

# TURBOCHARGE EUC WITH GENERATIVE AI



Ryan Mangan



# SESSION DISCLAIMER

While the Presenter has made every effort to ensure that the information in this presentation is accurate, we do not assume and hereby disclaim any liability to any party for any loss, damage, or disruption caused by errors or omissions. This holds true whether such errors or omissions result from negligence, accident, or any other cause.

# RYAN MANGAN

CEO @ EfficientEther Ltd

Ryan Mangan is a seasoned technologist. The founder of EfficientEther Ltd, he focuses on leveraging AI for cloud optimisation, sustainability, assistance, and AI tooling. Ryan is also the creator of EtherAssist.AI, which offers a 14-day free trial, and Efficient Ether Cloud, both designed to drive innovation and efficiency in cloud operations. In addition, he is an accomplished author and technical reviewer on a wide range of topics, including building cloud applications.



[WWW.ETHERASSIST.AI](http://WWW.ETHERASSIST.AI)





2ND EDITION

# Mastering Azure Virtual Desktop

A practical guide to designing, implementing, and managing Azure Virtual Desktop environments



RYAN MANGAN | NEIL MCLOUGHLIN | MARCEL MEURER

Foreword by Christiaan Brinkhoff,  
Principal PM, Azure Virtual Desktop and Windows 365



# AGENDA

- Introduction to Generative AI
- Foundations of AI Success
- Using and Building AI Capabilities
- Practical Use Cases
- Additional Applications
- Summary
- Q&A

# INTRODUCTION

The IT industry is facing a significant talent gap. According to IDC, by 2027, there will be a shortage of 10 million IT operations professionals.

## Manual Processes Are Inefficient

Tasks like log analysis, compliance checks, and documentation consume valuable time.

## Increasing Complexity

Hybrid and multi-cloud setups require expertise across various platforms, leading to delays and errors.

## Compliance and Security Risks

Meeting regulatory standards with limited resources adds additional workload and stress to IT teams.

## Knowledge Gaps

Teams struggle to find fast, reliable answers without centralized expertise.



# AI ?



Raise Your Hands If you are using ML ?



Raise Your Hands If you are using NLP ?



Raise Your Hands If you are using other  
AI?



# HIGH-LEVEL AI STAGES

AI Stages	Artificial Narrow Intelligence (ANI)	Artificial General Intelligence (AGI)	Artificial Super Intelligence (ASI)
Description	Execute specific focused tasks, without ability to self-expand functionality	Perform broad tasks, reason, and improve capabilities comparable to humans	Demonstrate intelligence beyond human capabilities
Timing	Today	2050 vs 2075	TBC
Implications	Outperform humans in specific repetitive functions, such as driving, medical diagnosis, and financial advice	Compete with humans across all endeavours, such as earning university degrees and convincing humans that it is human	Outperform humans, helping to achieve societal objectives or threatening the human race

<https://research.aimultiple.com/artificial-general-intelligence-singularity-timing/>

# OPENAI'S 5 LEVELS OF 'SUPER AI'...

AI Levels	Description	Timing	Implications	Impact on Jobs
<b>Level One: Conversational AI</b>	Simple AI that can handle basic conversations and interactions	Today	Enhances customer service and user interactions	Jobs enhanced
<b>Level Two: Reasoning AI</b>	Today	2050 vs 2075	TBC	Some jobs at risk due to automation
<b>Level Three: Autonomous AI</b>	Outperform humans in specific repetitive functions, such as driving, medical diagnosis, and financial advice	Compete with humans across all endeavours, such as earning university degrees and convincing humans that it is human	Outperform humans, helping to achieve societal objectives or threatening the human race	Significant job displacement potential
<b>Level Four: Innovating AI</b>	AI that can create new ideas, innovations, and improve existing processes	Long-term Future	Drives innovation and improvements across various fields	Jobs at risk; new job roles created
<b>Level Five: Organizational AI</b>	AI that can manage and optimize entire organizations	Distant Future	Transforms organizational structures and efficiency	Major reorganization of job roles and functions

**Enterprise Adoption: A McKinsey survey indicates that AI adoption in organisations has risen to 72% in 2024, with a notable increase in generative AI applications.**

**McKinsey & Company**

**Industry-Specific Models: By 2027, over 50% of the generative AI models used by enterprises will be tailored to specific industries or business functions, up from approximately 1% in 2023.**

Gartner

# 2024 HIGHLIGHTS: GEN AI

## OpenAI's GPT-4o and o1 Models:

GPT-4o: Multimodal processing of text, images, and audio; supports up to 32,768 tokens.

o1: Enhanced reasoning capabilities for complex tasks in science and programming.

## Anthropic's Claude 3.5 Sonnet:

Excels in reading comprehension, coding, mathematics, and vision tasks. Emphasizes safety through Constitutional AI, reducing reliance on human feedback.

## Nvidia's Fugatto AI Model:

Modifies voices and generates novel sounds for creators in music, film, and gaming.

Transforms audio, such as converting piano lines into human vocals.

## Alibaba's Qwen 2.5 Models:

Open-source AI models ranging from 0.5 to 72 billion parameters. Excel in mathematics, coding, and multi-language support.

<https://www.theverge.com/2024/5/13/24155493/openai-gpt-4o-launching-free-for-all-chatgpt-users>

<https://www.theverge.com/2024/9/12/24242439/openai-o1-model-reasoning-strawberry-chatgpt>

<https://venturebeat.com/ai/anthropic-unveils-claude-3-5-sonnet-pushing-the-boundaries-of-ai-capabilities-and-affordability/>

<https://www.reuters.com/technology/artificial-intelligence/openai-rivals-seek-new-path-smarter-ai-current-methods-hit-limitations-2024-11-11/>

<https://neoteric.eu/blog/gpt-o1-vs-gpt-4o-comparison/>

<https://techcrunch.com/2024/09/12/openai-unveils-a-model-that-can-fact-check-itself/>

<https://www.anthropic.com/news/3-5-models-and-computer-use>

<https://help.openai.com/en/articles/9624314-model-release-notes>

[https://www.theverge.com/2024/10/4/24261990/meta-movie-gen-ai-video-generator-openai-sora?utm\\_source=chatgpt.com](https://www.theverge.com/2024/10/4/24261990/meta-movie-gen-ai-video-generator-openai-sora?utm_source=chatgpt.com)

## Meta's Movie Gen:

AI video-generation tool creating realistic videos from text instructions. Generates 16-second videos, edits existing footage, matches sound, and personalizes content.

## Runway's Gen-3 Alpha:

Generates 10-second video clips from text, image, or video prompts. Autonomously learns 3D dynamics, paving the way for photorealistic video production.

## Advancements in Token Sizes:

Models like GPT-4o support up to 32,768 tokens, allowing processing of longer and more complex inputs.

## Microsoft's Copilot Enhancements:

Copilot Actions: Automates tasks in Microsoft 365 apps, such as summarising meetings.

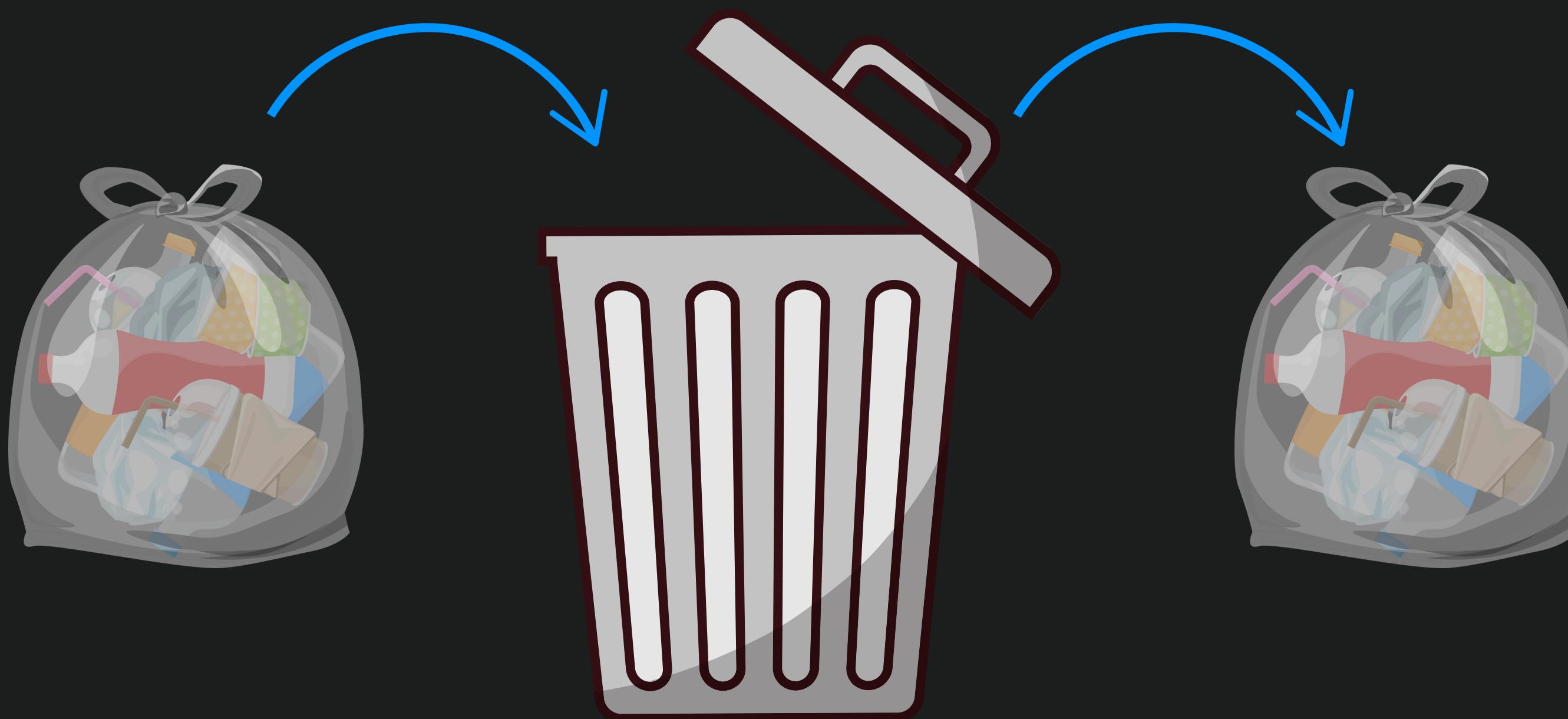
Copilot Vision: Incorporates voice and vision capabilities for natural interactions.

Copilot+ PCs: AI-integrated personal computers designed for on-device AI processing.



# Before We Start





**Garbage in..... Garbage out.....**

# Misunderstandings

Example 1: "An English-speaking person offers a 'gift,' meaning a present, to a German speaker. However, the German speaker thinks it means 'poison,' leading to a misunderstanding"



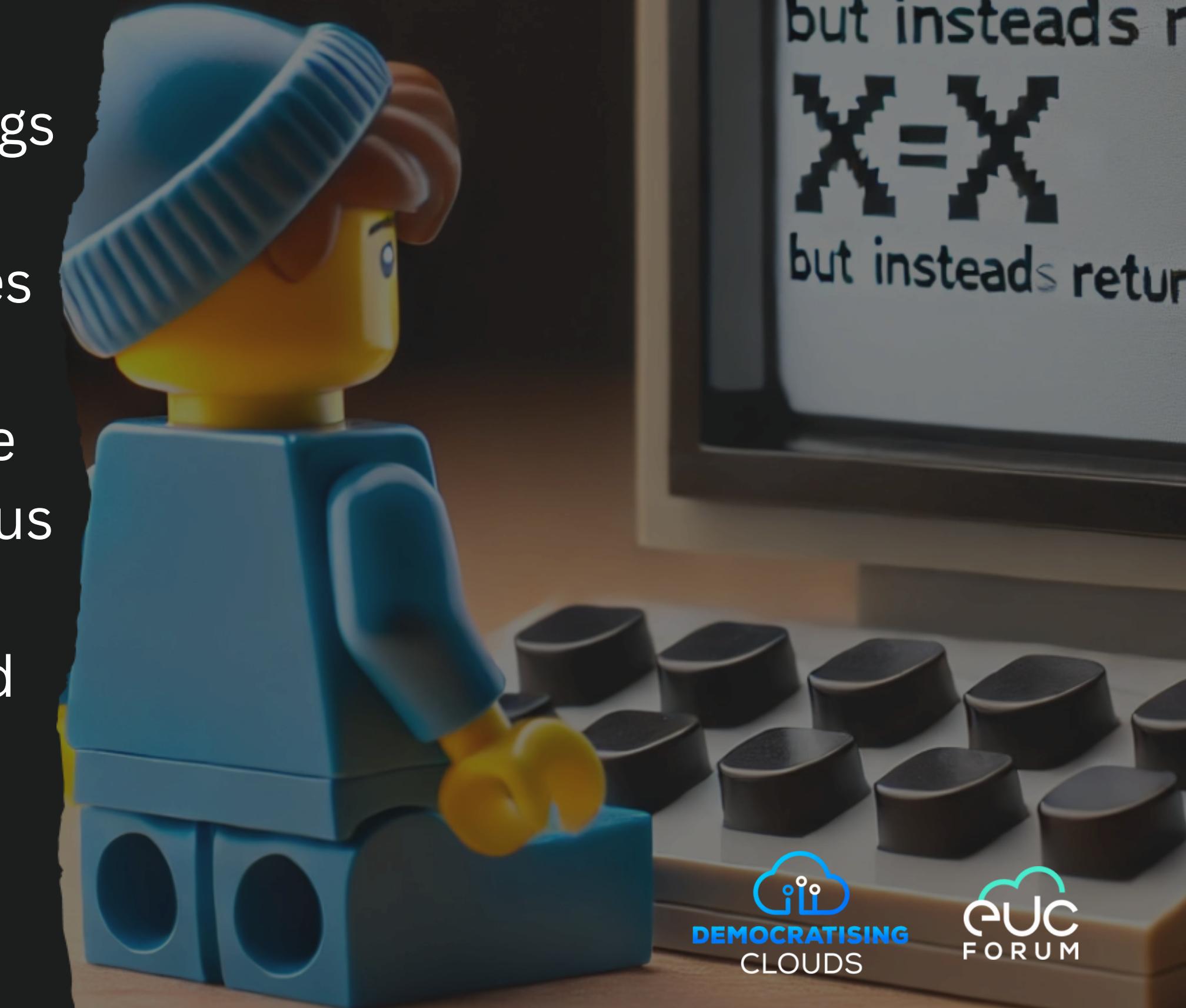
# Misunderstandings

"Example 2: "An American speaker uses the word 'football,' referring to American football, but a British speaker thinks they are talking about soccer, leading to confusion."



# Misunderstandings

Just as language misunderstandings can occur between people from different cultures, the same issues can arise when using Natural Language Processing (NLP). If the input is not clear or uses ambiguous wording, the NLP system can produce incorrect or unexpected responses.

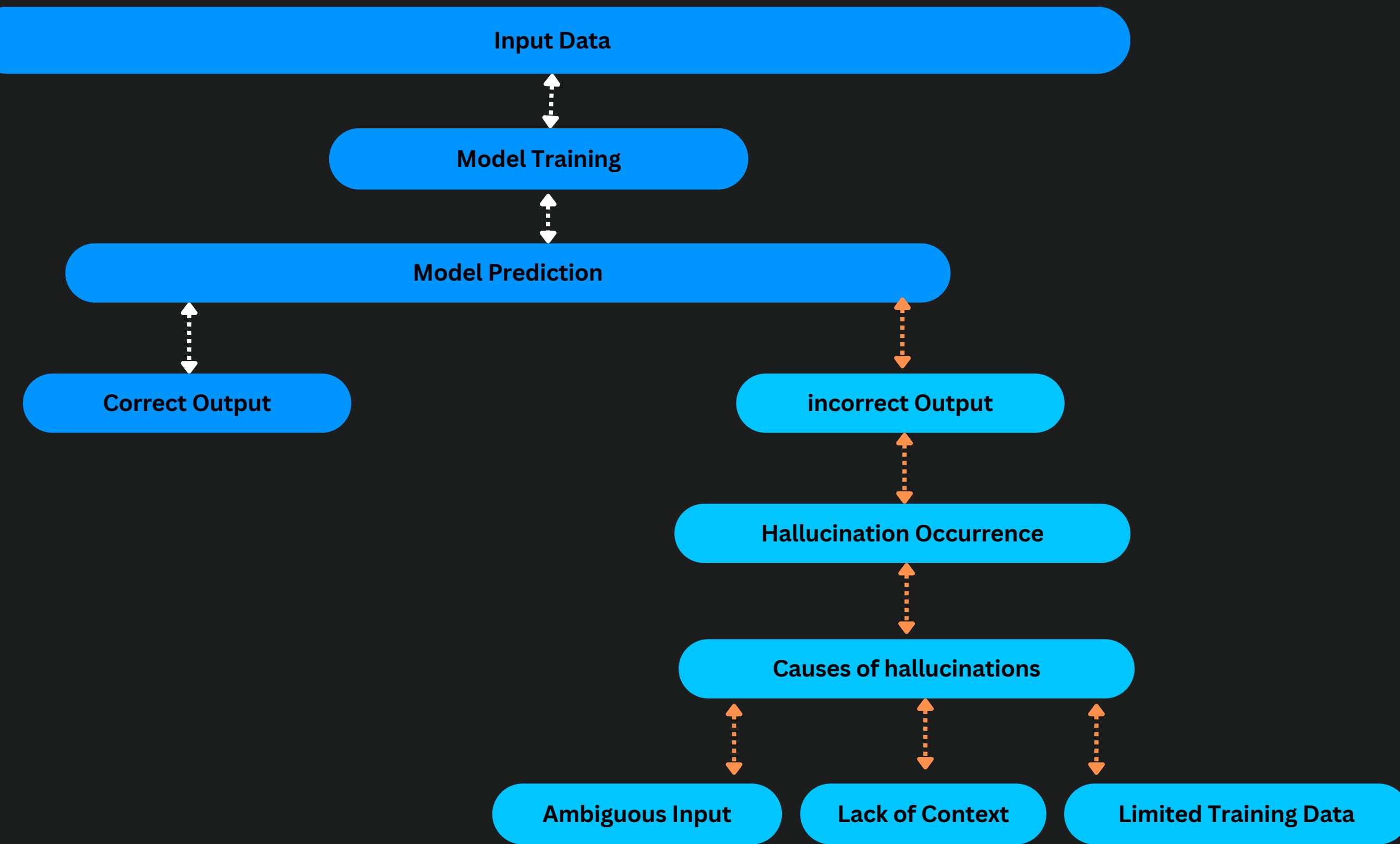


# Hallucinations and why

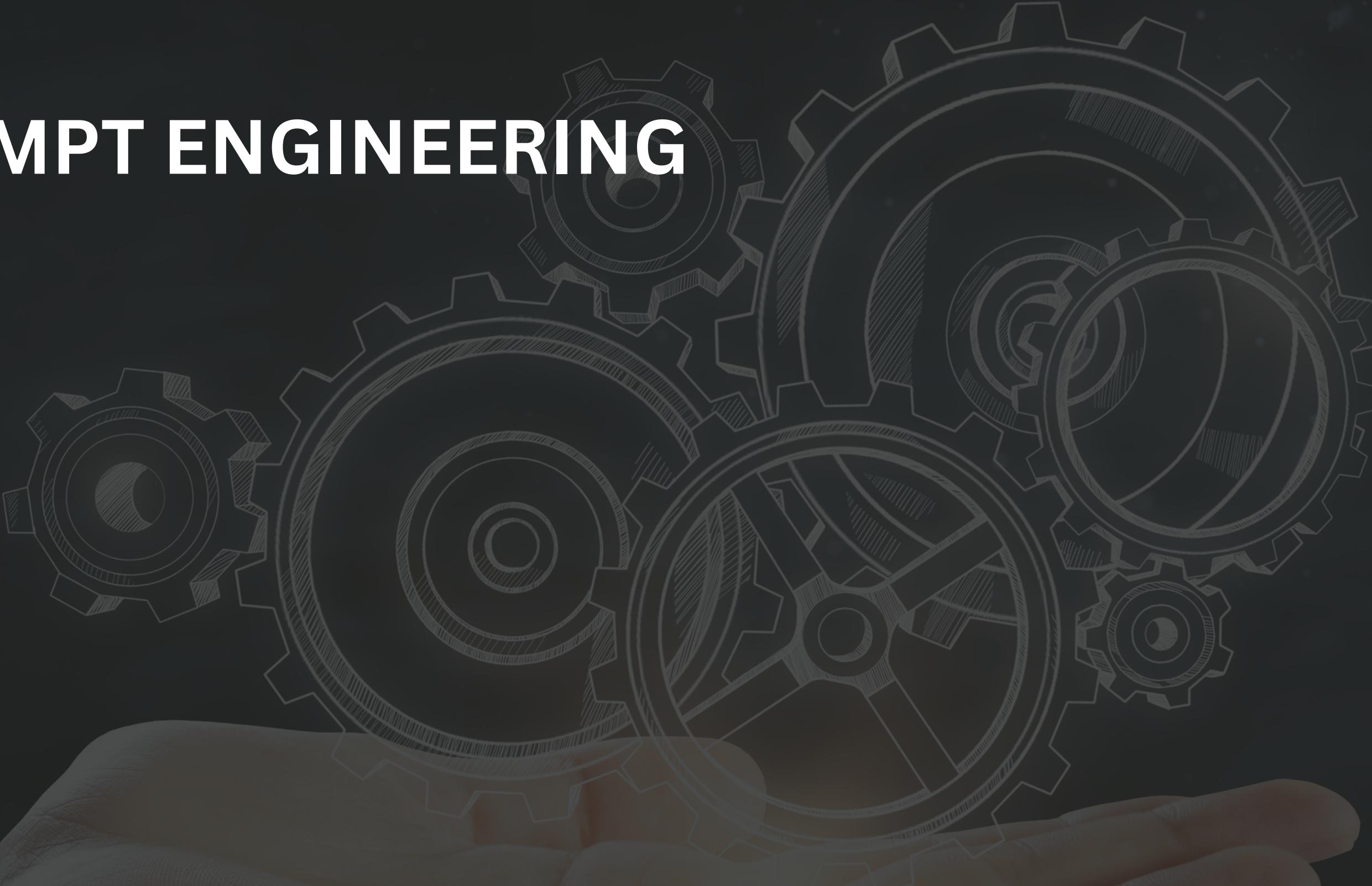
- Core limitation of Generative AI
- Predictive Nature (sequence-based)
- Training Data Limitations
- Lack of Real-World Understanding
- Ambiguity and Open-Ended Queries
- Inference Under Constraints
- Presenting out-of-date or generic information when the user expects a specific, current response.
- Creating a response from non-authoritative sources.



# Example Hallucination Flow



# PROMPT ENGINEERING



# CO-STAR PROMPT FRAMEWORK

- C: Context: Provide background and information on the task
- O: Objective: Define the task that you want the LLM to perform
- S: Style: Specify the writing style you want the LLM to use
- T: Tone: Set the attitude and tone of the response
- A: Audience: Identify who the response is for
- R: Response: Provide the response format and style

# SYSTEM VS. USER ROLES IN AI PROMPTS

System Role

VS

User Role

"role": "system", "content": "You are an IT log analysis assistant."

"role": "user", "content": "Can you summarise the errors from the logs?"

# AI PARAMETER'S

## **max\_tokens = 1000**

Defines the maximum number of tokens (words or parts of words) the AI can generate in a response.  
Larger values allow longer responses.

## **temperature = 0.3**

Controls the creativity of the AI.  
Lower values make it more deterministic and focused, while higher values make it more random.

## **top\_p = 0.95**

Controls the probability mass considered in the response.  
0.95 means the AI considers the top 95% most likely responses.

## **frequency\_penalty = 0**

Controls how much the AI penalises repetition.  
0 means no penalty; higher values reduce repetition.

## **presence\_penalty = 0**

Encourages or discourages introducing new topics.  
0 means no penalty; higher values discourage new topics.

# EXAMPLE

```
$body = @{
    "messages" = @(
        @{
            "role" = "system"
            "content" = "You are an IT log analysis assistant. Summarise the provided log analysis summaries and provide concise, to-the-point key insights and conclusions."
        },
        @{
            "role" = "user"
            "content" = $prompt
        }
    )
    "max_tokens" = 1000          # Defines the maximum number of tokens (words or parts of words) the AI can generate in the response. Larger values allow longer responses.

    "temperature" = 0.3         # Controls the "creativity" of the AI. Lower values (closer to 0) make the output more deterministic and focused, while higher values make it more random.

    "top_p" = 0.95             # This is an alternative to temperature. It controls how much of the probability mass is considered when generating the response. 0.95 means the AI will consider 95% of the highest probable choices.

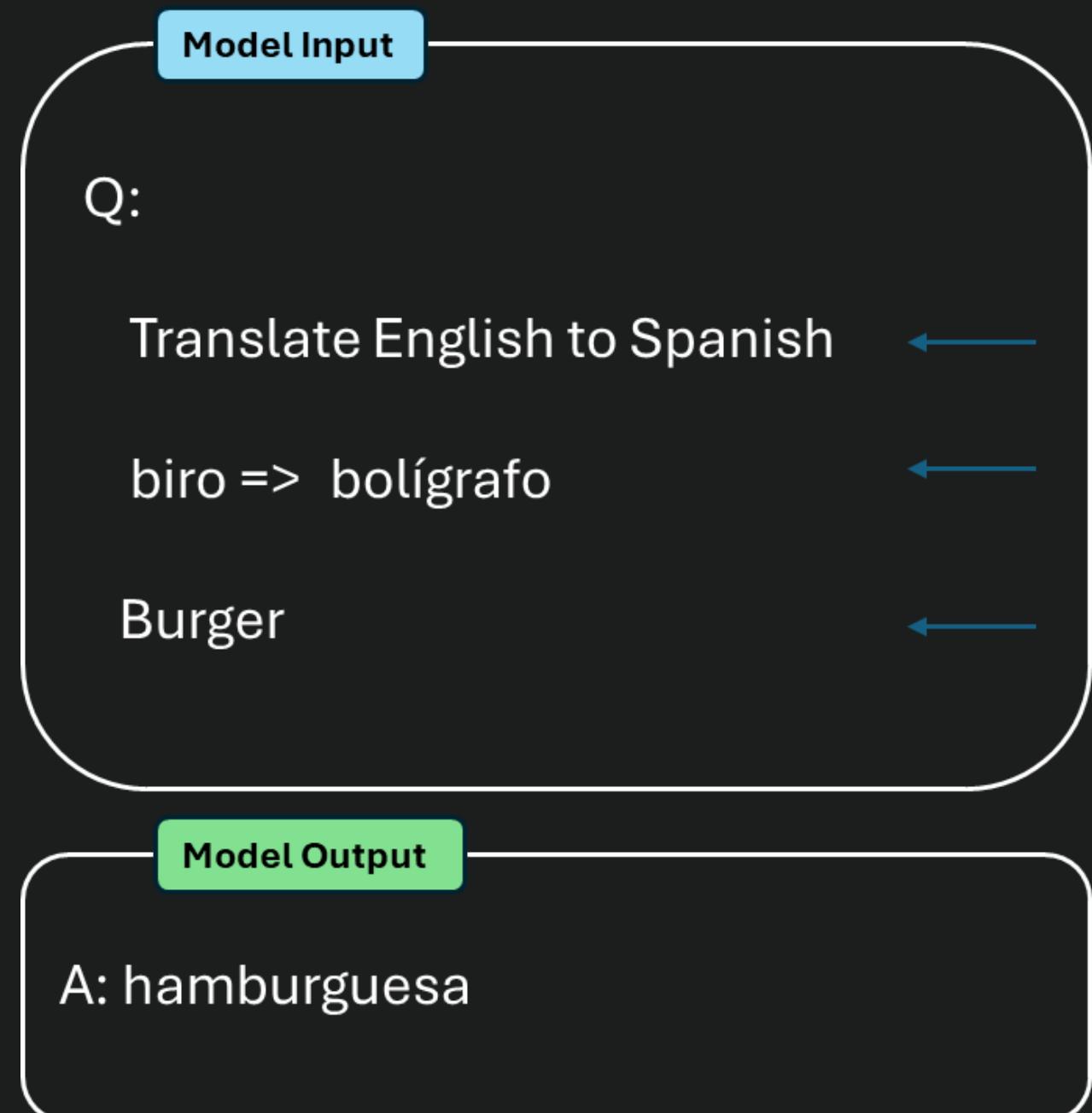
    "frequency_penalty" = 0     # Controls how much the AI should penalize repeated phrases or words. 0 means no penalty; higher values reduce repetition.

    "presence_penalty" = 0      # Encourages or discourages the model from talking about new topics not already in the text. 0 means no penalty, higher values make it more likely to talk about new topics.

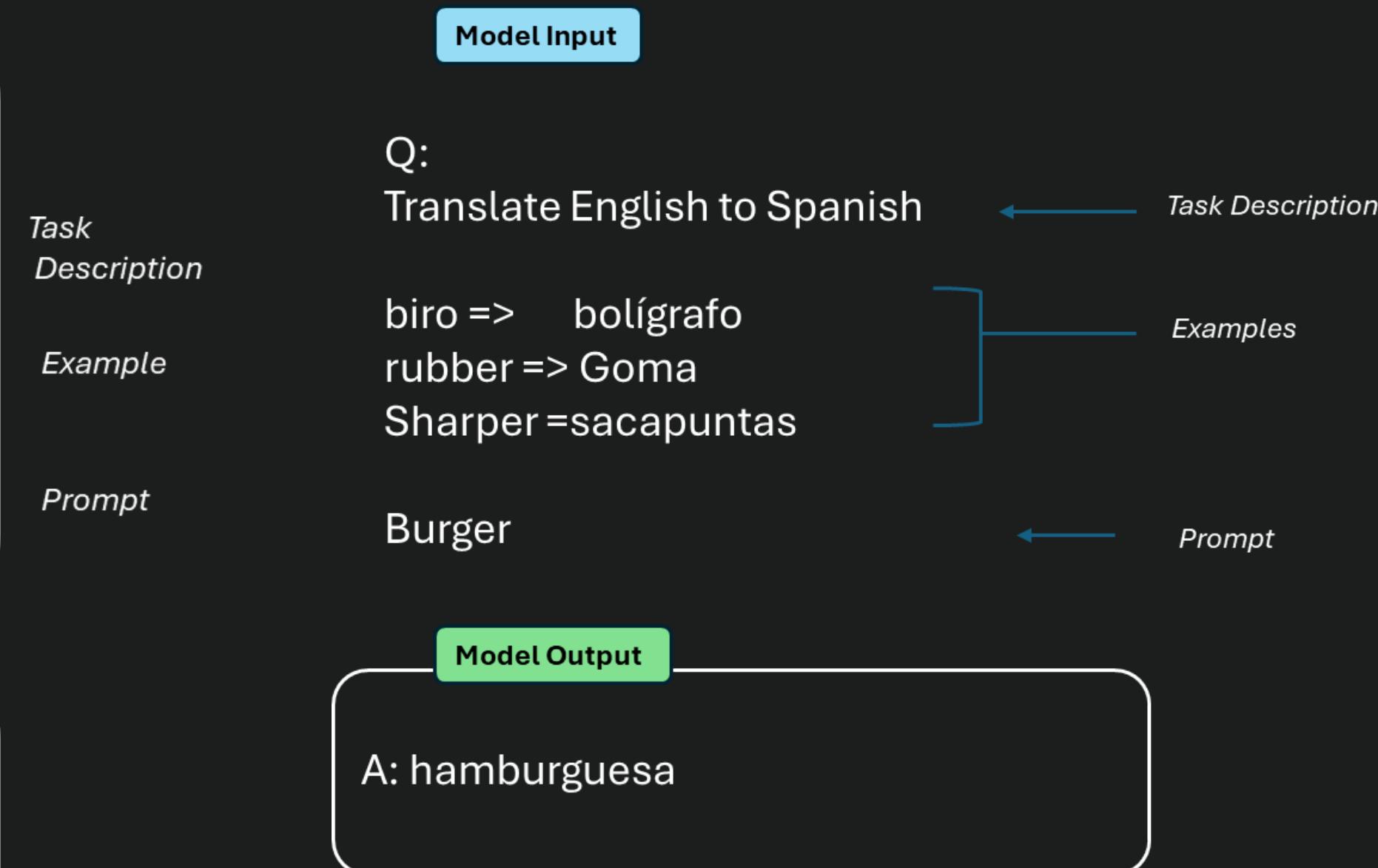
} | ConvertTo-Json -Depth 4    # Converts the body into a JSON format with a depth level of 4, allowing for nested structures to be preserved.
```

# PROMPT EXAMPLES

## One Shot



## Few Shot



## Standard Prompting

Model Input

Q: Jane has 8 flowers. She buys 3 more bouquets of flowers. Each bouquet has 4 flowers. How many flowers does she have now?

A: The answer is 20.

Q: There are 40 students in a class. If 12 students go on a field trip and 8 new students join the class, how many students are in the class now?

Model Output

A: The answer is 34.



## Chain-of-Thought Prompting

Model Input

Q: Jane has 8 flowers. She buys 3 more bouquets of flowers. Each bouquet has 4 flowers. How many flowers does she have now?

A: Jane started with 8 flowers. 3 bouquets of 4 flowers each is 12 flowers.  $8 + 12 = 20$ . The answer is 20.

Q: There are 40 students in a class. If 12 students go on a field trip and 8 new students join the class, how many students are in the class now?

Model Output

A: The class had 40 students originally. 12 went on a field trip. So they had  $40 - 12 = 28$  students. 8 new students joined the class, so they have  $28 + 8 = 36$ . The answer is 36.





# USE CASES

# Microsoft Copilot

The image displays two side-by-side screenshots of the Microsoft Copilot application interface, illustrating its conversational capabilities.

**Screenshot 1 (Left):** A user asks, "Can you create a powershell script to list all processes running on my client?". Copilot responds with a PowerShell script:

```
# This script lists all running processes on the client  
Get-Process | Select-Object Name, Id, CPU, WorkingSet, StartTime
```

Copilot also provides instructions on how to run the script:

1. Open PowerShell as an administrator:
  - Click on the Start menu, type PowerShell, right-click on Windows PowerShell, and select "Run as administrator".
2. Copy and paste the script into the PowerShell window.
3. Press Enter to execute the script.

This script will display the process name, process ID, CPU usage, memory usage (WorkingSet), and start time for each running process.

**Screenshot 2 (Right):** A user asks, "How do i review running processes in windows 11.". Copilot responds with a general guide:

To review running processes in Windows 11, you have a few options. Here are the most common methods:

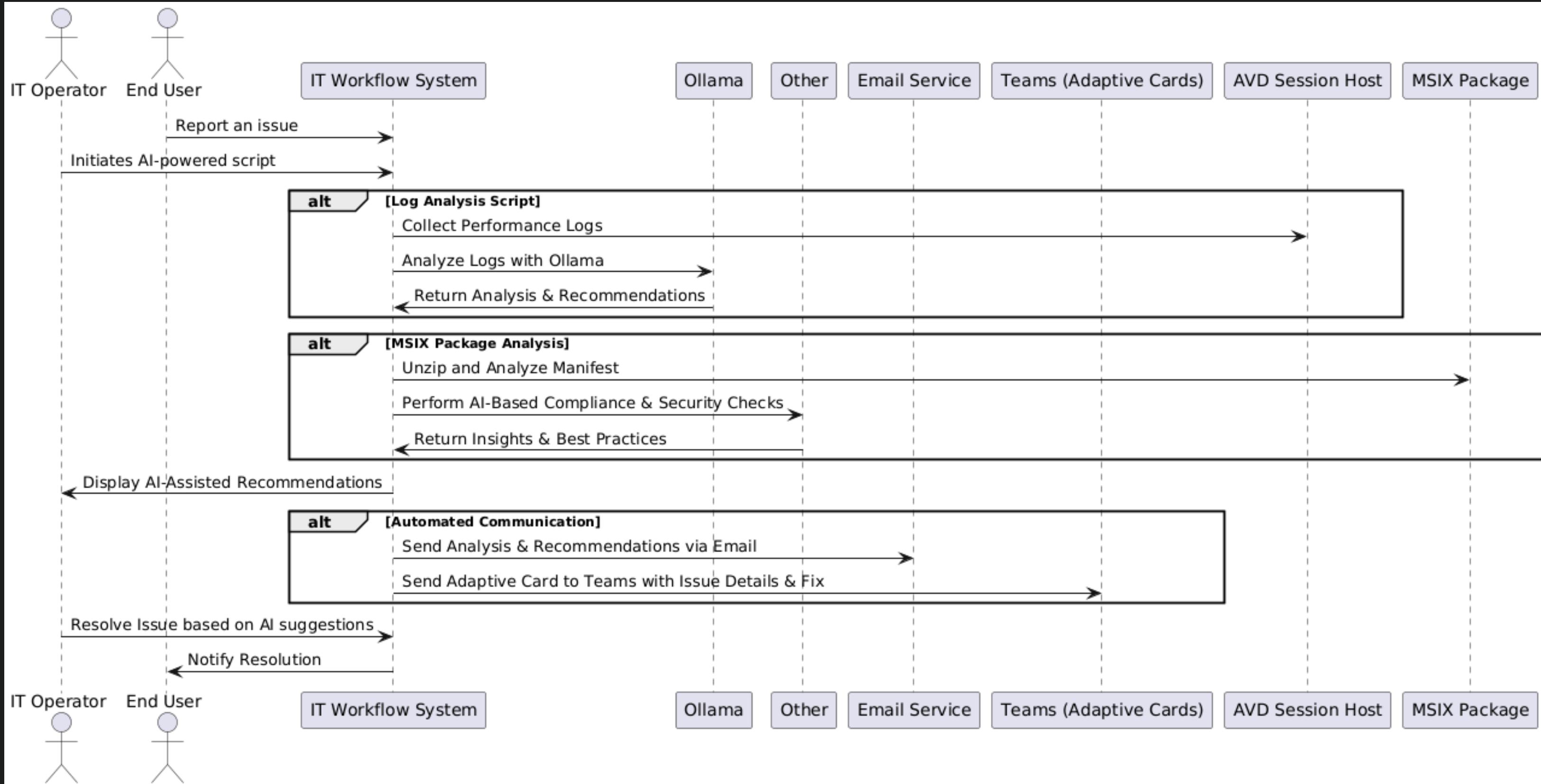
- Using Task Manager**
  1. Press Ctrl + Shift + Esc to open Task Manager directly.
  2. Click on the "Processes" tab to see a list of all running processes.
  3. You can sort the processes by CPU, Memory, Disk, or Network usage by clicking on the respective column headers.
  4. To end a process, select it and click "End task".
- Using Command Prompt**
  1. Open Command Prompt as an administrator:
    - Click on the Start menu, type cmd, right-click on Command Prompt, and select "Run as administrator".
  2. Type `tasklist` and press Enter. This will display a list of all running processes along with their Process IDs (PIDs) and memory usage.
- Using PowerShell**
  1. Open PowerShell as an administrator:
    - Click on the Start menu, type PowerShell, right-click on Windows PowerShell, and select "Run as administrator".

Below the main content, there are three additional prompts:

- How to end a process safely?
- What is the difference between processes and services?
- Can I see resource usage for each process?

# TECH STACK - FOR BUILDING TOOLING





# AI-ASSISTED LOG ANALYSIS (FSLOGIX EXAMPLE)

## Challenges:

- Manual Process: Time-consuming, error-prone.
- Complex Logs: Difficult to identify key issues quickly.

```
|FSLogix Service Version 2.9.8884.27471 Status: RUNNING  
FSLogix Kernel Driver Version 2.9.8884.27471 Status: RUNNING  
FSLogix Kernel Virtualization Driver Version 2.9.8884.27471 Status: RUNNING  
-----  
[06:09:07.845][tid:0000100c.00000750][WARN: 00000003] Copy redirections.xml failure (The  
system cannot find the path specified.)
```

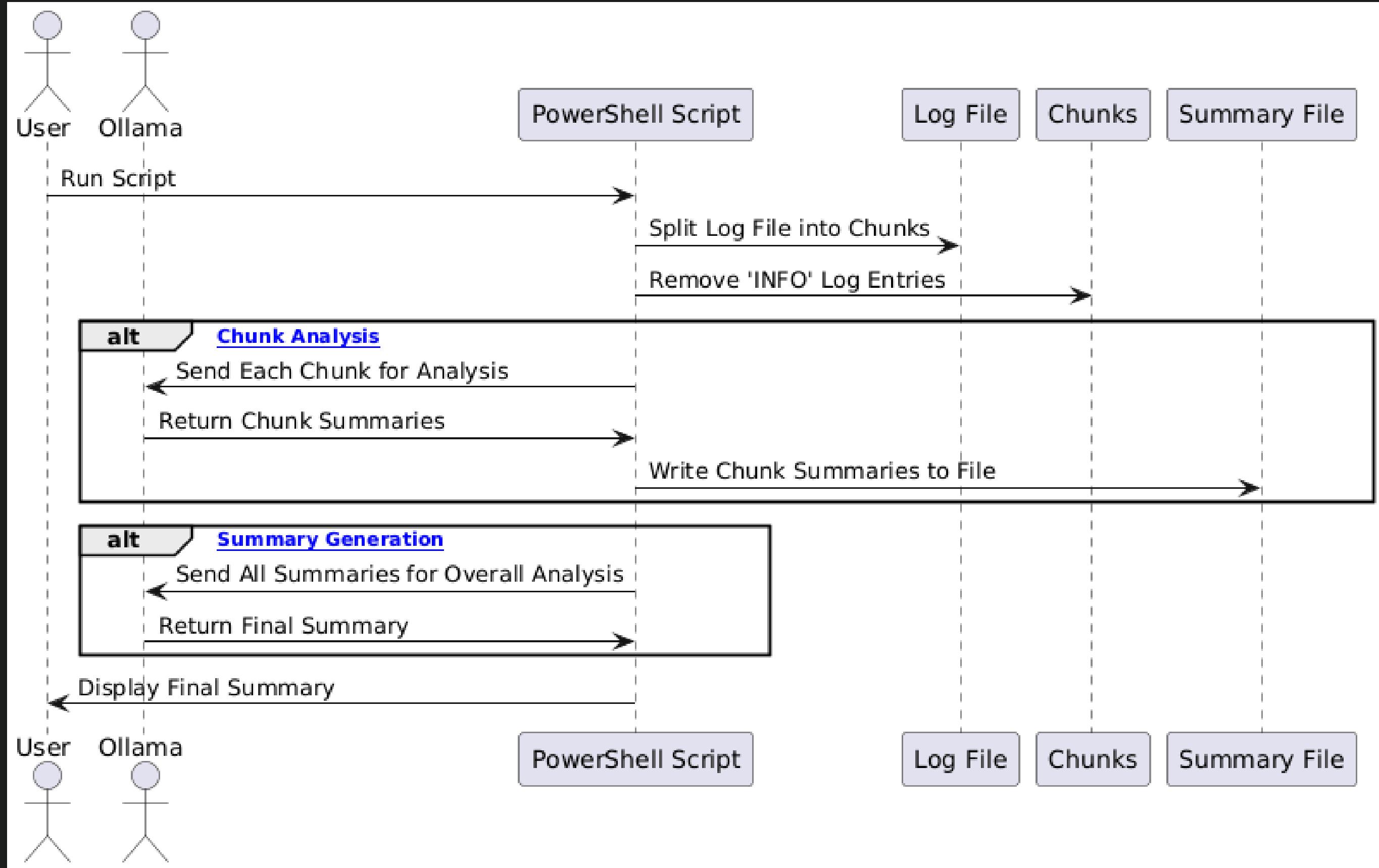
## AI Approach:

- Automated Log Categorisation: Filters errors, warnings, and info.
- Summarised Insights: Provides quick, concise summaries.
- Actionable Recommendations: AI suggests fixes based on log data.

## Outcome:

- Faster Resolution: Quicker identification of root causes.
- Efficiency Gains: Reduces manual effort and human error.
- Improved Troubleshooting: Focuses on resolving critical issues.

Earlier this week				
 log_chunk_0	20/09/2024 18:18	Text Document	1 KB	
 log_chunk_1	20/09/2024 18:18	Text Document	1 KB	
 log_chunk_2	20/09/2024 18:18	Text Document	1 KB	
 log_chunk_3	20/09/2024 18:18	Text Document	1 KB	



# APP PACKAGE AUDITING WITH AI

## Challenges with MSIX Packages:

- Verifying package configurations before deployment is critical to avoid misconfigurations, limitations, security configurations, etc.
- Manually checking package dependencies, capabilities, and compliance can be time-consuming.

The screenshot shows the Windows 365.exe Properties dialog box. The 'Details' tab is selected. The following table summarizes the file information:

Property	Value
Description	
File description	
Type	Application
File version	
Product name	
Product version	
Copyright	
Size	27,6 MB
Date modified	20-09-24 11:33
Language	

Ryan Mangan · Sep 20, 2024

@RyMangan · Follow

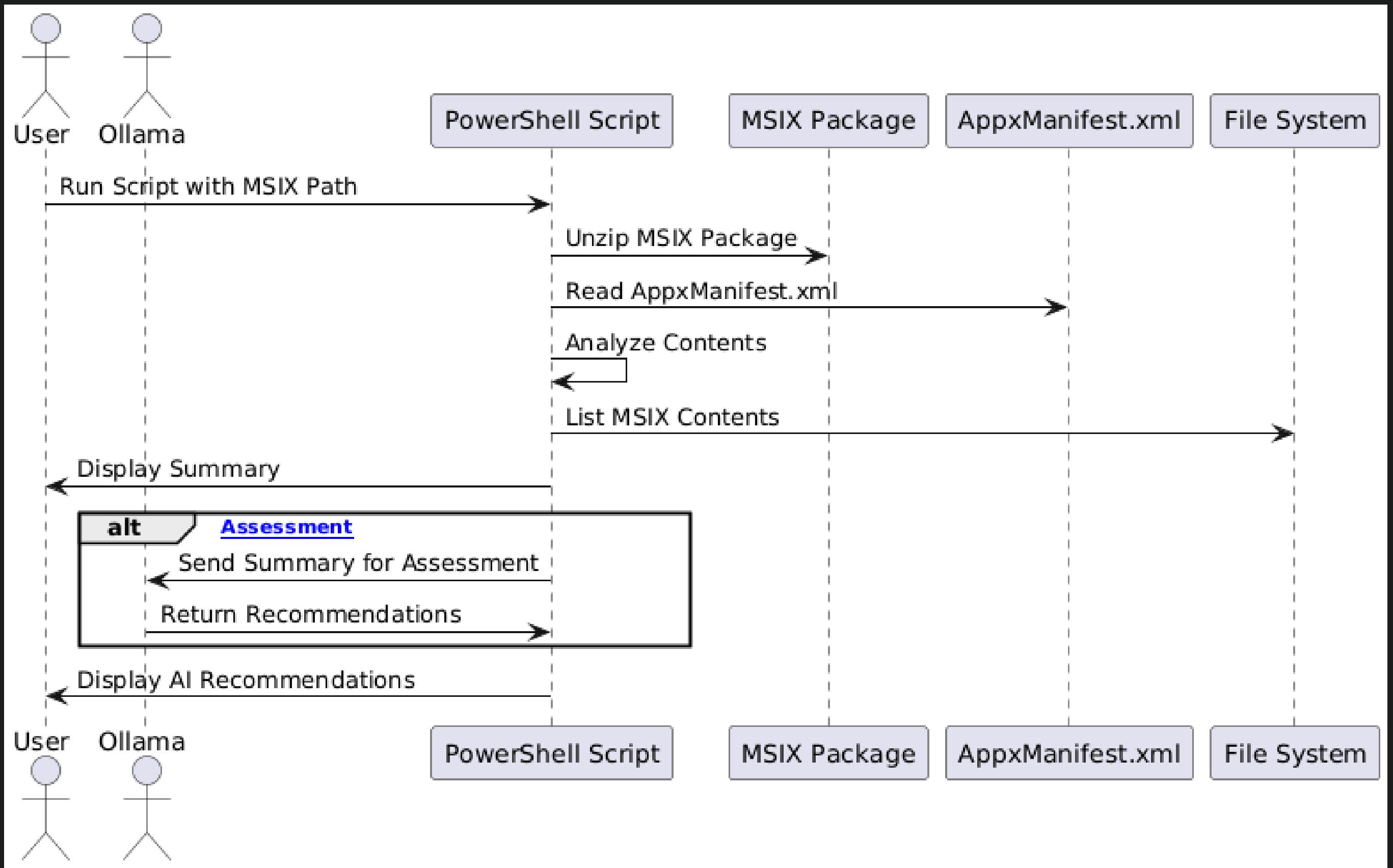
Replying to @RemkoWeijnen @CyberCakeX and 2 others

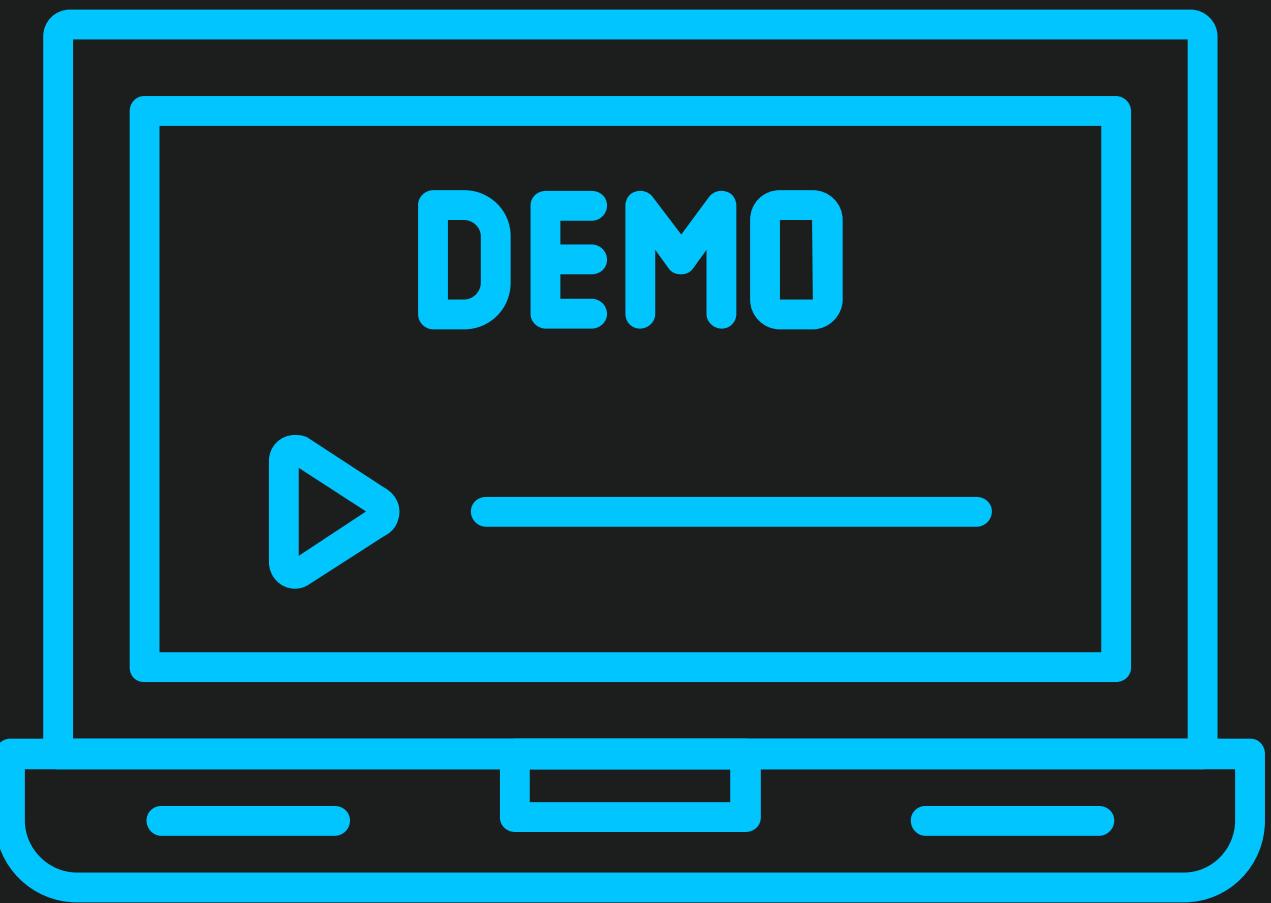
Do you have other examples of the same ?

Guy Leech · Follow

Of the 92 different processes running in my Windows 11 session, only 4 of the exes are not signed, 2 of which are 3rd party & the other unsigned @Microsoft exe is WidgetService.exe but at least it has a version resource even if it is v0.5 on 23H2 🤪

```
name | Get-AuthenticodeSignature -FilePath $_.Group[0].Path |where {$_select -ExpandProperty VersionInfo -Unique  
022\crashpad_handler.exe  
err to parameter 'FilePath' because it is null.  
$Path $_.Group[0].Path |where ...  
} |Get-AuthenticodeSignature, ParameterBindingValidationException  
ValidationExceptionNullNotAllowed, Microsoft.PowerShell.Commands.GetAuth  
0A\WhatsAppDesktop_2.2436.6.0_x64\cvlg1gvny\g\WhatsApp.exe  
soft.WidgetsPlatformRuntime_1.5.0.0_x64\_0ekyb3d5bbe\WidgetService  
oftCorporation\Windows365_2.0.205.0_x64\_0ekyb3d5bbe\Windows365  
name|measure).Count  
files\windowsapps\\*\group path|sort path  
  
istics.Process (Widgets))  
istics.Process (WhatsApp))  
istics.Process (pwsh), System.Diagnostics.Process (pwsh)  
istics.Process (WindowsTerminal), System.Diagnostics.Process (Windows  
istics.Process (Windows365))  
istics.Process (WidgetService), System.Diagnostics.Process (WidgetSer  
istics.Process (CloudOS))  
istics.Process (CrossDeviceService))  
istics.Process (ApplePhotaStreams))  
istics.Process (PhoneExperienceHost))  
istics.Process (OpenConsole), System.Diagnostics.Process (OpenConsole  
istics.Process (ms-teams), System.Diagnostics.Process (ms-teams), Sys  
Copyright. All rights reserved.  
Microsoft Corporation. All rights reserved.
```





# REAL-TIME PERFORMANCE DATA COLLECTION WITH AI

## Challenges with Performance Monitoring:

- Ensuring optimal performance of virtual machines (VMs) is crucial to avoid resource bottlenecks, system crashes, and degraded user experience.
- Manually monitoring CPU, memory, disk, and network usage across multiple VMs can be time-consuming and prone to errors.

## Outcome:

- Real-Time Monitoring: AI provides real-time insights into VM performance, detecting anomalies and potential issues and advising on next steps.
- Proactive Resource Optimisation: AI recommendations help optimise resource allocation, preventing system overloads and ensuring efficient operation.

Sending performance data to Azure OpenAI for analysis...

#### AI-Assisted Performance Analysis and Recommendations:

Based on the provided performance metrics from the Azure Virtual Desktop (AVD) session host, here is a detailed analysis and recommendations:

#### ### Analysis:

##### 1. \*\*Network Performance:\*\*

- \*\*Realtek PCIe 2.5GbE Family Controller:\*\*
  - \*\*Bytes Received/Sec:\*\* 517,309.12
  - \*\*Bytes Sent/Sec:\*\* 15,821.84
  - \*\*Packets Received/sec:\*\* 366.27
  - \*\*Packets Sent/Sec:\*\* 202.65
- \*\*Intel Wi-Fi 6 AX200 160MHz:\*\*
  - \*\*Bytes Sent/Sec:\*\* 0
  - \*\*Bytes Received/Sec:\*\* 0
  - \*\*Packets Sent/Sec:\*\* 0
  - \*\*Packets Received/Sec:\*\* 0

\*\*Observation:\*\* The Realtek network interface is actively handling network traffic, while the Intel Wi-Fi interface shows no activity. This indicates that the system is likely using a wired connection exclusively.

##### 2. \*\*Memory Usage:\*\*

- \*\*Available MBytes:\*\* 91,054.12
- \*\*% Committed Bytes in Use:\*\* 34.23

\*\*Observation:\*\* Memory availability is high, and the percentage of committed bytes in use is low, indicating no memory pressure.

##### 3. \*\*CPU Usage:\*\*

- \*\*% Idle Time:\*\* 93.09
- \*\*% Processor Time:\*\* 0.79

\*\*Observation:\*\* The CPU is largely idle, suggesting that the system is not under CPU stress.

##### 4. \*\*Disk Usage:\*\*

- \*\*Disk Read Bytes/Sec:\*\* 9,670.66
- \*\*Disk Write Bytes/Sec:\*\* 1,860,493.6

\*\*Observation:\*\* There is a significant amount of disk write activity compared to read activity. This could indicate logging, caching, or other write-heavy operations.

#### ### Potential Issues and Causes:

##### 1. \*\*Network Configuration:\*\*

- The lack of activity on the Intel Wi-Fi interface suggests that it may not be configured or utilized. This could be intentional if the environment prefers wired connections for stability and performance.

##### 2. \*\*Disk Write Activity:\*\*

- High disk write activity could lead to performance degradation over time if not managed properly. This might be due to applications with high logging or caching requirements.

#### ### Recommendations:

##### 1. \*\*Network Optimization:\*\*

- Ensure that the Realtek network interface is configured

Performance log collection and analysis completed.

PS C:\WINDOWS\system32>



# OTHER EXAMPLES

- Kusto log extraction and inspection.
- Risk-based alerting
- Recurring incident/anomaly identification.
- Application compatibility assessment E.g MSIX App Attach
- Knowledge repository.
- Many More

# QUESTIONS





2ND EDITION

# Mastering Azure Virtual Desktop

A practical guide to designing, implementing, and managing Azure Virtual Desktop environments



RYAN MANGAN | NEIL MCLOUGHLIN | MARCEL MEURER

Foreword by Christiaan Brinkhoff,  
Principal PM, Azure Virtual Desktop and Windows 365

