disso, prefira configurações mais restritivas para proteção de dados e ofereça treinamento de conscientização que instrua os usuários sobre as melhores práticas de segurança de dados.

#6 Siga as melhores práticas

Ao projetar e integrar serviços de IA em seus ambientes, sempre consulte as melhores práticas recomendadas pelo provedor de serviços e aplique suas configurações mais restritivas. Isso permite que você use adequadamente o serviço de IA e proteja seus ambientes.

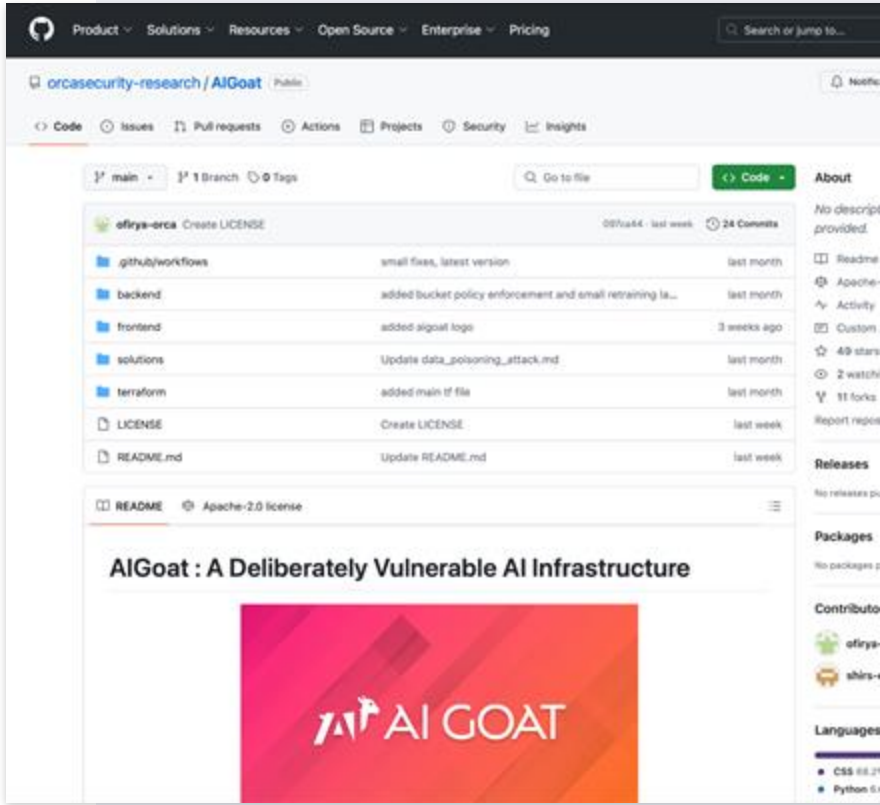
Coloque a mão na massa: AI Goat

O Orca's AI Goat é o primeiro ambiente de aprendizado prático de segurança de IA de código aberto baseado nos 10 principais riscos de ML da OWASP. Fornecido como uma ferramenta de código aberto no repositório GitHub da Orca Research, o Orca's AI Goat é um ambiente de IA intencionalmente vulnerável que inclui várias ameaças e vulnerabilidades para fins de teste e aprendizado.

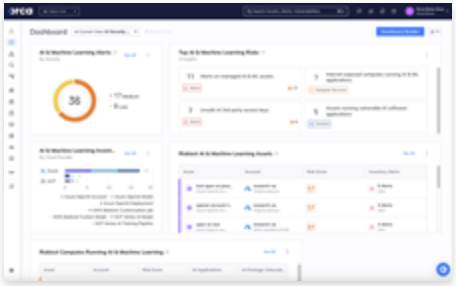
O ambiente de aprendizado foi criado para ajudar os profissionais de segurança e os pentesters a entender como as vulnerabilidades específicas de IA - baseadas nos dez principais riscos de segurança de aprendizado de máquina da OWASP - podem ser exploradas e como as organizações podem se defender melhor contra esses tipos de ataques.

A implantação do AI Goat é simples e totalmente automatizada usando o Terraform na infraestrutura da AWS. Essa abordagem garante que você possa configurar rapidamente o ambiente e começar a explorar as vulnerabilidades e como elas podem ser aproveitadas.

Saiba mais sobre a AI Goat, acesse orca.security



Como a Orca ajuda?

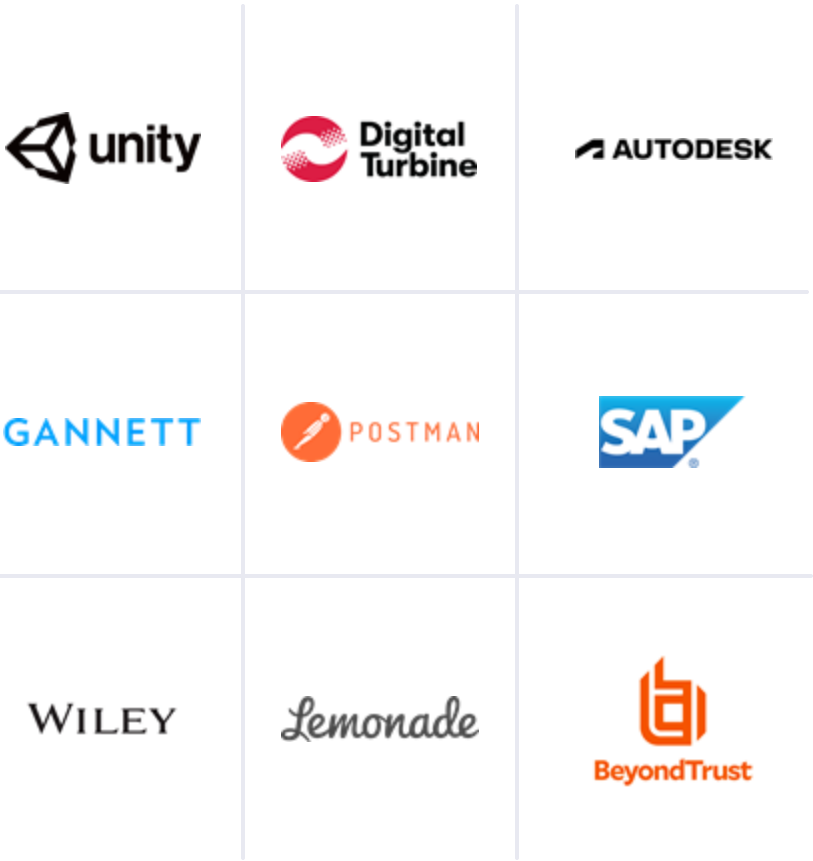


Gerenciamento de postura de segurança de IA (AI-SPM)

A solução de Gerenciamento de Postura de Segurança de IA (AI-SPM) da Orca fornece visibilidade incomparável e análise de risco profunda para serviços, modelos, recursos e casos de uso de IA. A solução AI-SPM da Orca abrange mais de 50 modelos de IA e pacotes de software, permitindo que você adote ferramentas de IA com confiança enquanto fortalece a segurança em toda a sua pilha de tecnologia - não são necessárias soluções pontuais. O AI-SPM da Orca fornece um inventário completo de todos os modelos de IA em seu ambiente (incluindo qualquer IA sombra), bem como segurança de IA de ponta a ponta para proteger seus modelos de IA.

Segurança na nuvem orientada por IA

A própria Plataforma Orca aproveita amplamente a IA geradora integrada para simplificar as investigações e acelerar a correção. Por exemplo, a pesquisa com IA do Orca suporta consultas em linguagem simples em mais de 50 idiomas, eliminando o conhecimento especializado da terminologia do provedor de nuvem. Enquanto isso, a remediação com IA do Orca gera códigos e instruções de remediação detalhados, adaptados ao seu processo exclusivo. O Orca também oferece recursos alimentados por IA para gerenciamento de políticas de IAM e geração de descrições de alertas e ativos.



Sobre a Orca Security

A plataforma de segurança em nuvem sem agente da Orca se conecta ao seu ambiente em minutos e fornece 100% de visibilidade de todos os seus ativos no AWS, Azure, Google Cloud, Kubernetes e muito mais.

A Orca detecta, prioriza e ajuda a remediar os riscos da nuvem em todas as camadas de sua propriedade de nuvem, incluindo vulnerabilidades, malware, configurações incorretas, risco de movimento lateral, riscos de API, dados confidenciais em risco, riscos de IA e identidades excessivamente permissivas.

Para saber mais, agende uma demonstração personalizada da plataforma Orca.