



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)

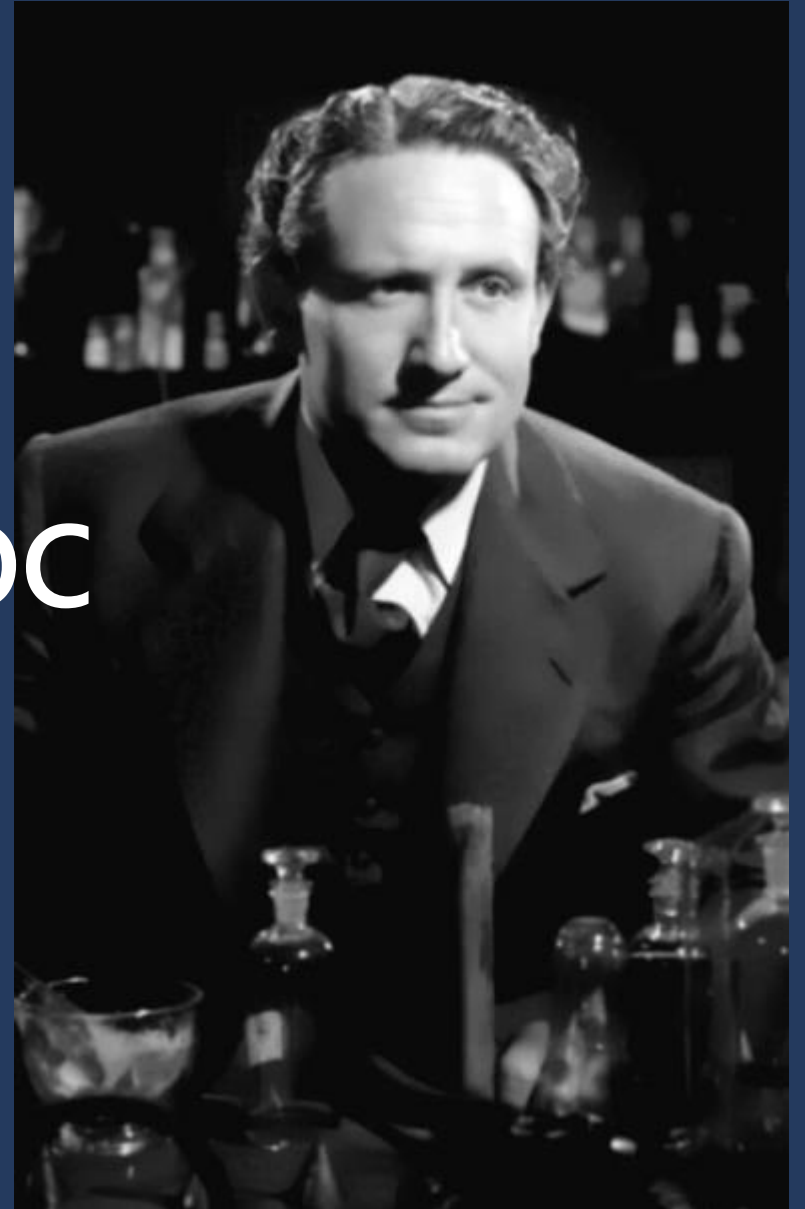
<https://aka.ms/AzOpenAINews>



myITforum.com
Powered by you

Part 1: Using AI to Deliver SOC Intelligence and Efficiency

Dr Sentinel



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"



Generative means relating to or capable of production or reproduction ¹. It can refer to the production of offspring or new life ^{2 3}, or to other generative forces or powers ^{2 3}. In linguistics, it denotes 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 ¹.

Learn more:

- 1. [bing.com](https://www.bing.com)
- 2. [merriam-webster.com](https://www.merriam-webster.com)
- 3. [dictionary.com](https://www.dictionary.com)
- 4. [dictionary.cambridge.org](https://www.dictionary.cambridge.org)
- +1 more



What is generative linguistics?

What are some examples of generative forces?

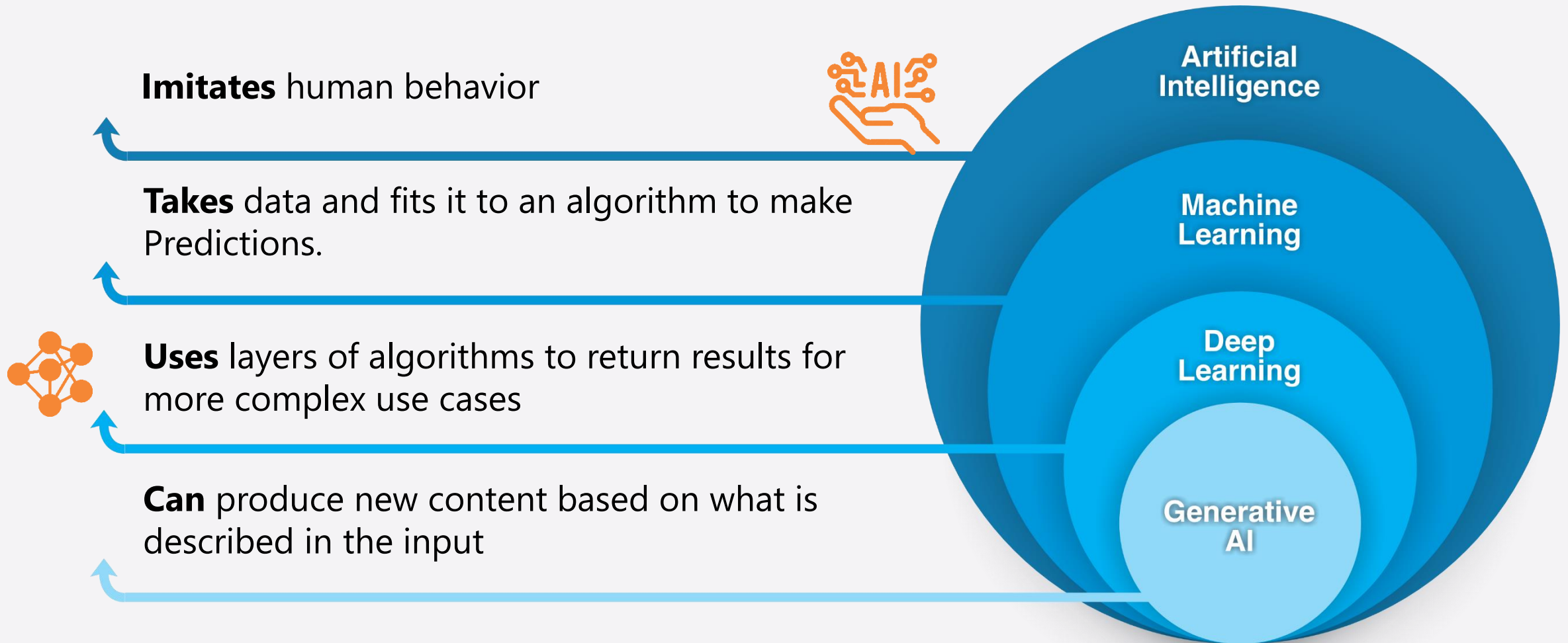
What are some examples of generative powers?



Let's chat

What is generative AI?

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



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

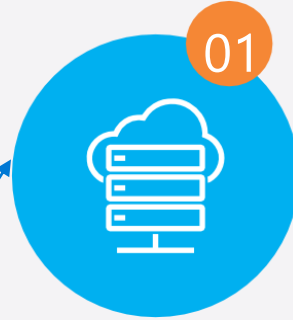
Prompt: A white Siamese cat

Response:



What is Azure OpenAI?

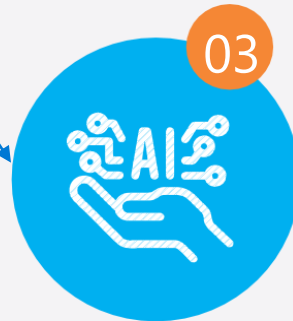
Microsoft has partnered with OpenAI to deliver on three main goals:



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



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



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



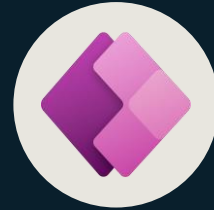
Better threat detection, identification, and mitigation

GitHub Copilot



Better code development

Power Platform Copilot



Better creation of apps, workflows, and agents

Copilots for Business

Copilots for Analytics

Copilot for Security

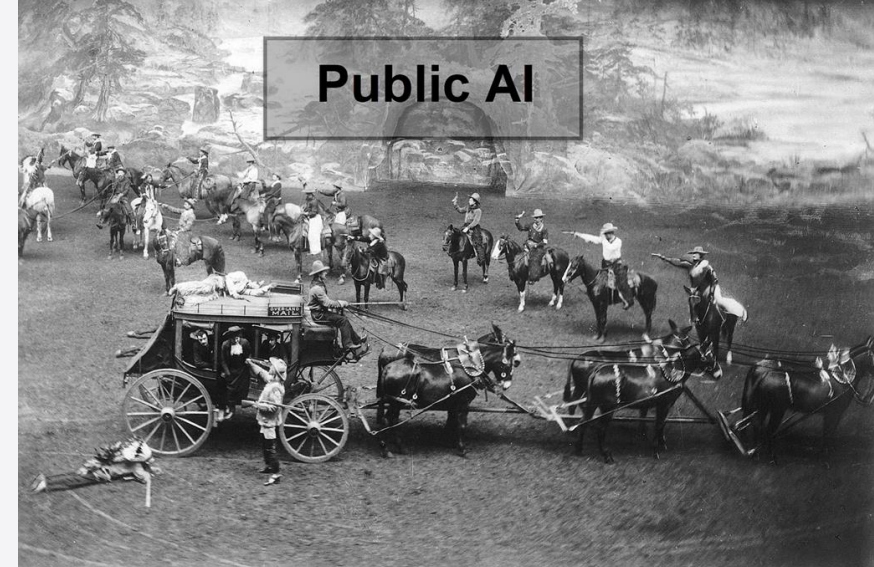
Copilot for Development

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.

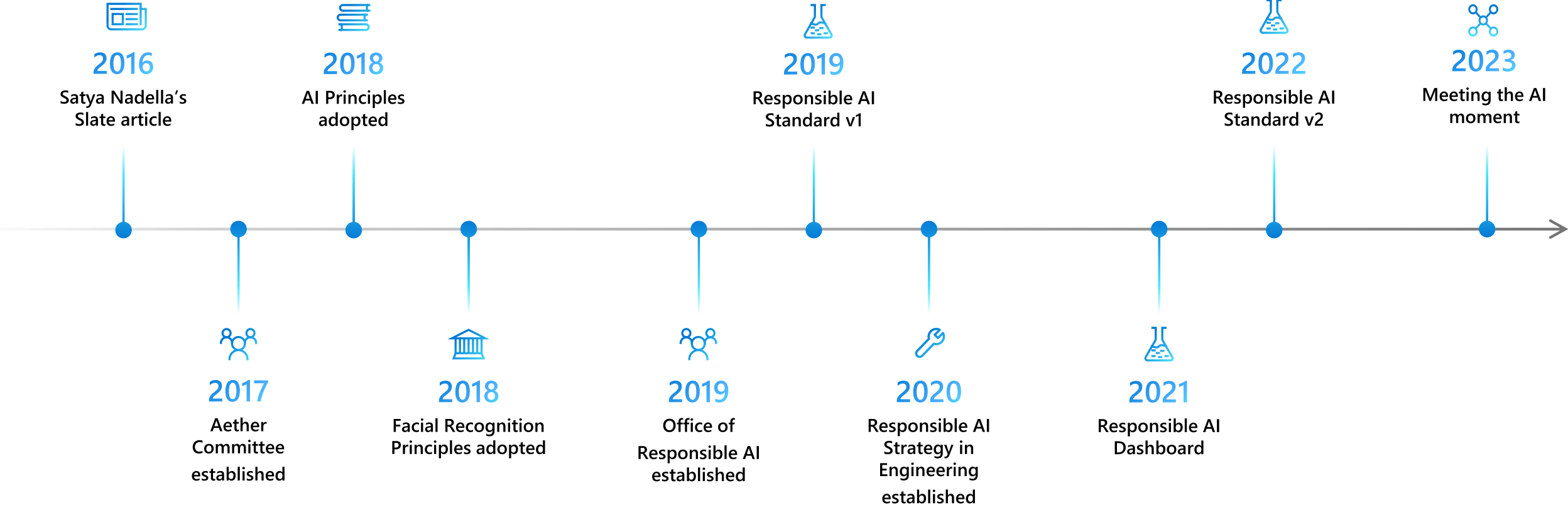
05 Accountability: People must be accountable for how AI systems operate.

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



DALL-E playground (Preview)

Playground

[View code](#) [Settings](#)

Prompt ⓘ

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

Generate

⊗ DALL-E text to image

Your task failed as a result of our safety system.

Tile Size

Search

Medium tiles ▾

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?

Azure AI Studio > DALL-E playground (Preview)

DALL-E playground (Preview)

Playground

[View code](#) [Settings](#)

Prompt ⓘ

darth vader and luke skywalker in a pillow fight in space

⊗ **DALL-E text to image**
Your task failed as a result of our safety system.

Search Tile Size

DALL-E playground (Preview)

Playground

[View code](#) [Settings](#)

Prompt ⓘ

darth vader and luke skywalker hitting each other with pillows in space

⊗ **DALL-E text to image**
Your task failed as a result of our safety system.

Search Tile Size: [Medium tiles](#) ▾

Generate

Access to Azure OpenAI

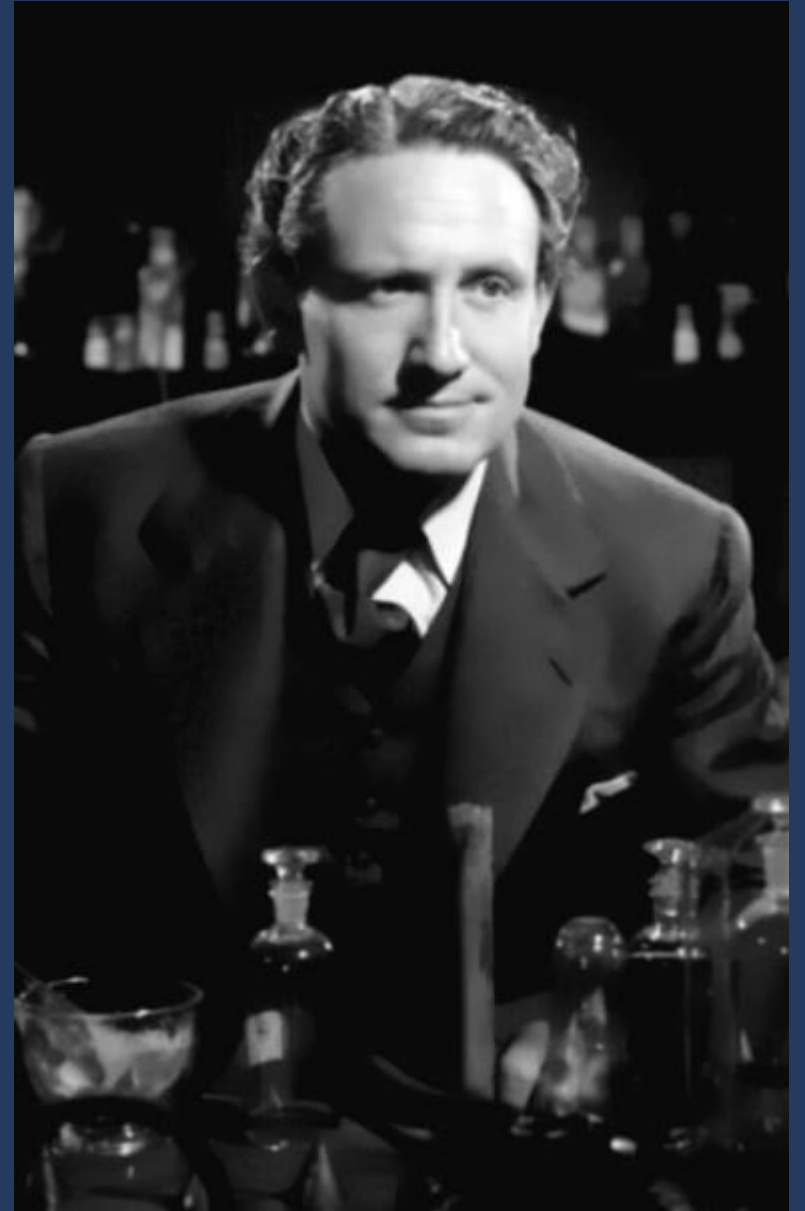


You need to [apply](#) for access to Azure OpenAI.

<https://aka.ms/oaiapply>

Using AI to Deliver SOC Intelligence and Efficiency

Dr Sentinel



Copilot

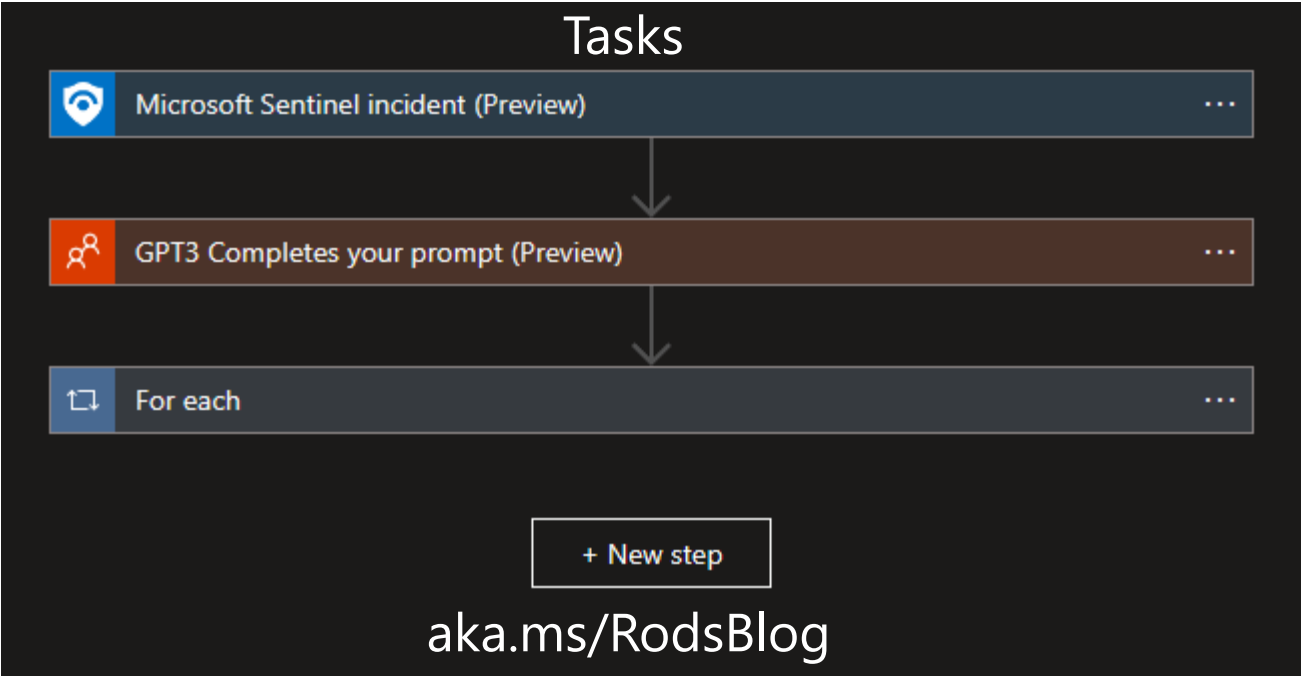
An **experience** using **generative AI**
to **assist humans** with **complex cognitive**
tasks^a

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

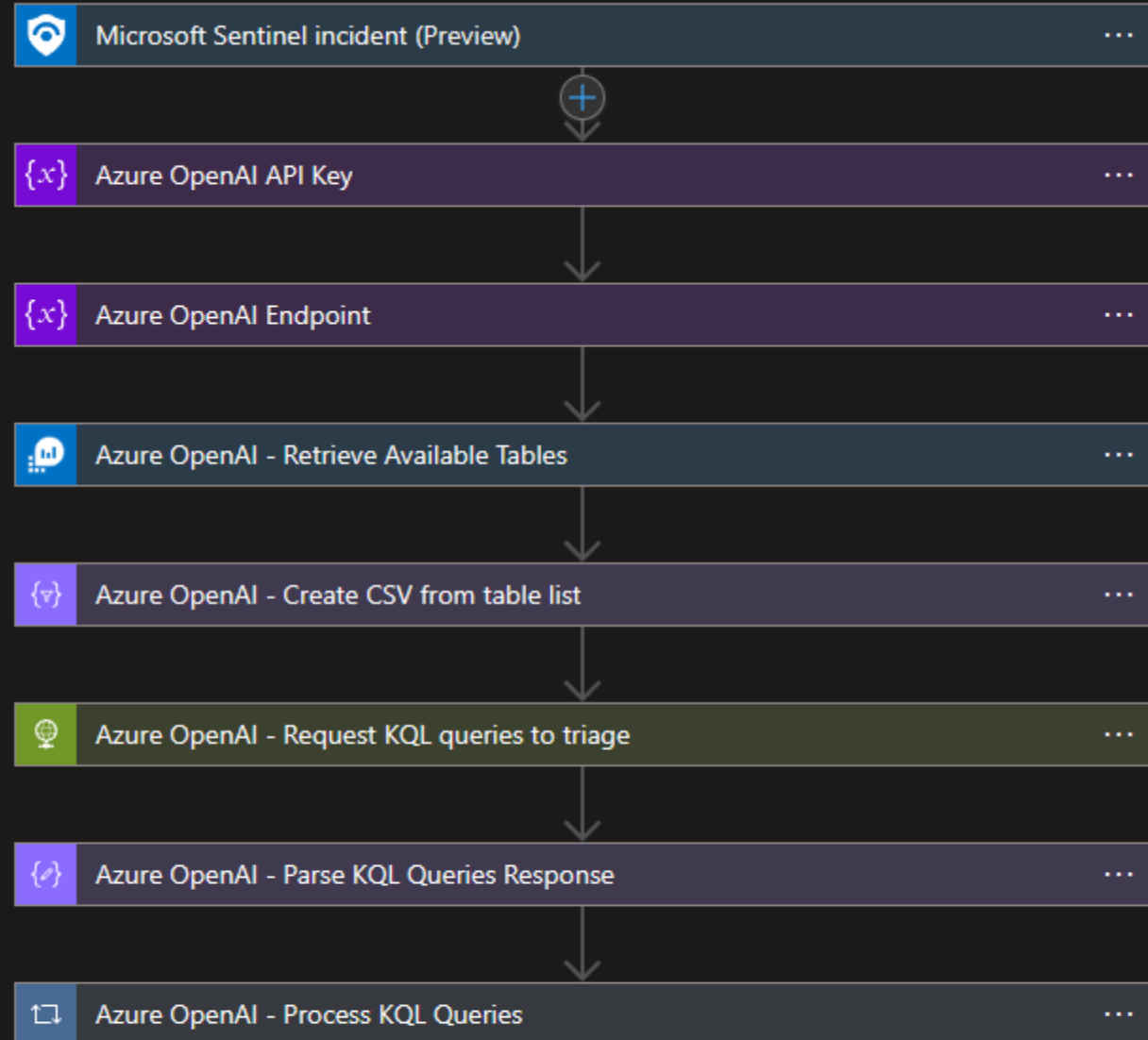


KQL

aka.ms/MustLearnKQL

AzureOpenAI-Enrichment

Activity Log



+ New step

aka.ms/RodsBlog

Assistant (Copilot)

aka.ms/RodsBlog

The screenshot shows the Microsoft Sentinel Incidents dashboard. At the top, it displays '11 Open incidents', '11 New incidents', and '0 Active incidents'. Below this is a bar chart titled 'Open incidents by severity' showing 6 High, 4 Medium, 0 Low, and 1 Informational incidents. A table lists incidents with columns for Severity, Incident ID, and Title. The selected incident is 'Successful logon from IP and failure...' with Incident ID 10035. The incident details panel shows a description, alert product names (Microsoft Sentinel), evidence (2 Events, 1 Alerts, 0 Bookmarks), last update and creation times (07/17/23, 09:03 AM), and entities (Credential Access (1)).

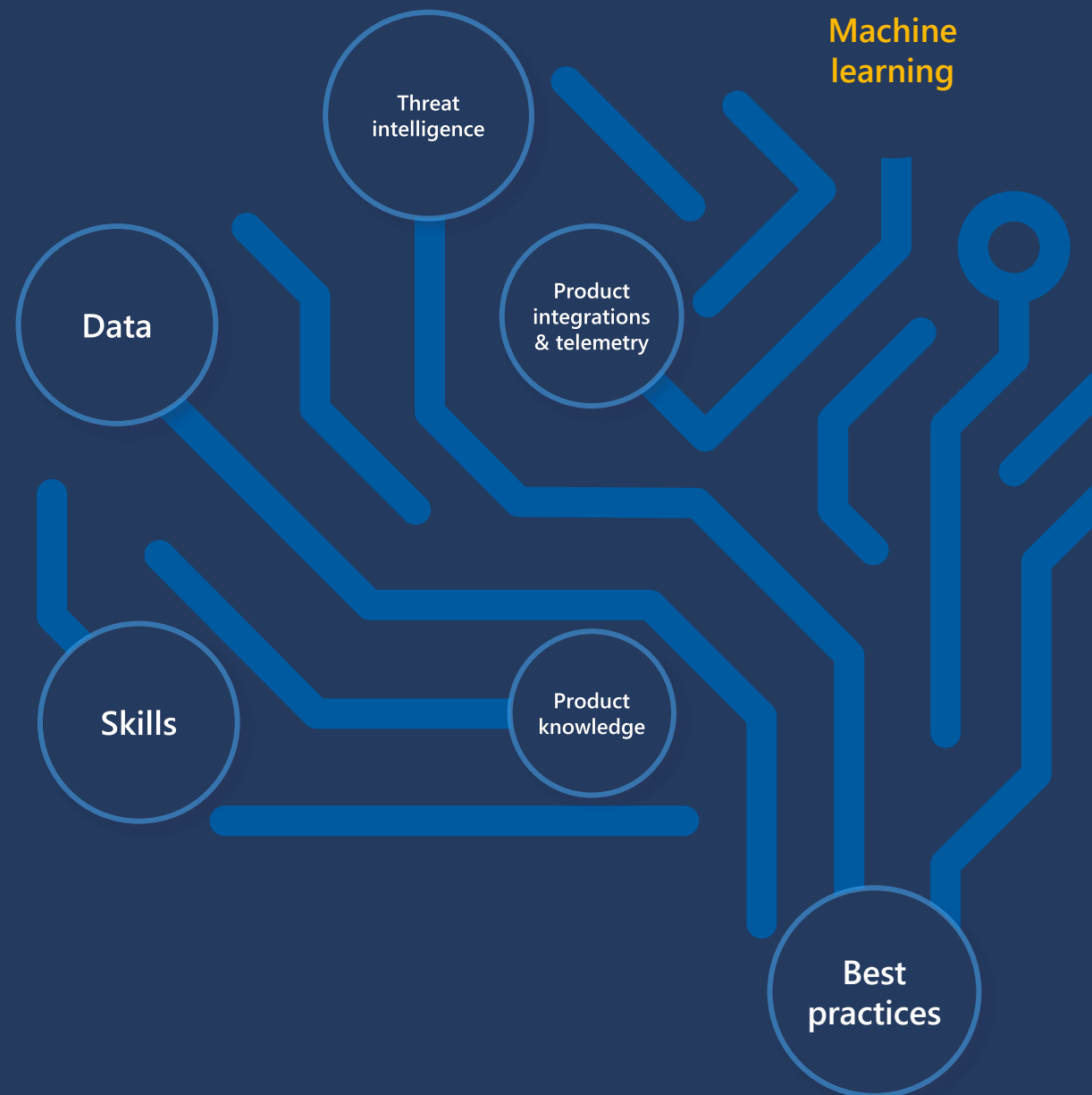
Severity	Incident ID	Title
High	10036	Office365 Con
Medium	10035	Successful log
Medium	10034	Intune Device
Medium	10033	Rare RDP Con
High	10032	Office365 Con
High	10031	Office365 Con
High	10030	Office365 Con
High	10029	Office365 Con
Medium	10028	(Preview)M21
High	10027	Office365 Con
Informational	10026	URL Detonatic

The screenshot shows the Azure AI chatbot interface. It features the Azure logo at the top left and a 'Share' button at the top right. The main content area displays the text 'Start chatting' and 'This chatbot is configured to answer your questions'. At the bottom, there is a text input field with the placeholder 'Type a new question...' and a send button.



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?

How can I improve my security posture

Jan 10, 12:38 PM

Here are three evidence-based recommendations to improve your security posture:

- 1. Use Microsoft Intune ASR rules to prevent common attack techniques:** This will help protect your devices from common attack techniques. Currently, 438,174 out of 758,659 devices are impacted. Learn more about ASR rules [here](#).
- 2. Use Microsoft Intune to enforce the latest Defender security baseline:** This will ensure that your devices are up-to-date with the latest security settings. Currently, 438,174 out of 758,659 devices are impacted. Learn more about the Defender security baseline [here](#).
- 3. Use Microsoft Intune Endpoint Privilege Management to reduce exposure by enabling standard users and limiting local admin accounts:** This will help minimize the attack surface by reducing the number of local admin accounts. Currently, 438,174 out of 758,659 devices are impacted. Learn more about Endpoint Privilege Management [here](#).

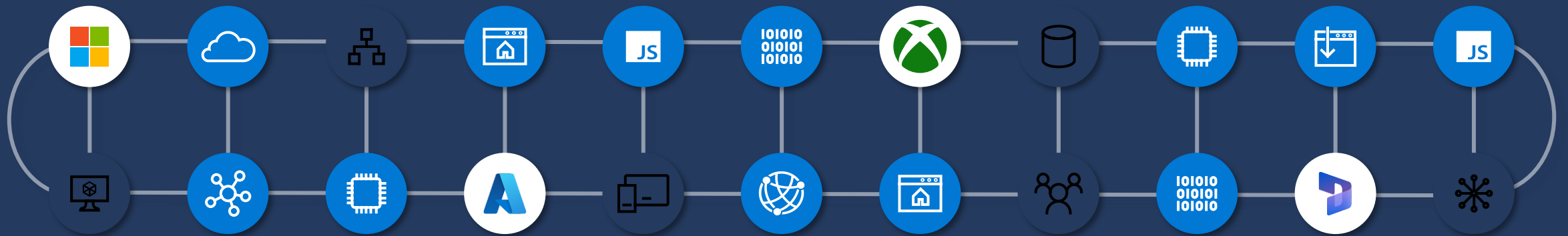
▼ Sources

Microsoft Intune

Confirm Off-target Report Pin

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 endpoints¹ 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

101010
010101
101010



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

Storage Container (Blob)

Step 1

The screenshot displays the Microsoft Azure portal interface for a storage account named 'sixmdmaistorage'. The main view is 'Containers', showing a list of containers with the following details:

Name	Last modified	Public access level	Lease state
<input type="checkbox"/> \$logs	6/20/2023, 10:42:30 AM	Private	Available
<input type="checkbox"/> fileupload-sixmdmepisodes	6/20/2023, 12:13:23 PM	Private	Available
<input type="checkbox"/> rodbotcontainer	7/12/2023, 1:34:01 PM	Private	Available

The left-hand navigation pane includes categories such as Overview, Activity log, Tags, Diagnose and solve problems, Access Control (IAM), Data migration, Events, Storage browser, Data storage, Containers, File shares, Queues, Tables, Security + networking, Networking, Access keys, Shared access signature, Encryption, Microsoft Defender for Cloud, Data management, Redundancy, Data protection, Blob inventory, and Static website.

At the bottom of the page, the URL is: <https://portal.azure.com/#@sixmilliondollarman.onmicrosoft.com/resource/subscriptions/2959f496-711b-447e-9b85-53a1da341c4c/resourceGroups/6MDMOpenAIResGroup/providers/Microsoft.Storage/storageAccounts/sixmdmaistorage/containersList>

Indexer (Cognitive Search)

Step 2

The screenshot displays the Microsoft Azure portal interface for a Cognitive Search service named 'sixmdmrodbotsearch'. The left-hand navigation pane includes sections for Overview, Activity log, Access control (IAM), Tags, Diagnose and solve problems, Settings, Semantic search (Preview), Knowledge Center, Keys, Scale, Search traffic analytics, Identity, Networking, Properties, Locks, Monitoring (Alerts, Metrics, Diagnostic settings, Logs), Automation (Tasks (preview), Export template), and Support + troubleshooting.

The main content area shows the 'Essentials' section with the following details:

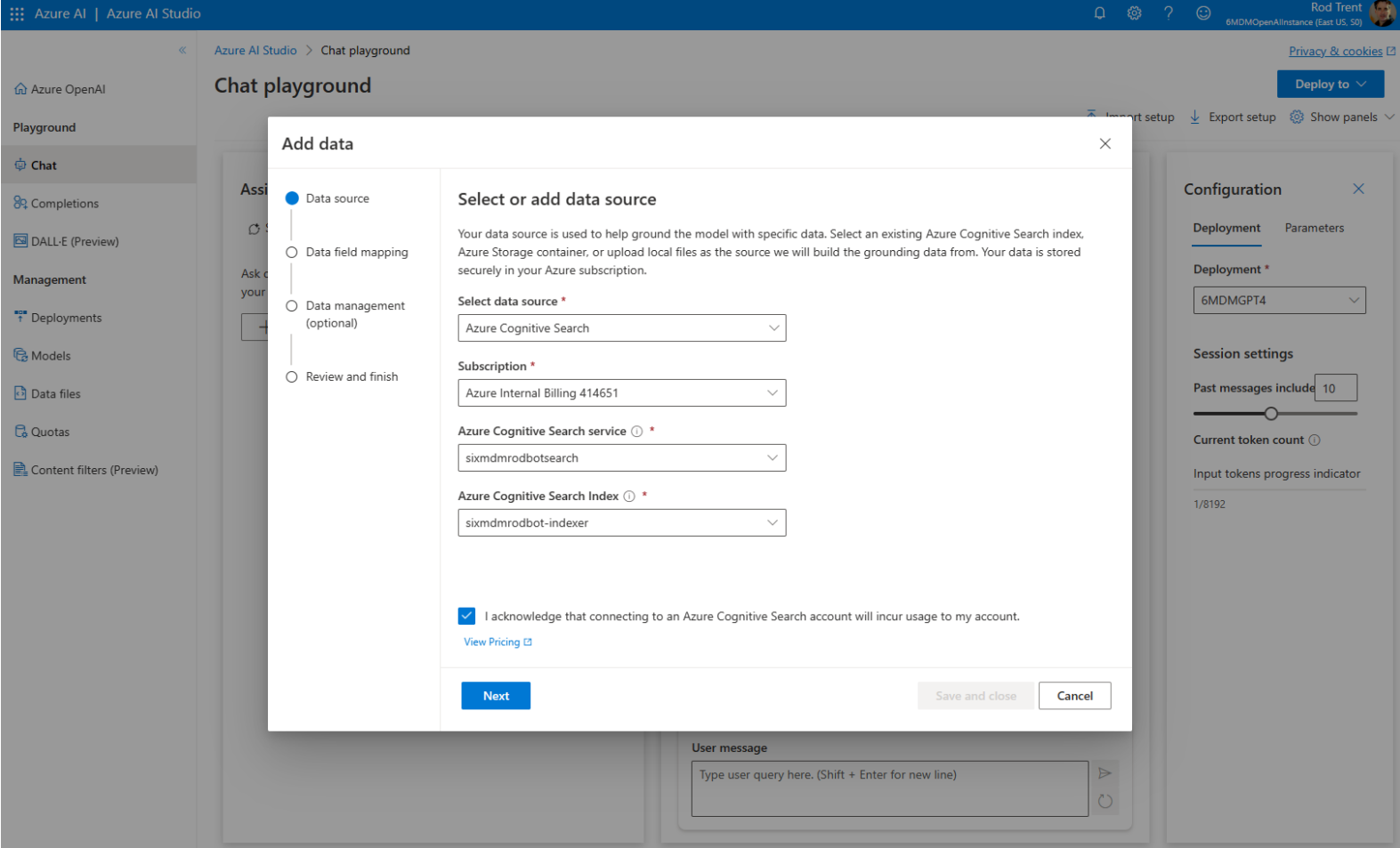
- Resource group: 6MDMOpenAIResGroup
- Location: East US
- Subscription: Azure Internal Billing 414651
- Subscription ID: 2959f496-711b-447e-9b85-53a1da341c4c
- Status: Running
- Url: https://sixmdmrodbotsearch.search.windows.net
- Pricing tier: Standard
- Replicas: 1 (No SLA)
- Partitions: 1
- Search units: 1

Below the Essentials section, there is a 'Tags' section with a tag labeled 'ProjectType : aoi-your-data-service'. A navigation bar includes 'Get started', 'Usage', 'Monitoring', 'Indexers', 'Data sources', 'Aliases', 'Skillsets', and 'Debug sessions'. The 'Indexers' tab is active, showing a table with the following data:

Name	Document Count	Storage Size
sixmdmrodbot-indexer	28	69.91 KB

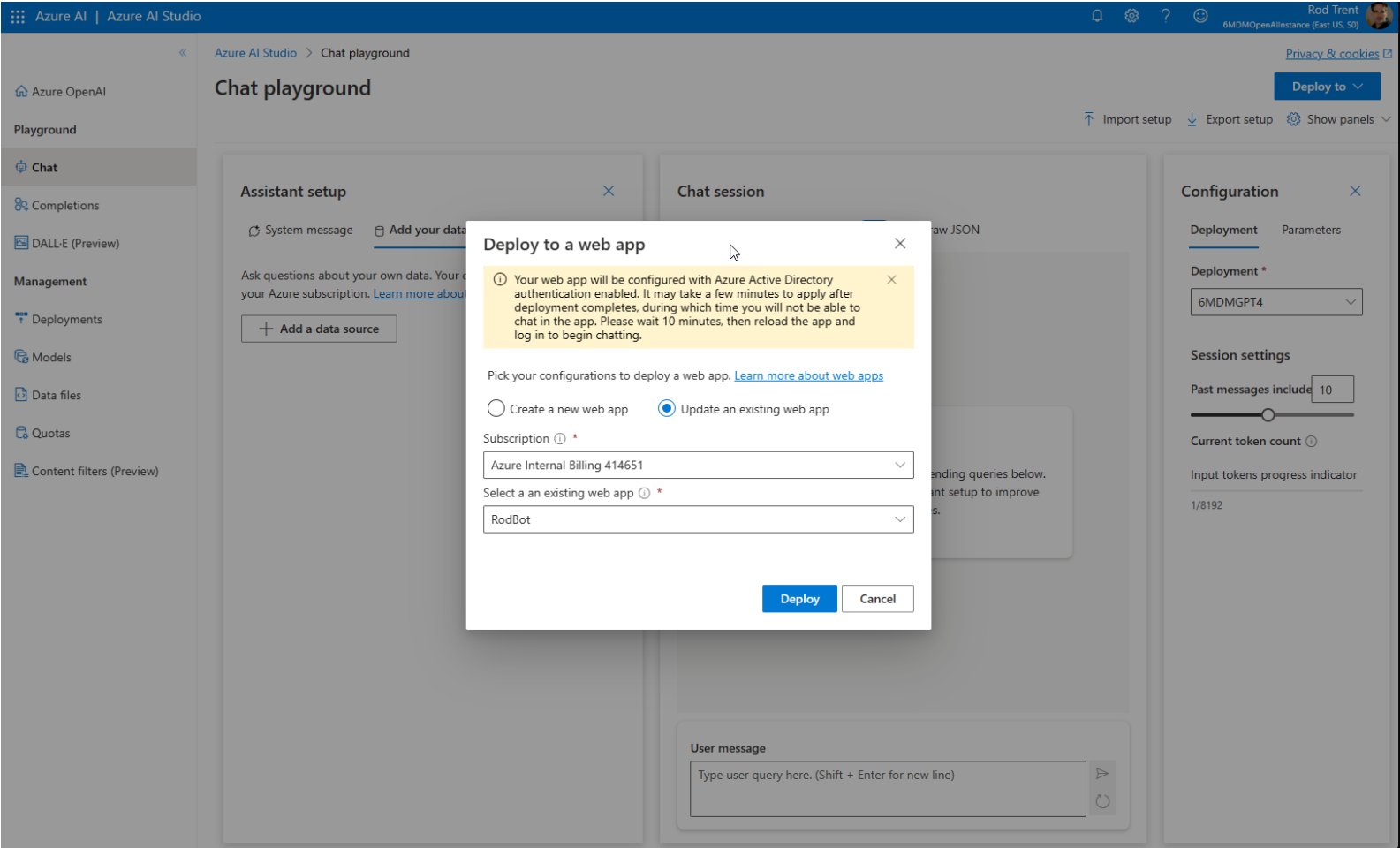
Add Your Data (Indexer)

Step 3



Deploy to App (New or Existing)

Step 4



Add to Edge Sidebar

Step 5

The image shows two browser windows side-by-side. The left window is Microsoft Sentinel, displaying the 'Incidents' page. The right window is the Azure AI chatbot interface.

Microsoft Sentinel Incidents Page:

- Header: Microsoft Azure, Search resources, services, and docs (G+)
- Breadcrumbs: Home > Microsoft Sentinel > Microsoft Sentinel
- Page Title: Microsoft Sentinel | Incidents
- Selected workspace: 'rodazuresentinelworkspace'
- Search bar and actions: Create incident (Preview), Refresh, Last 24 hours, Actions, Delete
- General section: Overview (Preview), Logs, News & guides, Search
- Threat management section: Incidents, Workbooks, Hunting, Notebooks, Entity behavior, Threat intelligence, MITRE ATT&CK (Preview)
- Content management section: Content hub, Repositories (Preview), Community
- Configuration section: Workspace manager (Preview), Data connectors, Analytics, Watchlist, Automation, Settings
- Incidents summary: 11 Open incidents, 11 New incidents, 0 Active incidents
- Open incidents by severity: High (6), Medium (4), Low (0), Informational (1)
- Table of incidents:

Severity	Incident ID	Title
High	10036	Office365 Con
Medium	10035	Successful log
Medium	10034	Intune Device
Medium	10033	Rare RDP Con
High	10032	Office365 Con
High	10031	Office365 Con
High	10030	Office365 Con
High	10029	Office365 Con
Medium	10028	(Preview)M21
High	10027	Office365 Con
Informational	10026	URL Detonatic

Azure AI Chatbot Interface:

- Header: Azure AI, Share
- Content: Start chatting, This chatbot is configured to answer your questions
- Input field: Type a new question...
- Send button: Paper plane icon

Maintaining Source (Improvements)

Upload to Container

The screenshot shows the Microsoft Azure portal interface for a storage account named 'rodbotcontainer'. The left sidebar contains navigation options like Overview, Diagnose and solve problems, Access Control (IAM), and Settings. The main content area shows a list of blobs with the following columns: Name, Modified, Access tier, and Archive. The 'Upload blob' dialog is open on the right, featuring a large dashed box for file upload, an 'Overwrite if files already exist' checkbox, and an 'Advanced' section with an 'Upload' button.

Name	Modified	Access tier	Archive
<input type="checkbox"/> Aggregate_functions.pdf	7/12/2023, 1:15:11 PM	Hot (Inferred)	
<input type="checkbox"/> Best_Practice.pdf	7/12/2023, 1:14:53 PM	Hot (Inferred)	
<input type="checkbox"/> Cross_Cluster.pdf	7/12/2023, 1:14:53 PM	Hot (Inferred)	
<input type="checkbox"/> Data_types.pdf	7/12/2023, 1:15:00 PM	Hot (Inferred)	
<input type="checkbox"/> Entity_types.pdf	7/12/2023, 1:14:56 PM	Hot (Inferred)	
<input type="checkbox"/> Function_types.pdf	7/12/2023, 1:14:54 PM	Hot (Inferred)	
<input type="checkbox"/> Geospatial.pdf	7/12/2023, 1:15:54 PM	Hot (Inferred)	
<input type="checkbox"/> Intro.pdf	7/12/2023, 1:14:56 PM	Hot (Inferred)	
<input type="checkbox"/> Limits_and_errors.pdf	7/12/2023, 1:15:00 PM	Hot (Inferred)	
<input type="checkbox"/> Plugins.pdf	7/12/2023, 1:14:59 PM	Hot (Inferred)	
<input type="checkbox"/> Query_statement_types.pdf	7/12/2023, 1:15:06 PM	Hot (Inferred)	
<input type="checkbox"/> Query_Tools.pdf	7/12/2023, 1:15:14 PM	Hot (Inferred)	
<input type="checkbox"/> Reference_material.pdf	7/12/2023, 1:15:11 PM	Hot (Inferred)	
<input type="checkbox"/> Scalar_function_types_1.pdf	7/12/2023, 1:15:20 PM	Hot (Inferred)	
<input type="checkbox"/> Scalar_function_types_2.pdf	7/12/2023, 1:15:26 PM	Hot (Inferred)	
<input type="checkbox"/> Scalar_function_types_3.pdf	7/12/2023, 1:15:30 PM	Hot (Inferred)	
<input type="checkbox"/> Scalar_function_types_4.pdf	7/12/2023, 1:15:29 PM	Hot (Inferred)	
<input type="checkbox"/> Scalar_function_types_5.pdf	7/12/2023, 1:15:29 PM	Hot (Inferred)	
<input type="checkbox"/> Scalar_operators_1.pdf	7/12/2023, 1:15:30 PM	Hot (Inferred)	
<input type="checkbox"/> Scalar_operators_2.pdf	7/12/2023, 1:15:37 PM	Hot (Inferred)	
<input type="checkbox"/> Scalar_operators_3.pdf	7/12/2023, 1:15:40 PM	Hot (Inferred)	

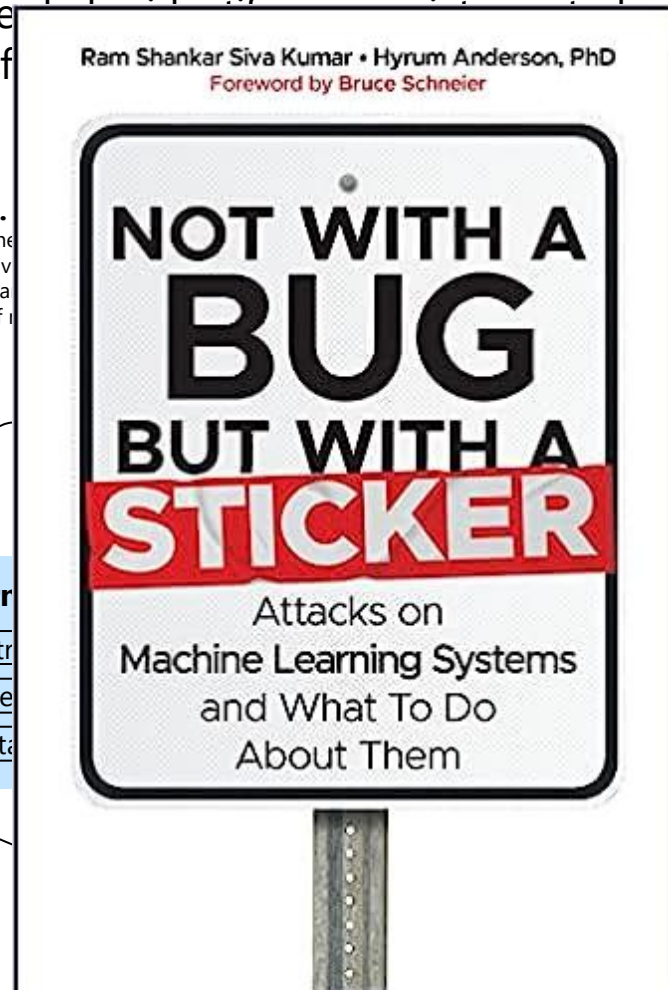
Monitoring and Securing AI Using Modern Tools

Mr AI

A black silhouette of a man in a top hat and suit, holding a cane. He is shown from the waist up, facing right. His right hand is raised, palm facing forward. The background features large, semi-transparent blue circles and a white horizontal line at the bottom.

Threats across AI system layers

Assigning AI Risks to these three categories... available today, and what we may need to develop in the future



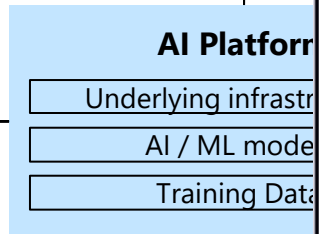
4. Inversion

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 Inverse Learning of...



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

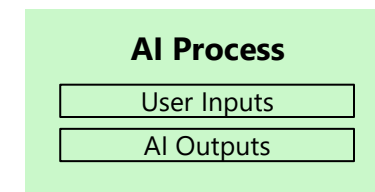
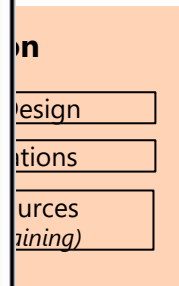
As with any computer system, interference in the infrastructure, operating systems, and applications that support the AI system will lead to software vulnerability exploitation

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.



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 & 5 techniques	Resource Development & 7 techniques	Initial Access & 4 techniques	ML Model Access 4 techniques	Execution & 2 techniques	Persistence & 2 techniques	Defense Evasion & 1 technique	Discovery & 3 techniques	Collection & 3 techniques	ML Attack Staging 4 techniques	Exfiltration & 2 techniques	Impact & 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

The screenshot shows a GitHub repository for user 'rod-trent'. The repository name is 'Add files via upload' and it has 175 commits. The file list includes folders like '.github/workflows', 'Code', 'Datasets', 'Docs', 'Misc', 'PowerShell', 'Resources', 'Security/Sentinel', and a file 'README.md'. The README.md content is visible below the list, featuring the title 'Open AI Security' and a description: 'Scripts, code, and content for working with Azure Open AI Security.'

File/Folder	Commit Message	Time Ago
.github/workflows	Add or update the Azure App Service build and deployment workflow c...	3 months ago
Code	Update WebChatBot.py	3 months ago
Datasets	Add files via upload	5 days ago
Docs	Update Security Advocacy - Responsible and Secure AI.md	last week
Misc	Update AICoffeeStory.md	4 months ago
PowerShell	Add files via upload	3 months ago
Resources	Update Readme.md	4 months ago
Security/Sentinel	Update AzureOpenAIChatGPTResponse.kqI	3 weeks ago
README.md	Update README.md	3 months ago

README.md

Open AI Security

Scripts, code, and content for working with Azure Open AI Security.

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).

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

Create content filtering configuration ✕

Content filtering configurations are created within a Resource and can be associated with Deployments.
[Learn more about configurability here.](#)

The default content filtering configuration is set to filter at the medium severity threshold for all four content harms categories for both, prompts and completions. That means that content that is detected at severity level medium or high is filtered, while content detected at severity level low is not filtered by the content filters.

Create custom configuration name

Set severity levels

Severity	User prompts (Input)			Model completions (Output)				
	Low	Medium	High	Low	Medium	High		
Hate	<input type="checkbox"/> On	✓	⊖	⊖	<input type="checkbox"/> On	✓	⊖	⊖
Sexual	<input type="checkbox"/> On	✓	⊖	⊖	<input type="checkbox"/> On	✓	⊖	⊖
Self-harm	<input type="checkbox"/> On	✓	⊖	⊖	<input type="checkbox"/> On	✓	⊖	⊖
Violence	<input type="checkbox"/> On	✓	⊖	⊖	<input type="checkbox"/> On	✓	⊖	⊖

[Learn more about content filters here](#)

Save

Cancel

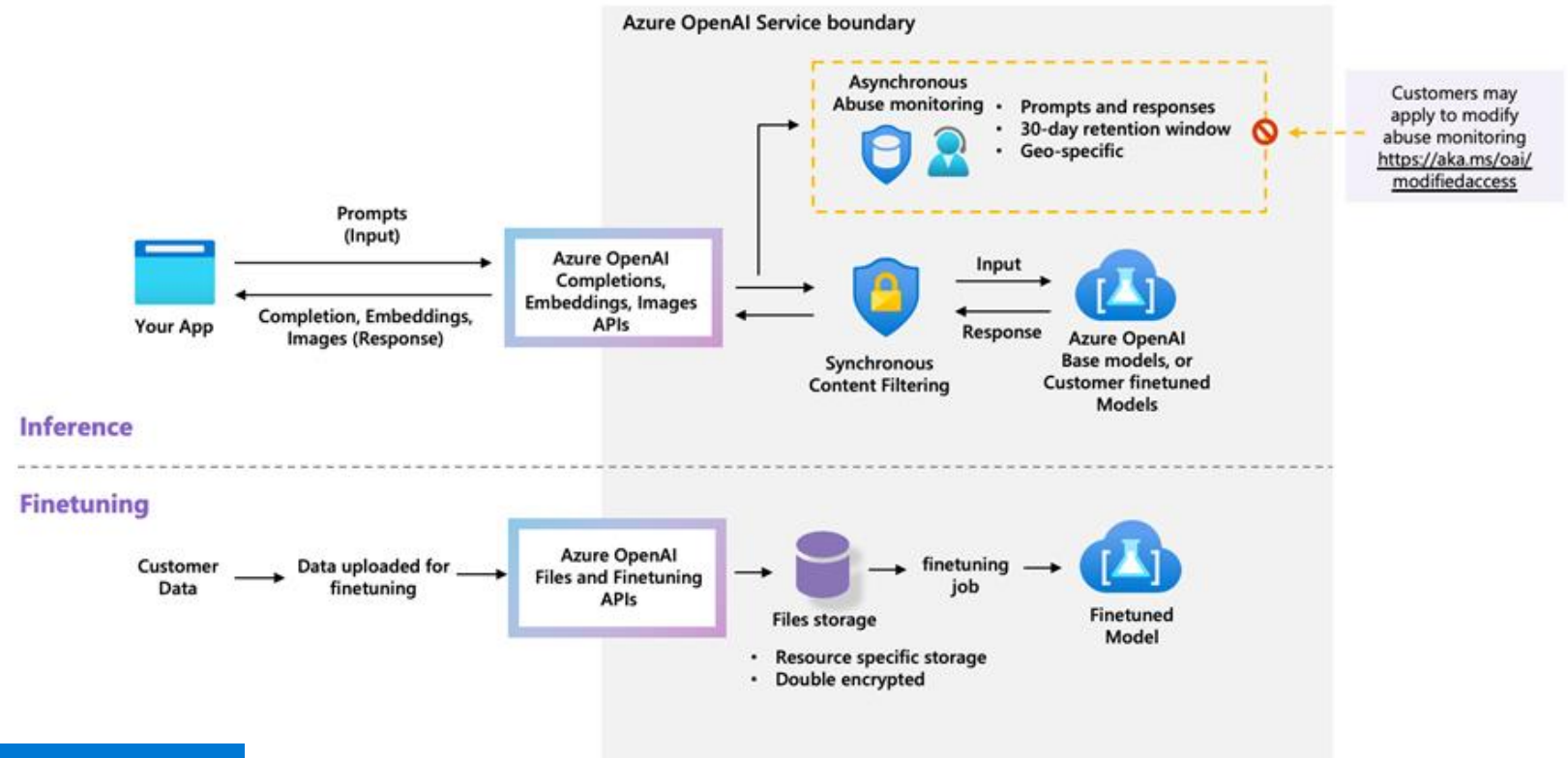
Compliance

*Abuse
Monitoring/Content
Filtering*

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

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

Azure OpenAI | Data flows for inference and training



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

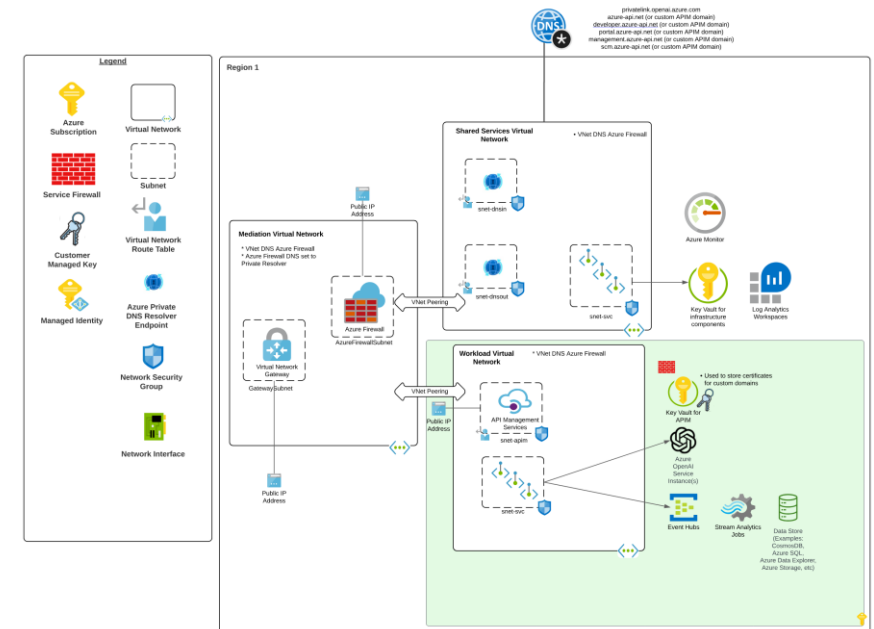
Advanced Logging

<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

Microsoft 365 Defender

Search

Cloud app catalog

Filters: App tag: Sanctioned Unsanctioned None Risk score: 0 10 Compliance risk factor: **Select factors** Advanced filters

Security risk factor: **Select factors**

Browse by category:

- Business intelligence 1
- Productivity 1

App	Risk score	Actions
OpenAI ChatGPT Productivity	MONITORED 8	<input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/>
ChatGPT Pro Business intelligence	2	<input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/>

Couldn't find what you were looking for?
[Suggest new app...](#)

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 conform 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>



Thank you!
