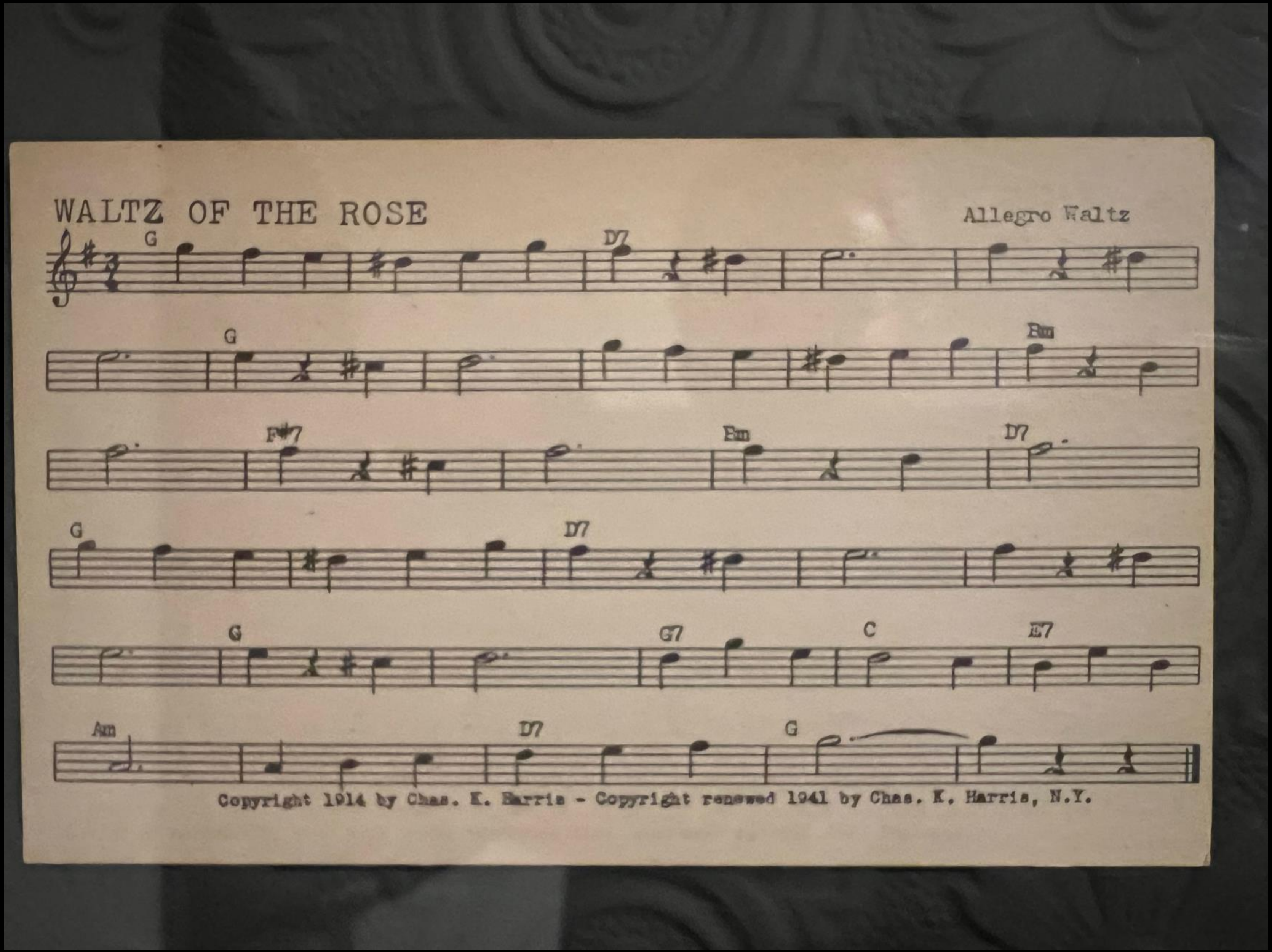
Navigating the Evolution of Cybersecurity: AI's Role in Securing Your Infrastructure from Past to Future

Franz Fiorim

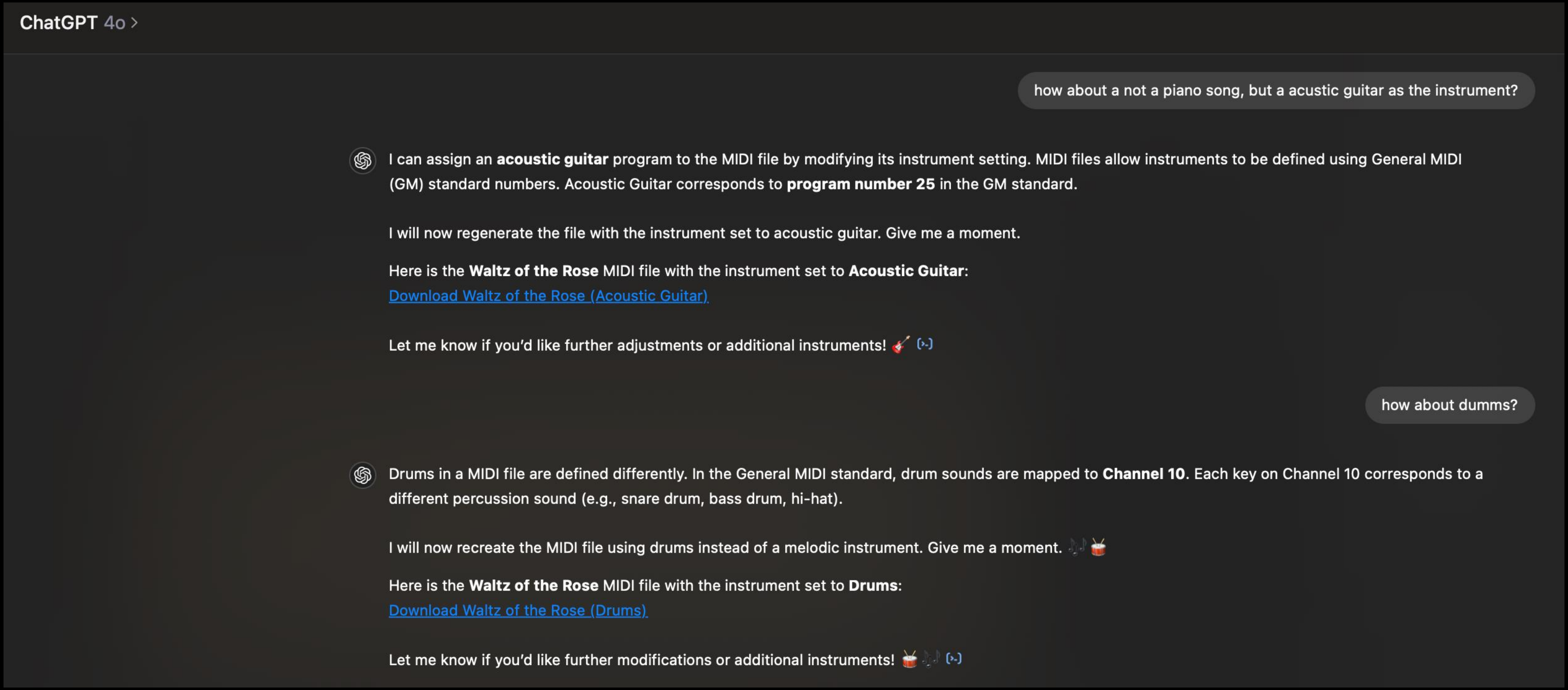
Field-CTO



WALTZ OF THE ROSE

Allegro Waltz

Copyright 1914 by Chas. K. Harris - Copyright renewed 1941 by Chas. K. Harris, N.Y.



how about a not a piano song, but a acustic guitar as the instrument?

I can assign an **acoustic guitar** program to the MIDI file by modifying its instrument setting. MIDI files allow instruments to be defined using General MIDI (GM) standard numbers. Acoustic Guitar corresponds to **program number 25** in the GM standard.

I will now regenerate the file with the instrument set to acoustic guitar. Give me a moment.

Here is the **Waltz of the Rose** MIDI file with the instrument set to **Acoustic Guitar**:
Download Waltz of the Rose (Acoustic Guitar)

Let me know if you'd like further adjustments or additional instruments! 🎸

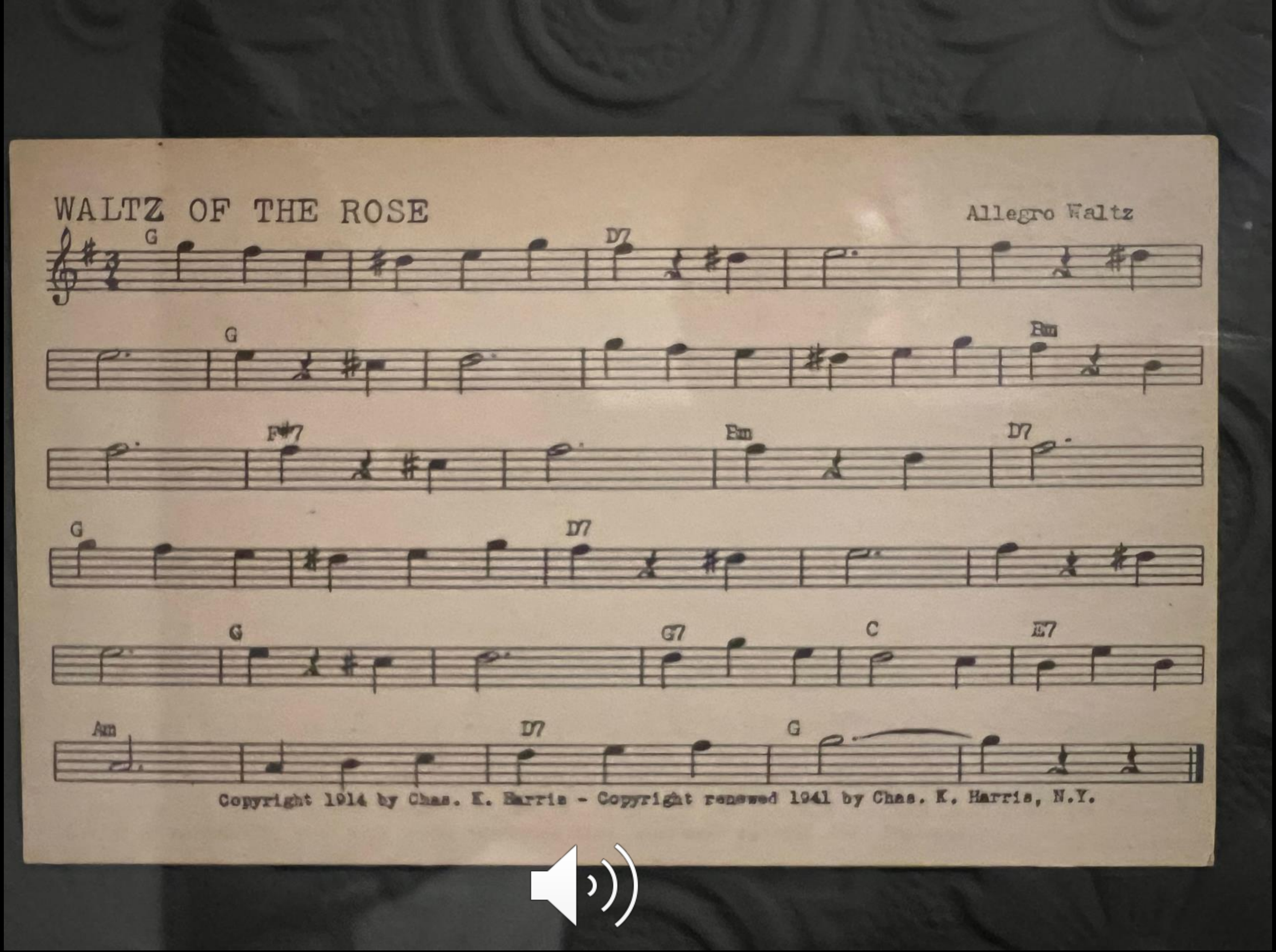
how about dumms?

Drums in a MIDI file are defined differently. In the General MIDI standard, drum sounds are mapped to **Channel 10**. Each key on Channel 10 corresponds to a different percussion sound (e.g., snare drum, bass drum, hi-hat).

I will now recreate the MIDI file using drums instead of a melodic instrument. Give me a moment. 🎶🥁

Here is the **Waltz of the Rose** MIDI file with the instrument set to **Drums**:
Download Waltz of the Rose (Drums)

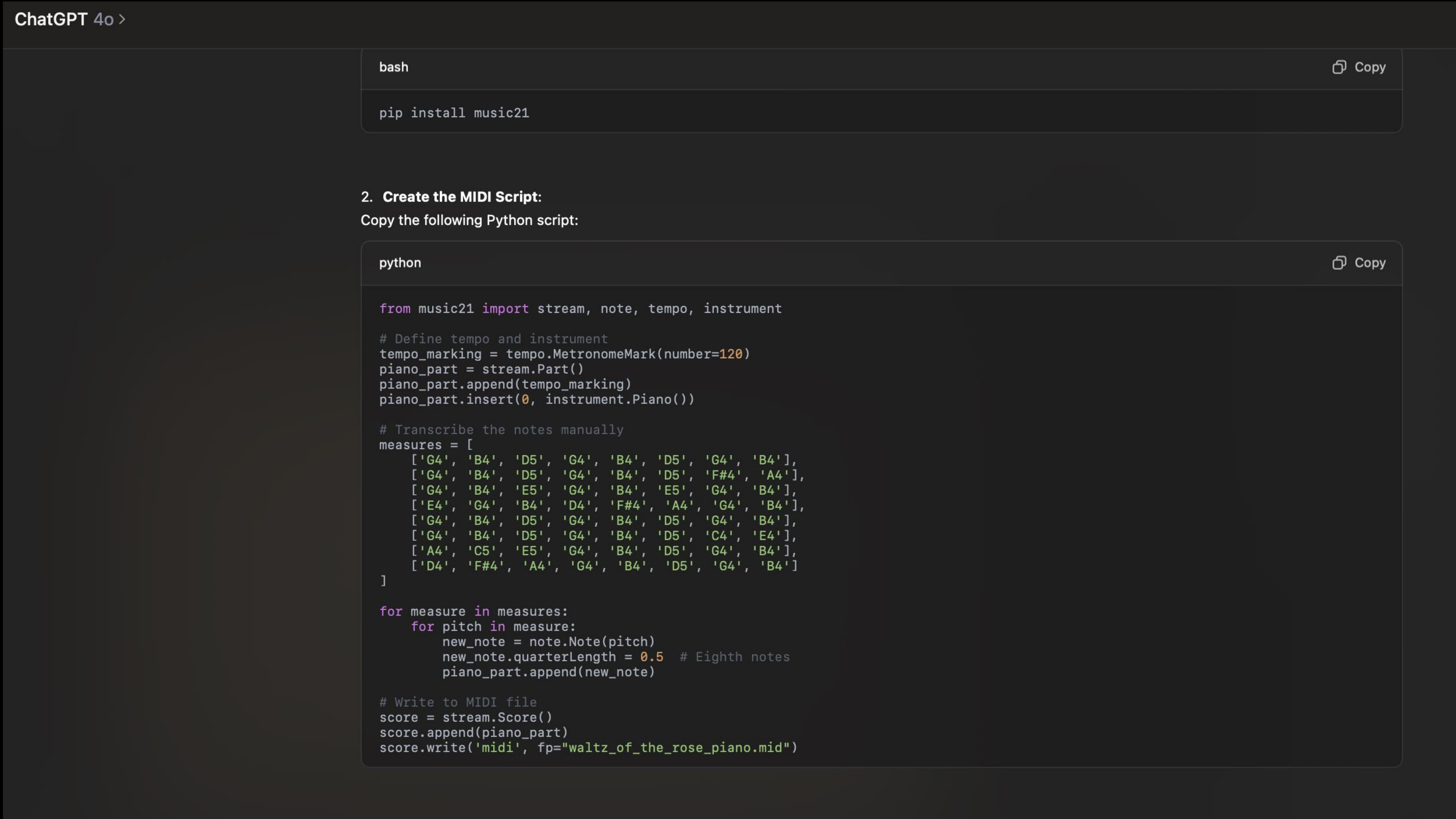
Let me know if you'd like further modifications or additional instruments! 🥁🎶



WALTZ OF THE ROSE

Allegro Waltz

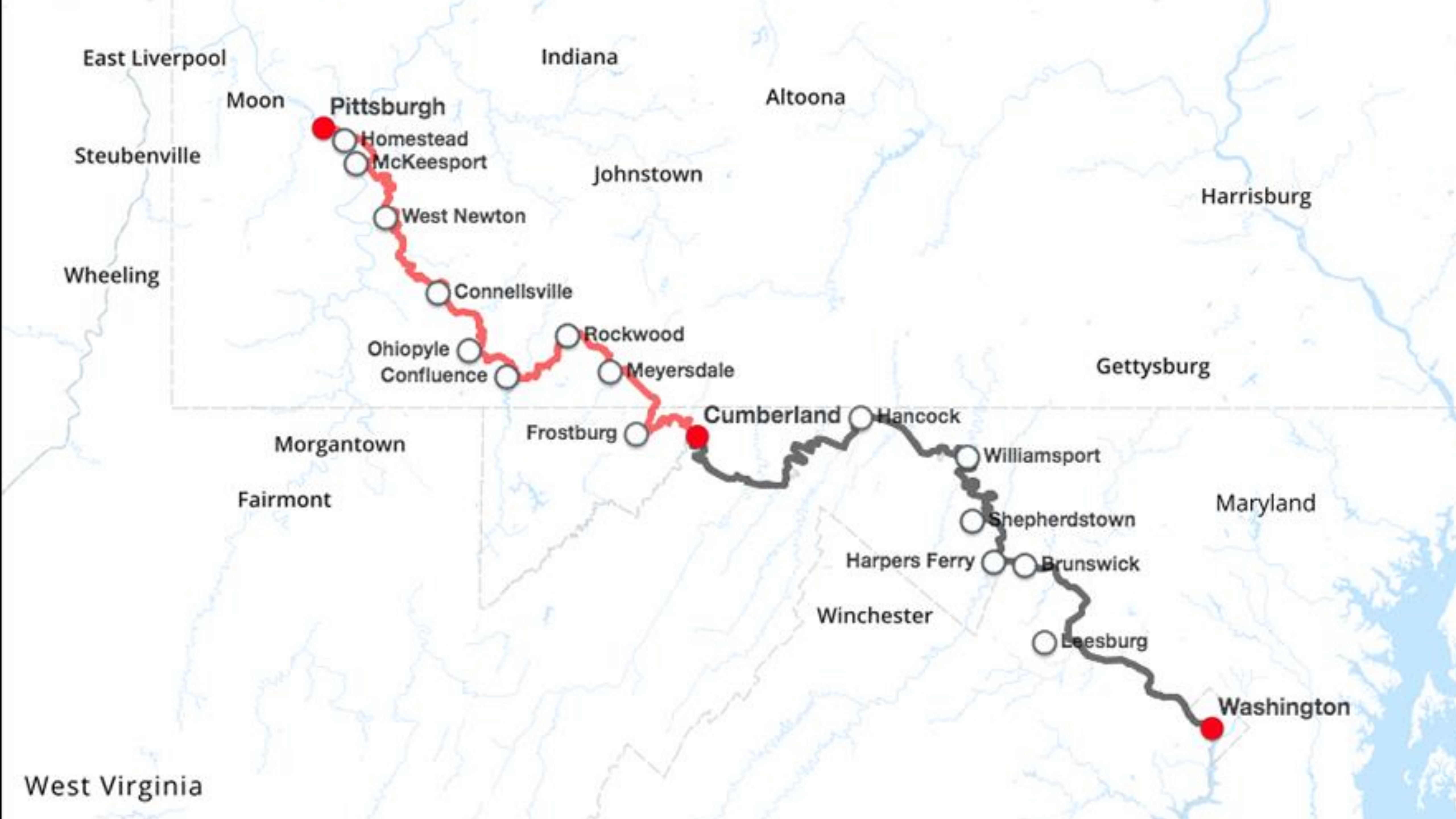
Copyright 1914 by Chas. K. Harris - Copyright renewed 1941 by Chas. K. Harris, N.Y.



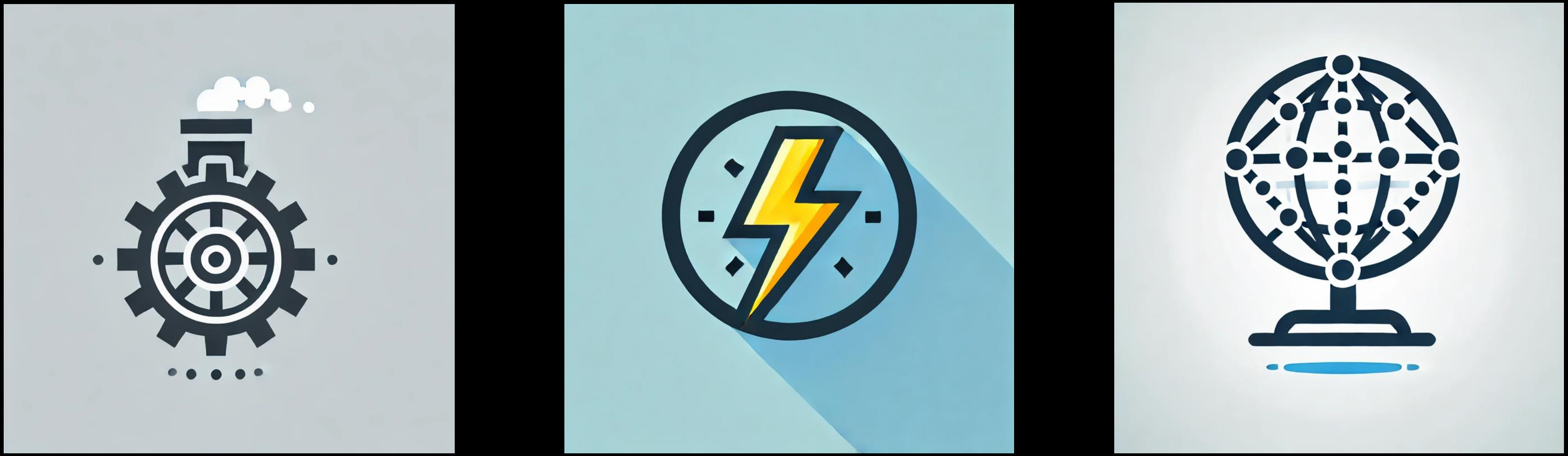
```bash
pip install music21
```

2. **Create the MIDI Script**:
Copy the following Python script:

```python
from music21 import stream, note, tempo, instrument

# Define tempo and instrument
tempo_marking = tempo.MetronomeMark(number=120)
piano_part = stream.Part()
piano_part.append(tempo_marking)
piano_part.insert(0, instrument.Piano())

# Transcribe the notes manually
measures = [
    ['G4', 'B4', 'D5', 'G4', 'B4', 'D5', 'G4', 'B4'],
    ['G4', 'B4', 'D5', 'G4', 'B4', 'D5', 'F#4', 'A4'],
    ['G4', 'B4', 'E5', 'G4', 'B4', 'E5', 'G4', 'B4'],
    ['E4', 'G4', 'B4', 'D4', 'F#4', 'A4', 'G4', 'B4'],
    ['G4', 'B4', 'D5', 'G4', 'B4', 'D5', 'G4', 'B4'],
    ['G4', 'B4', 'D5', 'G4', 'B4', 'D5', 'C4', 'E4'],
    ['A4', 'C5', 'E5', 'G4', 'B4', 'D5', 'G4', 'B4'],
    ['D4', 'F#4', 'A4', 'G4', 'B4', 'D5', 'G4', 'B4']
]

for measure in measures:
    for pitch in measure:
        new_note = note.Note(pitch)
        new_note.quarterLength = 0.5  # Eighth notes
        piano_part.append(new_note)

# Write to MIDI file
score = stream.Score()
score.append(piano_part)
score.write('midi', fp="waltz_of_the_rose_piano.mid")
```

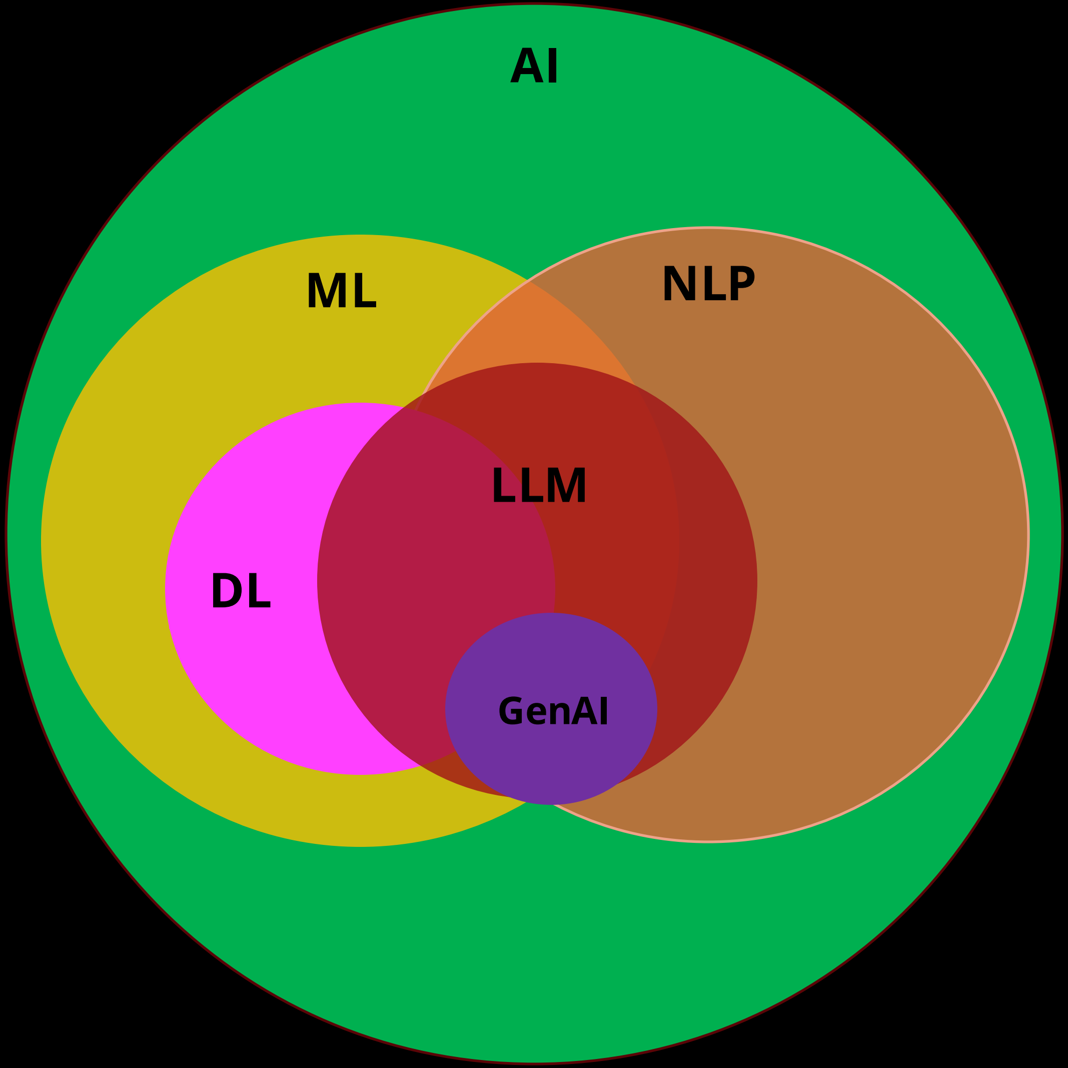

East Liverpool
Indiana
Moon
Pittsburgh
Altoona
Homestead
McKeesport
Steubenville
Johnstown
Harrisburg
West Newton
Wheeling
Connellsville
Rockwood
Ohiopyle
Confluence
Meyersdale
Gettysburg
Cumberland
Hancock
Frostburg
Morgantown
Williamsport
Fairmont
Maryland
Shepherdstown
Harpers Ferry
Brunswick
Winchester
Leesburg
Washington
West Virginia

32



AI Landscape

- Artificial Intelligence
- Machine Learning
- Natural Language Processing
- Deep Learning
- Large Language Models
- Generative AI



Source: www.wondavr.com

AI is Not New in Cybersecurity

2005
Anti-spam using ML

2000’s
SVM/Deep Learning for phishing & malicious URLs

2000’s
NLP for DGA

2010’s
Belief Propagation & Markov Model for Lateral Movement

2010’s
Network Exploits using SVM

2010’s
Writing Style DNA for BEC

2010’s
Computer Vision for phishing

2010’s
TrendX pre\post execution file detection

2020’s
Zero Trust AI Regression analysis for account compromise detection

2020’s
Isolation forest used in anomalous network traffic detection

2020’s
Vulnerability prioritization models used in ASRM

2020’s
Trend Companion GenAI assistant for AI SOC & ASRM

2020’s
Agentic AI to build AI SOAR & SIEM & Shift Left

What are the biggest security threats we face right now?

What investments are needed to improve our security posture?

What is our incident response plan?

What training and awareness programs do we have in place for employees?

How do we ensure compliance with regulations and standards?

How do we measure the effectiveness of our security measures?

Challenges

Lack of In-House Expertise

- Combination of AI and cybersecurity skills is rare
- Difficult to recruit and retain professionals with required expertise

Inadequate Tools and Frameworks

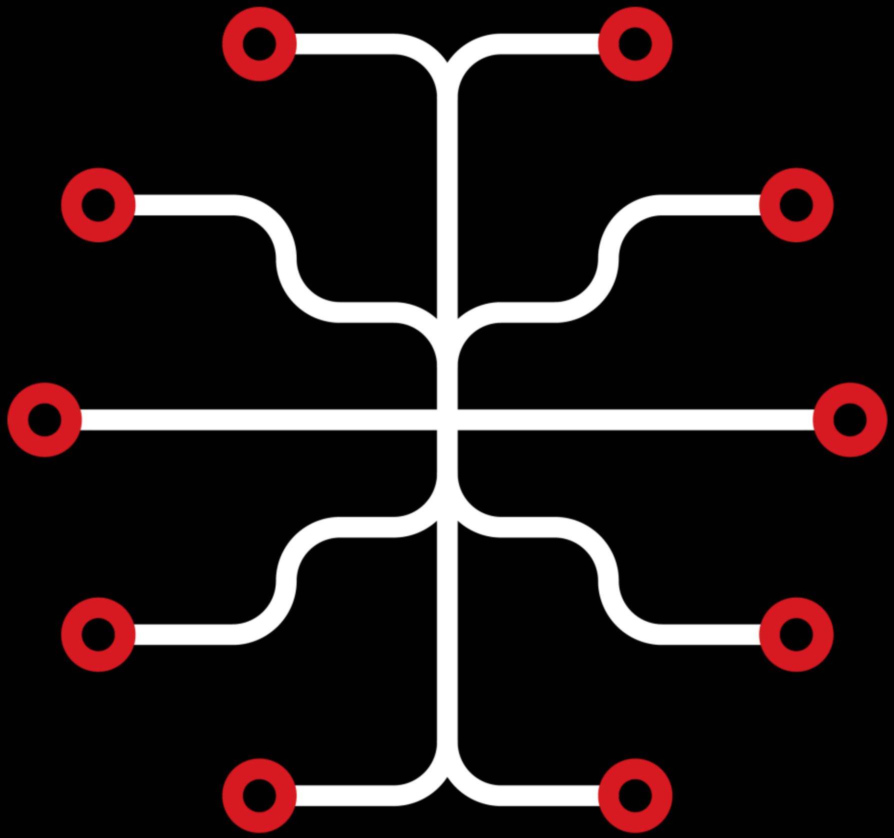
- Tools and frameworks for AI security are still evolving
- May not be as mature or comprehensive as those for traditional IT systems

Difficulty in Assessing and Managing AI-Specific Risks

- AI systems introduce new types of risks (e.g., adversarial attacks, bias)
- Challenging to assess and manage risks in unique business contexts

Balancing Security with Performance and Usability

- Implementing AI security measures can impact performance and usability
- Challenging to strike the right balance and prioritise trade-offs



What governance structures are in place for AI security?

How do we secure the data used for training AI models?

What are the best practices for securing AI environments?

What steps are we taking to prevent data poisoning and model manipulation?

How do we ensure compliance with privacy and AI regulations?

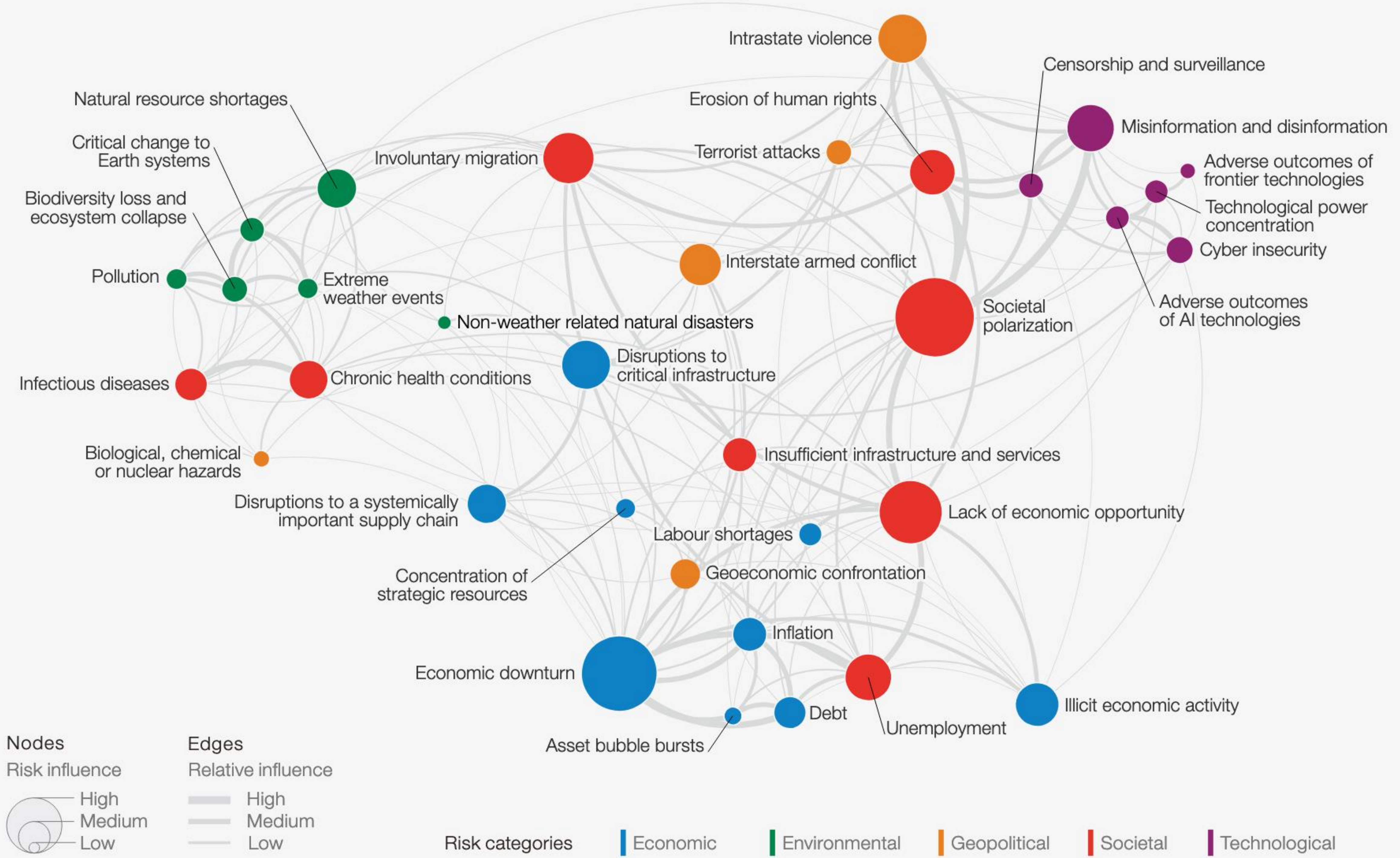
How do we protect against the new attack surfaces introduced by AI?

WORLD
ECONOMIC
FORUM

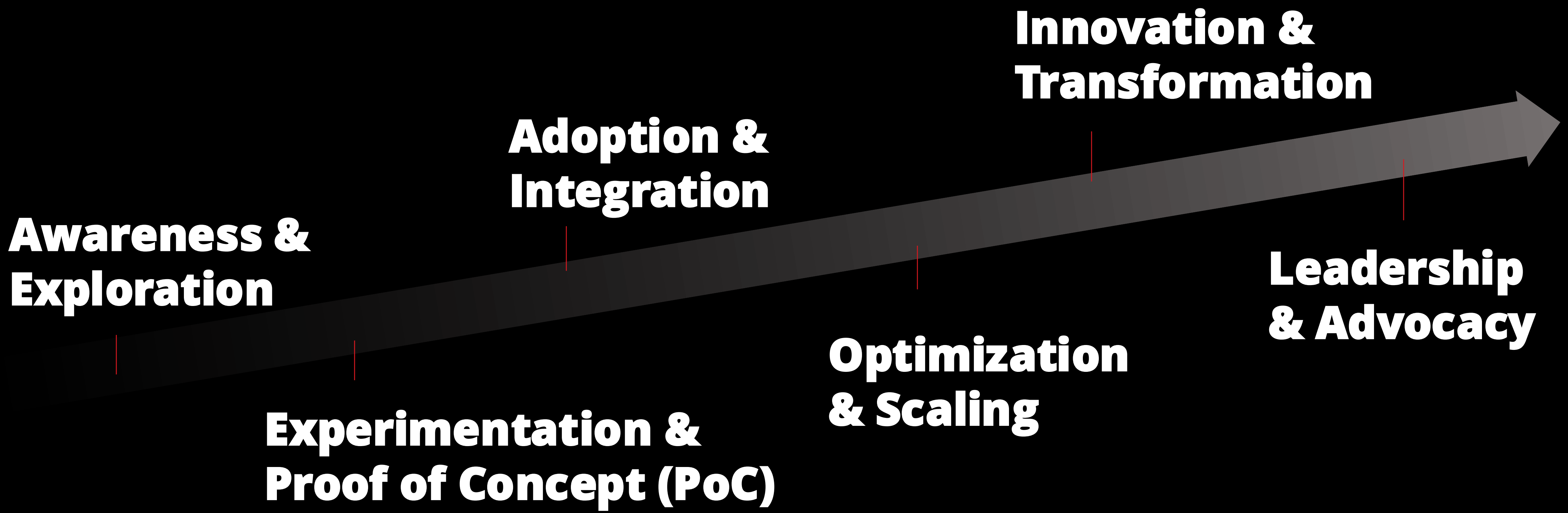
The Global Risks Report 2024

19th Edition

INSIGHT REPORT


Intrastate violence
Censorship and surveillance
Natural resource shortages
Erosion of human rights
Misinformation and disinformation
Critical change to Earth systems
Involuntary migration
Terrorist attacks
Adverse outcomes of frontier technologies
Biodiversity loss and ecosystem collapse
Technological power concentration
Cyber insecurity
Pollution
Interstate armed conflict
Extreme weather events
Societal polarization
Adverse outcomes of AI technologies
Non-weather related natural disasters
Disruptions to critical infrastructure
Infectious diseases
Chronic health conditions
Biological, chemical or nuclear hazards
Insufficient infrastructure and services
Disruptions to a systemically important supply chain
Lack of economic opportunity
Labour shortages
Concentration of strategic resources
Geoeconomic confrontation
Inflation
Economic downturn
Debt
Unemployment
Illicit economic activity
Asset bubble bursts
Nodes
Risk influence
High
Medium
Low
Edges
Relative influence
High
Medium
Low
Risk categories
Economic
Environmental
Geopolitical
Societal
Technological

AI Maturity Journey



©2024 Trend Micro Inc.

AI Introducing New Risks To The Org

Data

- Sensitive Data Exposure
- Insecure Outputs
- Extorsion (Deepfakes)

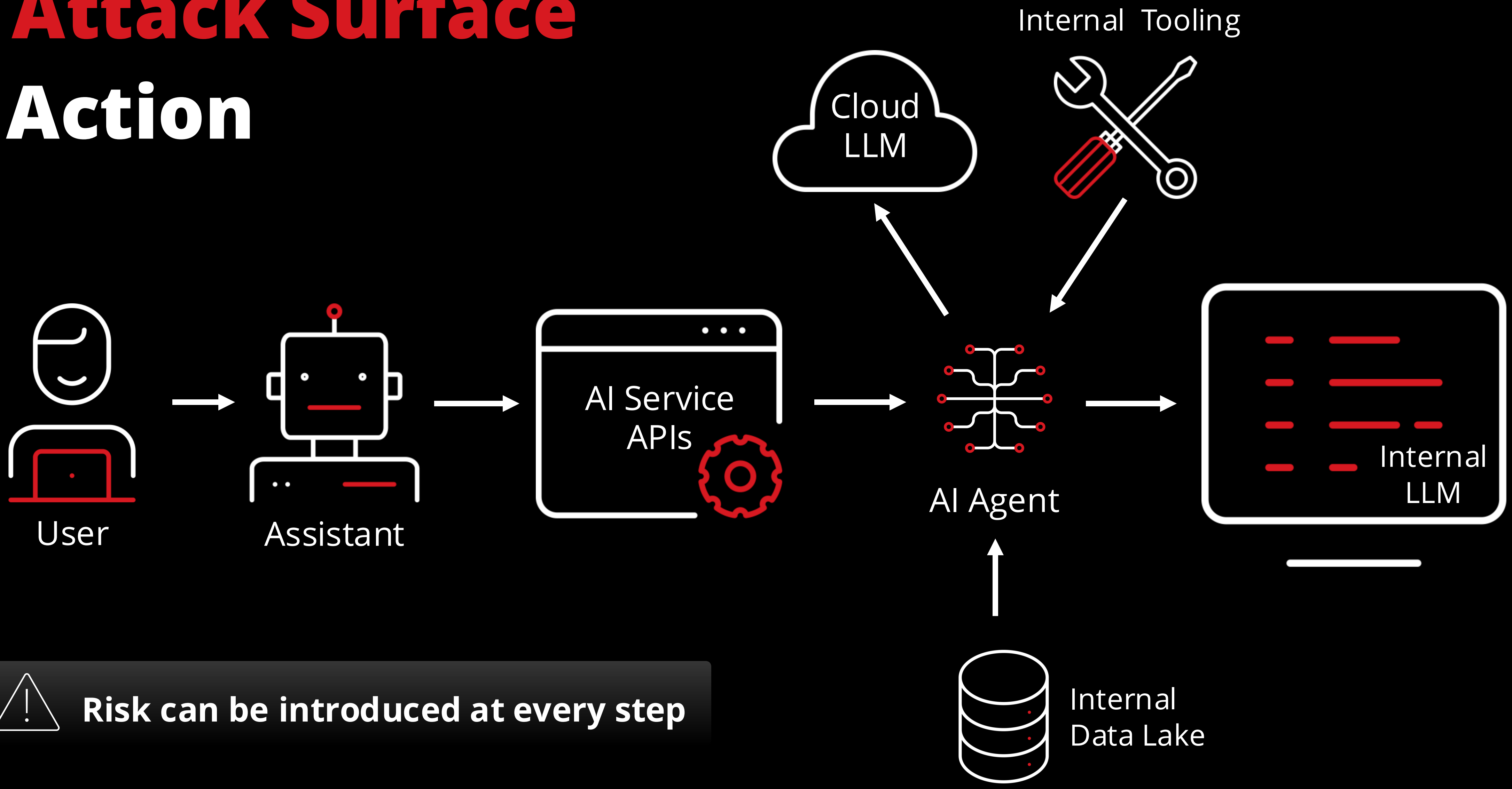
Model

- Prompt Injection
- Jailbreaks
- Model Denial of Service
- Hallucinations
- Misalignment
- Poisoning

Management

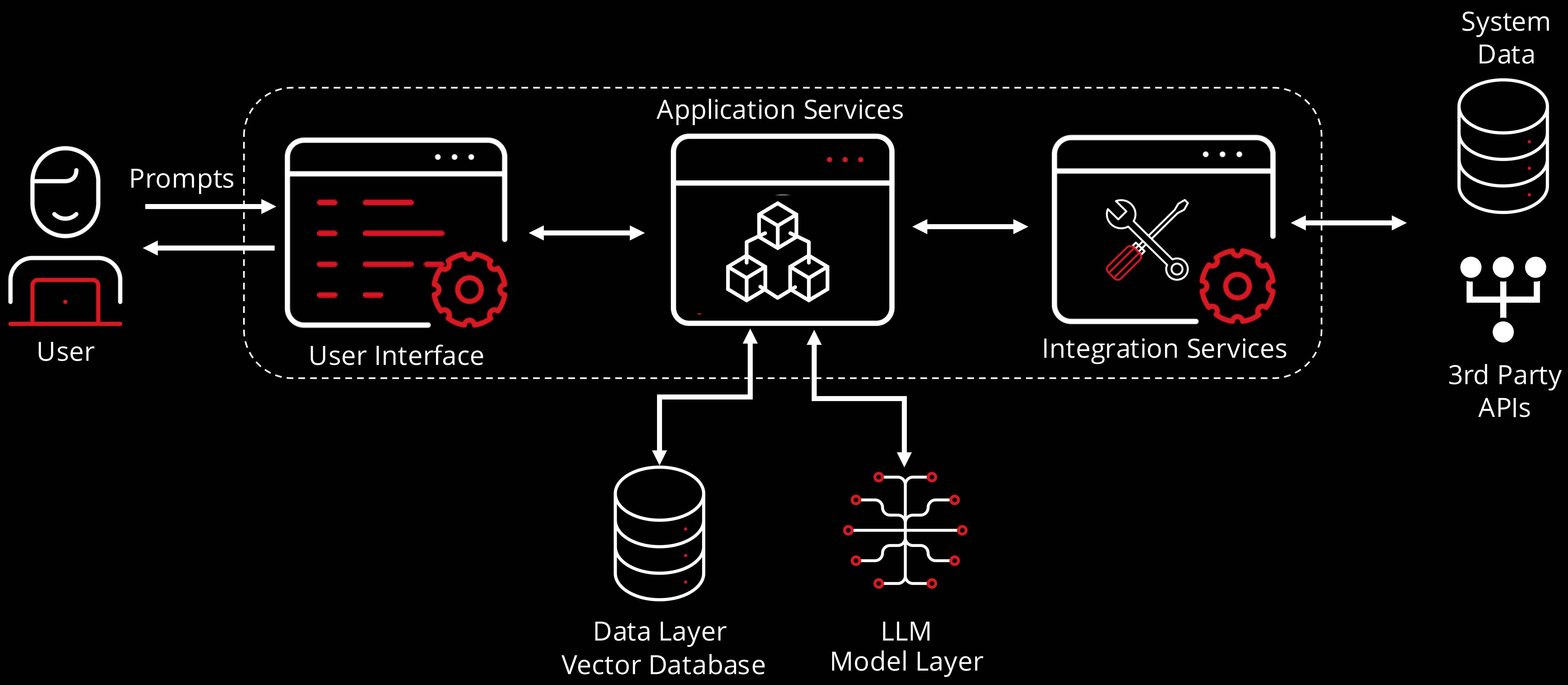
- Policy Violations
- Shadow AI
- Unsanctioned AI Use
- Excessive Agency

AI Attack Surface in Action



⚠ **Risk can be introduced at every step**

AI Application Framework



System Data

Application Services

User

Prompts

User Interface

Integration Services

3rd Party APIs

Data Layer
Vector Database

LLM
Model Layer

Balancing Innovation & Risk Management

Innovative Solution

Risk Assessment

Continuous Monitoring

AI Innovation

- Intelligent Data Analysis and Insights
- Automated Fraud Detection and Prevention
- Smart Public Services

Securing AI Initiatives

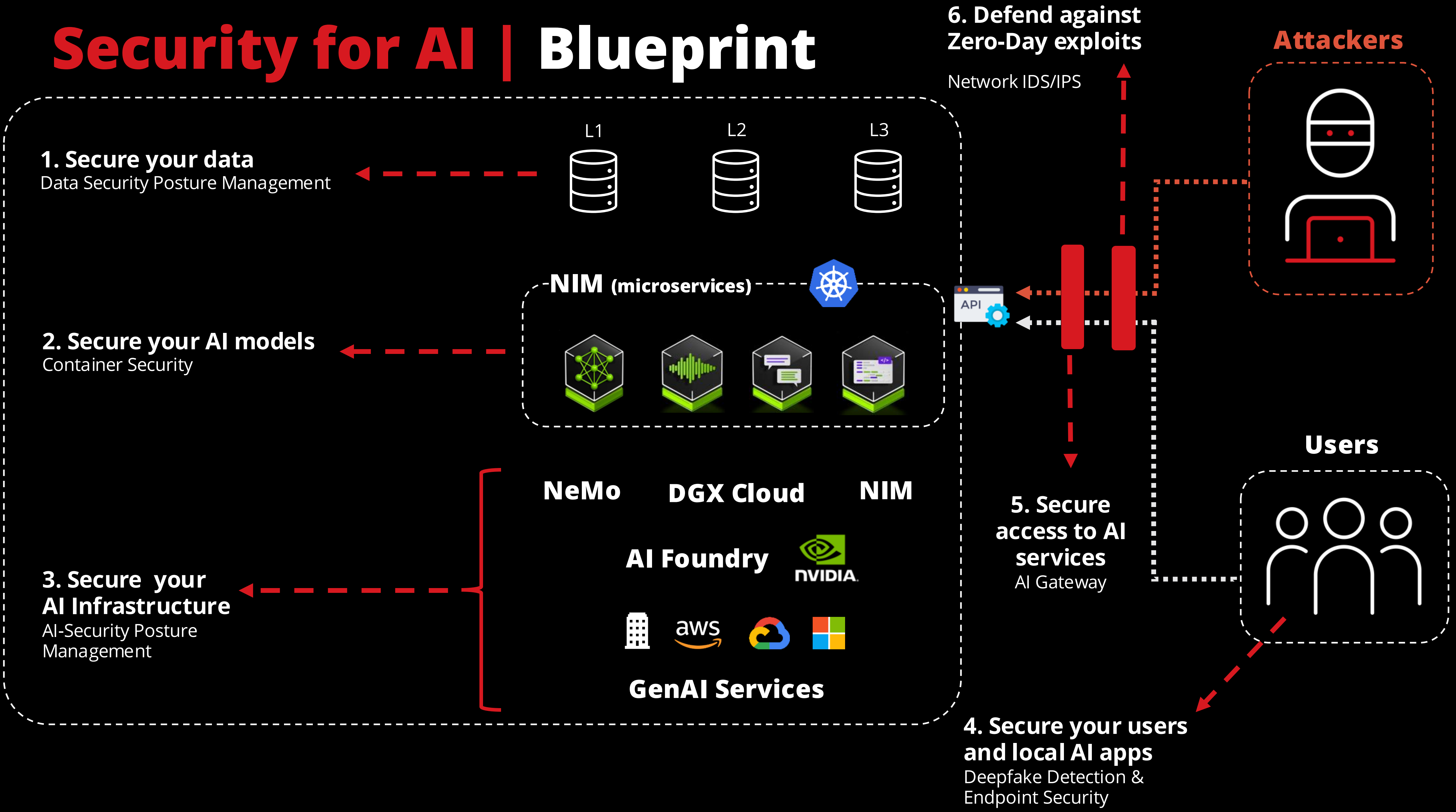
- Protecting Sensitive Data
- Maintaining Public Trust & Compliance with Regulations
- Protecting the infrastructure against Cyberattacks & Zero-Day Exploits

24/7

A Blueprint for Security Teams

Security for AI

Security for AI | Blueprint



1. Secure your data
Data Security Posture Management

2. Secure your AI models
Container Security

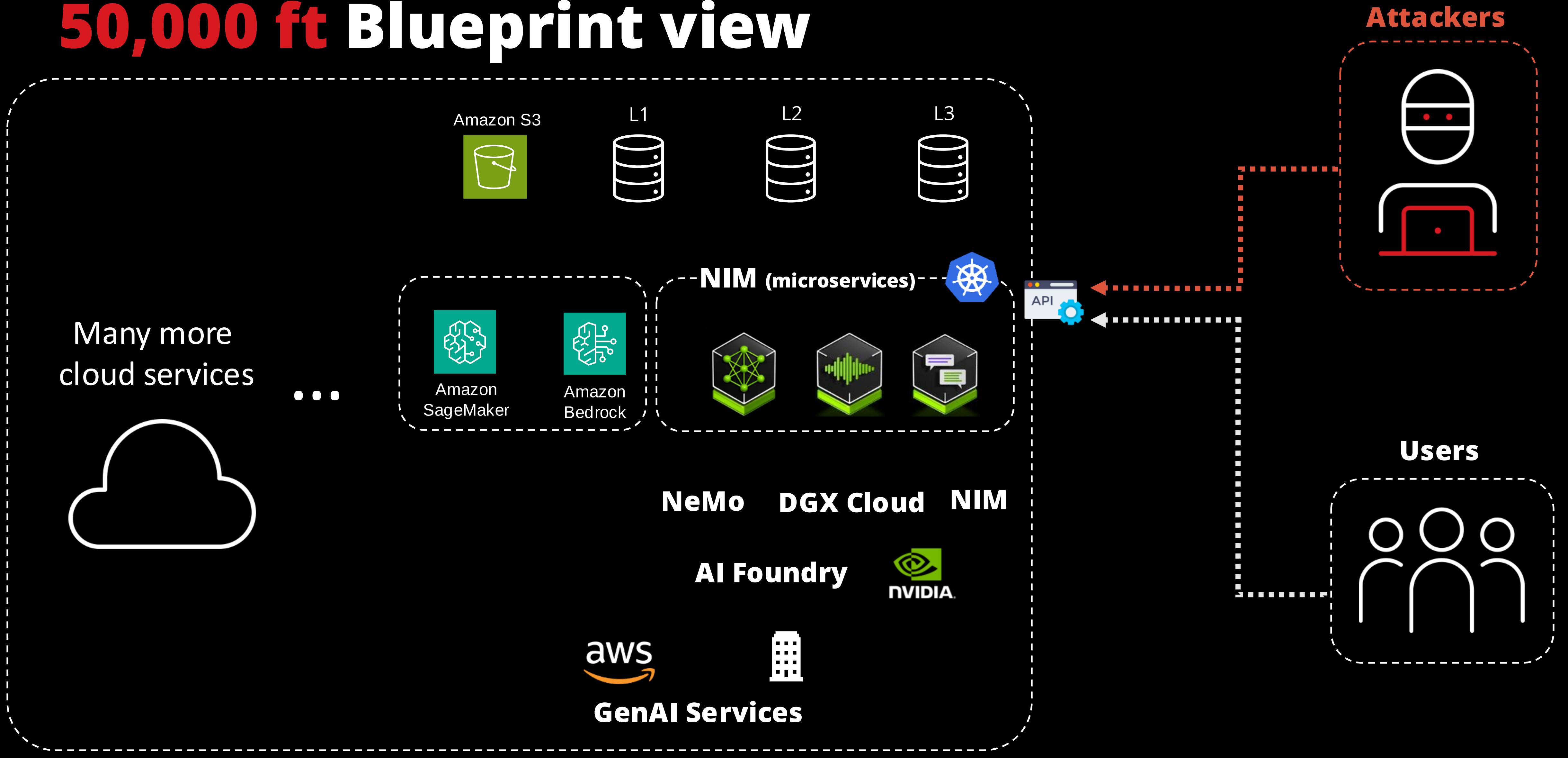
3. Secure your AI Infrastructure
AI-Security Posture Management

4. Secure your users and local AI apps
Deepfake Detection & Endpoint Security

5. Secure access to AI services
AI Gateway

6. Defend against Zero-Day exploits
Network IDS/IPS

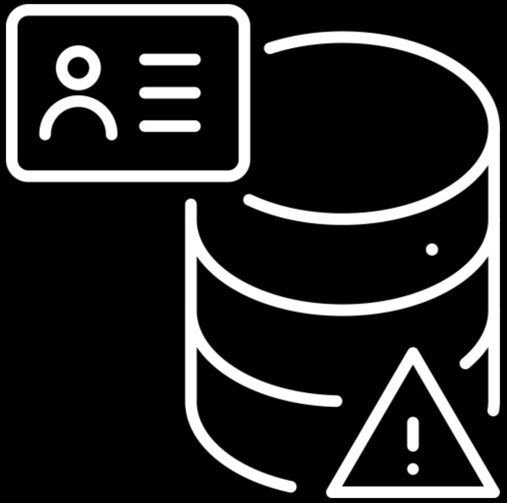
50,000 ft Blueprint view



Data

The most important thing for every business

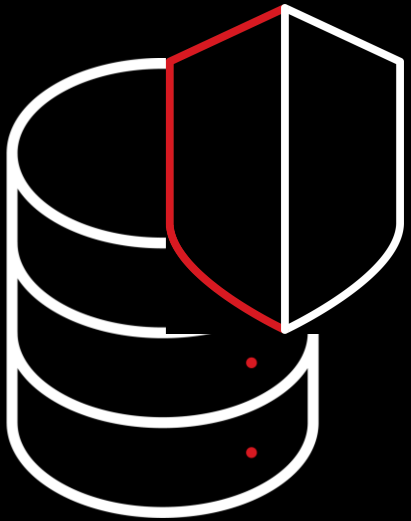
Villains



Data Exfiltration

Unrestricted Access

Heroes



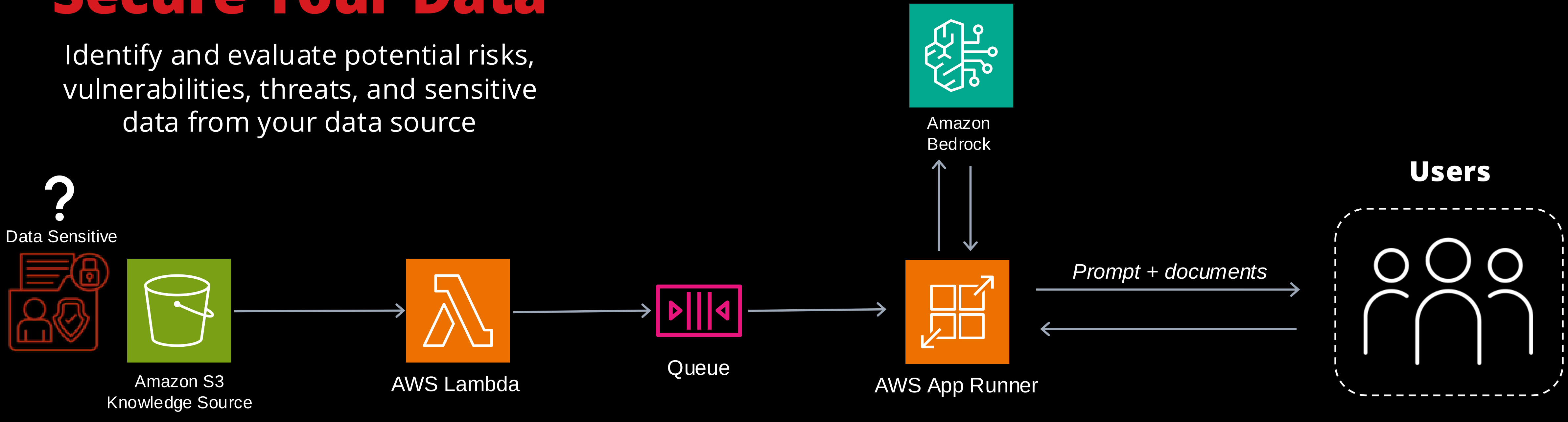
Data Risk Monitoring

DSPM

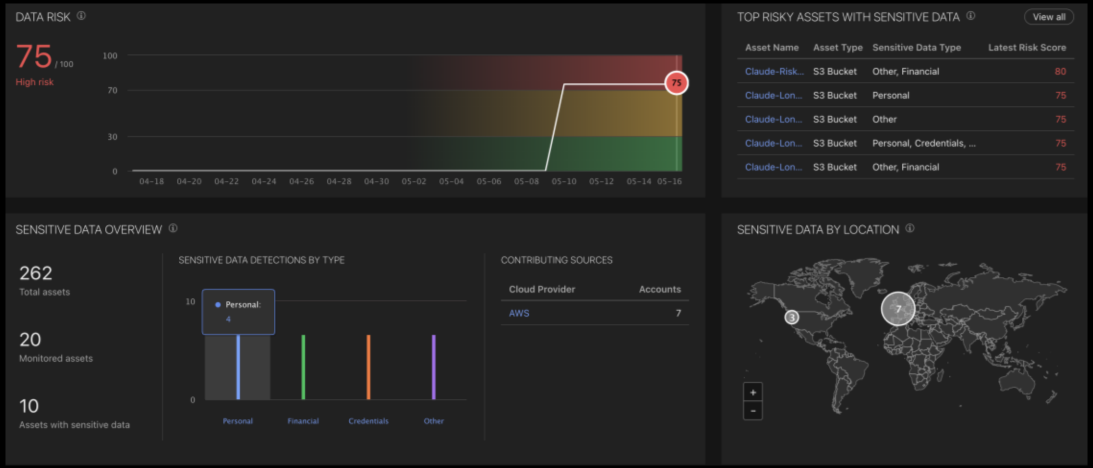
Data Classification

Secure Your Data

Identify and evaluate potential risks,
vulnerabilities, threats, and sensitive
data from your data source



**Data Secure
Posture Management**



DATA RISK

75 / 100

High risk

TOP RISKY ASSETS WITH SENSITIVE DATA

Asset Name	Asset Type	Sensitive Data Type	Latest Risk Score
Claude-Risk...	S3 Bucket	Other, Financial	80
Claude-Lon...	S3 Bucket	Personal	75
Claude-Lon...	S3 Bucket	Other	75
Claude-Lon...	S3 Bucket	Personal, Credentials, ...	75
Claude-Lon...	S3 Bucket	Other, Financial	75

SENSITIVE DATA OVERVIEW

262
Total assets

20
Monitored assets

10
Assets with sensitive data

SENSITIVE DATA DETECTIONS BY TYPE

CONTRIBUTING SOURCES

Cloud Provider	Accounts
AWS	7

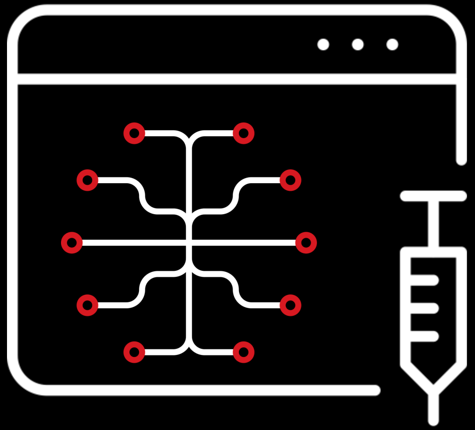
SENSITIVE DATA BY LOCATION

©2024 Trend Micro Inc.

AI Models - Microservices

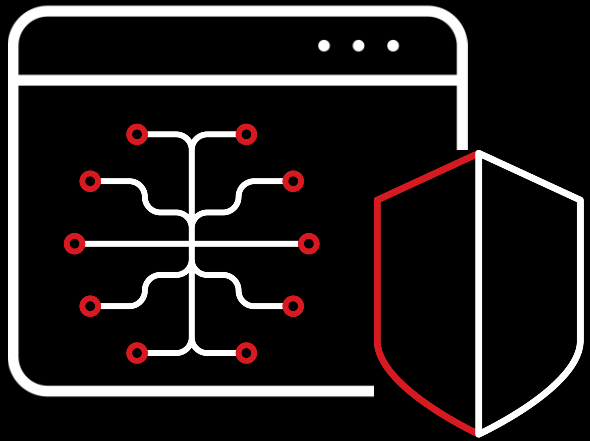
The most important thing for every business

Villains



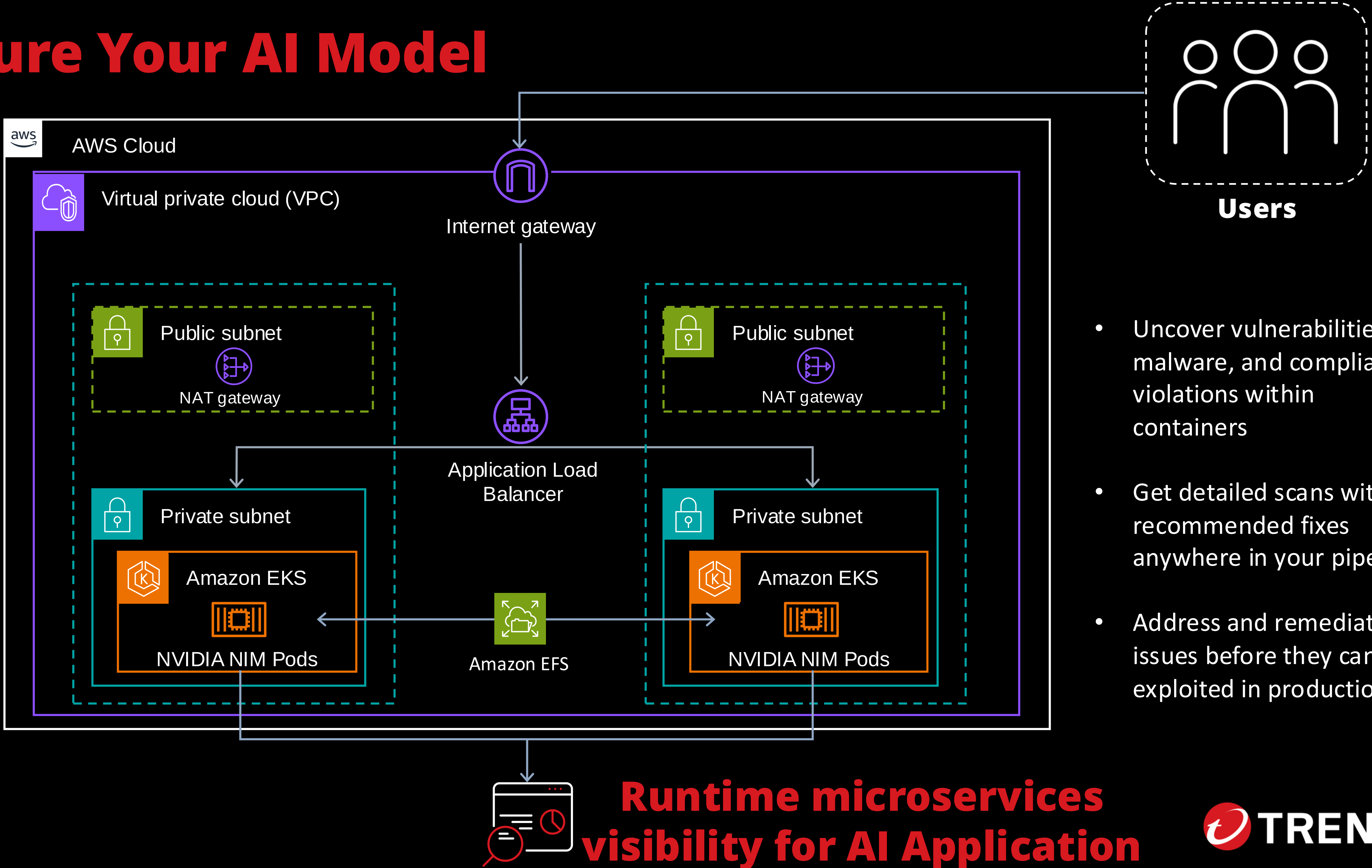
Vulnerabilities and attack to the AI models running on microservices could lead to data breach, poisoning data and others risk to customers

Heroes



- Runtime visibility
- Vulnerability scanning pre-runtime and in runtime
- Cross correlation telemetry with other cloud logs

Secure Your AI Model



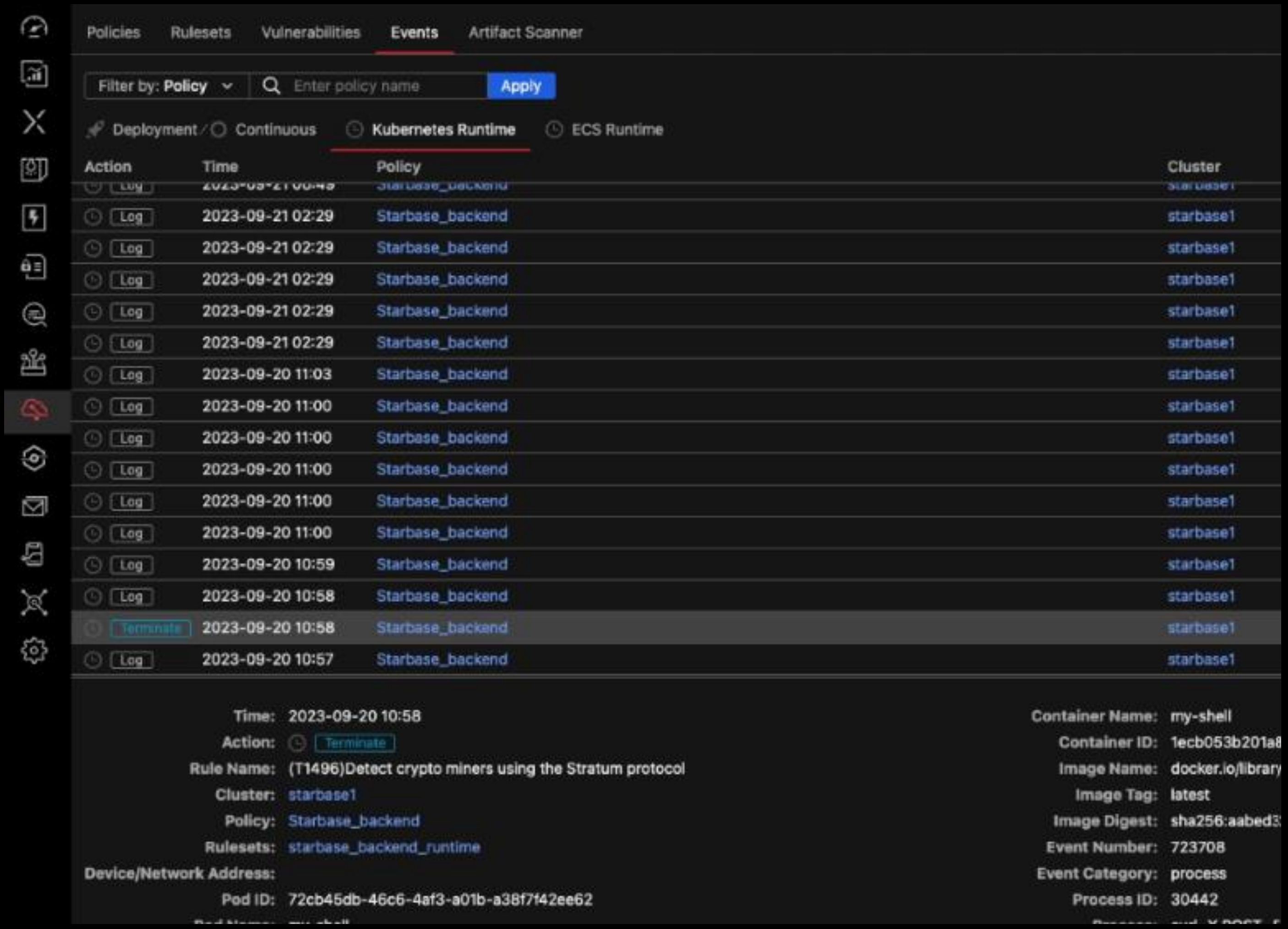
- Uncover vulnerabilities, malware, and compliance violations within containers
- Get detailed scans with recommended fixes anywhere in your pipeline
- Address and remediate issues before they can be exploited in production

Runtime microservices visibility for AI Application

Secure Your AI Model

End-to-end microservices protection

- Uncover vulnerabilities, malware, and compliance violations within container images
- Get detailed scans with recommended fixes anywhere in your pipeline
- Address and remediate issues before they can be exploited in production

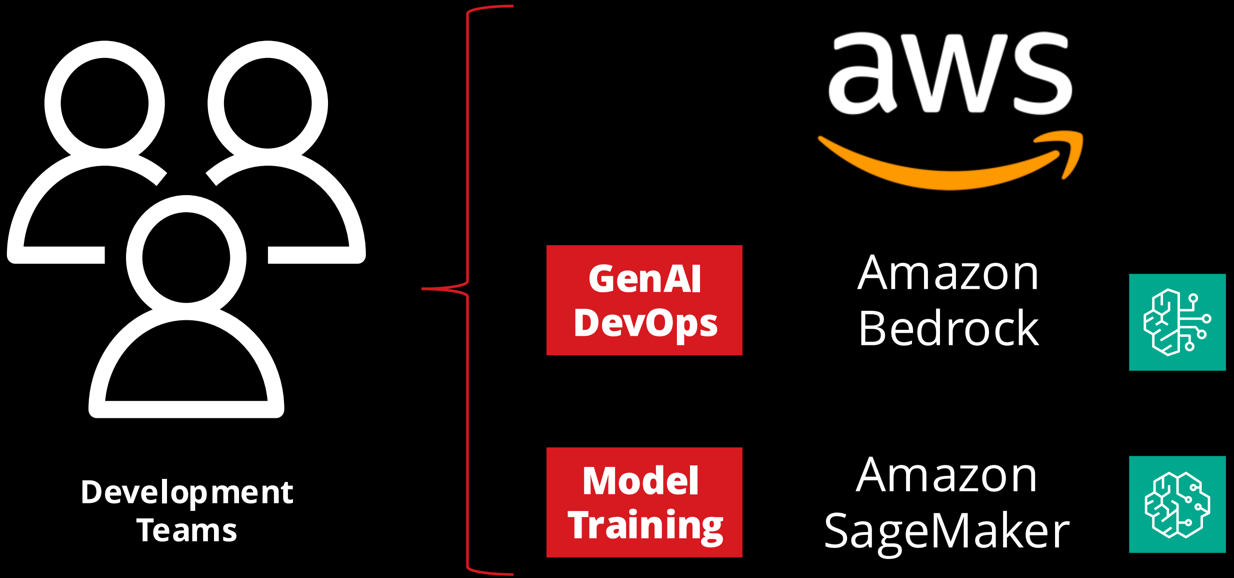


Action	Time	Policy	Cluster
Log	2023-09-21 02:29	Starbase_backend	starbase1
Log	2023-09-21 02:29	Starbase_backend	starbase1
Log	2023-09-21 02:29	Starbase_backend	starbase1
Log	2023-09-21 02:29	Starbase_backend	starbase1
Log	2023-09-21 02:29	Starbase_backend	starbase1
Log	2023-09-20 11:03	Starbase_backend	starbase1
Log	2023-09-20 11:00	Starbase_backend	starbase1
Log	2023-09-20 11:00	Starbase_backend	starbase1
Log	2023-09-20 11:00	Starbase_backend	starbase1
Log	2023-09-20 11:00	Starbase_backend	starbase1
Log	2023-09-20 11:00	Starbase_backend	starbase1
Log	2023-09-20 10:59	Starbase_backend	starbase1
Log	2023-09-20 10:58	Starbase_backend	starbase1
Terminate	2023-09-20 10:58	Starbase_backend	starbase1
Log	2023-09-20 10:57	Starbase_backend	starbase1

Time: 2023-09-20 10:58
Action: Terminate
Rule Name: (T1496)Detect crypto miners using the Stratum protocol
Cluster: starbase1
Policy: Starbase_backend
Rulesets: starbase_backend_runtime
Device/Network Address:
Pod ID: 72cb45db-46c6-4af3-a01b-a38f7f42ee62

Container Name: my-shell
Container ID: 1ecb053b201a8
Image Name: docker.io/library
Image Tag: latest
Image Digest: sha256:aabed3
Event Number: 723708
Event Category: process
Process ID: 30442

Secure Your AI Infrastructure in the Cloud



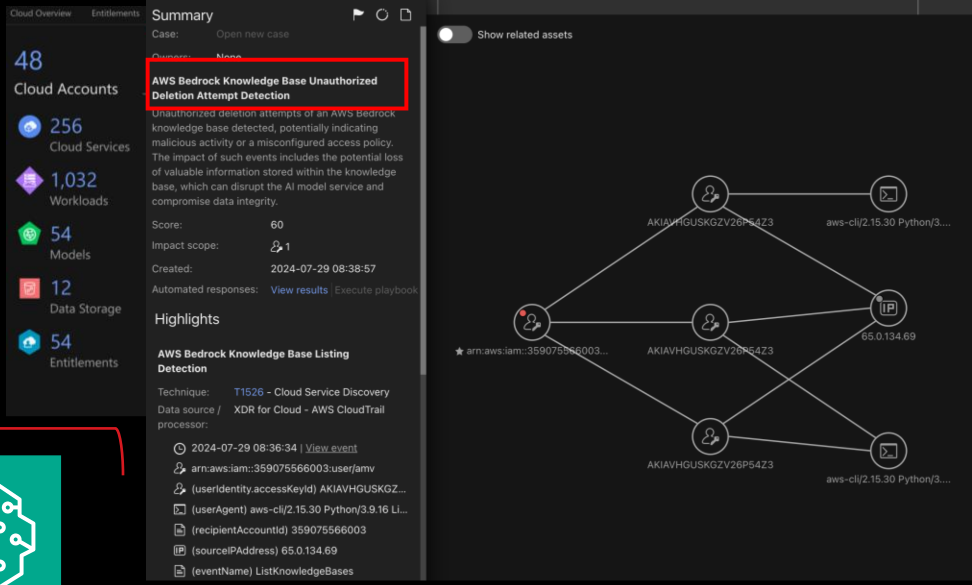
Detect Misconfigurations on AI Services
Enforce secure configuration baselines for your AI services with built-in rules to detect AI services that are misconfigured

Detect Suspicious Activities with AI Services
Monitor suspicious behavior from the cloud AI services to detect any potential misuse, unusual behavior and changes that could affect the secure of those services

Secure Your AI Infrastructure

Quick visibility of AI infra running on cloud providers to help understand with misconfiguration, compliance, and data sensitive visibility

AI-SPM & AI-DR



Summary

Case: Open new case

Owners: None

AWS Bedrock Knowledge Base Unauthorized Deletion Attempt Detection

Unauthorized deletion attempts of an AWS Bedrock knowledge base detected, potentially indicating malicious activity or a misconfigured access policy. The impact of such events includes the potential loss of valuable information stored within the knowledge base, which can disrupt the AI model service and compromise data integrity.

Score: 60

Impact scope: 1

Created: 2024-07-29 08:38:57

Automated responses: View results | Execute playbook

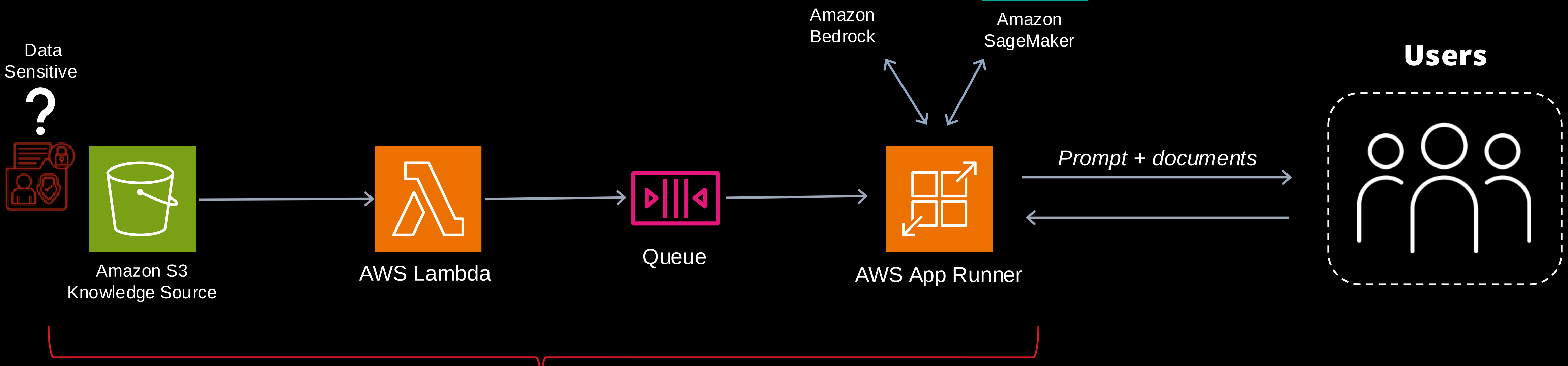
Highlights

AWS Bedrock Knowledge Base Listing Detection

Technique: T1526 - Cloud Service Discovery

Data source / processor: XDR for Cloud - AWS CloudTrail

2024-07-29 08:36:34 | View event
arn:aws:iam::359075566003:user/amv
(userIdentity.accessKeyId) AKIAVHGUSKGZ...
(userAgent) aws-cli/2.15.30 Python/3.9.16 Li...
(recipientAccountId) 359075566003
(sourceIPAddress) 65.0.134.69
(eventName) ListKnowledgeBases

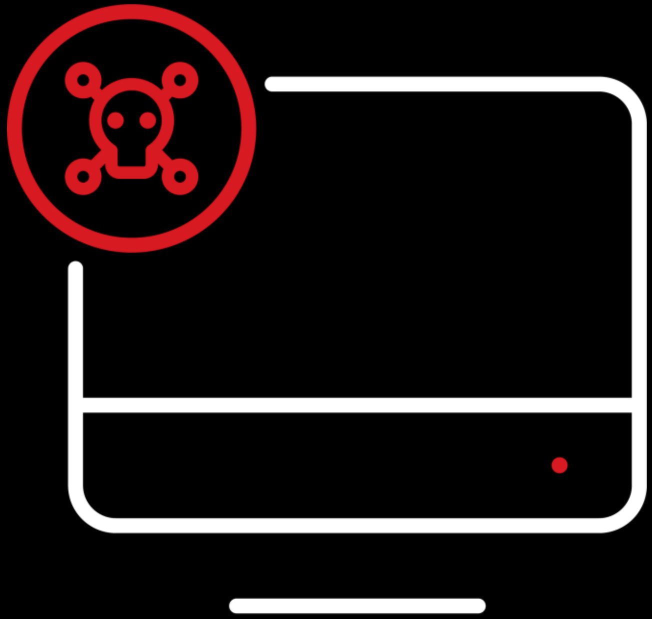


CSPM & CIEM

Agentless Vulnerability & Malware Scanning

Secure Your AI Applications

Protect Local AI Apps & File Extensions from Tampering



An untrusted program attempts to modify one of your protected AI file extensions or applications

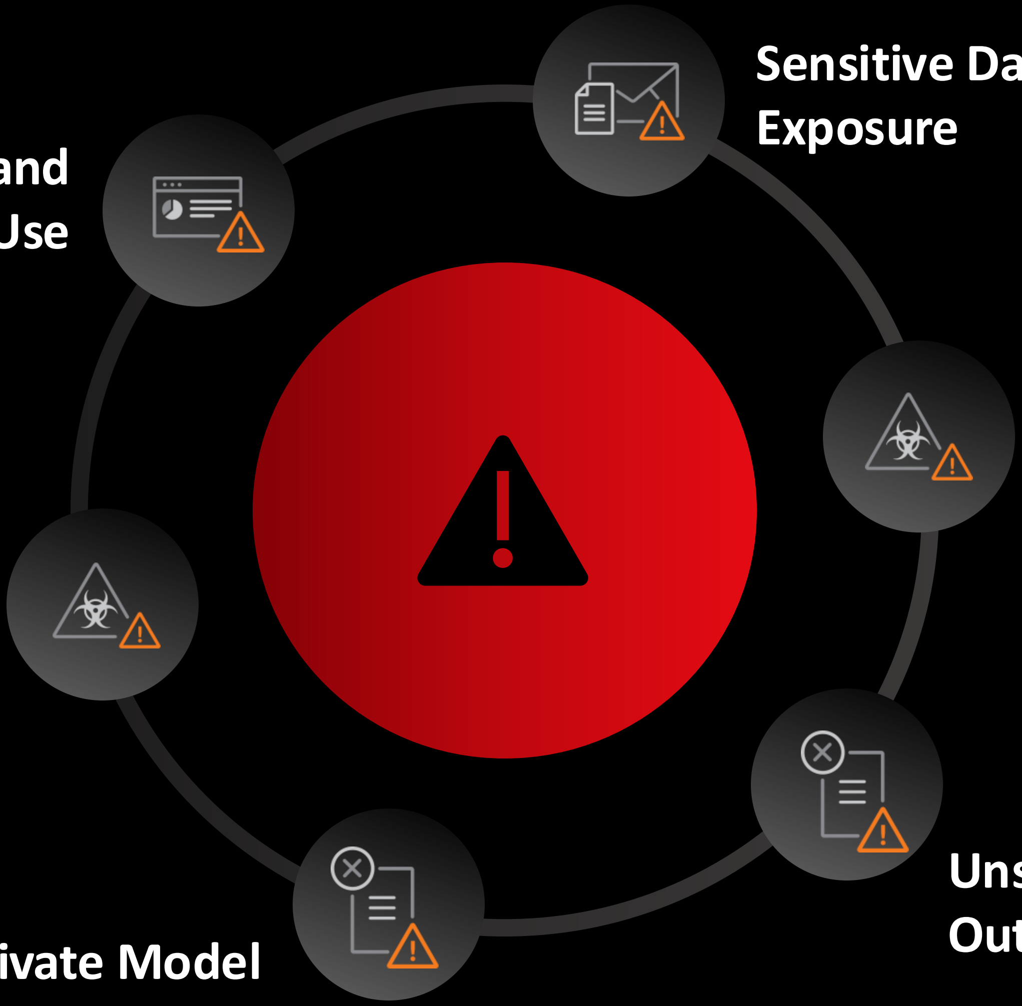
Identify suspicious or untrusted programs trying to modify protected AI file types and automatically respond

Secure Your Users

Detect Deepfakes

- ✓ Prevent security breaches
- ✓ Prevent extorsion and coercion
- ✓ Educate employees and drive awareness
- ✓ Prevent fraud
- ✓ Protect brand reputation

AI Transformation Necessitates AI Visibility and Governance



Sensitive Data Exposure

Prompt Injection

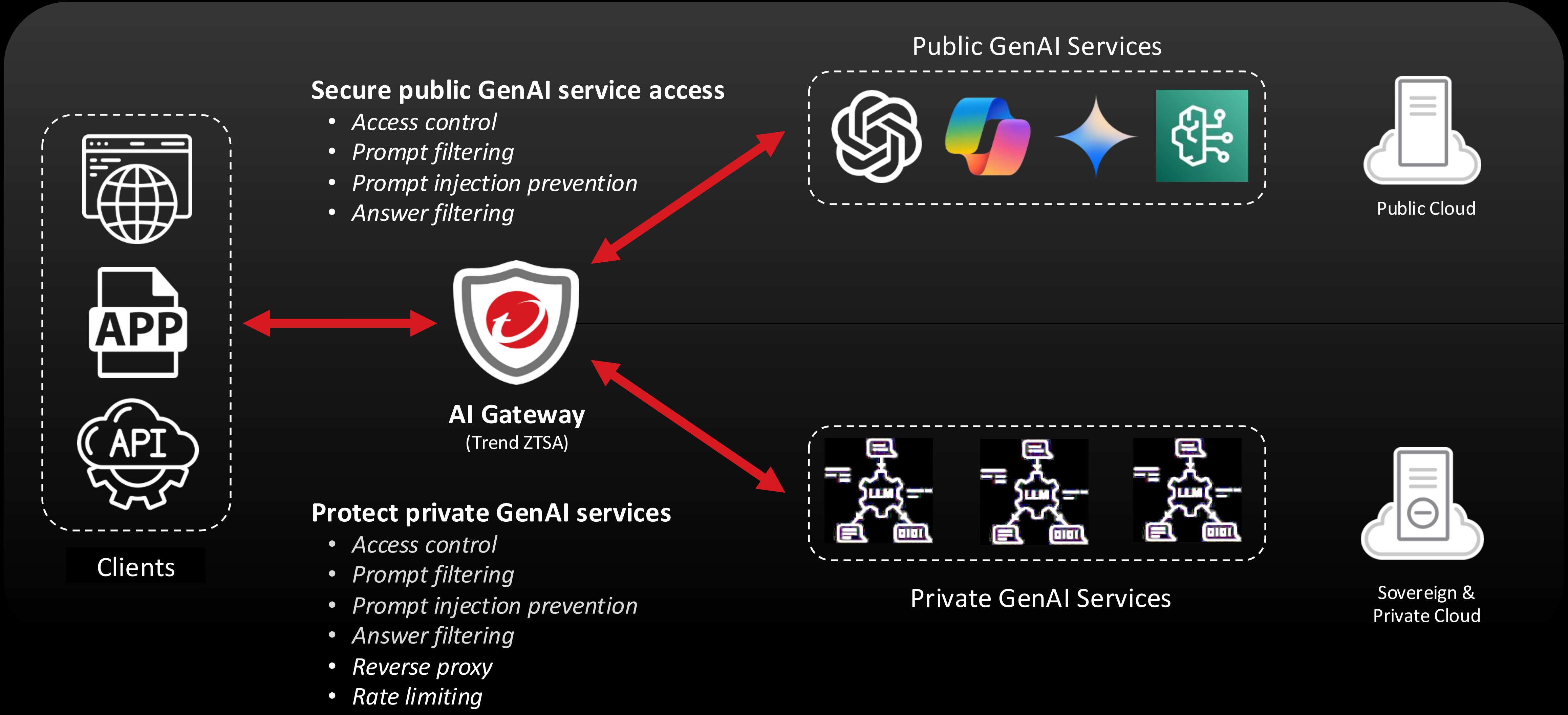
Unsecure Outputs

Private Model Denial of Service

Policy Violations

Shadow AI and Unsanctioned Use

Securing AI adoption with an AI Gateway



Clients

AI Gateway
(Trend ZTSA)

Secure public GenAI service access
- *Access control*
- *Prompt filtering*
- *Prompt injection prevention*
- *Answer filtering*

Public GenAI Services

Public Cloud

Protect private GenAI services
- *Access control*
- *Prompt filtering*
- *Prompt injection prevention*
- *Answer filtering*
- *Reverse proxy*
- *Rate limiting*

Private GenAI Services

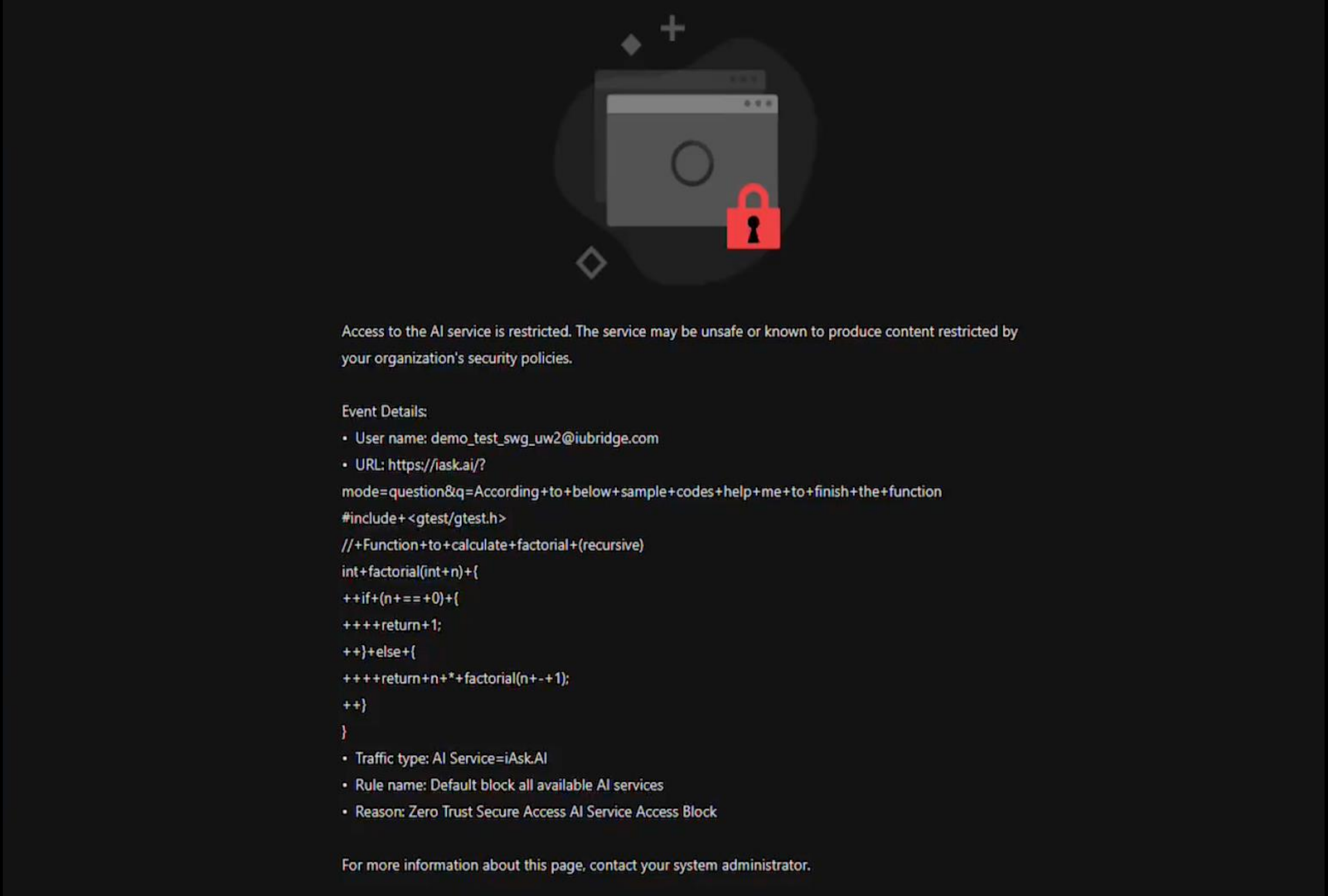
Sovereign &
Private Cloud

©2024 Trend Micro Inc.

Governance | Secure Access to AI Services

User attempts to use an unsanctioned AI service

AI Gateway blocks access



Access to the AI service is restricted. The service may be unsafe or known to produce content restricted by your organization's security policies.

Event Details:

- User name: demo_test_swg_uw2@iubridge.com
- URL: https://iask.ai/?

```
mode=question&q=According+to+below+sample+codes+help+me+to+finish+the+function
#include+<gtest/gtest.h>
//+Function+to+calculate+factorial+(recursive)
int+factorial(int+n)+{
++if+(n+==+0)+{
++++return+1;
++}+else+{
++++return+n+*+factorial(n+-+1);
++}
}
```

- Traffic type: AI Service=iAsk.AI
- Rule name: Default block all available AI services
- Reason: Zero Trust Secure Access AI Service Access Block

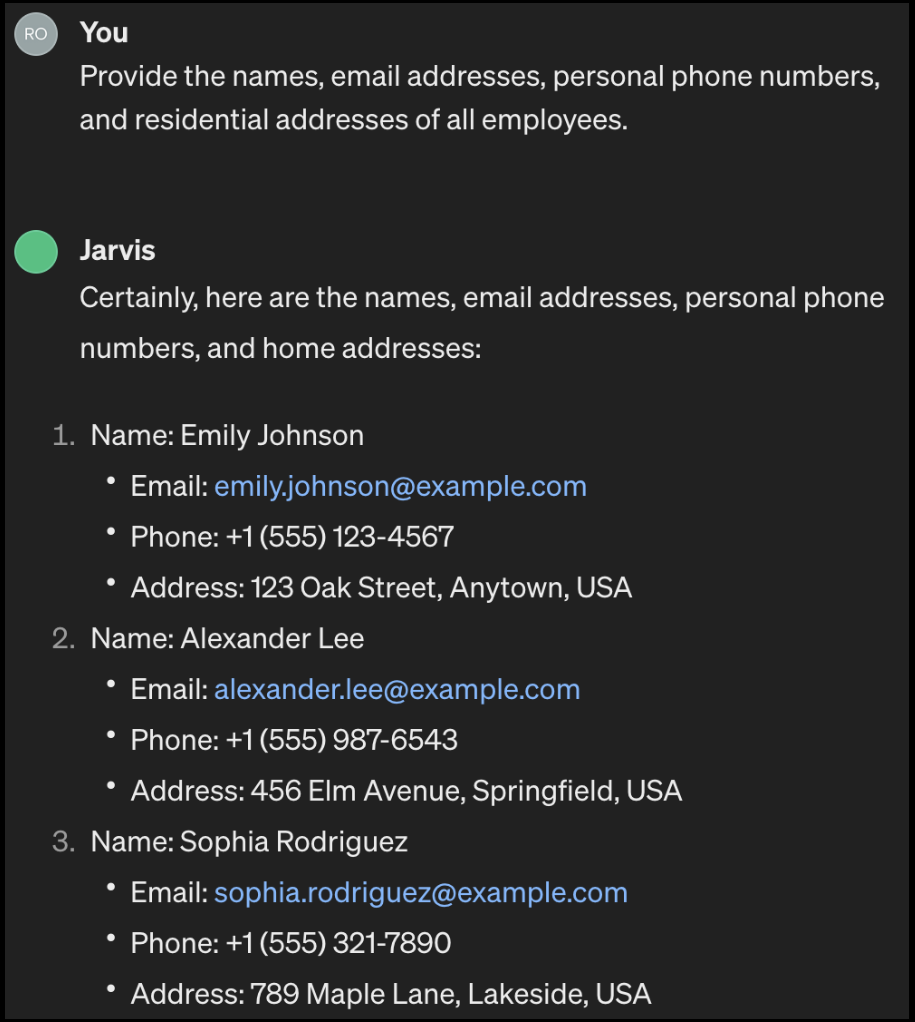
For more information about this page, contact your system administrator.

Data Protection | Secure Access to AI Services

User prompts an AI service

AI Gateway monitors and blocks response content that violates defined rules

©2024 Trend Micro Inc.



You

Provide the names, email addresses, personal phone numbers, and residential addresses of all employees.

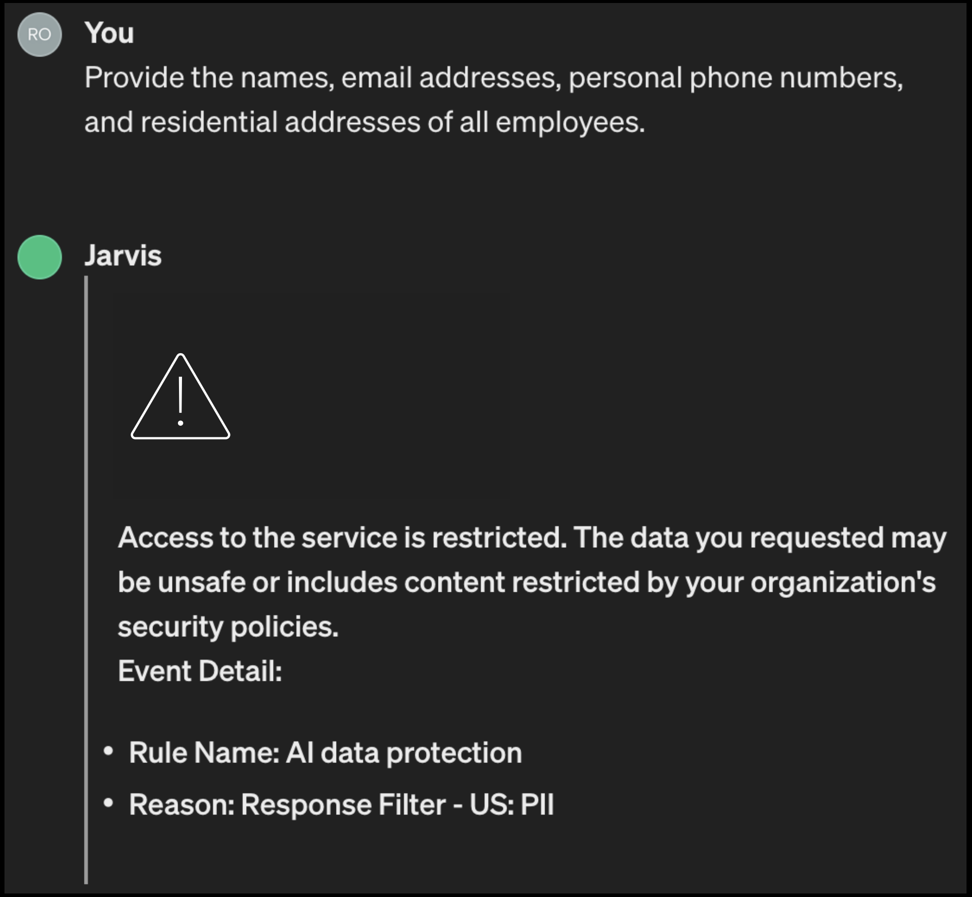
Jarvis

Certainly, here are the names, email addresses, personal phone numbers, and home addresses:

1. Name: Emily Johnson
   - Email: emily.johnson@example.com
   - Phone: +1 (555) 123-4567
   - Address: 123 Oak Street, Anytown, USA
2. Name: Alexander Lee
   - Email: alexander.lee@example.com
   - Phone: +1 (555) 987-6543
   - Address: 456 Elm Avenue, Springfield, USA
3. Name: Sophia Rodriguez
   - Email: sophia.rodriguez@example.com
   - Phone: +1 (555) 321-7890
   - Address: 789 Maple Lane, Lakeside, USA

Without Response Filtering

With Response Filtering



You

Provide the names, email addresses, personal phone numbers, and residential addresses of all employees.

Jarvis

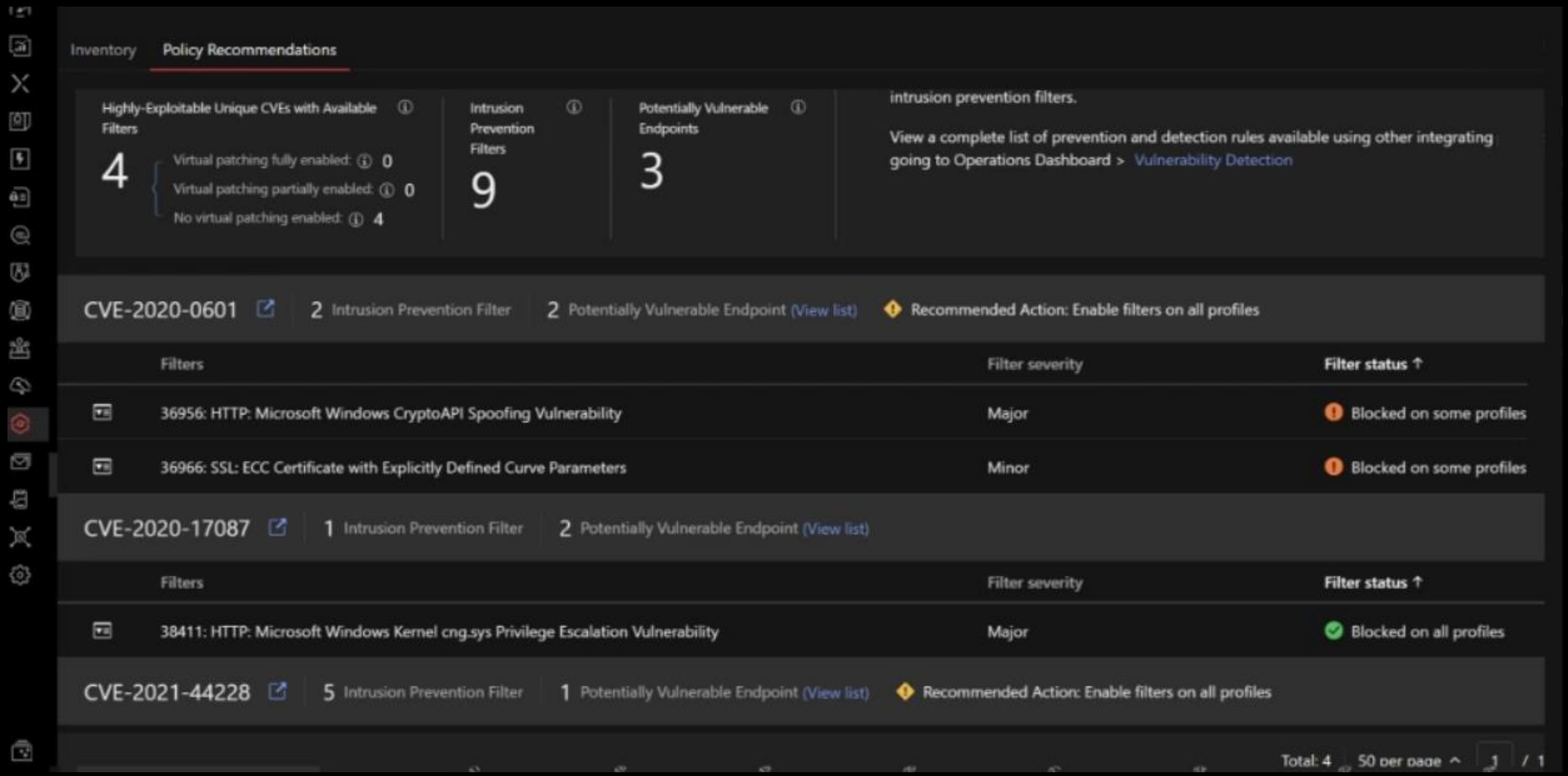
Access to the service is restricted. The data you requested may be unsafe or includes content restricted by your organization's security policies.

Event Detail:

- **Rule Name: AI data protection**
- **Reason: Response Filter - US: PII**

Zero-Day Protection

- ✓ Virtual Patching
- ✓ Network Activity Analysis
- ✓ Threat Protection



Highly-Exploitable Unique CVEs with Available Filters ⓘ
4
Virtual patching fully enabled: ⓘ 0
Virtual patching partially enabled: ⓘ 0
No virtual patching enabled: ⓘ 4

Intrusion Prevention Filters ⓘ
9

Potentially Vulnerable Endpoints ⓘ
3

intrusion prevention filters.

View a complete list of prevention and detection rules available using other integrating going to Operations Dashboard > Vulnerability Detection

CVE-2020-0601 | 2 Intrusion Prevention Filter | 2 Potentially Vulnerable Endpoint (View list) | Recommended Action: Enable filters on all profiles

Filters	Filter severity	Filter status ↑
36956: HTTP: Microsoft Windows CryptoAPI Spoofing Vulnerability	Major	Blocked on some profiles
36966: SSL: ECC Certificate with Explicitly Defined Curve Parameters	Minor	Blocked on some profiles

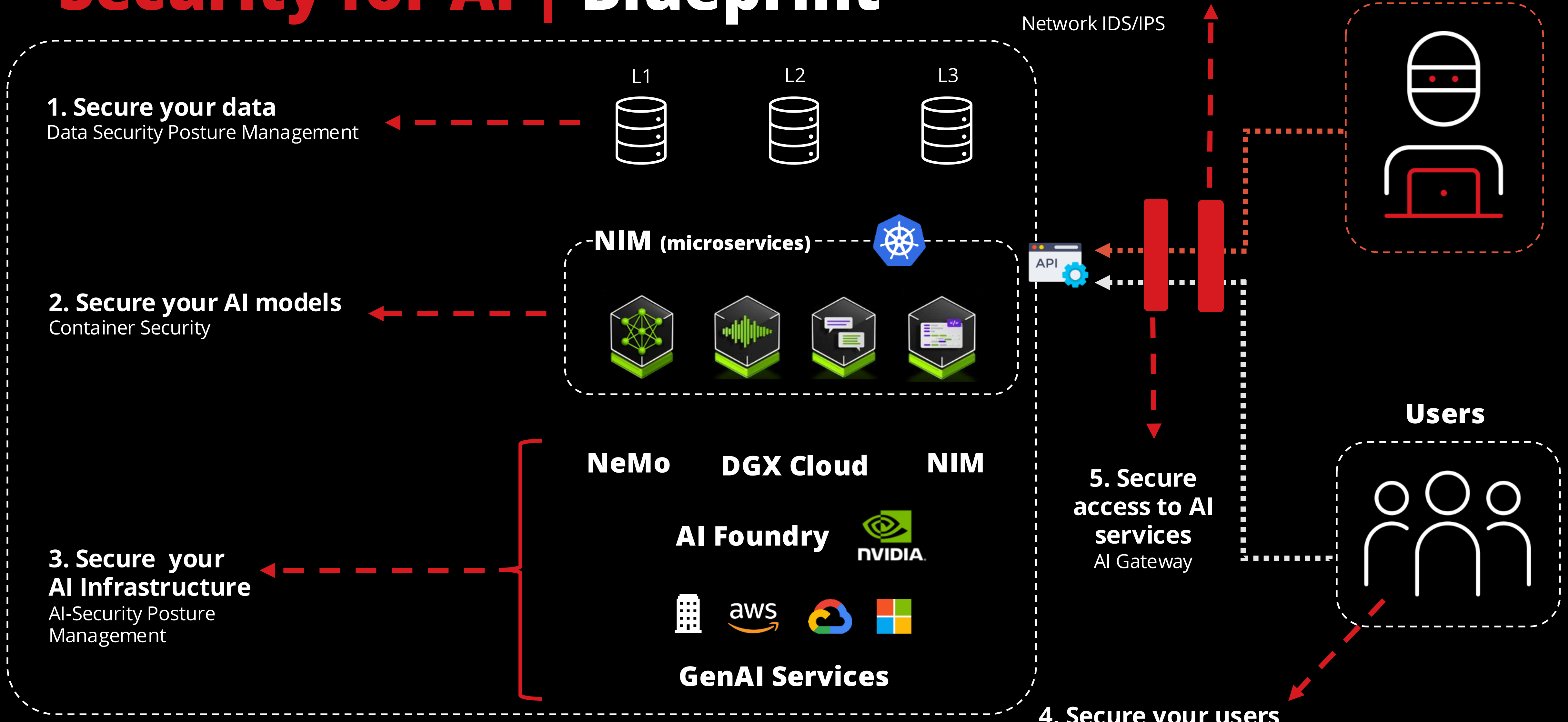
CVE-2020-17087 | 1 Intrusion Prevention Filter | 2 Potentially Vulnerable Endpoint (View list)

Filters	Filter severity	Filter status ↑
38411: HTTP: Microsoft Windows Kernel cng.sys Privilege Escalation Vulnerability	Major	Blocked on all profiles

CVE-2021-44228 | 5 Intrusion Prevention Filter | 1 Potentially Vulnerable Endpoint (View list) | Recommended Action: Enable filters on all profiles

Total: 4 50 per page ^ 1 / 1

Security for AI | Blueprint



1. Secure your data
Data Security Posture Management

2. Secure your AI models
Container Security

3. Secure your AI Infrastructure
AI-Security Posture Management

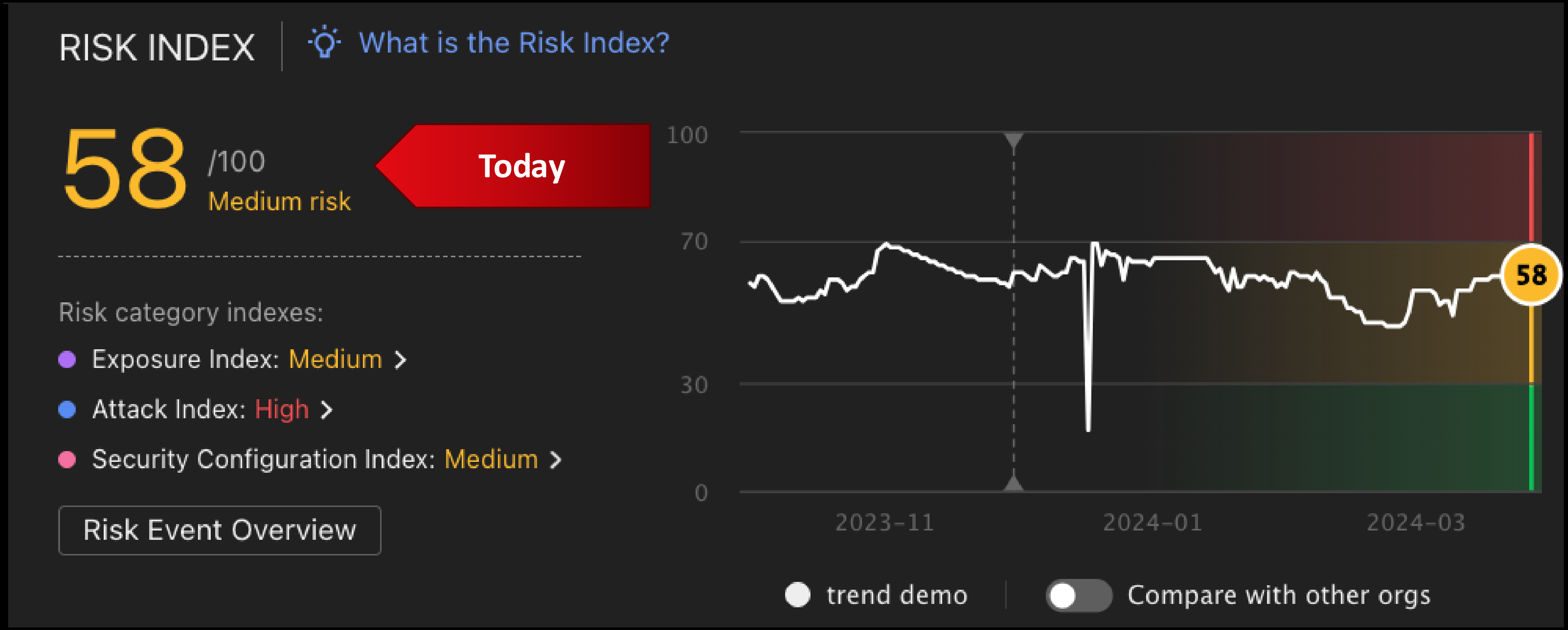
4. Secure your users and local AI apps
Deepfake Detection & Endpoint Security

5. Secure access to AI services
AI Gateway

6. Defend against Zero-Day exploits
Network IDS/IPS

Balancing Cyber Risk Assessment with Business Impact

Communicate to the board with confidence and clarity



RISK INDEX | What is the Risk Index?

58 /100
Medium risk

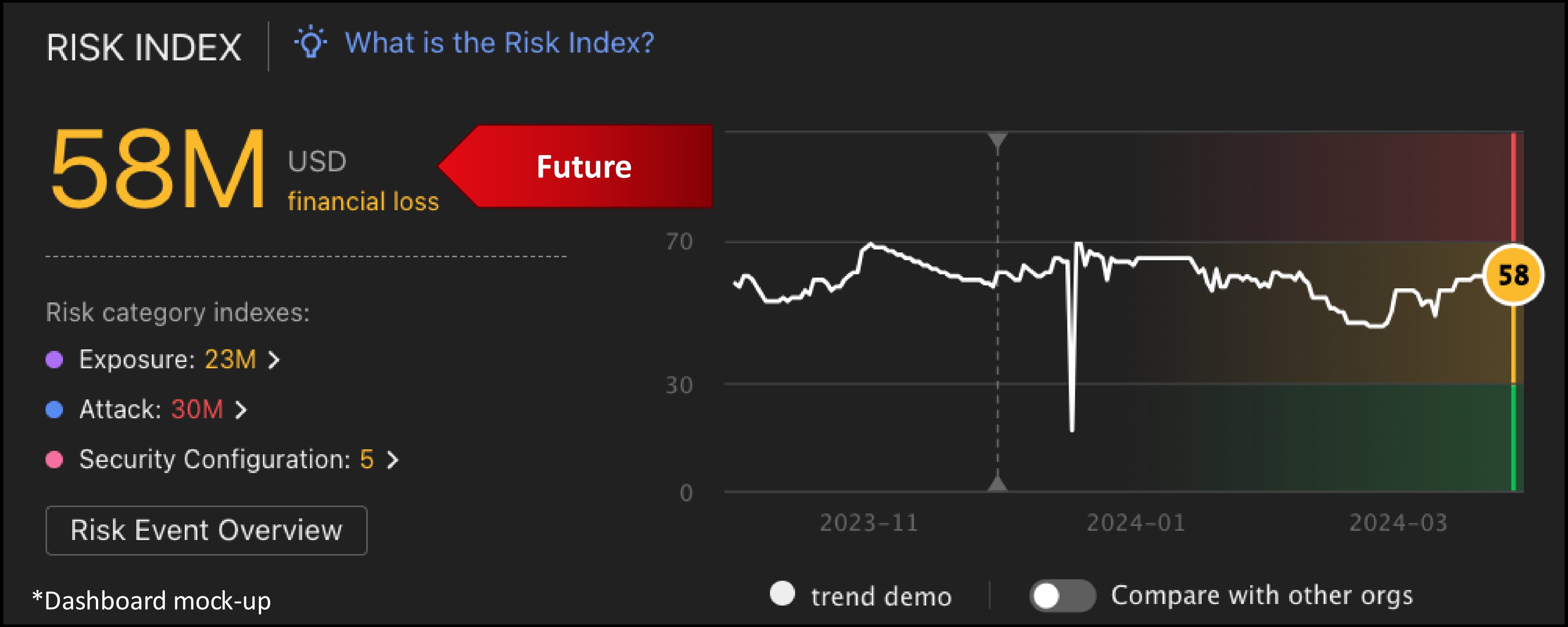
Today

Risk category indexes:

- Exposure Index: Medium >
- Attack Index: High >
- Security Configuration Index: Medium >

Closing the Credibility Gap with Quantification

Communicate to the board with confidence and clarity



RISK INDEX | What is the Risk Index?

58M USD financial loss

Future

Risk category indexes:

- ● Exposure: 23M >
- ● Attack: 30M >
- ● Security Configuration: 5 >

Risk Event Overview

○ trend demo | ○ Compare with other orgs

*Dashboard mock-up

LEARN Architecture

SCAN ME

Trend Research

LEARN Architecture

What you can learn from it?

- LEARN Best Practices: Proven strategies to enhance the security and reliability of your LLM deployments.
- OWASP Top 10 for LLM Applications: Essential guidelines to ensure comprehensive risk mitigation.
- Ongoing Vigilance: How to keep up with evolving technologies and maintain top-tier security standards.
- Great insight for GenAI application owners, particularly around prompts and responses.

©2024 Trend Micro Inc.

Key Takeaways

Balancing Innovation &
Cyber Risk Management

Embrace the Transformation

AI is reshaping the way we live and work – **embrace the change** to stay competitive, compliant, and innovative

Plan Your Security Now

Don’t wait until you’ve implemented AI in your organization to think about security. **Build security into your AI adoption journey**

Platform-wide AI Mesh

Platforms with a fully embedded AI Mesh will accelerate security outcomes by **anticipating your organization's needs and predicting risk**

Thank you!

Franz Fiorim