Mastering AI Security Boot Camp

TTAI2820: Hands-on AI Security: Essentials, Threat Detection, Vulnerabilities, Forensics, Incident Response & Future Trends

Trivera Technologies www.triveratech.com

20240429

Experience is

Jumping right In…

- Welcome!
  - **Mastering AI Security Boot Camp (TTAI2820)**
  - Geared for technical professionals eager to deepen their knowledge in machine learning and AI security. Roles include Data Scientists, Machine Learning Engineers, IT Security Professionals, and DataOps Engineers or similar.
  - Topics, labs and agenda may adjust during delivery based on your interests, roles and goals.
- Hours:
  - 10:00 to 6:00 PM Eastern; One Hour for Lunch; A few breaks as needed
- A Bit About Me: Dr. Ernesto Lee, Dr.Lee@triveratech.com
  - Chief Innovation Officer, Trivera Technologies www.triveratech.com
- A Bit About You:
  - What's your role / day to day?
  - Are you working with these skills already?
  - What kinds of related things are you working on?
  - What are you most excited to learn about in this class?

Teaming for Success

- Course Portal: Trivera's SkillJourneys LXP www.skilljourneys.com
  - Quick Look at the Learning Experience Platform / Course Portal
  - Where to find the Courseware: Course Guide, Deck & Resources
  - Feedback Surveys
  - Access is live for 60 days
- Sharing Feedback – We're Here to Provide Value!
  - Feedback is welcome & always encouraged
  - Real time is best
  - Other ways to connect
  - Course Check In & End of Course feedbacks – complete right in the LXP
- Course Recordings
  - Provided by separate link a few days after class; Live for 60 days
- Course Certificates
  - Will be sent out a few days after class after End of course survey is completed.

Agenda Review

1. Introduction to AI in Security:
   - Explore foundational AI security, threat identification, and protective strategies through practical examples.
2. Playing Detective:
   - Explores AI system vulnerabilities, different threat types, and data privacy concerns.
3. Building the AI Fortress: Defense Mechanisms 101
   - Teaches design and implementation of robust AI-driven defense systems..
4. CSI Cyber: Exploring AI Forensics
   - Focuses on applying forensic techniques and analyzing AI security incidents.

Agenda

5. AI Adversarial Attacks and Defenses:
   - Covers strategies to tackle adversarial threats to AI systems.
6. Crisis Averted: Crafting Your AI Incident Response Plan:
   - Develop and execute effective incident response plans for AI system breaches.
7. AI Privacy and Ethical Considerations:
   - Addresses privacy risks and ethical considerations in AI applications.
8. What's Next? Preparing for Future AI Security Challenges:
   - Explore future AI security trends and prepare for emerging threats like deepfakes.

Additional Resources

These Resources are in the back of your Course Guide

- Course Site References & Additional Information
- Glossary of Main Terms, Skills and Key Topics
- Next Steps, Follow on Courses & SkillJourneys

Getting Hands-On

- Demos & Activities
  - We'll focus activities on things that will be useful to you and provide value.
  - *ADD A few sentences about what the demos will show*

TriveraTech®
TECHNOLOGY TRAINING

Any Questions?

Let's Dive In!

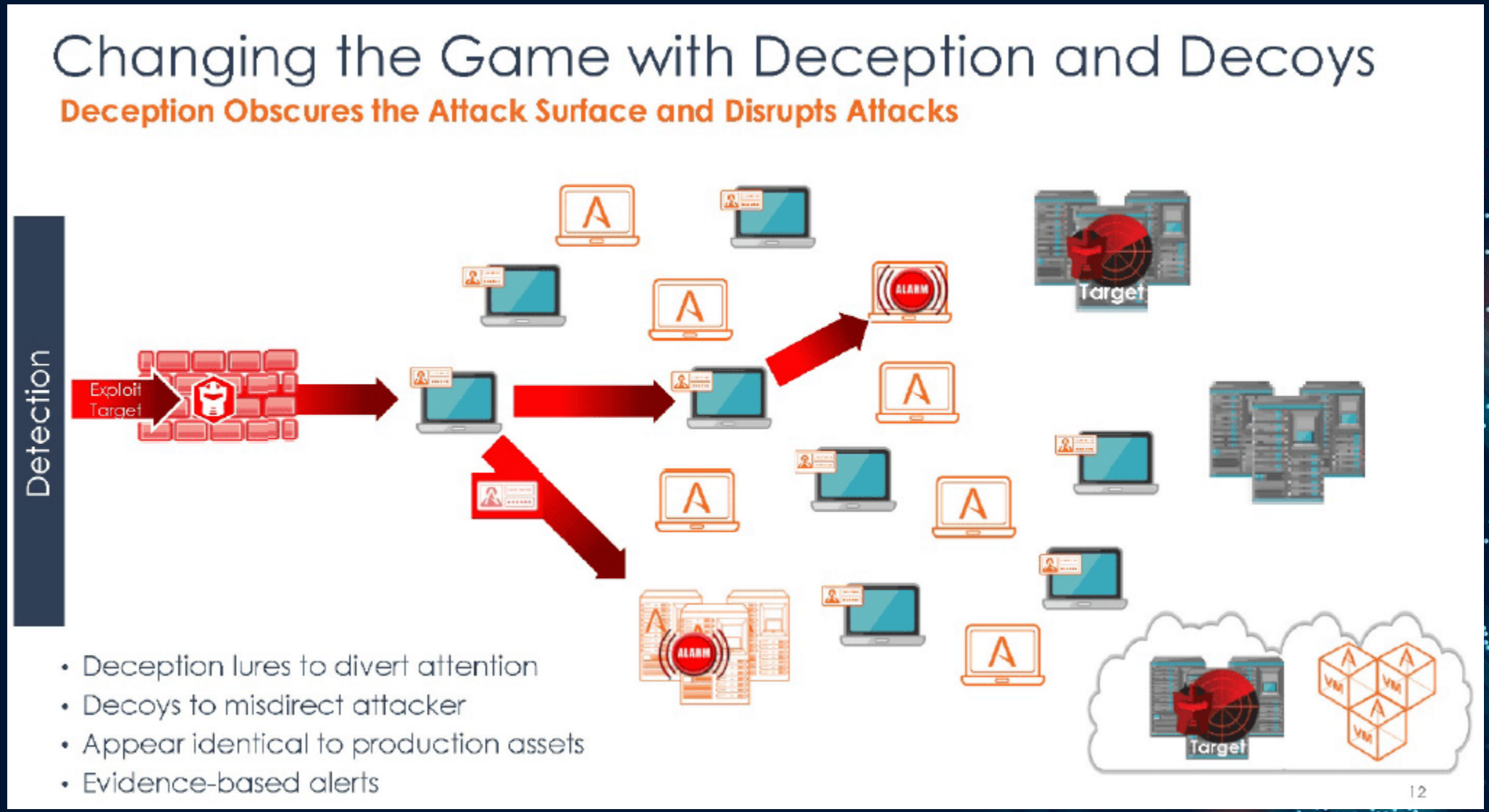
Experience is

Chapter 1:

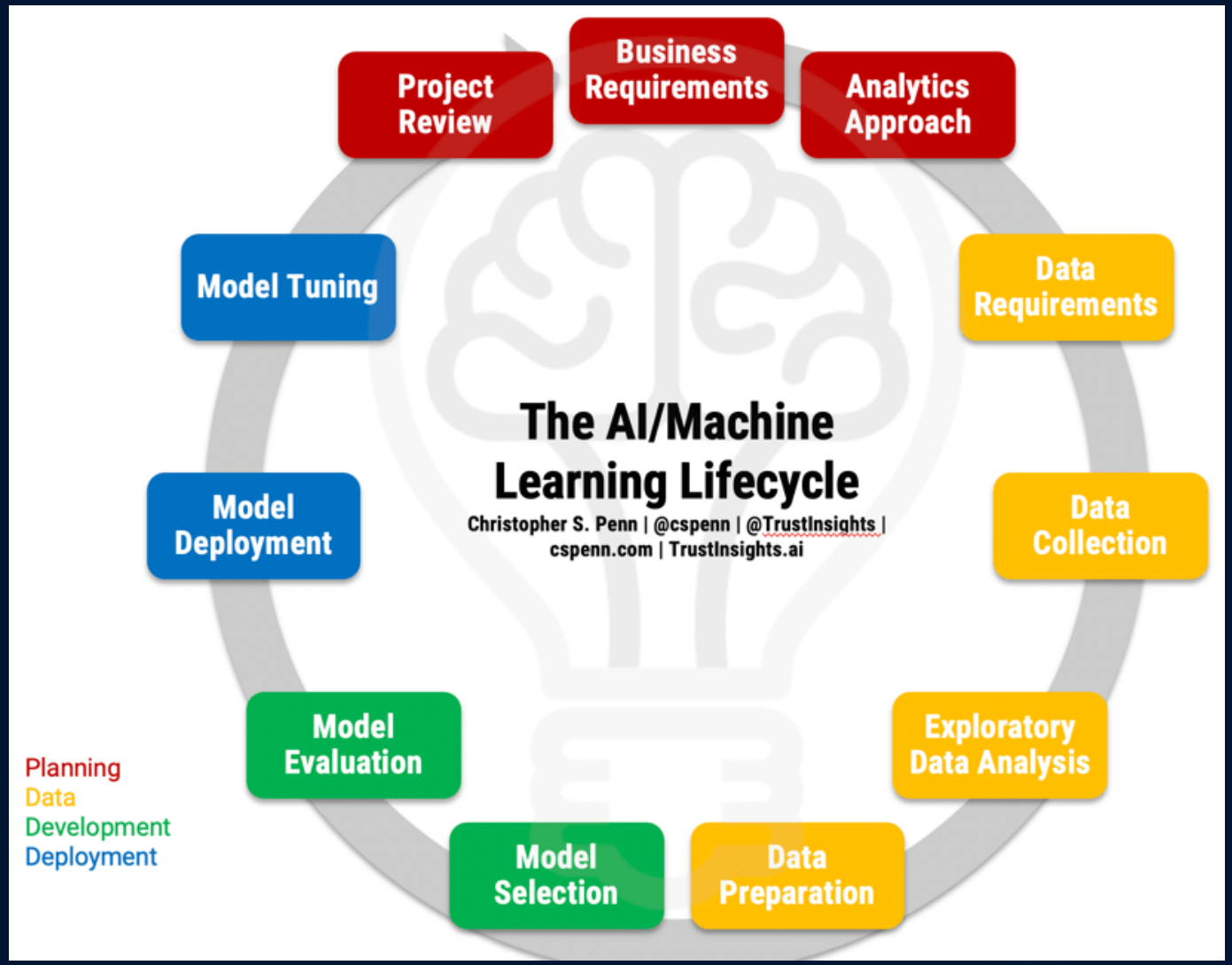
Introduction to AI & Security

Explore foundational AI security, threat identification, and protective strategies

Defining Machine Learning Security



A Picture of ML



ML Only Works When…

- The source data is untainted
- The training and testing data are unbiased
- The correct algorithms are selected
- The model is created correctly
- Any goals are clearly defined and verified against the training and test data

Identifying the ML Security Domain

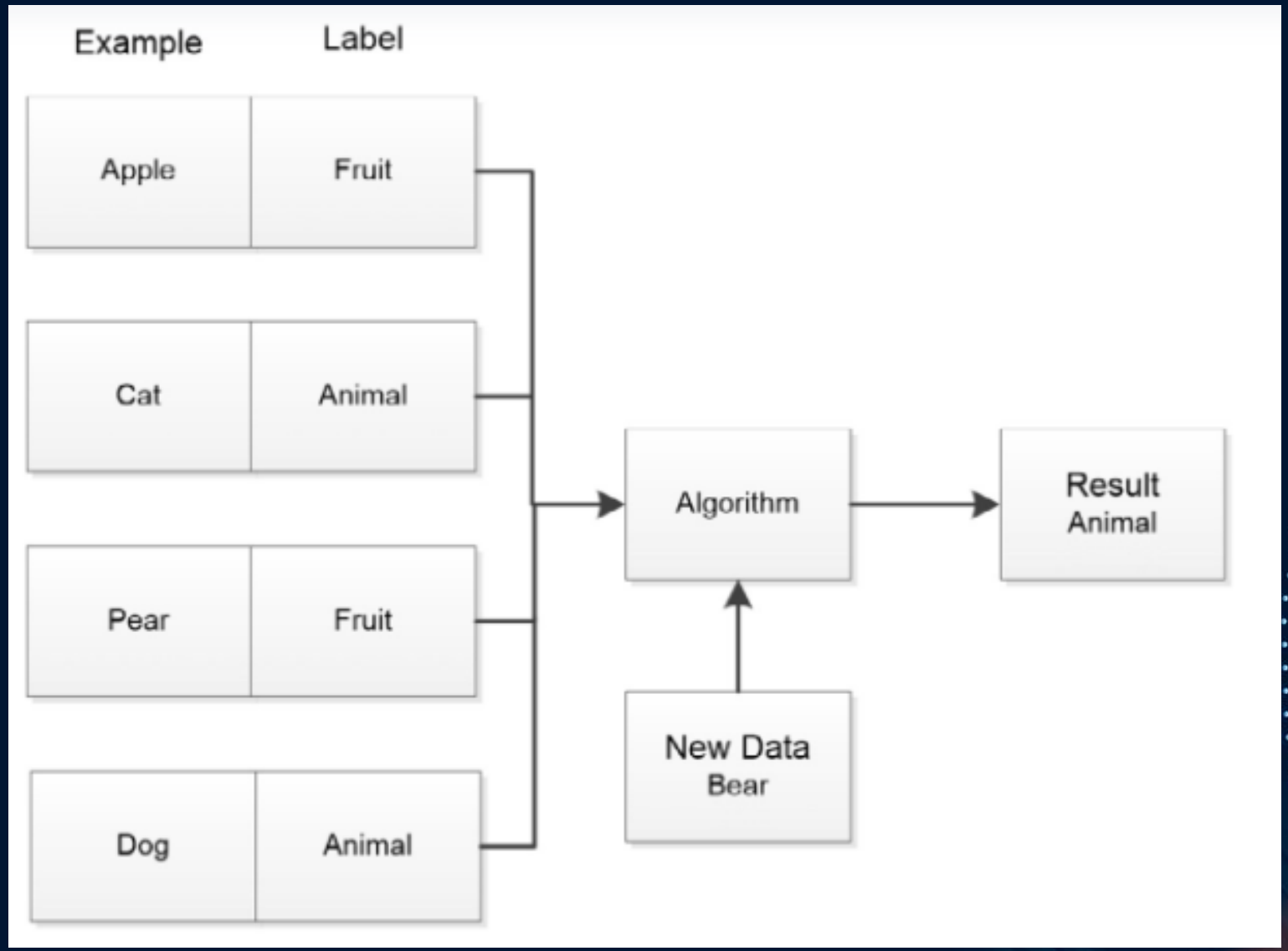
- Data Bias
- Data Corruption
- Missing critical data
- Errors in Data
- Algo correctness
- Algorithmic Bias
- Repeatable Results

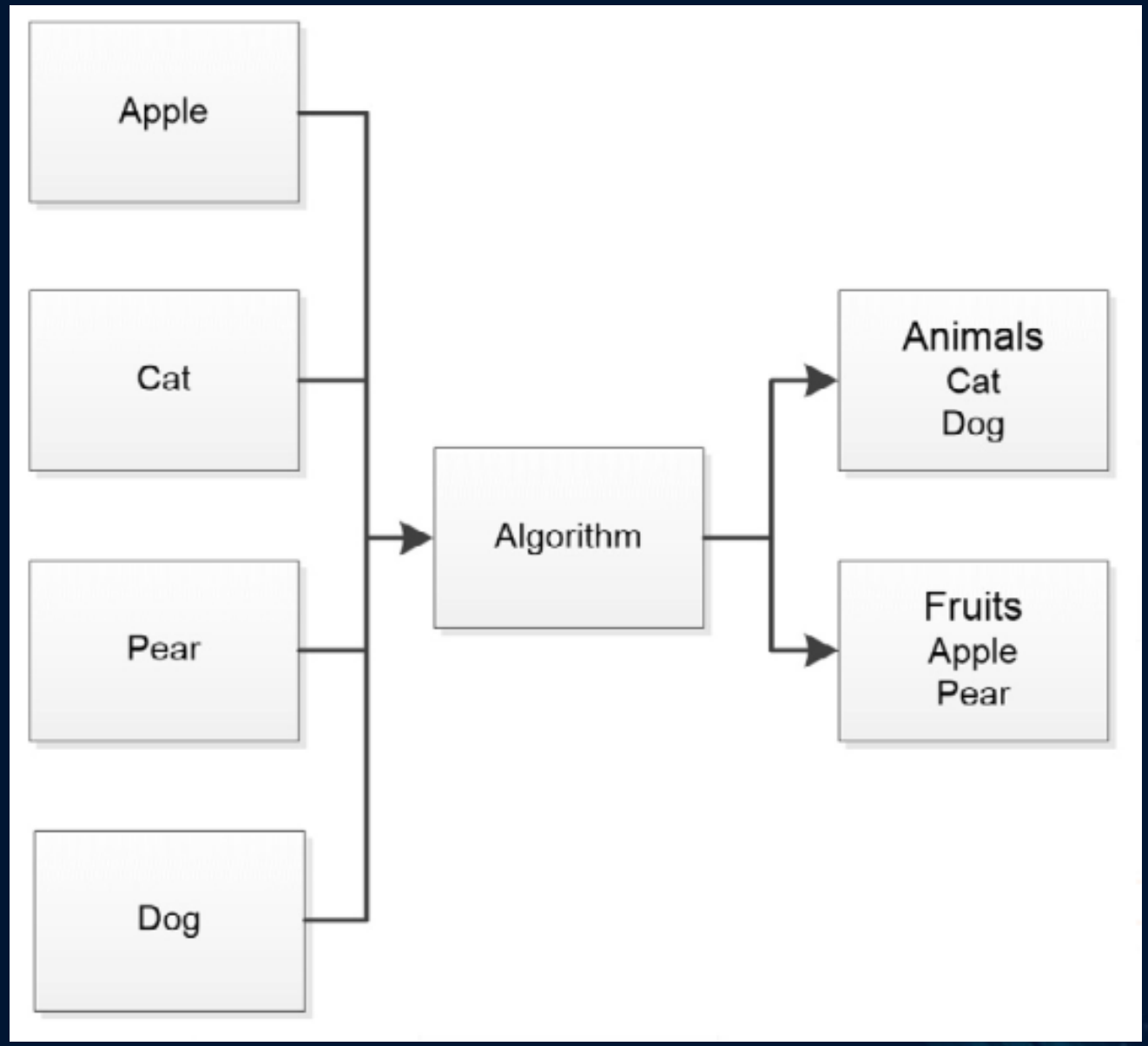
Vulnerabilities

- Evasion
- Poisoning
- Inference
- Trojans
- Backdoors
- Espionage
- Sabotage
- Fraud

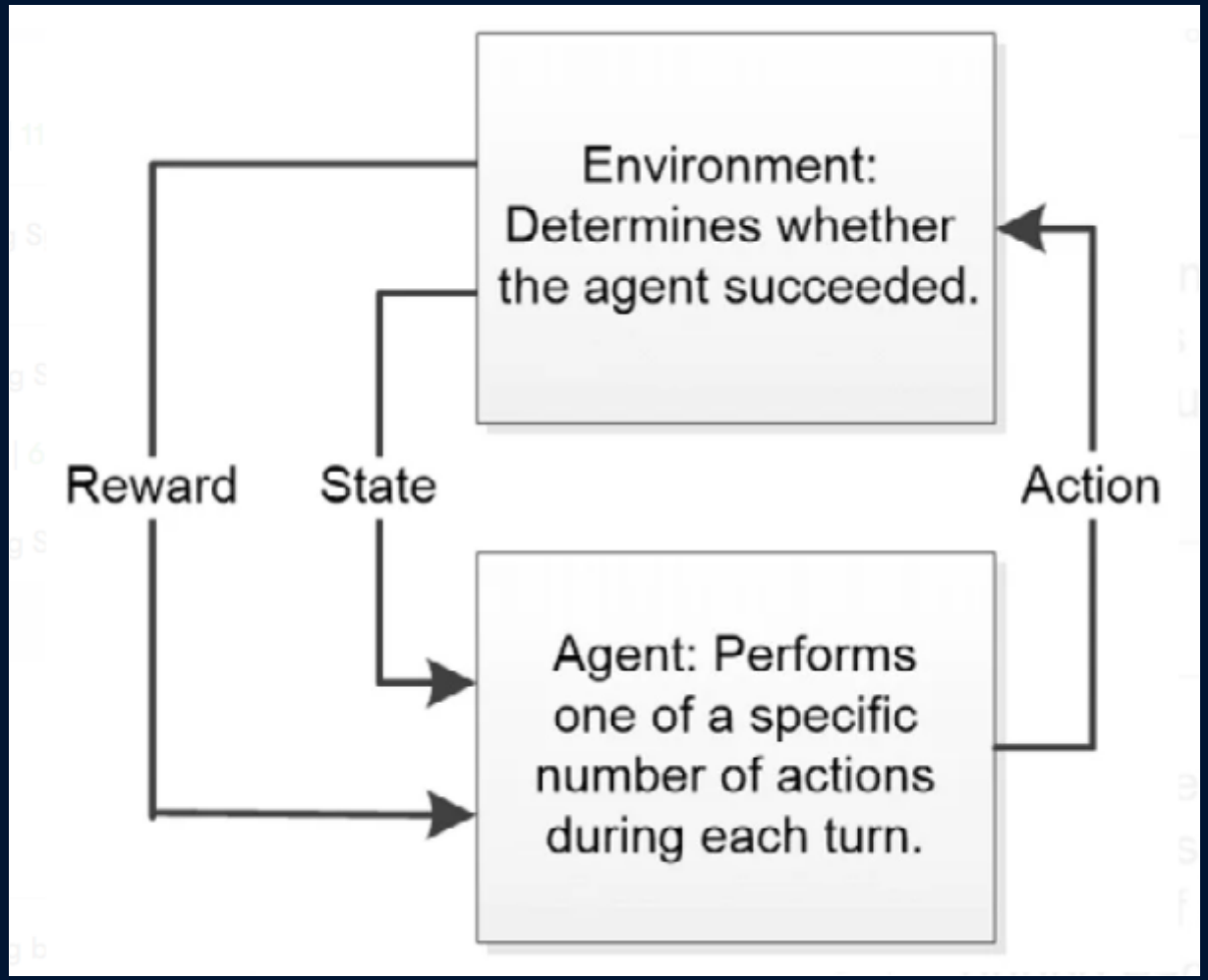
Types of ML

- Supervised
- Unsupervised
- Reinforcement





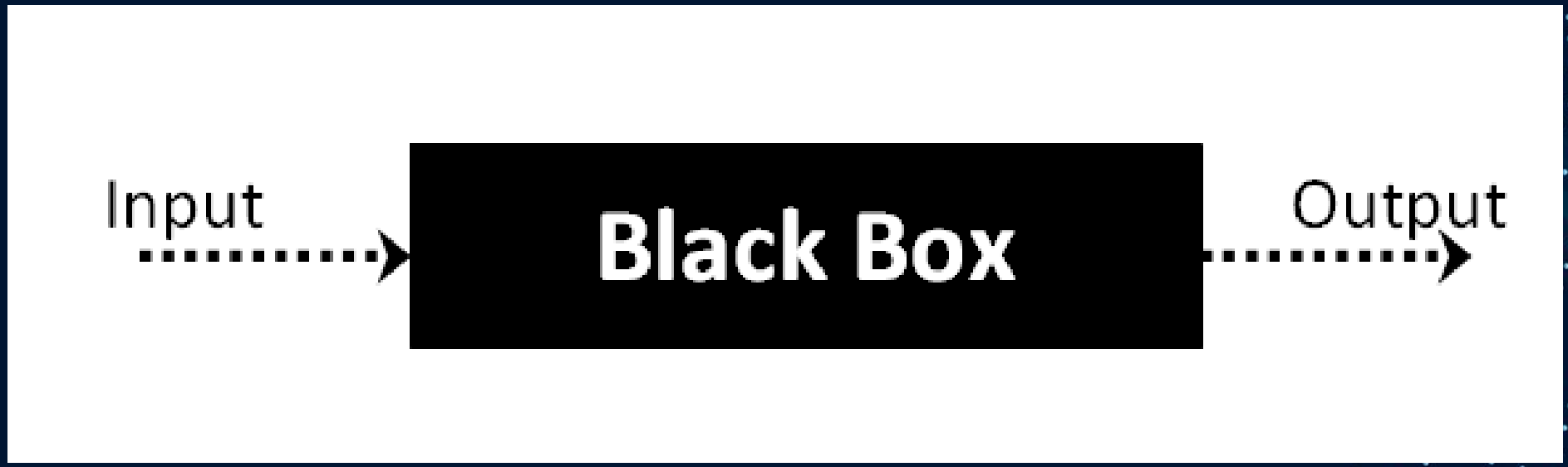
Reinforcement Learning



Add Security to ML

- Commission
- Omission
- Bias
- Perspective
- Frame of Reference

Compromising the integrity and availability of ML



Types of Attacks against ML

- https://portswigger.net/daily-swig/vulnerabilities
- Adversarial Attacks
- ML relies on Statistics!

What Can be Achieved with ML Security

- Set understandable and achievable result goals that are verifiable, consistent, and answer specific needs
- Train personnel (which means everyone in the organization, along with consultants and third parties) to interact with the application and its data appropriately
- Ensure that data passes all of the requirements for proper format, lack of missing elements, absence of bias, and lack of various forms of corruption
- Choose algorithms that actually perform tasks in a manner that will match the goals set for the ML application
- Use training techniques that create a reliable model that won't overfit or underfit the data
- Perform testing that validates the data, algorithms, and models used for the ML application
- Verify the resulting application using real-world data that the ML application hasn't seen in the past

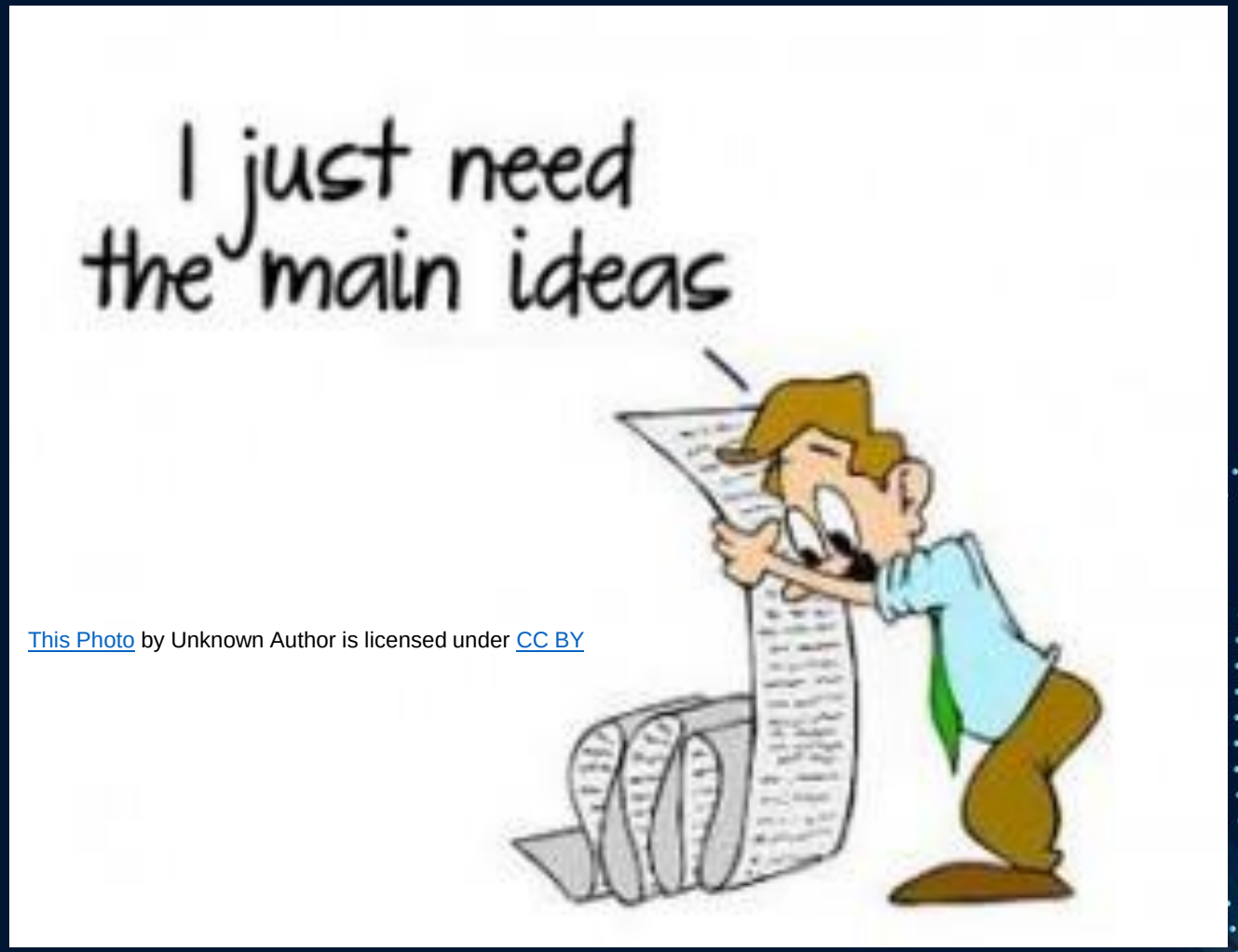
Setup for the class

- Google Colab
- To a lessor extent Microsoft Azure

What do you need to know?

- DEMO

Summary



TriveraTech®
TECHNOLOGY TRAINING

Lab: Validate Colab

Hands-on Lab: Please refer to your Lab Guide
and follow the instructions provided by your Instructor

Experience is

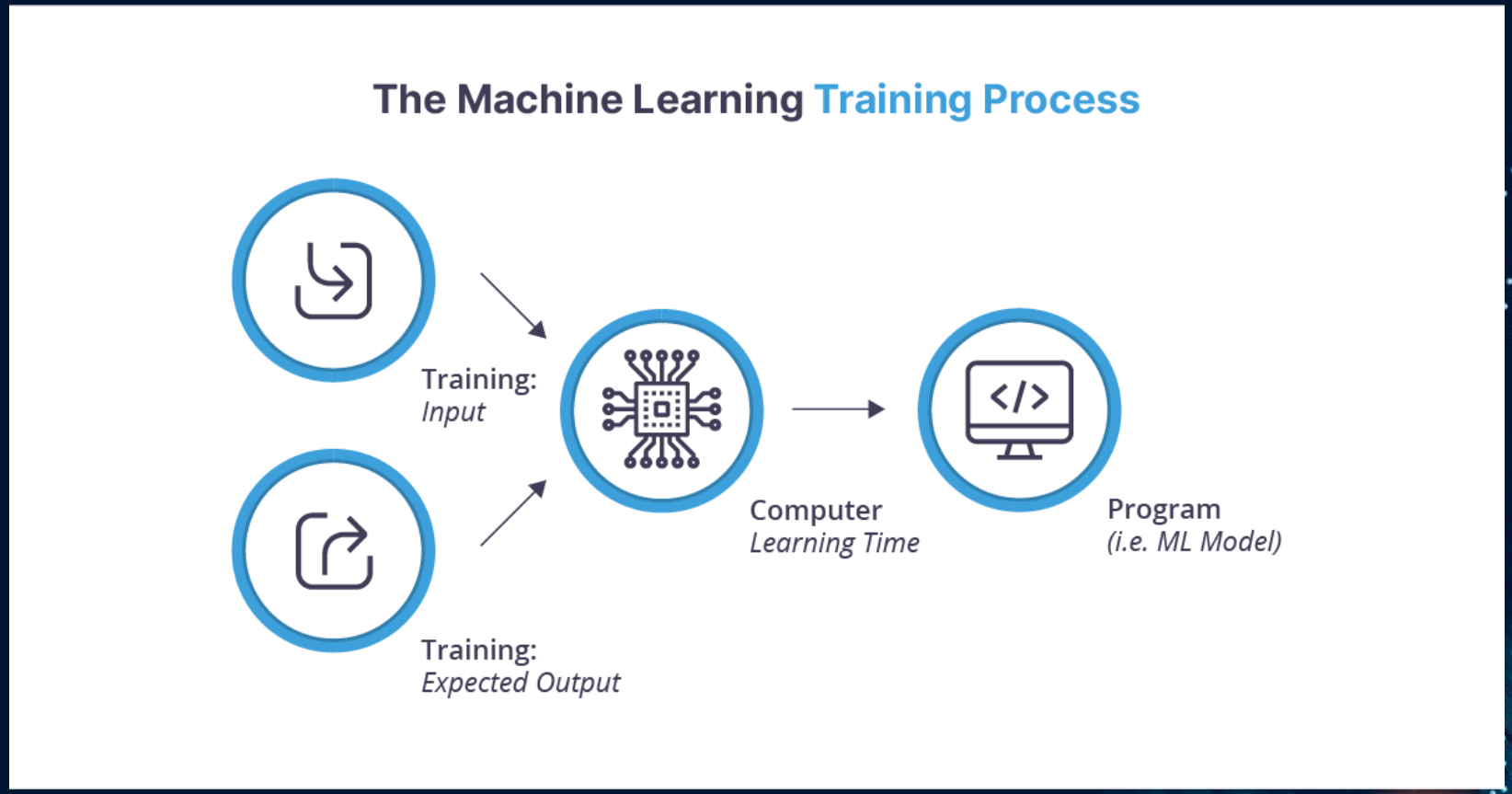
Chapter 2:

Playing Detective

Identify Threats and Vulnerabilities

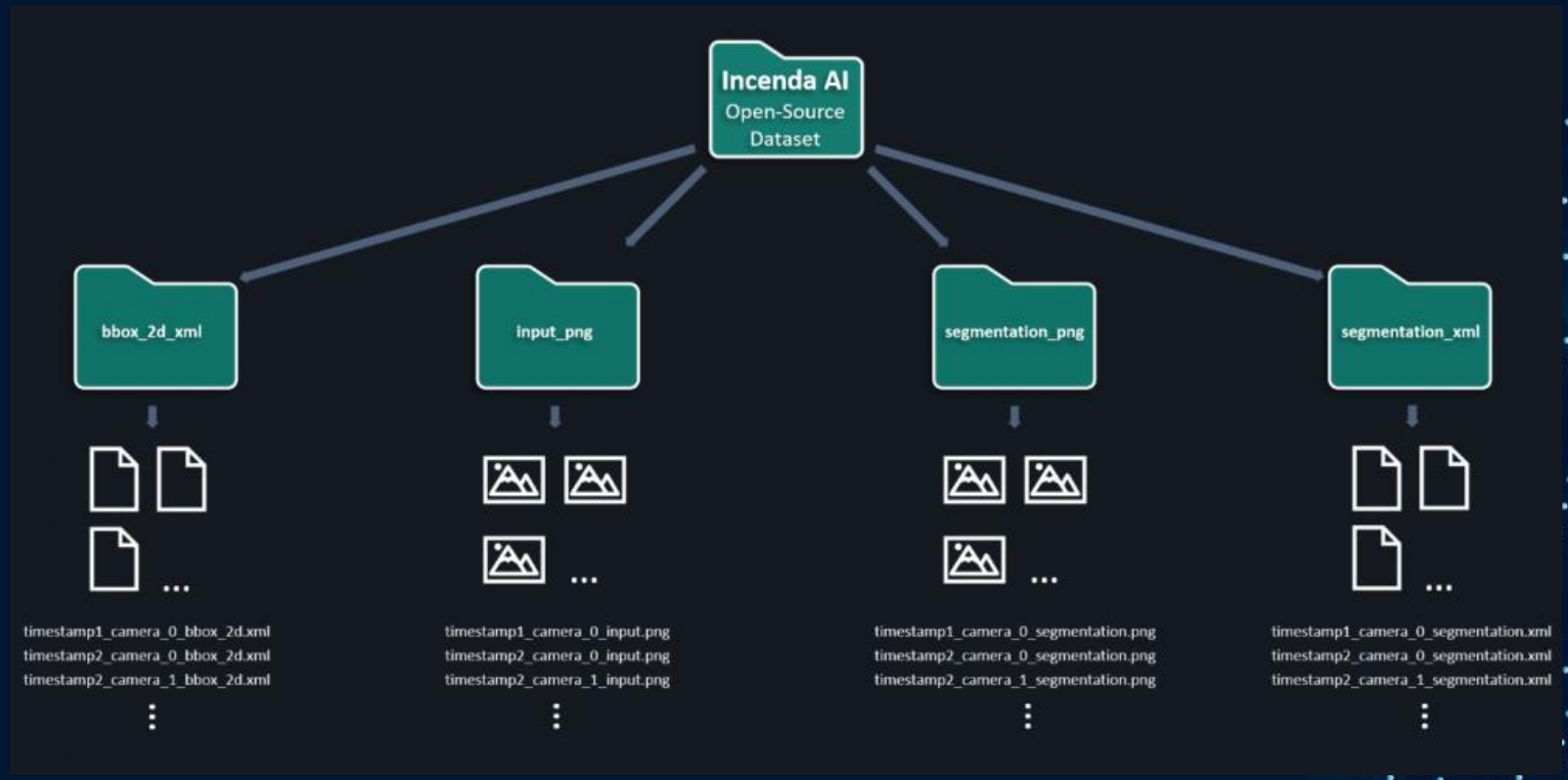
Threats at Training – Dataset vulnerabilities

- Defining dataset threats
- Detecting dataset modification
- Mitigating dataset corruption



Dataset Threats

- Dataset Modification
- Dataset Corruption



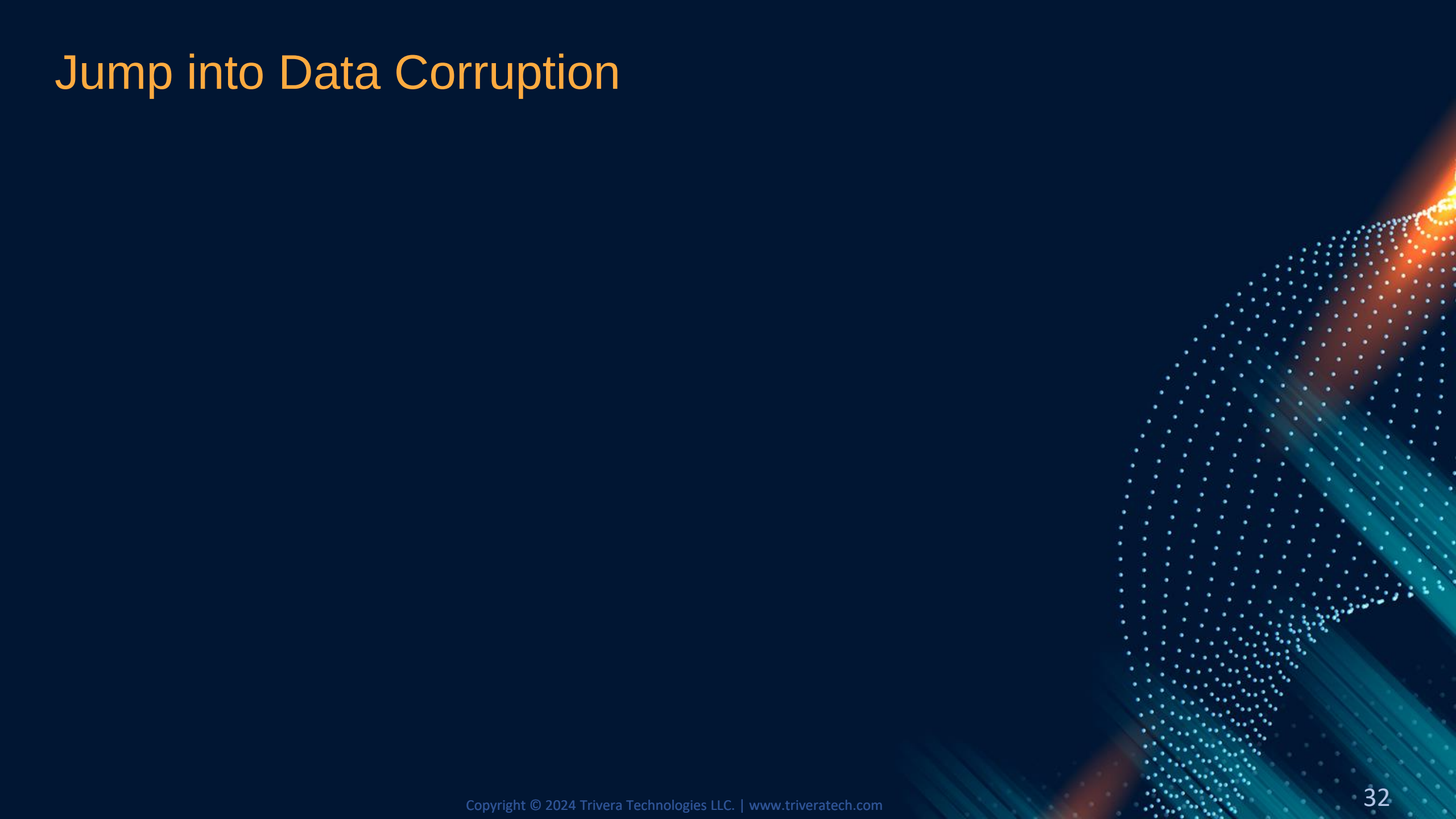
Dataset Threat Sources

Task	Learning Type	ML Consideration
Automatic language translation	Supervised	Translates one language into another language using a sequence-to-sequence learning algorithm. The results are often less useful than expected due to variations between languages and the fact that languages generally contain words that don't have equivalents in other languages. Susceptible to data errors, missing data, data corruption, algorithm bias, and an inability to repeat and verify results due to naturally occurring evolution in languages. This kind of application is also sensitive to speech patterns and misidentifying terms when words aren't enunciated clearly.
Email spam and malware filtering	Supervised	Marks, moves, or deletes email that meets the criteria of spam or malware from an inbox as it's received from a server. There are usually several levels of filtering including Content, Header, Blacklist, Rule-based, and Permission. Susceptible to a number of potential attacks including backdoors, Trojans, espionage, sabotage, fraud, evasion, inference, data errors, and data corruption. This is one of the more reliable forms of ML applications, but users still regularly find spam in their inboxes and useful messages in their spam folders.
Image recognition	Supervised	Identification of objects, persons, places, patterns, and other elements within an image. Susceptible to a variety of attack types, but also prone to misidentification when the image contains elements the application didn't expect or when those objects appear in positions that the application isn't trained to recognize.
Medical diagnosis	Supervised and unsupervised	Predicts the progression and characteristics of diseases and other conditions, along with locating and identifying potential patient illnesses. Susceptible to data bias, data corruption, data errors, incorrect algorithm selection, and algorithm bias. This particular application type can never operate alone; it always assists a physician with the required experience to make a diagnosis.
Online fraud detection	Supervised	Reduces the risk of conducting transactions online by detecting conditions such as fake accounts, fake IDs, compromised sites, compromised security certificates, and so on. Susceptible to a wide range of attacks, some of which have nothing to do with the application. For example, a compromised certificate authority could cause the application to fail by allowing the hacker access to the underlying infrastructure, even if the application itself isn't at fault. This kind of application is also known to display false positives and false negatives depending on the reliability of the code used to create it and the model training.
Product recommendation	Unsupervised	Outputs product recommendations based on previous buying habits, associated goods, and direct queries. It's one of the most widely used and common ML applications. Susceptible to data errors, data bias, missing data, algorithm bias, fraud, sabotage, and a wealth of other issues. This kind of application often provides irrelevant information along with useful product recommendations because the application has no method of judging user needs and wants.
Self-driving cars	Supervised, unsupervised, and reinforcement	Allows a vehicle to drive itself by m various cameras and detectors for the of obstacles, interpreting the conten signs, and so on. Susceptible to so many different kinds it's truly amazing that self-driving at all. In addition to ML, self-drivin rely on other AI technologies such systems (https://www.aitr com/ai-insider/expert-sy ai-self-driving-cars-cr innovative-techniques/). It possible that self-driving cars will e become completely successful, but d for this advance anytime soon.
Speech recognition	Supervised	Translation of spoken or written speech i that the computer can recognize and Susceptible to data errors and use of un terms. This kind of application is also s speech patterns and misidentifying te words aren't enunciated clearly.
Stock market trading	Supervised	Predicts trends in the stock market past and current data. This is one of th applications that relies heavily on sh memory and weighting processes to ma data count for more than past data. Susceptible to data bias, data corruptio data, data errors, incorrect algorithm and algorithm bias. Attackers will a gain access by any means possible wit emphasis on evasion, inference, Tro backdoors. Reliability is a prime conce application type, but incredibly hard t given the variability of the stock mark

Jump Into Data Exchange

- Automated software makes an unwanted update to a value
- Company policy or procedure changes so that the value that used to be correct is no longer correct
- Aging and archiving software automatically removes values that are deemed too old, even when they aren’t
- New sensors report data using a different range, format, or method that creates a data misalignment
- Someone changes the wrong record

Jump into Data Corruption



Discover Feature Manipulation

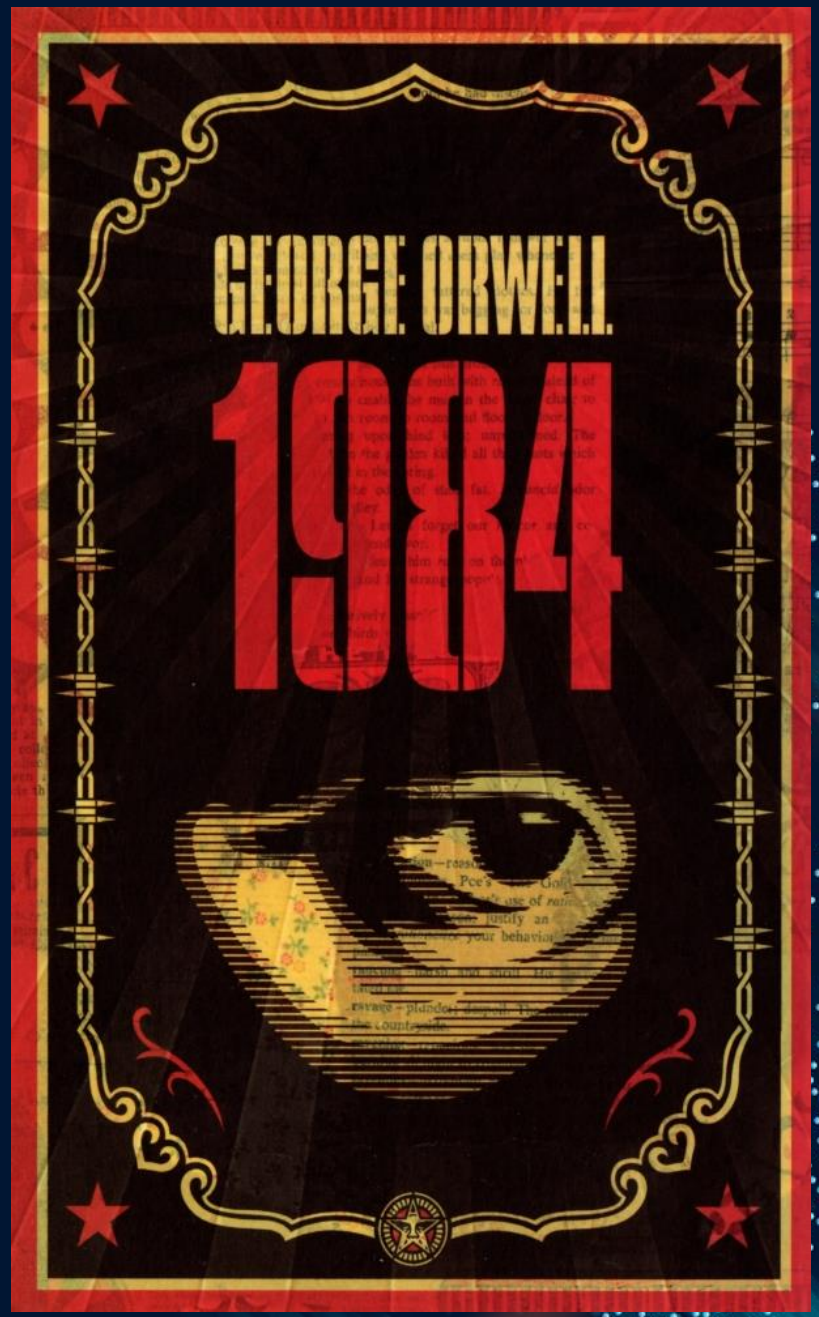
- Keep personal data out of the dataset when possible
- Use aggregate values where it's difficult to reconstruct the original value, but the aggregate still provides useful information
- Perform best practices feature reduction studies to determine whether a feature really is needed for a calculation

Source Modification

- Source modification attacks occur when a hacker successfully modifies a data source you rely on for input to your model.
- It doesn’t matter how you use the data, but rather how the attacker modifies the site.

Thwarting Privacy Attacks

- Membership inference attack
- GAN
- Language Generation Models
- Federated ML System
- Aggregate location data
- Data extraction
- Genomic information
- Facial Recognition
- Unintended Memory
- Model Extraction

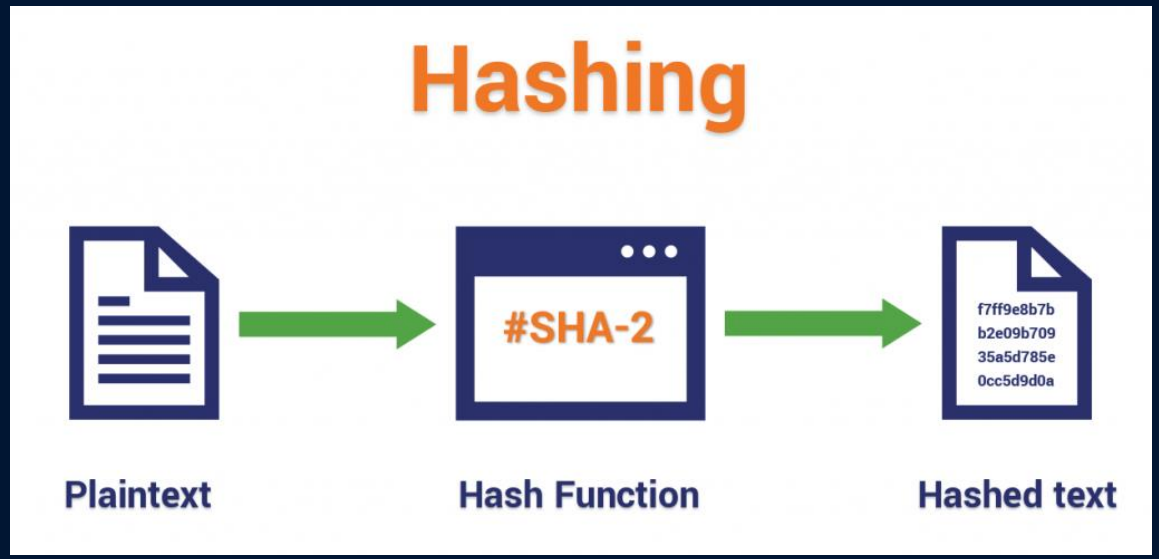


Detecting Dataset Modification

1. Hackers want to create an environment where products from Organization A, a competitor of Organization B, receive better placement on a sales site because the competitor is paying them to do so
2. The hackers discover that buyer product reviews and their product ratings are directly associated with the site's ranking mechanism
3. The hackers employ zombie systems (computers they have taken over) to upload copious reviews to the site giving Organization B's products a one-star review
4. The site's ML application begins to bring down the product rankings for Organization B and the competitor begins to make a ton of money

Rely on traditional methods… Hashes

- Data scientists, DBAs, and developers understand the underlying methodologies
- The cost of implementing this kind of solution is usually low
- Because people understand the methods so well, this kind of system is usually robust and reliable



Code Along…

Summary

TriveraTech®
TECHNOLOGY TRAINING

Lab: Hash Your Dataset

Hands-on Lab: Please refer to your Lab Guide
and follow the instructions provided by your Instructor

Experience is

Chapter 2:

Building the AI Fortress

Design and implement robust AI-driven defense and intrusion systems

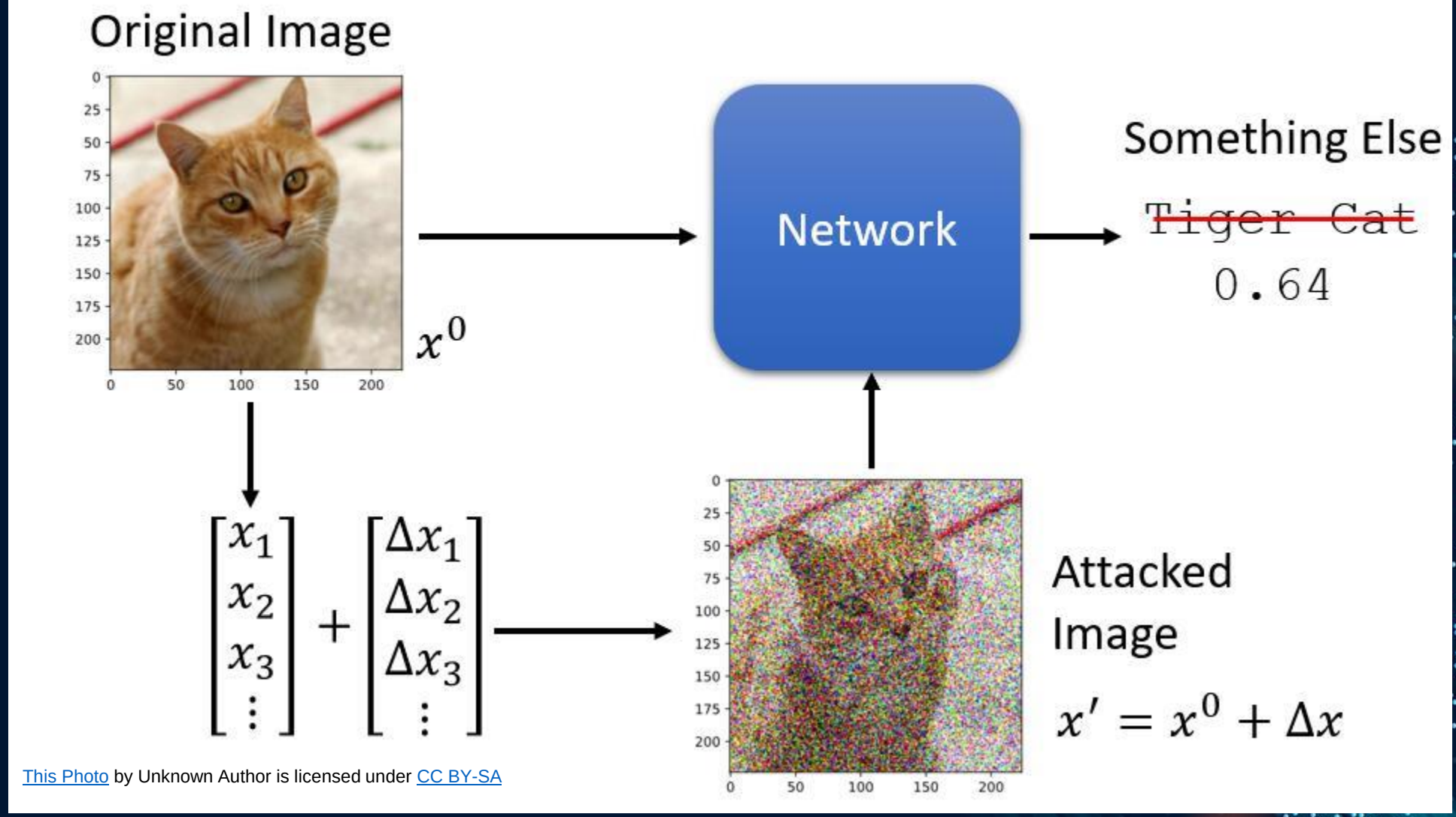
Avoid Adversarial Machine Learning Attacks

- Many adversarial attacks don’t occur directly through data
- Attackers often rely on attacking the machine learning (ML) algorithms through the resulting models.
- Such an attack is termed adversarial ML because it relies on someone purposely attacking the software.

Let’s do the following now:

1. Define adversarial attack
2. Consider security issues in ML
3. Describe the most common attack techniques
4. Mitigate threats to the algorithm

Define Adversarial ML



Categorize Attack Vectors

The Hackers Mindset

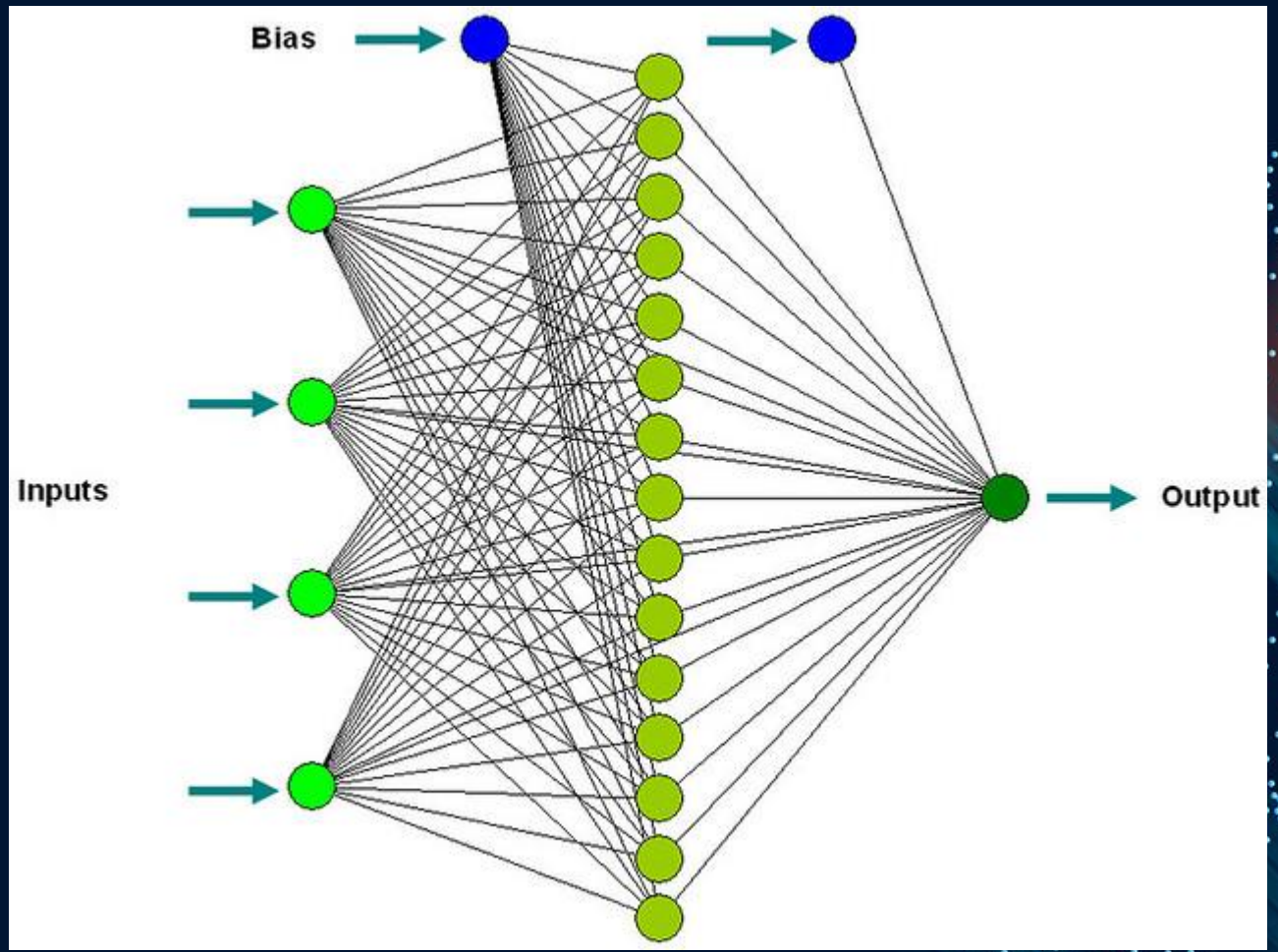
- To obtain money or power
- To take revenge on another party
- Because they need or want attention
- Because there is a misunderstanding as to the purpose of the application
- To make a political statement or create distrust
- Because there is a disagreement over how to accomplish a task

Hacker Goals

- Fly under the security radar
- Stay on the network as long as possible
- Perform specific tasks without being noticed
- Spend as little time as possible breaking into an individual site
- Reuse research performed before the break-in
- Employ previous datasets and statistical analysis to improve future efforts

Trial and Error and Humans as the Weakest Link

- Social Engineering
- Phishing Attacks
- Spoofing



Don’t Help the Attackers!

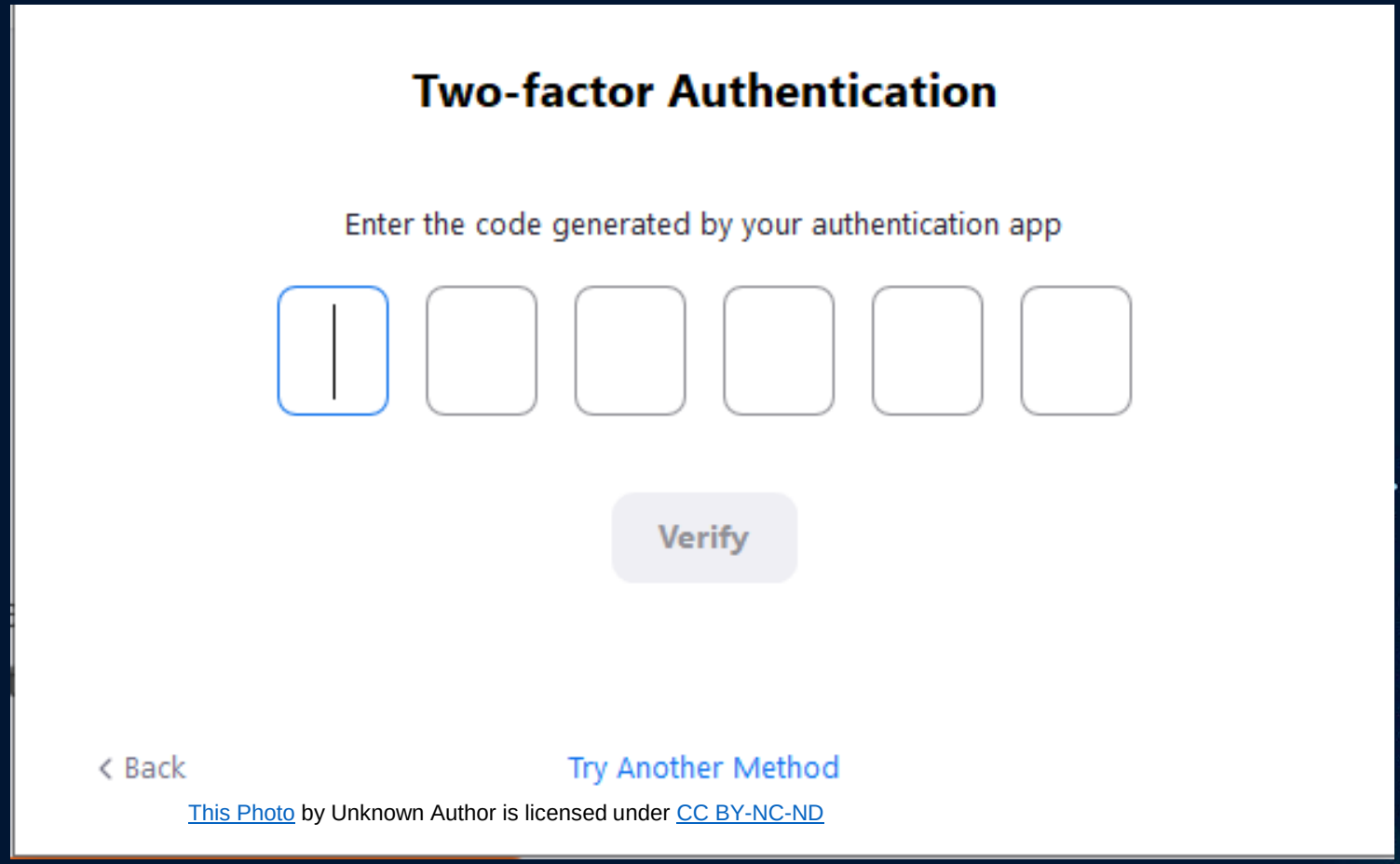
- Keep your secrets by not telling anyone (or keeping the list incredibly small)
- Eliminate clues
- Make the hacker jump through hoops
- Feed the hacker false information
- Learn from the hacker
- Create smarter models

Limit Probing

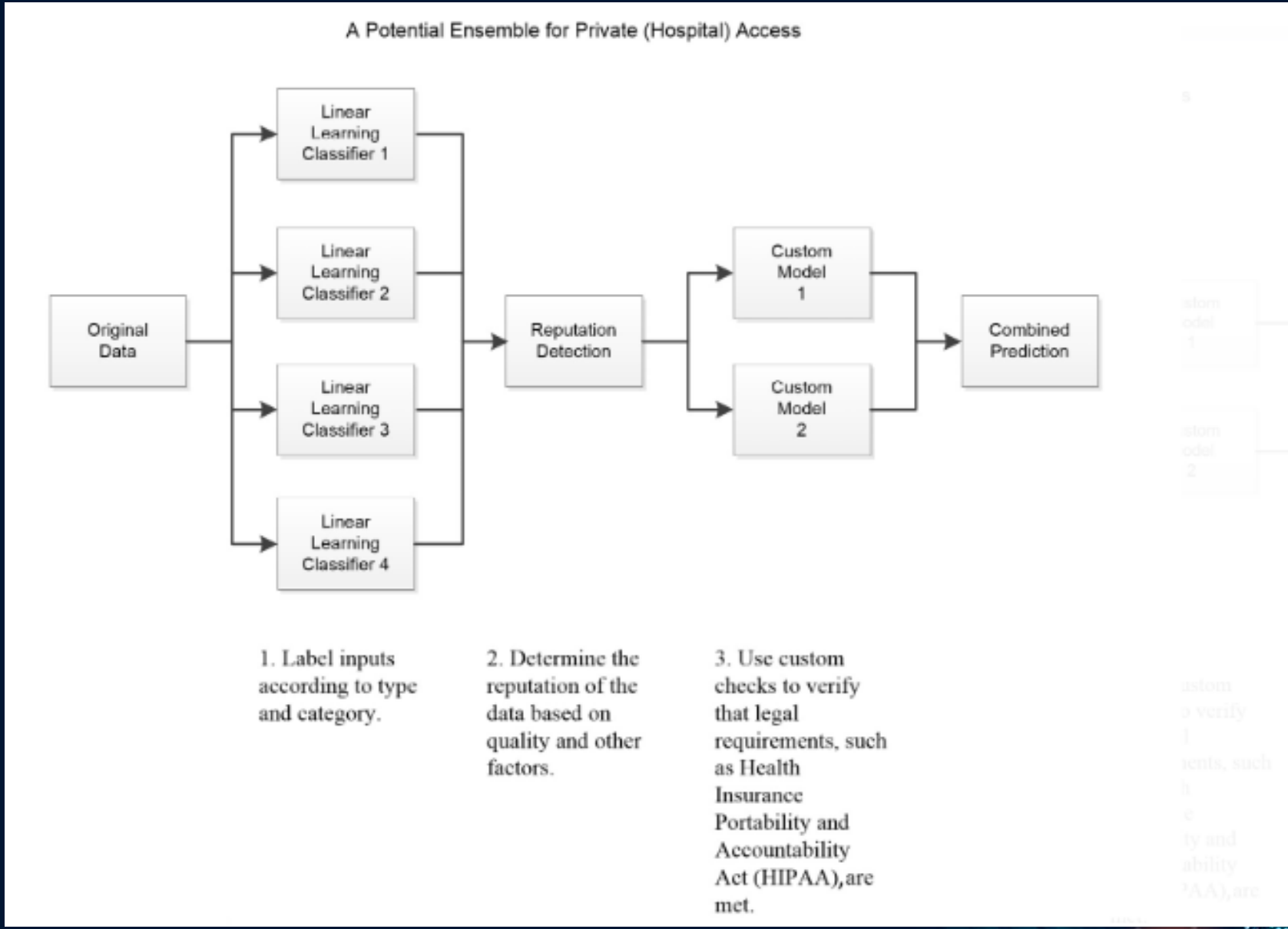
- Probing is the act of interacting with your application in a manner that allows observation of specific results that aren't necessarily part of the application's normal output.
- A hacker could keep trying scripts, control characters, odd data values, control key combinations, or other kinds of inputs and actions to see if an error occurs.

- CONSIDER CAPTCHAS

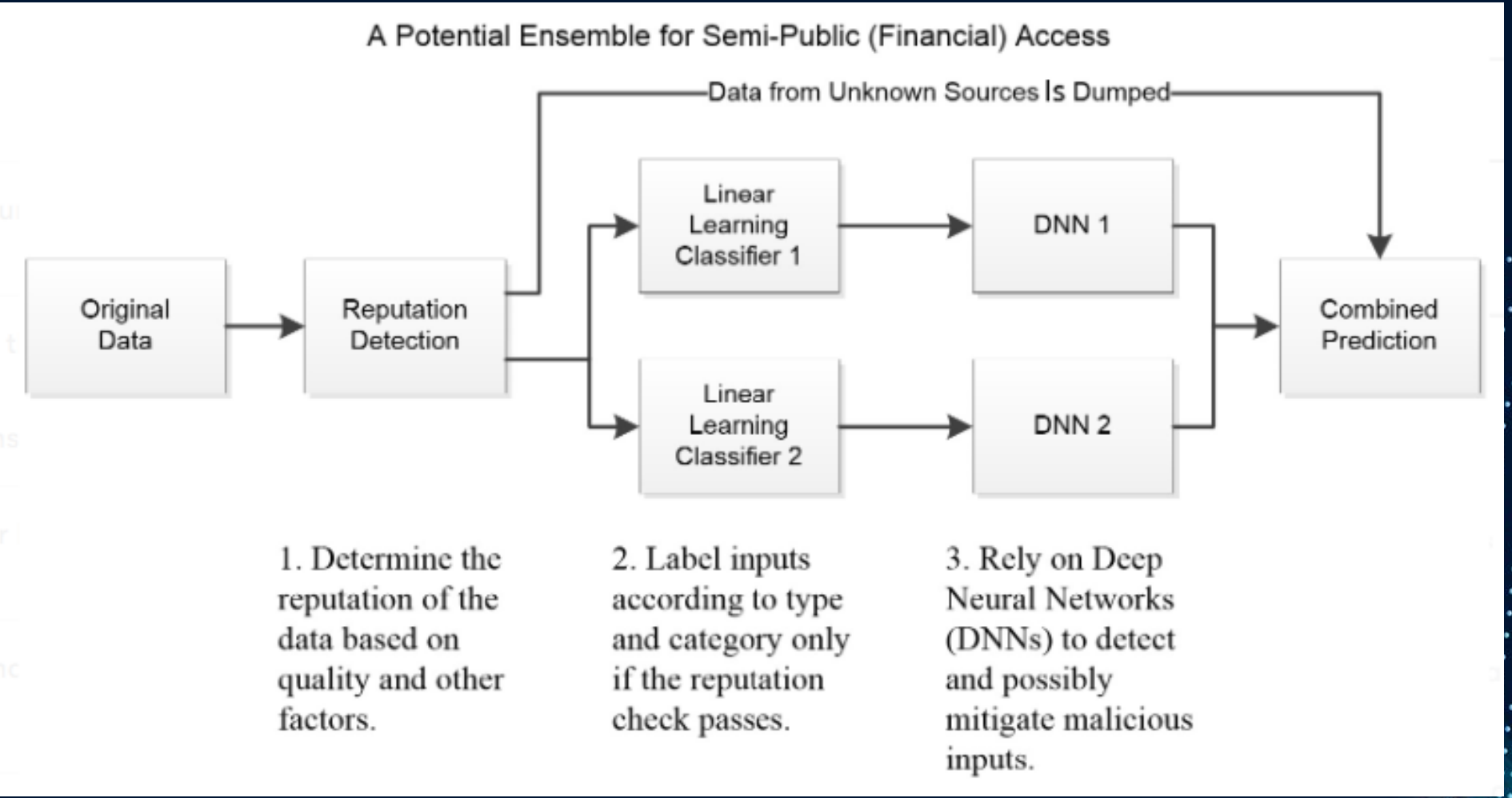
Use 2FA with your ML



Use Ensemble Learning



More Ensemble

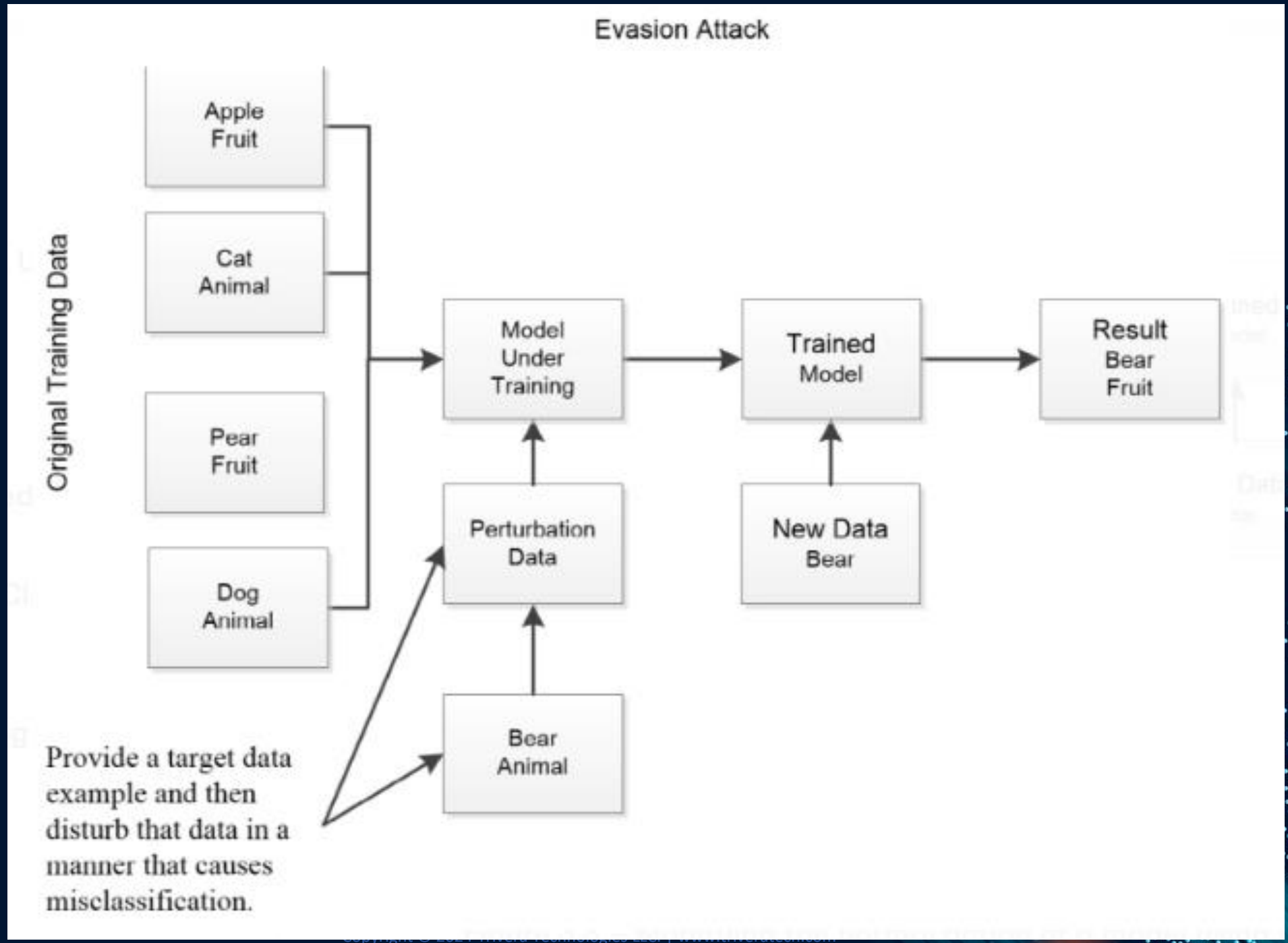


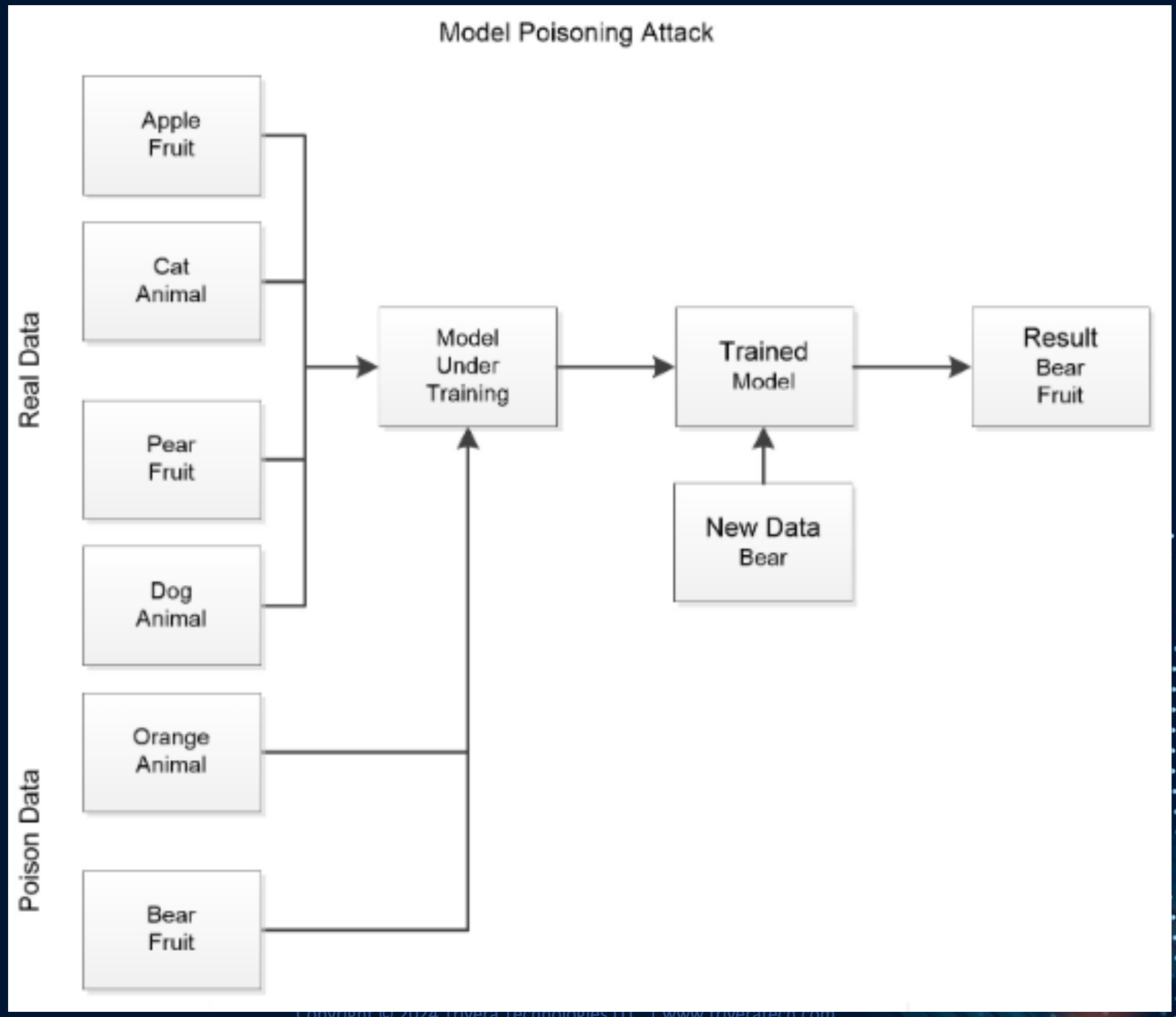
Understand Black Swan Theory

- High-profile, hard-to-predict, and rare events that history, science, finance, and technology can’t explain
- Rare events that modern statistical methods can’t calculate due to the small sample size
- Psychological biases that prevent people from seeing a rare event’s massive effects on historical events

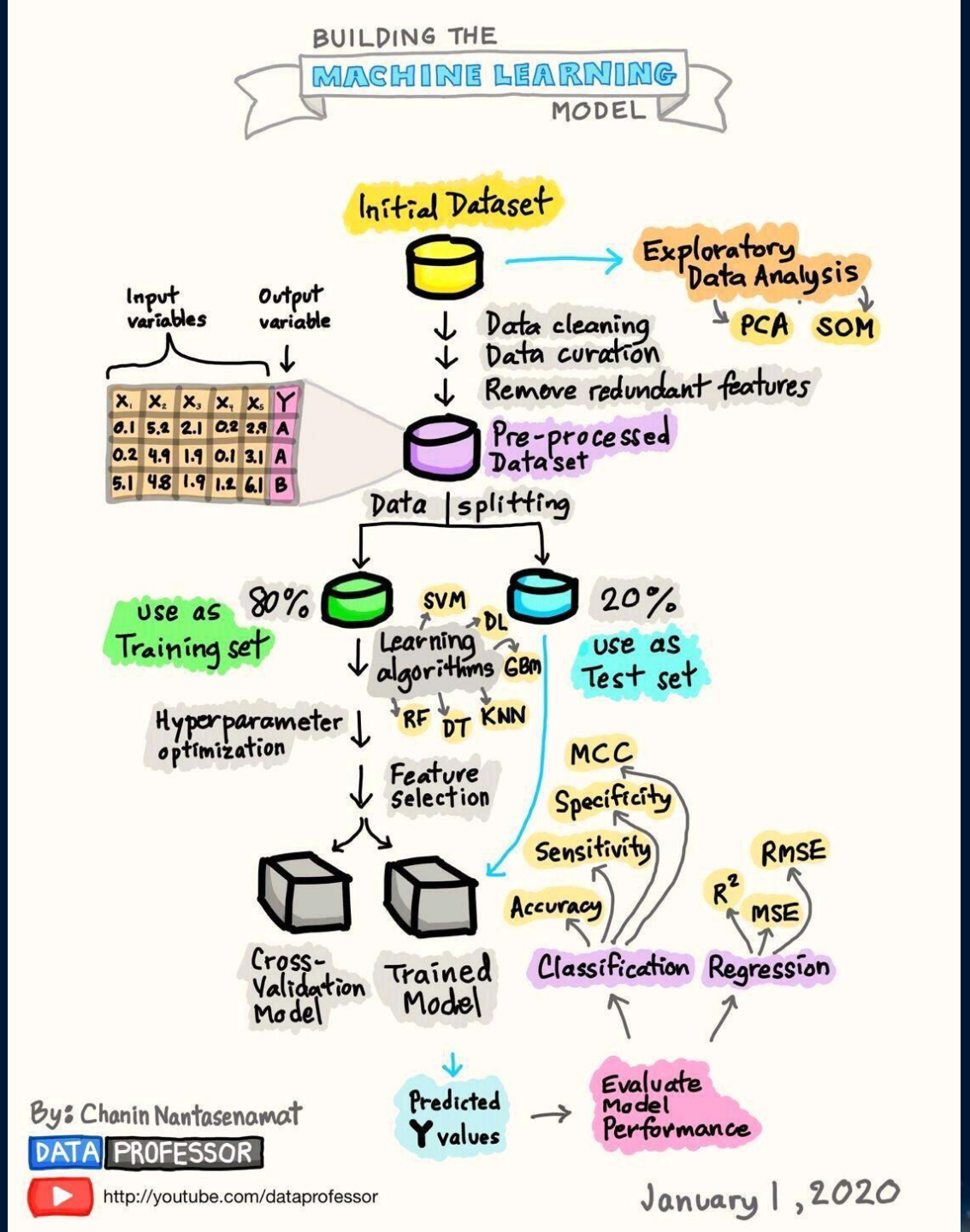
Antiknowledge

- Antiknowledge refers to any agent that reduces the level of knowledge available in a group or society.
- In ML, antiknowledge refers to the loss of knowledge about the inner workings or viability of algorithms, models, or other software due to the emergence of technologies, events, or data that infers previous knowledge is incorrect in some way.

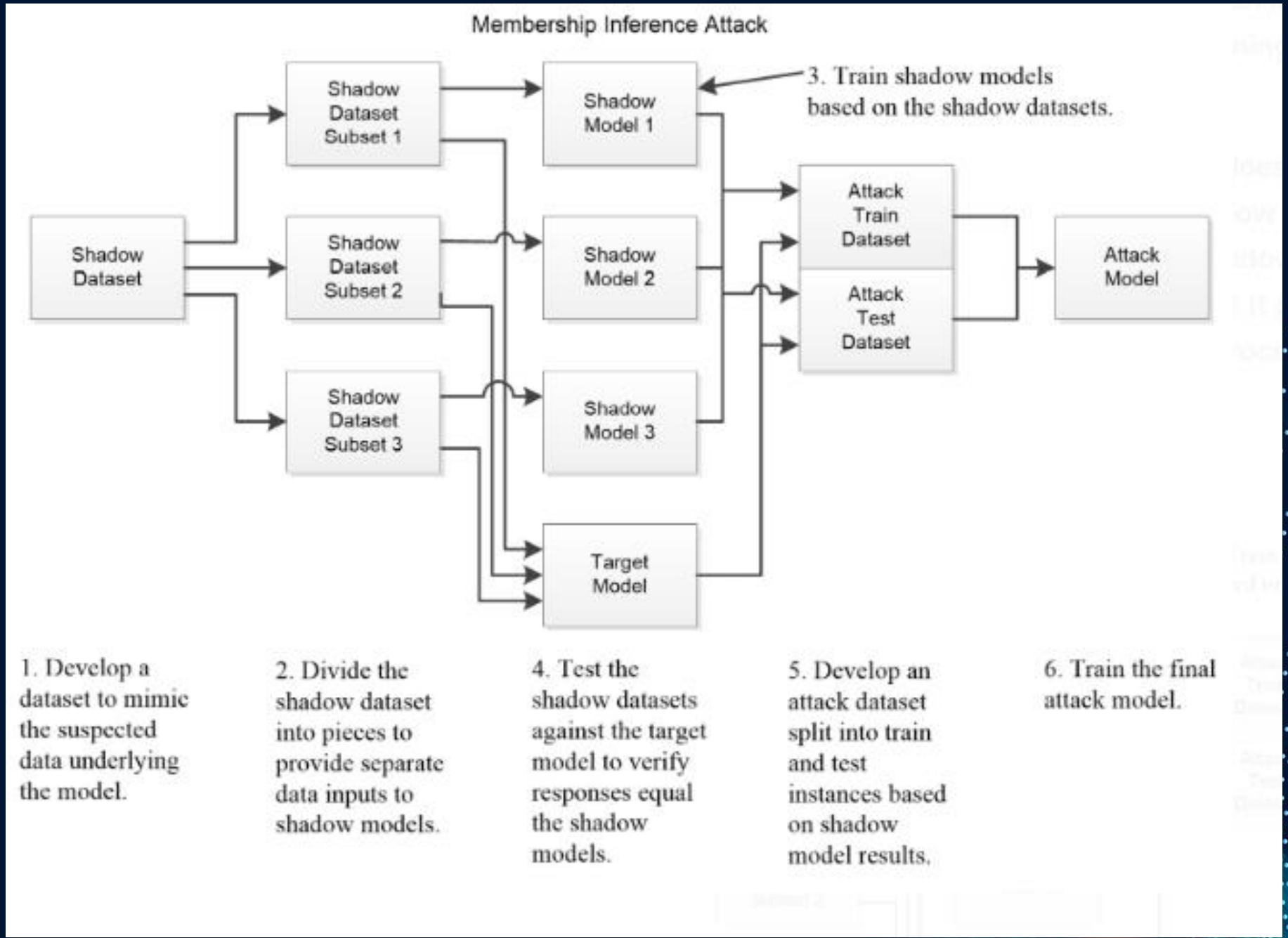

Evasion Attack
Original Training Data
Apple Fruit
Cat Animal
Pear Fruit
Dog Animal
Model Under Training
Trained Model
Result Bear Fruit
Perturbation Data
New Data Bear
Bear Animal
Provide a target data example and then disturb that data in a manner that causes misclassification.

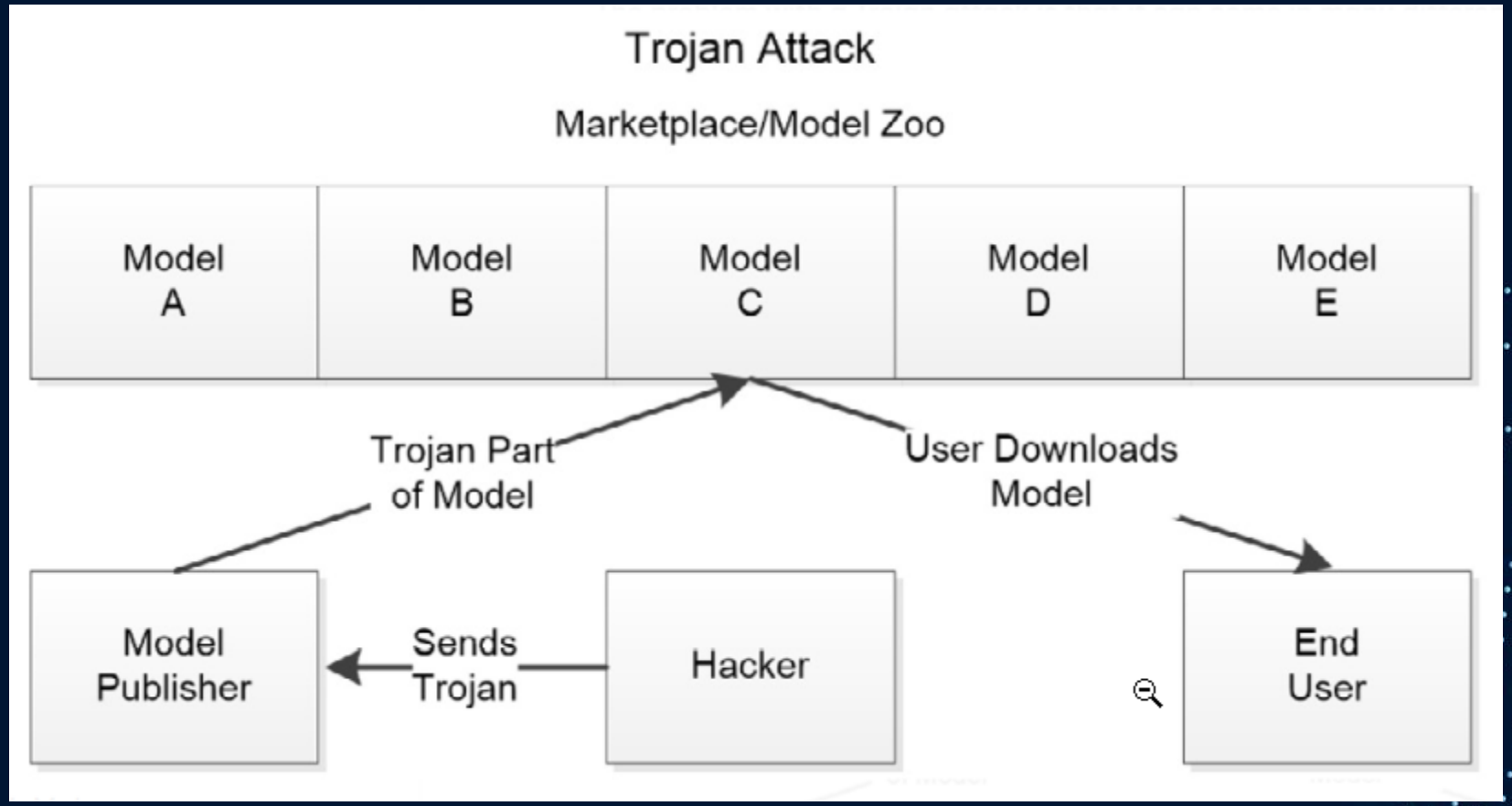


Model Skewing

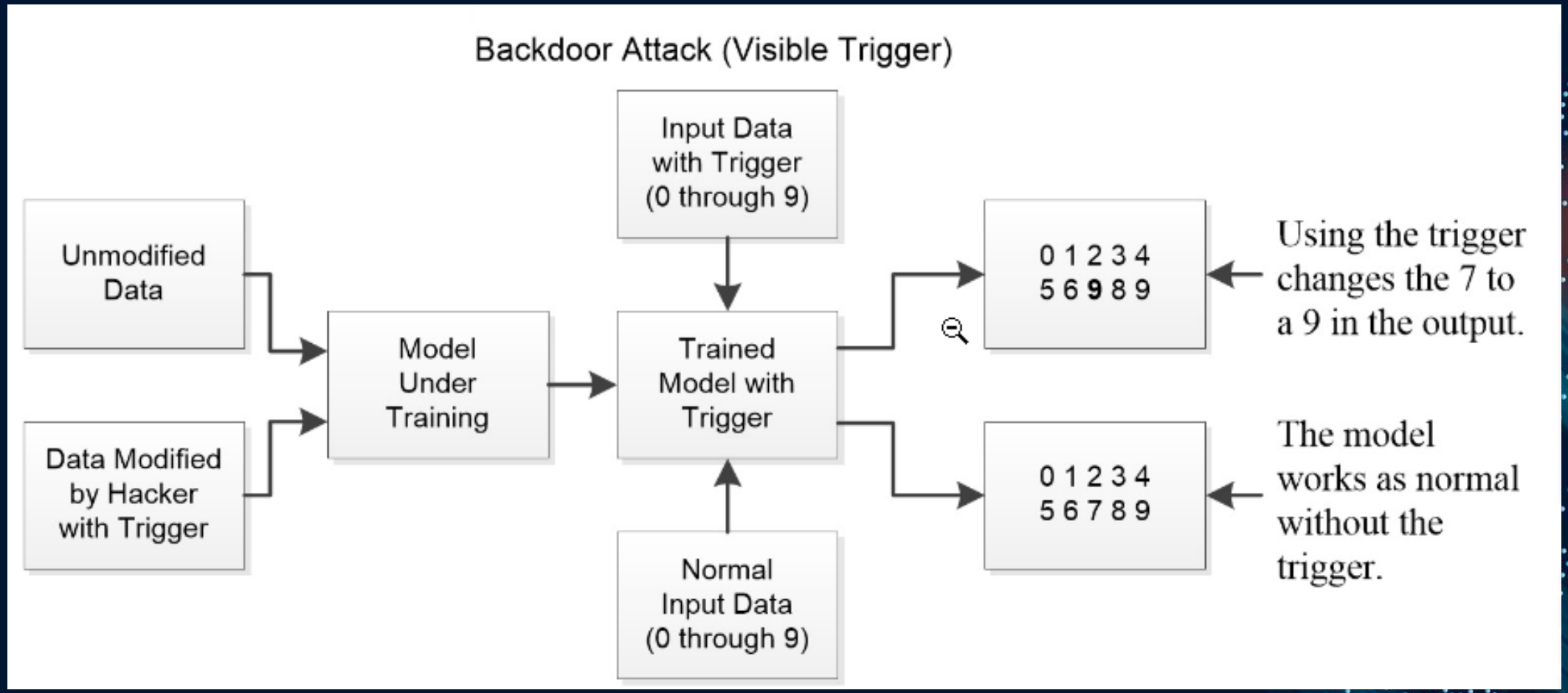


Feedback Weaponization

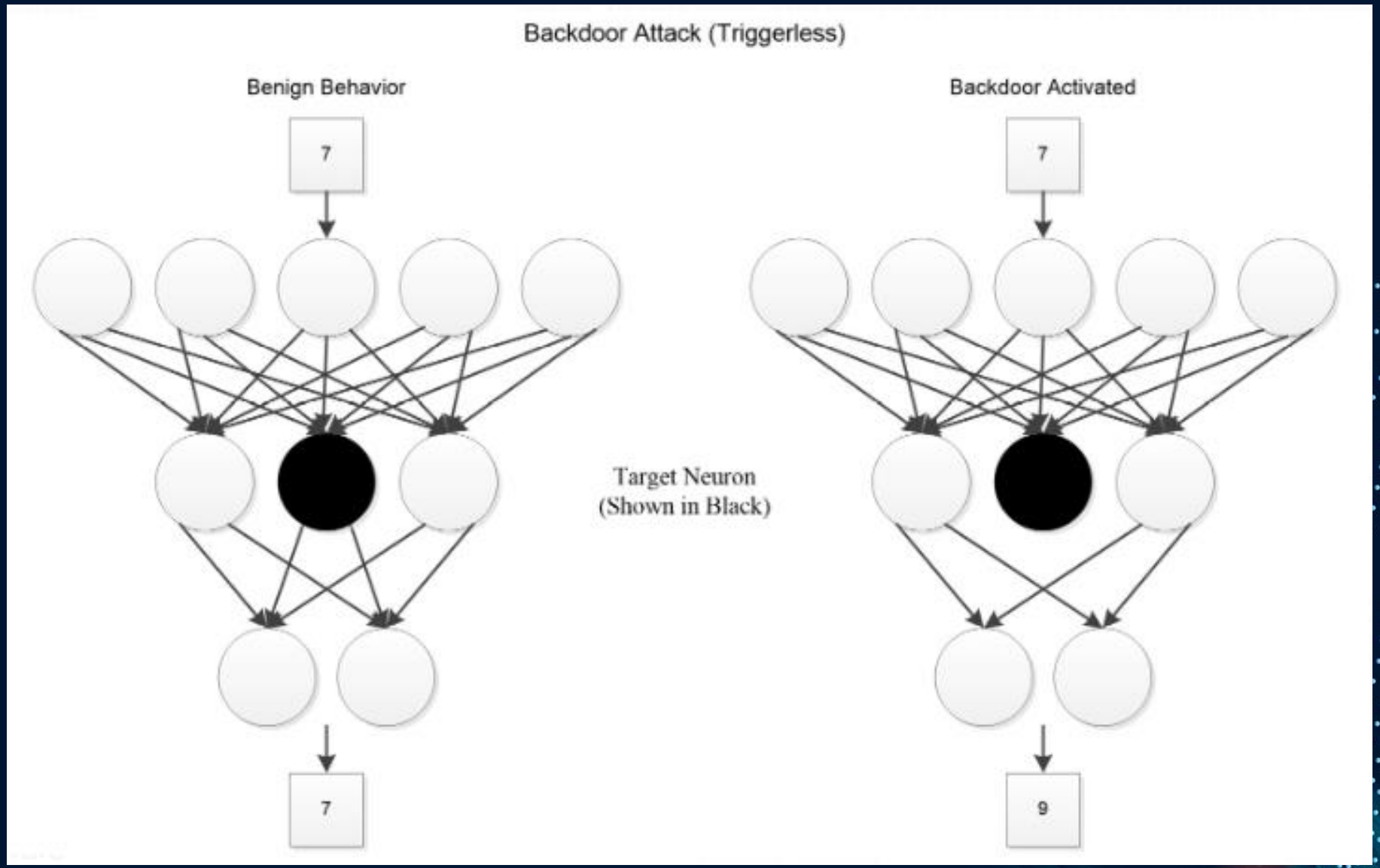




Backdoor (neural) attacks



Triggerless Backdoor Attack



See it in action…

- https://kennysong.github.io/adversarial.js/

Attack Types by Strength (Carlini & Wagner Strongest)

- Carlini and Wagner: See details at https://arxiv.org/pdf/1608.04644.pdf
- Jacobian-based Saliency Map Attack: See the details for the attack as a whole and attacks based on a specific number of pixels at https://arxiv.org/abs/2007.06032 and https://arxiv.org/pdf/1808.07945.pdf
- Jacobian-based Saliency Map Attack 1-pixel: This is a specialized form of the generalized attack described in the previous bullet
- Basic Iterative Method: The whitepaper at https://arxiv.org/pdf/1607.02533.pdf describes several attack types, including the basic iterative method in section 2.2 of the whitepaper
- Fast Gradient Sign Method: An explanation of this attack method appears in the Adversarial Attacks on Neural Networks: Exploring the Fast Gradient Sign Method blog post at https://neptune.ai/blog/adversarial-attacks-on-neural-networks-exploring-the-fast-gradient-sign-method

ML Security in the Real World

- Ensuring user authentication (the validation that the user's identity is real) and authorization (giving the user the correct rights) go as planned
- Filtering out potentially hazardous data before the user even gets to see it

Understanding the kinds of App Security

- Role Based
- Attribute Based
- Resource Based
- Group Based
- Identity Based

The Minimum

1. Requesting authorization
2. Authenticating the individual
3. Monitoring and logging their access
4. Verifying that each action is allowed by the application security profile

Code Along…

Summary

TriveraTech®
TECHNOLOGY TRAINING

Lab: None for this section

Hands-on Lab: Please refer to your Lab Guide
and follow the instructions provided by your Instructor

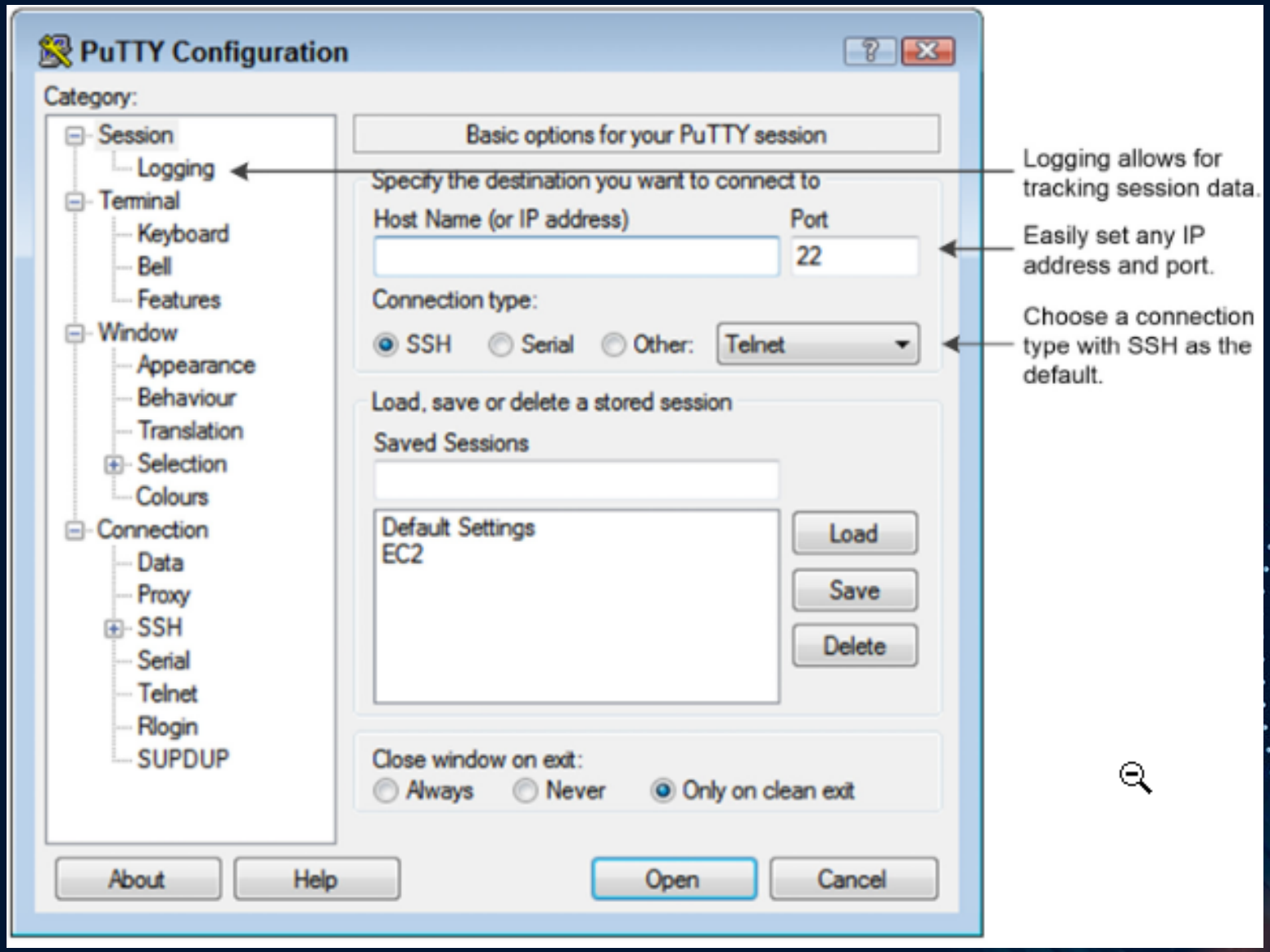
Experience is

Chapter 4:

CSI Cyber

Keep Your Network Clean

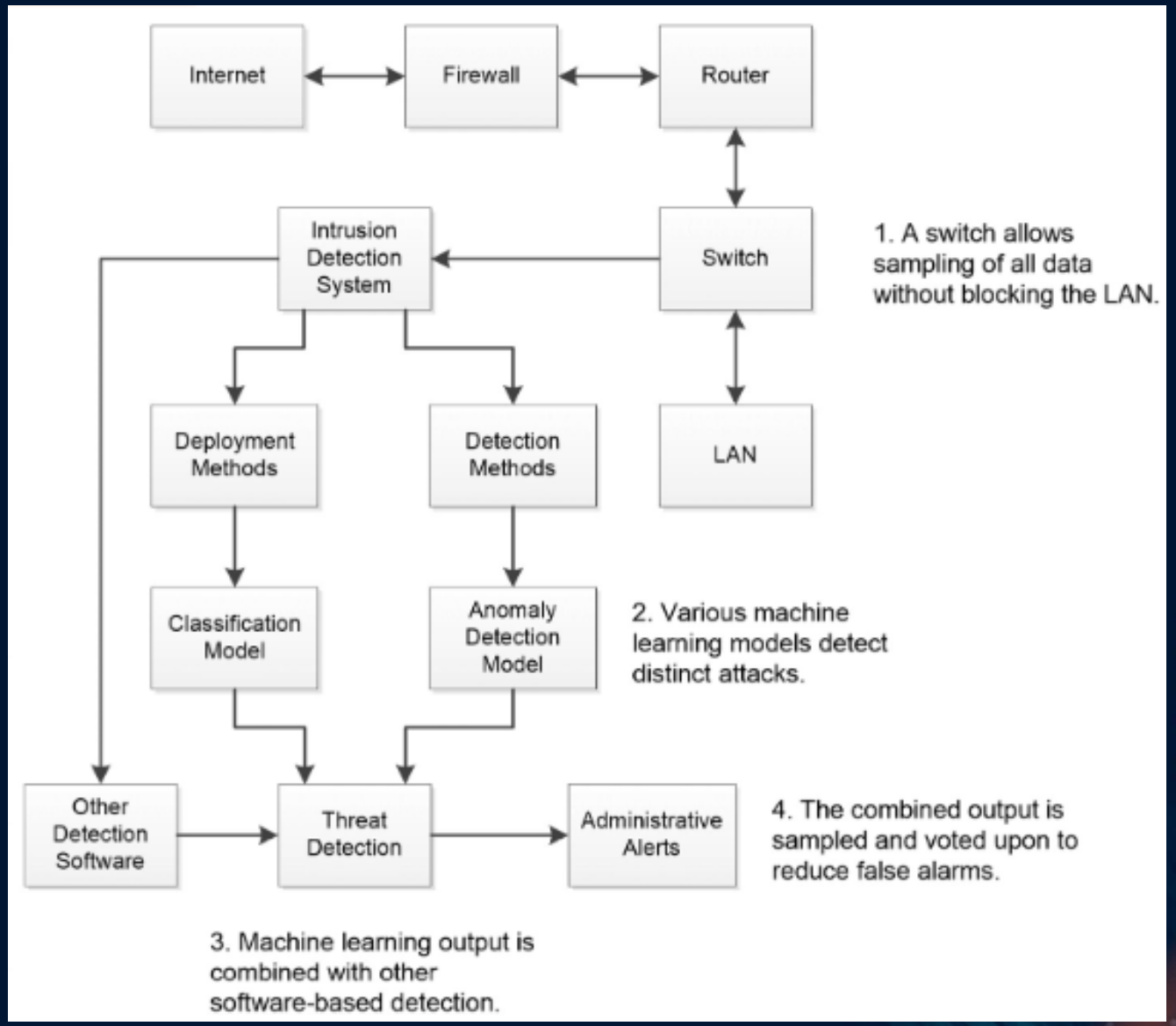
Access Control



Authentication

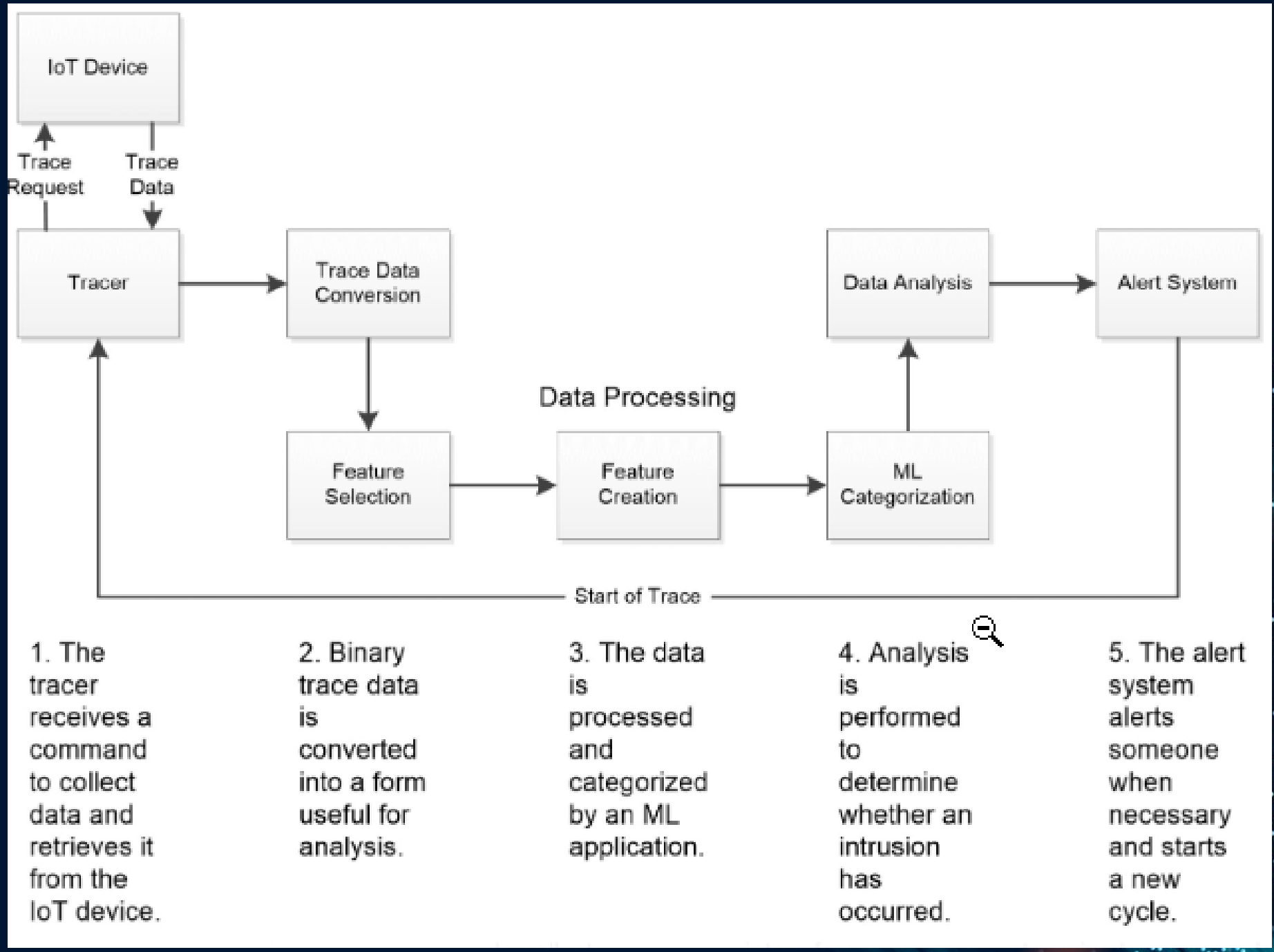
- Device type and/or name
- Location
- Network type
- Operating system
- User risk profile
- User tendencies

Intrusion Detection



Subtle Intrusion Detectors

1. A hacker gains entry to a smart device connected to the internet for monitoring purposes.
2. The hacker changes data in an unobtrusive manner on the device, such that the data will produce an unexpected result when processed by analytics software.
3. The user or host service accesses the device from a desktop system, tablet, or other device attached to the network.
4. The data modifications produce an unexpected result.
5. The network is now potentially open to attack due to the result produced by the analytics software.



Add ML to the AI Security Mix

- Perform regression analysis to determine whether certain packets are somehow flawed compared to normal packets from a given source. In other words, you're not dealing with absolutes but, rather, determining what is normal from a particular sender. Anything outside the normal pattern is suspect.
- Rely on classification to detect whether incoming data matches particular suspect patterns. Unlike signature matching, this form of analysis relies on training a neural network to recognize classes of data that it hasn't seen before. Consequently, even if an attacker changes a signature, the model can still likely recognize the data class.
- Use clustering to detect attack patterns and as part of forensic analysis in real time. For example, suddenly seeing groups of requests from a particular set of IP addresses that all have the same characteristics is a type of suspect pattern.

Develop an Updated Security Plan

- Determining what sort of data to collect before, during, and after an attack
- Tracking user activities for analysis as part of detecting attack vectors
- Creating and testing models specifically designed for security needs and then providing instructions for deploying them during an attack

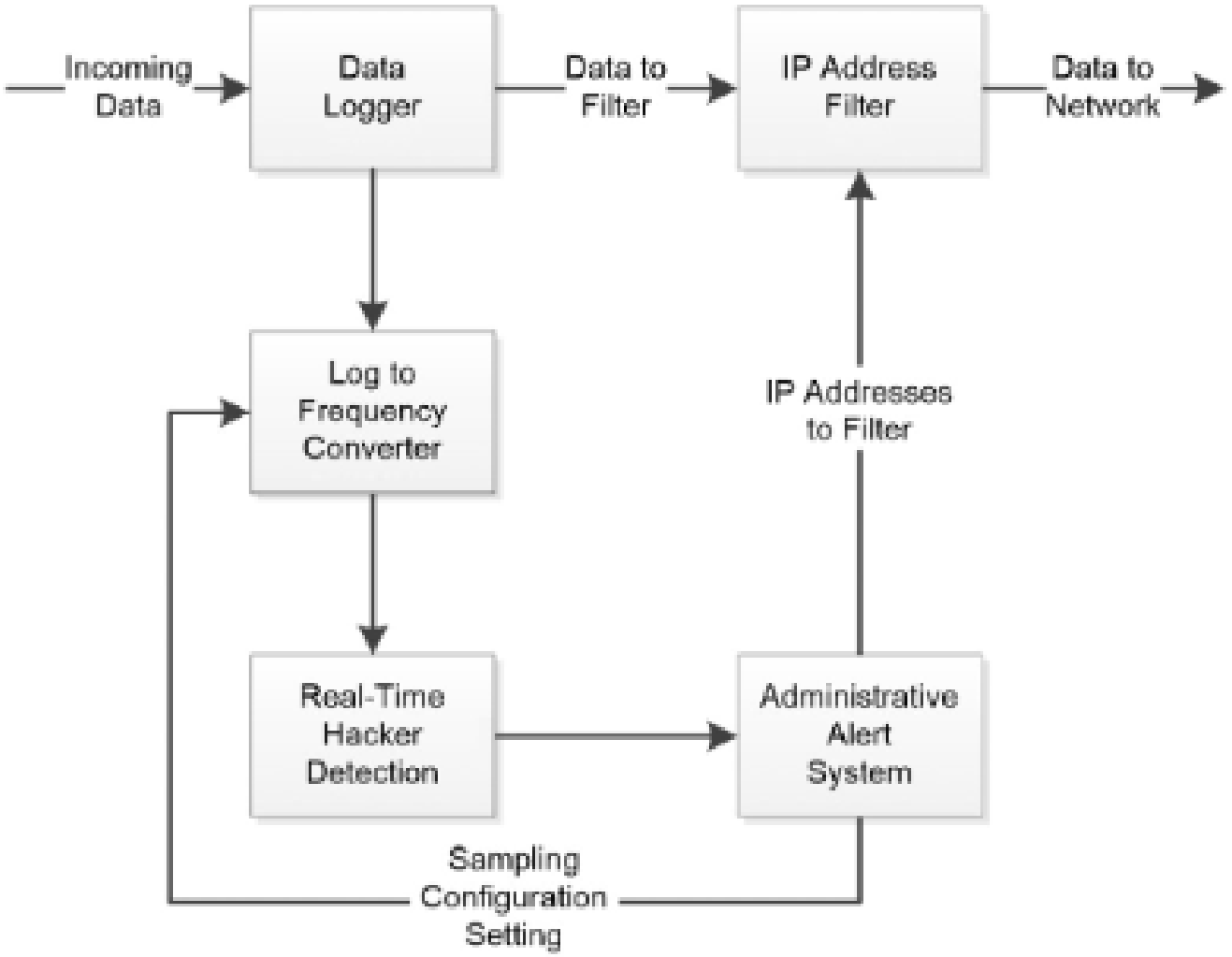
Which Features to Track

- Track user behaviors such as login time, the time between breaks, and other factors using regression
- Employ known user factors, such as meeting times, to classify users by peer group (such as a workgroup or users who exercise during lunch)
- Use clustering techniques to detect users who have unusual habits or aren't part of known groups (the outliers)

Real Time Defenses

- Detect
- Analyze
- Mitigate
- (DAM)

1. Incoming data is logged by time and address.



5. Hacker calls are filtered out.

2. Data is converted to show call frequency.

3. Suspicious call patterns are detected.

4. Administrator reacts to the hacker attack.

Code Along…

TriveraTech®
TECHNOLOGY TRAINING

Lab: Lab Name (or Demo)

Hands-on Lab: Please refer to your Lab Guide
and follow the instructions provided by your Instructor

Experience is

Chapter 5:

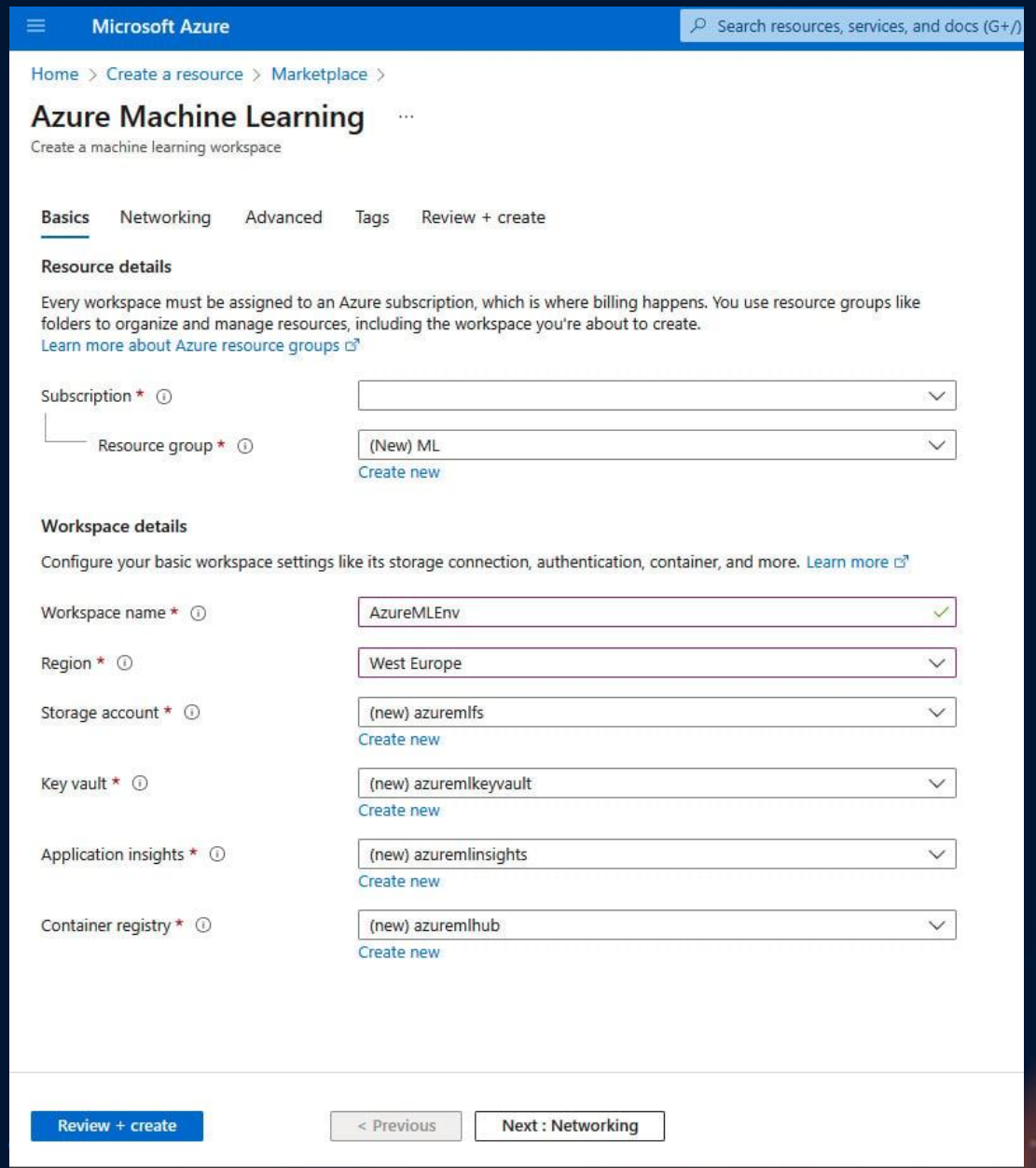
AI Adversarial Attacks and Defenses

Exploring Strategies and Defenses

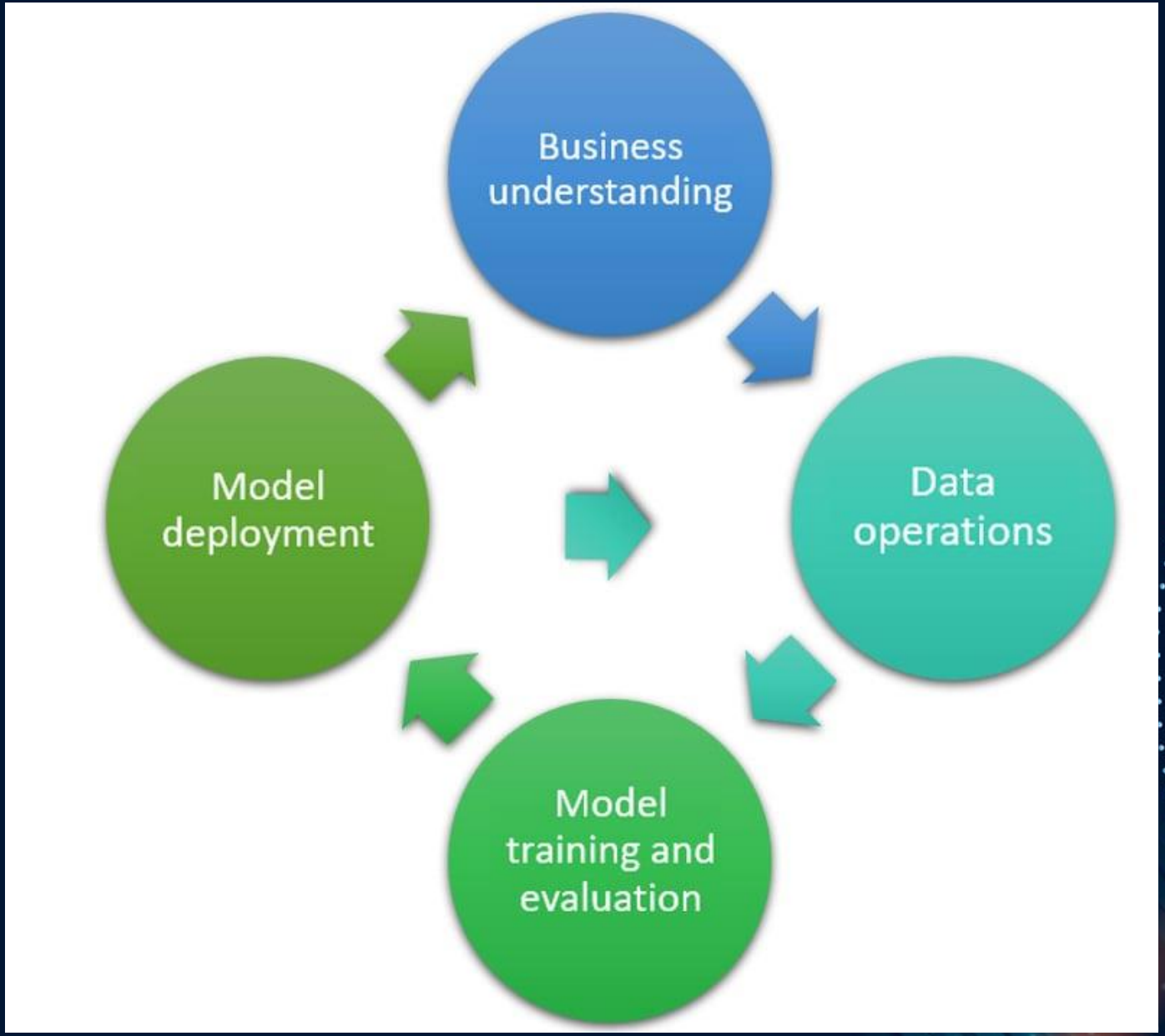
Assessing the Vulnerability of Your Algorithms, Models, and AI Environments

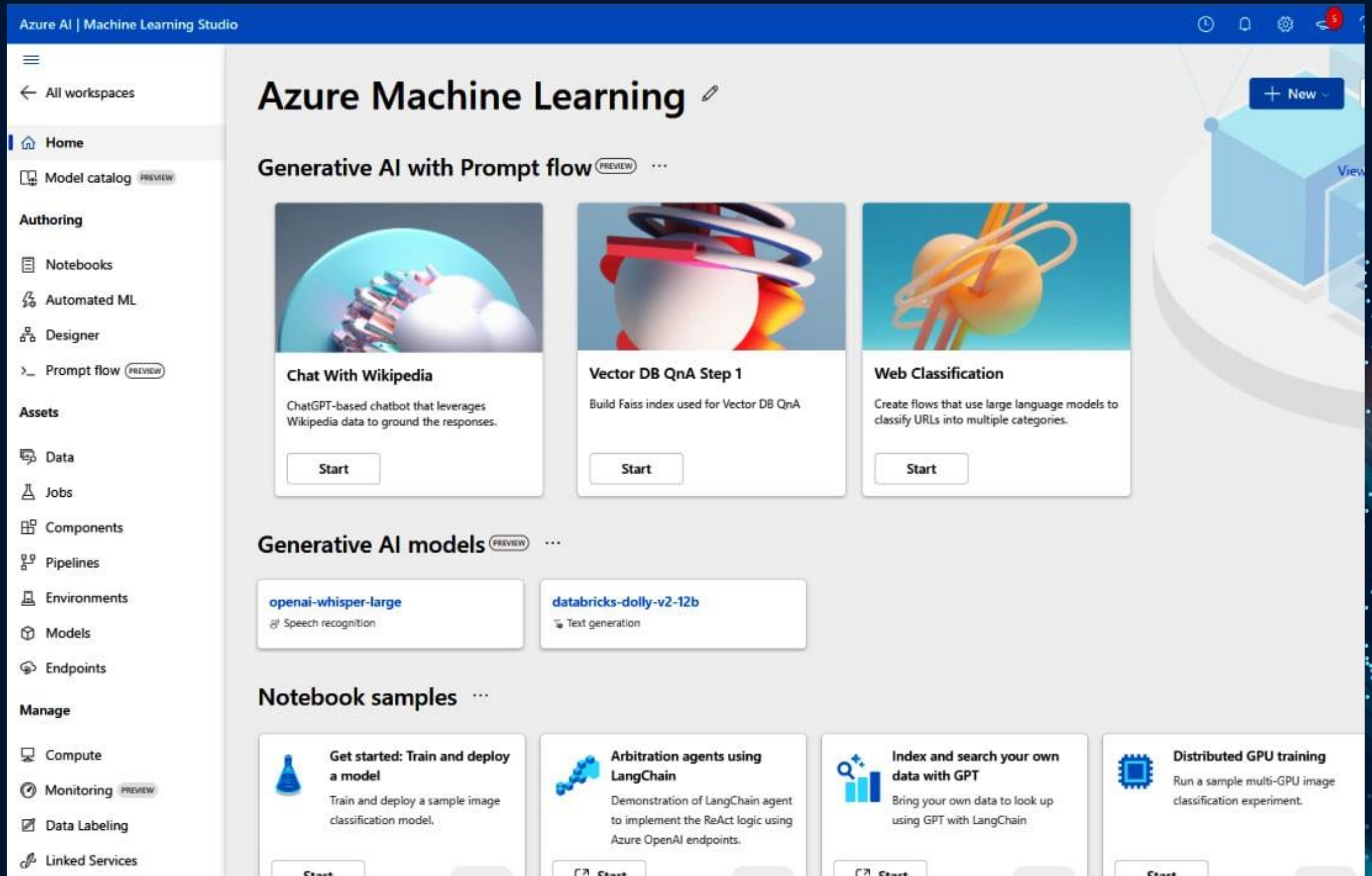
- Reviewing the Azure Machine Learning life cycle
- Introducing an ML project
- Exploring the Zero Trust model
- Assessing the vulnerability of ML assets and apps

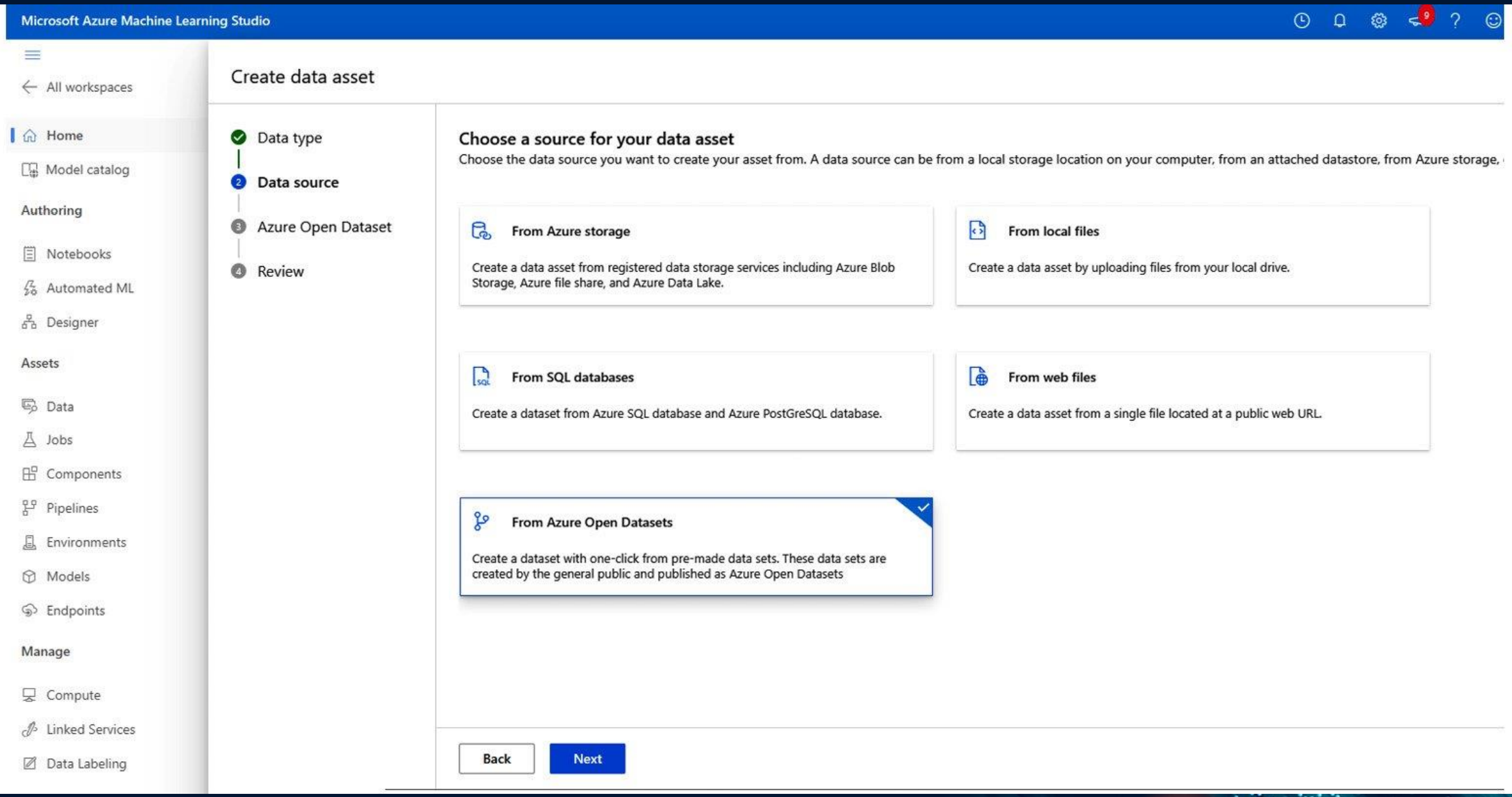
Cloud Machine Learning (Azure)

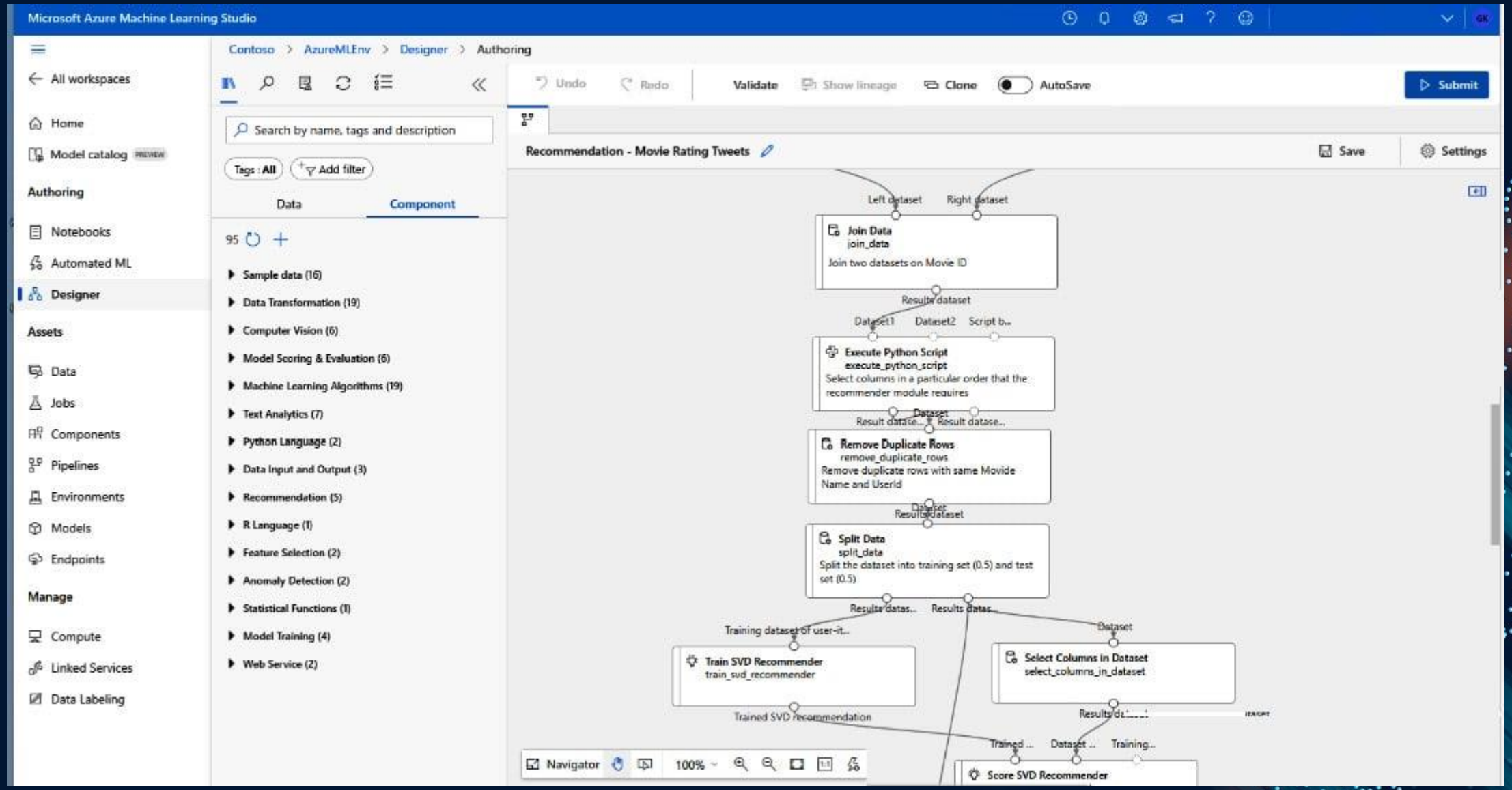


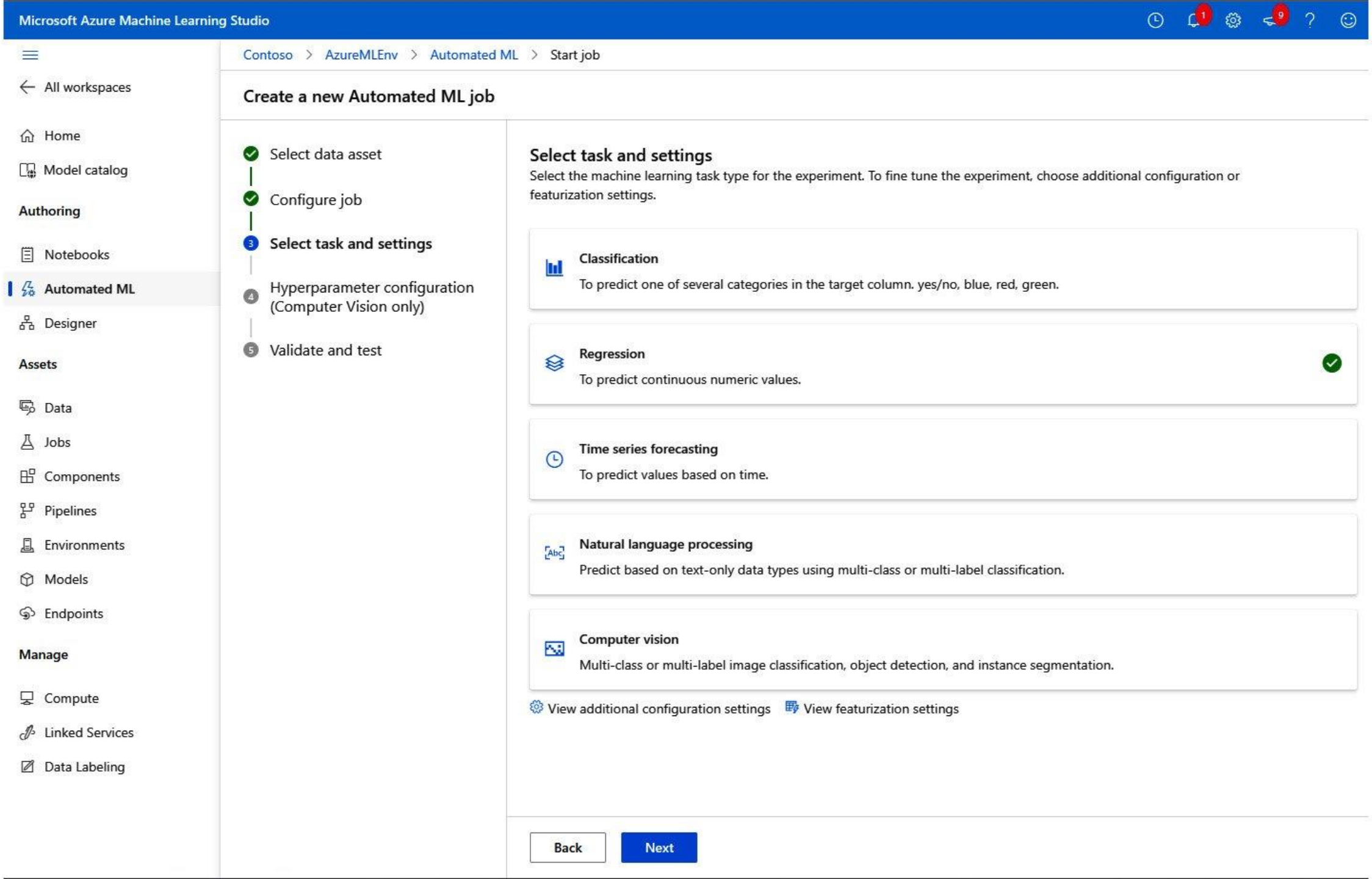
Cloud ML Life Cycle



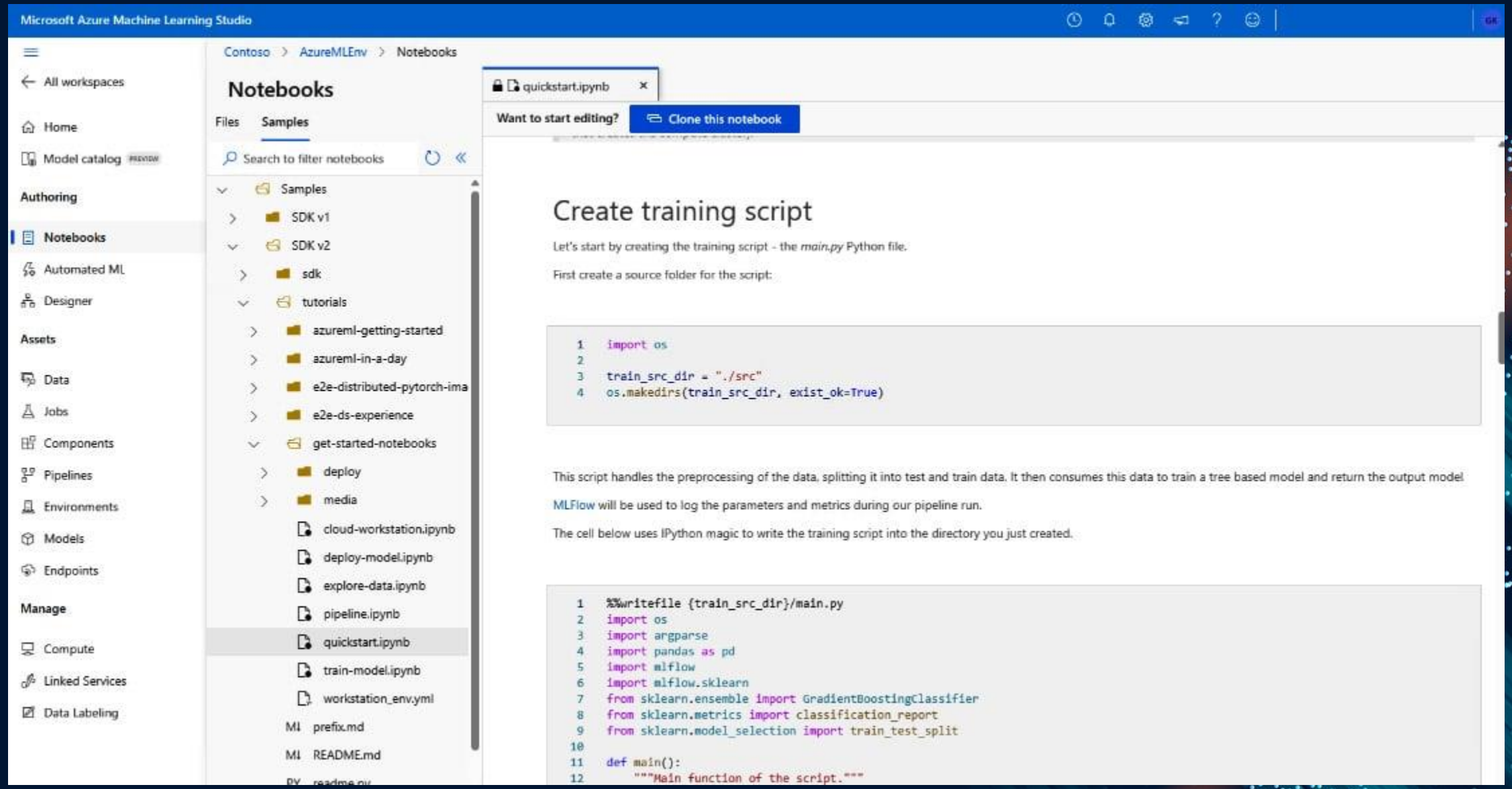

Azure AI | Machine Learning Studio
All workspaces
Home
Model catalog PREVIEW
Authoring
Notebooks
Automated ML
Designer
Prompt flow PREVIEW
Assets
Data
Jobs
Components
Pipelines
Environments
Models
Endpoints
Manage
Compute
Monitoring PREVIEW
Data Labeling
Linked Services
Azure Machine Learning
New
Generative AI with Prompt flow PREVIEW
Chat With Wikipedia
ChatGPT-based chatbot that leverages Wikipedia data to ground the responses.
Start
Vector DB QnA Step 1
Build Faiss index used for Vector DB QnA
Start
Web Classification
Create flows that use large language models to classify URLs into multiple categories.
Start
Generative AI models PREVIEW
openai-whisper-large
Speech recognition
databricks-dolly-v2-12b
Text generation
Notebook samples
Get started: Train and deploy a model
Train and deploy a sample image classification model.
Arbitration agents using LangChain
Demonstration of LangChain agent to implement the ReAct logic using Azure OpenAI endpoints.
Index and search your own data with GPT
Bring your own data to look up using GPT with LangChain
Distributed GPU training
Run a sample multi-GPU image classification experiment.



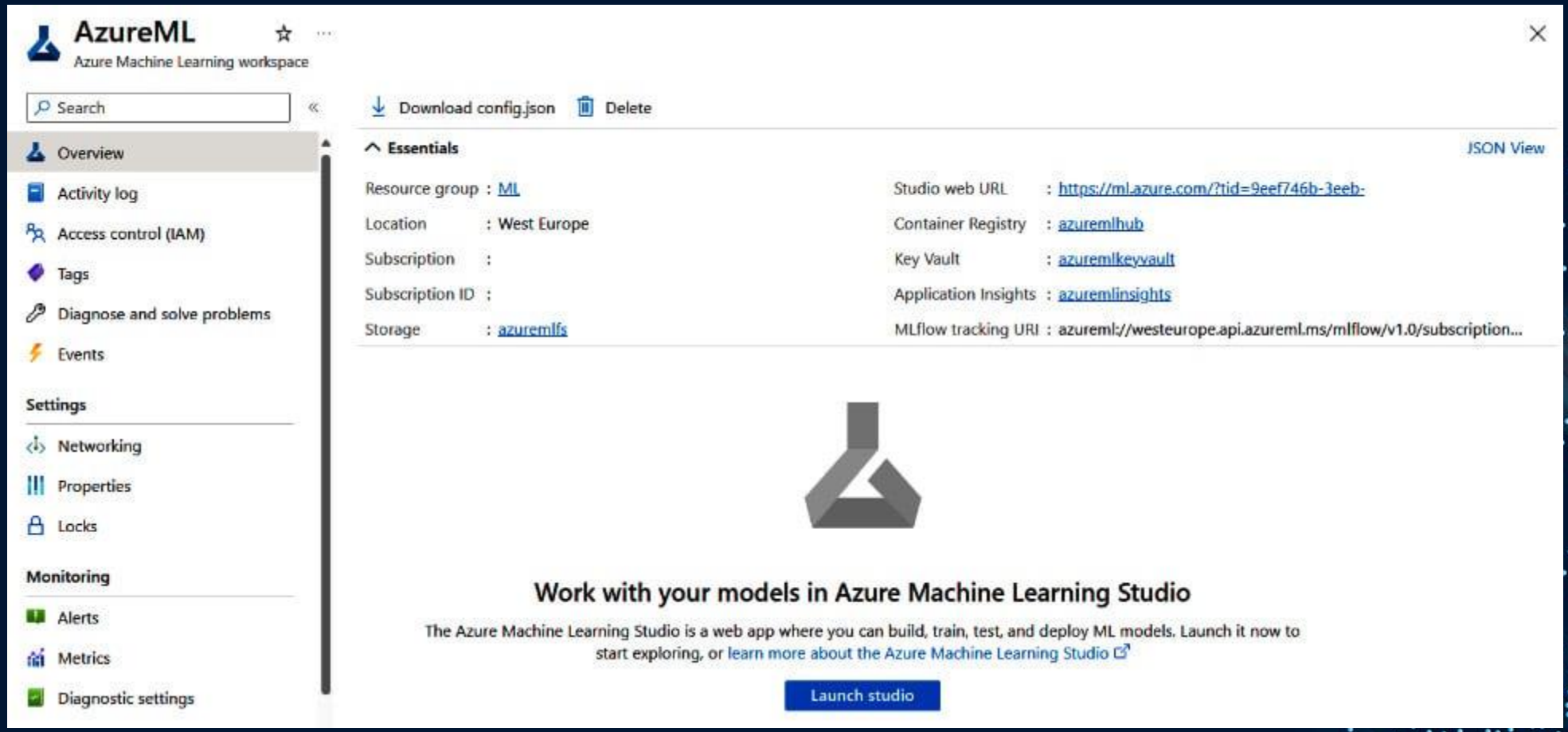



Microsoft Azure Machine Learning Studio
All workspaces
Home
Model catalog
Authoring
Notebooks
Automated ML
Designer
Assets
Data
Jobs
Components
Pipelines
Environments
Models
Endpoints
Manage
Compute
Linked Services
Data Labeling
Contoso > AzureMLEnv > Automated ML > Start job
Create a new Automated ML job
Select data asset
Configure job
3 Select task and settings
4 Hyperparameter configuration (Computer Vision only)
5 Validate and test
Select task and settings
Select the machine learning task type for the experiment. To fine tune the experiment, choose additional configuration or featurization settings.
Classification
To predict one of several categories in the target column. yes/no, blue, red, green.
Regression
To predict continuous numeric values.
Time series forecasting
To predict values based on time.
Natural language processing
Predict based on text-only data types using multi-class or multi-label classification.
Computer vision
Multi-class or multi-label image classification, object detection, and instance segmentation.
View additional configuration settings
View featurization settings
Back
Next

Notebooks



ML Project



Please complete Lab 1: Setup your Azure Environment

- We will break into groups of 2 or 3
- Please download the Lab 1 Lab Guide from here:

- https://github.com/fenago/ai-security/tree/main/lab-guides

Microsoft Azure Machine Learning Studio

All workspaces

Home
Model catalog

Authoring

Notebooks
Automated ML
Designer

Assets

Data
Jobs
Components
Pipelines
Environments
Models
Endpoints

Manage

Compute
Linked Services
Data Labeling

Create data asset

1. Data type
2. Data source

Set the name and type for your data asset

Name *

Diabetes

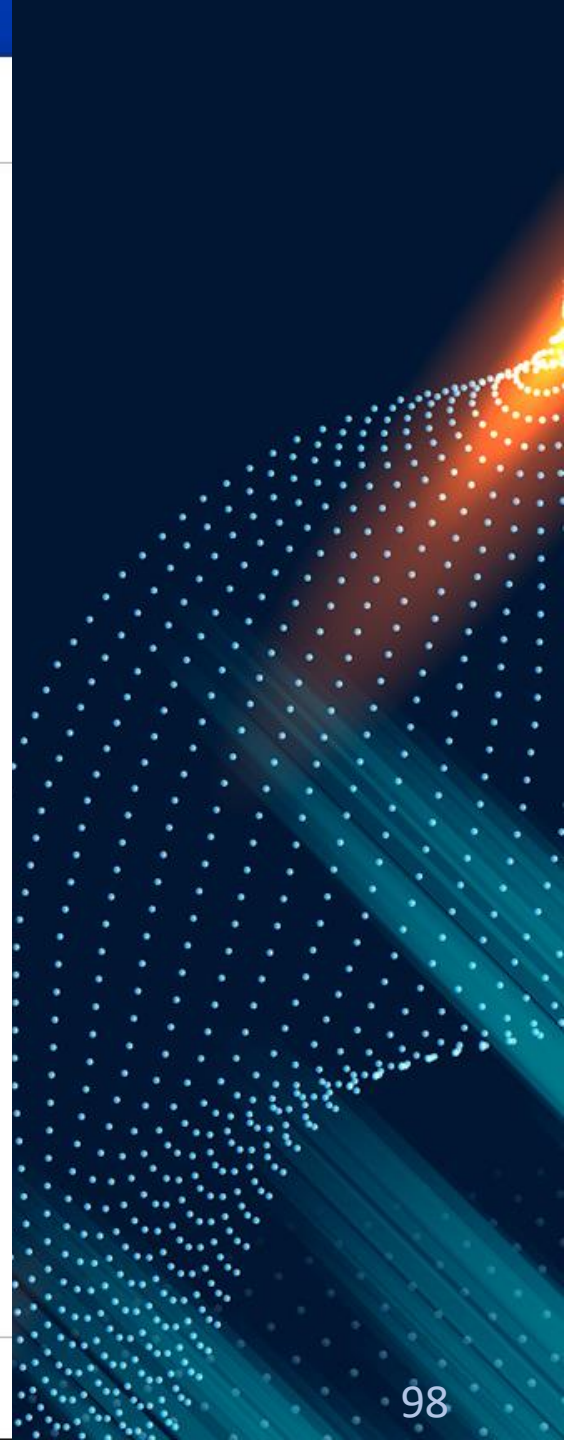
Description

The Diabetes dataset has 442 samples with 10 features, making it ideal for getting started with machine learning algorithms.

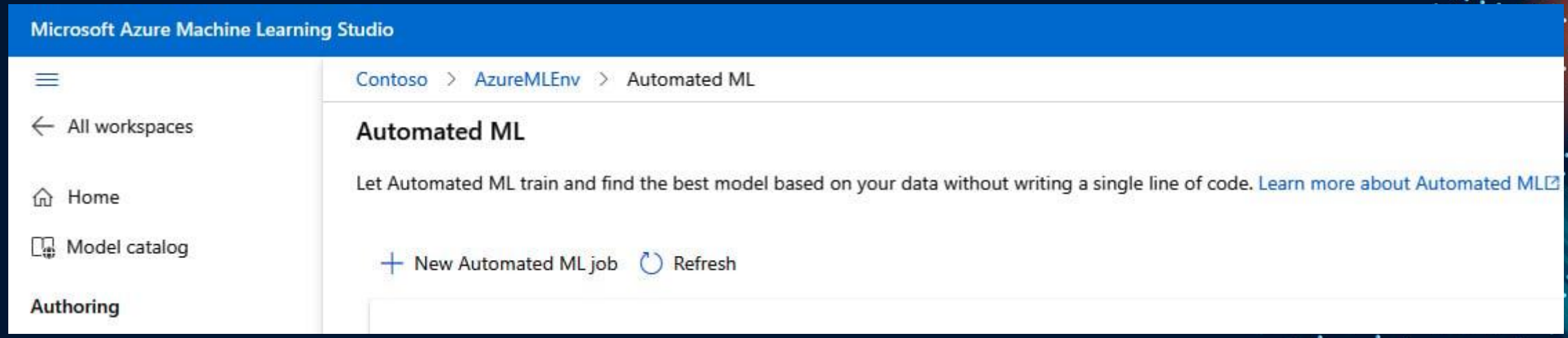
Type *

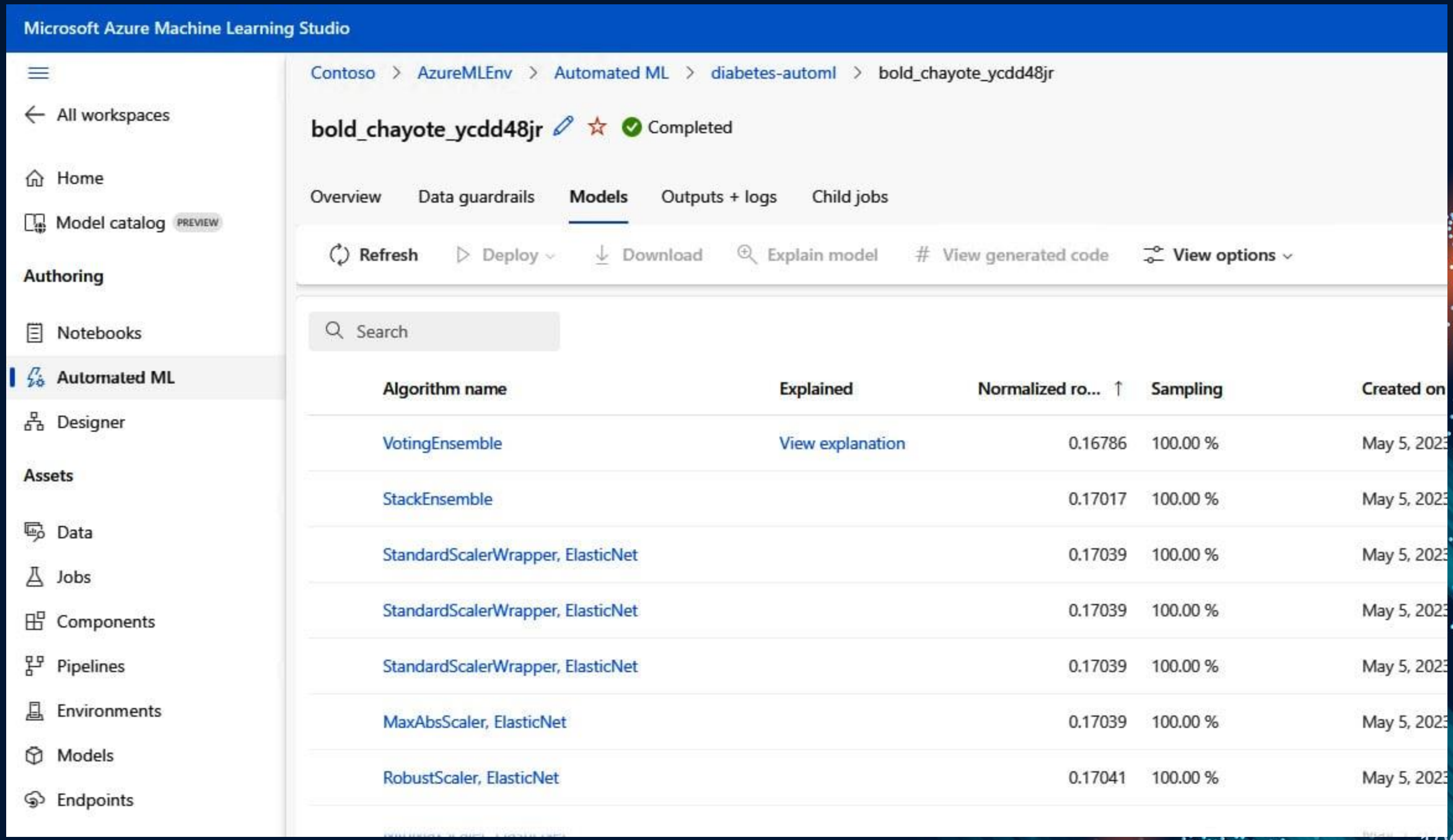
Tabular

Back | Next

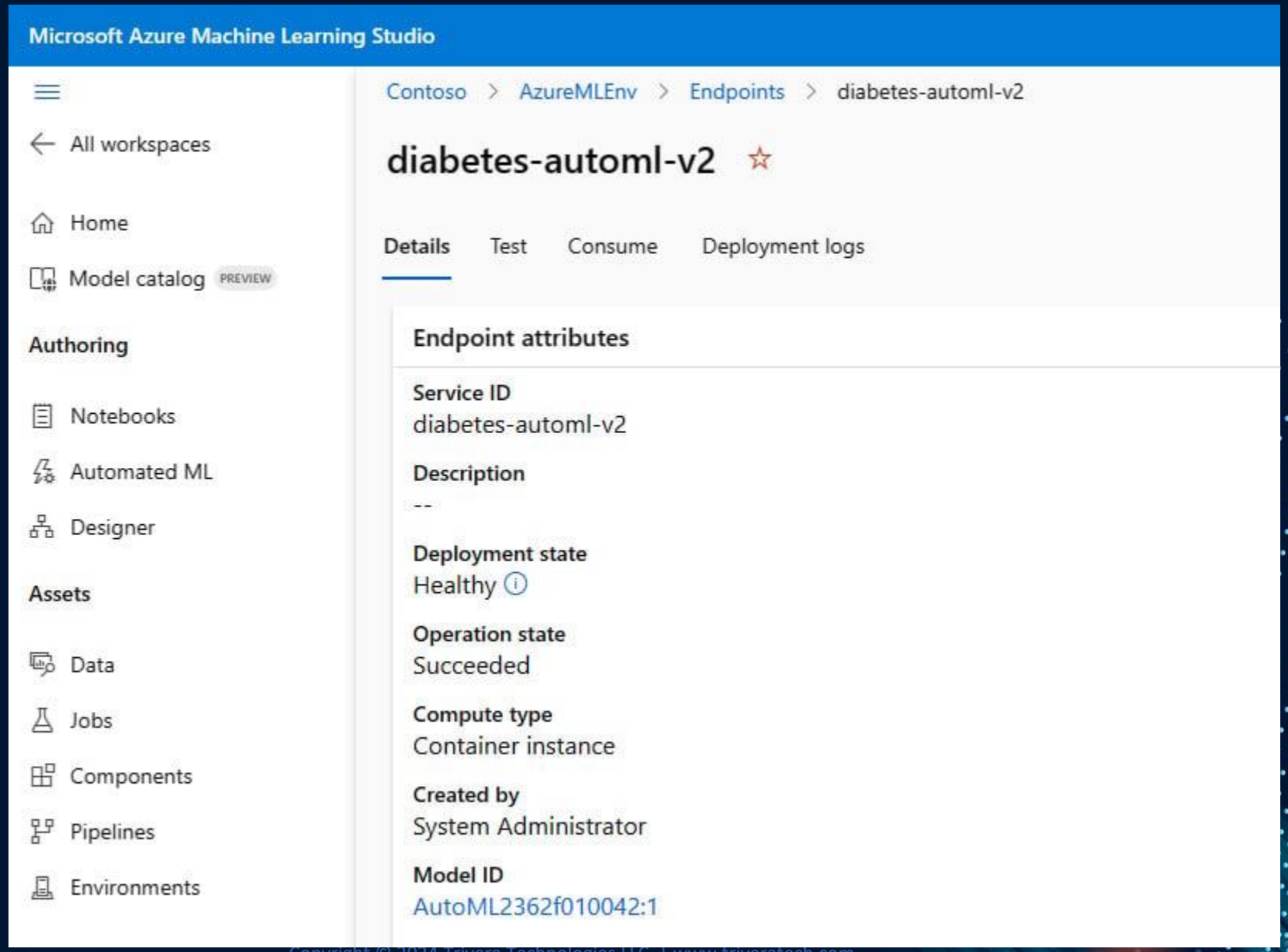


Training the Model in the Cloud





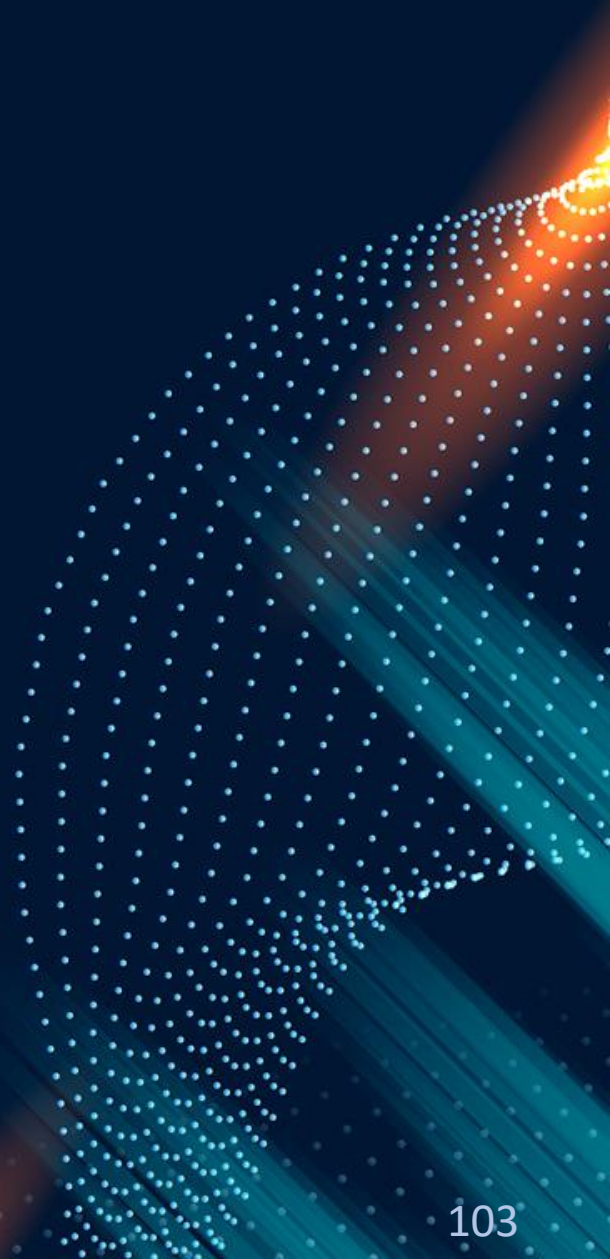
Make Predictions from the Deployed Model

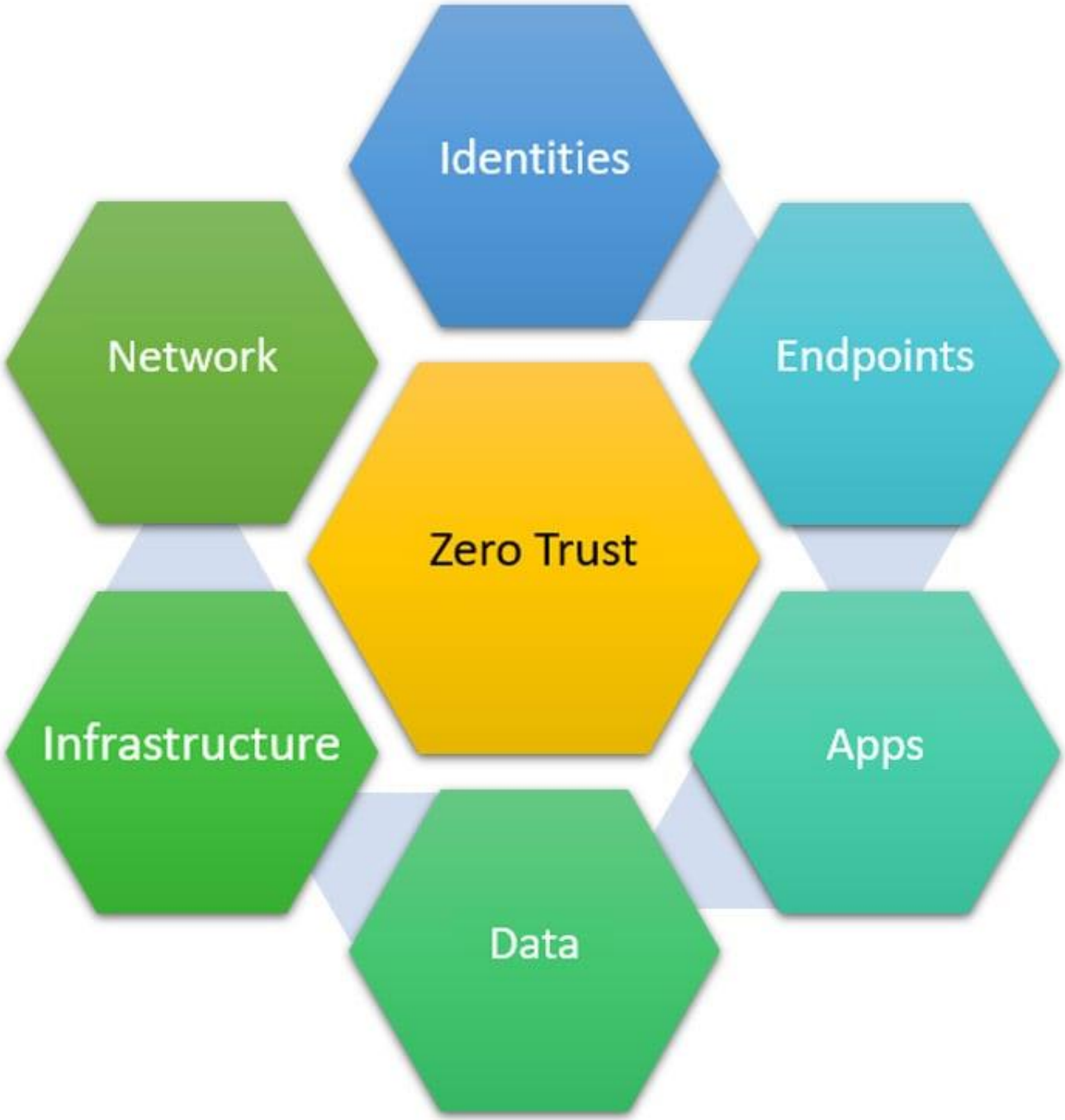


Labs 2,3,4

Zero Trust Model

- Verify Explicitly
- Use Least Privilege Access
- Assume Breach

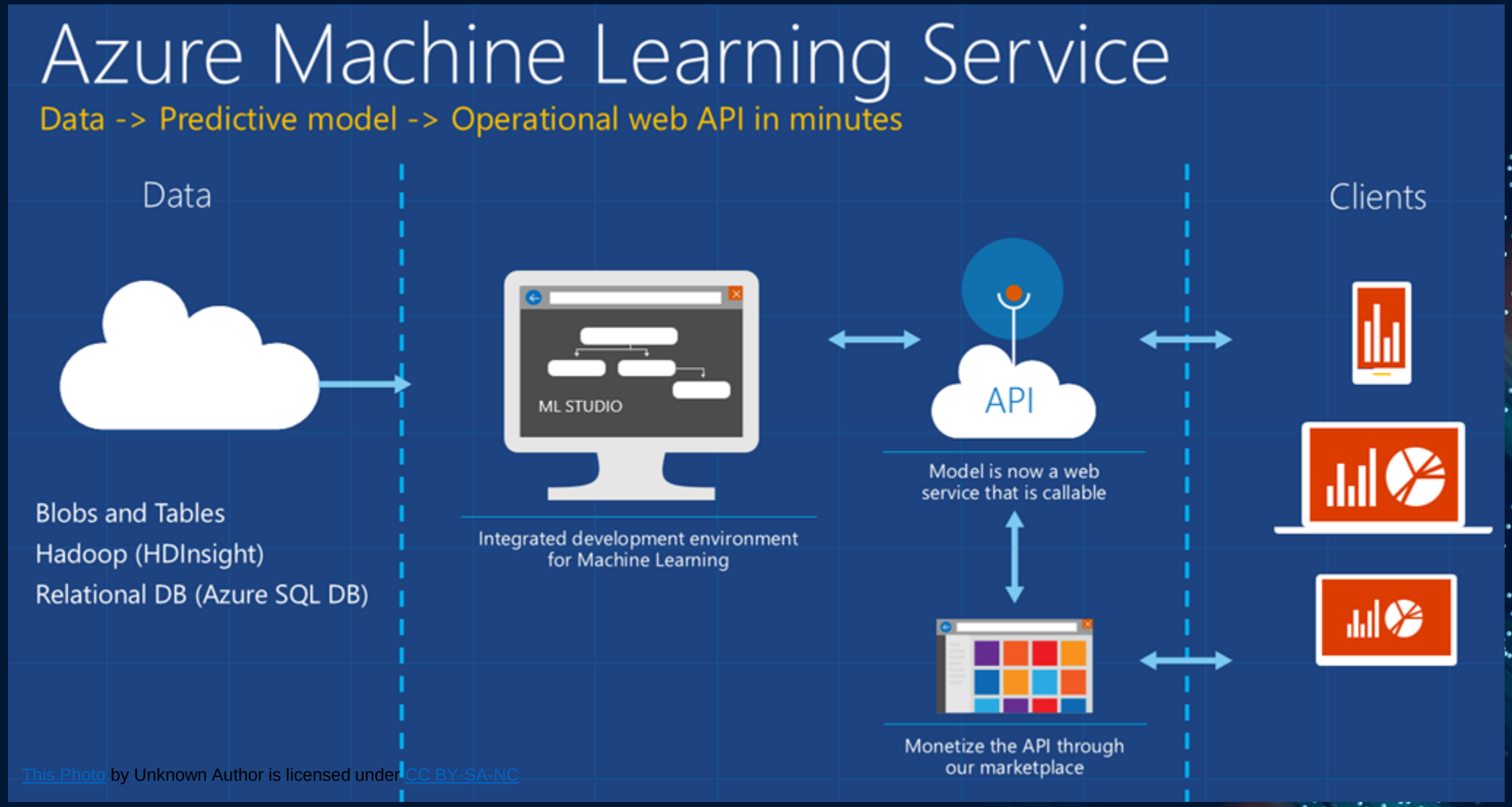



Identities
Network
Endpoints
Zero Trust
Infrastructure
Apps
Data

Assessing the vulnerability of ML assets and apps

1. Identity Management
2. Data and Data Sources
3. Infrastructure
4. Network and End-Points
5. Monitoring and Maintenance

AI/ML Applications



Lab 5

TriveraTech®
TECHNOLOGY TRAINING

Summary

Experience is

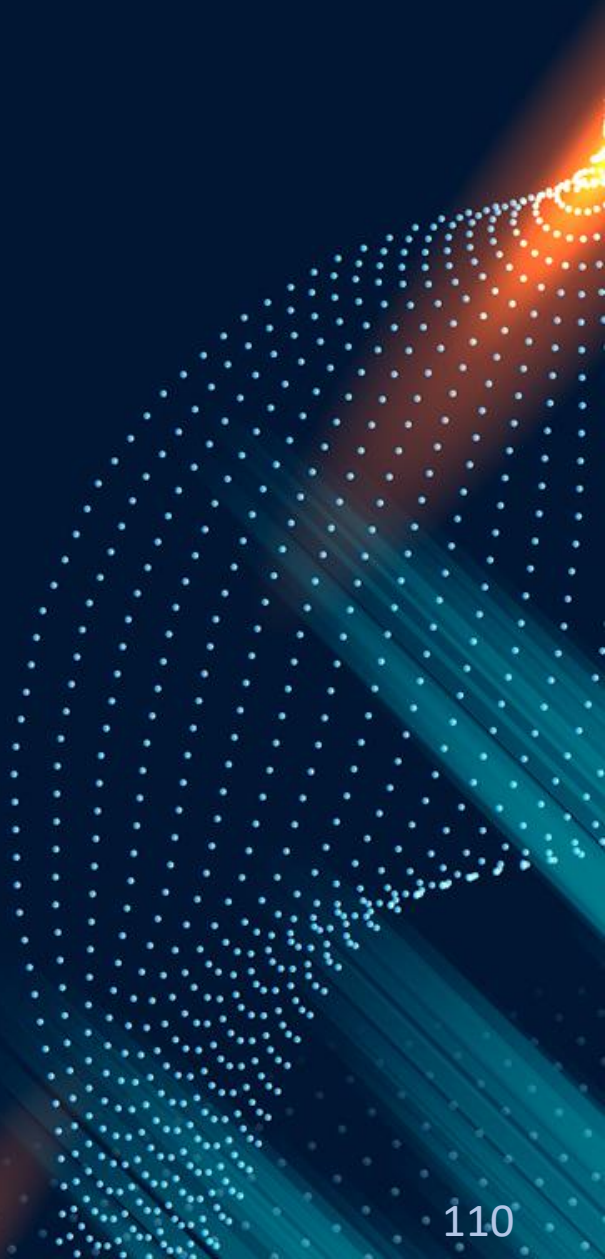
Chapter 6:

Crisis Averted: AI Incident Response Planning

Develop and Implement effective incident response plans for AI system breaches: Compliance and Regulatory

Compliance and Regulatory

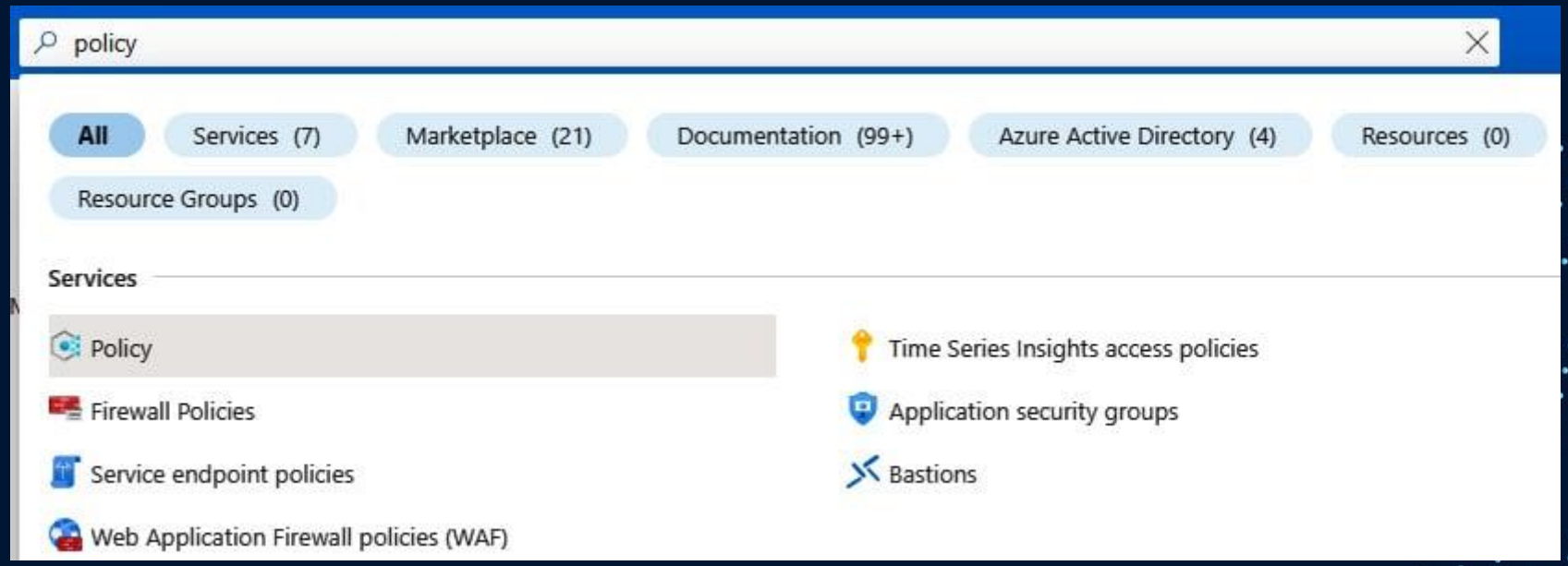
- Exploring Responsible AI development
- Regulatory compliance in Azure Policy for Azure Machine Learning
- Compliance auditing and reporting
- Compliance automation in Azure



Re

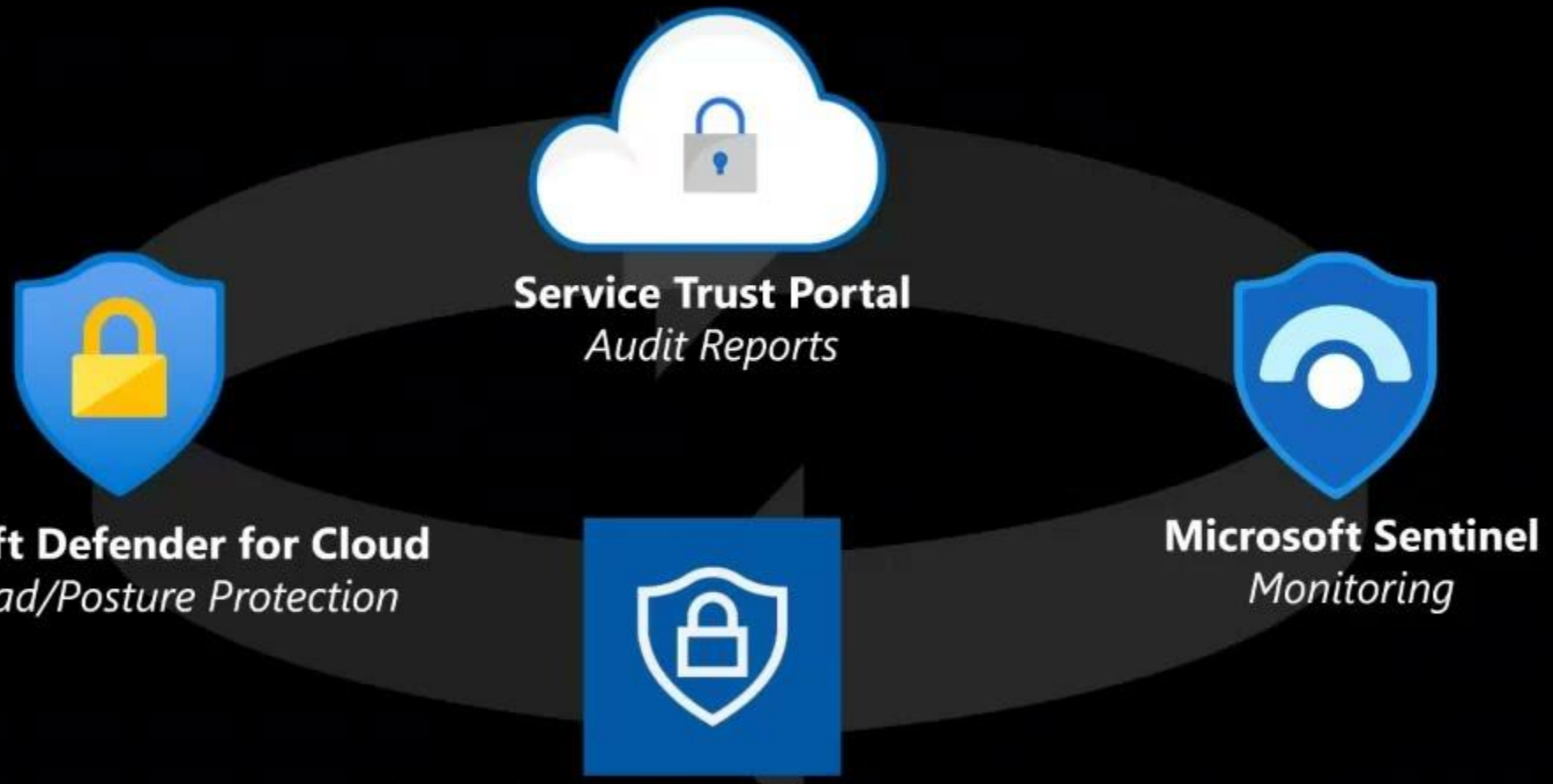
Responsible AI Principles

Regulatory Compliance in Azure Policy for Azure ML



Lab 6

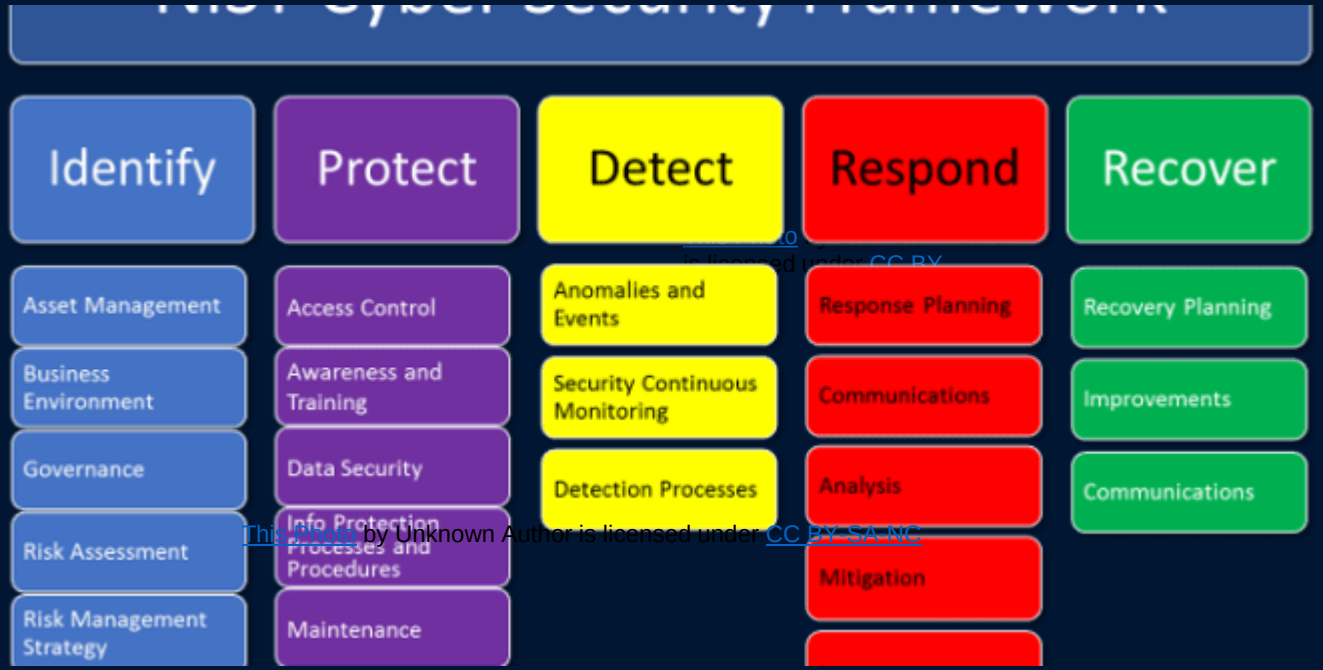
Microsoft Compliance Ecosystem



FedRAMP

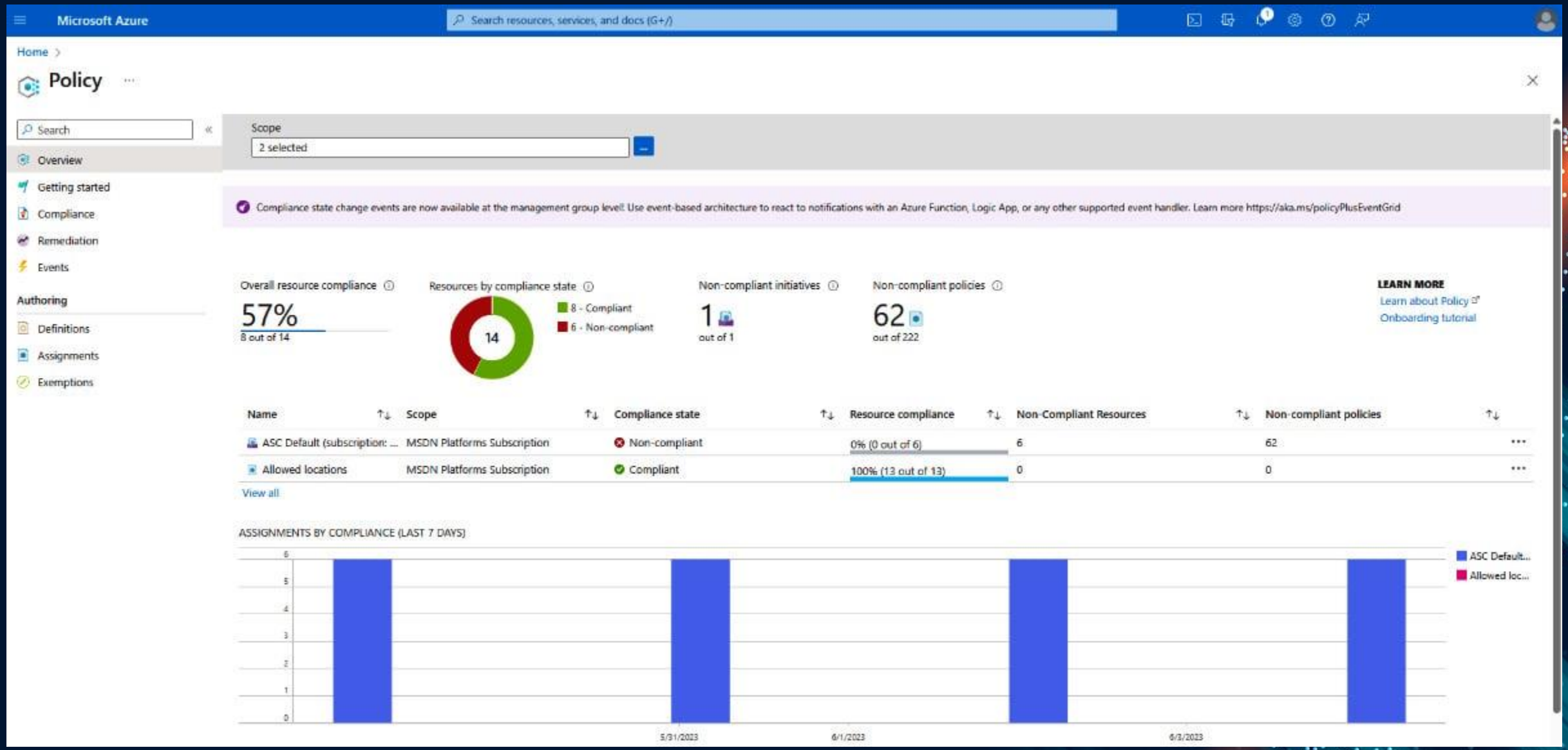
FedRAMP

NIST SP 800-53 rev. 5

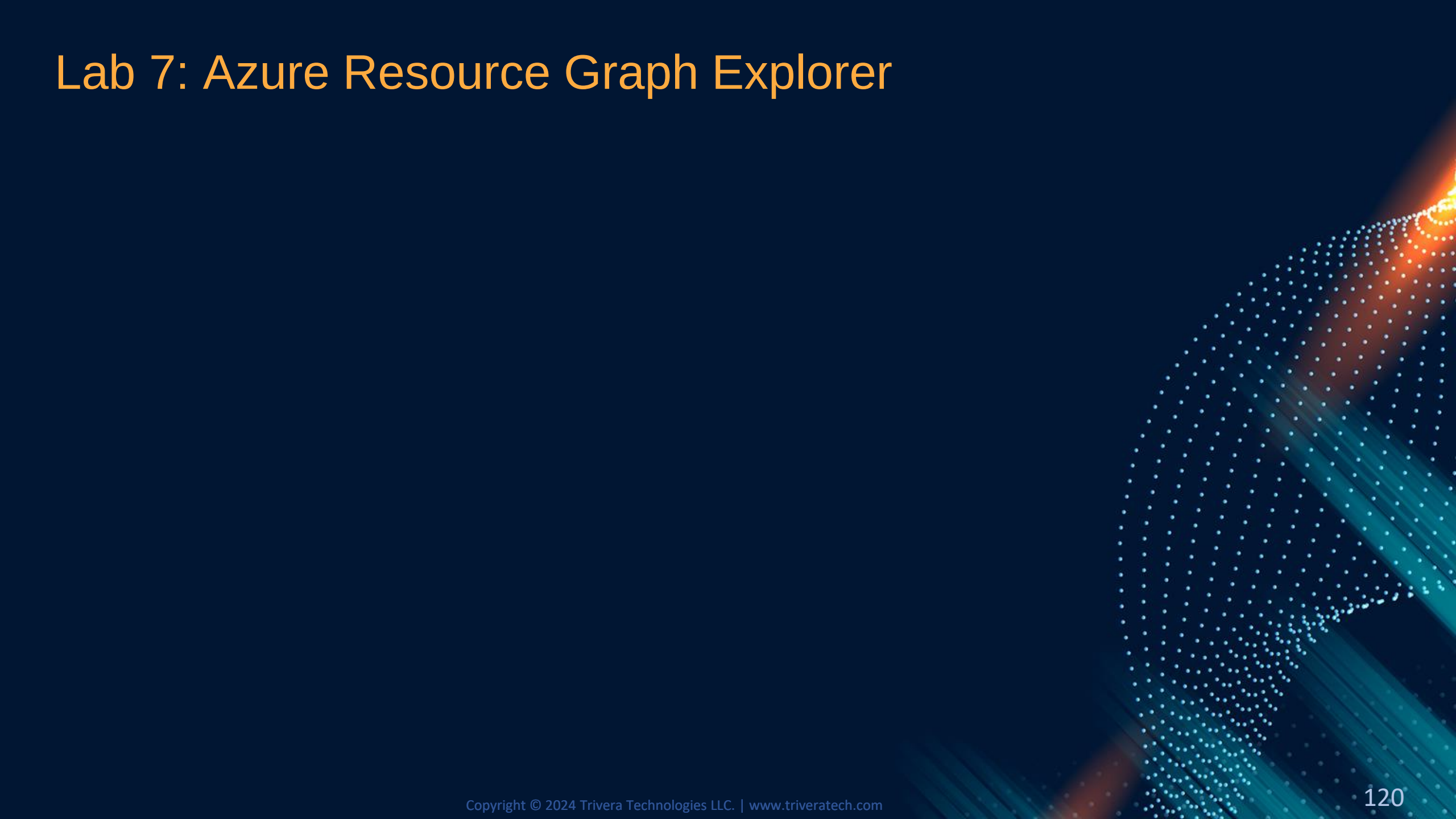



AUDIT

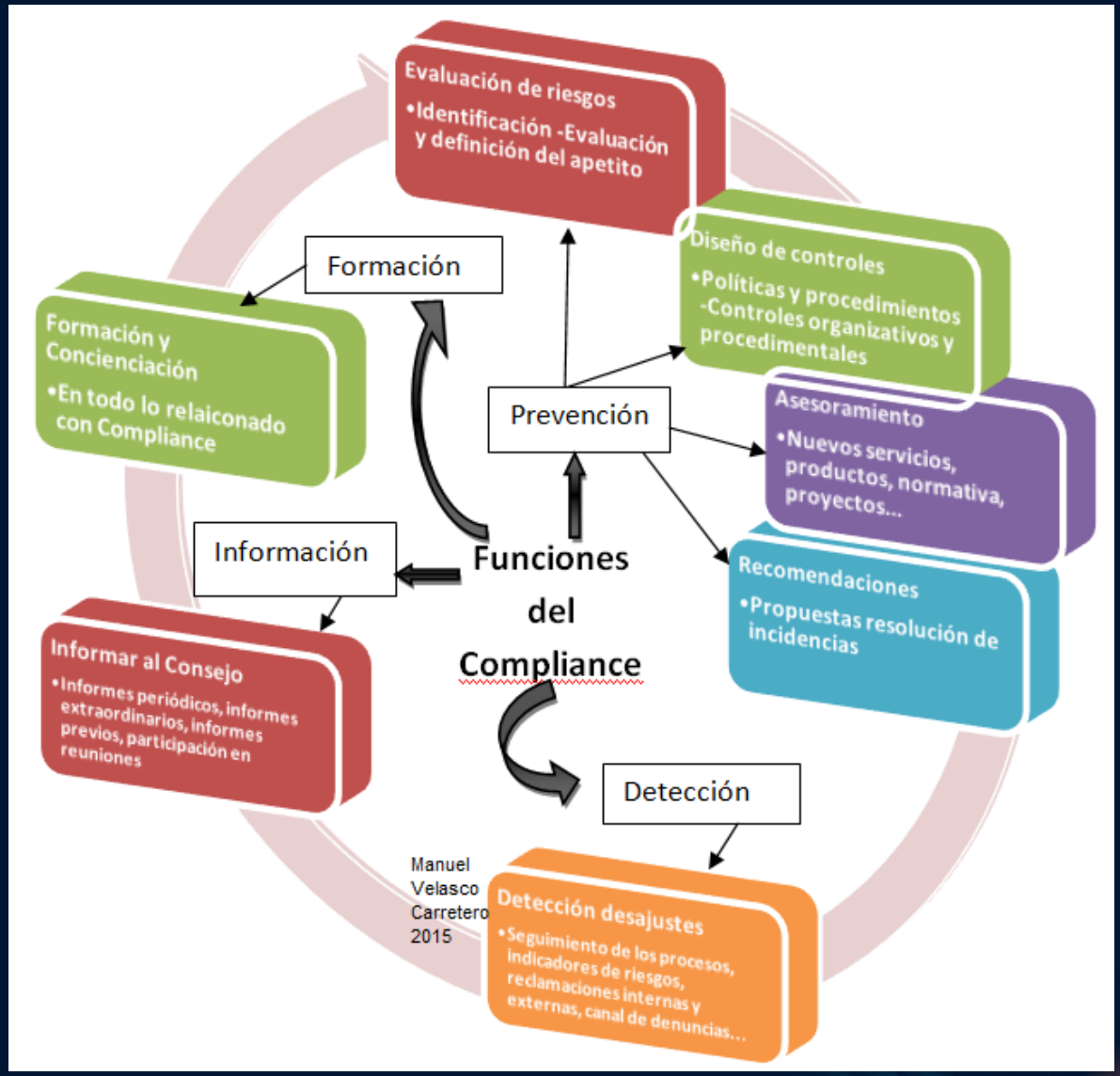
Azure Portal



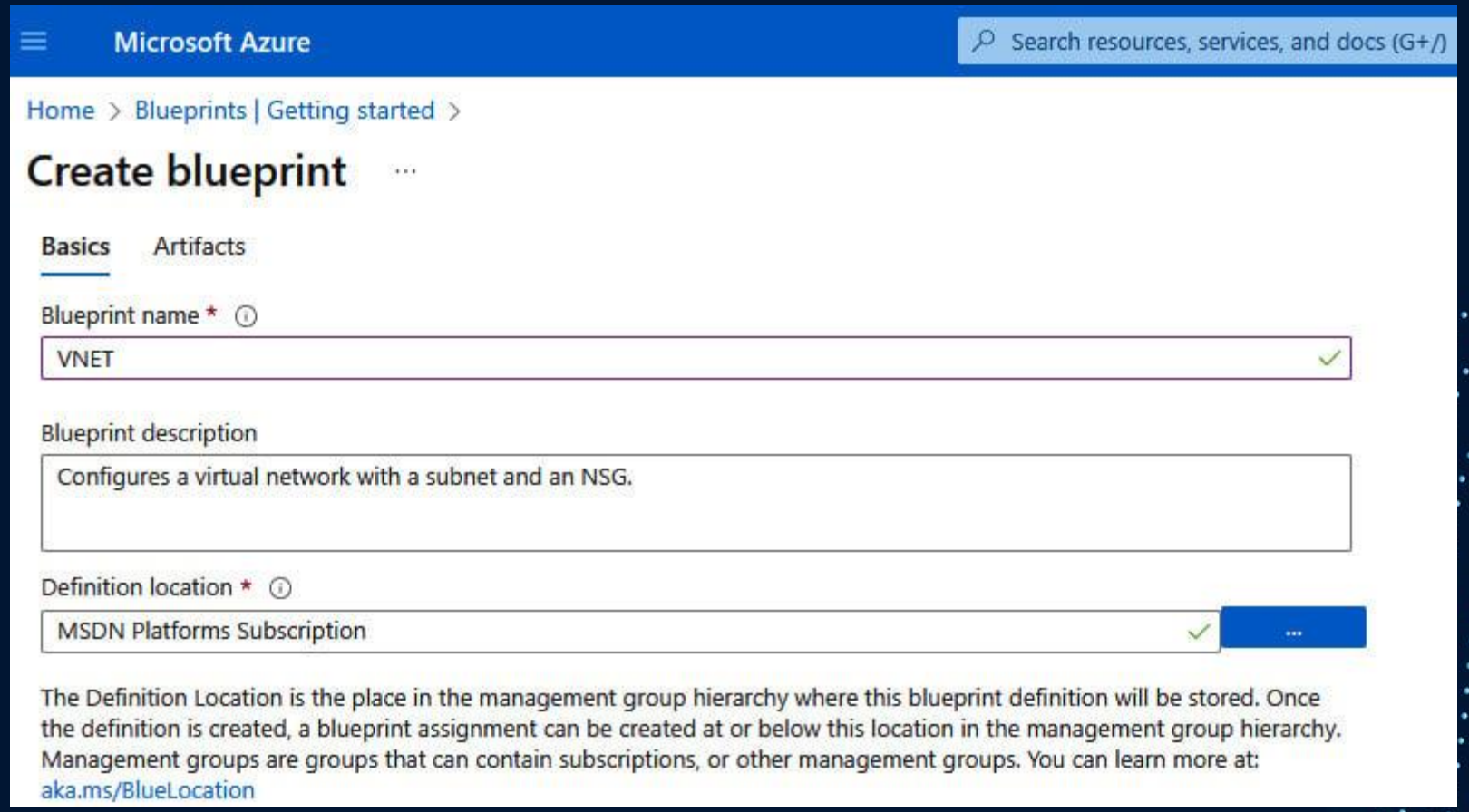
Lab 7: Azure Resource Graph Explorer



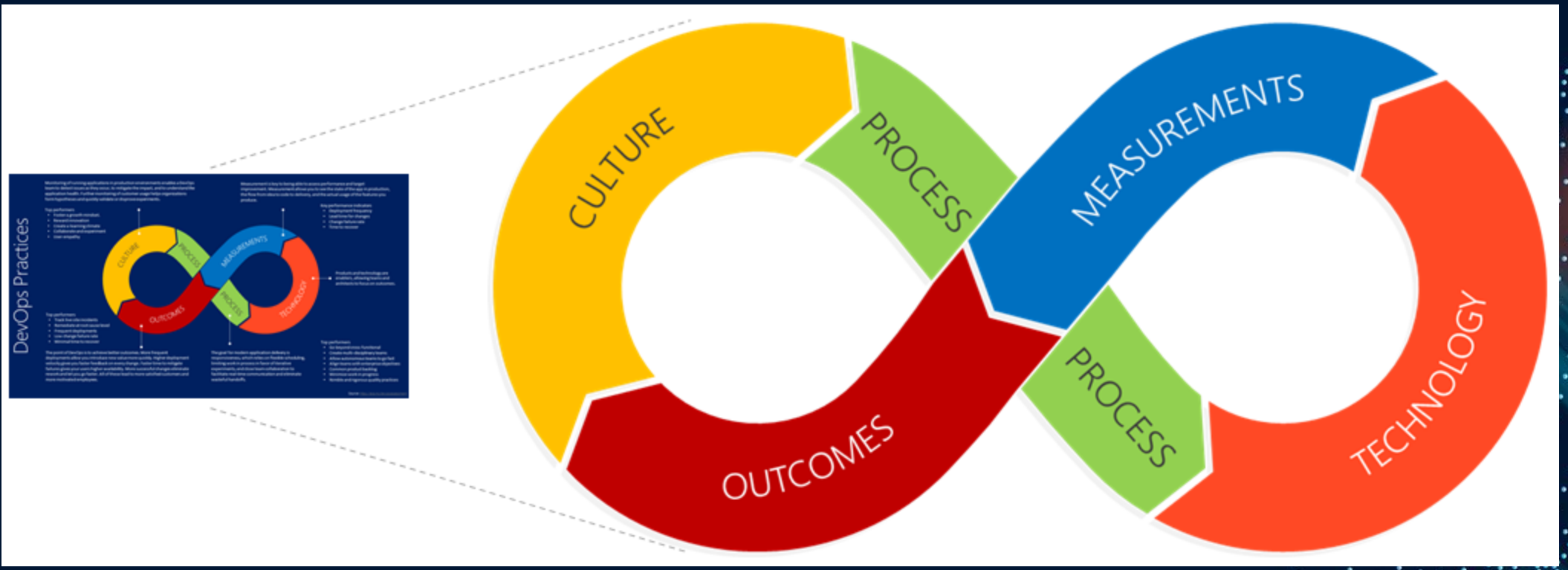
Compliance Automation



Azure Blueprints



IaC



Summary

Chapter 7:

AI Privacy & Ethical Considerations

Data Privacy and Responsible AI Best Practices

- Discovering and protecting sensitive data
- Introducing differential privacy
- Mitigating fairness
- Working with model interpretability
- Exploring FL and secure multi-party computation

Lab 9

Discovering and protecting sensitive data

- **Personally identifiable information (PII)**: Information that can be used to identify an individual, such as full name, date of birth, social security number, driver's license number, passport number
- **Financial information:** Credit card numbers, bank account details, financial transaction records, and so on
- **Health information**: Medical records, health insurance information, mental health records, and other health-related data
- **Passwords and authentication data**: Usernames, passwords, security questions, or any other credentials used to access systems or accounts
- **Biometric data**: Fingerprints, retinal scans, facial recognition data, and other biometric identifiers
- **Confidential business information**: Trade secrets, intellectual property (IP), financial reports, customer lists, proprietary algorithms, and so on
- **Government classified information**: Information classified by governments for national security reasons
- **Personal communications**: Private messages, emails, and other communications that individuals expect to be confidential
- **Social and demographic information**: Race, ethnicity, religion, sexual orientation, and other sensitive demographic data
- **Geolocation data**: Precise location data of individuals or assets

Data Anonymization

- Removing direct identifiers
- Pseudonymization
- Data masking / Tokenization
- Generalization / Aggregation

Differential Privacy

Chapter 8:

What's Next?

Preparing for Future AI Security Challenges: Managing and Securing Access

Lesson Agenda: What We Will Cover

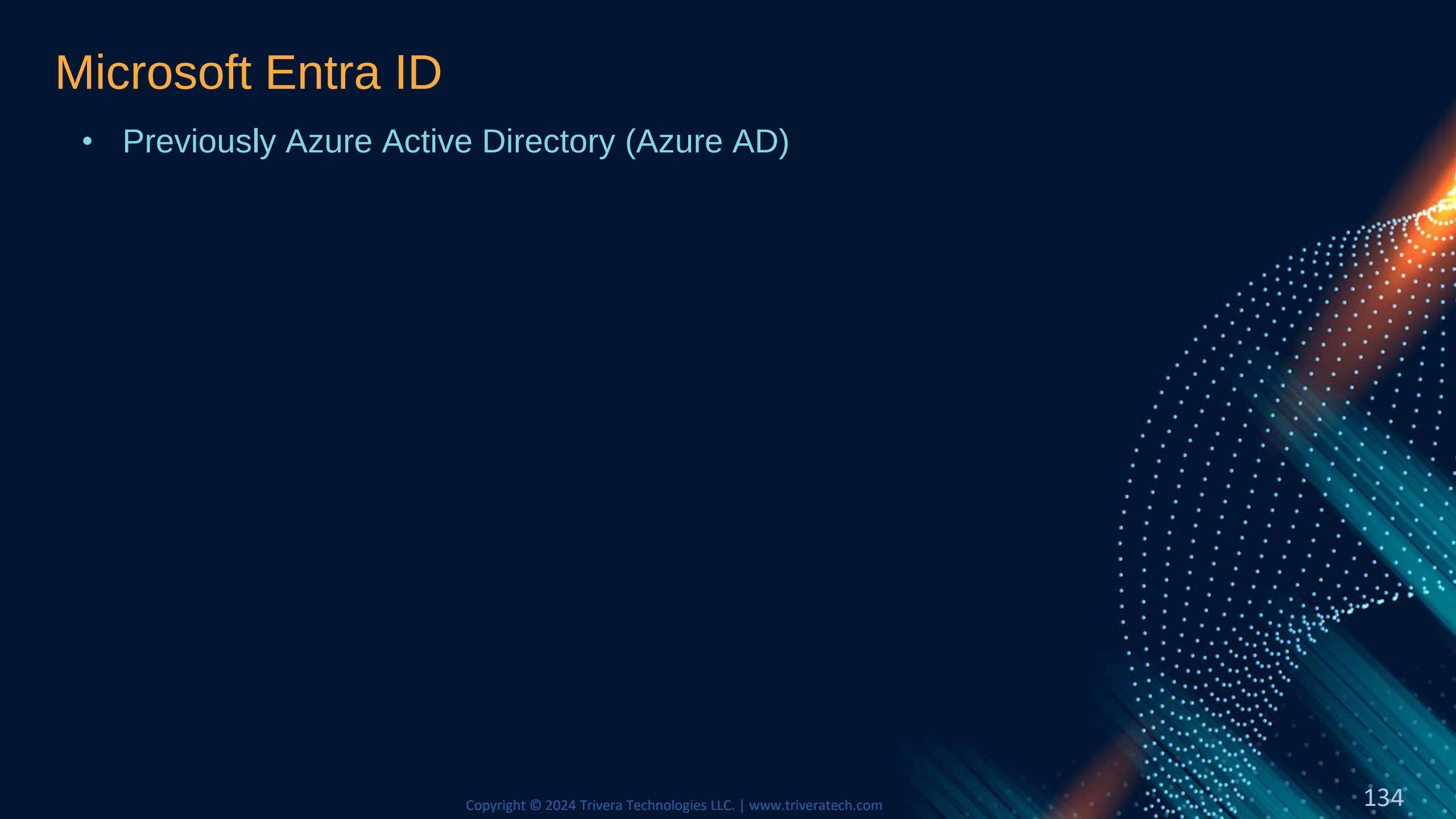
- Working with the PoLP
- Authenticating with Microsoft Entra ID
- Implementing RBAC
- Authenticating with application identities
- Enhancing access security

PoLP

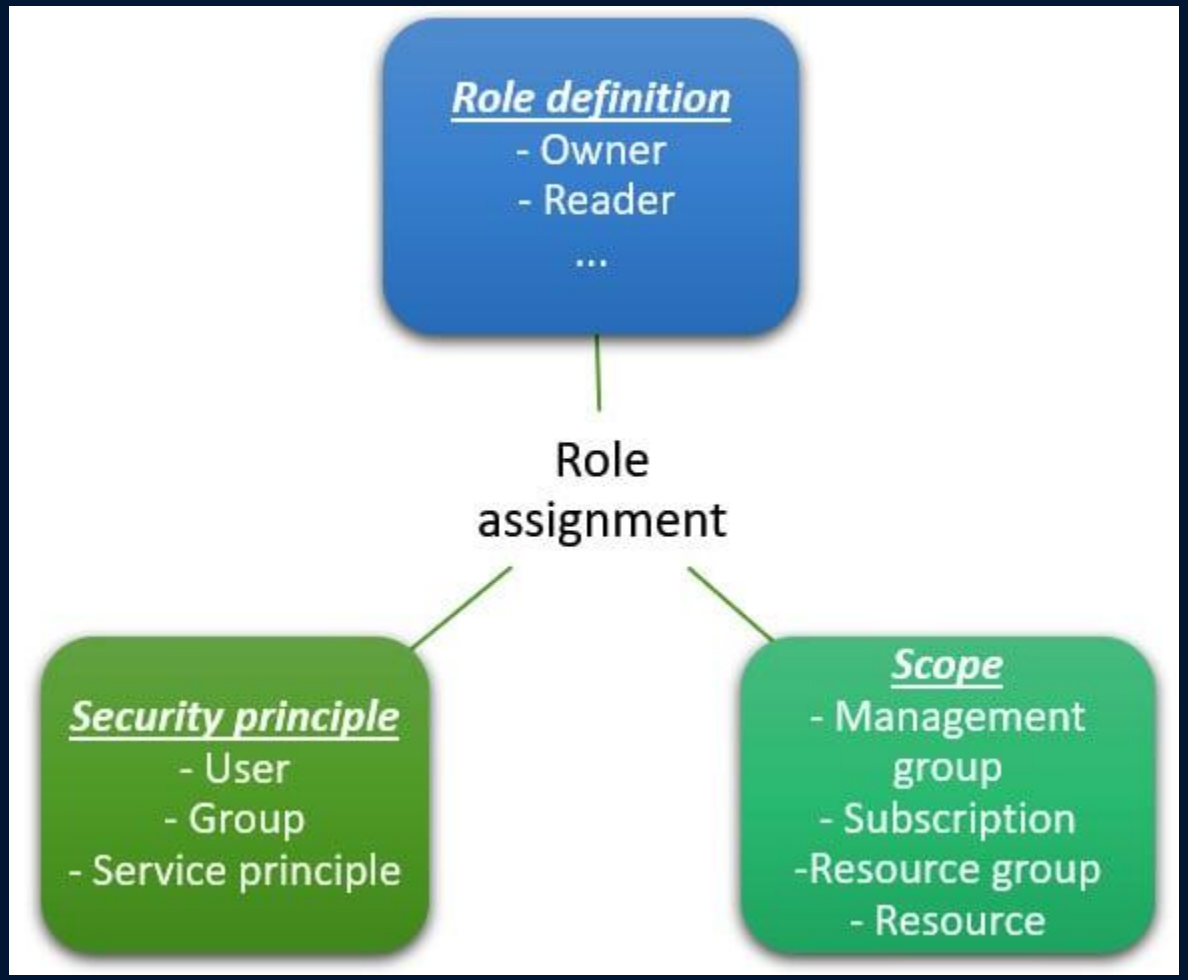
- Application Inventory
- RBAC
- Regular reviews
- Automated tools
- Zero Trust Architecture
- Continuous Monitoring
- Training!

Microsoft Entra ID

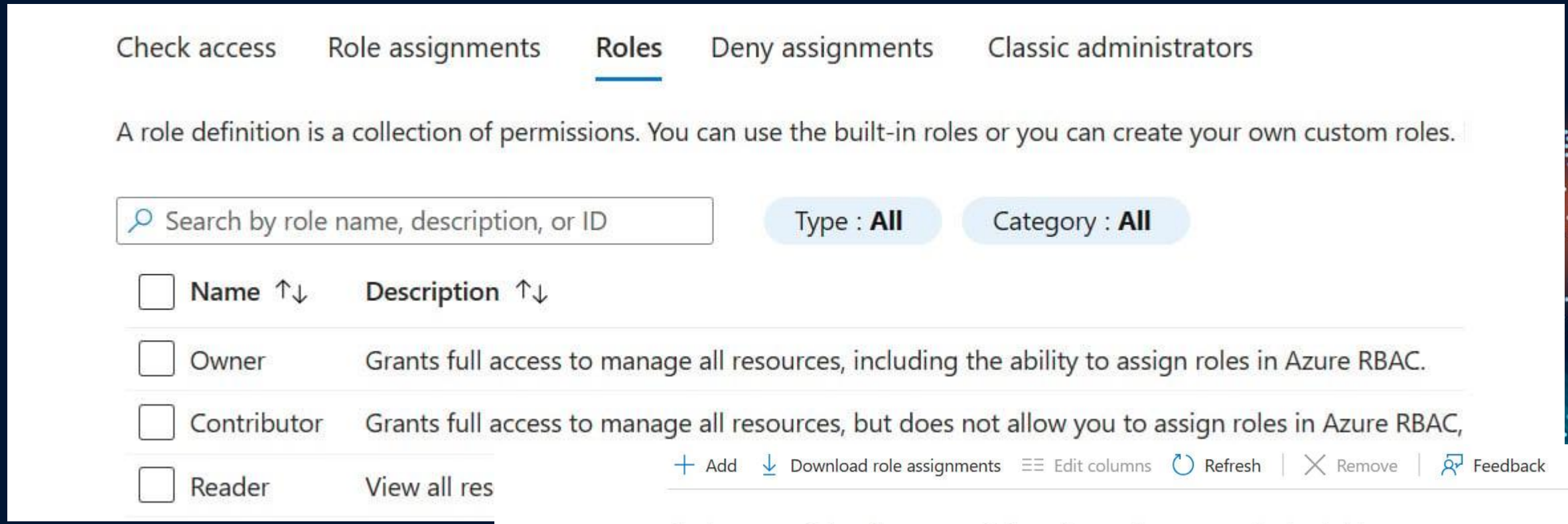
- Previously Azure Active Directory (Azure AD)

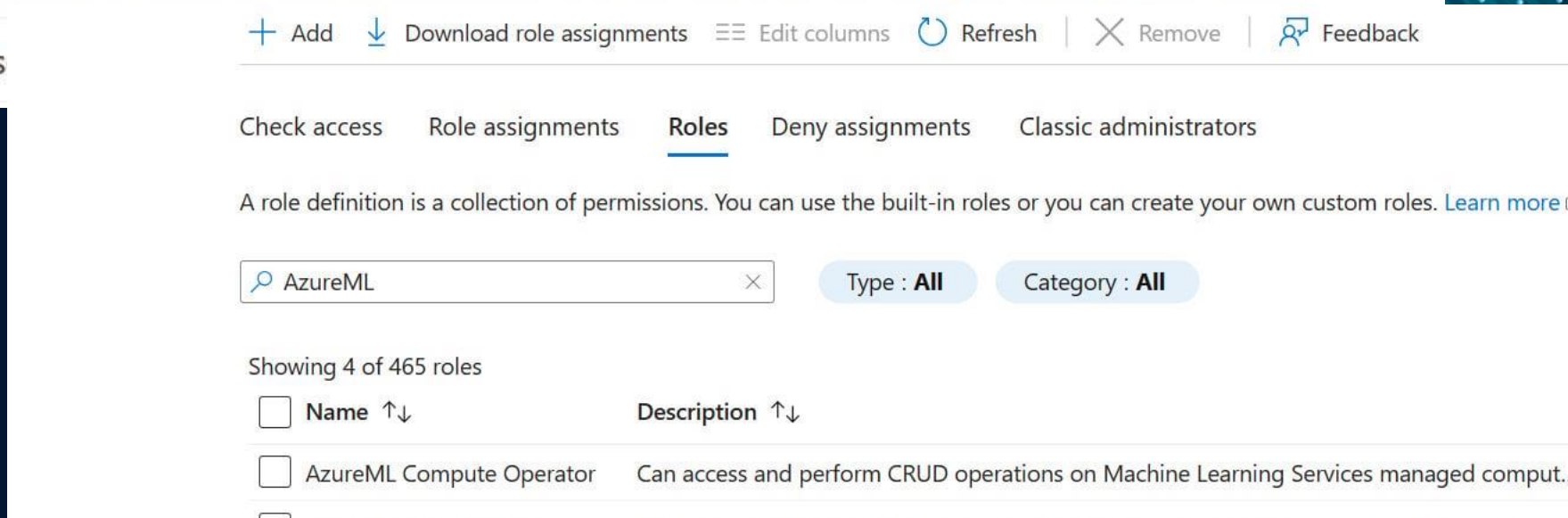


RBAC

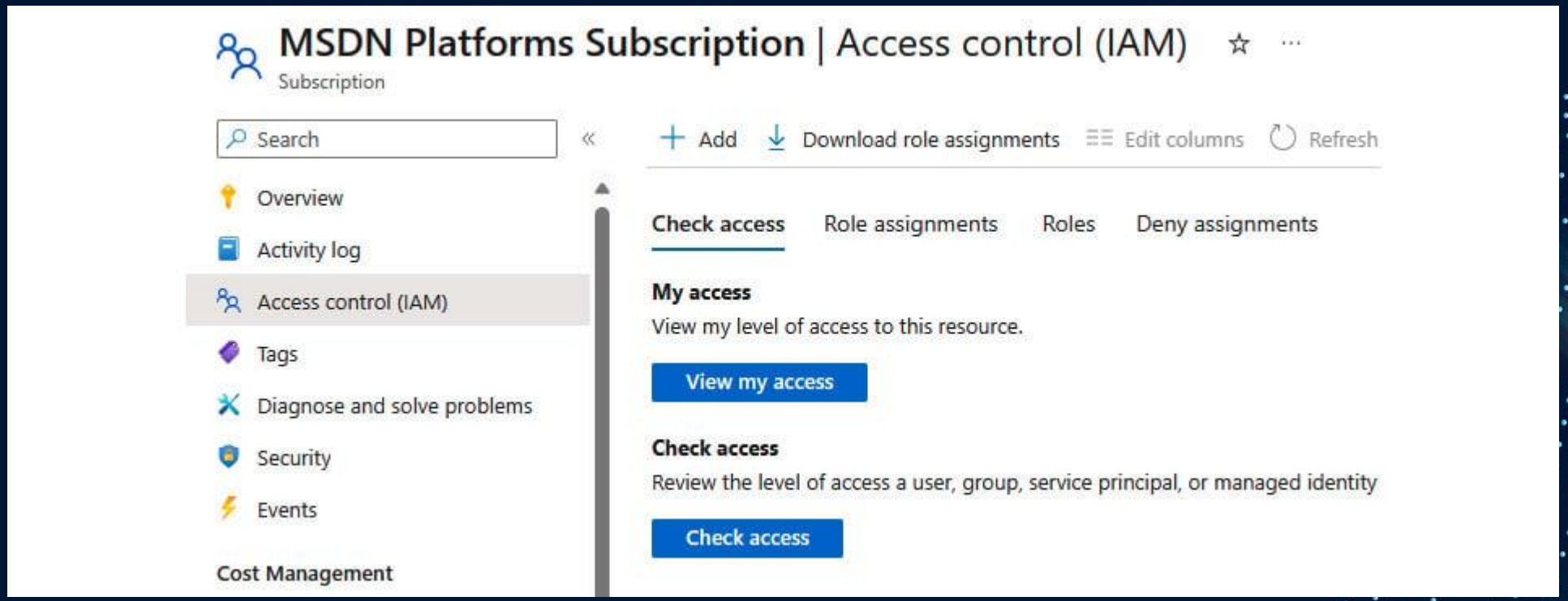


Built in Roles





Add a Role Assignment



Lab 11, 12

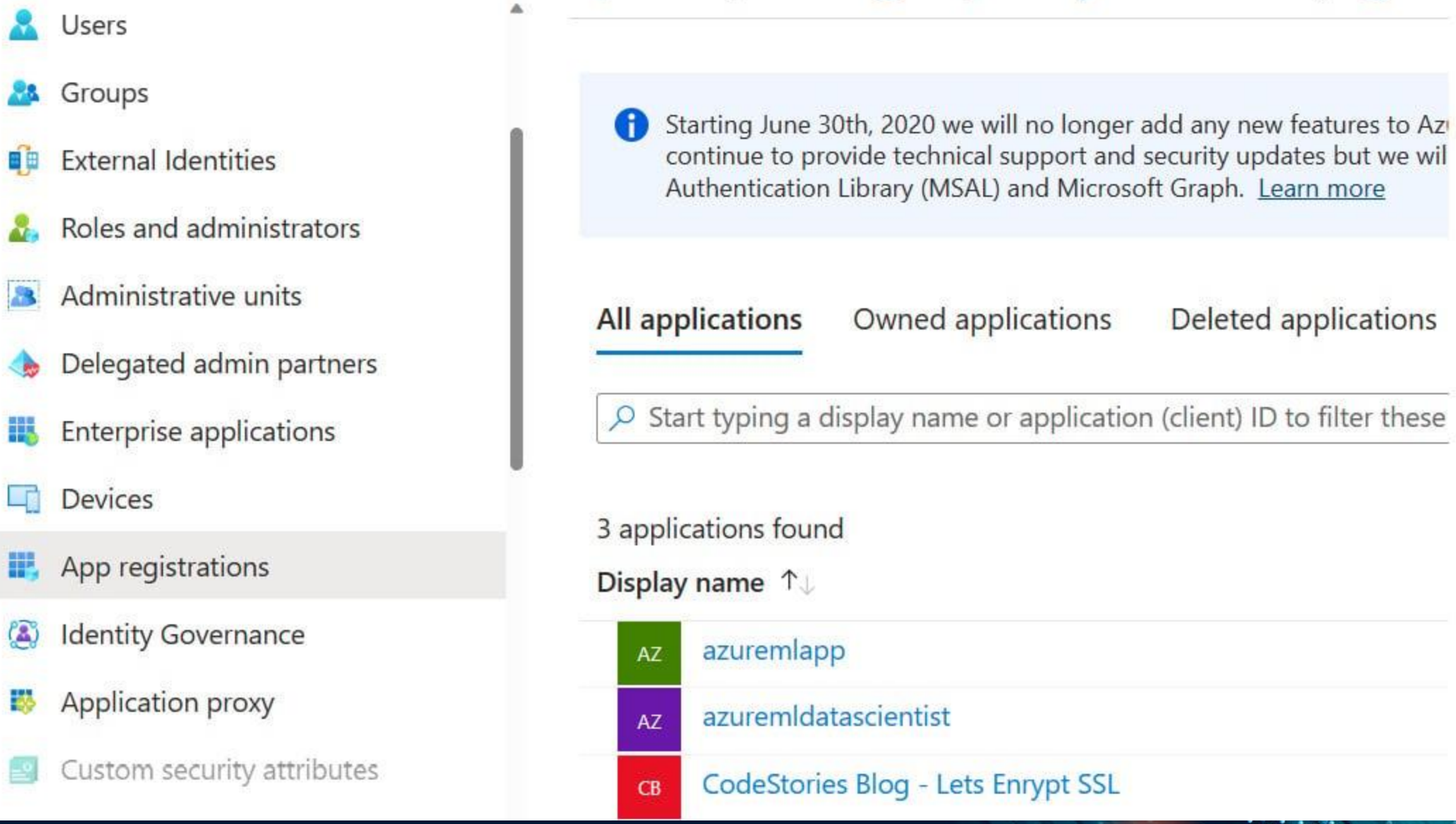
Create a Service Principle

- In order to authenticate when working with applications or scripts that train and test a model, for example, you can use service principal authentication. To achieve this, we need to create a service principal in the Microsoft Entra ID workspace. If you are creating a service principal, you need to create an application registration.

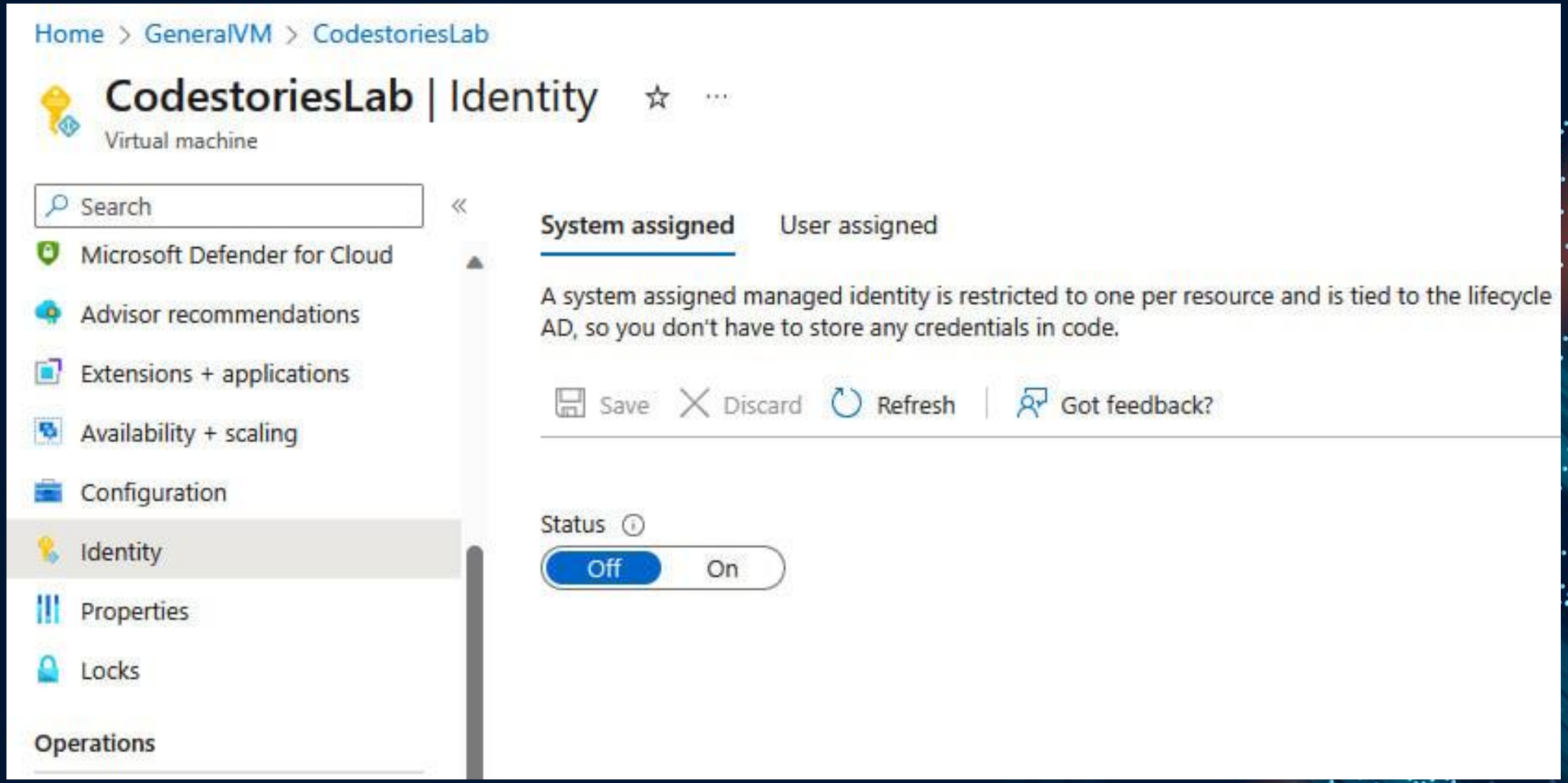
```
georgia [ ~ ]$ az ad sp create-for-rbac --sdk-auth --name azuremldatascientist --role Reader --scopes /subscriptions/
Option '--sdk-auth' has been deprecated and will be removed in a future release.
Creating 'Reader' role assignment under scope '/subscriptions/
The output includes credentials that you must protect. Be sure that you do not include these credentials in your code
{
  "clientId": "
  "clientSecret": "
  "subscriptionId": "
  "tenantId": "
  "activeDirectoryEndpointUrl": "https://login.microsoftonline.com",
  "resourceManagerEndpointUrl": "https://management.azure.com/",
  "activeDirectoryGraphResourceId": "https://graph.windows.net/",
  "sqlManagementEndpointUrl": "https://management.core.windows.net:8443/",
  "galleryEndpointUrl": "https://gallery.azure.com/",
  "managementEndpointUrl": "https://management.core.windows.net/"
}
```

Default Directory | App registrations

Microsoft Entra ID



Managed Identities



Conditional Access

- Requiring for multi-factor authentication (MFA) to be enabled for users with highly privileged roles or from unrecognized locations
- Blocking users that are flagged by the system as high risk
- Allowing or blocking access from specific IPs or countries
- Requiring a password change to allow access

Summary

Chapter 9:

Managing Your Cloud Workspace

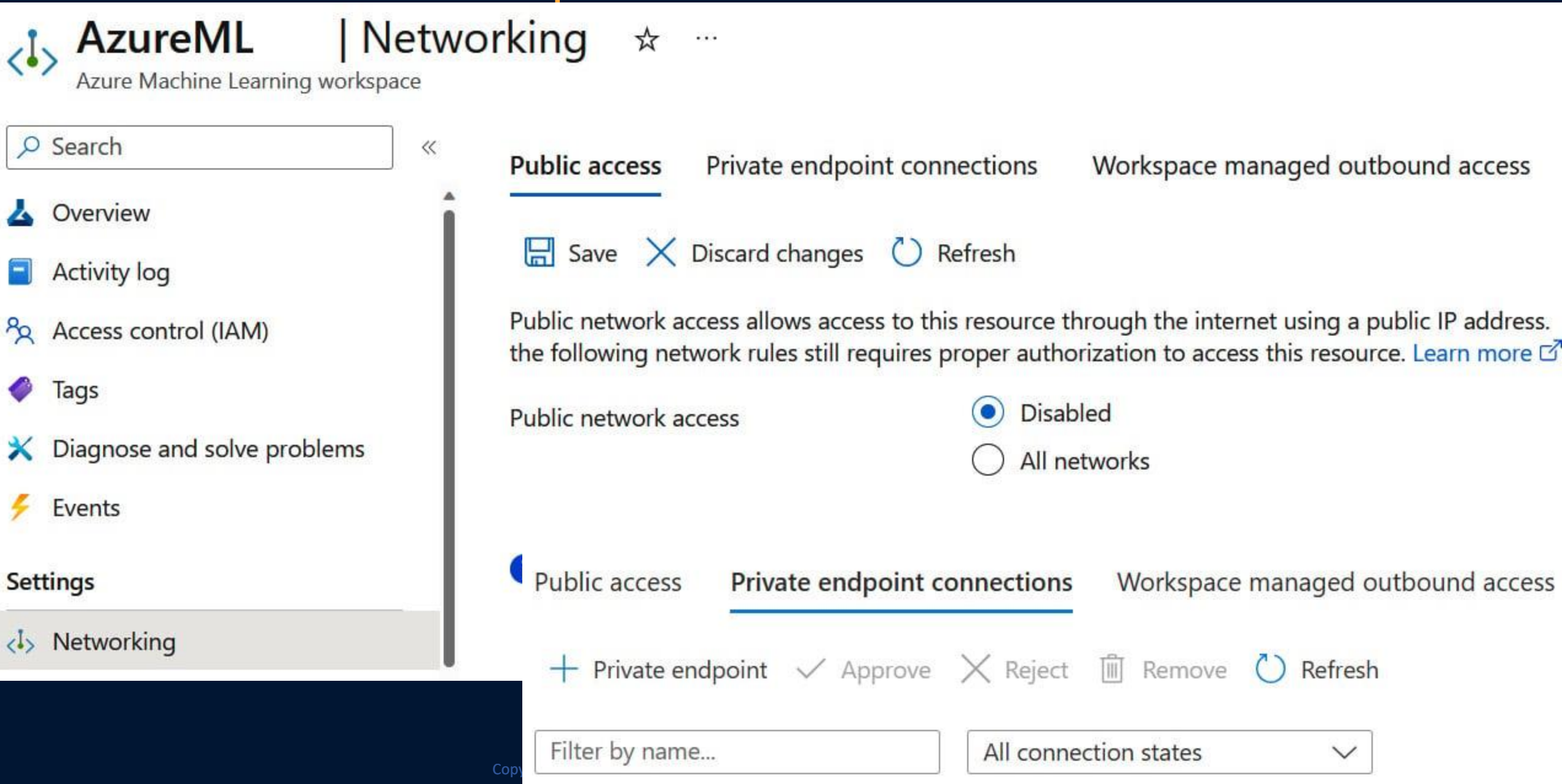
Azure ML Workspace

Lesson Agenda: What We Will Cover

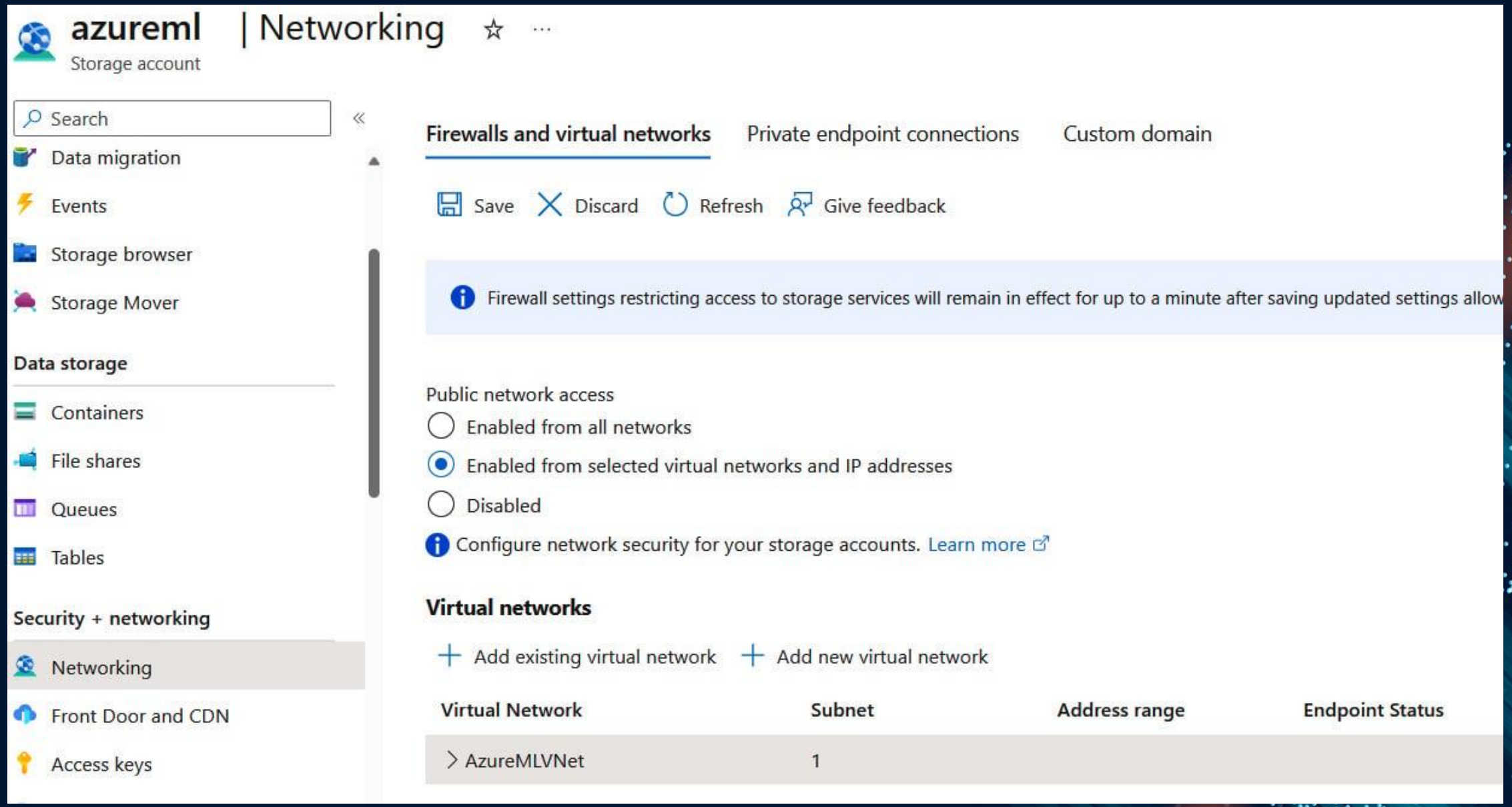
- Exploring network security
- Working with Azure Machine Learning compute
- Managing container registries and containers

ML Network Security

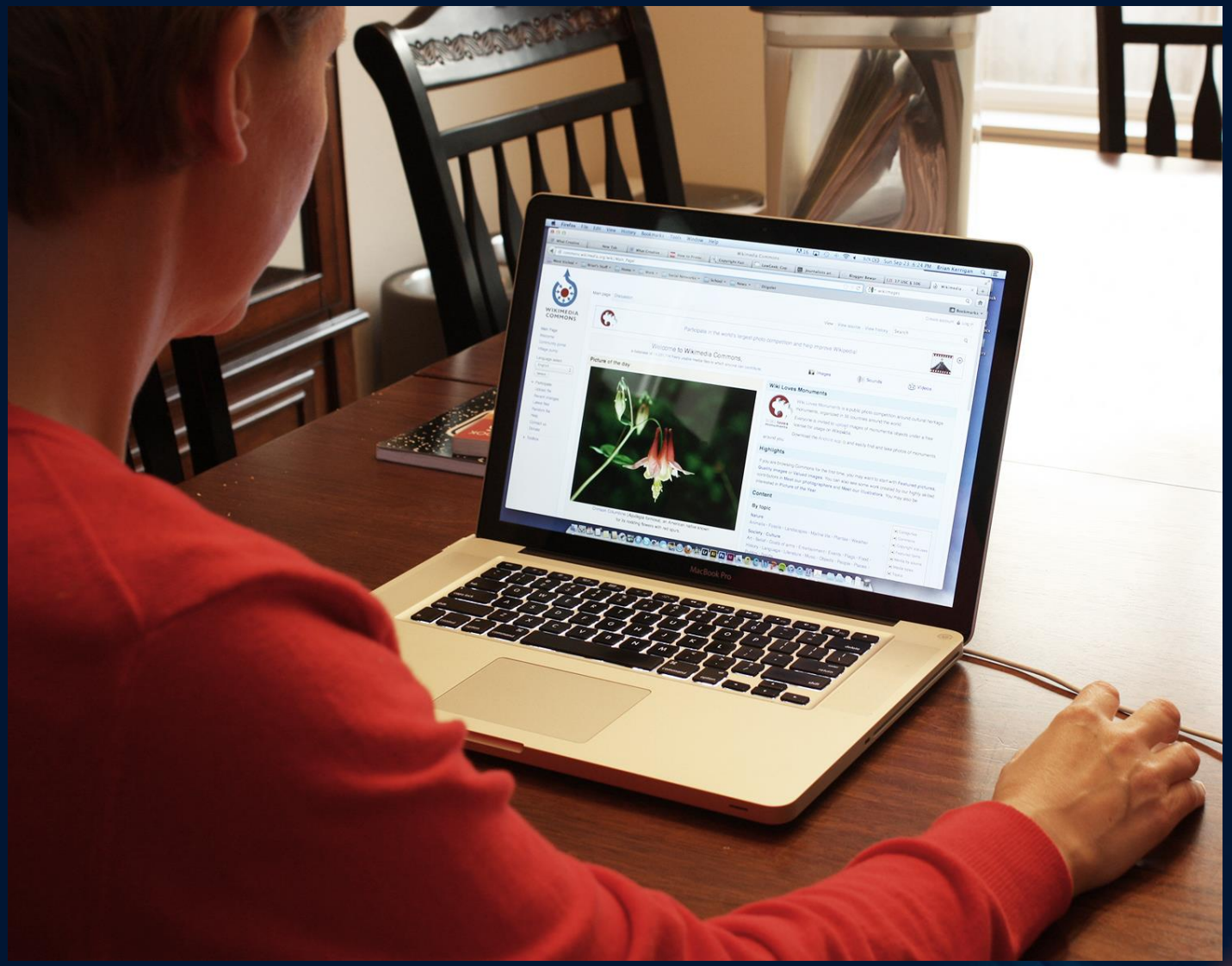
Secure the ML Workspace



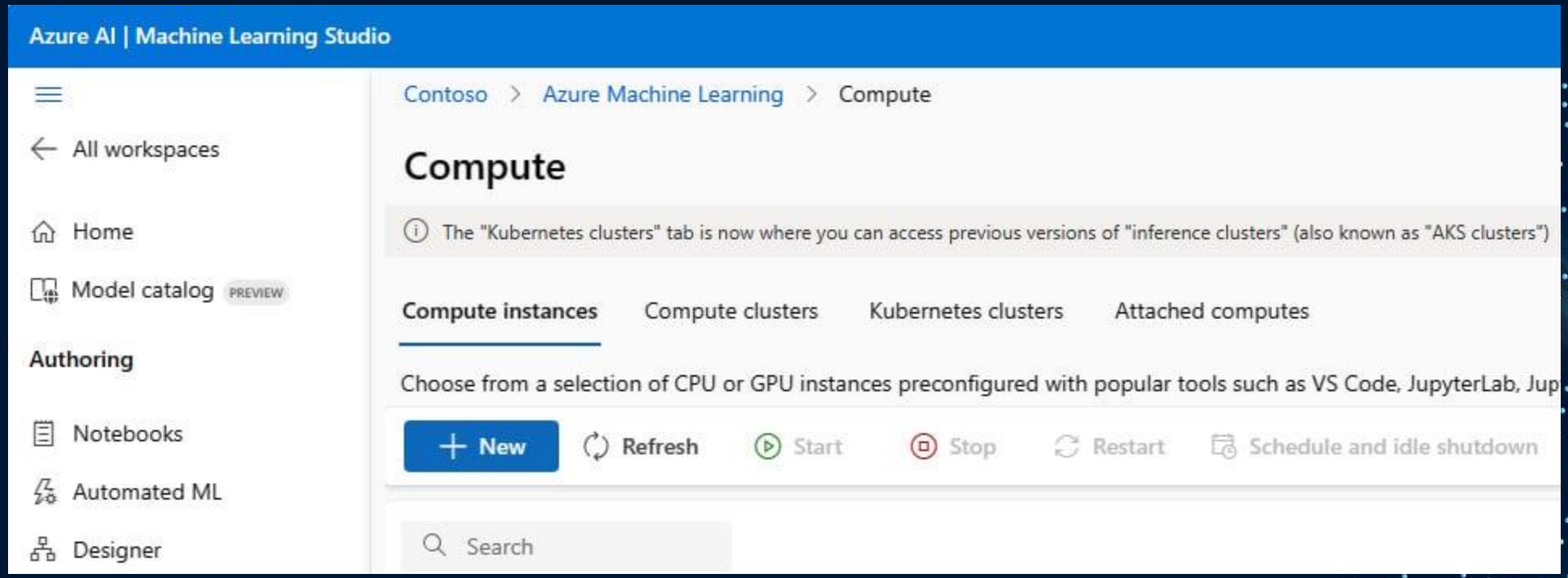
Securing Associated Resources



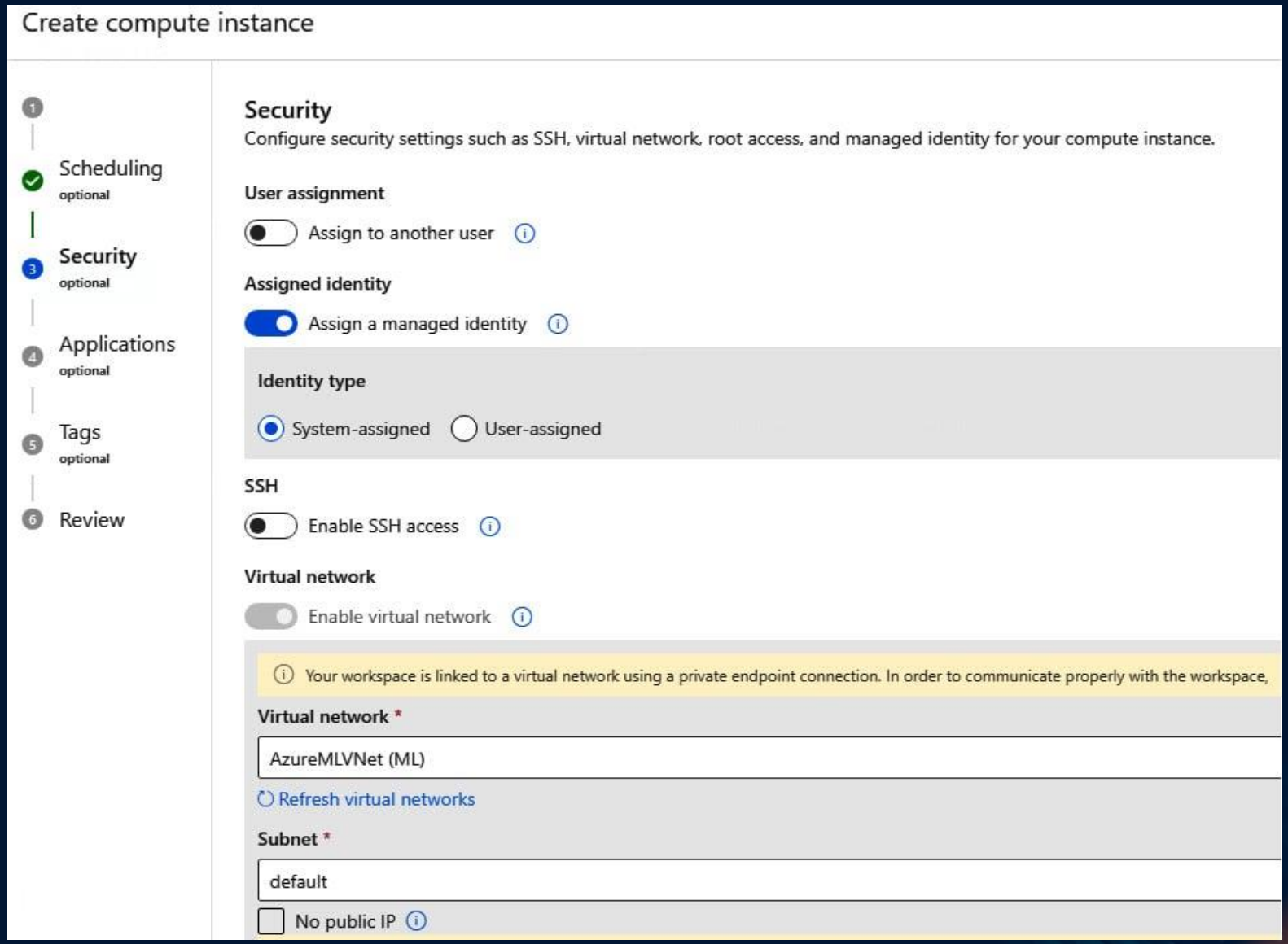
Validate Access



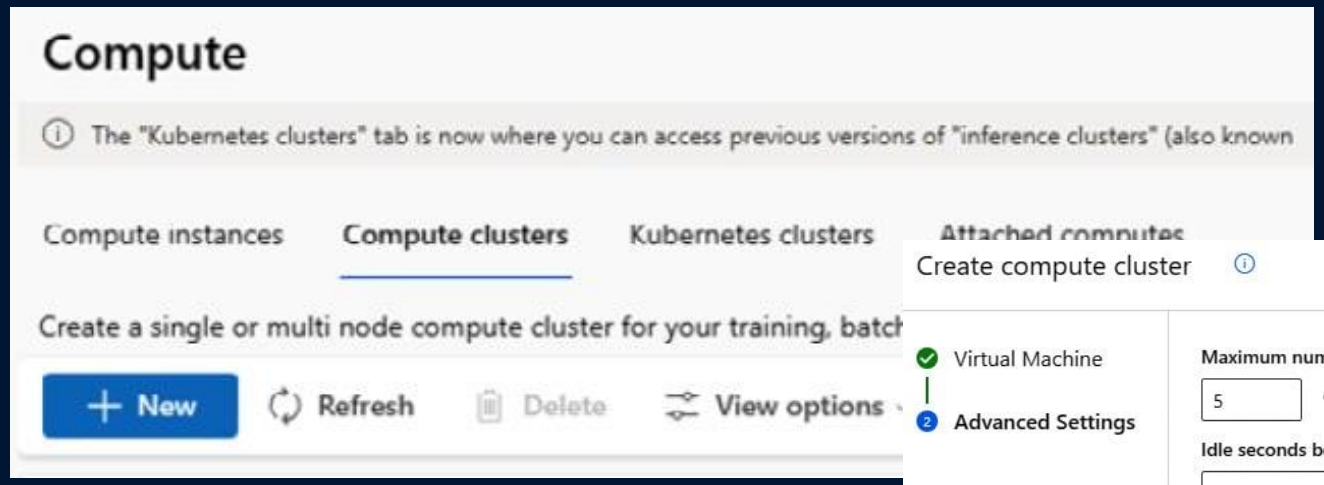
Working with Azure ML Compute

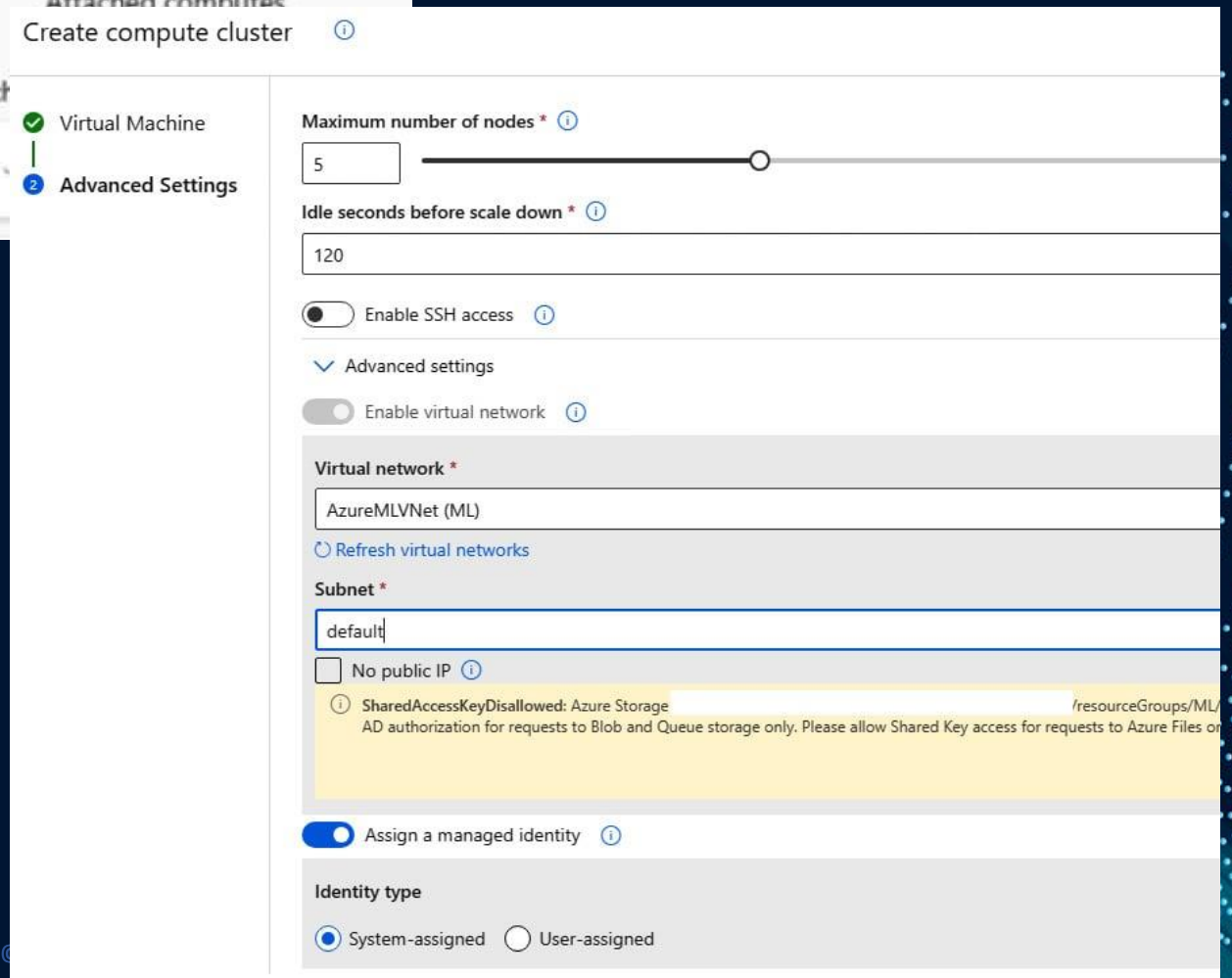


Secure Compute Instances

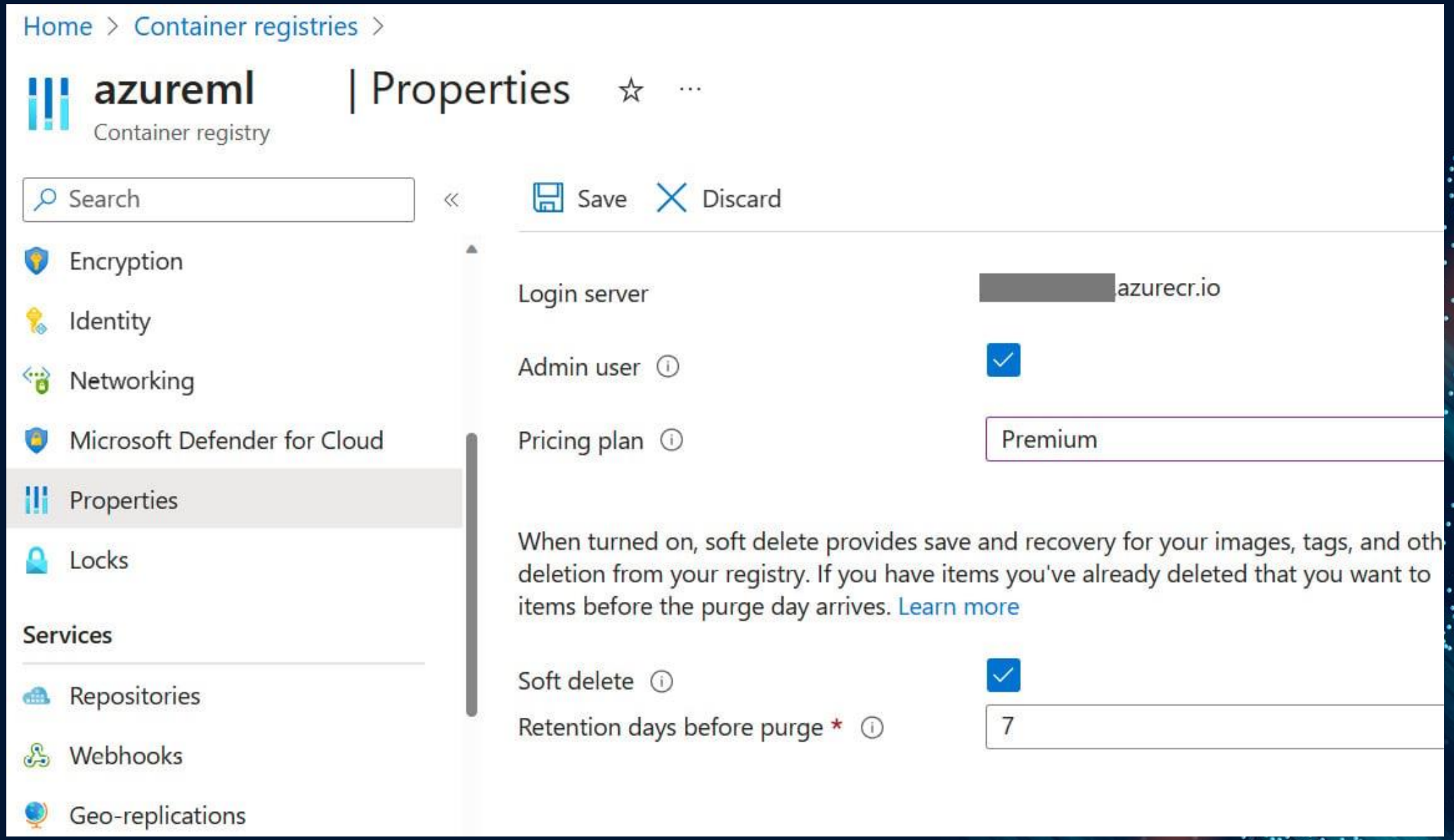


Secure Compute Clusters

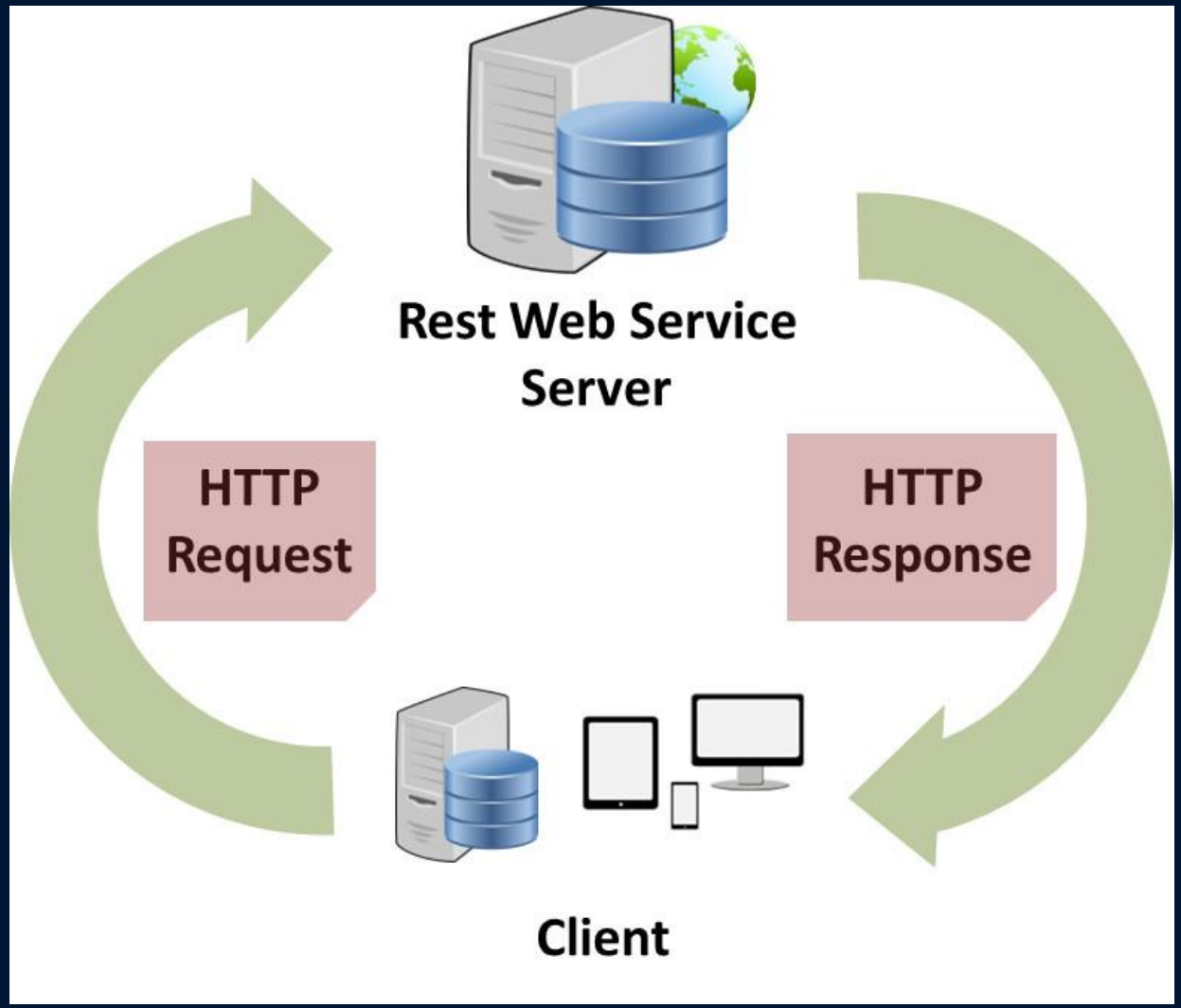




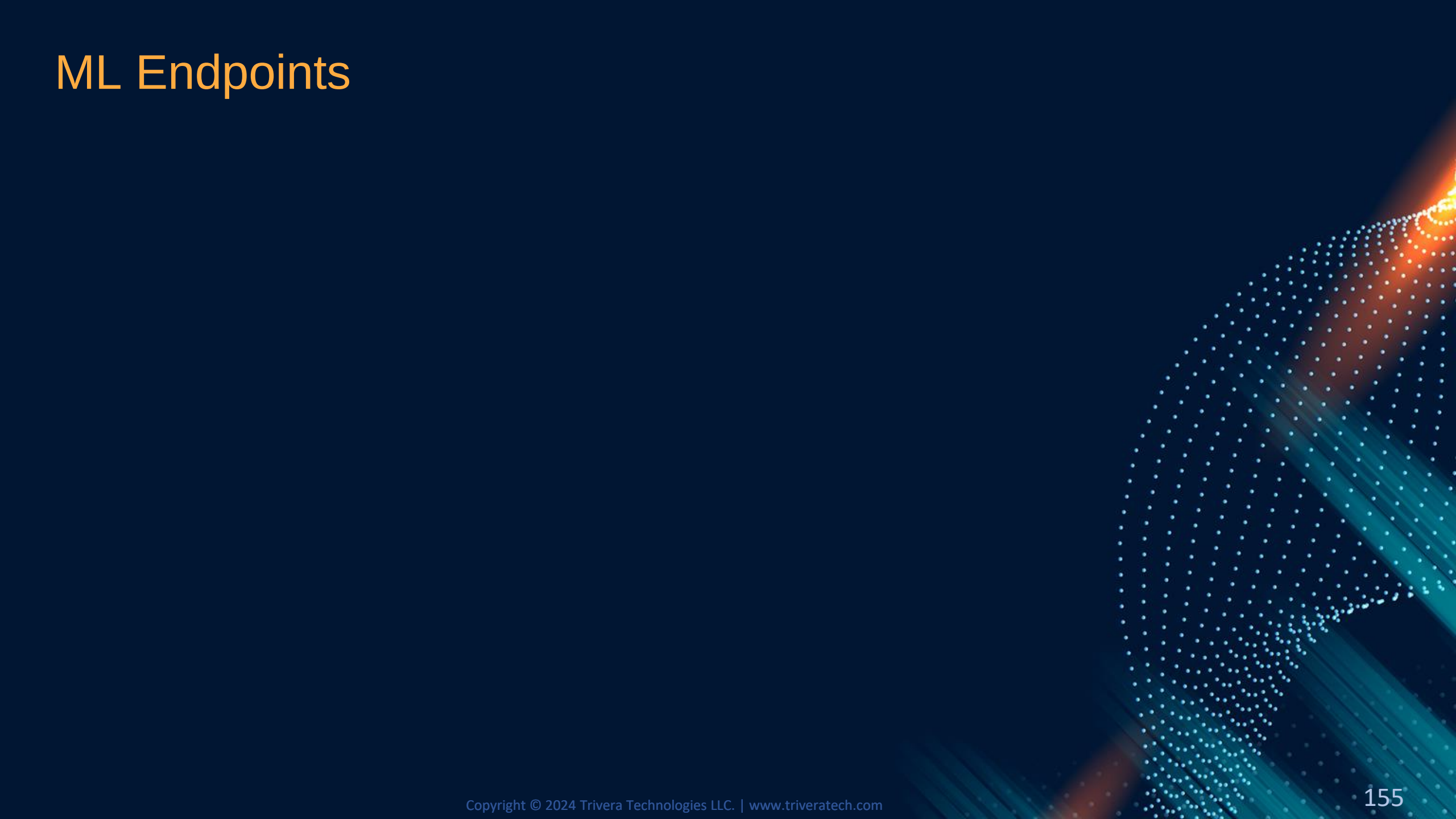
Securing with Container Registry



ML Endpoints



ML Endpoints



Chapter 10:

MLOps

Azure MLFlow

Lesson Agenda: What We Will Cover

- Working with MLOps in Azure Machine Learning
- Leveraging IaC
- Implementing CI/CD
- Exploring event-driven workflows in Azure

MLOps

- Collaboration
- Versioning
- Model Validation
- CI/CD
- Monitoring and Logging
- Scalability
- Reproducibility
- Automation

IaC

- ARM
- Bicep
- CLI
- Blueprints
- Terraform

IaC with Azure ML

- Infrastructure
- Datastores / Datasets
- Model Deployment
- Automation and Orchestration

CI/CD with ML

- Training the model
- Logging metrics
- Retrieving the registered model
- Packaging and deploying the model

Azure DevOps

- Azure Boards: This provides work tracking with Kanban boards, backlogs, team dashboards, and custom reporting. It allows teams to plan, track, and discuss work across the entire development life cycle.
- Azure Repos: This is a version control system that provides Git repositories for source control. It supports pull requests, branching, and searching.
- Azure Pipelines: This is a CI/CD platform for deploying and testing applications to different platforms both for cloud and on-premises offerings.
- Azure Test Plans: This is a platform that includes multiple testing tools for different scenarios such as exploratory or continuous testing.
- Azure Artifacts: This allows collaboration between teams to share packages from private or public repositories such as NuGet or Maven into their pipelines.

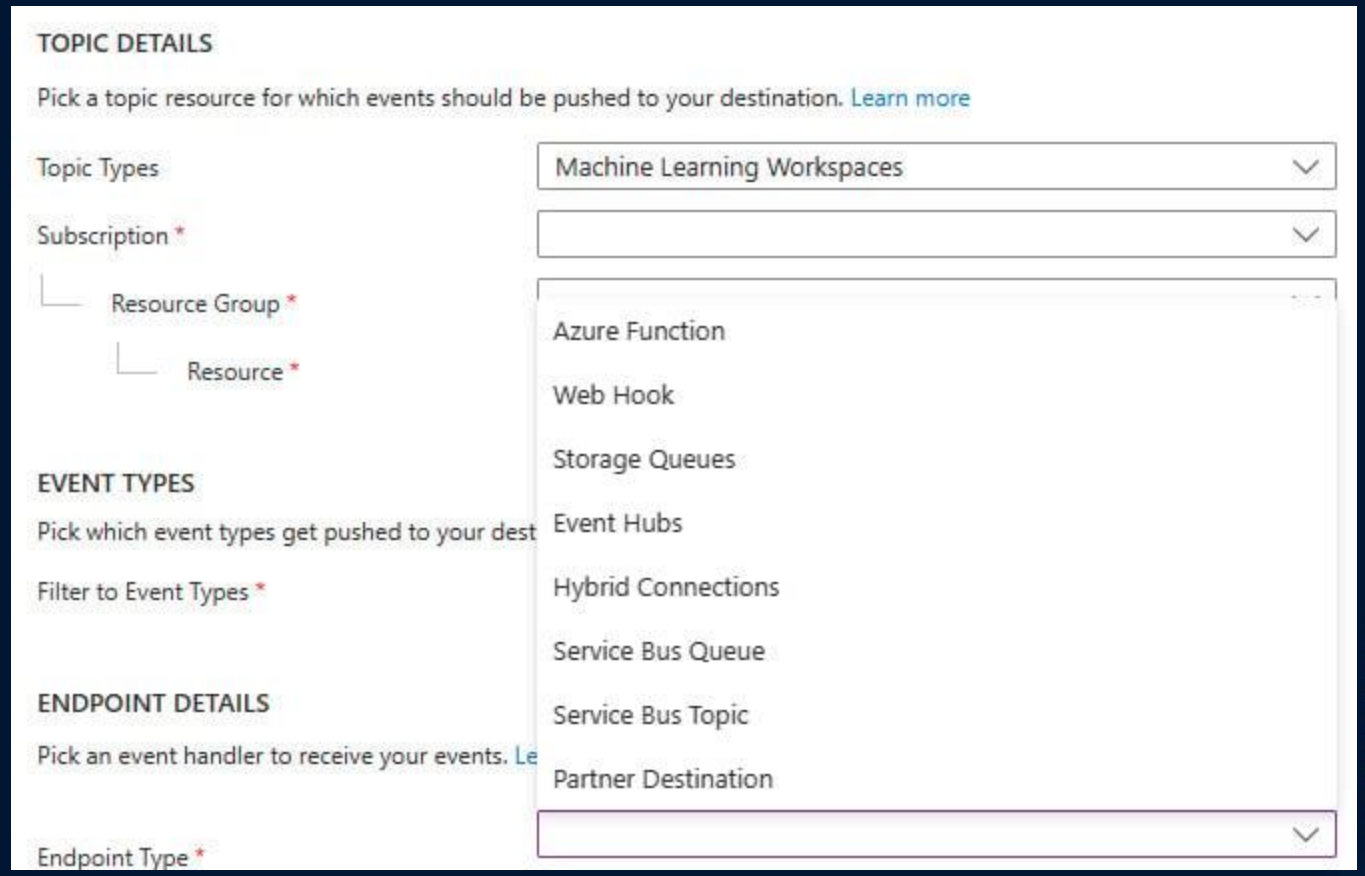
Free Azure DevOps

https://learn.microsoft.com/en-us/azure/devops/user-guide/sign-up-invite-teammates?view=azure-devops.

Event Grid

- Azure Event Grid is a fully managed event routing service.
- It enables you to easily build applications that react to changes or events happening within Azure services or even on-premises.

Event Handlers



Summary

Chapter 11:

Monitoring

Azure Monitoring, Logging, Threat Detection

Lesson Agenda: What We Will Cover

- Enabling logging and configuring data retention for Azure services
- Securing resources with Microsoft Defender
- Exploring threat management with Sentinel

Home >

Monitor | Overview

Microsoft

...

Search

- Overview
- Activity log
- Alerts
- Metrics
- Logs
- Change Analysis
- Service health
- Workbooks

Insights

- Applications
- Virtual Machines
- Storage accounts
- Containers
- Networks
- SQL (preview)
- Azure Cosmos DB
- Key Vaults
- Azure Cache for Redis

The Log Analytics agents, used by VM Insights, won't be supported as of August 31, 2024. Plan to migrate to VM Insights on Azure Monitor agent prior to this date. →

Overview | Tutorials

Insights

Use curated monitoring views for specific Azure resources. View all insights

Application insights
Monitor your app's availability, performance, errors, and usage.
View ··· More

Container Insights
Gain visibility into the performance and health of your controllers, nodes, and containers.
View ··· More

VM Insights
Monitor the health, performance, and dependencies of your VMs and VM scale sets.
View ··· More

Network Insights
View the health and metrics for all deployed network resources.
View ··· More

Detection, triage, and diagnosis

Visualize, analyze, and respond to monitoring data and events. Learn more about monitoring

Metrics
Create charts to monitor and investigate the usage and performance of your Azure resources.
View ··· More

Alerts
Get notified and respond using alerts and actions.
View ··· More

Logs
Analyze and diagnose issues with log queries.
View ··· More

Workbooks
View, create and share interactive reports.
View ··· More

Change Analysis
Investigate what changed to triage

Diagnostic Settings
Route monitoring metrics and logs to

Azure Monitor SCOM managed instance

Managed Prometheus
Collect Prometheus metrics from your

Home > Azure Cache for Redis > gkcache >

Metrics

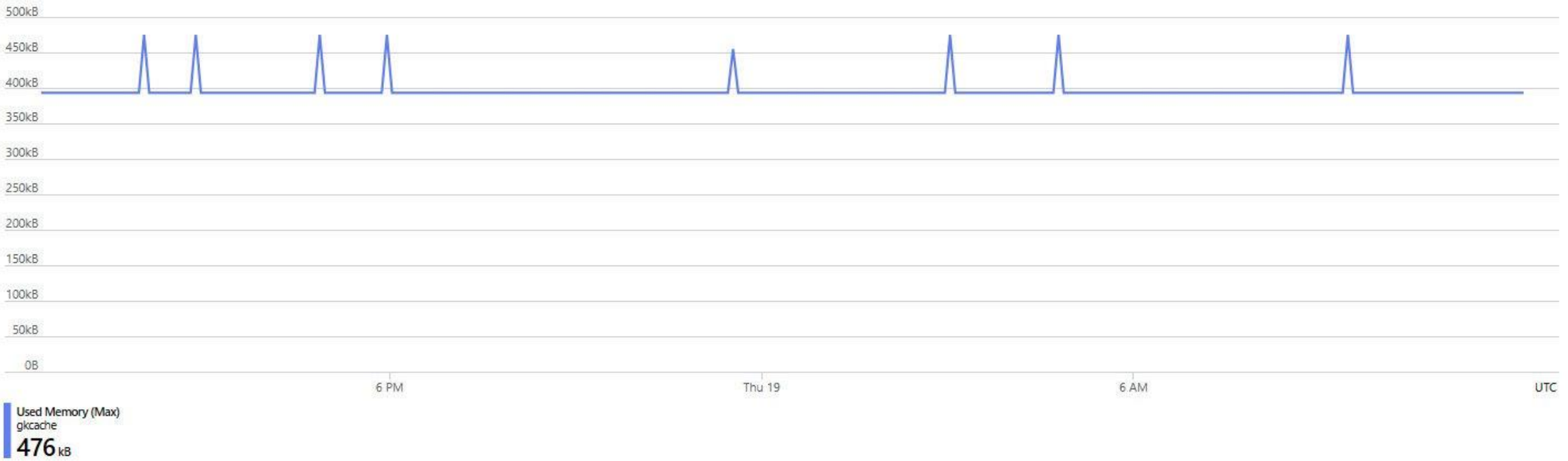
Azure Monitoring

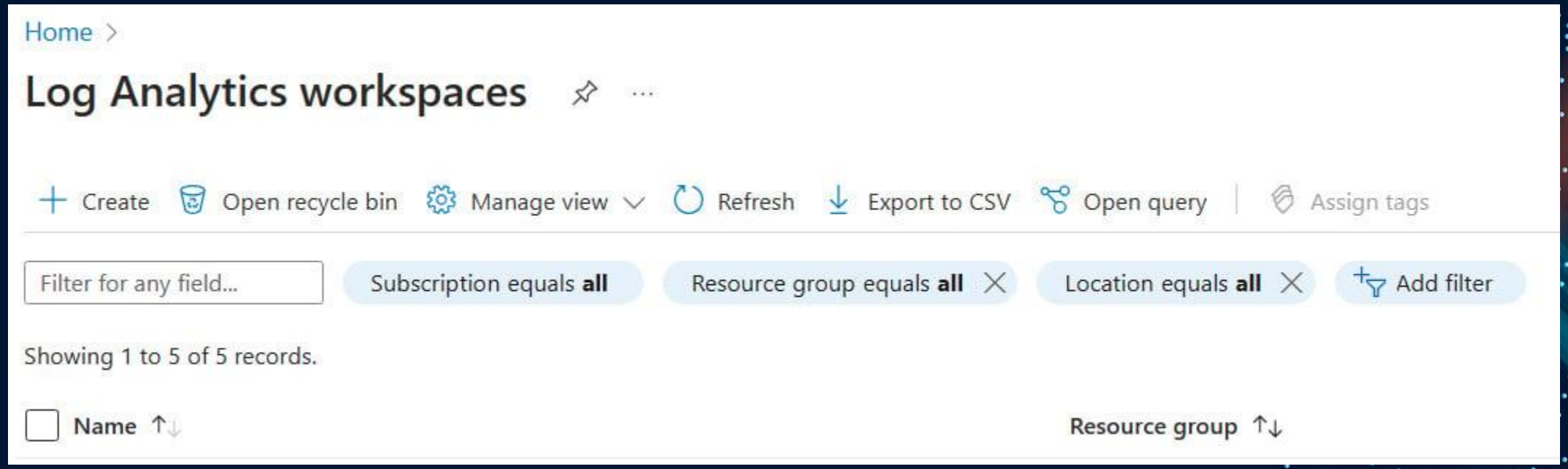
New chart | Refresh | Share | Feedback | UTC Time: Last 24 hours (Automatic - 5 minutes)

Memory Usage

Add metric | Add filter | Apply splitting | Line chart | Drill into Logs | New alert rule | Save to dashboard

gkcache, **Used Memory**, Max





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 * Allsettings

Logs

Category groups

- [x] allLogs
- [x] audit

Categories

- [x] AmlComputeClusterEvent
- [x] AmlComputeClusterNodeEvent
- [x] AmlComputeJobEvent
- [x] AmlComputeCpuGpuUtilization
- [x] AmlRunStatusChangedEvent
- [x] ModelsChangeEvent
- [x] ModelsReadEvent
- [x] ModelsActionEvent

Destination details

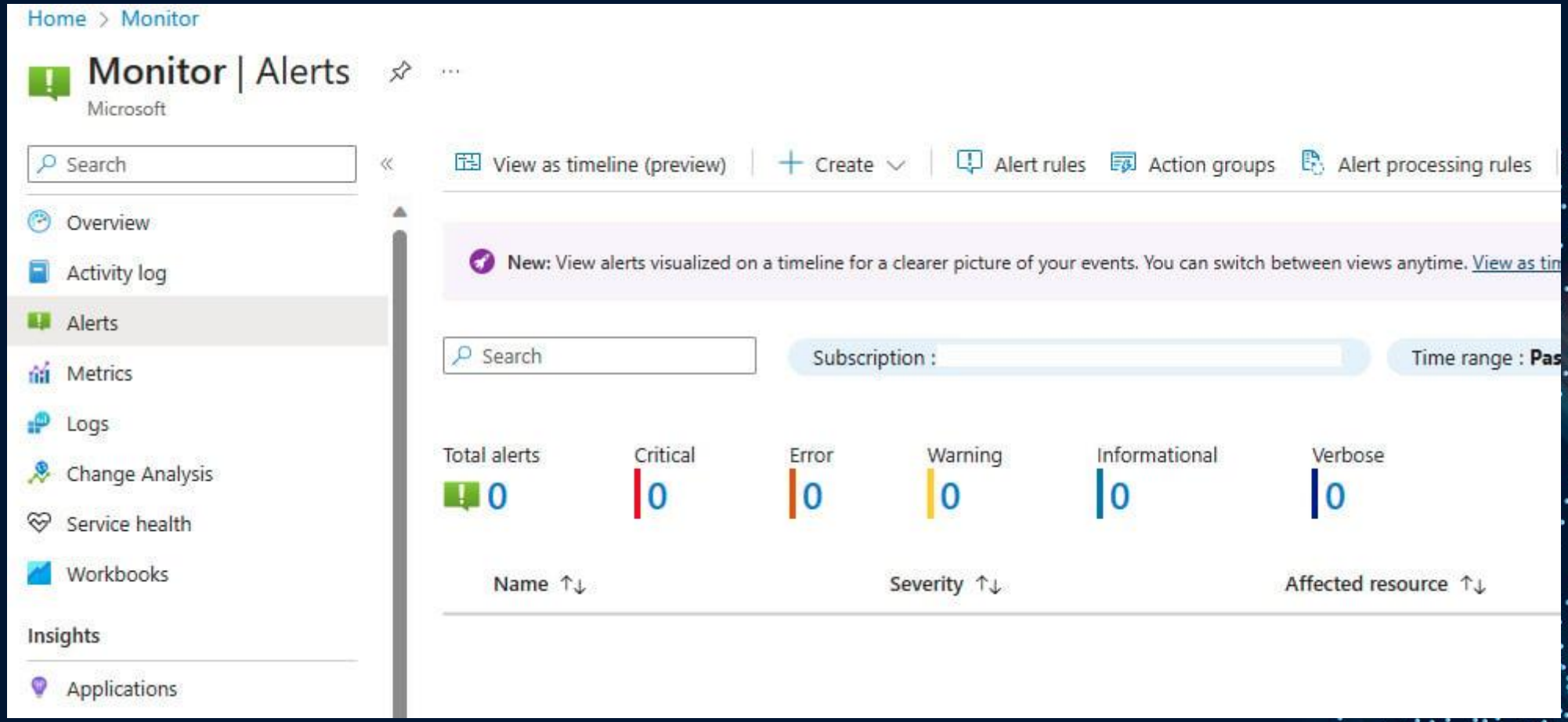
- [x] Send to Log Analytics workspace

Subscription

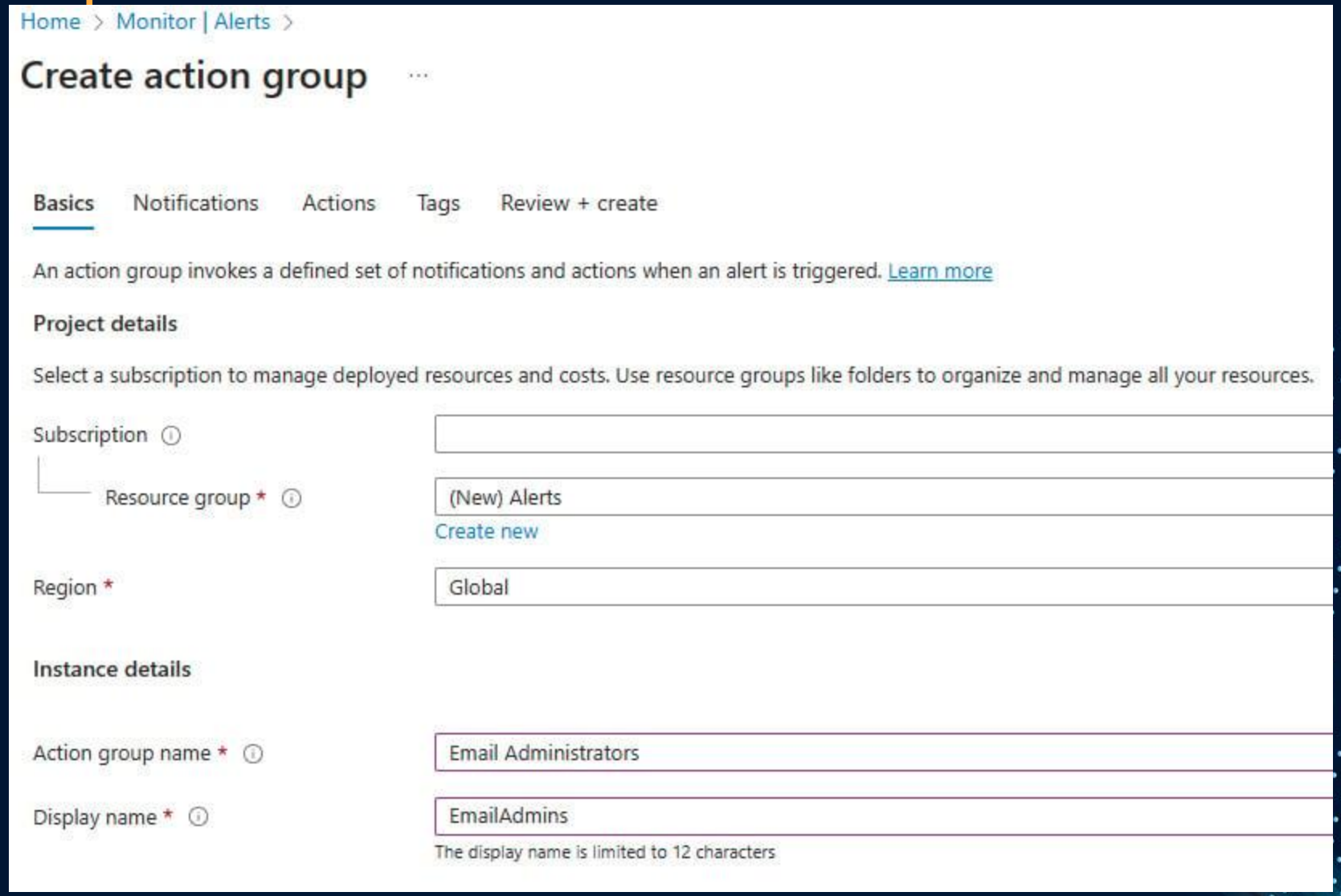
Log Analytics workspace

AzureMLAnalytics (westeurope)

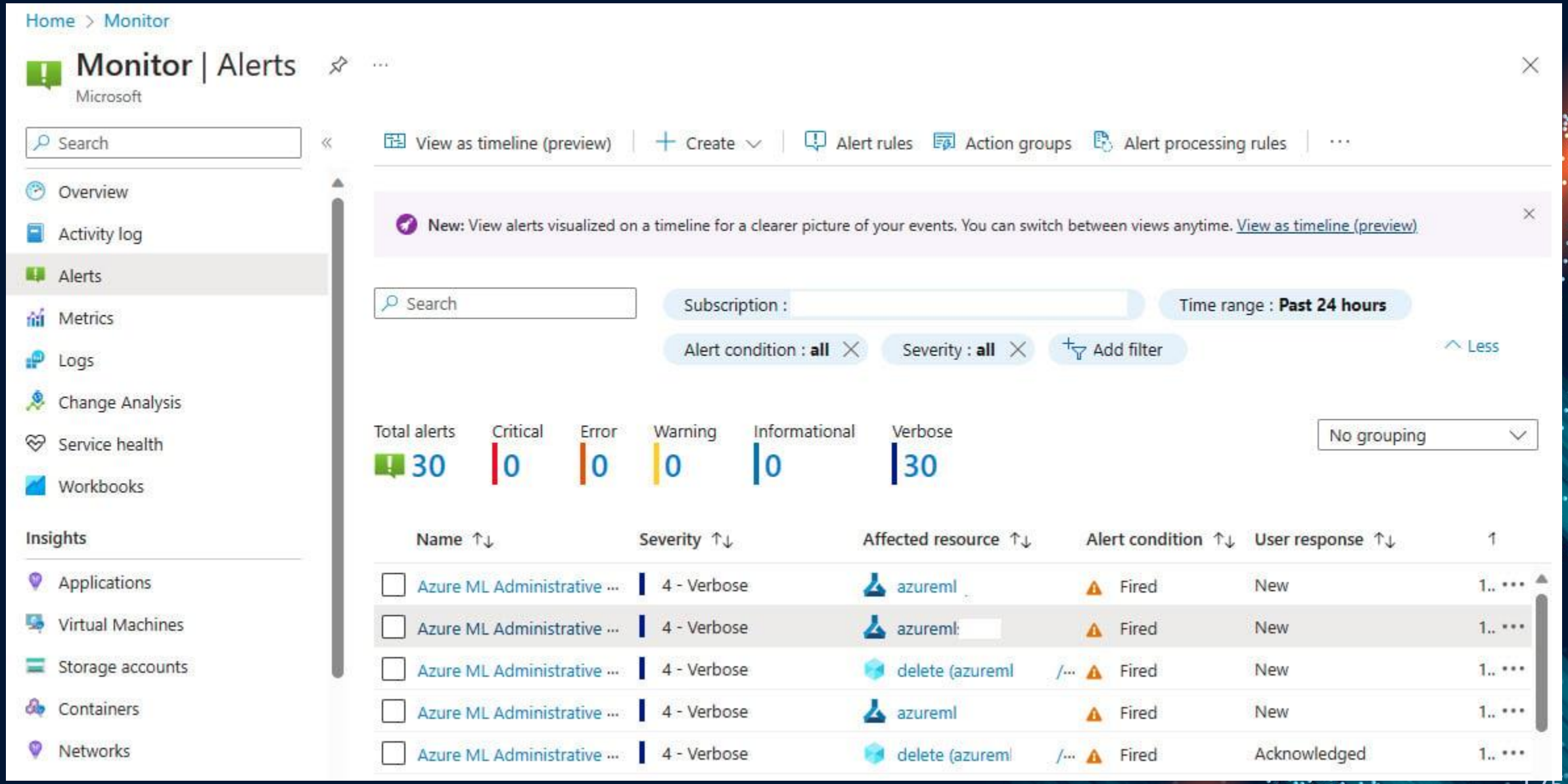
- [ ] Archive to a storage account
- [ ] Stream to an event hub
- [ ] Send to partner solution



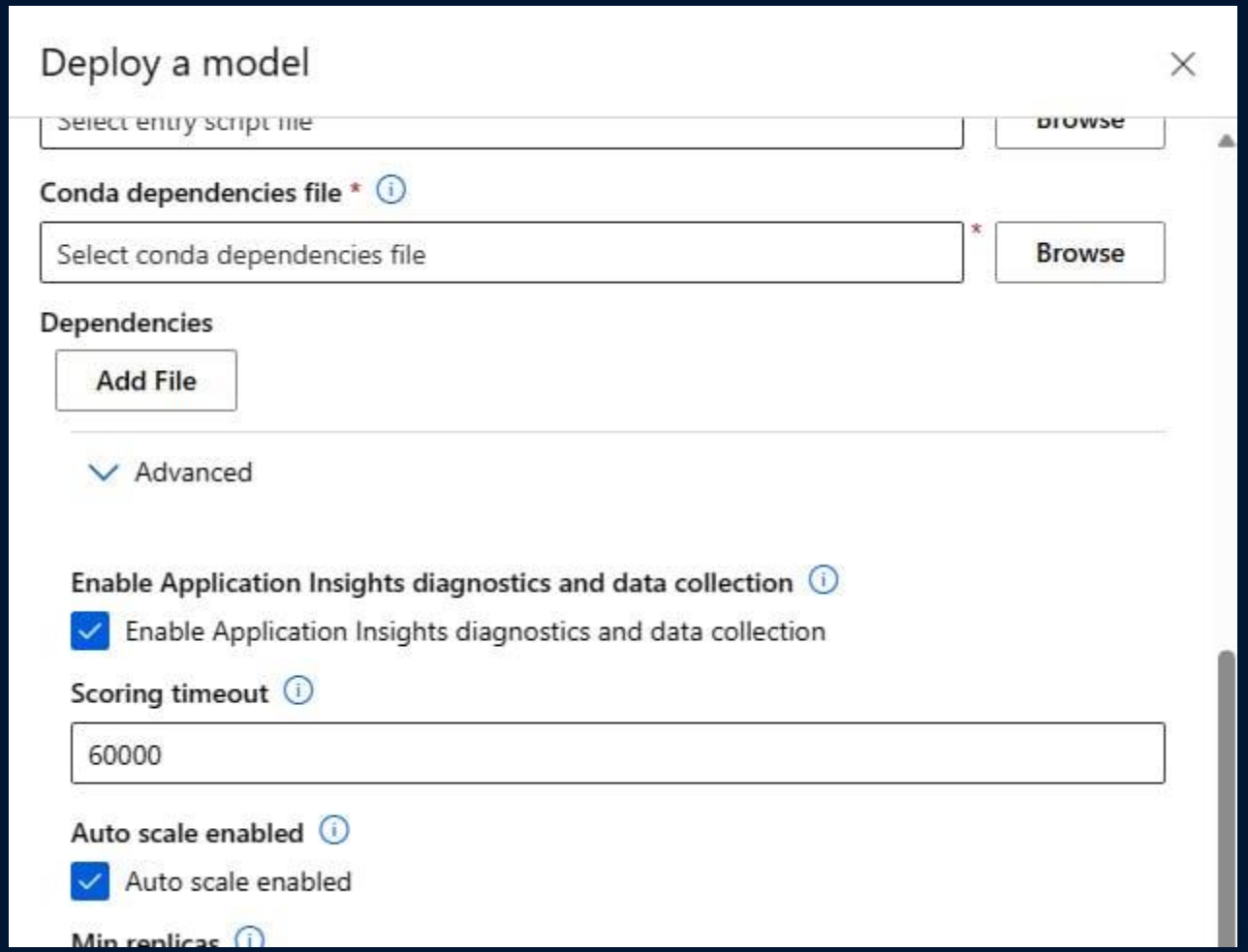
Action Groups



Managing Alerts



Application Insights



Microsoft Defender for Cloud | Overview

Search

Subscriptions | What's new

General

- Overview
- Getting started
- Recommendations
- Attack path analysis
- Security alerts
- Inventory
- Cloud Security Explorer
- Workbooks
- Community
- Diagnose and solve problems

Cloud Security

- Security posture
- Regulatory compliance
- Workload protections
- Data security (Preview)
- Firewall Manager
- DevOps security (preview)

Management

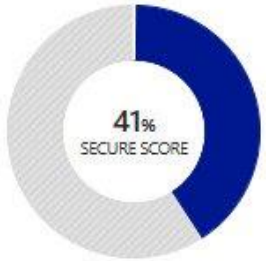
- Environment settings
- Security solutions
- Workflow automation

1	96	20	--	--
Azure subscriptions	Assessed resources	Active recommendations	Attack paths	Security alerts

Security posture

20/20	0/0	0
Unassigned recommendation	Overdue recommendations	Attack paths

Secure score



Azure	41%
AWS	-
GCP	-

Explore your security posture >

Regulatory compliance

Microsoft cloud security benchmark

34 of 63 passed controls

Lowest compliance regulatory standards
by passed controls

No additional standards are currently monitored.

Open policy settings to manage additional compliance policies

Improve your compliance >

Workload protections

Resource coverage

100% For full protection, enable 1 resource plans

Inventory

Unmonitored VMs

3 To better protect your organization, we recommend installing agents

Home > Microsoft Defender for Cloud | Overview >

Security posture

Secure score over time | Governance report | Guides & Feedback

Azure environment

Azure | AWS | GCP | AzureDevOps

Secure score

Environment

1 Management groups

1 Subscriptions

28/46 Unhealthy resources

59 Recommendations

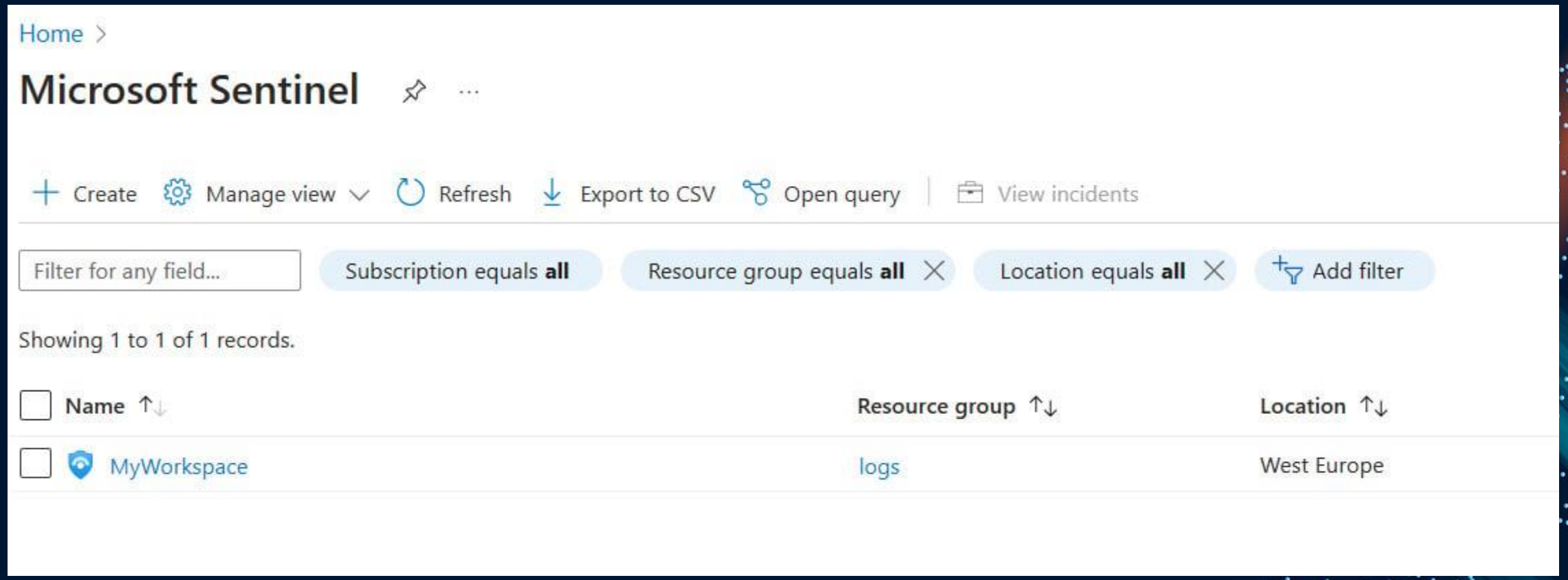
Governance

Assign ownership and drive recommendations remediation using governance. To create your first rule, click here.

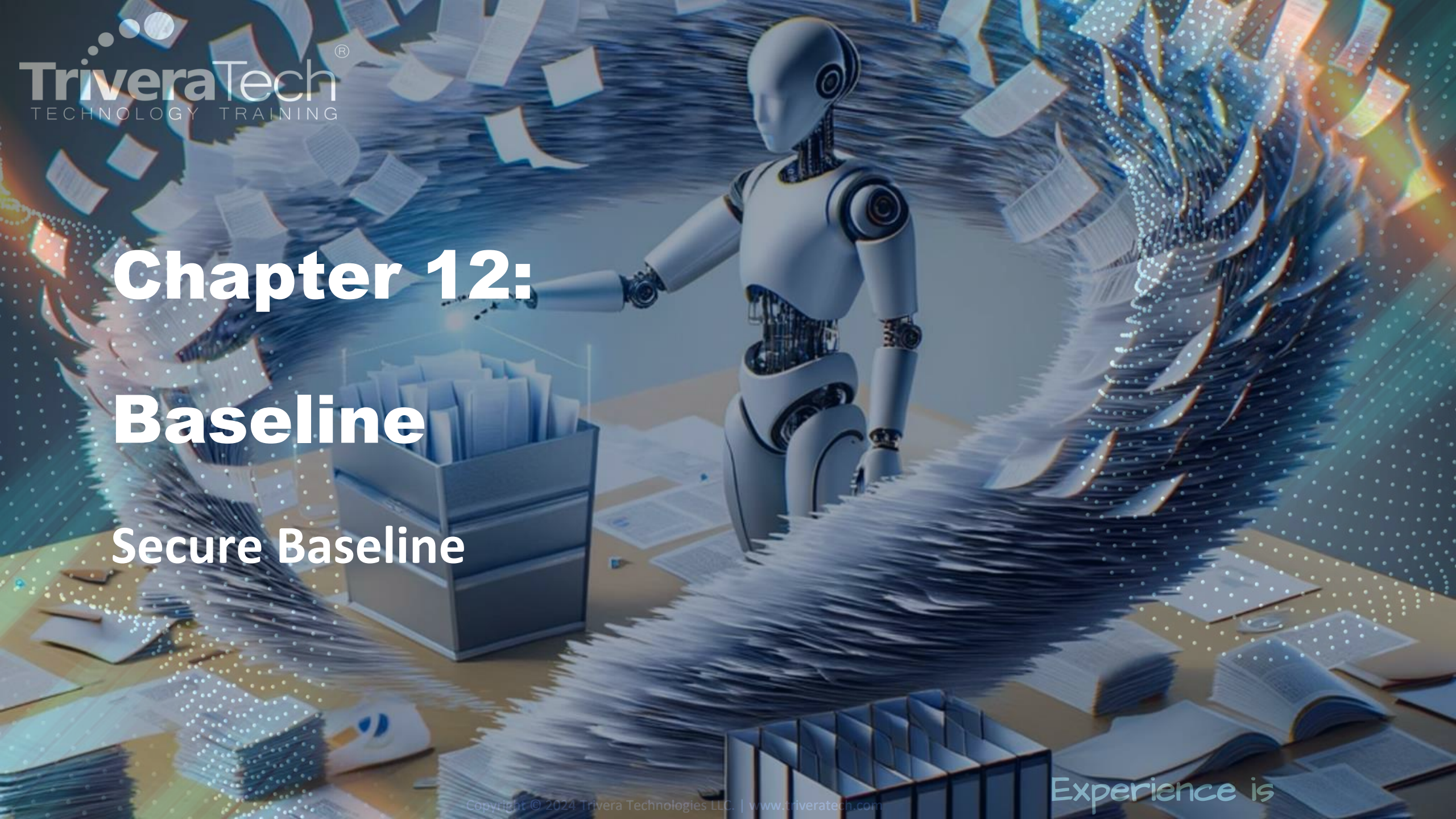
Environment | Owner

Search by name | Environment == All | Group by environment

Name	Secure score	Unhealthy resour...	Attack paths	Recommendations
Azure subscription	41%	28 of 46	0	View recommendations...



Summary



Chapter 12:

Baseline

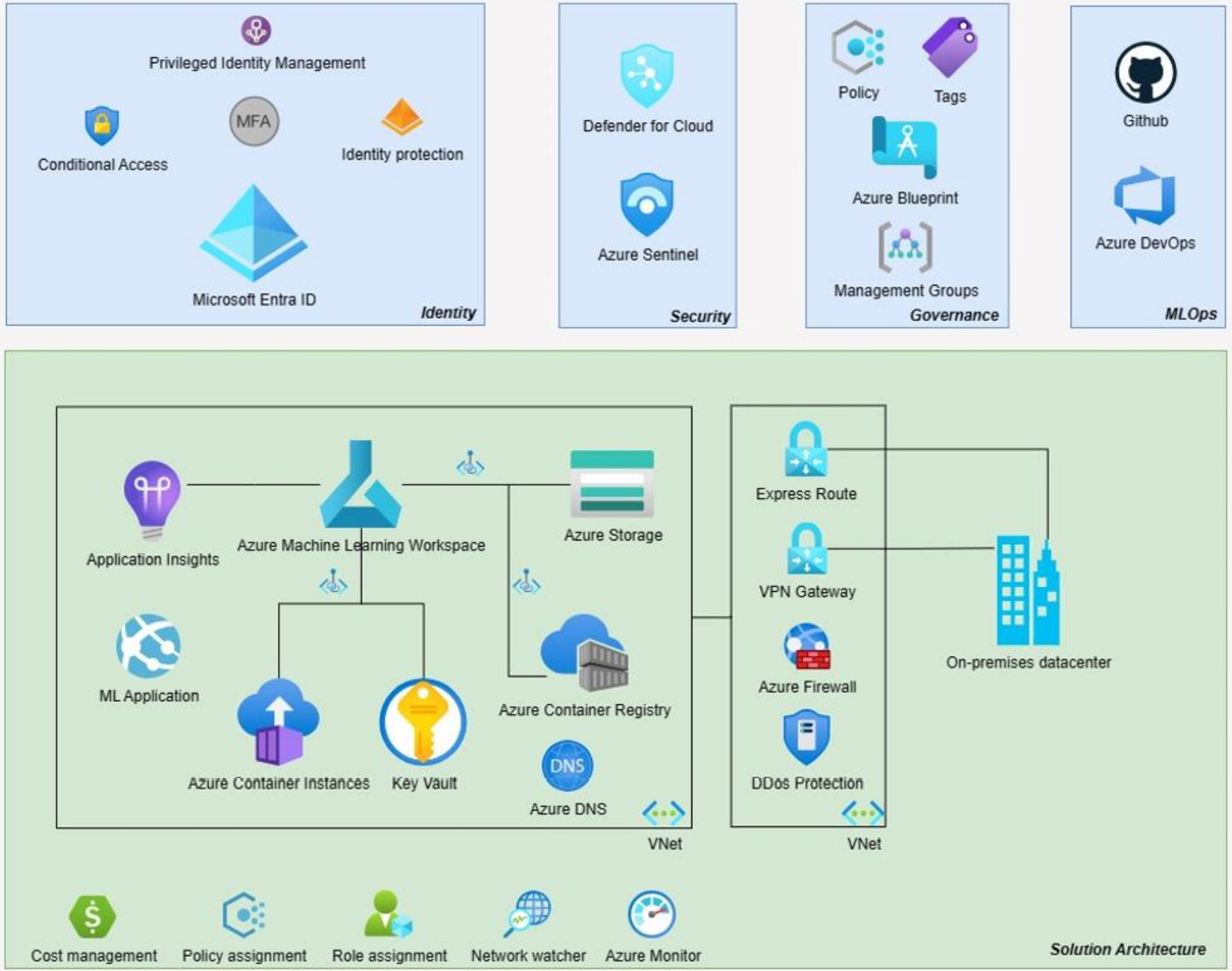
Secure Baseline

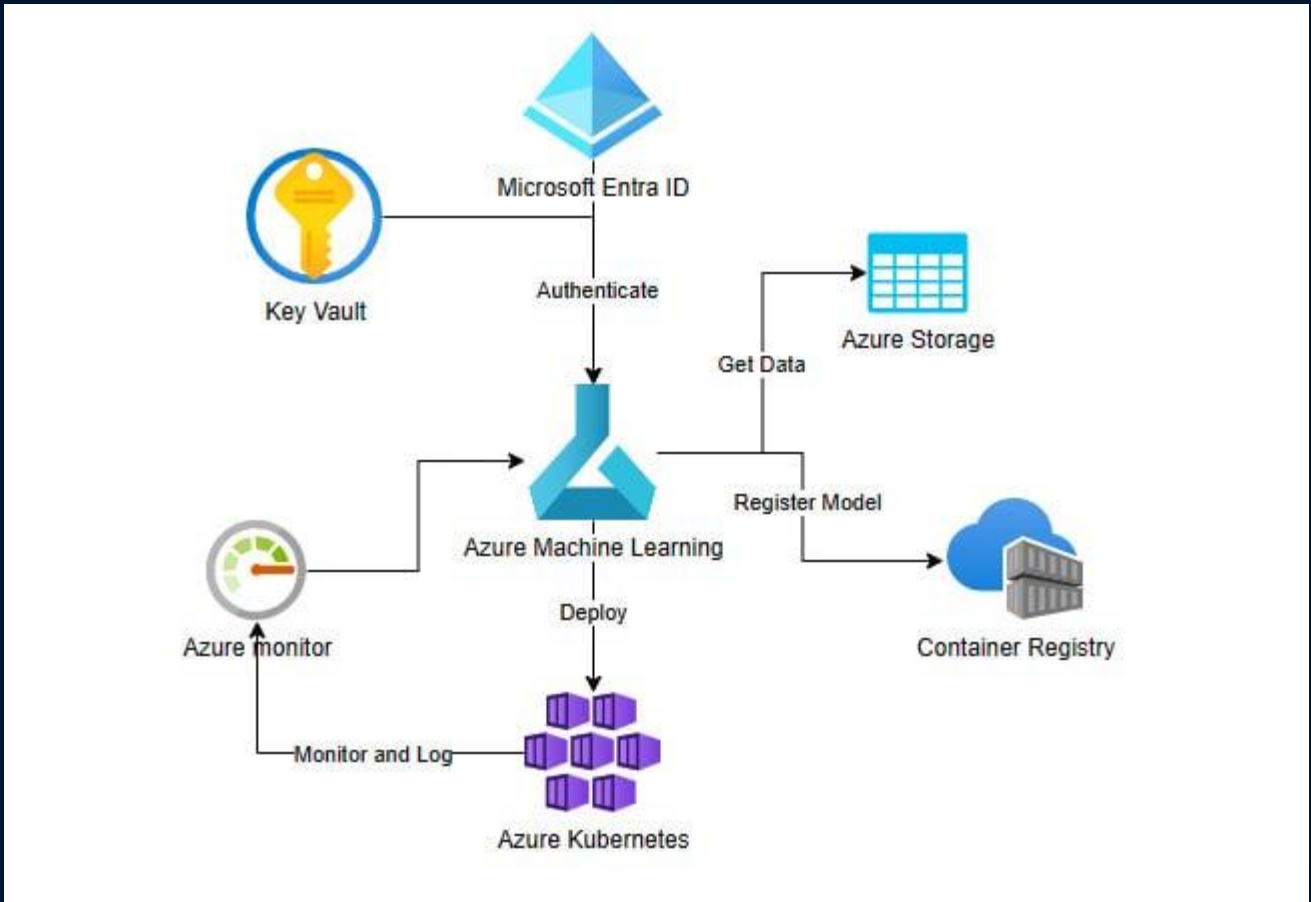
Lesson Agenda: What We Will Cover

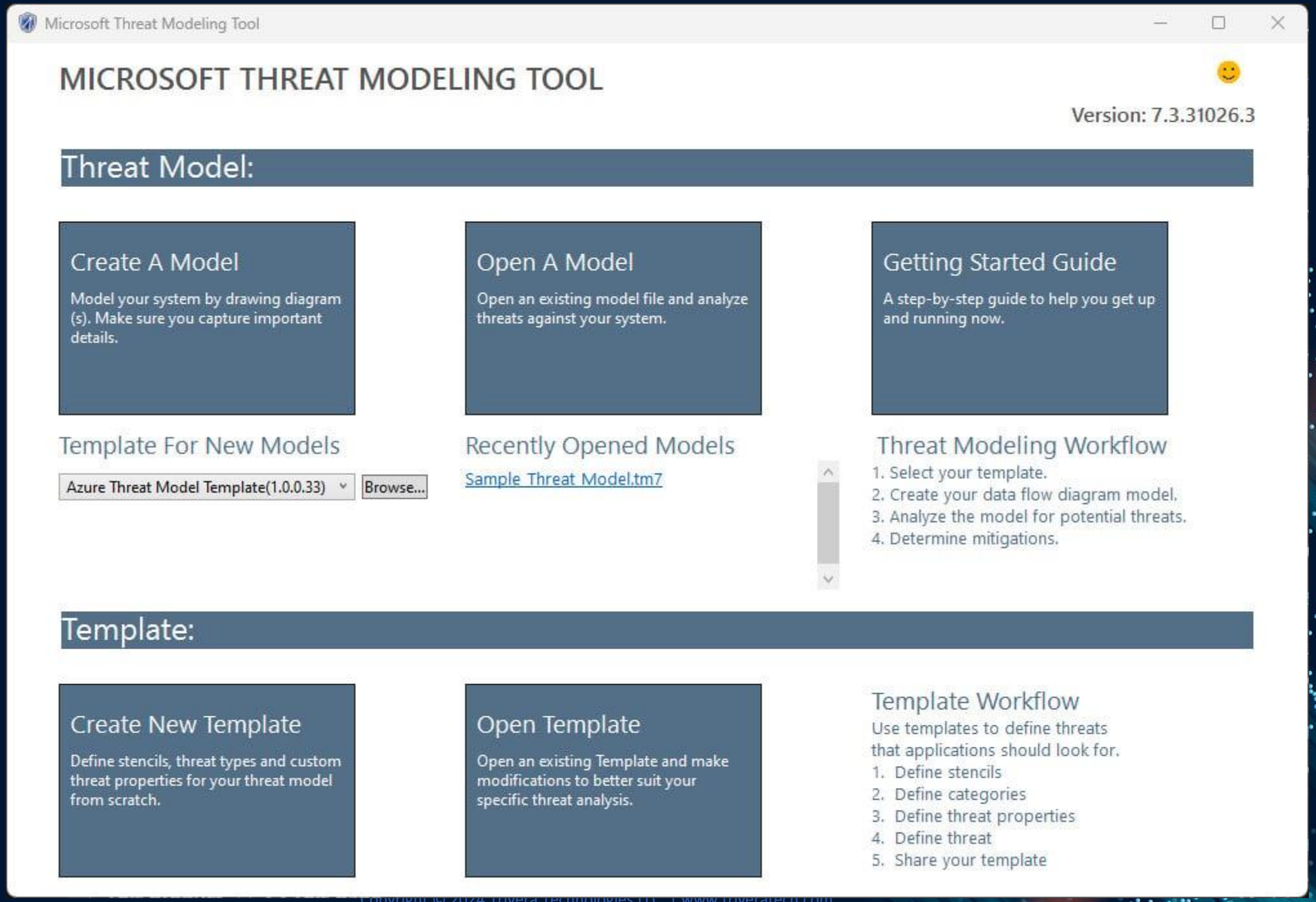
- Setting a baseline for Azure Machine Learning
- Theat modeling for Azure Machine Learning
- Reviewing the share responsibility model for cloud security

Discover Services for Added Security

- NSG
- VPN Gateways
- ExpressRoute
- Azure Firewall
- VWAN
- WAF
- DDoS
- Azure Front Door
- Defender
- Azure Site Recovery


Privileged Identity Management
Conditional Access
MFA
Identity protection
Microsoft Entra ID
Identity
Defender for Cloud
Azure Sentinel
Security
Policy
Tags
Azure Blueprint
Management Groups
Governance
Github
Azure DevOps
MLOps
Application Insights
Azure Machine Learning Workspace
Azure Storage
ML Application
Azure Container Registry
Azure Container Instances
Key Vault
DNS
Azure DNS
VNet
Express Route
VPN Gateway
Azure Firewall
DDos Protection
VNet
On-premises datacenter
Cost management
Policy assignment
Role assignment
Network watcher
Azure Monitor
Solution Architecture





Responsibility	SaaS	PaaS	IaaS	On-Premises
Information and data	Customer	Customer	Customer	Customer
Devices (Mobile and PCs)	Customer	Customer	Customer	Customer
Accounts and identities	Customer	Customer	Customer	Customer
Identity and directory infrastructure	Shared	Shared	Shared	Customer
Applications	Provider	Shared	Shared	Customer
Network controls	Provider	Shared	Shared	Customer
Operating system	Provider	Provider	Shared	Customer
Physical hosts	Provider	Provider	Provider	Customer
Physical network	Provider	Provider	Provider	Customer
Physical datacenter	Provider	Provider	Provider	Customer

Summary

Thanks again for joining us!

- We truly appreciate your time. Please complete the End of Course Survey.
- Any questions?
  - Review the full Course Guide for Course Tips, Resources & Next Step Learning Plans
  - Feel Free to Reach Out: Info@triveratech.com / Dr.Lee@triveratech.com
  - See full list of AI, Python, Coding, Security & Full Stack Courses & SkillJourneys: www.triveratech.com
- Free Courses, Articles, Resources & Offers

Let's Connect! Follow Us for Free Courses, Articles, Resources & Offers:
LinkedIn: @TriveraTech

Subscribe to our Channel for Free Courses & Events
YouTube: @TriveraTech

TriveraTech
TECHNOLOGY TRAINING