Microsoft

Security Focus: Using and Monitoring AI

Strange Case of Dr Sentinel and Mr AI

Rod Trent
Senior Program Manager
Microsoft

About Me…

Name: Rod Trent
Professional Title: Senior Program Manager, Cybersecurity and AI
LinkedIn Profile: https://www.linkedin.com/in/rodtrent/
Twitter: https://twitter.com/rodtrent



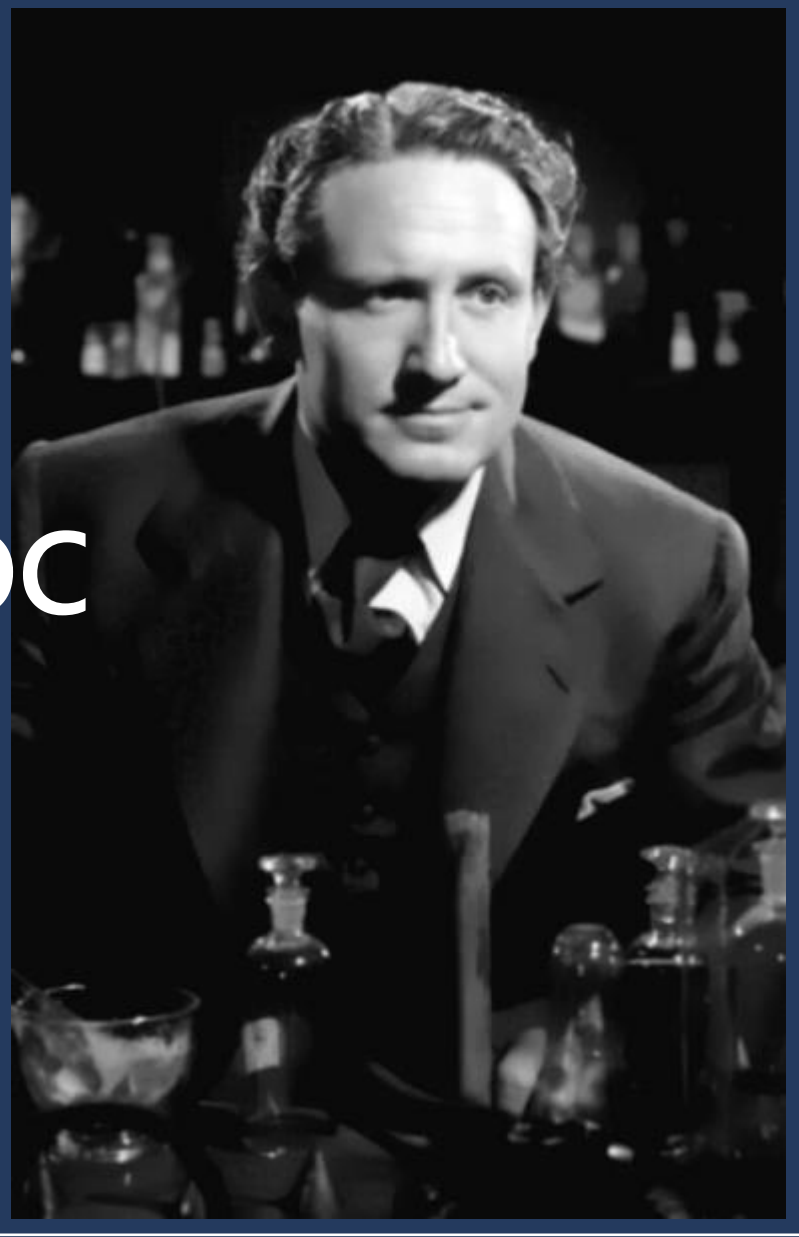

- Husband
- Dad
- Grandfather (G-pop)
- Running freak
- KQL nut
- 6MDM

My blog: https://aka.ms/RodsBlog
Microsoft Sentinel this Week (newsletter):
http://aka.ms/MicrosoftSentinelNewsletter
Microsoft Defender (newsletter)
http://aka.ms/MicrosoftDefenderNewsletter
Azure Open AI Weekly Community Copilot (newsletter)
https://aka.ms/AzOpenAINews

myITforum.com
Powered by you

Microsoft

Part 1: Using AI to Deliver SOC Intelligence and Efficiency



Dr Sentinel

Microsoft

Part 2: Monitoring and Securing AI Using Modern Tools

Mr AI

Level Set

Will AI blow up the world and take over humanity?

Super-ugly AI-generated picture

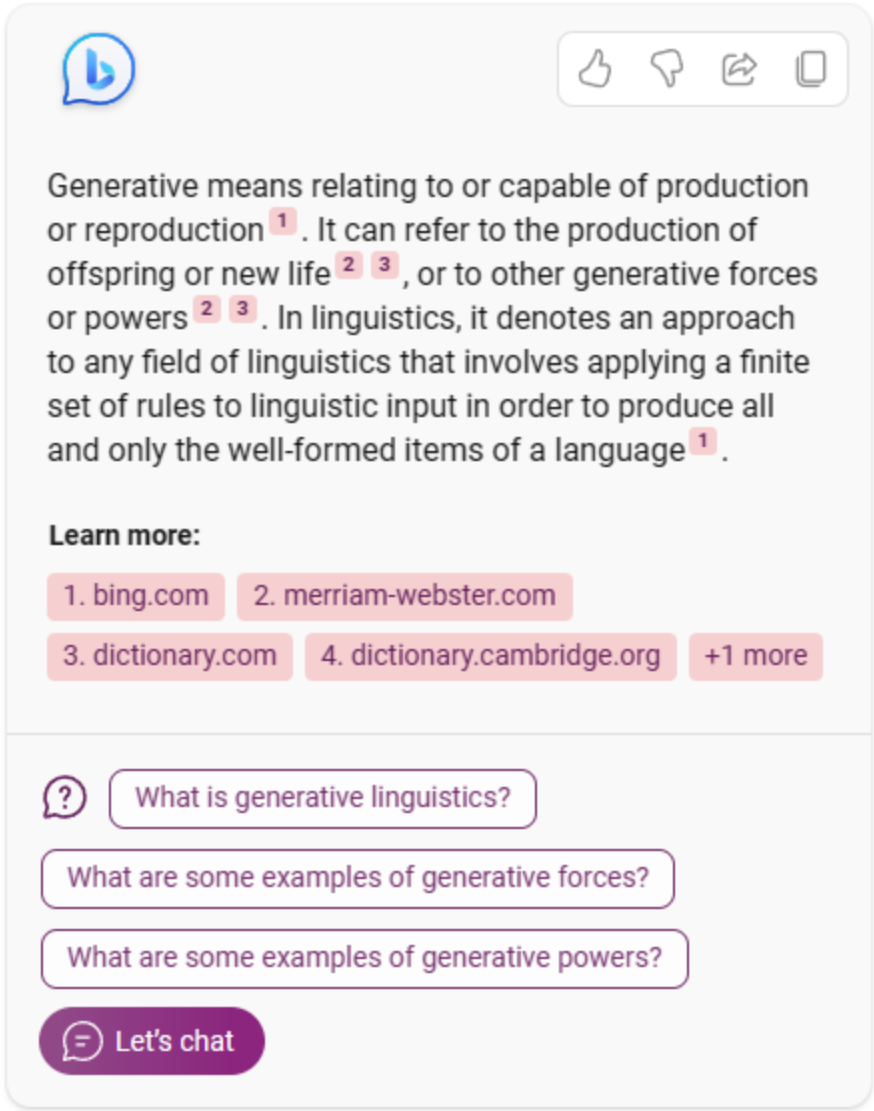
Definition

gen·er·a·tive

ADJECTIVE

relating to or capable of production or reproduction

denoting an approach to any field of linguistics that involves applying a finite set of rules to linguistic input in order to produce all and only the well-formed items of a language: "generative phonology"



What is generative AI?

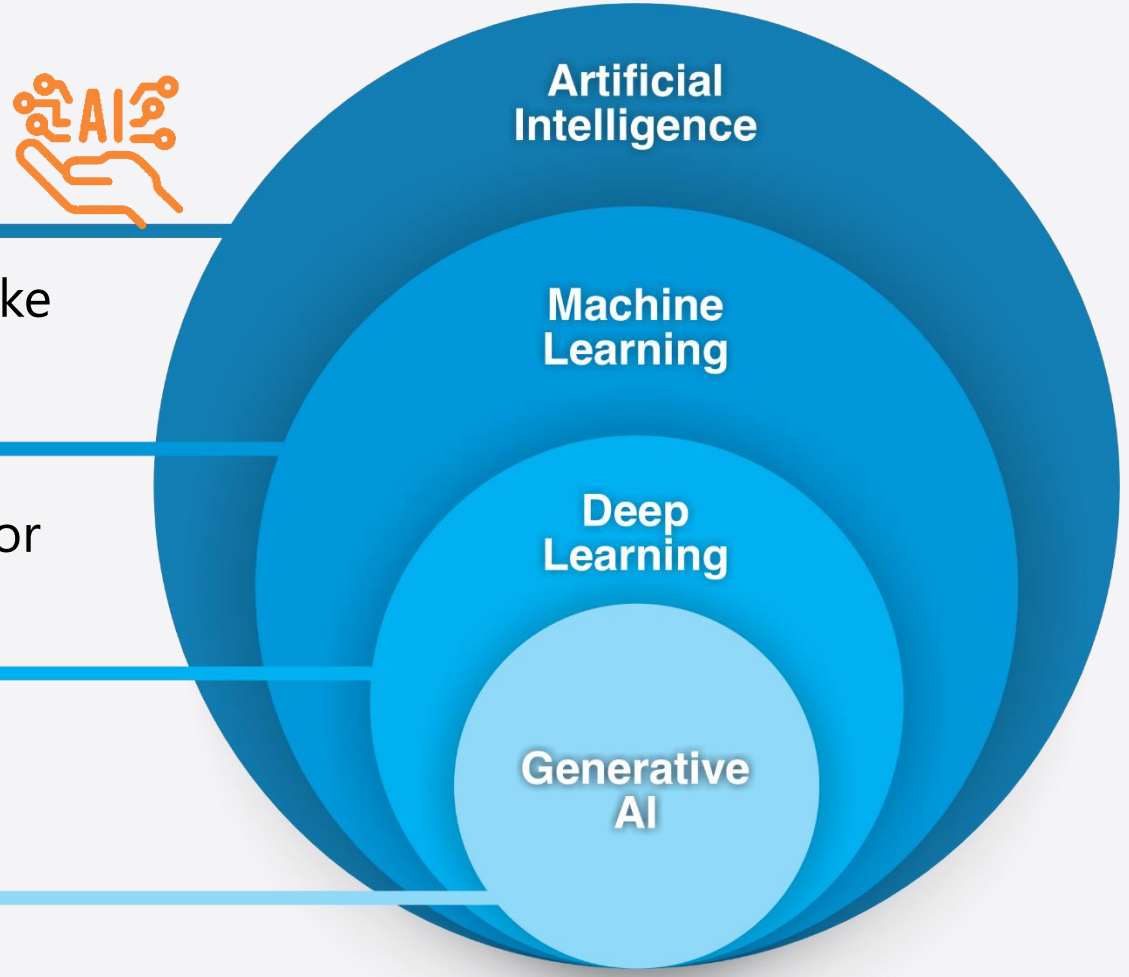
OpenAI makes its AI models available to developers to build powerful software applications, such as ChatGPT.

Imitates human behavior

Takes data and fits it to an algorithm to make Predictions.

Uses layers of algorithms to return results for more complex use cases

Can produce new content based on what is described in the input



OpenAI Microsoft

Generative AI

GPT-3

Prompt:

```
Write a tagline for an ice cream shop.
```

Response:

```
We serve up smiles with every scoop!
```

Codex

Prompt:

```
Table customers, columns = [CustomerId, FirstName, LastName, Company, Address, City, State, Country, PostalCode]

Create a SQL query for all customers in Texas named Jane
query =
```

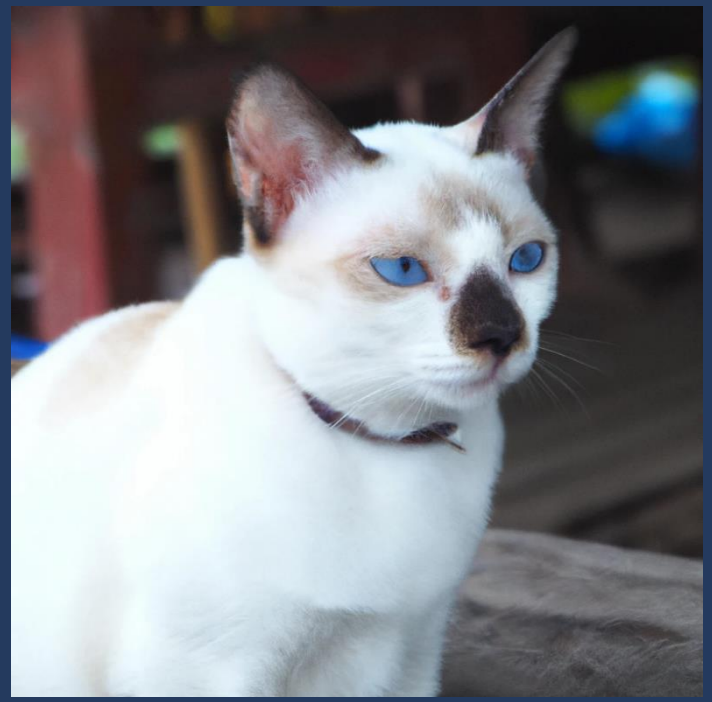
Response:

```
SELECT *
FROM customers
WHERE State = 'TX' AND
FirstName = 'Jane'
```

DALL·E

Prompt: A white Siamese cat

Response:



What is Azure OpenAI?

Microsoft has partnered with OpenAI to deliver on three main goals:

01

To **utilize Azure's infrastructure**, including security, compliance, and regional availability, to help users build enterprise-grade applications.

02

To **deploy OpenAI AI model capabilities across Microsoft products**, including and beyond Azure AI products.

03

To **use Azure to power all of OpenAI's** workloads.

ChatGPT | Bing Chat | Microsoft 365 Copilot | Power Platform Copilot | Dynamics 365 Copilot | Windows Copilot

Microsoft runs on Azure AI

Microsoft Copilots

M365 Copilot

Bing

Better Q&A and task completion

Edge

Better interaction with web content

Word

Better reading and writing assistance

Outlook

Better e-mail management

Excel

Better data analysis

PowerPoint

Better presentations

Teams

Better Meetings

Business Chat

Better knowledge management

Designer

Better digital creations

Windows Copilot

Better interaction with OS, apps, and files

Copilots for Web

Copilots for Productivity

Copilot for Creativity

Copilot for Everyday

Dynamics Copilot

Better sales and customer support

Fabric Copilot

Better data analytics and business intelligence

Security Copilot

Better threat detection, identification, and mitigation

GitHub Copilot

Better code development

Power Platform Copilot

Better creation of apps, workflows, and agents

Copilots for Business

Copilots for Analytics

Copilot for Security

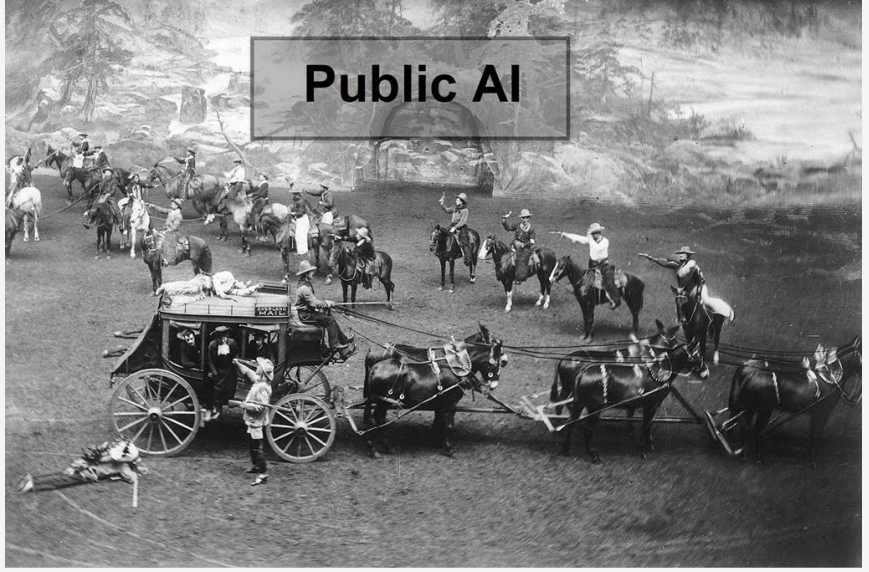
Copilot for Development

Copilot for Low/No Code Development

OpenAI versus Azure OpenAI

Key differences

- **OpenAI** - is a company that develops and provides access to GPT-3 and GPT-4 through their own APIs and web interface.
- **Azure OpenAI Service** - is a cloud service that co-develops and hosts GPT-3 and GPT-4, giving customers the security and enterprise promise of Azure.
  - Data submitted to the Azure OpenAI Service *remains* within Microsoft Azure
  - Is *not passed* to OpenAI (the company) for model predictions.
- Features of **Azure OpenAI Service**:
  - Private networking
  - Regional availability
  - Responsible AI content filtering
  - Fine-tuning
  - Virtual network support
  - Managed identity



Azure OpenAI's access and responsible AI policies

It's important to consider the ethical implications of working with AI systems.

01 **Fairness**: AI systems shouldn't make decisions that discriminate against or support bias of a group or individual.

02 **Reliability and Safety**: AI systems should respond safely to new situations and potential manipulation

03 **Privacy and Security**: AI systems should be secure and respect data privacy.

04 **Inclusiveness**: AI systems should empower everyone and engage people.

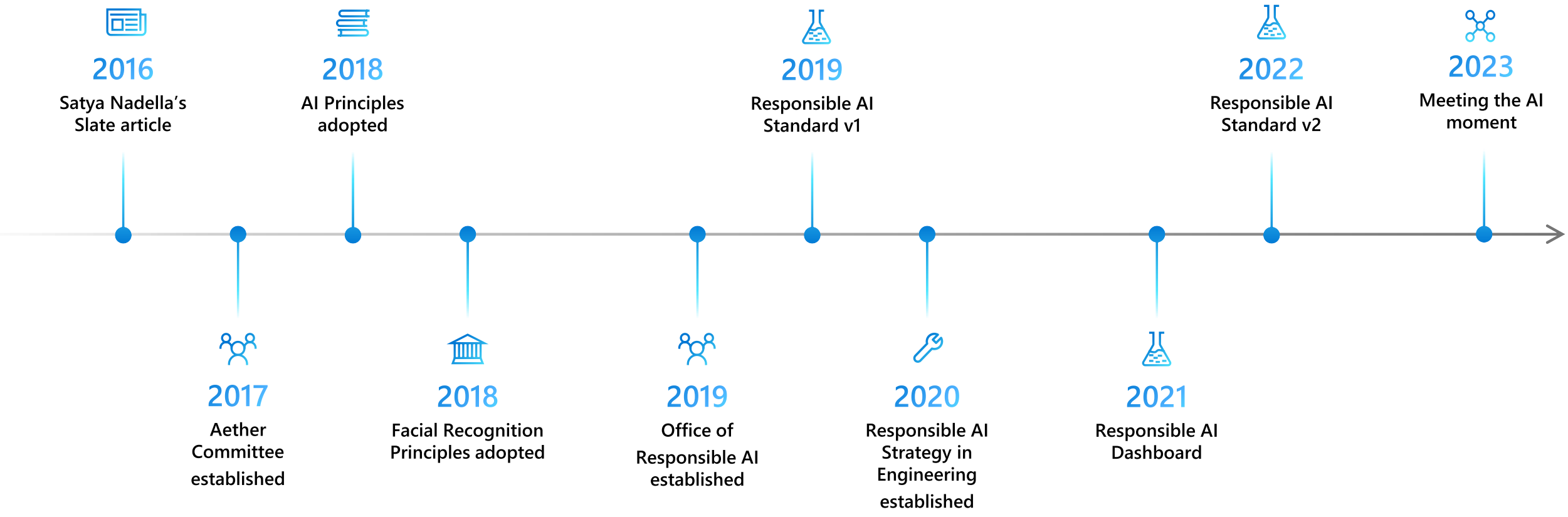
05 **Accountability**: People must be accountable for how AI systems operate.

06 **Transparency**: AI systems should have explanations so users can understand how they're built and used.

Note: As part of Microsoft's commitment to using AI responsibly, access to Azure OpenAI is currently limited.

aka.ms/ResponsibleAIResources

Our Responsible AI journey



Azure AI Studio > DALL·E playground (Preview)

Privacy & cookies

DALL·E playground (Preview)

Playground

View code Settings

DALL·E text to image

Your task failed as a result of our safety system.

Tile Size

Search Medium tiles

Prompt

a picture of AI blowing up the world and taking over humanity

Generate

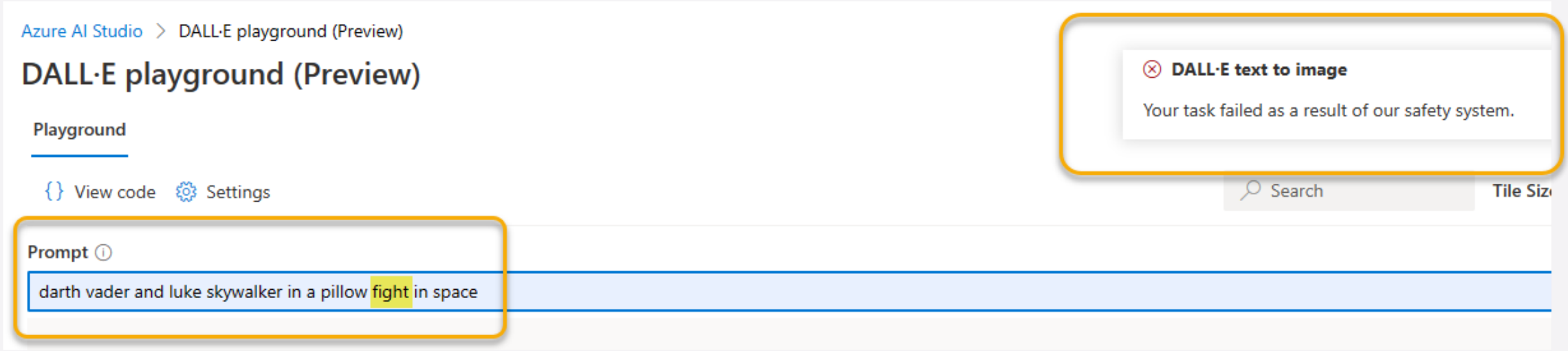
Example of responsible AI

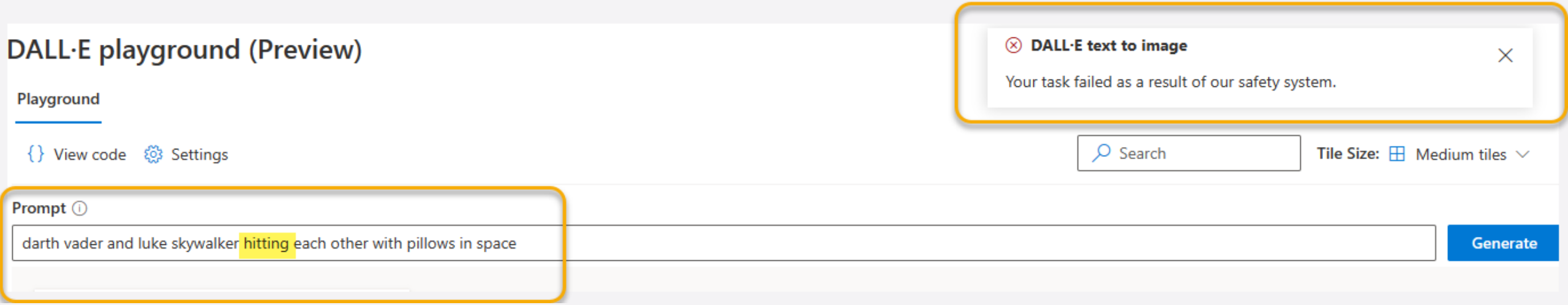
Prompt

an mushroom cloud and people running frantically and a cyborg standing with arms crossed in the foreground

an mushroom cloud and people running frantically and a cyborg standing with arms crossed in the foreground

AI Teaching Humans How to Think?





Access to Azure OpenAI

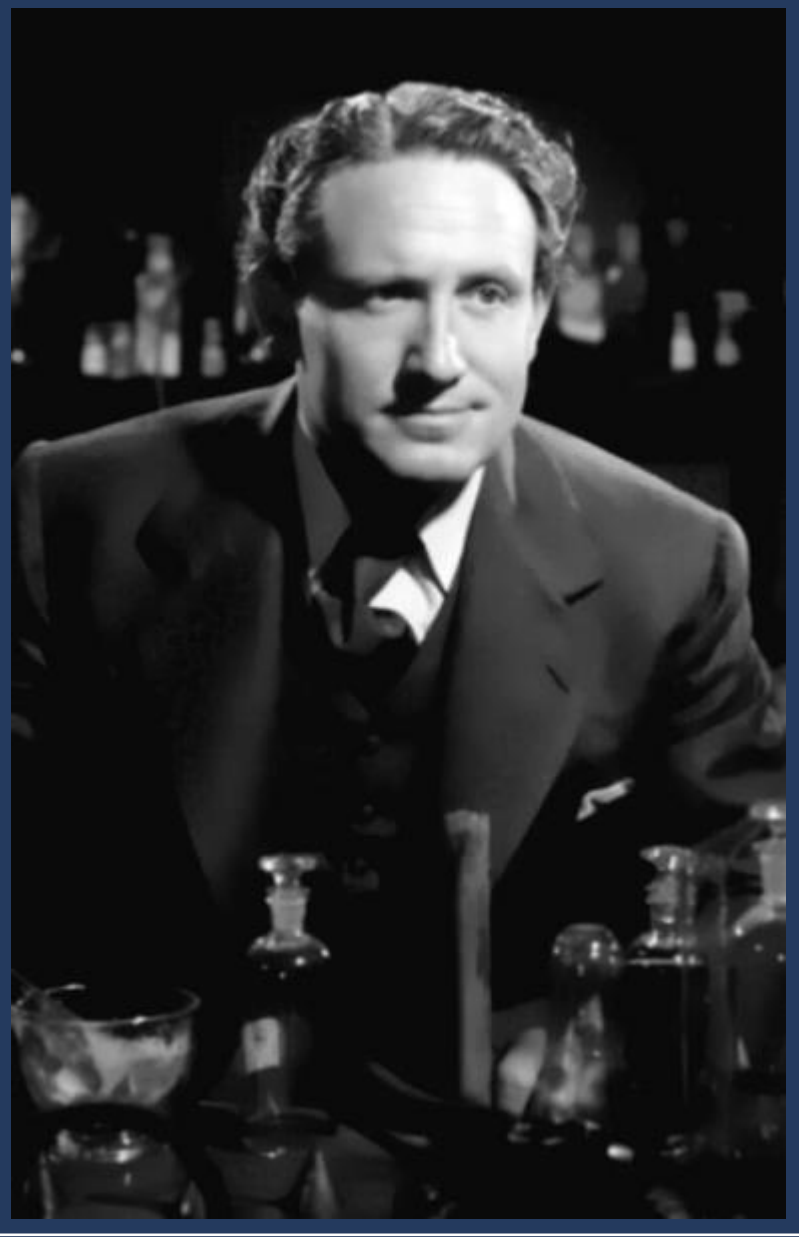
You need to apply for access to Azure OpenAI.

https://aka.ms/oaiapply

01

Using AI to Deliver SOC Intelligence and Efficiency

Dr Sentinel



Copilot

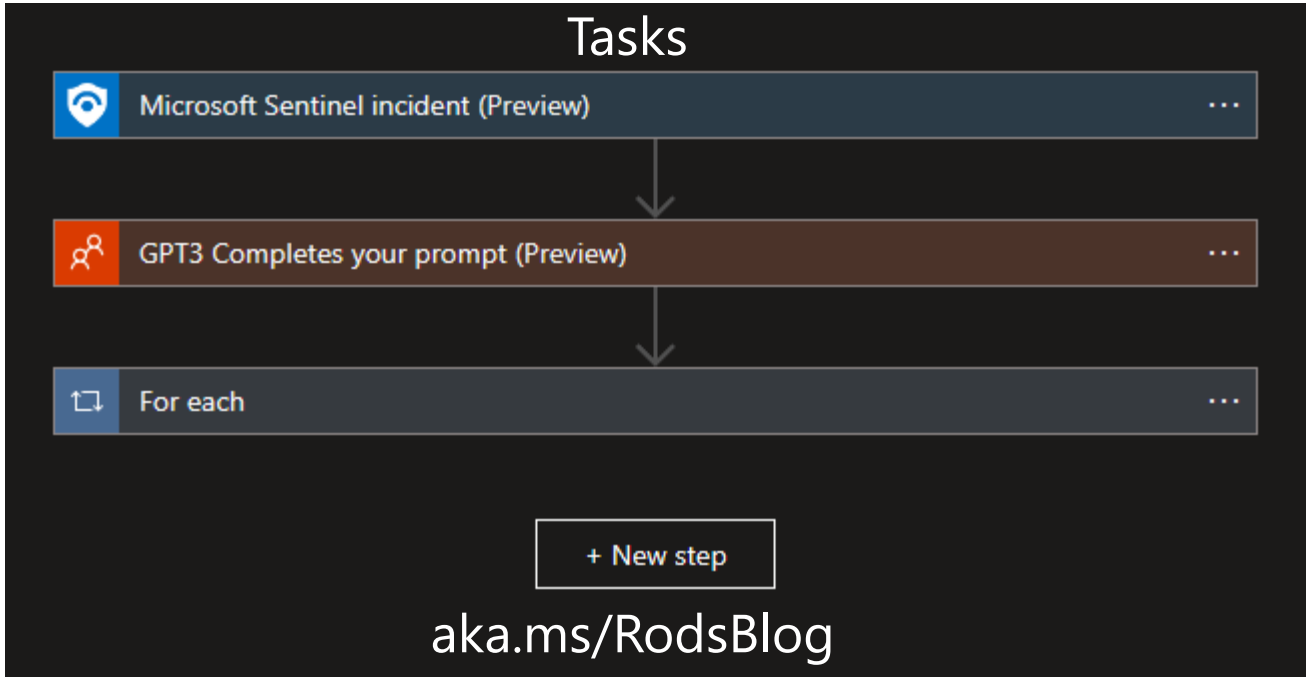
An experience using generative AI to assist humans with complex cognitive tasks

Top Efficiency Drivers for Security Teams

- Connecting data is brutally difficult
- Preparedness and Hunting is time consuming
- Continual Tuning and Evolution of SOC is nearly non-existent
- Adding Investigative Context Requires Rare Skillsets
- Separation of duties/expertise/Skills gap
- Learning KQL
- Locating and Maintaining Trusted Threat Intelligence Sources

Recommendations

6MDMChatGPT



KQL

aka.ms/MustLearnKQL

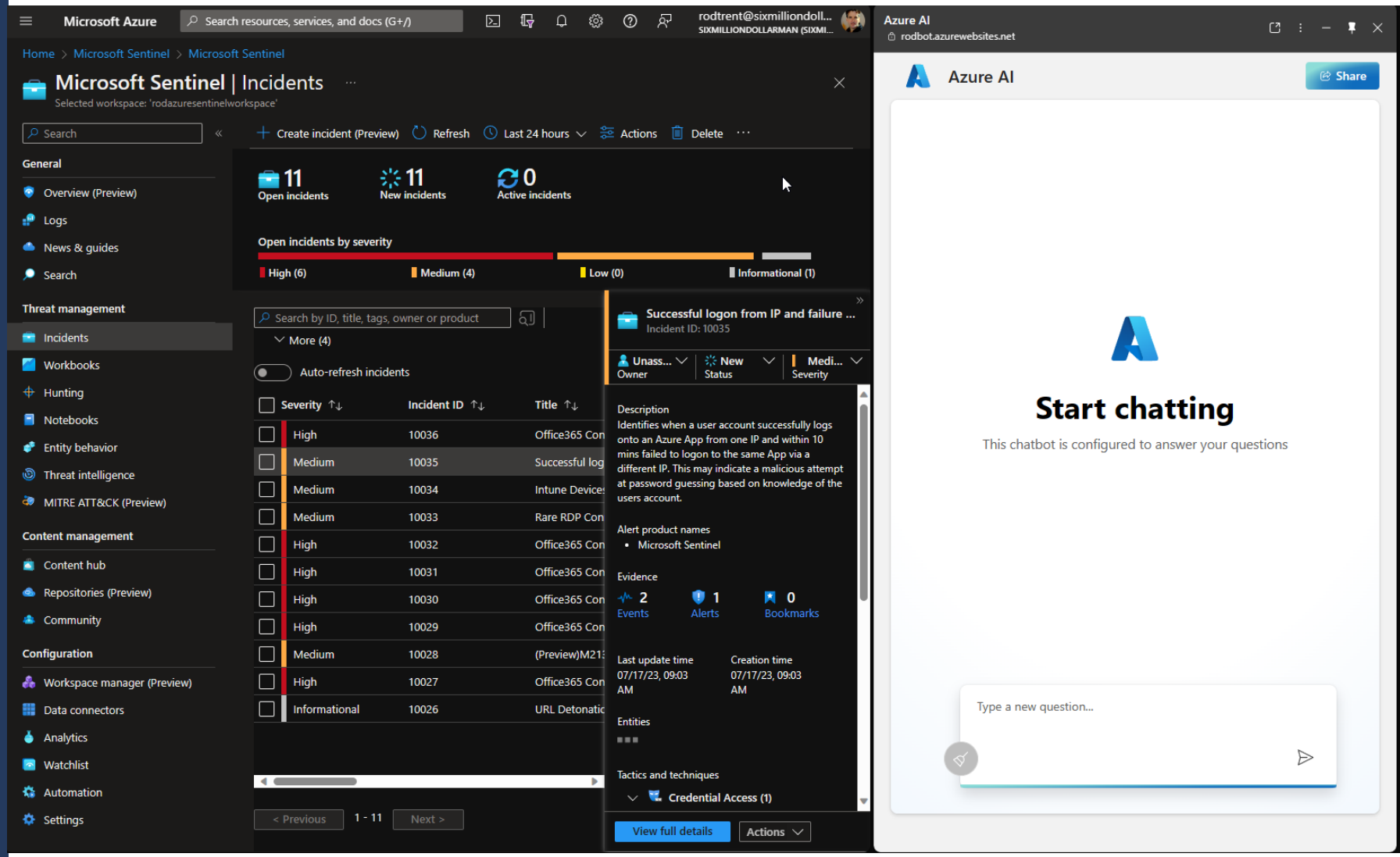
AzureOpenAI-Enrichment

aka.ms/RodsBlog

Assistant

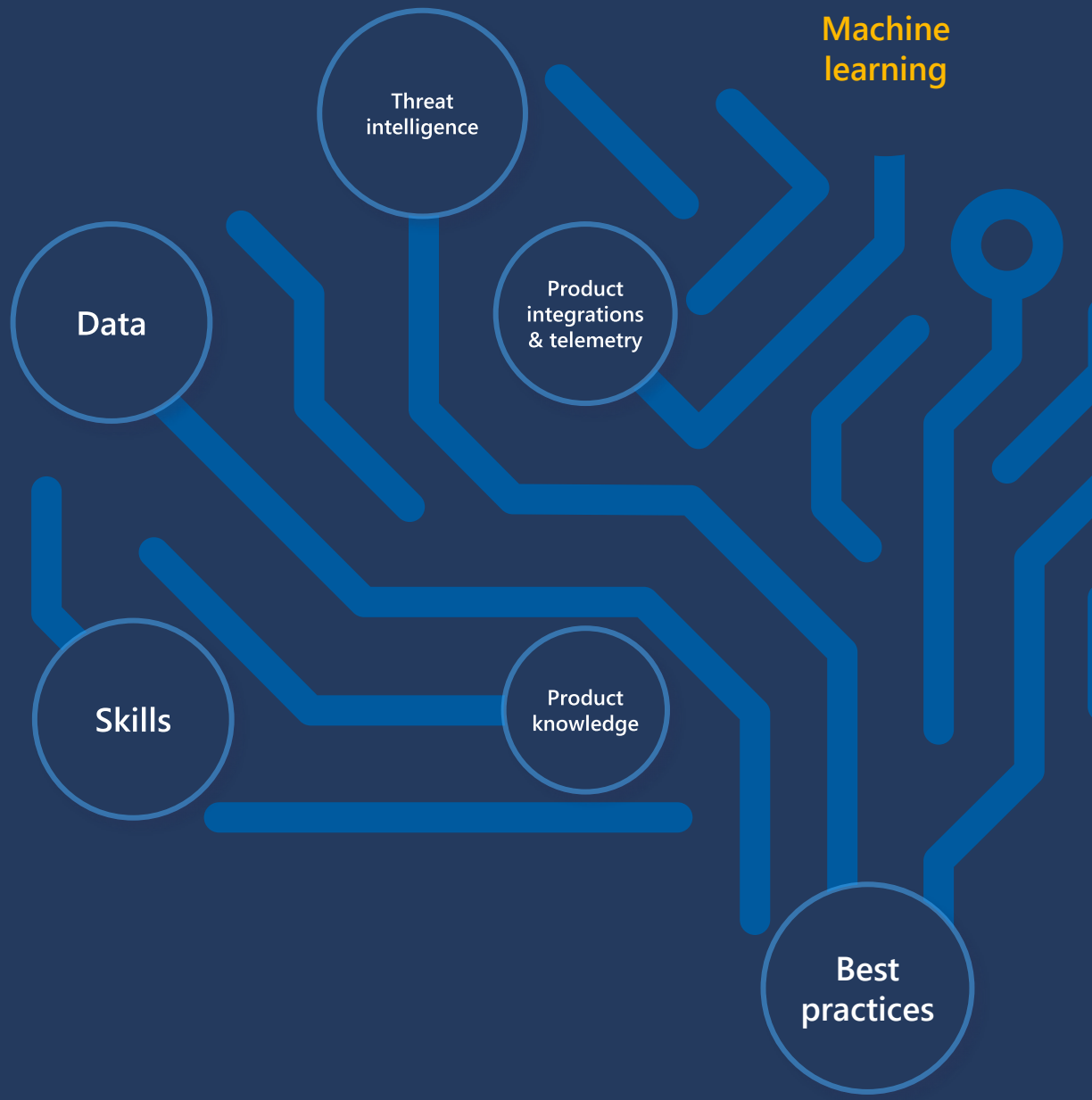
(Copilot)

aka.ms/RodsBlog



Microsoft Security Copilot

The first generative AI security product that empowers SOC analysts to defend their organizations at machine speed and scale



Microsoft Security Copilot

The first and only generative AI security product to help defend organizations at machine speed and scale.

Enables response in minutes, not hours

Simplifies the complex with natural language prompts and easy reporting

Catches what others miss with deeper understanding of events

Addresses talent shortage by extending human expertise

Tell me about my latest incidents.

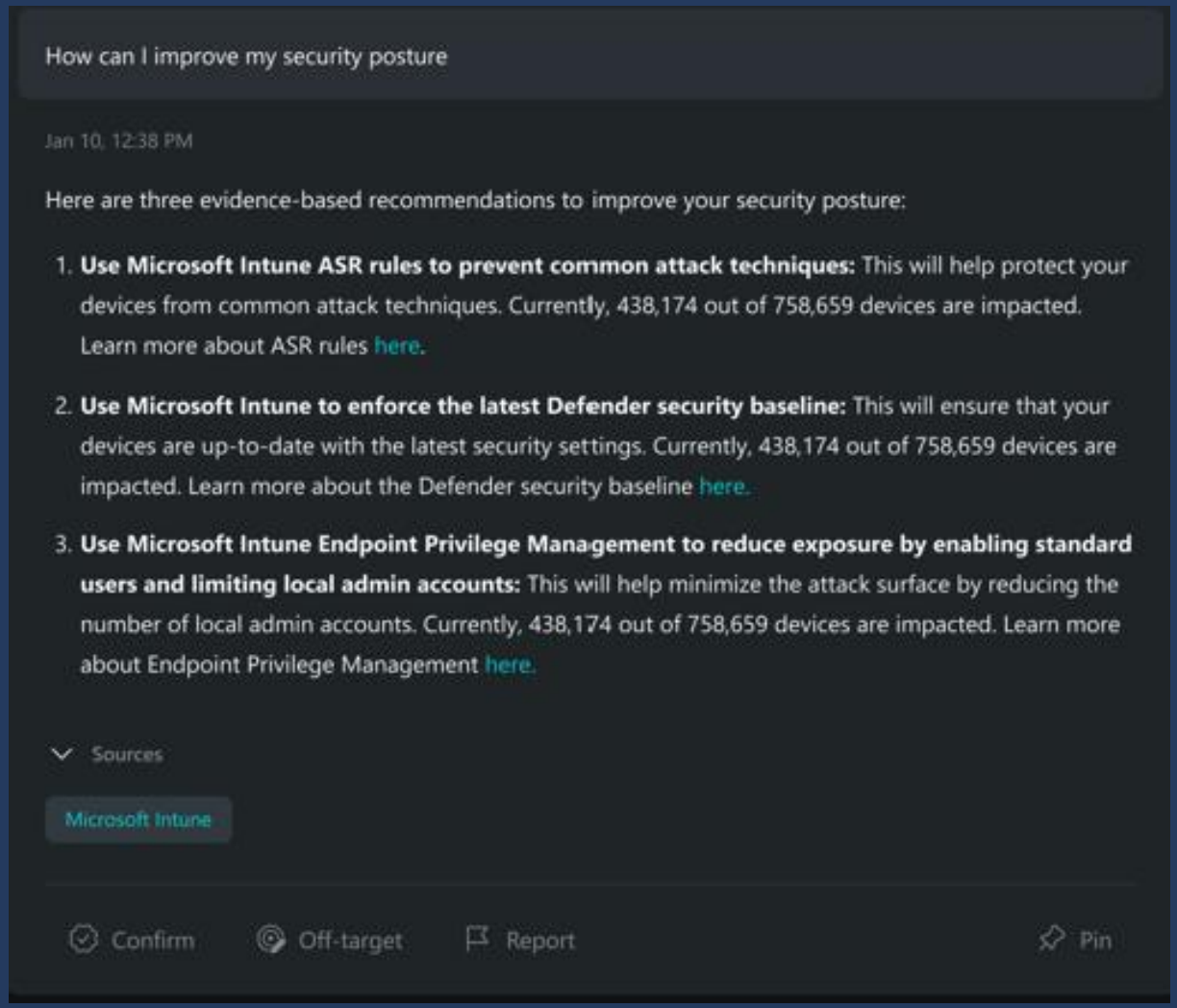
Are any of my machines affected?

Summarize this incident in PowerPoint.

Which alerts are being triggered the most?

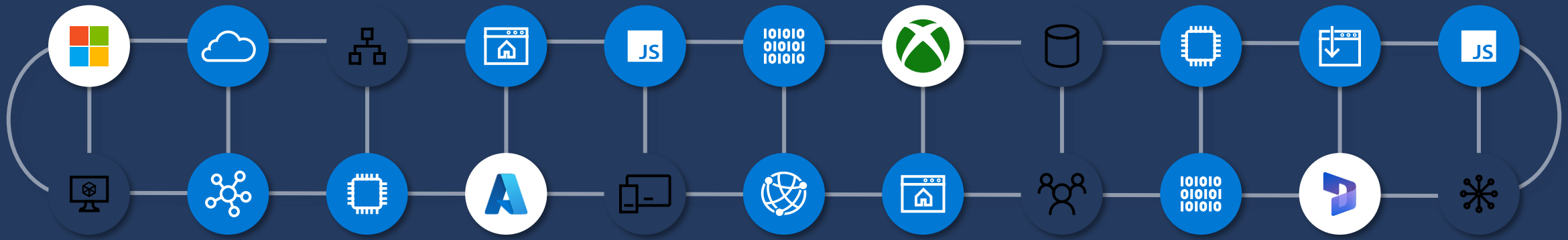
What is log4shell?

How can I improve my security posture?



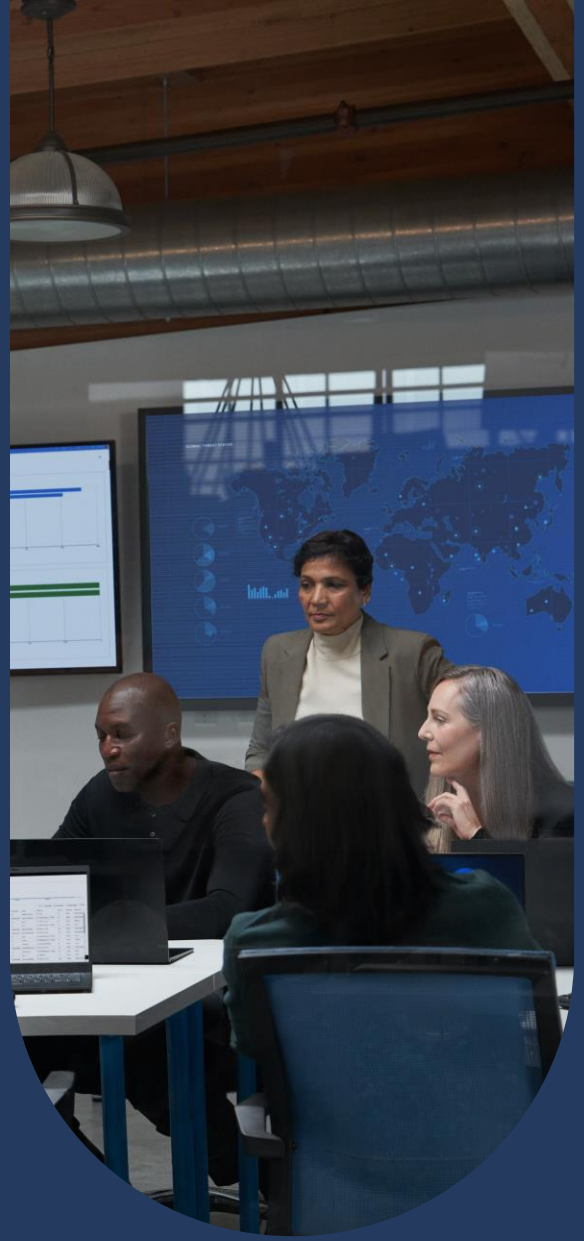
Microsoft Threat Intelligence

The industry's largest vector coverage powered by 65T daily signals



One of the world's largest clouds + **Signal from 1.4B endpoints[1] across the planet** + **Graphing global internet infrastructure**

1. "Microsoft by the Numbers". Microsoft Story Labs



Security Copilot boosting your SOC team

Security posture management

Discover whether your organization is susceptible to known vulnerabilities and exploits. Prioritize risks and address vulnerabilities with guided recommendations.

Incident response

Surface an ongoing incident, assess its scale, and get instructions to begin remediation based on proven tactics from real-world security incidents

Security reporting

Summarize any event, incident, or threat in seconds and prepare the information in a ready-to-share, customizable report for your desired audience

Built with security, privacy, and compliance.

Your data is **your** data

Your data is **not** used to train the foundation AI models

Your data is protected by the **most comprehensive** enterprise compliance and security controls

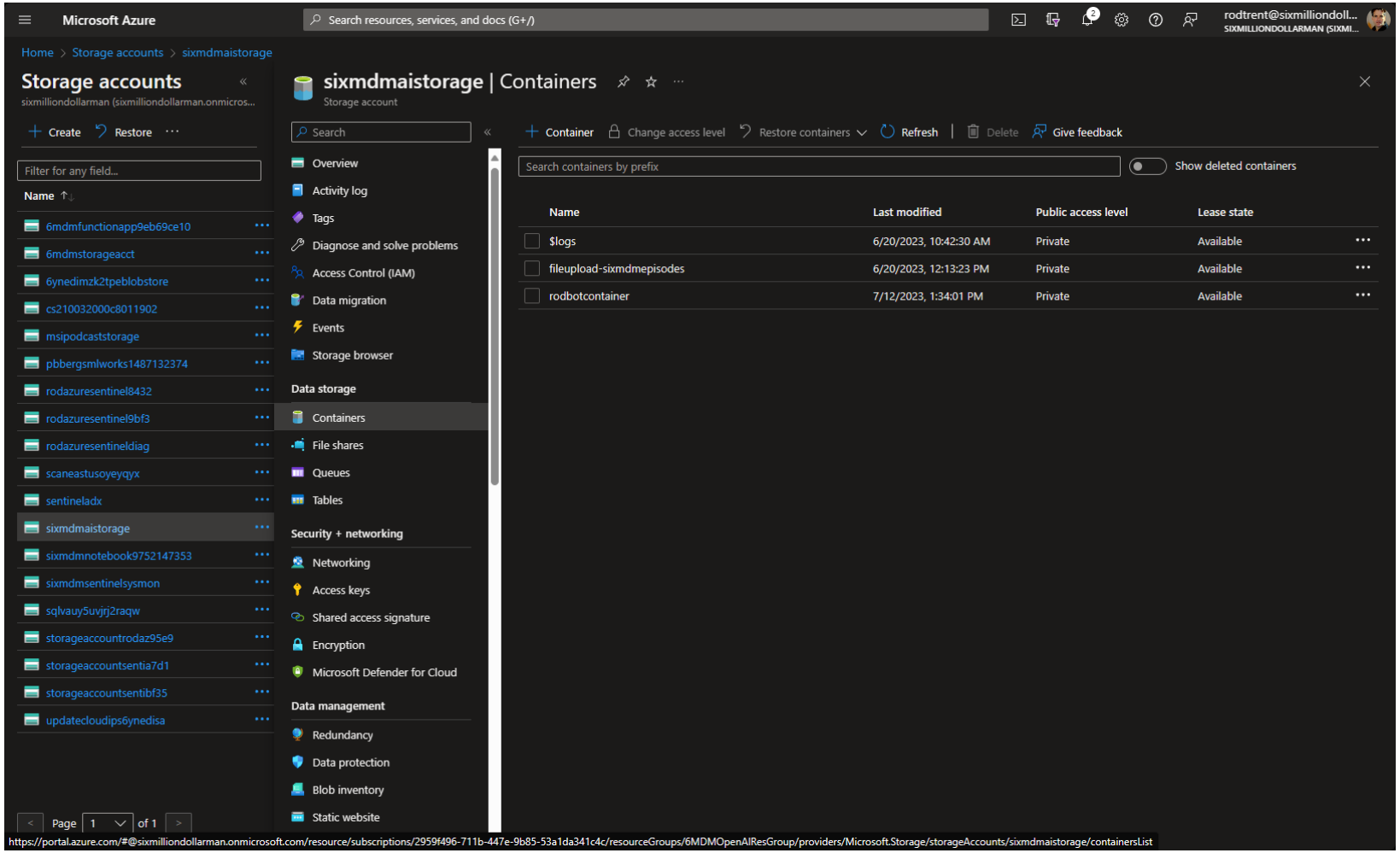
Demo of Security Copilot ????

https://aka.ms/SecurityCopilot

Build Your Own

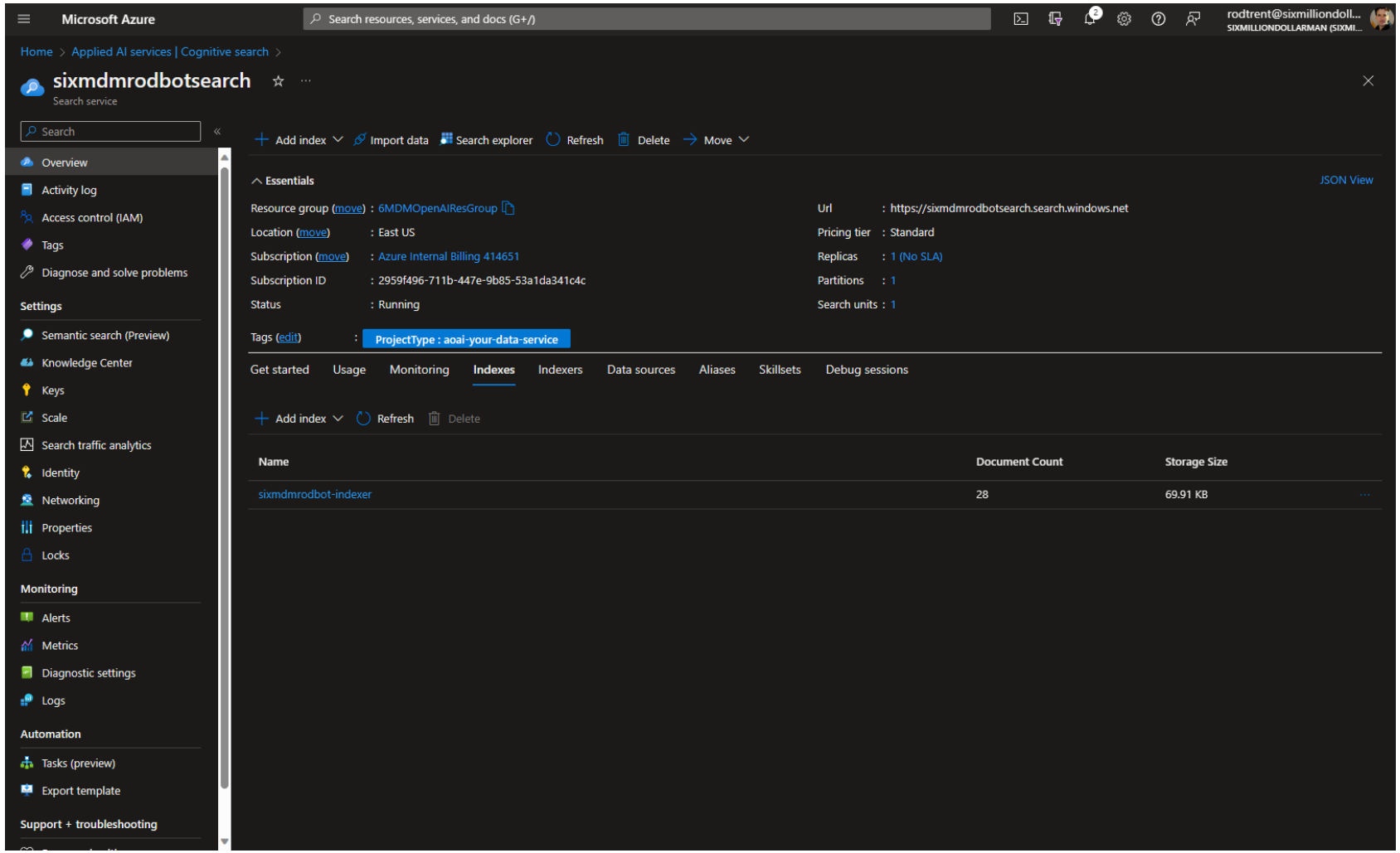
Step 1

Storage Container (Blob)



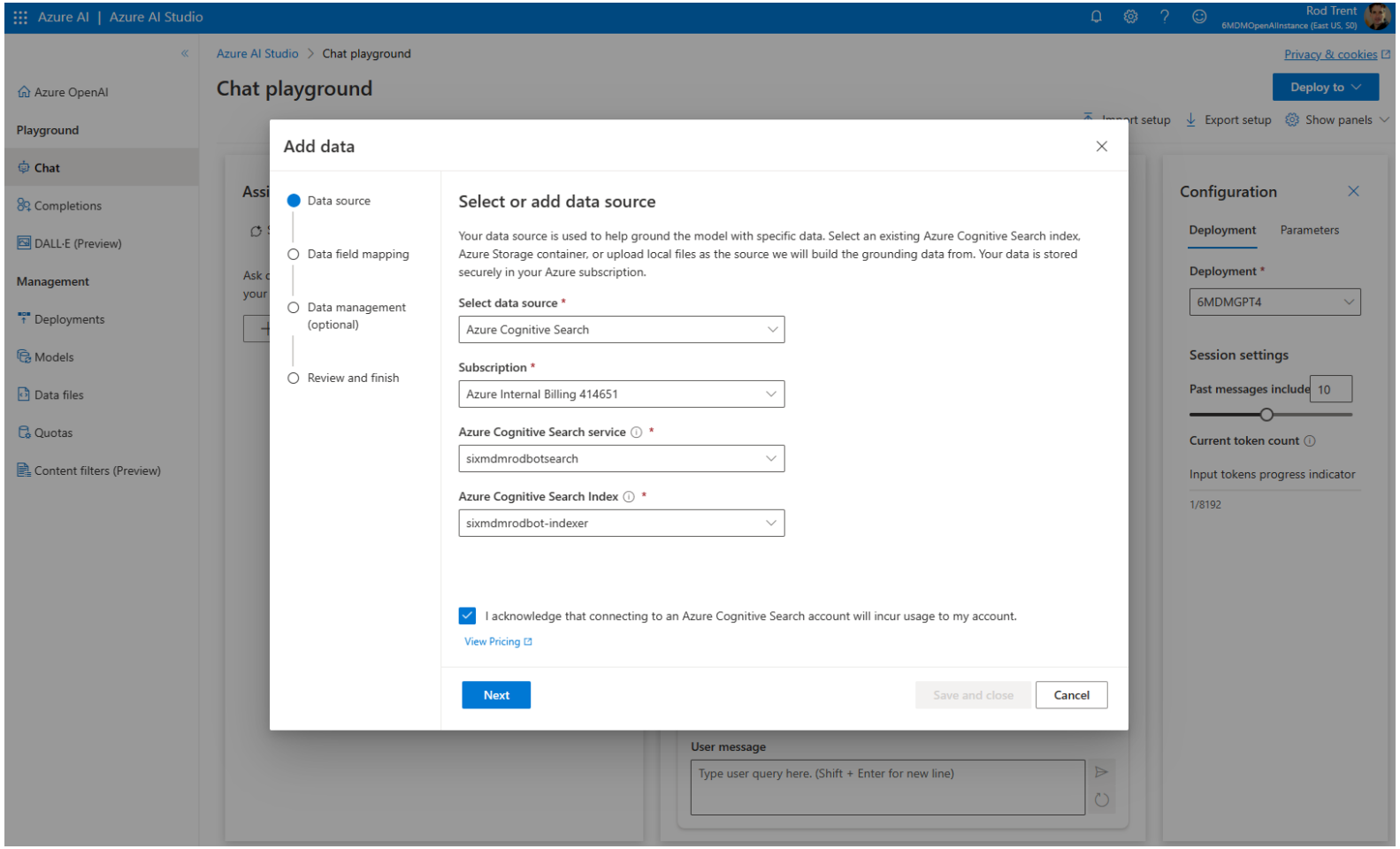
Step 2

Indexer (Cognitive Search)



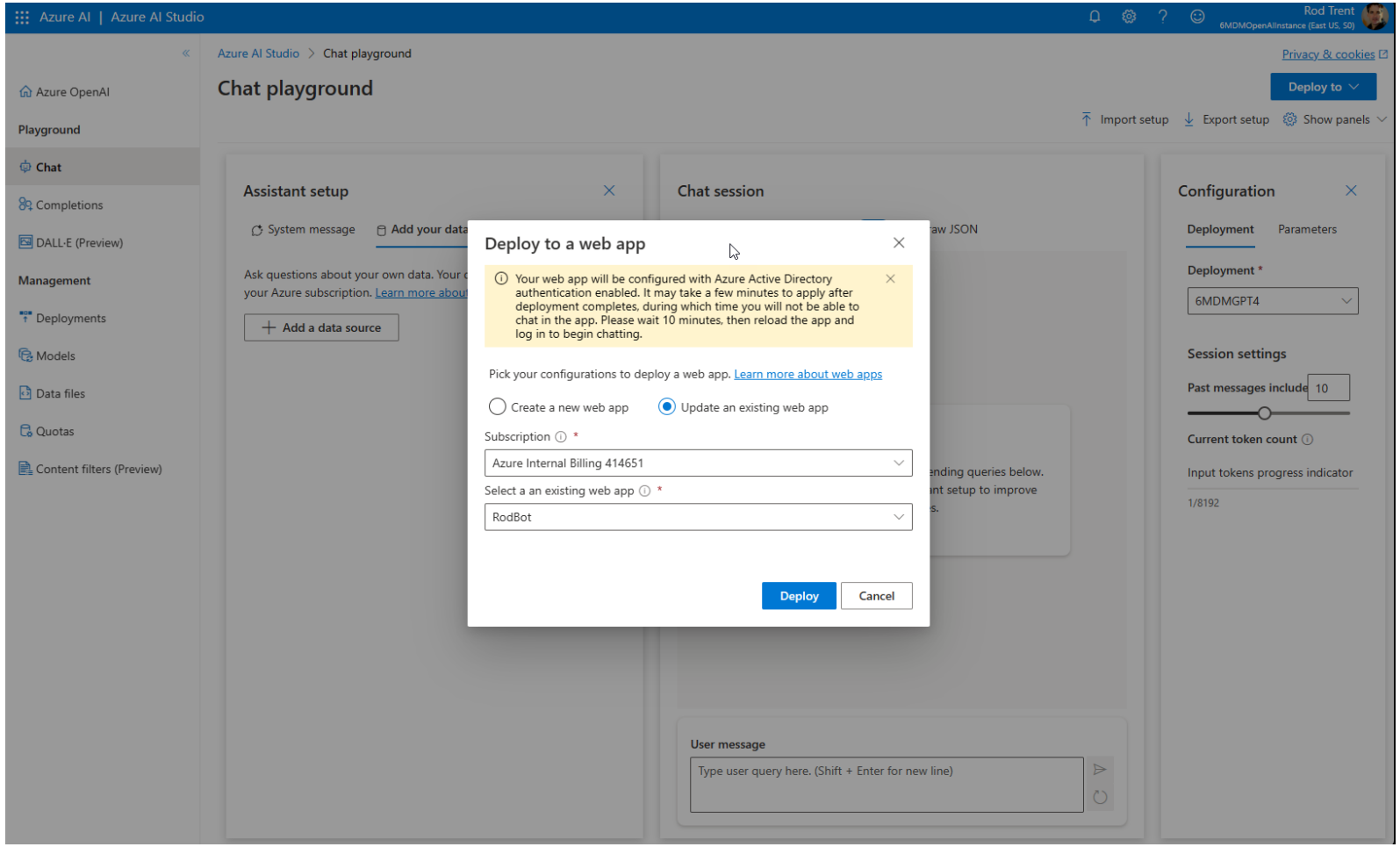
Step 3

Add Your Data (Indexer)



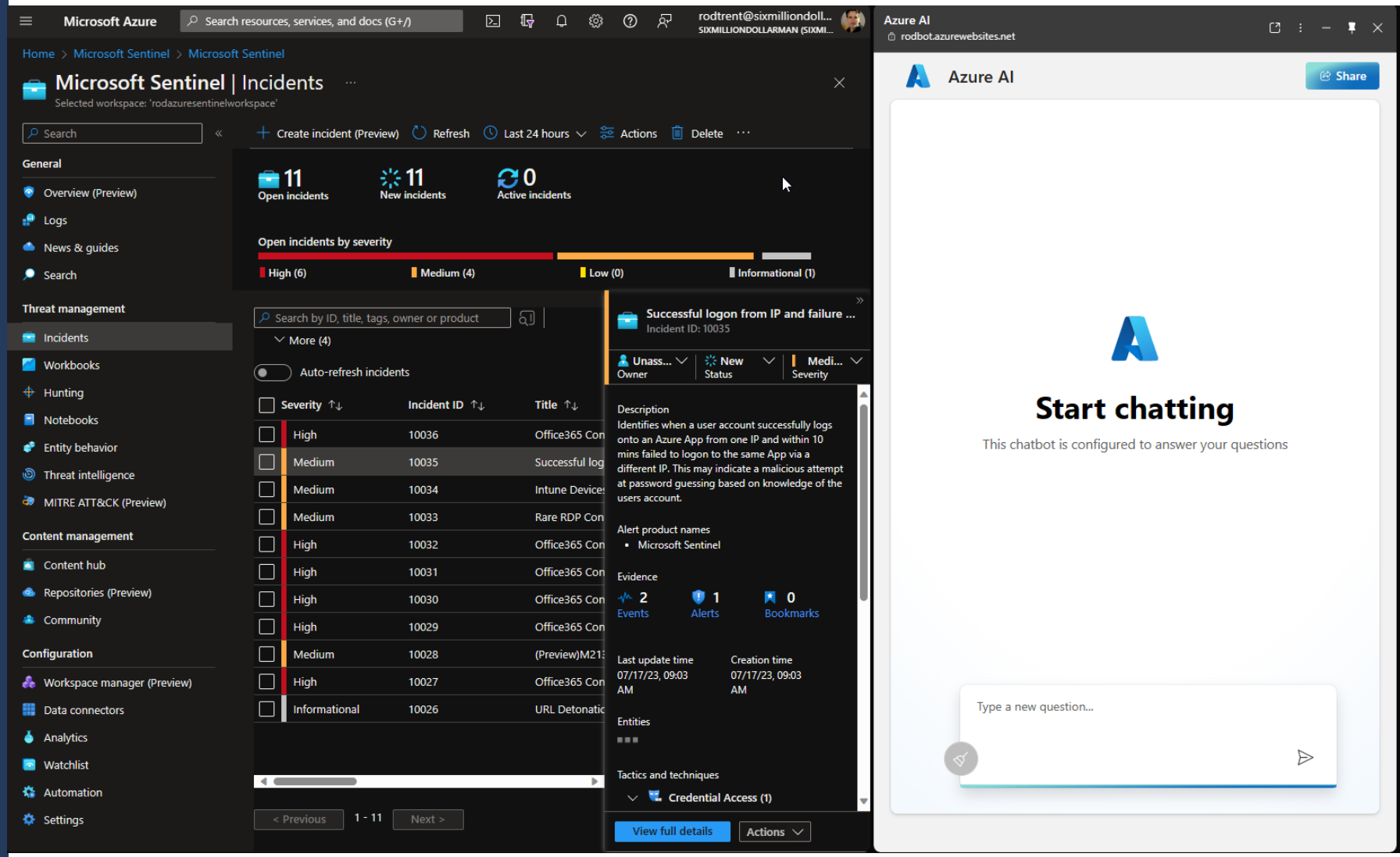
Step 4

Deploy to App (New or Existing)



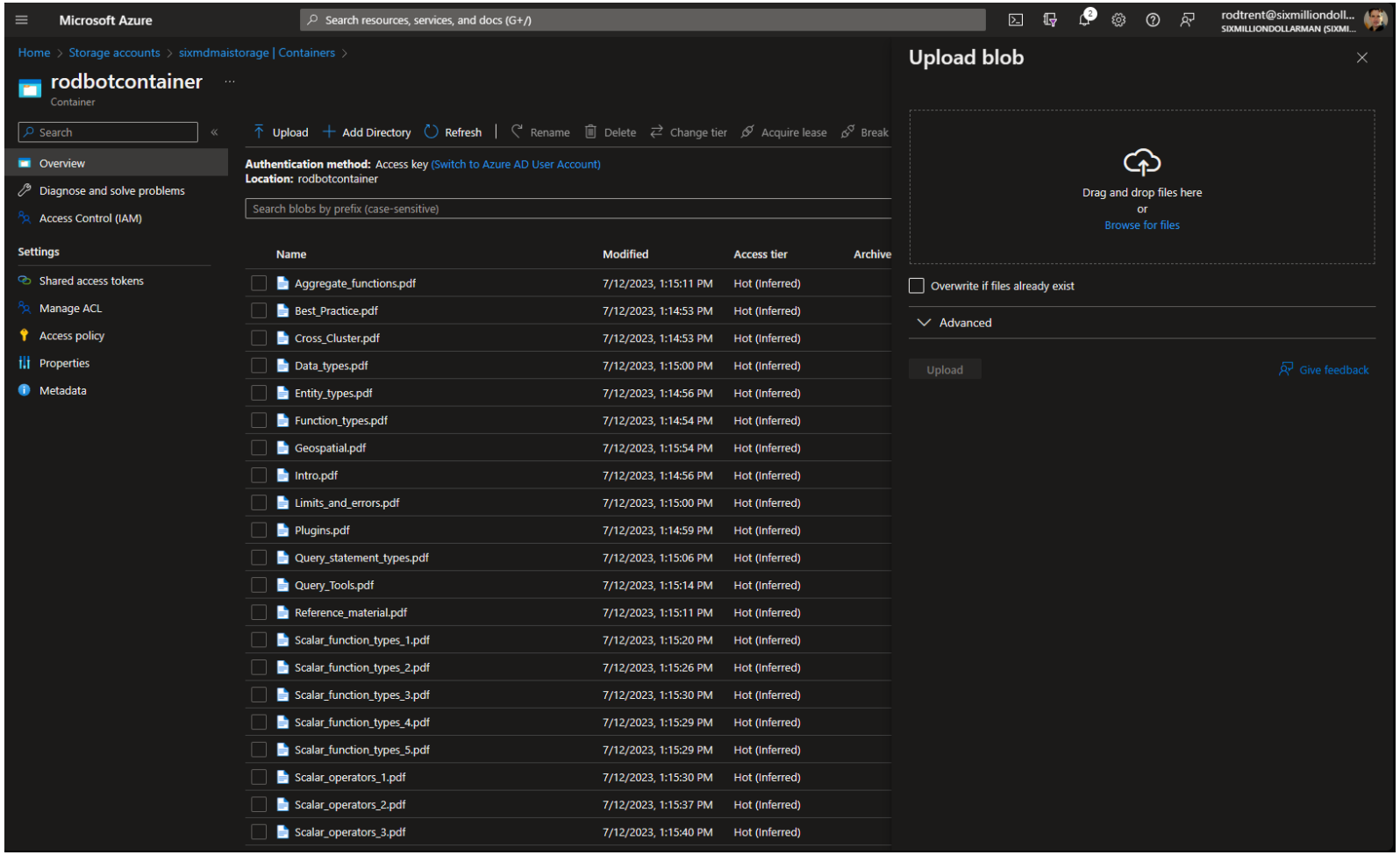
Step 5

Add to Edge Sidebar



Maintaining Source (Improvements)

Upload to Container



Monitoring and Securing AI Using Modern Tools

Mr AI

Threats across AI system layers

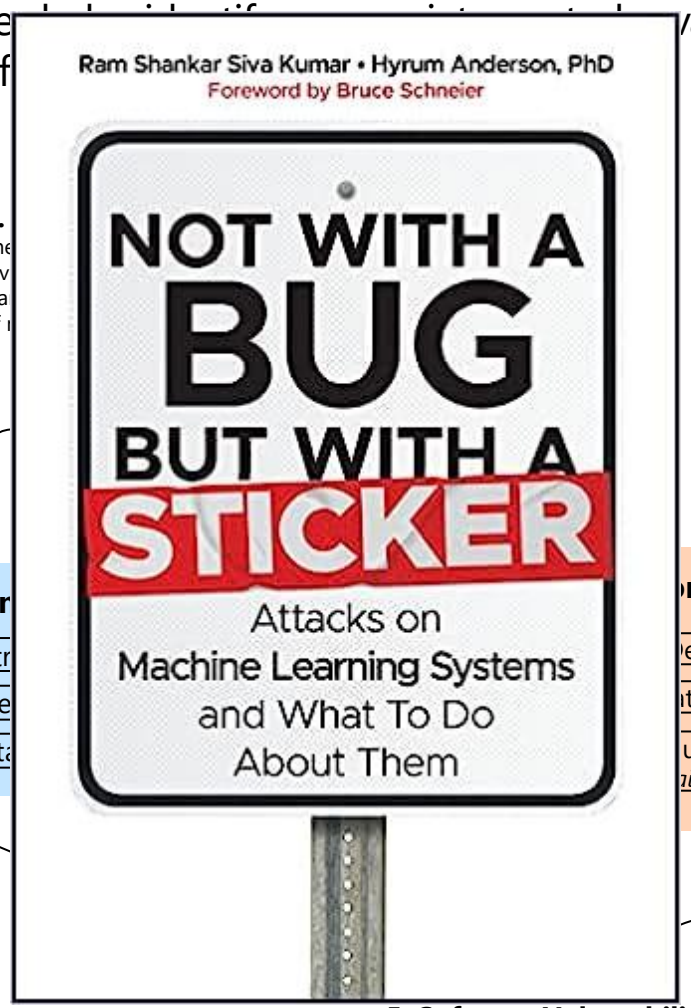
Assigning AI Risks to these three categorie vailable today, and what we may need to develop in the f

4. Inversion
Manipulate machine learning models by extracting information about the model's input data.

1. AI model bias
Refers to the phenomenon where an artificial intelligence model produces results that are systematically inaccurate or unfair towards a particular group of people.

2.
The
Inv
lea
of

AI Platforr
Underlying infrastr
AI / ML mode
Training Data



on
esign
tions
urces
ining)

10. Jailbreak
AI jailbreak refers to an attack in which an attacker exploits a vulnerability in an AI system to gain unauthorized access or control over the system.

8. Adversarial example attack
Aims to deceive AI models by inputting maliciously crafted data that is slightly different from regular data. The input is designed to trick the AI model into making a wrong prediction or classification.

AI Process
User Inputs
AI Outputs

6. Membership Inference
Adversary tries to determine whether a specific data record was used to train an AI model.

7. Model Stealing
Adversary tries to replicate or obtain the model of another machine learning system without permission.

9. Trojan attack
Malicious actor adds a backdoor or malicious code into an AI model during its development stage, which then gets incorporated into the deployed system, and can be triggered later to cause harm or disruption.

5. Software Vulnerabilities
As with any computer system, interference in the infrastructure, operating systems, and applications that support the AI system will lead to software vulnerability exploitation

ATLAS™

The ATLAS Matrix below shows the progression of tactics used in attacks as columns from left to right, with ML techniques belonging to each tactic below.

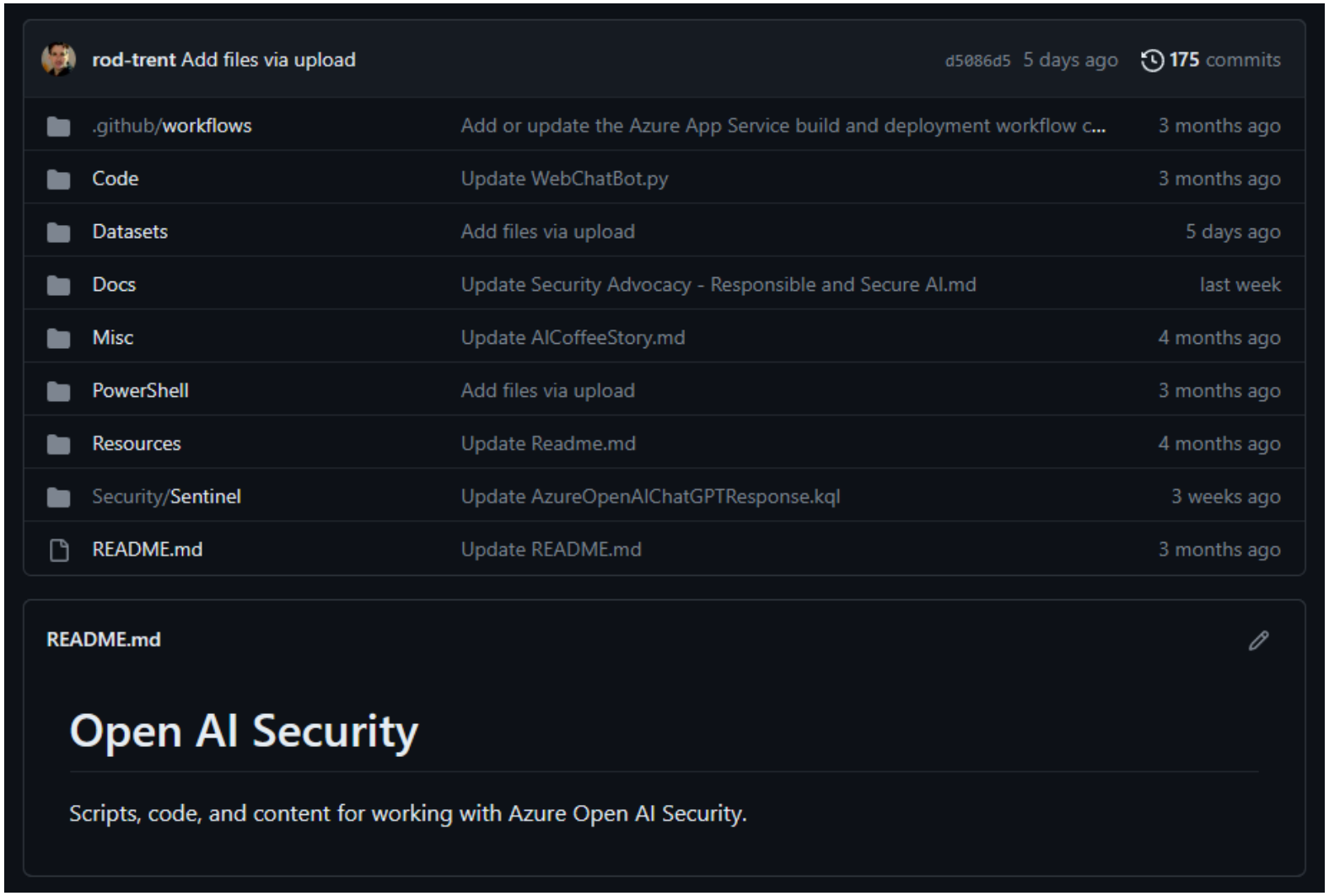
& indicates an adaptation from ATT&CK. Click on links to learn more about each item, or view ATLAS tactics and techniques using the links at the top navigation bar.

Reconnaissance &	Resource Development &	Initial Access &	ML Model Access	Execution &	Persistence &	Defense Evasion &	Discovery &	Collection &	ML Attack Staging	Exfiltration &	Impact &
5 techniques	7 techniques	4 techniques	4 techniques	2 techniques	2 techniques	1 technique	3 techniques	3 techniques	4 techniques	2 techniques	7 techniques
Search for Victim's Publicly Available Research Materials	Acquire Public ML Artifacts	ML Supply Chain Compromise	ML Model Inference API Access	User Execution &	Poison Training Data	Evade ML Model	Discover ML Model Ontology	ML Artifact Collection	Create Proxy ML Model	Exfiltration via ML Inference API	Evade ML Model
Search for Publicly Available Adversarial Vulnerability Analysis	Obtain Capabilities &	Valid Accounts &	ML-Enabled Product or Service	Command and Scripting Interpreter &	Backdoor ML Model		Discover ML Model Family	Data from Information Repositories &	Backdoor ML Model	Exfiltration via Cyber Means	Denial of ML Service
Search Victim-Owned Websites	Develop Adversarial ML Attack Capabilities	Evade ML Model	Physical Environment Access				Discover ML Artifacts	Data from Local System &	Verify Attack		Spamming ML System with Chaff Data
Search Application Repositories	Acquire Infrastructure	Exploit Public-Facing Application &	Full ML Model Access						Craft Adversarial Data		Erode ML Model Integrity
Active Scanning &	Publish Poisoned Datasets										Cost Harvesting
	Poison Training Data										ML Intellectual Property Theft
	Establish Accounts &										System Misuse for External Effect

https://atlas.mitre.org/

What to Monitor?

aka.ms/RodAI



Your prompts (inputs) and completions (outputs), your embeddings, and your training data:

- are **NOT** available to other customers.
- are **NOT** available to OpenAI.
- are **NOT** used to improve OpenAI models.
- are **NOT** used to improve any Microsoft or 3rd party products or services.
- are **NOT** used for automatically improving Azure OpenAI models for your use in your resource (The models are stateless, unless you explicitly fine-tune models with your training data).
- Your fine-tuned Azure OpenAI models are available exclusively for your use.

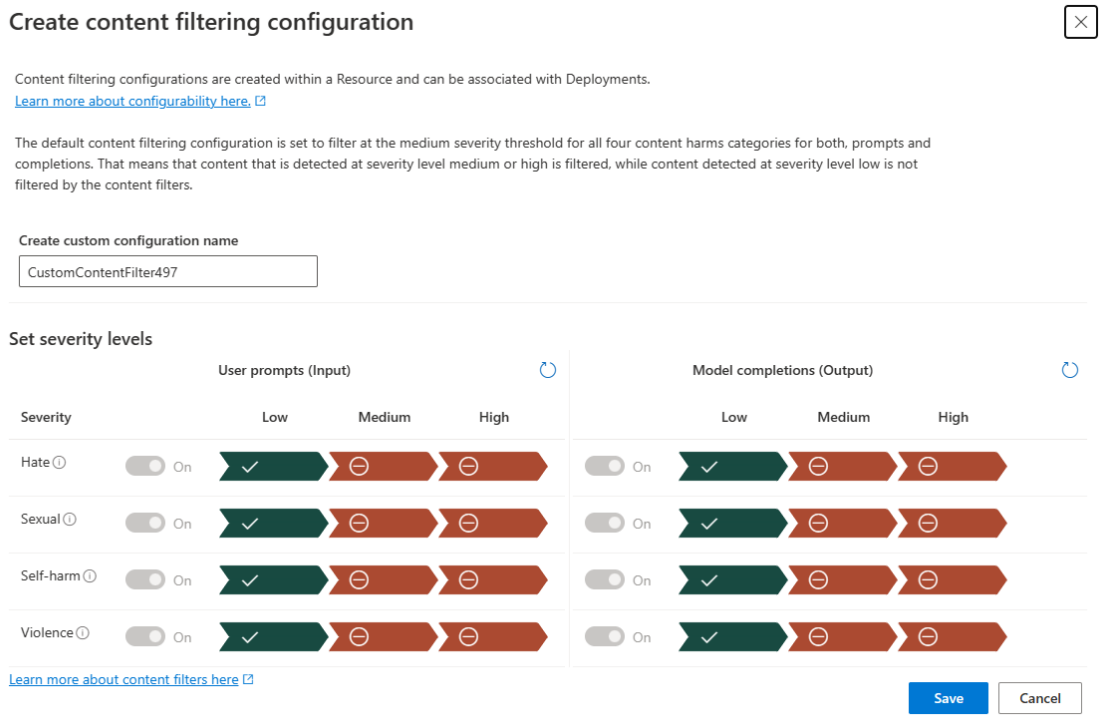
The Azure OpenAI Service is fully controlled by Microsoft; Microsoft hosts the OpenAI models in Microsoft's Azure environment and the Service does NOT interact with any services operated by OpenAI (e.g. ChatGPT, or the OpenAI API).

Compliance

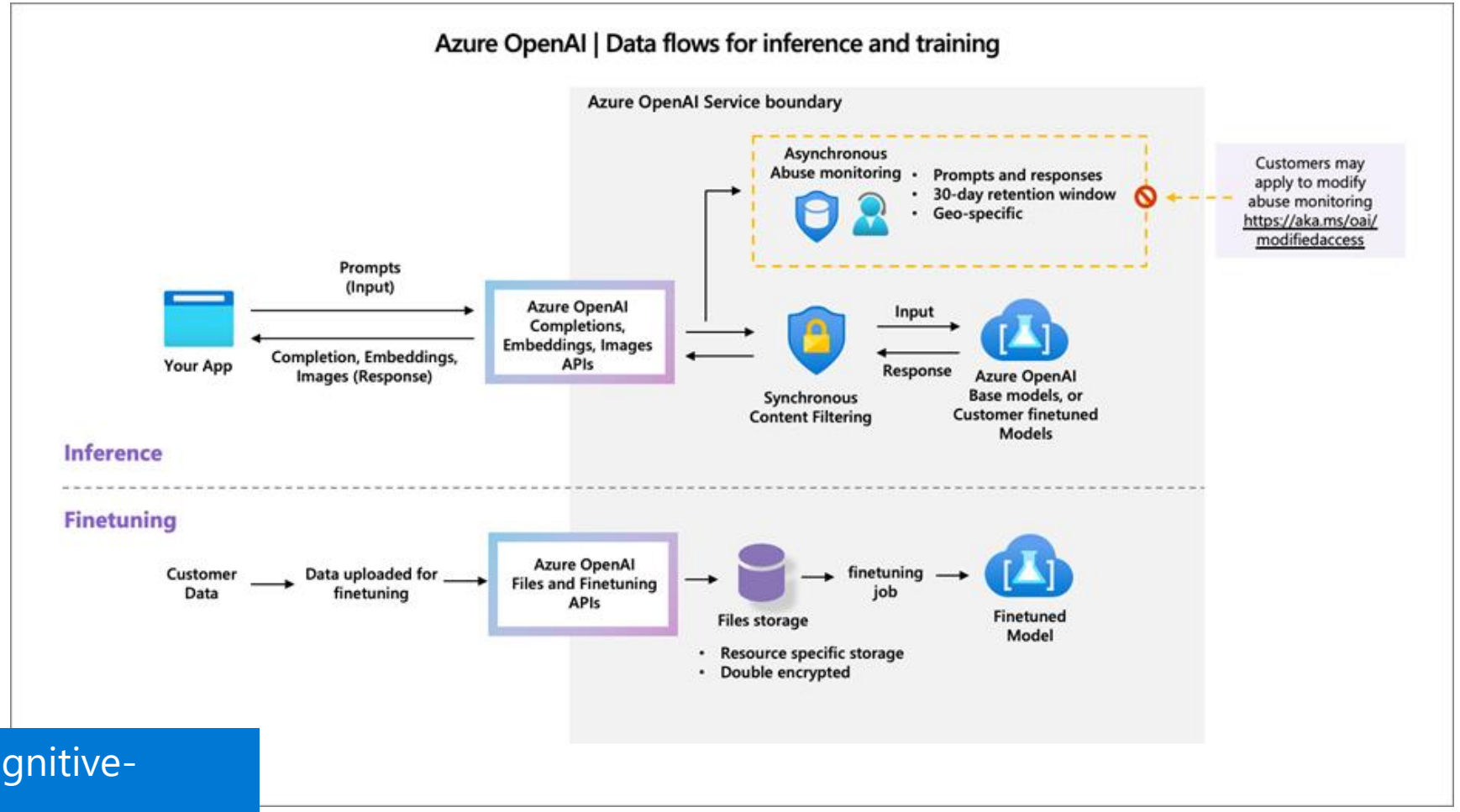
Abuse Monitoring/Content Filtering

https://learn.microsoft.com/legal/cognitive-services/openai/data-privacy

Azure OpenAI Service includes a content management system that works alongside core models to filter content.



“Safety” Data (captured and stored) for 30 days



https://learn.microsoft.com/legal/cognitive-services/openai/data-privacy

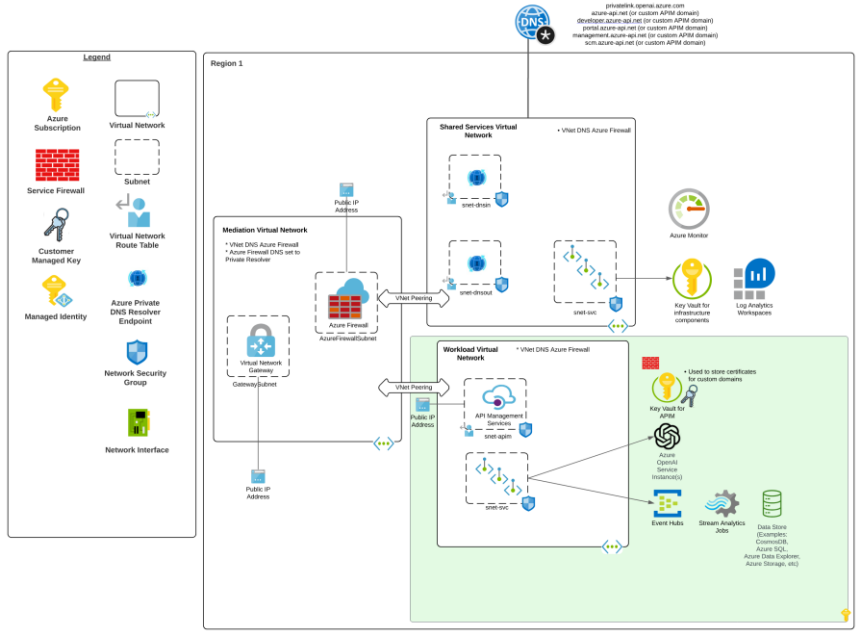
Advanced Logging

https://github.com/Azure-Samples/openai-python-enterprise-logging/tree/main/advanced-logging

Enterprise Azure OpenAI Advanced Logging

Enterprise logging of OpenAI usage metrics:

- Prompt Input
- Prompt Response
- Token Usage
- Model Usage
- Application Usage
- Model Response Times



Generative AI Monitoring

AzureDiagnostics Table

- API access
- Requests
- Responses
- Successes/Failures

AzureActivity Table

- Administrative activity
  - Mods (create, delete, edit, actions, publishing)

Azure Resource Graph Explorer

- Instances

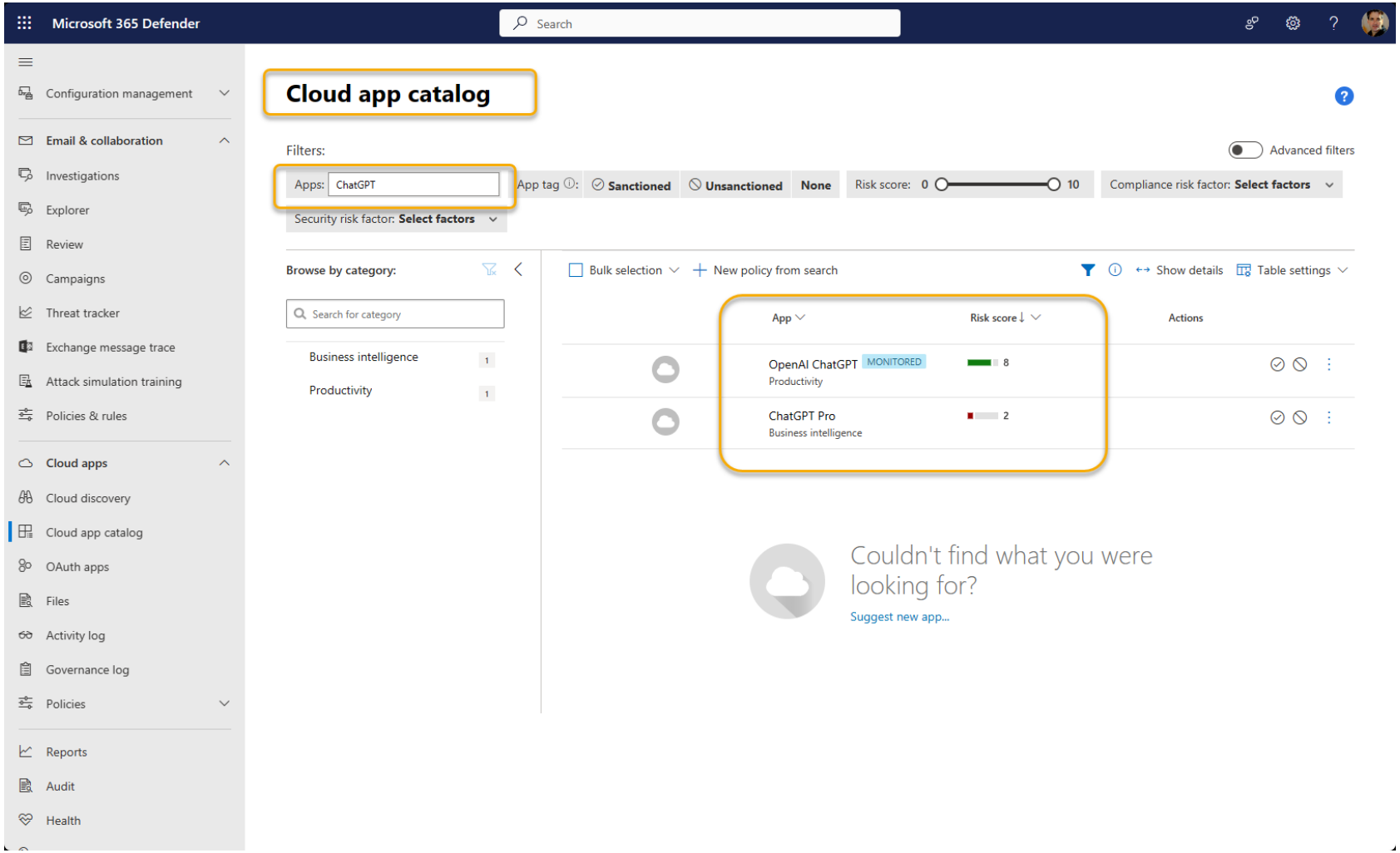
CloudAppEvents

- User activity

aka.ms/RodAI

365 Defender

Cloud Apps



Be prepared for incoming AI security standards

- Many jurisdictions are drafting AI standards: 21 draft standards were created in 2021.
- Two notable standards are:
  - NIST's AI Management Framework (USA)
  - The EU AI Act (AIA)
- The proposed EU AIA aims to be a comprehensive regulatory scheme for high-risk AI systems in already regulated industries e.g. medical, aviation, industrial control systems, etc.
- The NIST AI Management Framework is intended for voluntary use to better manage risks to individuals, organizations, and society associated with artificial intelligence (AI).
- As with GDPR, it is likely that the Brussels effect will happen to AI security standards with the AIA comes into force and many organizations worldwide will be obligated to confirm to AIA.

Information on Microsoft Copilots

- Bing and Edge
  https://blogs.windows.com/windowsexperience/2023/04/15/bing-and-edge-introduce-new-ai-features-to-enhance-your-search-and-browsing-experience/
- Microsoft 365 Copilot
  https://www.microsoft.com/en-us/microsoft-365/blog/2023/03/16/introducing-microsoft-365-copilot-a-whole-new-way-to-work/
- Windows Copilot:
  https://blogs.windows.com/windowsdeveloper/2023/05/23/bringing-the-power-of-ai-to-windows-11-unlocking-a-new-era-of-productivity-for-customers-and-developers-with-windows-copilot-and-dev-home/
- Dynamics 365 Copilot
  https://cloudblogs.microsoft.com/dynamics365/bdm/2023/03/06/introducing-microsoft-dynamics-365-copilot-bringing-next-generation-ai-to-every-line-of-business/
- Microsoft Fabric Copilot
  https://azure.microsoft.com/en-us/blog/introducing-microsoft-fabric-data-analytics-for-the-era-of-ai/
- Microsoft Security Copilot
  https://blogs.microsoft.com/blog/2023/03/28/introducing-microsoft-security-copilot-empowering-defenders-at-the-speed-of-ai/
- GitHub Copilot
  https://github.blog/2023-03-22-github-copilot-x-the-ai-powered-developer-experience/
- PowerApps Copilot
  https://powerapps.microsoft.com/en-us/blog/announcing-a-next-generation-ai-copilot-in-microsoft-power-apps-that-will-transform-low-code-development/

Resources

Weekly Newsletters

- Azure OpenAI: https://aka.ms/AzureOpenAINewsletter
- Microsoft Sentinel: https://aka.ms/MicrosoftSentinelNewsletter
- Microsoft Defender: https://aka.ms/MicrosoftDefenderNewsletter

LinkedIn Community Groups

- Azure OpenAI: https://aka.ms/AzureOpenAILinkedIn
- Microsoft Sentinel: http://aka.ms/SentinelLinkedIn
- Microsoft Defender: https://aka.ms/DefenderLinkedIn

Weekly Show/Podcast

Live every Wednesday (4 or 5pm EST). Distributed to all podcast networks every Friday

- https://aka.ms/GetMSIShow

Microsoft

Thank you!