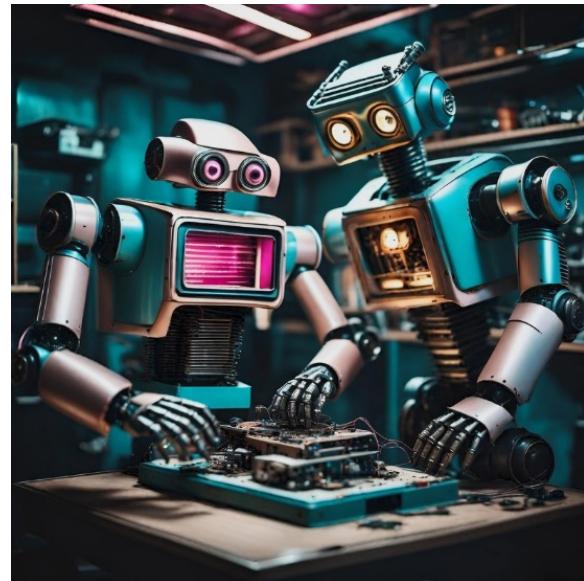


# Disrupting the Model: Abusing MLOps Platforms to Compromise ML Models and Enterprise Data Lakes

Brett Hawkins (@h4wkst3r)  
Adversary Services, IBM X-Force Red

Chris Thompson (@retBandit)  
Adversary Services, IBM X-Force Red



# Whitepaper and Tool



Whitepaper



MLOKit



# Agenda



1. Introduction
2. Background
3. MLOKit
4. Attacking MLOps Platforms
  - Azure ML, BigML, Vertex AI
    - Overview
    - Attack Scenarios
    - Demos
5. Configuration Guidance
6. Detection Guidance
7. Conclusion

# Introduction



# Who are we – Brett Hawkins

<https://h4wkst3r.github.io>



**Current Role**  
Capability Lead,  
Adversary Services  
**IBM X-Force Red**

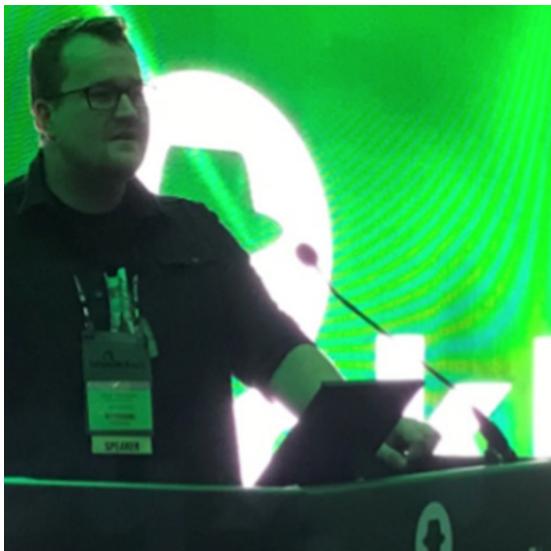


**Conference Speaker**  
Black Hat (US & EU),  
BlueHat, DerbyCon,  
Wild West Hackin'  
Fest, BSides,  
Hackers Teaching  
Hackers



**Open-Source Tool**  
**Author**  
SharPersist,  
InvisibilityCloak,  
SCMKit, ADOKit,  
MLOKIt

# Who are we – Chris Thompson



**Current Role**  
Global Head,  
**IBM X-Force Red**



## About

Chris has extensive experience performing red teaming for clients in a wide variety of industries.



**Conference Speaker**  
DEF CON, Black Hat,  
ToorCon, RSA,  
SecTor, BSides,  
SANS, and Wild  
West Hackin' Fest

# Research Drivers



Threat actors targeting MLOps platforms



Lack of research on attacking and defending MLOps platforms



Adoption of MLOps by enterprises



Lack of tooling to simulate attacks against MLOps platforms

# Research Goals



Inspire future  
MLOps research



Bring more  
attention to  
defending MLOps  
platforms



Open-source  
framework to  
simulate attacks  
and test detections  
against MLOps  
platforms

# Attendee Takeaways



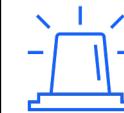
How to attack and defend MLOps platforms



Awareness of attack scenarios against MLOps platforms

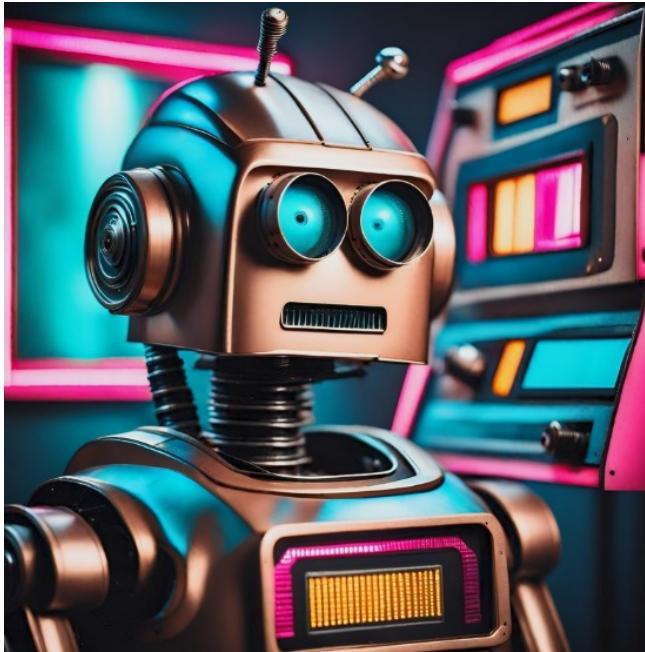


How to leverage toolkit to streamline and automate attacks against MLOps



How to build detections for attacks against MLOps platforms

# What is new in this research?



Practically perform  
ML-based attacks  
against popular  
MLOps platforms



REST API abuse of  
MLOps platforms



Dedicated toolkit  
for simulating  
attacks against  
MLOps platforms



Detection rules  
for ML-based  
attacks against  
MLOps platforms

# Our Perspective



We are

Offensive  
Cybersecurity  
Specialists

We are not

Data Scientists

AI/ML Engineers

Cloud Engineers

Detection Engineers

DevOps Engineers

Software Engineers

# Prior Work

Links to prior work are provided in whitepaper

Adrian Wood & Mary Walker – Black Hat Asia 2024

*Confused Learning: Supply Chain Attacks through Machine Learning Models*

Nitesh Surana – Black Hat USA 2023

*Uncovering Azure's Silent Threats: A Journey into Cloud Vulnerabilities*

Florian Tramer, Fan Zhang, Ari Juels, Michael K. Reiter, Thomas Ristenpart – Academic Whitepaper

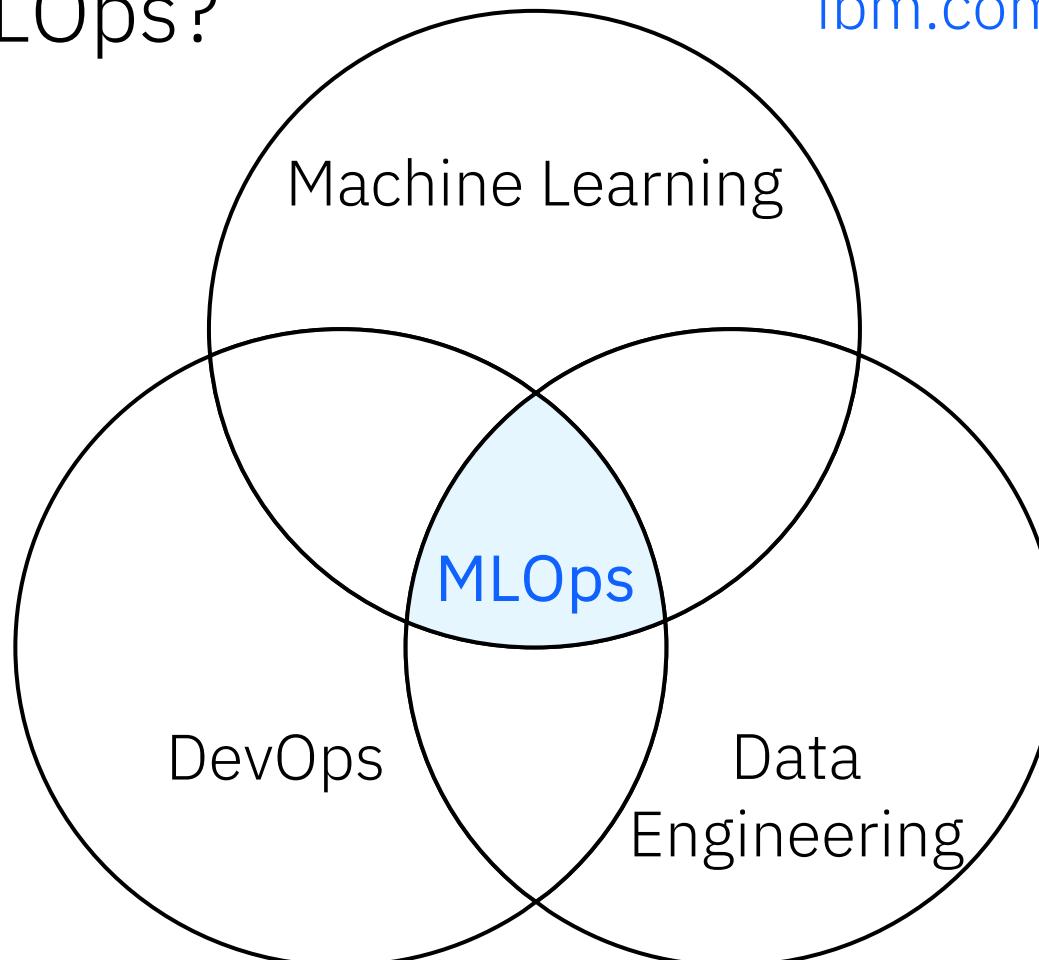
*Stealing Machine Learning Models via Prediction*

# Background



# What is MLOps?

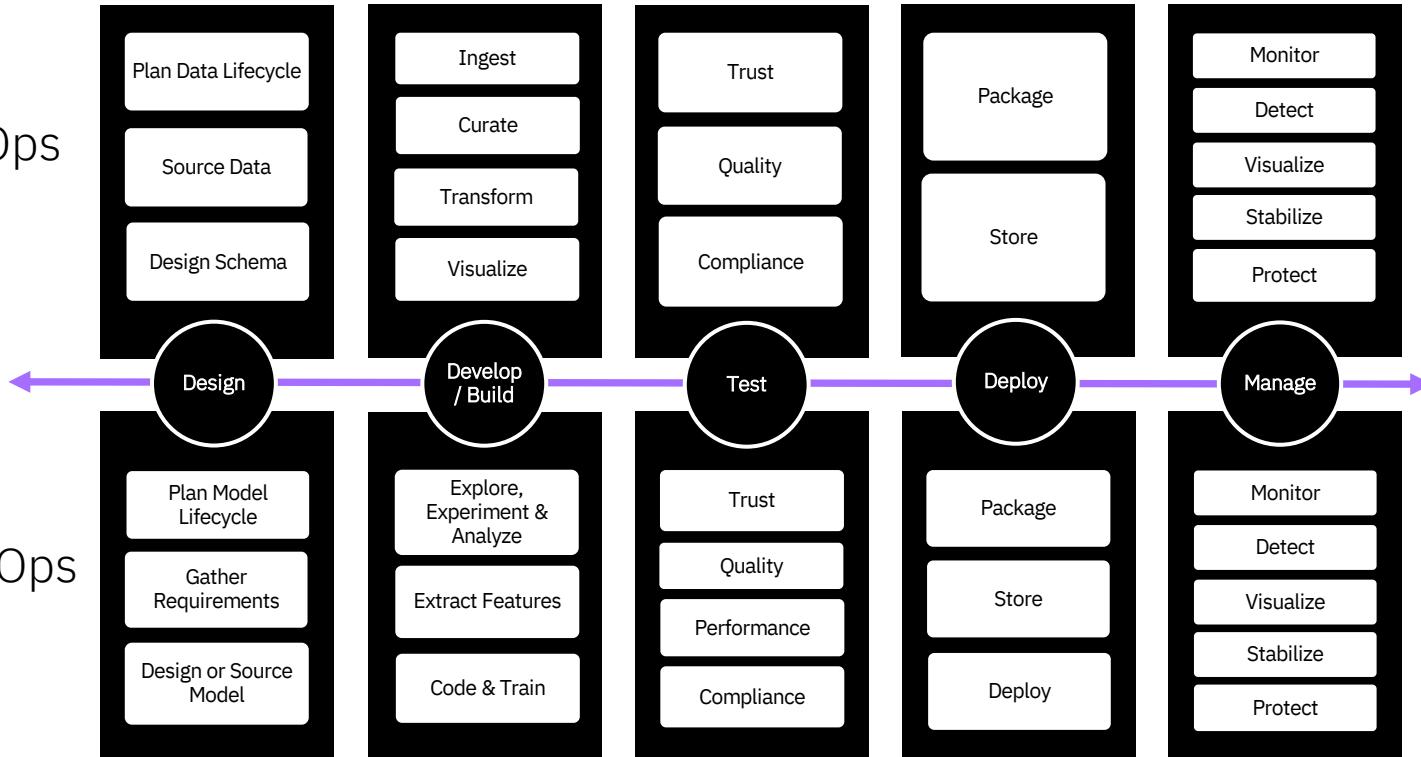
[ibm.com/topics/mlops](https://ibm.com/topics/mlops)



# MLOps Lifecycle

DataOps

ModelOps



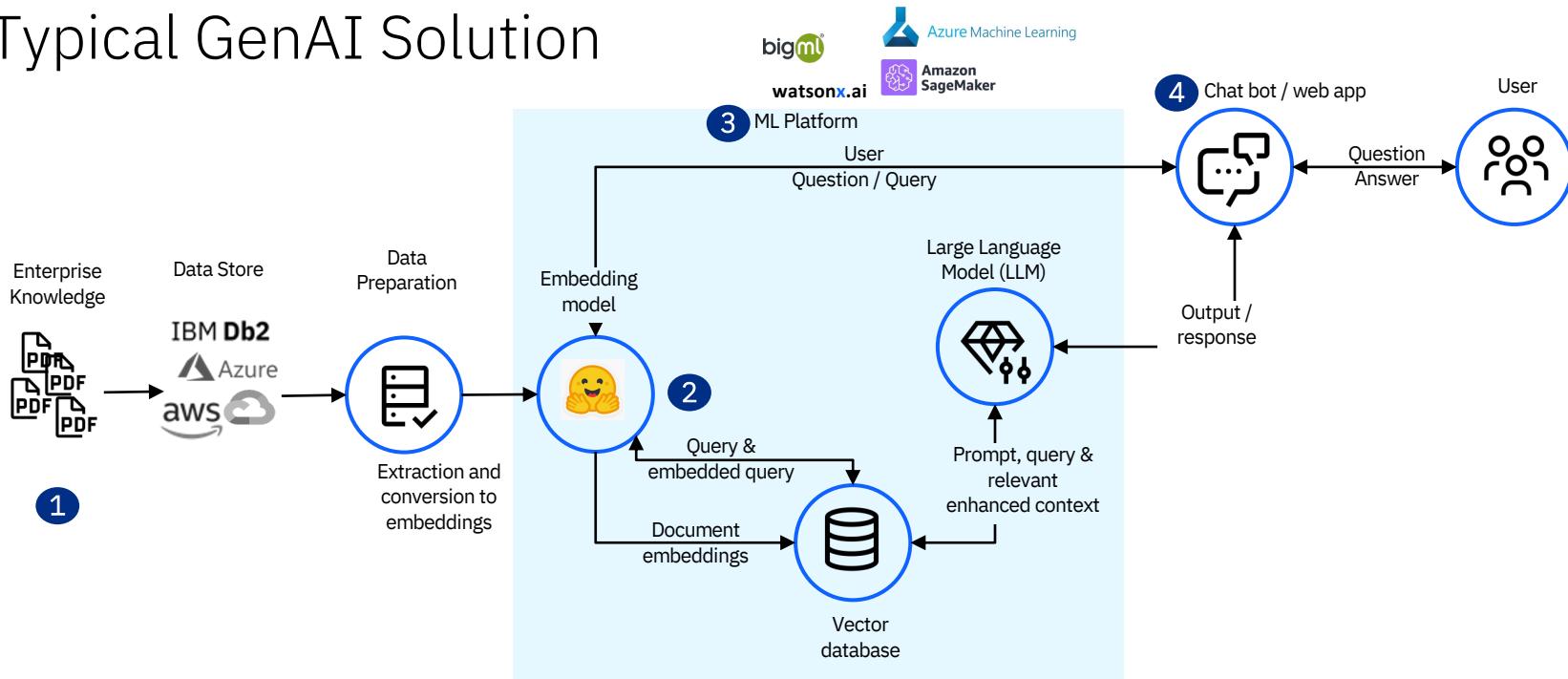
# Popular Open-Source MLOps Platforms

- Sematic
- Flyte
- Metaflow
- Seldon Core
- Kubeflow
- MLReef
- MLFlow
- TensorFlow

# Popular Commercial MLOps Platforms

- Amazon SageMaker
  - Domino Enterprise MLOps Platform
  - Azure ML
  - Databricks
  - BigML
  - DataRobot
  - Google Cloud Vertex AI
  - TrueFoundry
  - Qwak
  - Weights & Biases (W&B)
  - Palantir AIP
- Research Focus**

# Typical GenAI Solution



## 1 ML Pipeline

Testing of within the MLSecOps pipeline and production environments for attack paths from an adversary's perspective. Asses the potential impact of backdoored or compromised model or AI application environment and validate detections for attacks against datasets.

## 2 Models

Perform testing of GenAI applications, integrations, and API endpoints for security issues before production.

## 3 Platforms

Focused security testing on SaaS and PaaS platforms leveraged by GenAI applications to insecure security configurations and integrations with AI platforms such as Amazon SageMaker, Azure ML, BigML, Watsonx.ai, etc.

## 4 GenAI Apps

Perform testing of GenAI applications, integrations, and API endpoints for security issues.

# Threat Actor Motivation

The screenshot shows a web page from the Oligo platform. At the top, there's a navigation bar with links for Product, Solutions, Resources, Company, and Customers, along with a purple 'Book a Demo' button. Below the navigation is a breadcrumb trail with a back arrow. The main content area features two buttons: 'Research' and 'Shadow Vulnerability'. To the right of these buttons is a '17 min read' indicator. The main title of the article is 'ShadowRay: First Known Attack Campaign Targeting AI Workloads Actively Exploited In The Wild'. Below the title, there are three small profile pictures followed by the names 'Avi Lumelsky, Guy Kaplan, Gal Elbaz' and the date 'March 26, 2024'. To the right of the authors' names are sharing icons for LinkedIn, X (formerly Twitter), Facebook, and Email. A large, bold subtitle at the bottom of the article preview reads 'Thousands of publicly exposed Ray servers compromised as a result of Shadow Vulnerability'.

Theft of models and weights, backdooring models for initial access or persistence, expanding access via lateral movement and privsec, sensitive data theft or deploying ransomware, model modification/poisoning for misclassification, degradation, fraud or ml-based detection evasion.

# Example Goals

## Top Tier

Nation State

Top State-Sponsored Hacking Groups



- Poison trained models or model training data related to image recognition, predictive AI, early warning systems
- Modify models used for offensive tasking and targeting
- Extract sensitive data sets used to train models and their weights

## Advanced

State-Sponsored Hacking Groups

Well-Funded Ransomware Groups



- Compromise AI-based fraud analytic platforms to allow restricted actions, modify fraud algos, or ignore alert thresholds
- Steal/extract capital market trading algo and models
- Gain code execution in public AI platforms
- AI supply chain compromise (targeting high-value victims with backdoored HF models, python libraries, etc.), novel extensions
- Target MLSecOps for expanding and elevating corporate access

## Targeted

Commodity Ransomware Groups

Initial Access Brokers



- Focused on GenAI app and platform security issues to gain model code execution to compromise production or training environments
- Seed widespread malicious pytorch models on HF
- Compromise connected enterprise data lakes with financial and healthcare datasets

## Opportunistic

Hacktivists

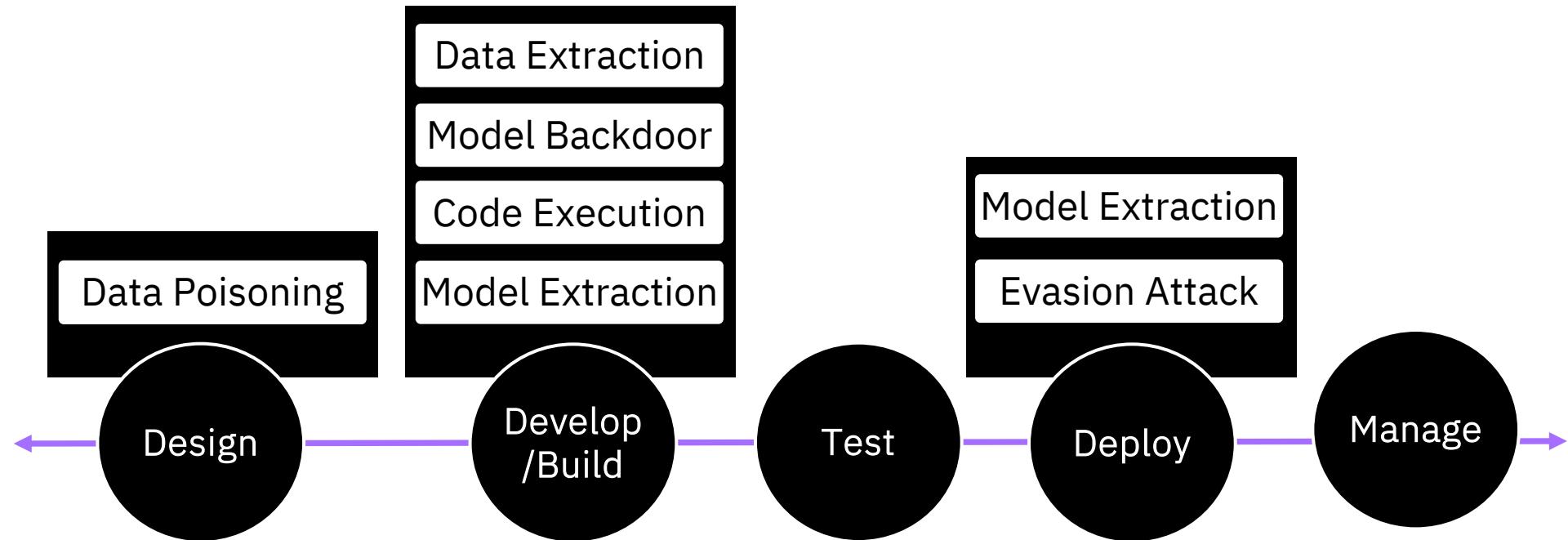
Solo Hackers

Insider Threats

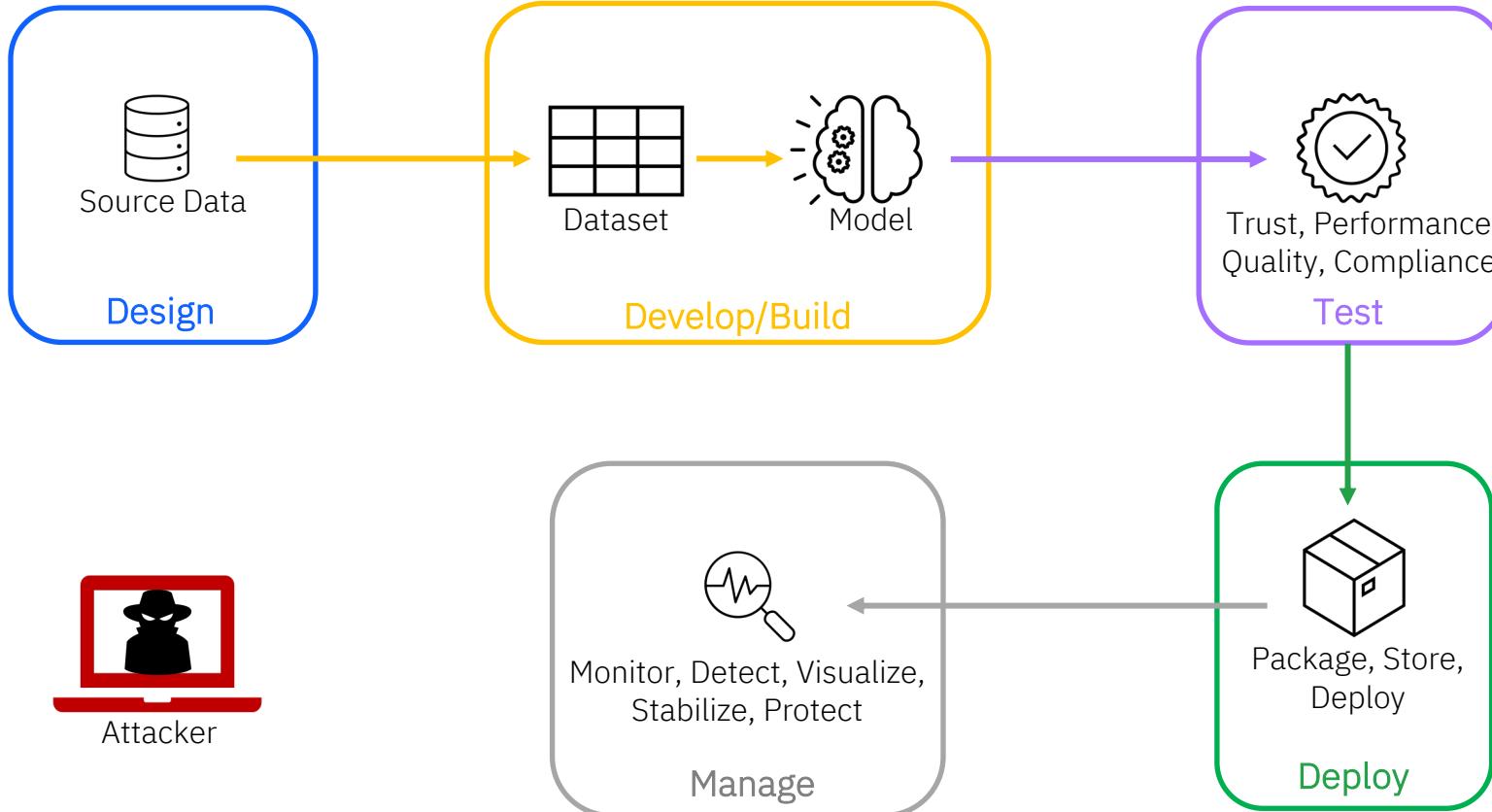


- Focused on published security attacks for code execution or service disruption.
- Focused on fraudulent use of AI such as manipulating sales bots, safety related content abuse
- Perform resource exhaustion/DOS

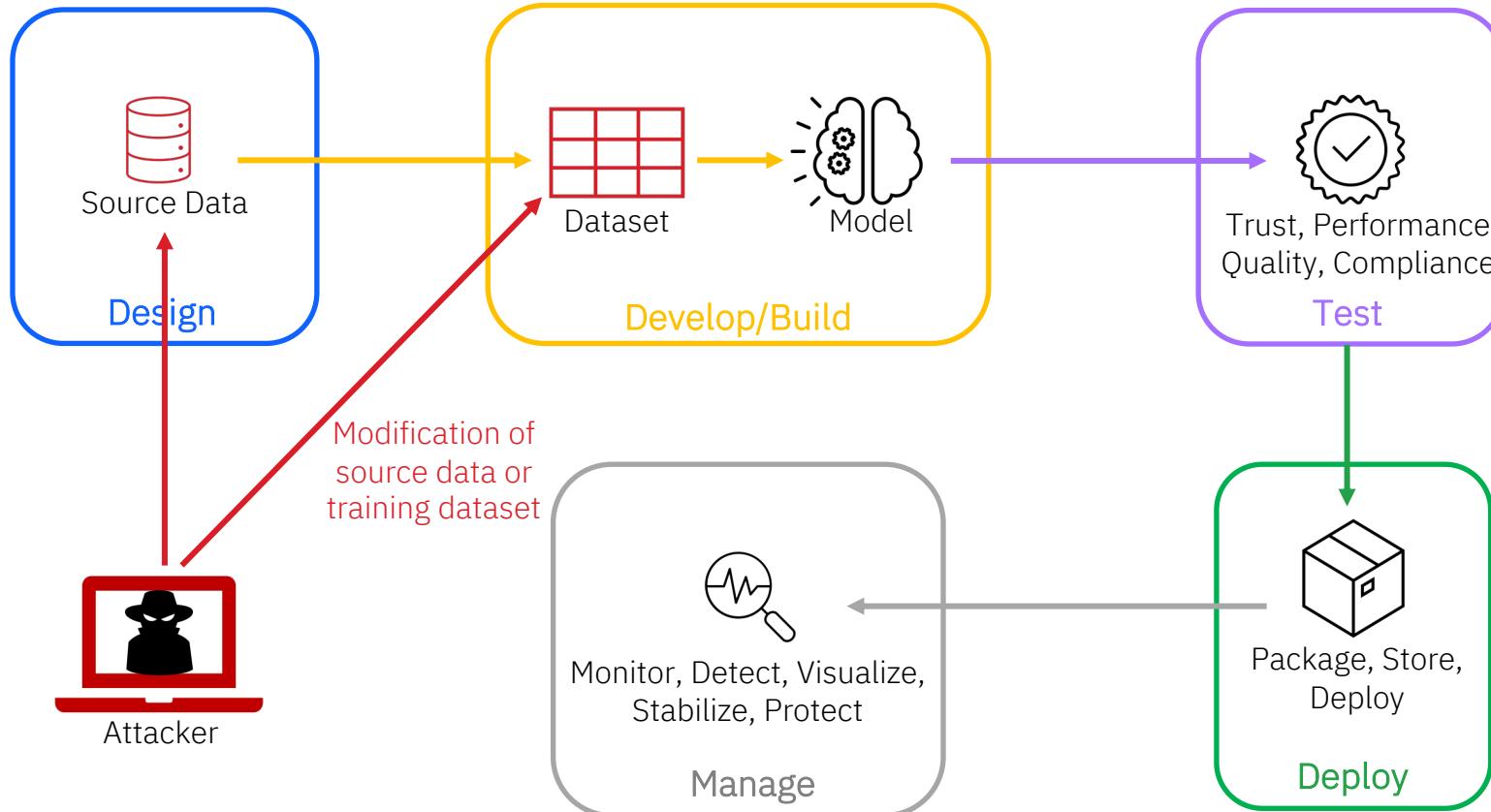
# Attack Scenarios against MLOps Lifecycle



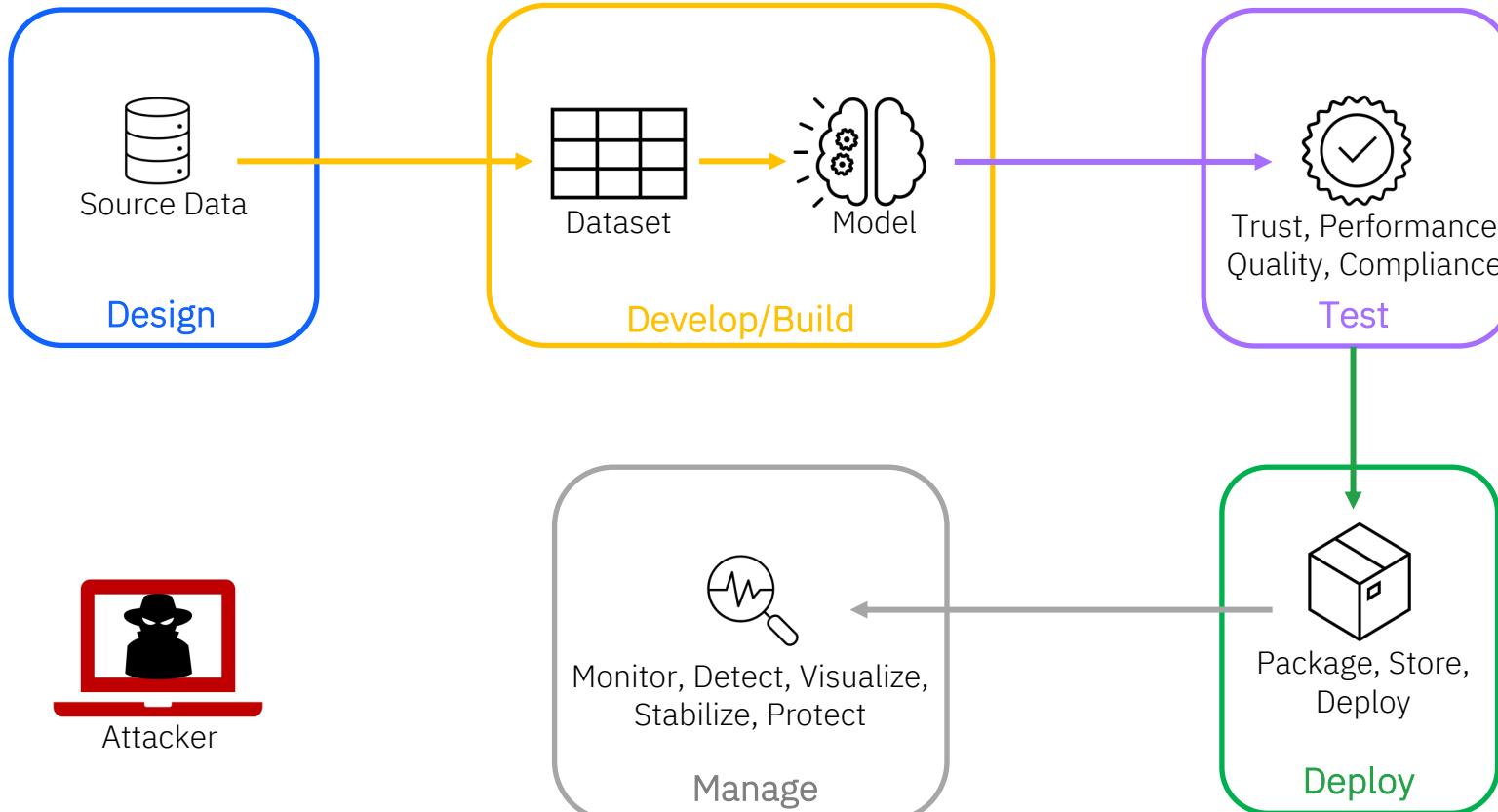
# Data Poisoning



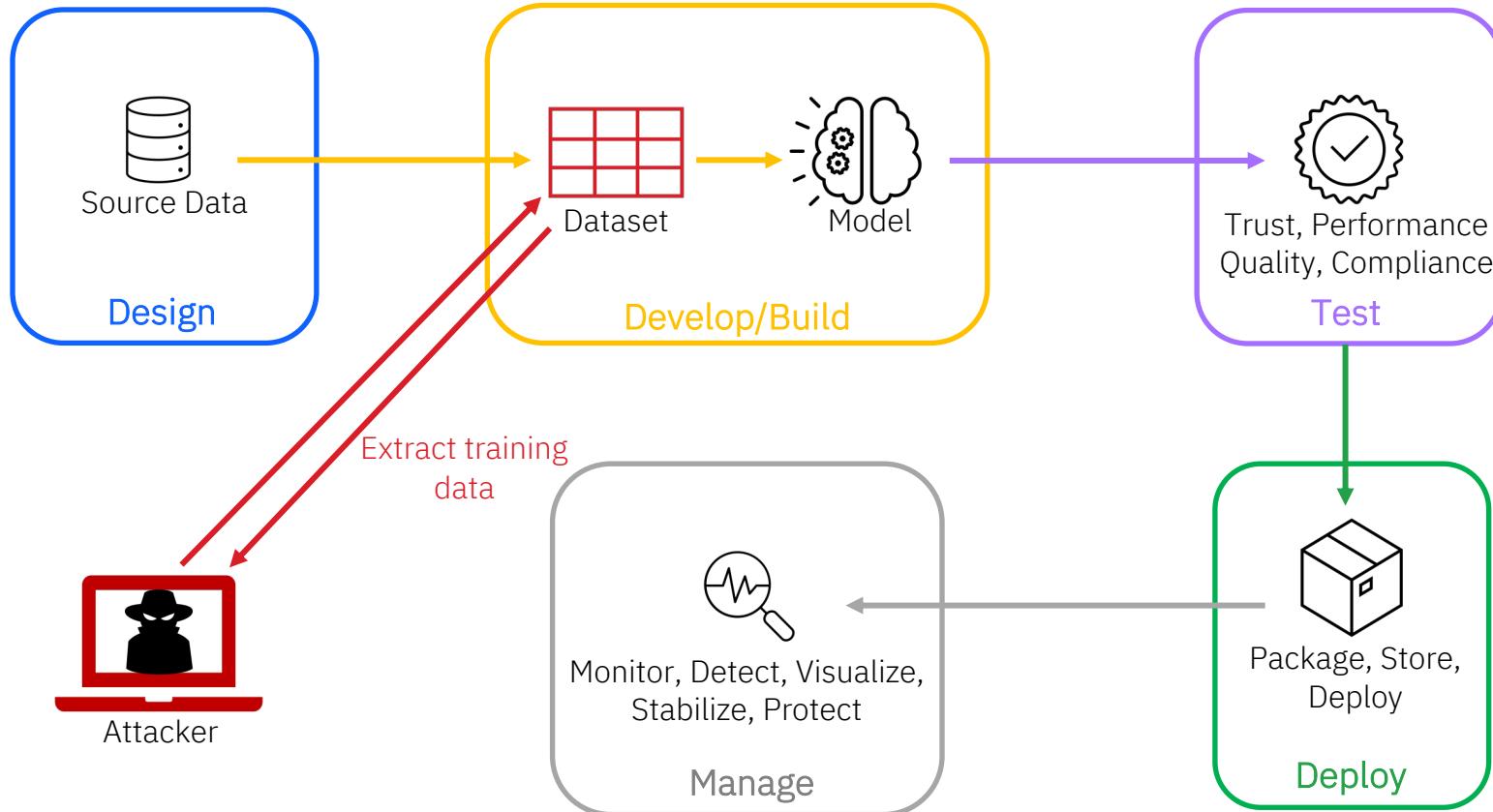
# Data Poisoning



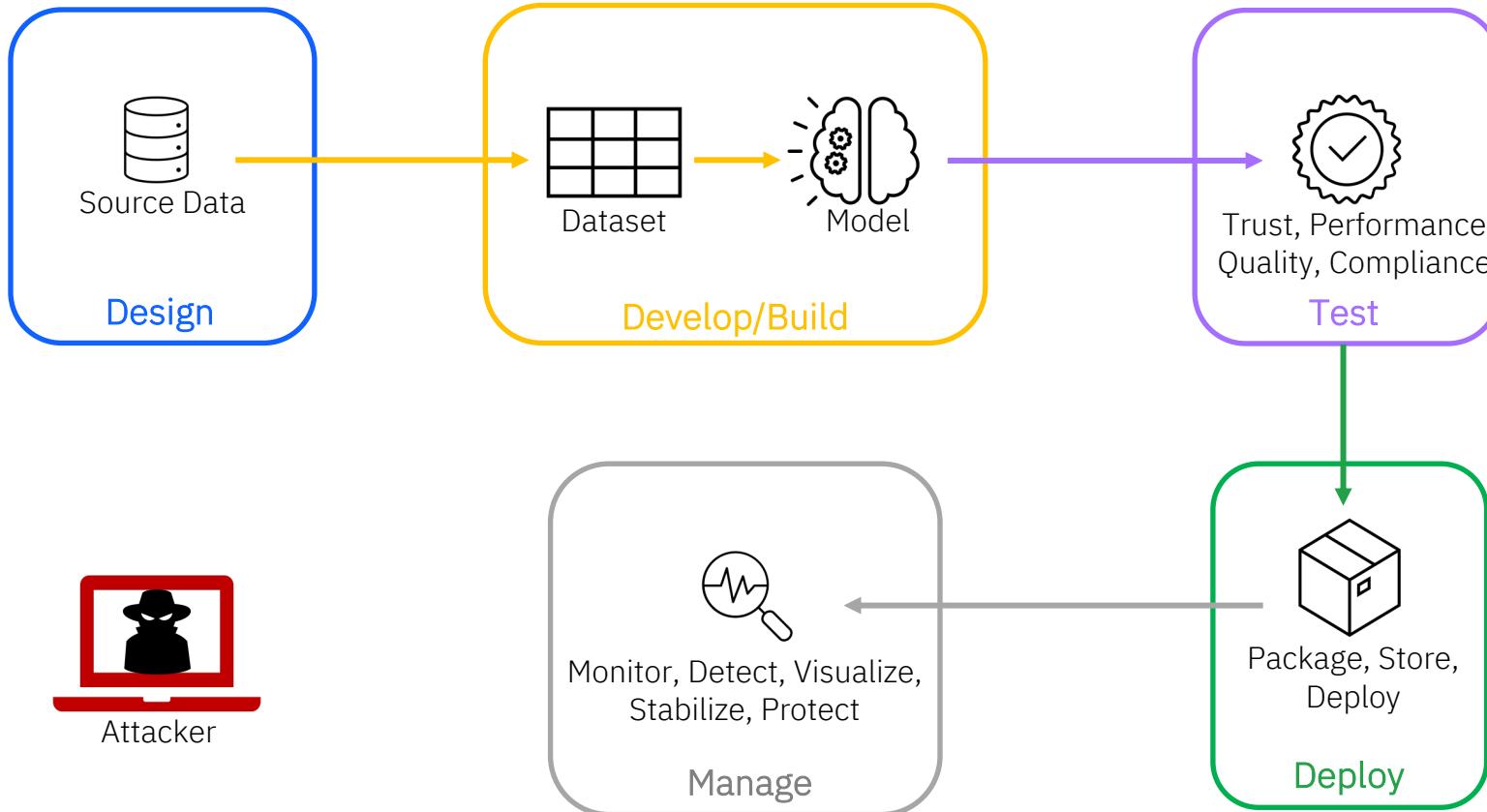
# Data Extraction



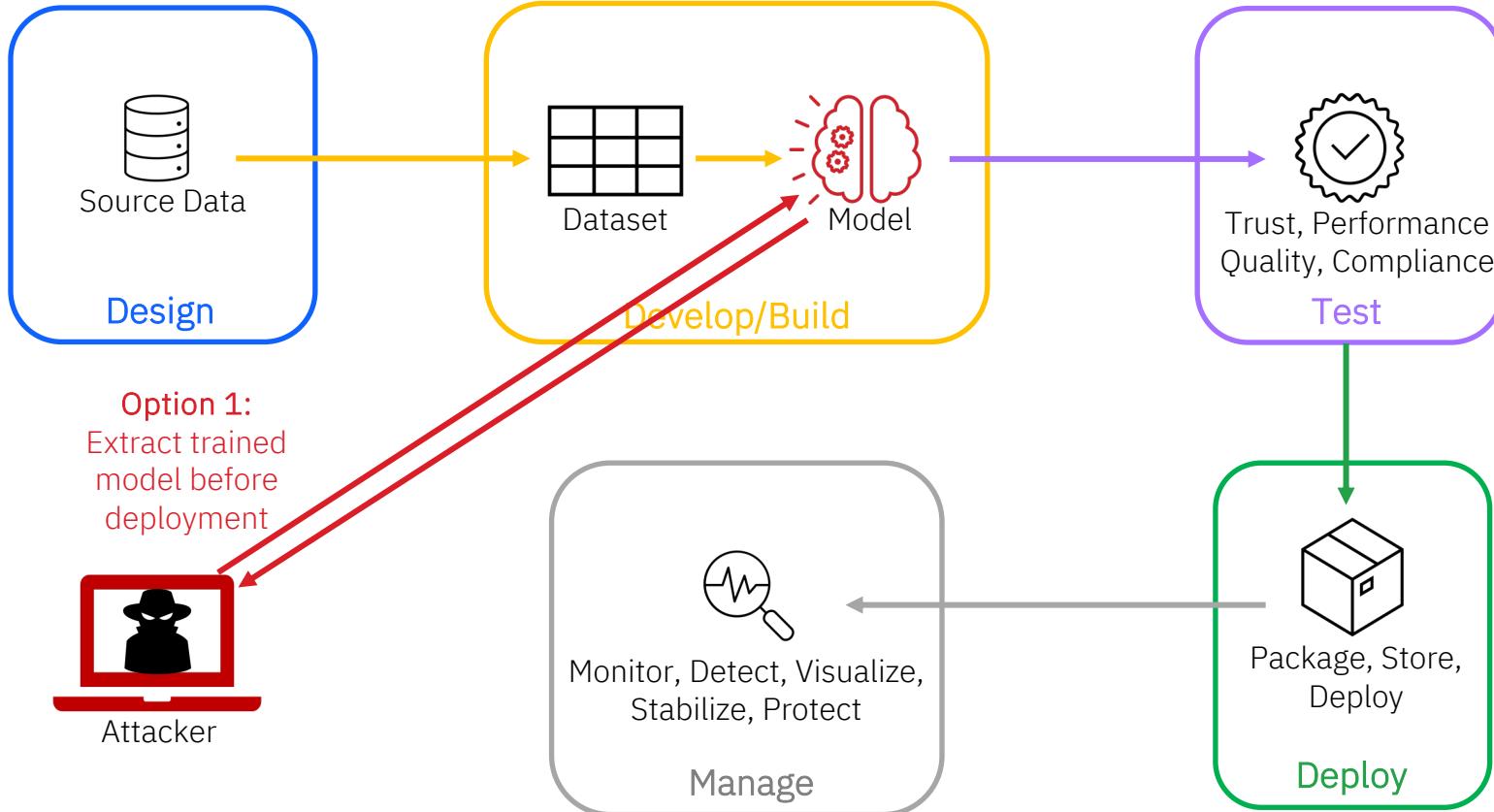
# Data Extraction



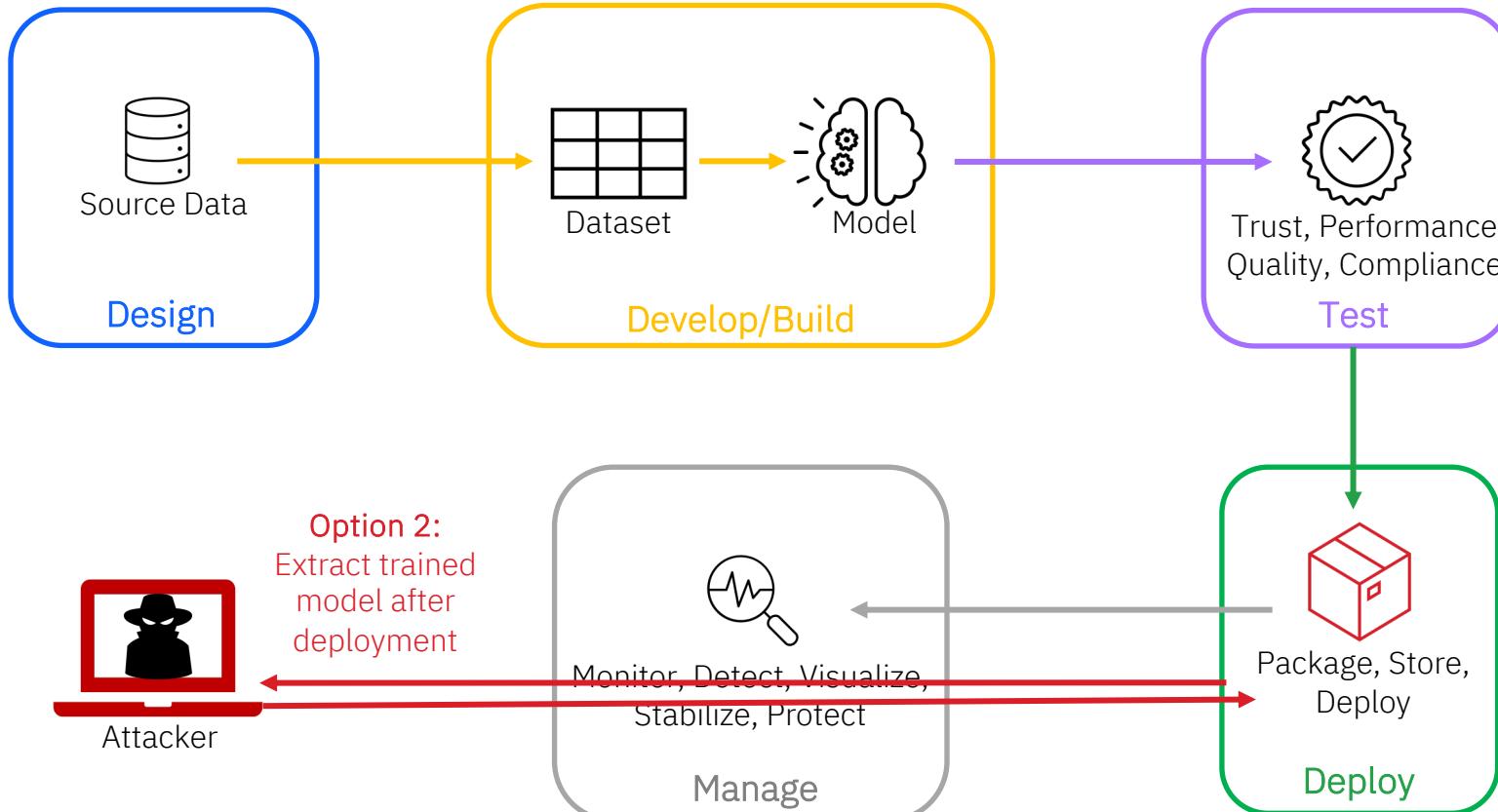
# Model Extraction



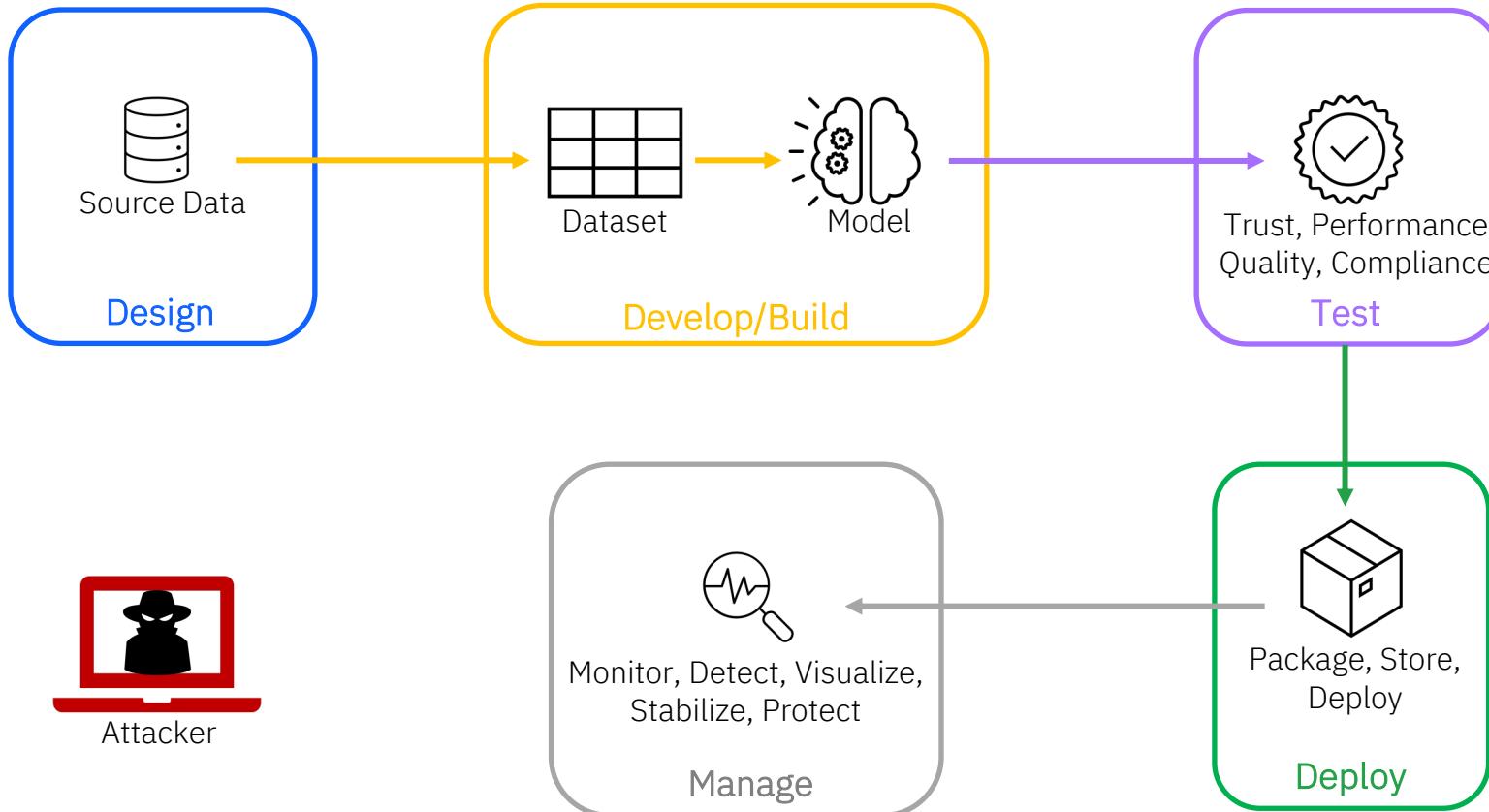
# Model Extraction



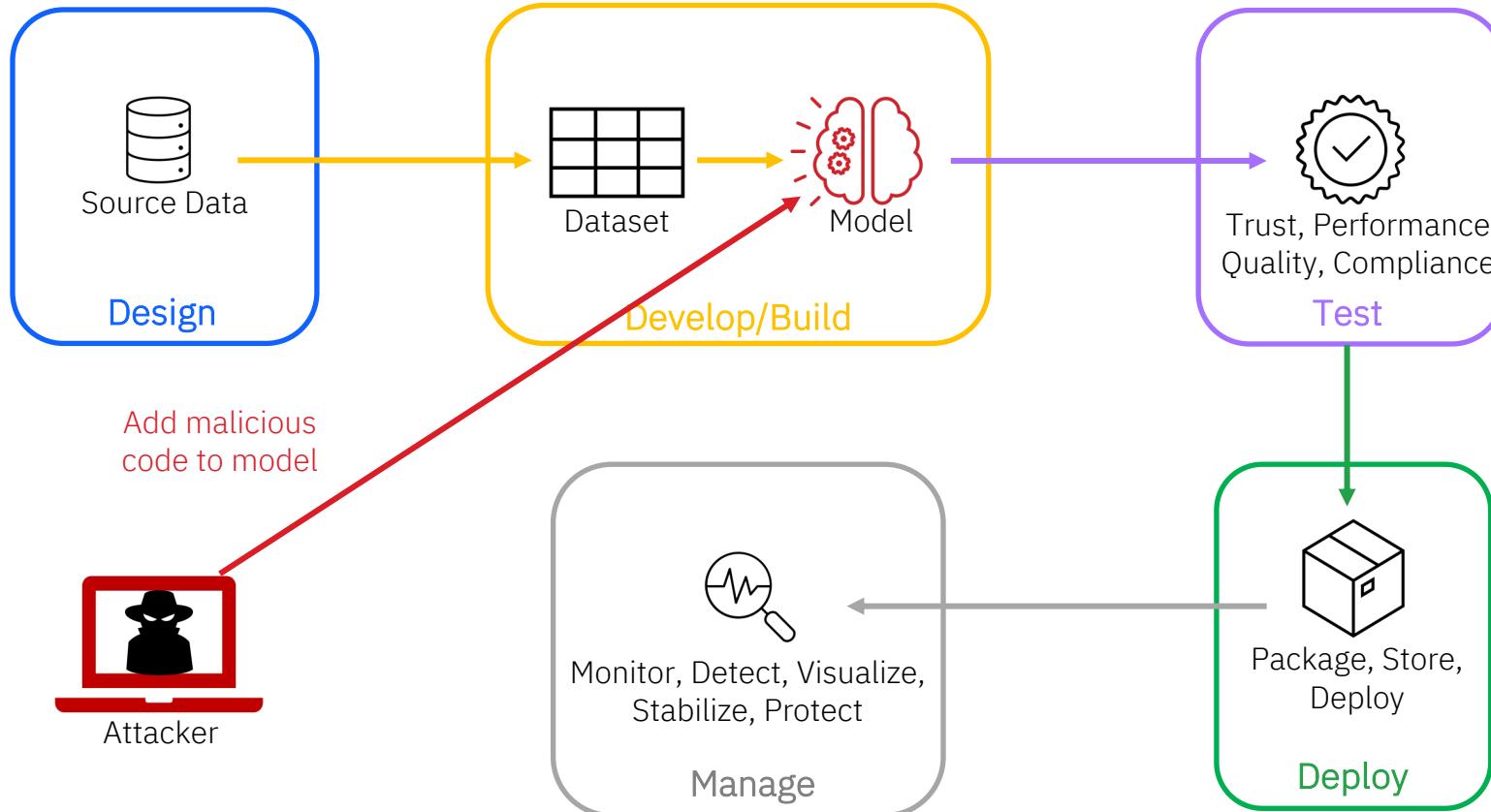
# Model Extraction



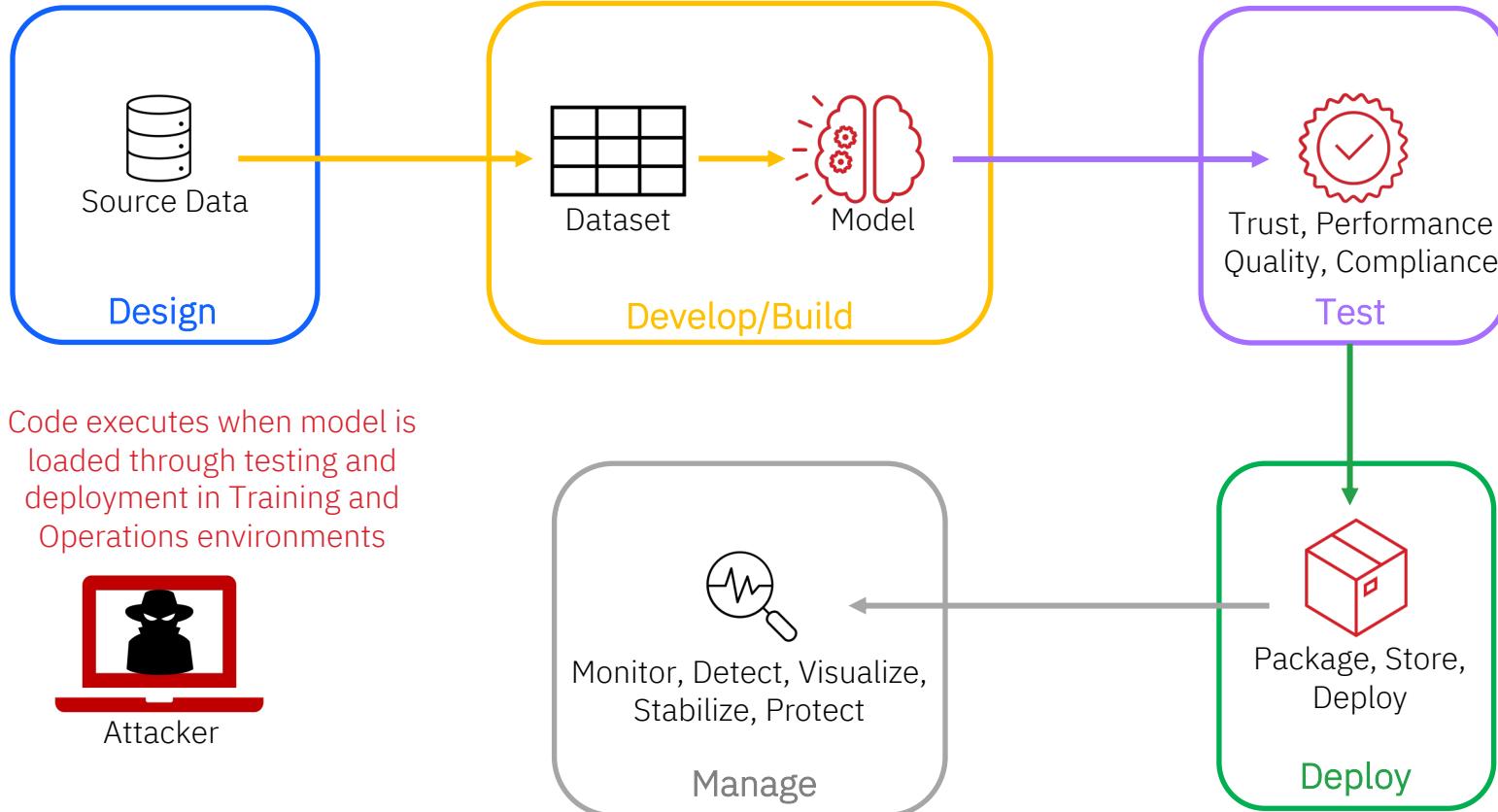
# Code Execution



# Code Execution

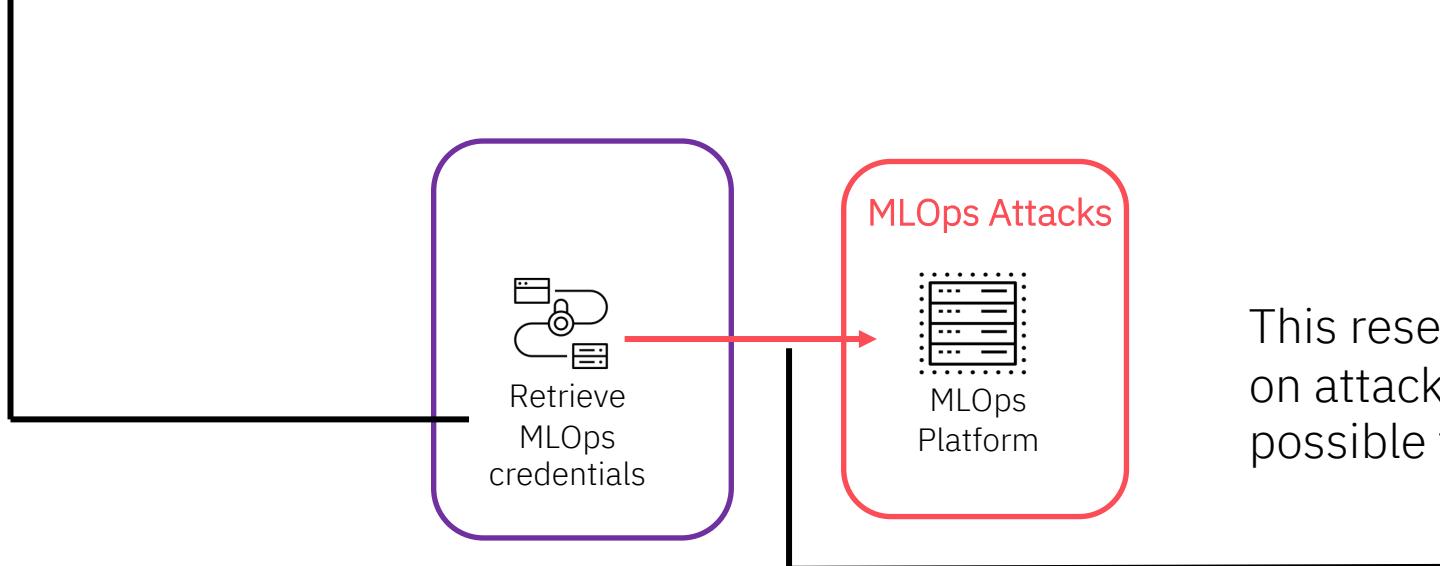


# Code Execution

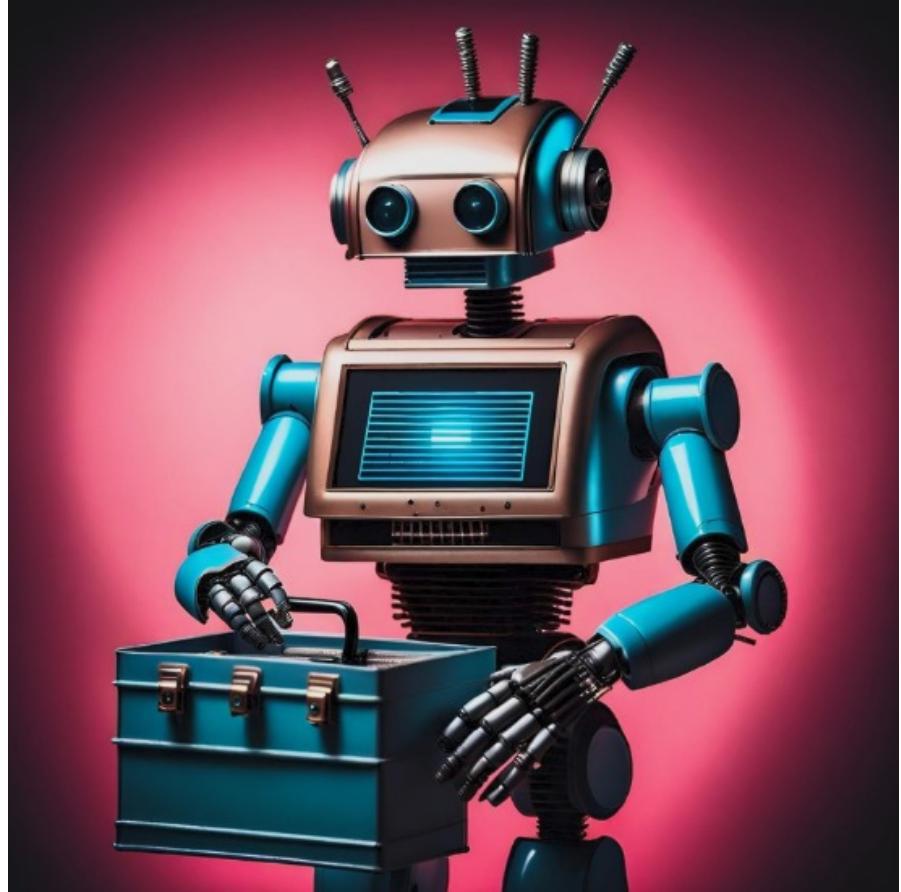


# Research Focus

File Shares  
Intranet Sites  
User Workstations  
Social Engineering  
Public Resources (e.g., Code Repos)  
Misconfigured Internal Network Recourses



# MLOKit



# Background

[github.com/xforced/MLOKit](https://github.com/xforced/MLOKit)

The screenshot shows the GitHub repository page for MLOKit. The 'Table of Contents' section lists various modules and examples, including MLOKit, Installation/Building, Command Modules, Arguments/Options, Authentication Options, Module Details Table, Examples, Detection, and References. The 'Release' section indicates that Version 1.0 is available in the Releases tab.

- Version 1.0 of MLOKit can be found in Releases

## Table of Contents

- [MLOKit](#)
- [Table of Contents](#)
- [Installation/Building](#)
  - [Libraries Used](#)
  - [Pre-Compiled](#)
  - [Building Yourself](#)
- [Command Modules](#)
- [Arguments/Options](#)
- [Authentication Options](#)
- [Module Details Table](#)
- [Examples](#)
  - [Check](#)
  - [List Projects/Workspaces](#)
  - [List Models](#)
  - [List Datasets](#)
  - [Download Model](#)
  - [Download Dataset](#)
- [Detection](#)
- [References](#)



**REST API Abuse**  
Conduct actions  
programmatically



**6 Modules**  
Recon, Data  
Extraction,  
Model Extraction



**Authentication**  
Supports API  
Key or Access  
Token

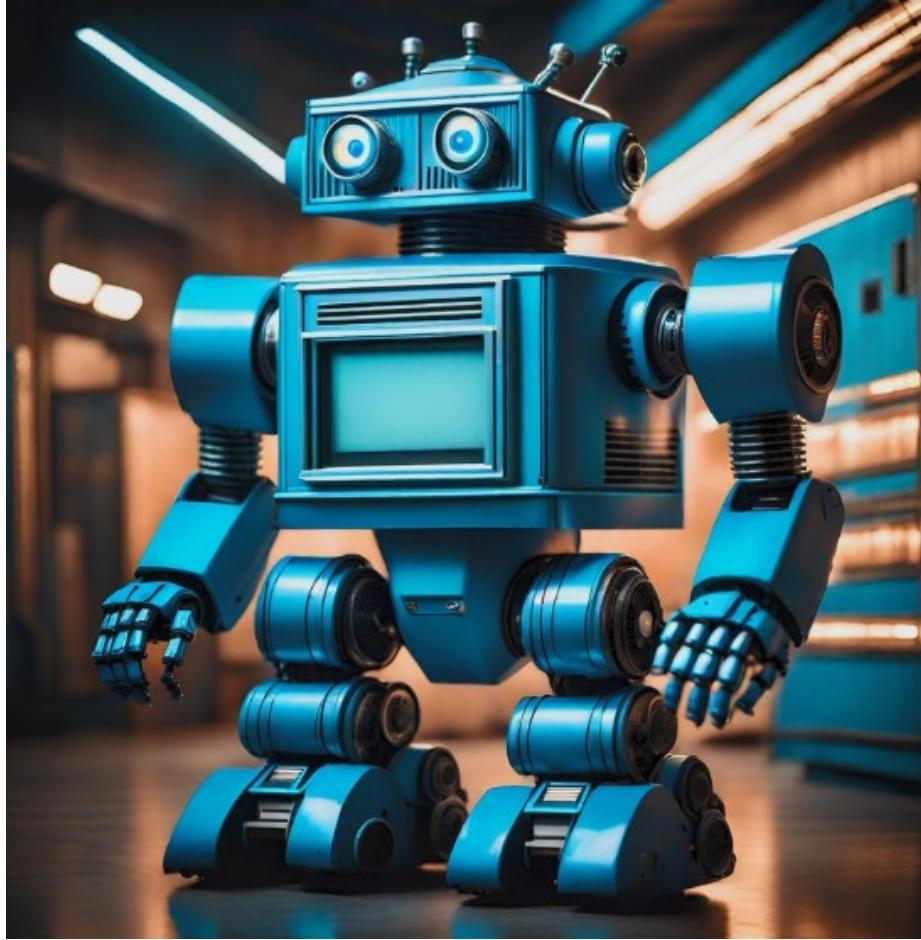


**3 Supported  
Platforms**  
Azure ML, BigML,  
Vertex AI

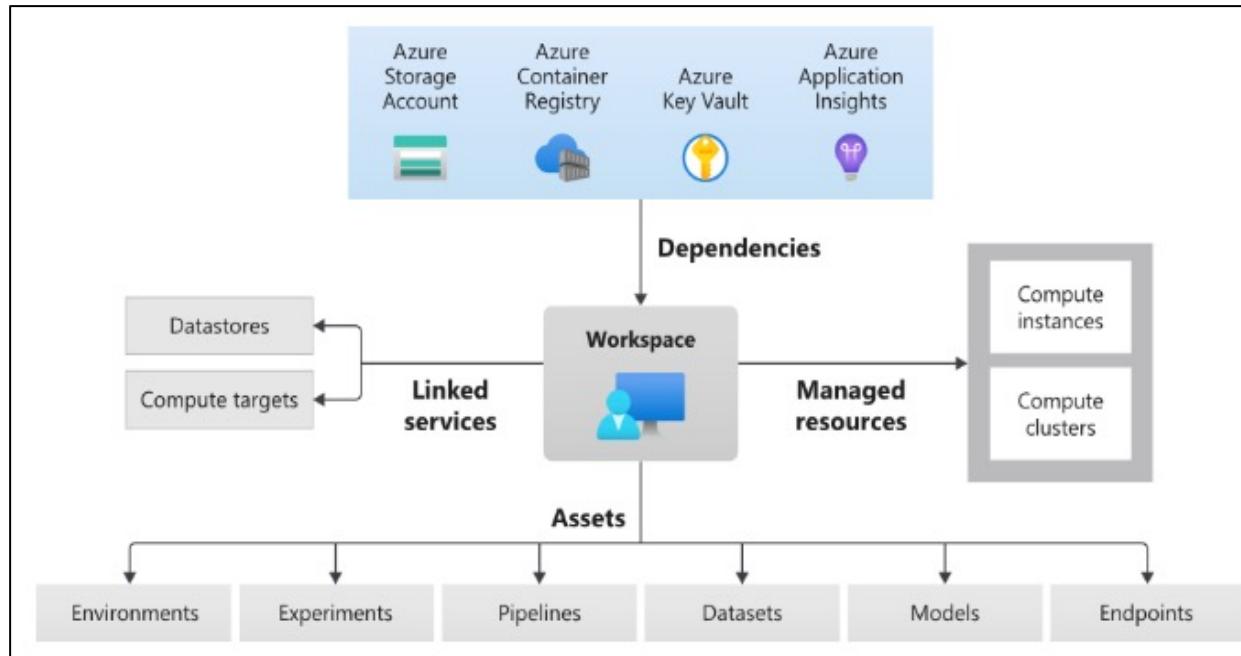
# Attacking MLOps Platforms



# Azure ML



# Key Terminology



<https://www.trendmicro.com/vinfo/gb/security/news/cybercrime-and-digital-threats/uncovering-silent-threats-in-azure-machine-learning-service-part-I>

# Authentication

<https://ml.azure.com/>



**Interactive**  
Authenticate with  
Entra ID  
authentication



**Service Principal**  
Use Service  
Principal ID and  
Key



**Azure CLI  
Session**  
Use AZ  
command-line  
tool with ML  
extension



**Managed Identity**  
Azure VM that can  
connect to  
workspace



**Access Token**  
Used to  
authenticate to  
Azure ML REST API

# Logging

Diagnostic setting name \* test

Logs

Category groups ⓘ

allLogs  audit

Categories

AmlComputeClusterEvent  
 AmlComputeClusterNodeEvent  
 AmlComputeJobEvent  
 AmlComputeCpuGpuUtilization  
 AmlRunStatusChangedEvent  
 ModelsChangeEvent

Destination details

Send to Log Analytics workspace

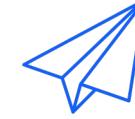
Subscription  
Azure subscription 1

Log Analytics workspace  
testing-sentinel ( eastus )

Archive to a storage account  
 Stream to an event hub  
 Send to partner solution



**Diagnostic Logging**  
Enable this as it is not by default

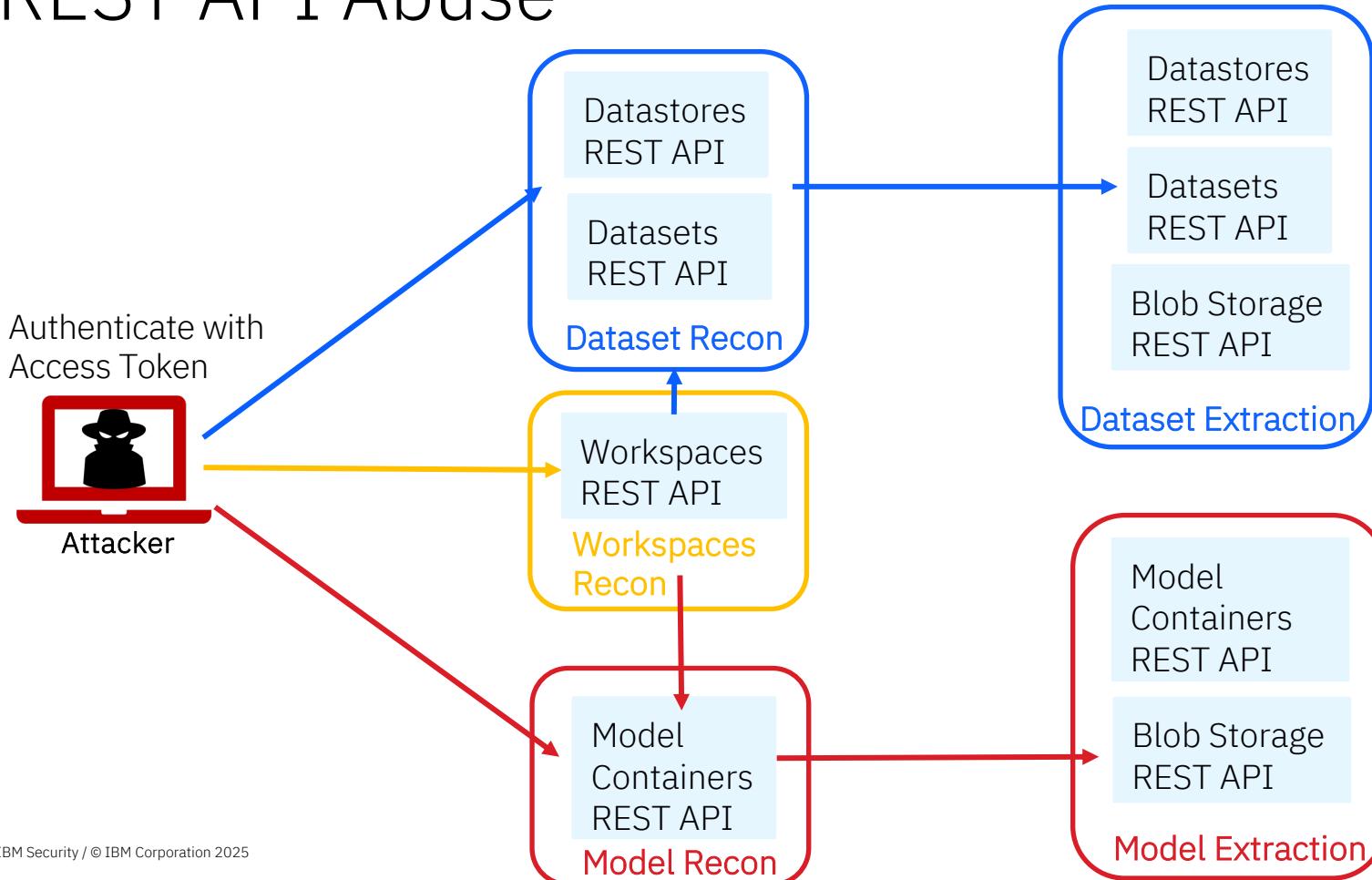


**Send Logs to Analytics Workspace**  
Detections can be built within Microsoft Sentinel

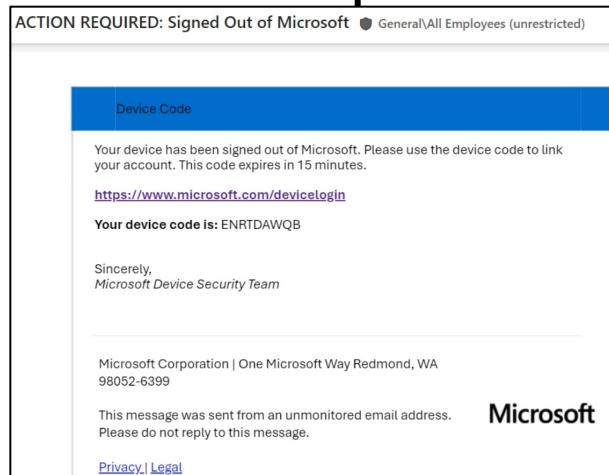
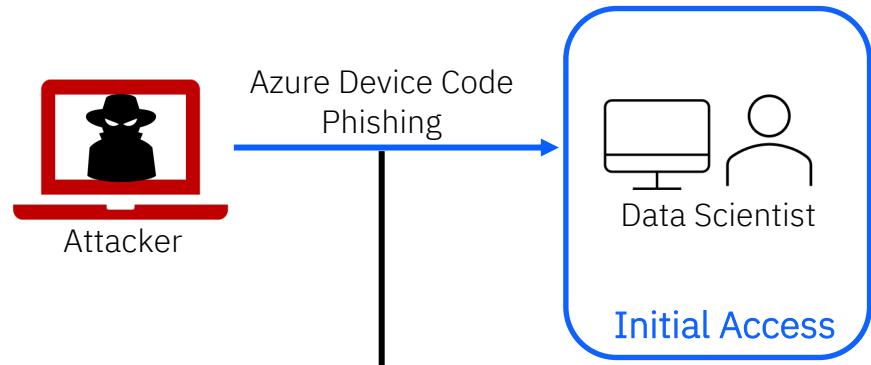


**Audit Logs Category**  
Required to log auditable events

# REST API Abuse

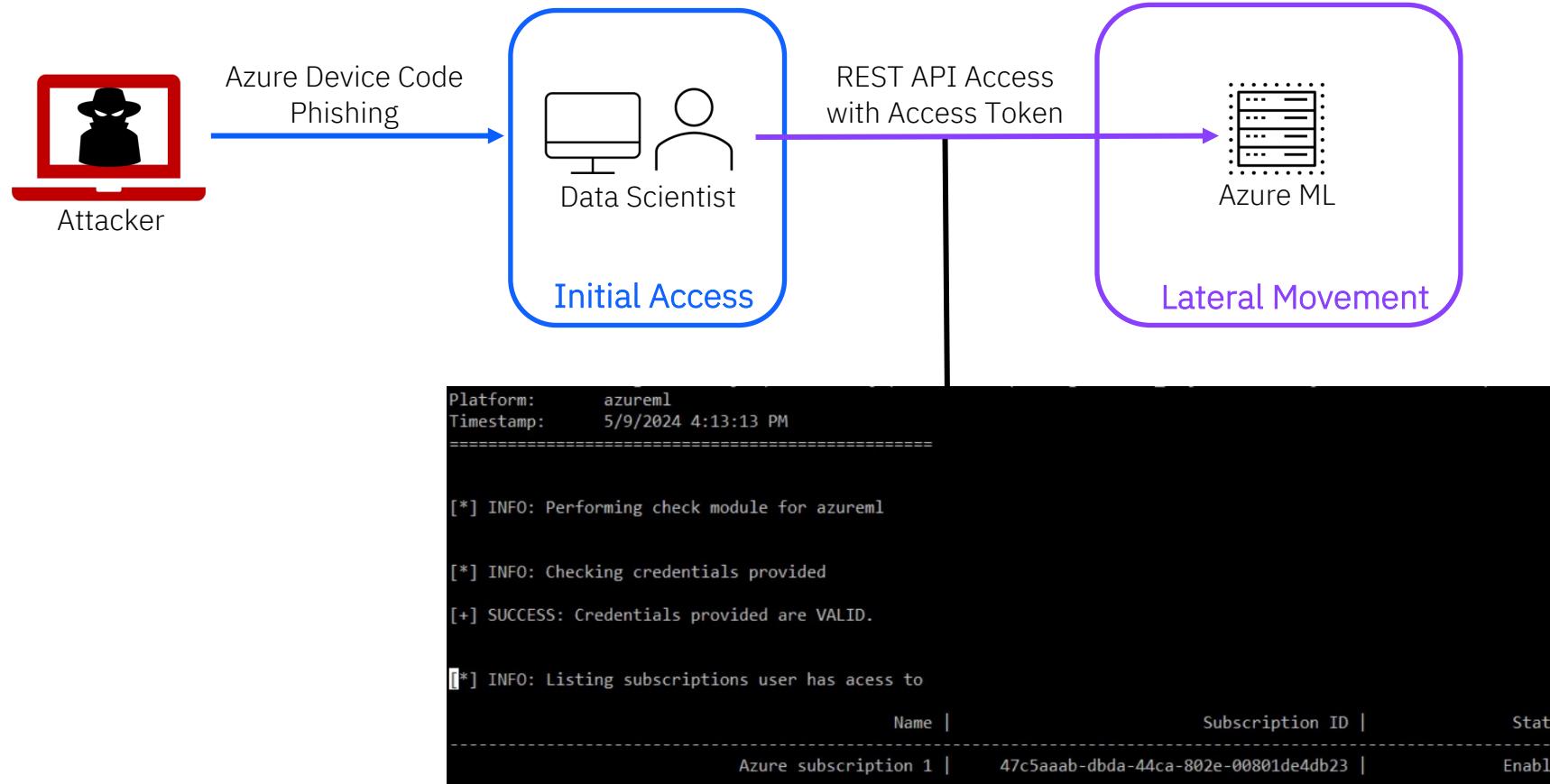


# Demo: Azure ML Model Extraction Attack

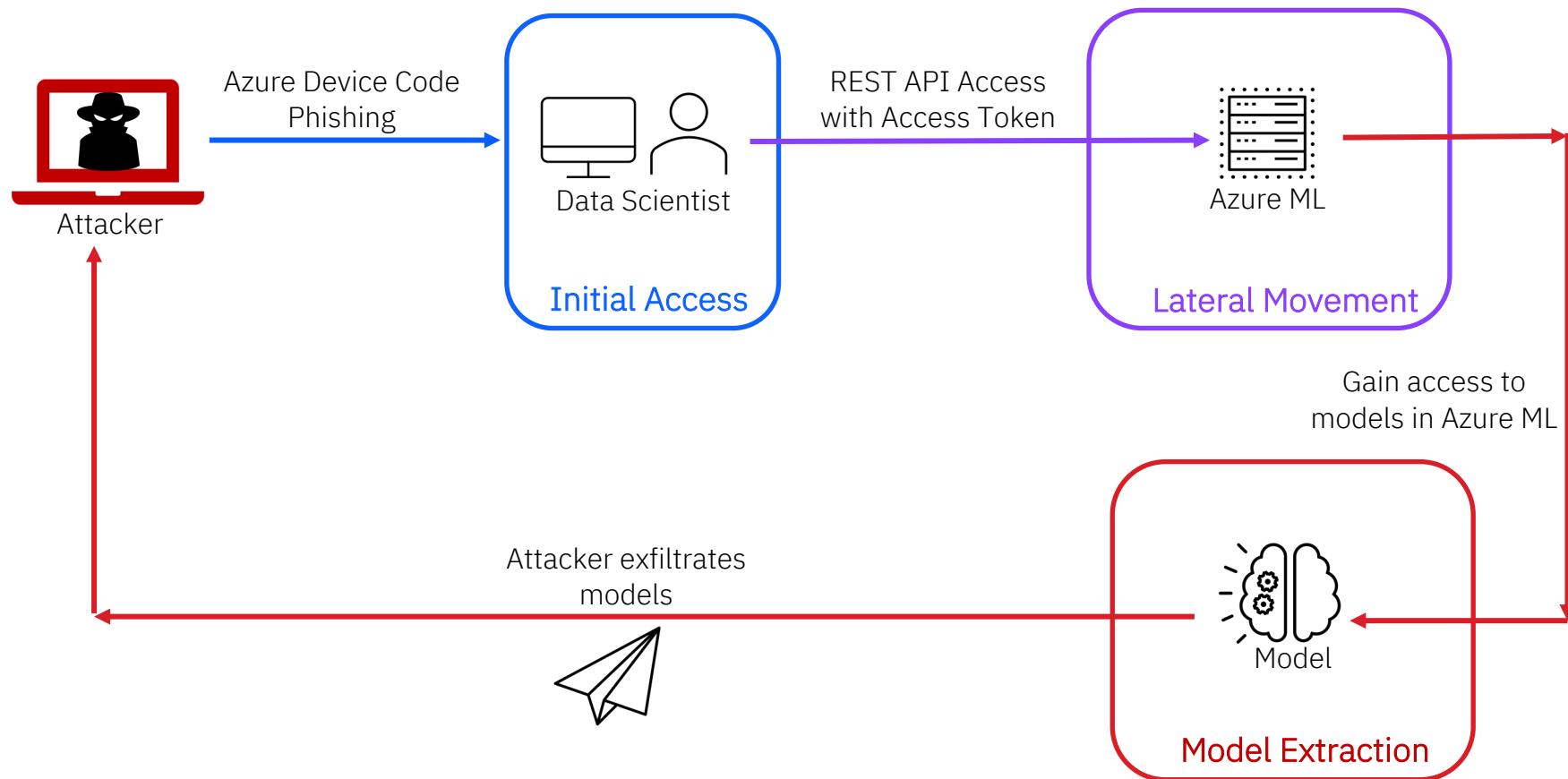


```
authorization_pending
authorization_pending
authorization_pending
token_type      : Bearer
scope          : user_impersonation
expires_in     : 7335
ext_expires_in : 7335
expires_on     : 1715292882
not_before     : 1715285246
resource        : https://management.core.windows.net/
access_token   : eyJ0eXAiOiJKV1QiLCJhbGciOiJSUzI1NiIsIng1dCI6IkwxS2ZLRk153aw5kb3dzLm5ldc8wMTdjZWJmMy0yMDYwLTQyOWEtOTJjM1hOTA3MTTNUdHl1SFdNZFp0SzdvNmVUz1BPdUNoTG52MEhMMjU2SEFoeVl3SEpYbH1fbmFtZSI6I1NjaWvudGlzdCIsImdpdmVuX25hbWUiOjJEYXRhIiw:wmzdGNTRDMDczIiwimcgiOiIwlkFWUE4LX04QVdbZ21rS1N3cWtIR11ViZjMtMjA2MC00MjhlTkYzItYTkwNzE5MDY3NjlmIiwiIw5pcXV1XGkiOizdEk5eG5CeUlrt2NWZ1FTaWNpX0FBiiwidmVyIjojMS4wIiwiIbbCbnIqdRQ0Mp0z_fXM7Sbc2S1U9LFXuhRWF04CmTKNDNtAK88rG-S:iwikSIWaBjr6Ve14I4ZEEmKRCw5epmNYn-3n7fGRPJLF6A-hNKR3Arefresh_token : 0.AXYA8-t8AWAgmkKSwkHGQZ2n9Y0WdOzUgJBrv-q0ikqsBy0ABo.Aq
```

# Demo: Azure ML Model Extraction Attack



# Demo: Azure ML Model Extraction Attack



C:\Demo>



# Listing stolen access token

# Model Extraction - Logs

## AmlModelsEvent

```
| where OperationName endswith "WORKSPACES/MODELS/READ"  
| project TimeGenerated, ResultType, Identity, OperationName, Aml modelName
```

The screenshot shows a log search interface with the following details:

Code Snippet:

```
1 AmlModelsEvent..  
2 | where OperationName endswith "WORKSPACES/MODELS/READ"  
3 | project TimeGenerated,ResultType,Identity,OperationName,AmlmodelName  
4
```

Results View:

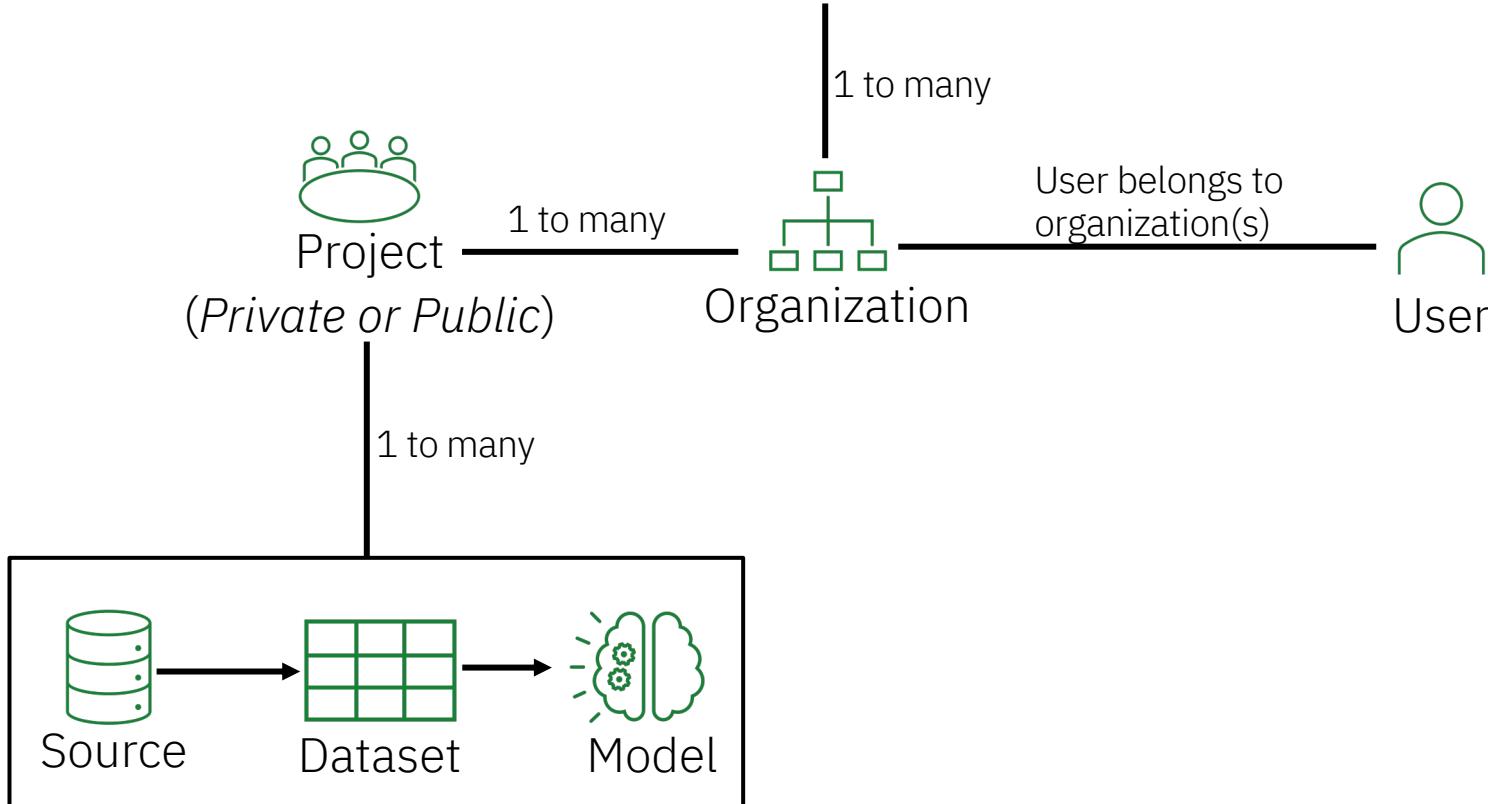
Results

TimeGenerated [UTC]	ResultType	Identity	OperationName	AmlmodelName
4/19/2024, 12:51:18.306 PM	Succeeded	{"UserObjectId": "d852e46b-de3a-4a39-8ec9-83d22327b00e", "UserName": "brett.hawkins@redacted.com"}	MICROSOFT.MACHINELEARNINGSERVICES/WORKSPACES/MODELS/READ	taxifare-output-model
TimeGenerated [UTC]	2024-04-19T12:51:18.306265Z			
ResultType	Succeeded			
Identity		{"UserObjectId": "d852e46b-de3a-4a39-8ec9-83d22327b00e", "UserName": "brett.hawkins@redacted.com"}		
OperationName			MICROSOFT.MACHINELEARNINGSERVICES/WORKSPACES/MODELS/READ	
AmlmodelName				taxifare-output-model

# BigML

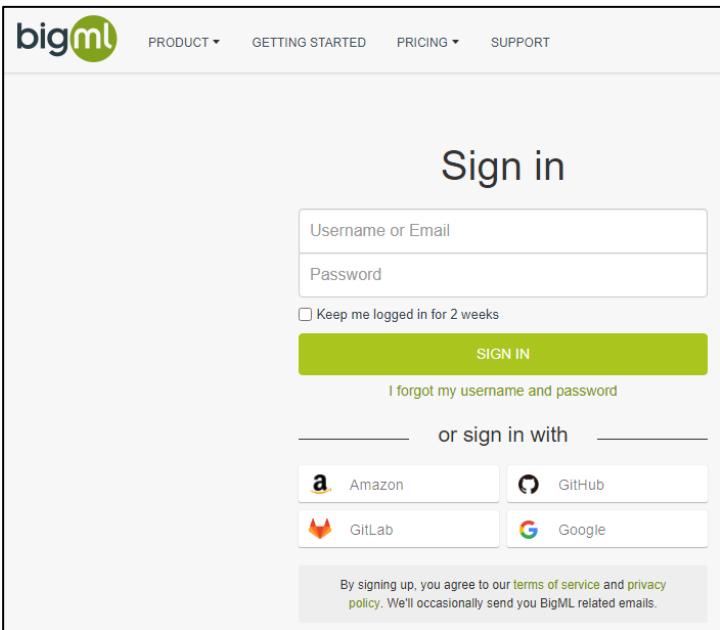


# Key Terminology



# Authentication

[https://bigml.com/dashboard/organization/\[ORGANIZATION\\_NAME\]](https://bigml.com/dashboard/organization/[ORGANIZATION_NAME])



The screenshot shows the BigML sign-in interface. At the top, there's a navigation bar with links for PRODUCT, GETTING STARTED, PRICING, and SUPPORT. Below the navigation is a large "Sign in" button. The main form has two input fields: "Username or Email" and "Password". There's also a checkbox for "Keep me logged in for 2 weeks". A green "SIGN IN" button is centered below the inputs. Below the button, there's a link to "I forgot my username and password". Underneath the sign-in section, there's a "or sign in with" section featuring social media logins for Amazon, GitHub, GitLab, and Google. At the bottom of the form, a small note states: "By signing up, you agree to our [terms of service](#) and [privacy policy](#). We'll occasionally send you BigML related emails."



**BigMLer**  
Authenticate with  
API Key



## Web Interface

- MFA
- User/Pass
  - Native
  - 3<sup>rd</sup> Party



**REST API**  
Authenticate with  
API Key

# Logging

- Private deployment required for logging ability
- No logging in cloud-based BigML

## Private deployments

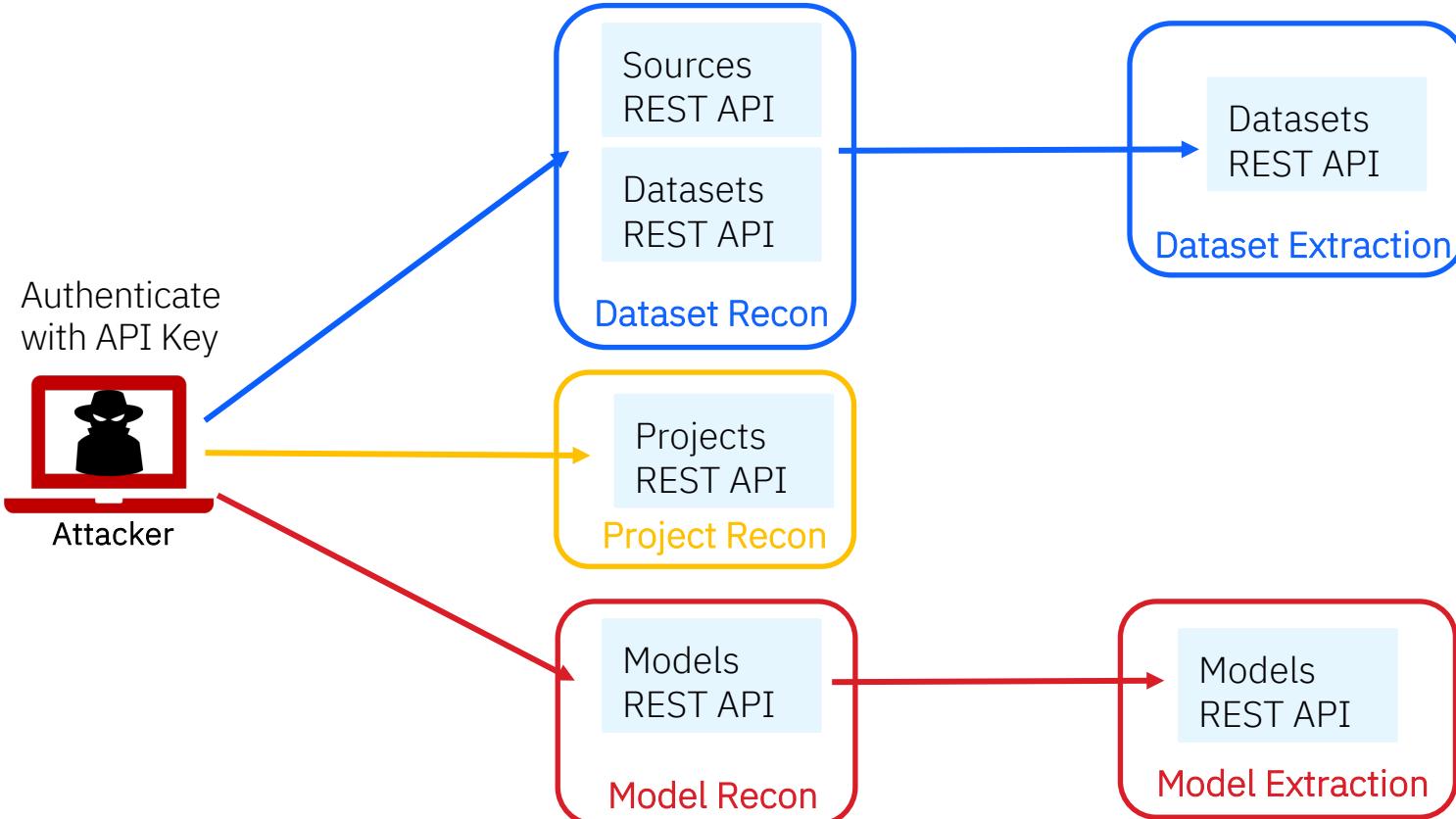
For companies with stringent data security, privacy, or regulatory requirements, BigML offers Private Deployments that can run on your preferred cloud provider, ISP, or own infrastructure with commodity servers to meet enterprise grade requirements such as traceability and repeatability for all your workflows. [More info.](#)

BigML Enterprise	BigML Lite
Accelerate the Machine Learning adoption in your company	All the power of BigML more accessible than ever
<ul style="list-style-type: none"><li>Unlimited users</li><li>Unlimited organizations</li><li>Personalized theme and logo · Prioritized feature request · Auto-scaling</li><li>Customized direct email and chat 24-hour max. response time</li></ul>	<ul style="list-style-type: none"><li>5 users</li><li>1 organization</li><li>BigML standard theme</li><li>Standard 8x5 via email and chat 48-hour max. response time</li></ul>
<p>Bronze Enterprise (Up to 1 server / 8 cores)</p> <p>\$45,000/year +\$10,000 setup fee</p> <p>REQUEST</p>	<p>1 server (8 cores) \$10,000/year \$1,000/month</p> <p>REQUEST</p>

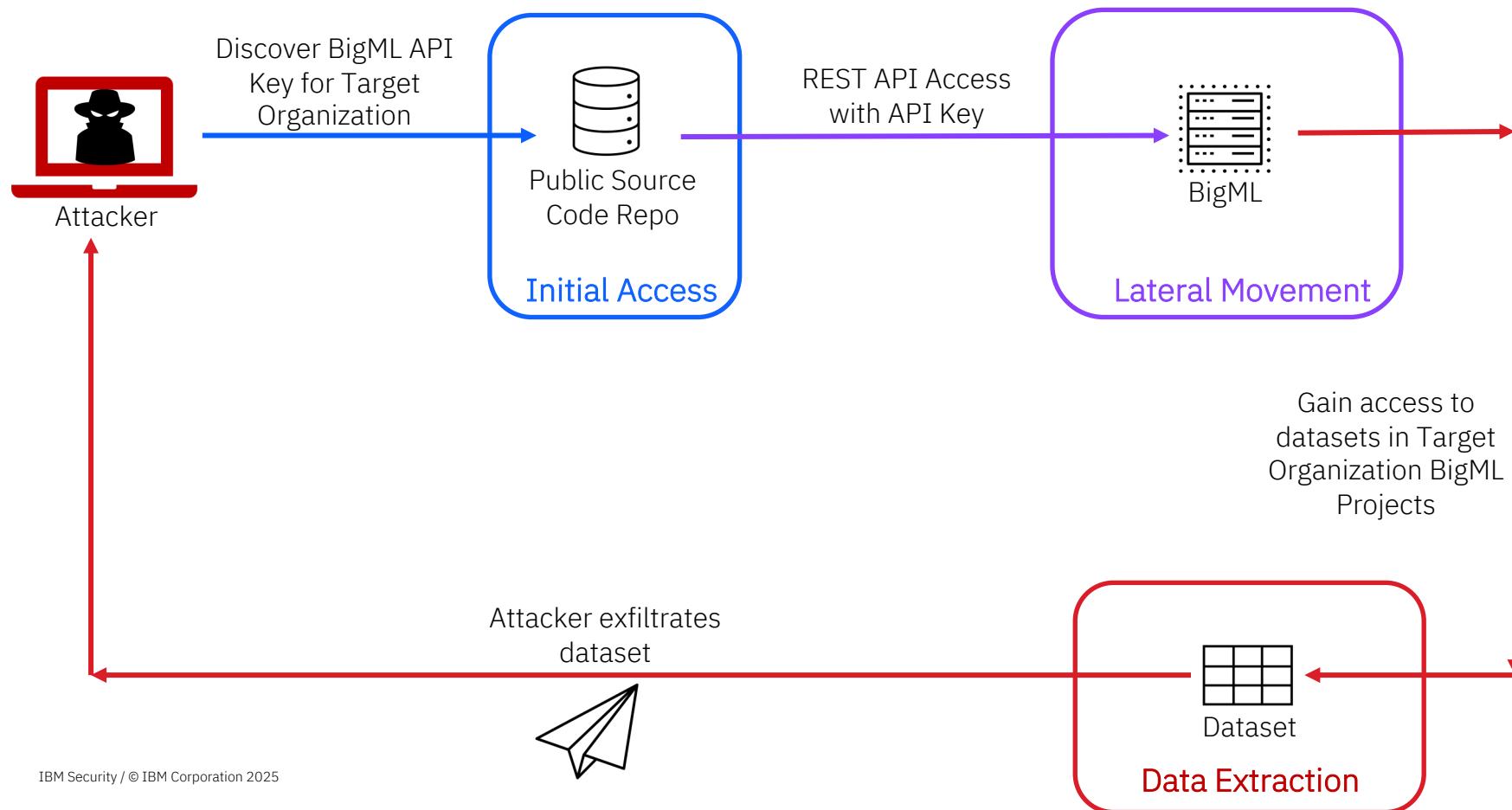
**ALL PRIVATE DEPLOYMENTS INCLUDE:**

- Unlimited tasks.
- Regular updates and upgrades of new features and algorithms.
- Priority access to customized assistance.
- Easy upgrades to bigger deployments.

# REST API Abuse



# Demo: BigML Data Extraction Attack



C:\Temp&gt;



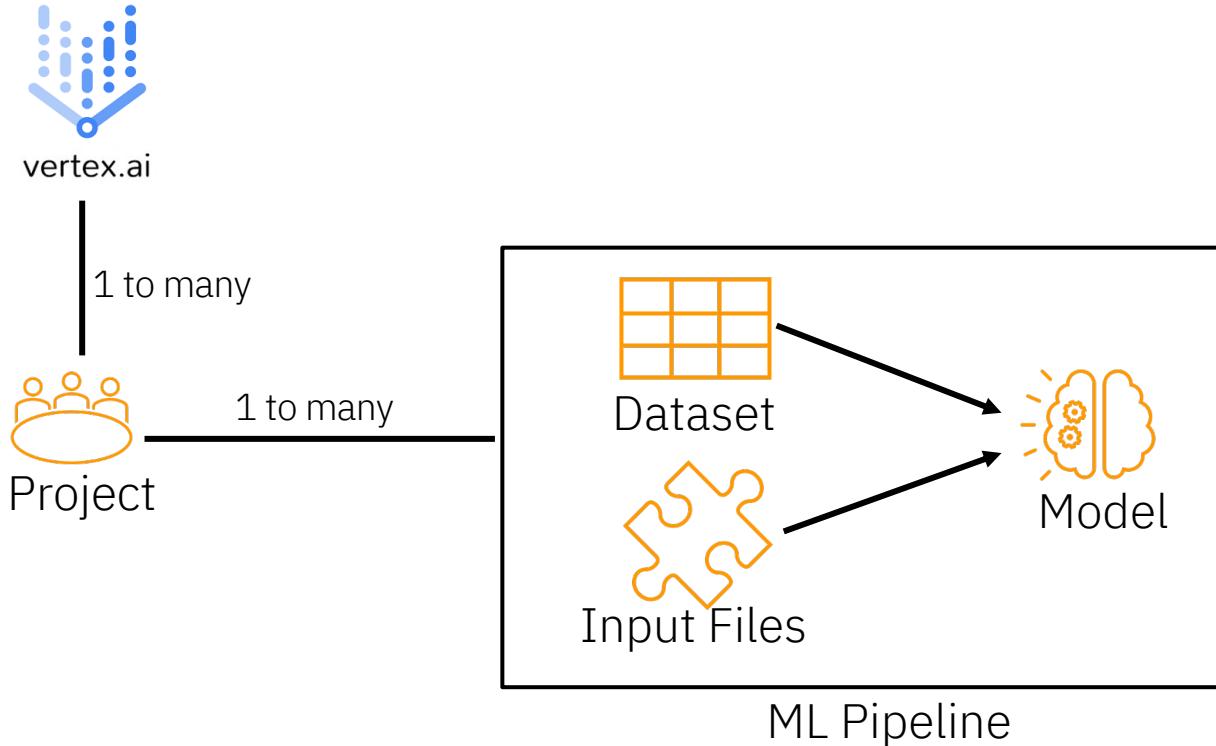
Type here to search



# Google Cloud Vertex AI

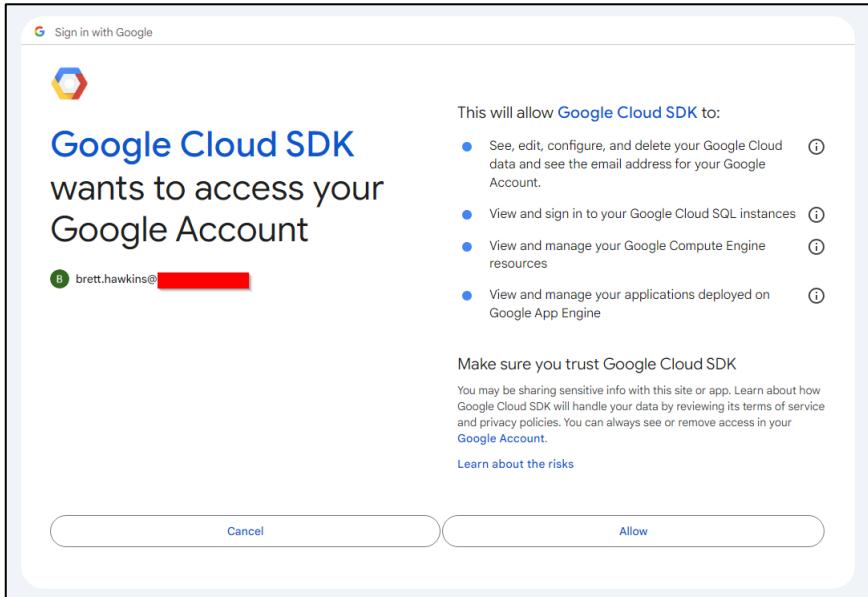


# Key Terminology



# Authentication

<https://console.cloud.google.com/vertex-ai>



## Client Libraries

Authenticate with Application Default Credentials (ADC)



## REST API

Authenticate with Access Token



## Google Cloud CLI

Authenticate with GCloud account

# Logging

Audit Logs [SET DEFAULT CONFIGURATION](#)

**Default configuration** [?](#) Admin Read: [Disabled](#)  
0 exempted principals

**Data Access audit logs configuration**  
The effective data access configuration below combines the configuration for the currently selected resources.

**Filter** [vertex](#) Enter property name or value

Service	Admin Read	Data Read	Data Write
<input checked="" type="checkbox"/> Vertex AI API	—	—	—

**Data Access audit logs configuration**  
The effective data access configuration below combines the configuration for the currently selected resources.

**Filter** [Google Cloud Storage](#) [X](#) Enter property name or value

Service	Admin Read	Data Read	Data Write
<input checked="" type="checkbox"/> Google Cloud Storage	—	—	—



**Vertex AI Data Access Audit Logs**  
Enable this as it is not by default

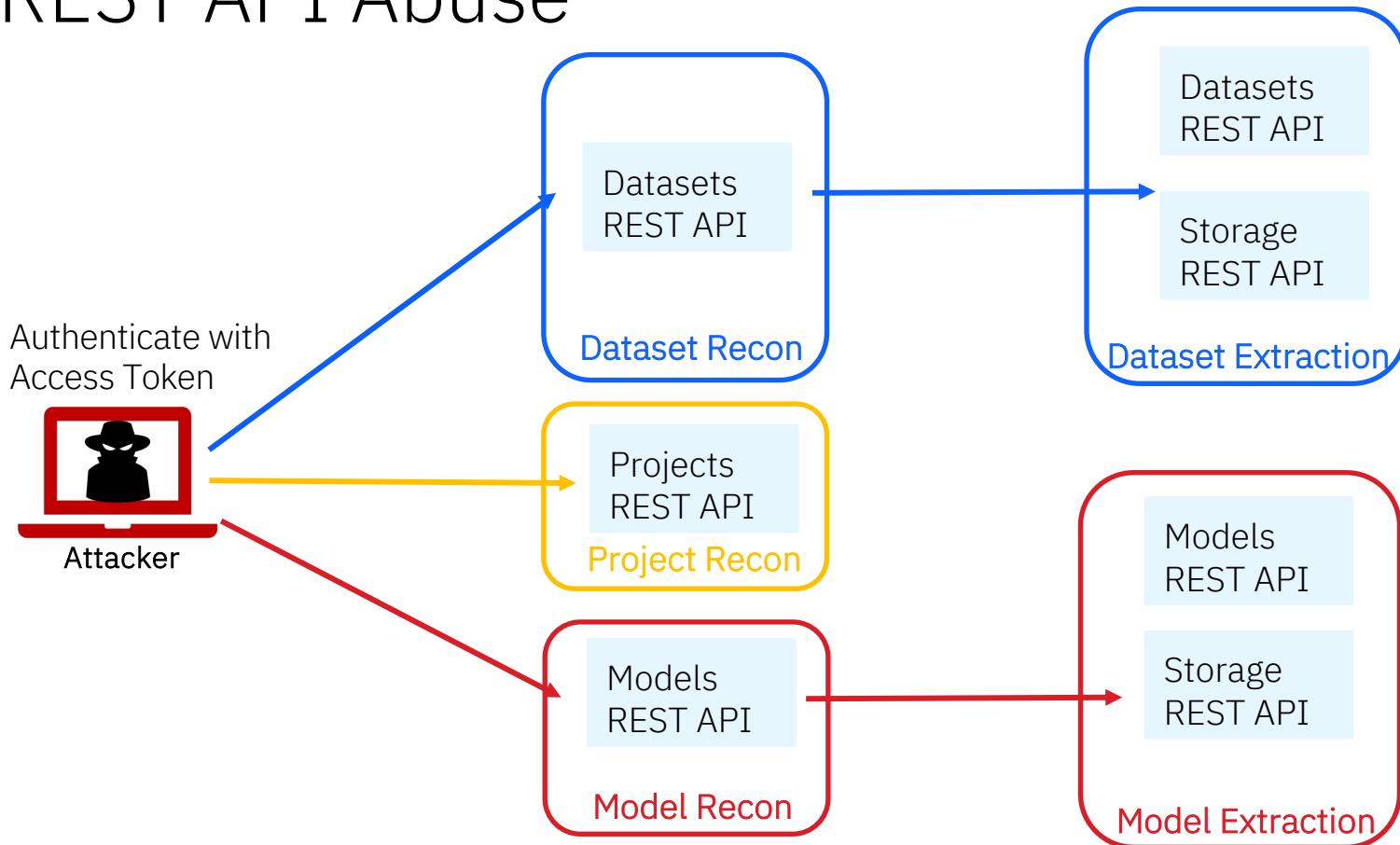


**Google Cloud Logs Explorer**  
Logs will be populated here that log alerts can be built

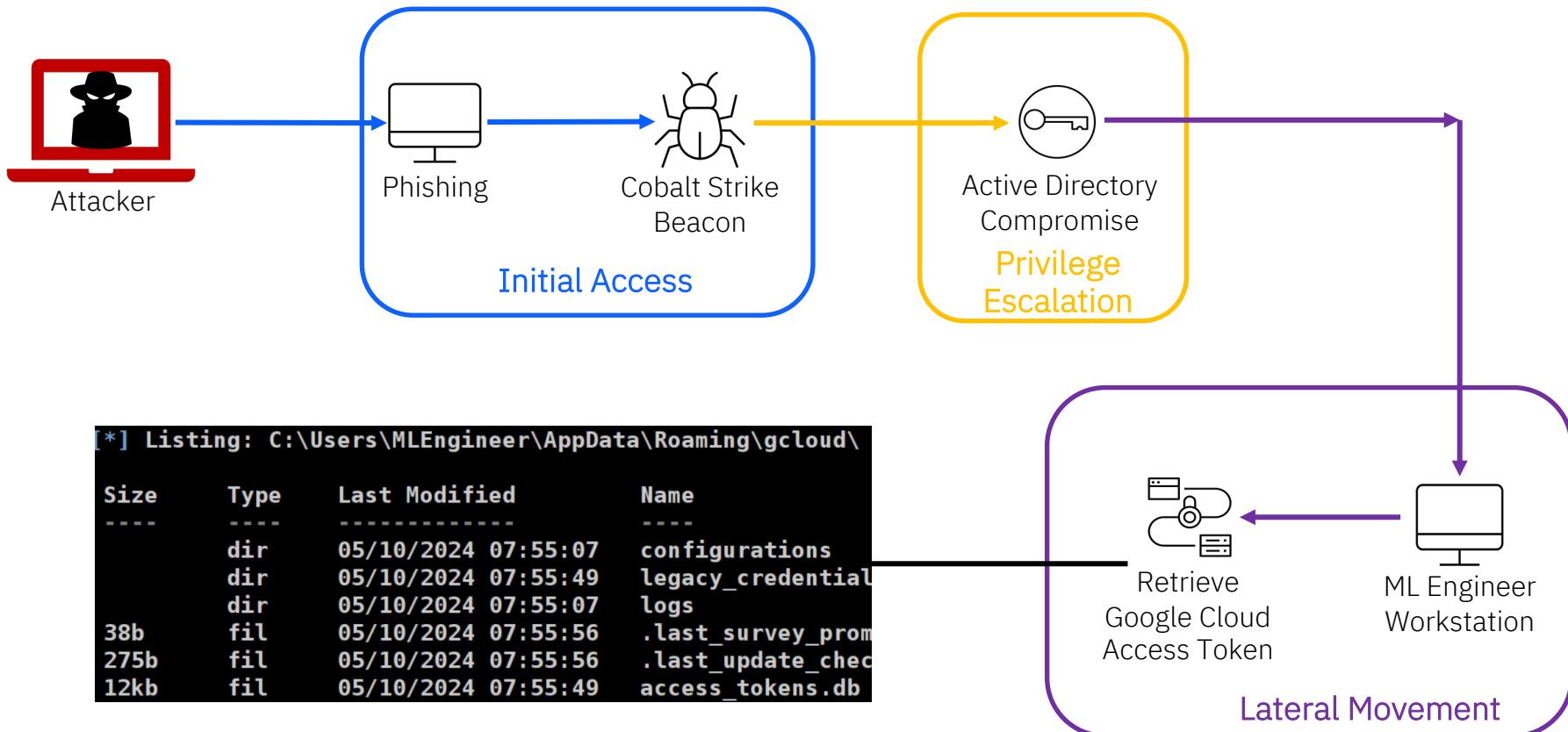


**Google Cloud Storage Audit Logs**  
Enable this as it is not by default

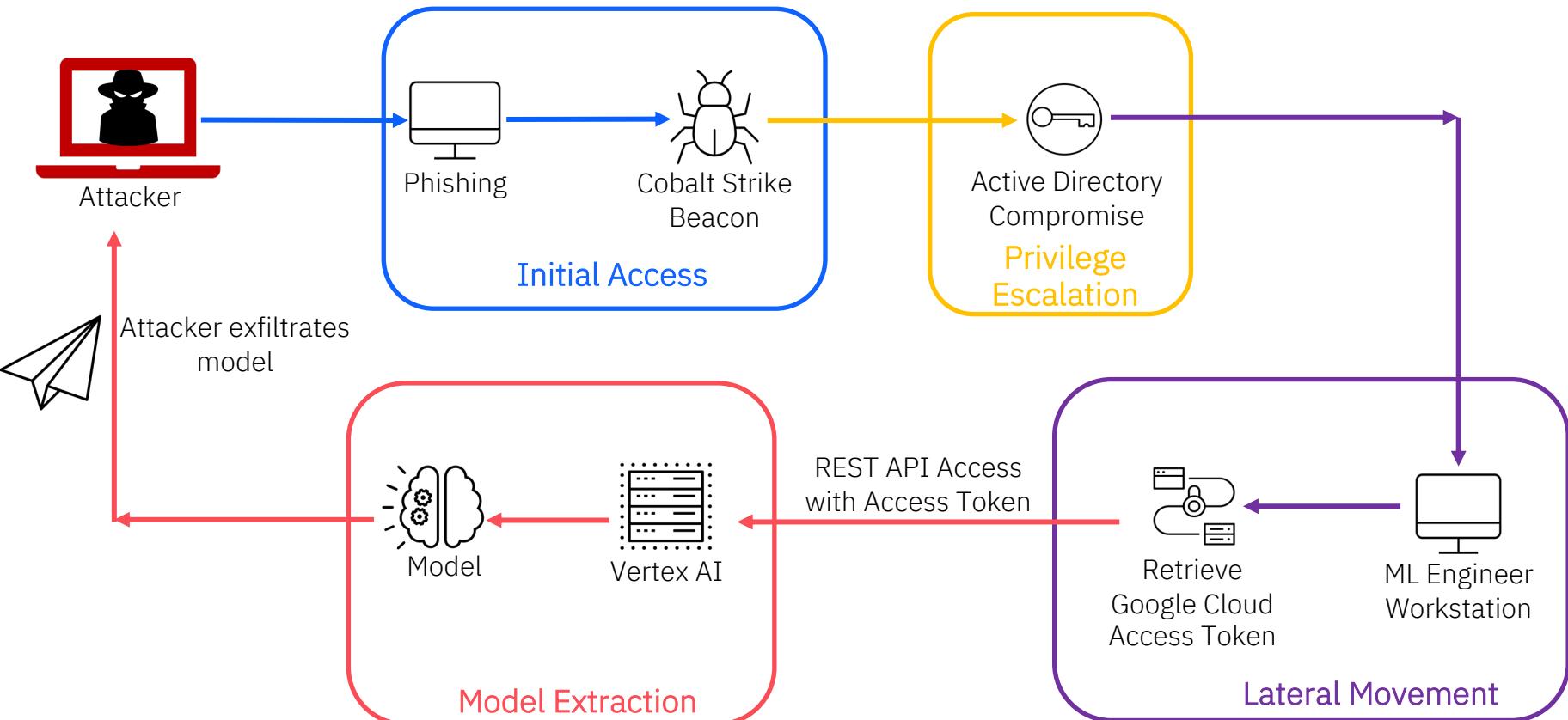
# REST API Abuse



# Demo: Vertex AI Model Extraction Attack



# Demo: Vertex AI Model Extraction Attack



Activities aggressor-Aggressor May 17 14:03 Cobalt Strike

Cobalt Strike View Payloads Attacks Site Management Reporting Help

external internal ^ listener user computer note process pid arch last sleep  
192.168.1.37 192.168.1.37 https MLEngineer DESKTOP-YFCE... werfault.exe 12524 x64 2s Interactive

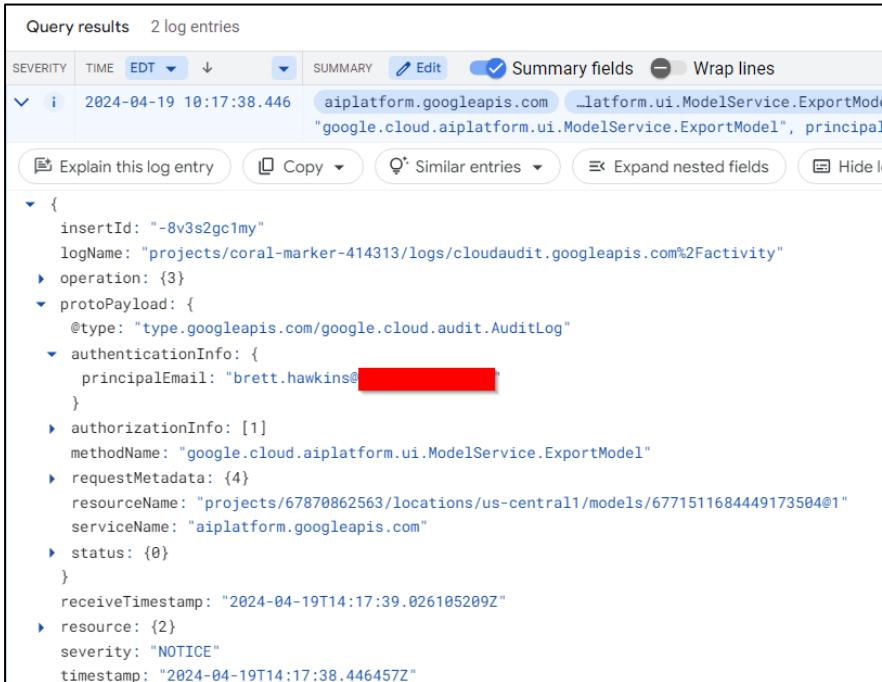
>\_ •

beacon>

# Dumped GCloud CLI Access Token

# Model Extraction - Logs

protoPayload.methodName="google.cloud.aiplatform.ui.ModelService.ExportModel"  
protoPayload.methodName="google.cloud.aiplatform.v1.ModelService.ExportModel"



The screenshot shows a log entry in the Google Cloud Logging interface. The log entry details a ModelService.ExportModel call made on 2024-04-19 at 10:17:38.446. The principal making the request is brett.hawkins@[REDACTED]. The log entry includes nested fields for insertId, logName, operation, protoPayload (which contains authenticationInfo, authorizationInfo, methodName, requestMetadata, status, receiveTimestamp, resource, severity, and timestamp), and status.

```
Query results 2 log entries
SEVERITY TIME EDT ↓ SUMMARY Edit Summary fields Wrap lines
i 2024-04-19 10:17:38.446 aiplatform.googleapis.com ...latform.ui.ModelService.ExportModel "google.cloud.aiplatform.ui.ModelService.ExportModel", principal_
Explain this log entry Copy Similar entries Expand nested fields Hide log
{
  insertId: "-8v3s2gc1my"
  logName: "projects/coral-marker-414313/logs/cloudaudit.googleapis.com%2Factivity"
  operation: {3}
  protoPayload: {
    @type: "type.googleapis.com/google.cloud.audit.AuditLog"
    authenticationInfo: {
      principalEmail: "brett.hawkins@[REDACTED]"
    }
    authorizationInfo: [1]
    methodName: "google.cloud.aiplatform.ui.ModelService.ExportModel"
    requestMetadata: {4}
    resourceName: "projects/67870862563/locations/us-central1/models/6771511684449173504@1"
    serviceName: "aiplatform.googleapis.com"
  }
  status: {0}
}
receiveTimestamp: "2024-04-19T14:17:39.026105209Z"
resource: {2}
severity: "NOTICE"
timestamp: "2024-04-19T14:17:38.446457Z"
```

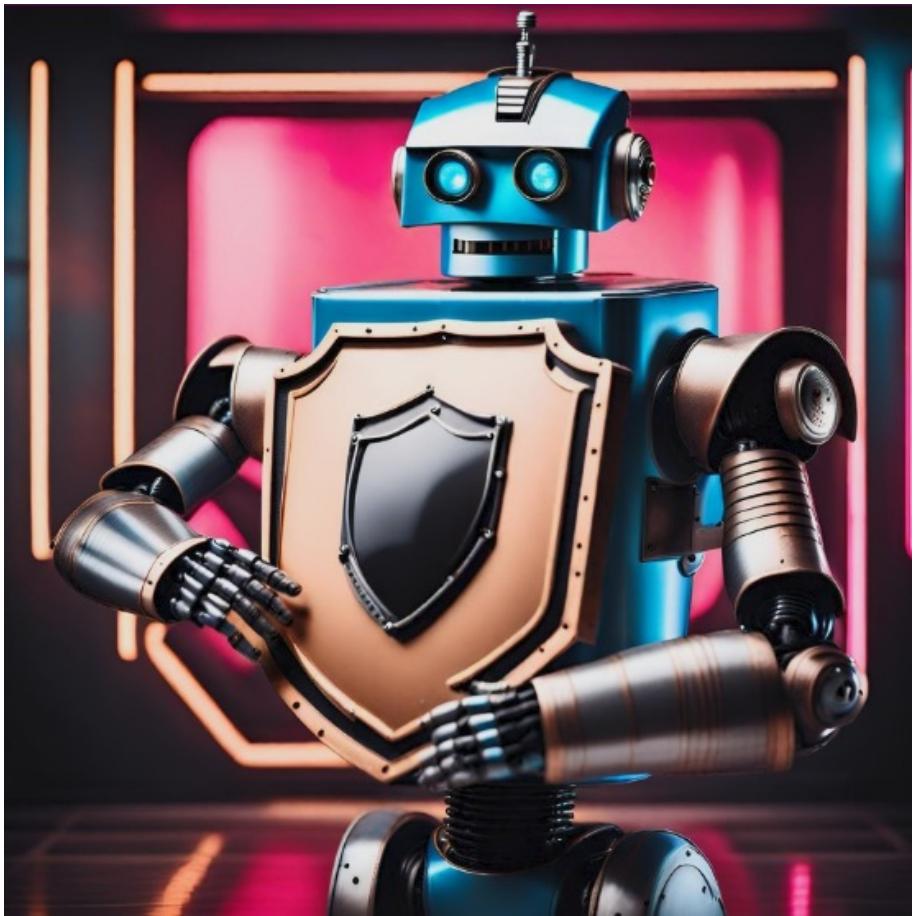
# Model Extraction - Logs

```
protoPayload.methodName = "storage.objects.get" AND (
    SEARCH(protoPayload.resourceName, ".pb") OR
    SEARCH(protoPayload.resourceName, ".mlmodel") OR
    SEARCH(protoPayload.resourceName, ".onnx") OR
    SEARCH(protoPayload.resourceName, ".pk1") OR
    SEARCH(protoPayload.resourceName, ".h5") OR
    SEARCH(protoPayload.resourceName, ".pmml")
)
```

The screenshot shows a log entry from the Cloud Audit Log interface. The log entry details a request for a storage object named 'model-67715116844491735' from a service account. The log includes fields such as insertId, logName, protoPayload (containing authentication and authorization info), and resourceLocation.

Field	Value
insertId	"1pvk7m9eliwt"
logName	"projects/coral-marker-414313/logs/claudaudit.googleapis.com%2Fdata_access"
protoPayload.authenticationInfo	{1}
protoPayload.authorizationInfo	[1]
protoPayload.methodName	"storage.objects.get"
protoPayload.requestMetadata	{4}
protoPayload.resourceLocation	{1}
protoPayload.resourceName	"projects/_/buckets/cloud-ai-platform-18195e29-682d-4d93-b7ee-e5c514f8eadd/objects/model-67715116844491735/saved-model/2024-04-19T14:17:29.126055Z/predict/001/saved_model.pb"
protoPayload.serviceName	"storage.googleapis.com"

# Configuration Guidance



# Azure ML

Home > Azure Machine Learning > Test-Workspace | Diagnostic settings >

## Diagnostic setting ...

Save Discard Delete Feedback

A diagnostic setting specifies a list of categories of platform logs and/or metrics that you want to collect from a resource, and one or more destinations that you would stream them to. Normal usage charges for the destination will occur. [Learn more about the different log categories and contents of those logs](#)

Diagnostic setting name \*

test

### Logs

Category groups ⓘ

allLogs

audit

### Categories

AmlComputeClusterEvent

AmlComputeClusterNodeEvent

AmlComputeJobEvent

AmlComputeCpuGpuUtilization

AmlRunStatusChangedEvent

ModelsChangeEvent

### Destination details

Send to Log Analytics workspace

Subscription

Azure subscription 1

Log Analytics workspace

testing-sentinel ( eastus )

Archive to a storage account

Stream to an event hub

Send to partner solution



## Asset Inventory

Manage ML asset inventory



## Logging

Enable audit logs and build detections from this research



## Threat Modeling

Include ML solutions in threat modeling exercises



## Data & Pipeline

### Security

Data and ML workflow security best practices

# BigML

Two-Factor Authentication

ACTIVATE TWO-FACTOR AUTHENTICATION  
Enhance your account security enabling Two-Factor Authentication

**Enable**

Two-Factor Auth doesn't apply when signing in with 3rd party auth providers (Amazon, Github, Google, ect.), only when you sign in with username/email and password.



## MFA

Enable MFA



## Credential Rotation

Regularly change passwords and API keys



## Granular Access

Apply granular access controls via alternative keys

# Vertex AI

Audit Logs [SET DEFAULT CONFIGURATION](#)

**Default configuration** [?](#) Admin Read:  
0 exempted principals  Disabled

**Data Access audit logs configuration**  
The effective data access configuration below combines the configuration for the currently selected

**Filter** [Vertex AI API](#)  OR  [Google Cloud Storage](#)  Enter property name

Service	Admin Read	Data Read	Data Write
Google Cloud Storage	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Vertex AI API	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>



## Access Control

Principal of least privilege



## IAM

Use IAM to manage roles and groups



## Logging

Enable audit logs and build detections from this research



## Security Command Center

Additional protection for Vertex AI

# Detection Guidance



# Azure ML

Attack Scenario	Log Schema(s)	Operation Name(s)
Data Poisoning	<b>AmlDataSetEvent</b>	<b>WORKSPACES/DATASETS/REGISTERED/WRITE</b>
Dataset Recon	<b>AmlDataSetEvent</b>	<b>WORKSPACES/DATASETS/REGISTERED/READ</b>
Model Recon	<b>AmlModelsEvent</b>	<b>WORKSPACES/MODELS/READ</b>
Model Extraction	<b>AmlModelsEvent</b>	<b>WORKSPACES/MODELS/READ</b>
	<b>StorageBlobLogs</b>	<b>GetBlob</b>
Data Extraction	<b>AmlDataSetEvent</b>	<b>WORKSPACES/DATASETS/REGISTERED/READ</b>
	<b>AmlDataStoreEvent</b>	<b>WORKSPACES/DATASTORES/READ</b>

# BigML

- Need private deployment to enable logging and develop detection rules
- We did not have access to private deployment for this research

# Vertex AI

Attack Scenario	Method Name(s)
Data Poisoning	<b>google.cloud.aiplatform.ui.DatasetService.UpdateDataset</b> <b>google.cloud.aiplatform.v1.DatasetService.UpdateDataset</b>
Dataset Recon	<b>google.cloud.aiplatform.v1.DatasetService.ListDatasets</b>
Model Recon	<b>google.cloud.aiplatform.v1.ModelService.ListModels</b>
Model Extraction	<b>storage.objects.get</b> <b>google.cloud.aiplatform.ui.ModelService.ExportModel</b> <b>google.cloud.aiplatform.v1.ModelService.ExportModel</b>
Data Extraction	<b>storage.objects.get</b>

# MLOKit

 MLOKit Usage  
Incident number 205

Owner: Unassigned | Status: New | Severity: High

Description:  
This rule will trigger when an operation against Azure Storage Blobs are conducted using MLOKit.

Alert product names:

- Microsoft Sentinel

Evidence:

 1 Events     1 Alerts     0 Bookmarks

Last update time: 05/08/24, 01:48 PM    Creation time: 05/08/24, 01:48 PM

Entities (0)

Tactics and techniques:

- Reconnaissance (0)
- Collection (0)



YARA Rule  
C# Project  
GUID



Snort Rule  
Hardcoded user agent string

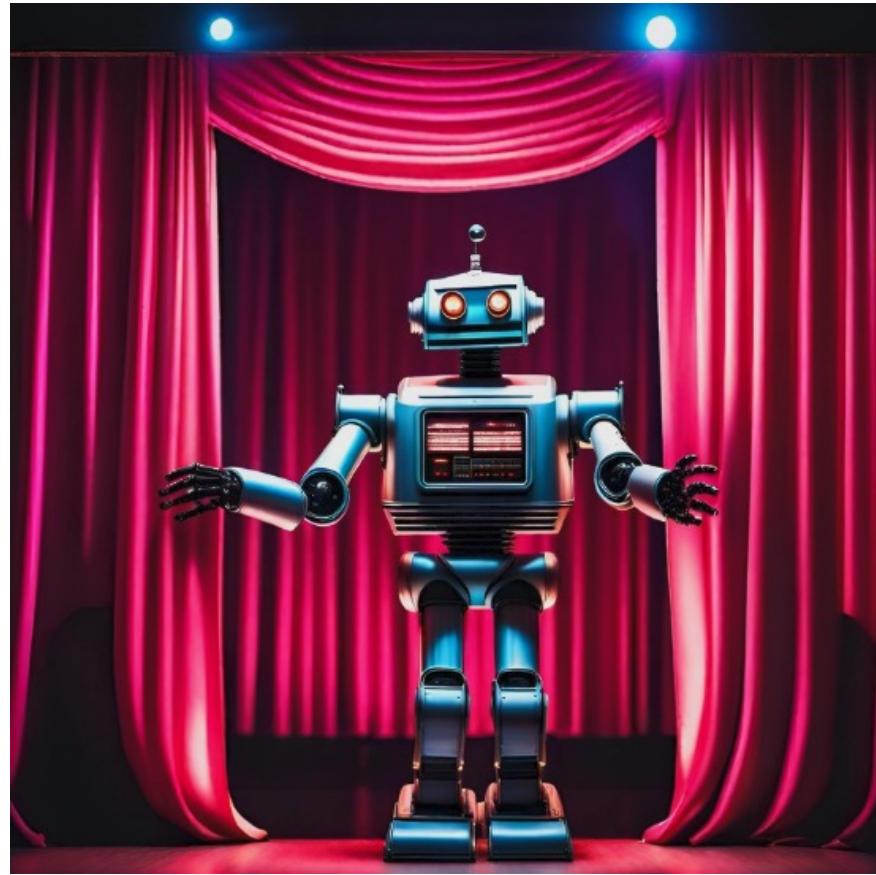


Google Cloud Alert Policy  
Actions against Google Cloud Storage



Sentinel Rules  
Actions against Azure Storage Blobs

# Conclusion



# Conclusion

01

Securing MLOps platforms and personnel is critical

02

Simulate attacks against MLOps platforms and build detections

03

Use of MLOps platforms will continue to grow as will attackers targeting them

# Questions?



Brett Hawkins  
[@h4wkst3r](https://twitter.com/h4wkst3r)  

Chris Thompson  
[@retBandit](https://twitter.com/retBandit)  

Whitepaper -



MLOKit Tool -





# Announcing the First Offensive AI Conference

Offensive AI Con is a cutting-edge, highly technical conference dedicated to exploring the use of AI within agentic frameworks for Cyber Network Attack (CNA), Cyber Network Exploitation (CNE), and Cyber Network Operations (CNO). The event will also delve into AI-assisted & autonomous vulnerability research and the development of offensive cyber capabilities, offering a platform for advanced discussion on the future of AI-driven cyber offense.

The topics will include:

- Agentic frameworks for offensive use
- Offensive model eval
- ML-assisted reversing, fuzzing, vulnerability research, and analysis
- ML-assisted weaponization & exploit development
- ML-based cyber effects development
- ML-assisted Offensive Security Tool (OST) & capability development
- ML-assisted target identification, analysis, tasking & targeting
- AI-based security testing, red teaming and software evaluation

# Thank you



# References

- <https://arxiv.org/pdf/1609.02943.pdf%20/>
- <https://twitter.com/WHITEHACKSEC>
- <https://twitter.com/mairebear>
- <https://www.blackhat.com/asia-24/briefings/schedule/index.html#confused-learning-supply-chain-attacks-through-machine-learning-models-37794>
- <https://www.trendmicro.com/vinfo/gb/security/news/cybercrime-and-digital-threats/uncovering-silent-threats-in-azure-machine-learning-service-part-I>
- <https://msrc.microsoft.com/update-guide/vulnerability/CVE-2023-23382>
- <https://www.blackhat.com/us-23/briefings/schedule/#uncovering-azures-silent-threats-a-journey-into-cloud-vulnerabilities-33073>
- <https://www.oligo.security/blog/shadowray-attack-ai-workloads-actively-exploited-in-the-wild>
- <https://neptune.ai/blog/mlops-tools-platforms-landscape>
- <https://aws.amazon.com/sagemaker/>
- <https://azure.microsoft.com/en-us/products/machine-learning>

# References

- <https://bigml.com/>
- <https://cloud.google.com/vertex-ai?hl=en>
- <https://www.qwak.com/>
- <https://www.dominodatalab.com/product/domino-enterprise-mlops-platform>
- <https://www.databricks.com/>
- <https://www.datarobot.com/platform/mlops/>
- <https://wandb.ai/site>
- <https://valohai.com/product/>
- <http://www.truefoundry.com/>
- <https://www.darkreading.com/application-security/hugging-face-ai-platform-100-malicious-code-execution-models>
- <https://huggingface.co/>
- <https://hiddenlayer.com/research/weaponizing-machine-learning-models-with-ransomware/#Overview-of-ML-Model-Serialization-Formats>

# References

- <https://docs.python.org/3/library/pickle.html>
- <https://learn.microsoft.com/en-us/azure/machine-learning/concept-workspace?view=azureml-api-2>
- <https://aadinternals.com/post/phishing/>
- <https://learn.microsoft.com/en-us/azure/machine-learning/azure-machine-learning-glossary?view=azureml-api-2>
- <https://learn.microsoft.com/en-us/azure/machine-learning/how-to-setup-authentication?view=azureml-api-2&tabs=sdk>
- <https://learn.microsoft.com/en-us/cli/azure/>
- <https://learn.microsoft.com/en-us/cli/azure/ml?view=azure-cli-latest>
- <https://learn.microsoft.com/en-us/rest/api/azureml/?view=rest-azureml-2023-10-01>
- <https://learn.microsoft.com/en-us/azure/machine-learning/how-to-assign-roles?view=azureml-api-2&tabs=labeler>
- <https://learn.microsoft.com/en-us/azure/ai-services/diagnostic-logging>
- <https://learn.microsoft.com/en-us/azure/azure-monitor/logs/log-analytics-workspace-overview>
- <https://learn.microsoft.com/en-us/cli/azure/ml?view=azure-cli-latest>

# References

- <https://learn.microsoft.com/en-us/azure/machine-learning/monitor-azure-machine-learning-reference?view=azureml-api-2>
- <https://medium.com/@maziarizadi/pickle-your-model-in-python-2bbe7dba2bbb>
- <https://curl.se/docs/manpage.html>
- <https://learn.microsoft.com/en-us/rest/api/azureml/workspaces?view=rest-azureml-2023-10-01>
- <https://github.com/xforceder/MLOKit>
- <https://learn.microsoft.com/en-us/rest/api/azureml/model-containers?view=rest-azureml-2023-10-01>
- <https://learn.microsoft.com/en-us/azure/ai-services/document-intelligence/create-sas-tokens?view=doc-intel-4.0.0>
- <https://bigml.com/customers/>
- <https://bigml.com/api/organizations>
- <https://bigml.com/api/>
- <https://bigmler.readthedocs.io/en/latest/>
- <https://bigml.com/api/quickstart>
- <https://bigml.com/api/sources>

# References

- <https://bigml.com/api/projects>
- <https://bigml.com/api/datasets>
- <https://bigml.com/api/models>
- <https://dmg.org/pmm/pmm-v4-1.html>
- <https://cloud.google.com/vertex-ai/docs/glossary>
- <https://cloud.google.com/vertex-ai/docs/workbench/reference/authentication>
- <https://cloud.google.com/docs/authentication/rest>
- <https://cloud.google.com/docs/authentication/rest#user-creds>
- <https://cloud.google.com/docs/authentication/rest#rest-request>
- <https://cloud.google.com/docs/authentication/rest#impersonated-sa>
- <https://cloud.google.com/docs/authentication/rest#metadata-server>
- <https://cloud.google.com/vertex-ai/docs/general/access-control>
- <https://cloud.google.com/vertex-ai/docs/general/audit-logging>

# References

- <https://cloud.google.com/logging/docs/view/logs-explorer-interface>
- <https://cloud.google.com/storage/docs/buckets>
- <https://cloud.google.com/logging/docs/view/building-queries>
- <https://towardsdatascience.com/guide-to-file-formats-for-machine-learning-columnar-training-inferencing-and-the-feature-store-2e0c3d18d4f9>
- <https://cloud.google.com/apis/docs/overview>
- <https://cloud.google.com/resource-manager/reference/rest/v1/projects>
- <https://cloud.google.com/vertex-ai/docs/reference/rest/v1/projects.locations.datasets>
- [https://cloud.google.com/storage/docs/json\\_api](https://cloud.google.com/storage/docs/json_api)
- <https://cloud.google.com/vertex-ai/docs/reference/rest/v1/projects.locations.models>
- <https://learn.microsoft.com/en-us/azure/machine-learning/concept-enterprise-security?view=azureml-api-2>
- <https://learn.microsoft.com/en-us/security/benchmark/azure/baselines/machine-learning-service-security-baseline>

# References

- <https://techcommunity.microsoft.com/t5/fasttrack-for-azure/six-security-considerations-for-machine-learning-solutions/ba-p/3718592>
- <https://blog.bigml.com/2013/05/03/alternative-keys-fine-grained-rest-api-access-to-your-machine-learning-resources/>
- [https://www.linkedin.com/pulse/secure-your-vertex-ai-workbench-enterprise-machine-learning-curtis-kbmof?trk=articles\\_directory](https://www.linkedin.com/pulse/secure-your-vertex-ai-workbench-enterprise-machine-learning-curtis-kbmof?trk=articles_directory)
- <https://cloud.google.com/blog/products/identity-security/introducing-security-command-center-protection-for-vertex-ai>
- <https://learn.microsoft.com/en-us/azure/data-explorer/kusto/query/>
- <https://learn.microsoft.com/en-us/azure/sentinel/detect-threats-custom?tabs=azure-portal>
- <https://learn.microsoft.com/en-us/azure/storage/blobs/monitor-blob-storage?tabs=azure-portal>
- <https://learn.microsoft.com/en-us/azure/azure-monitor/reference/tables/storagebloblogs>
- <https://cloud.google.com/monitoring?hl=en>
- <https://cloud.google.com/monitoring/alerts>
- <https://yara.readthedocs.io/en/stable/writingrules.html>
- <https://snort.org/>