THE H@CK SUMMIT

AI-Powered MITRE ATT&CK Mapping: Enhancing Cyber Threat Detection

Deniz SAKLI

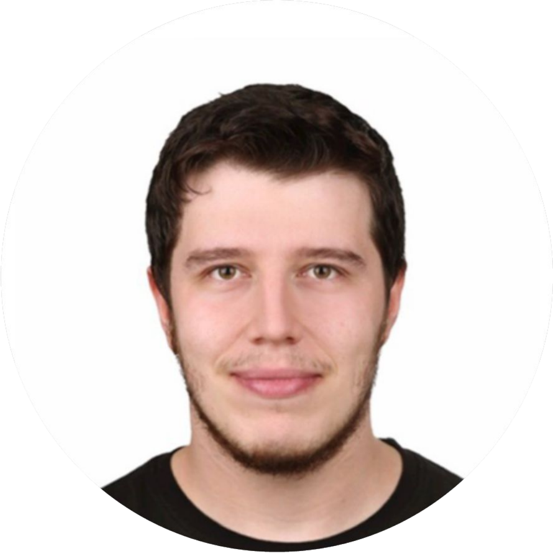
Sr. Blue Team Engineer, Picus Security

Fatih ERDOĞAN

Sr. Blue Team Engineer, Picus Security

Event organized by: Academic Partners Foundation

About us



Deniz Saklı
Senior
Blue Team Engineer

PICUS

in denizsakli

Fatih Erdoğan
Senior
Blue Team Engineer

PICUS

X FeCassie

in fatiherdogan1

Introduction

THE H@CK SUMMIT

ACADEMIC PARTNERS

MITRE ATT&CK Framework

ATT&CK®

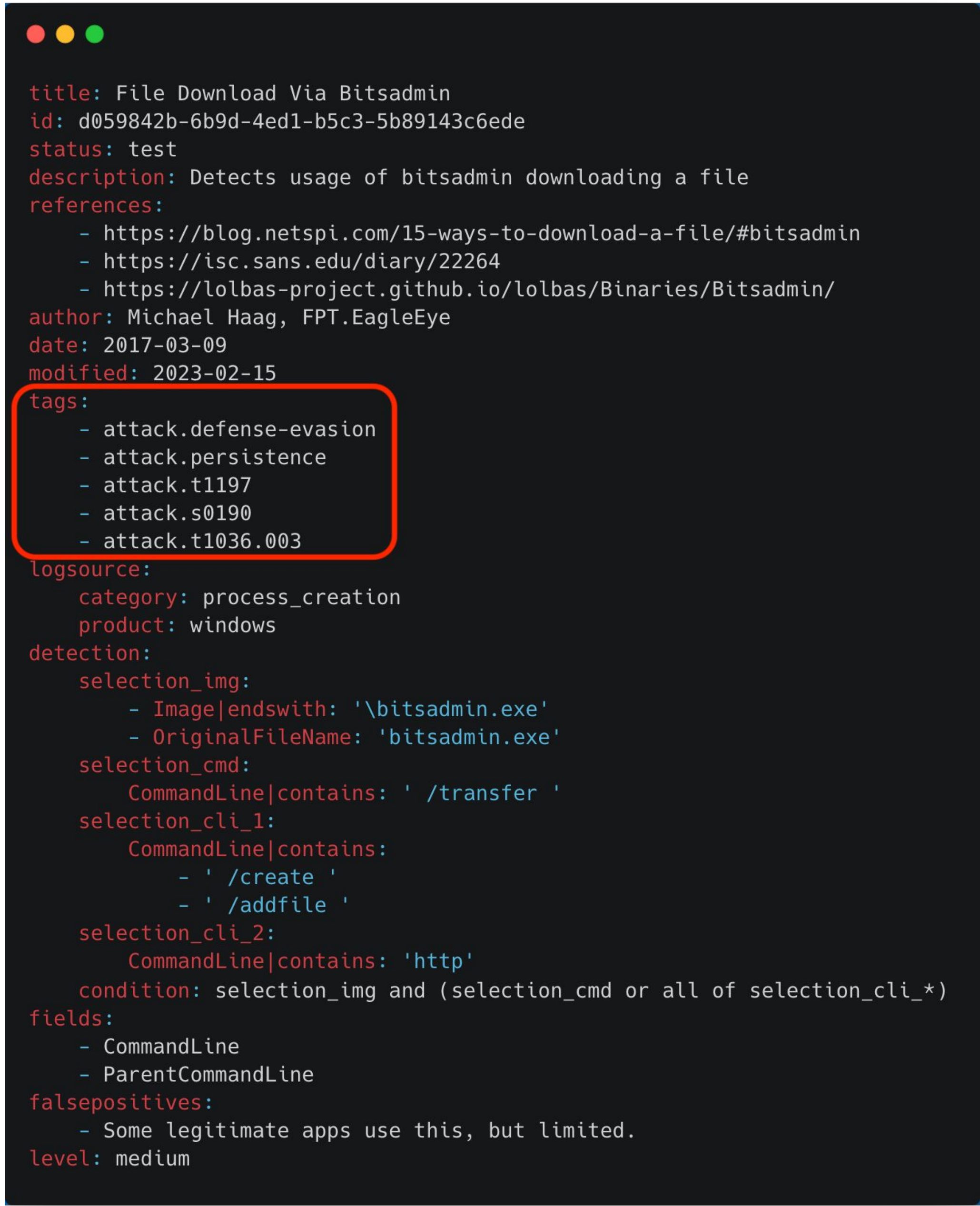
❏ https://attack.mitre.org/

Sigma Rule

- https://github.com/SigmaHQ/sigma

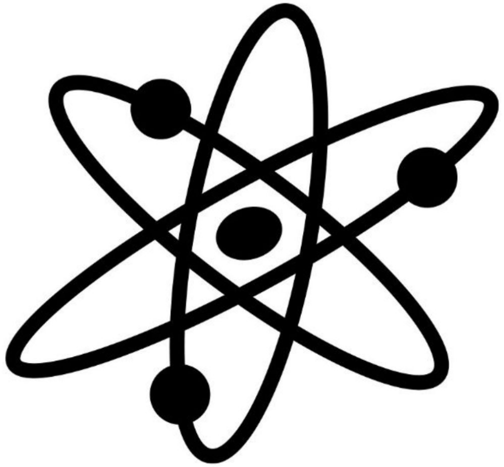
MITRE ATT&CK in Sigma Rule

- https://github.com/SigmaHQ/sigma/blob/master/rules/windows/process_creation/proc_creation_win_bitsadmin_download.yml
- Sigma rule profiling using MITRE ATT&CK
- Tactics, Techniques, related APT Groups, Softwares etc.
- Mitigations, Detections



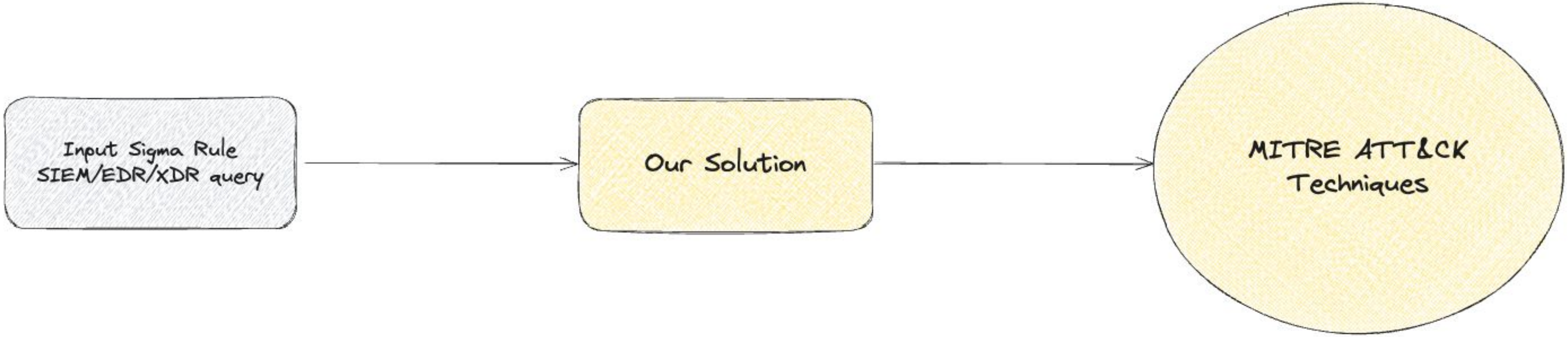
```yaml
title: File Download Via Bitsadmin
id: d059842b-6b9d-4ed1-b5c3-5b89143c6ede
status: test
description: Detects usage of bitsadmin downloading a file
references:
    - https://blog.netspi.com/15-ways-to-download-a-file/#bitsadmin
    - https://isc.sans.edu/diary/22264
    - https://lolbas-project.github.io/lolbas/Binaries/Bitsadmin/
author: Michael Haag, FPT.EagleEye
date: 2017-03-09
modified: 2023-02-15
tags:
    - attack.defense-evasion
    - attack.persistence
    - attack.t1197
    - attack.s0190
    - attack.t1036.003
logsource:
    category: process_creation
    product: windows
detection:
    selection_img:
        - Image|endswith: '\bitsadmin.exe'
        - OriginalFileName: 'bitsadmin.exe'
    selection_cmd:
        CommandLine|contains: ' /transfer '
    selection_cli_1:
        CommandLine|contains:
            - ' /create '
            - ' /addfile '
    selection_cli_2:
        CommandLine|contains: 'http'
    condition: selection_img and (selection_cmd or all of selection_cli_*)
fields:
    - CommandLine
    - ParentCommandLine
falsepositives:
    - Some legitimate apps use this, but limited.
level: medium
```

What’s the Big Bang of our Journey?

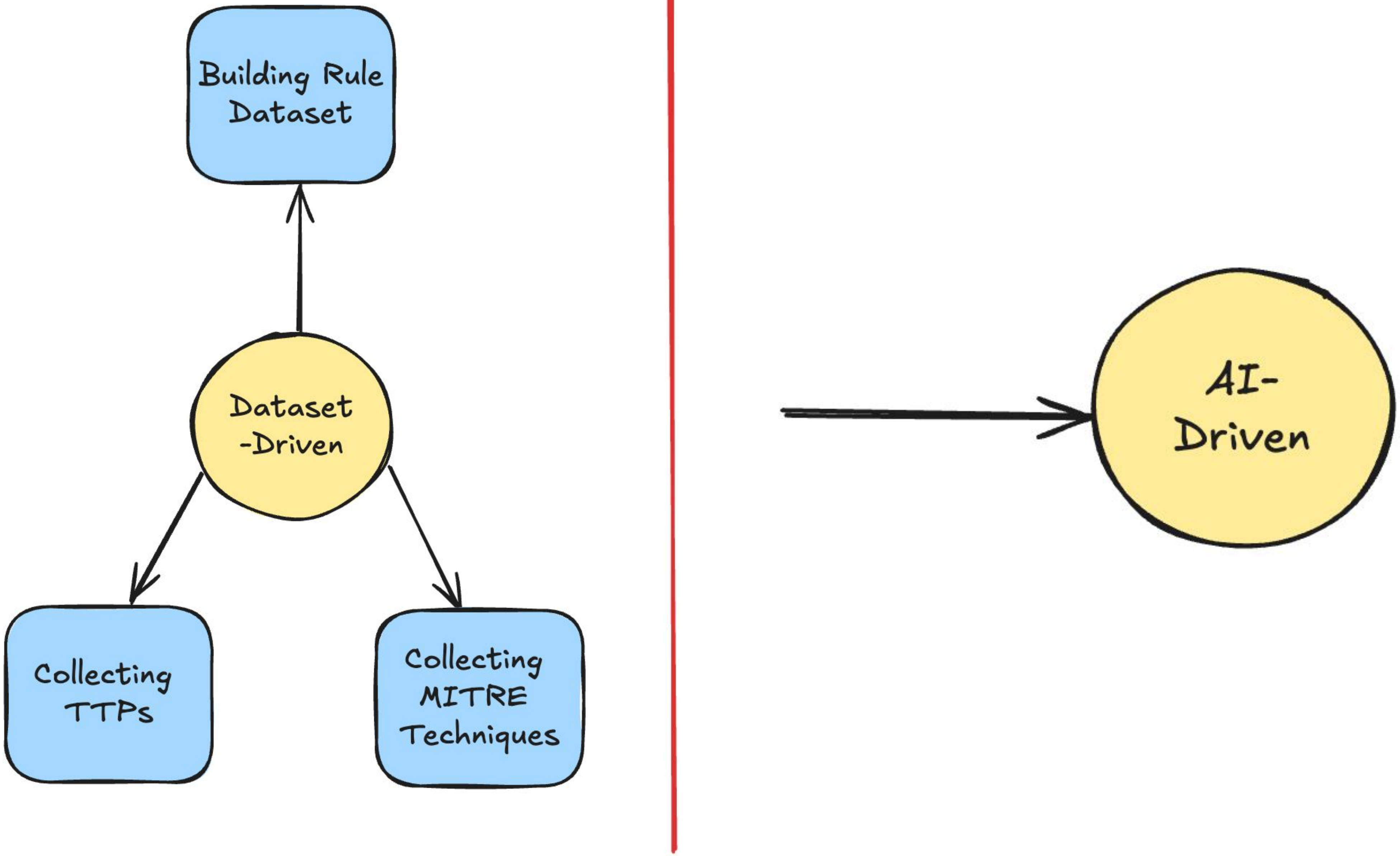


- ❑ Why we did this research?
- ❑ What was the problem?
- ❑ Detection Engineering Procces
- ❑ TTP to MITRE ATT&CK

What we aimed ??



Our Journey



Building Rule Dataset

Dataset -Driven

Collecting TTPs

Collecting MITRE Techniques

AI- Driven

Dataset-Driven MITRE Techniques Identification

THE H@CK SUMMIT

ACADEMIC PARTNERS

Finding Detection Rules

- ~3800 rules
  - Azure Sentinel
  - MITRE CAR
  - Joe Sandbox
  - Splunk Rules
  - SigmaHQ
  - Elastic Rules
  - Picus Security - Detection Rules

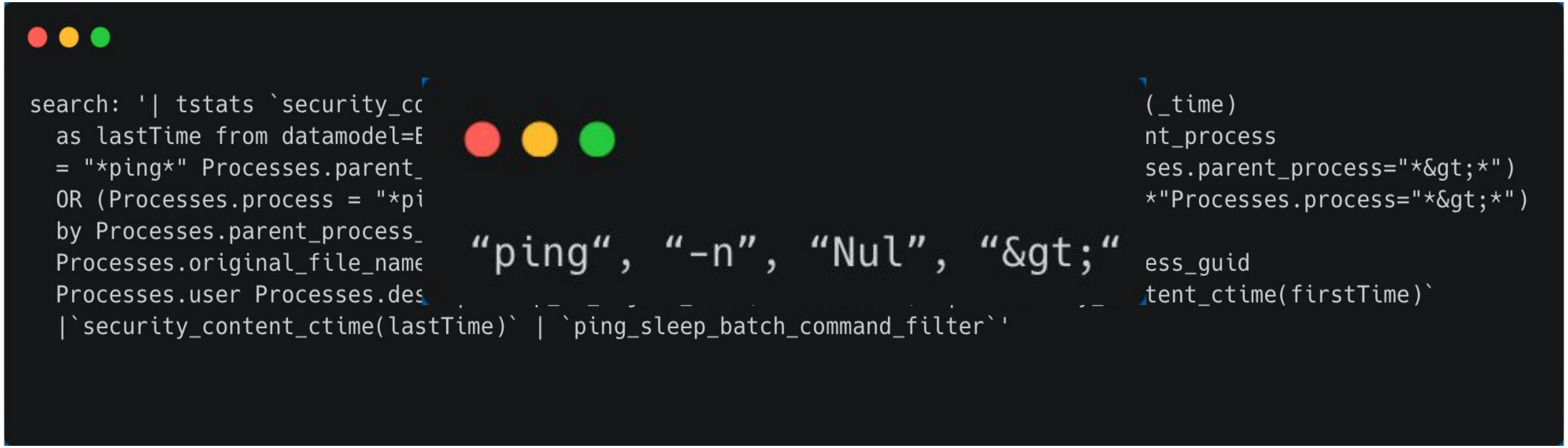
Parsing Detection Rules

- ❑ Rule normalization
- ❑ **fields** no, **values** ok!
- ❑ **field=value**, **field in value** and etc.
- ❑ and, or, not, stats, limit, dedup, count by, and etc.
- ❑ We extracted the strings by eliminating the parts in the “key=value” part, which is the general structure of the rules.

Parsing Detection Rules

Before

After

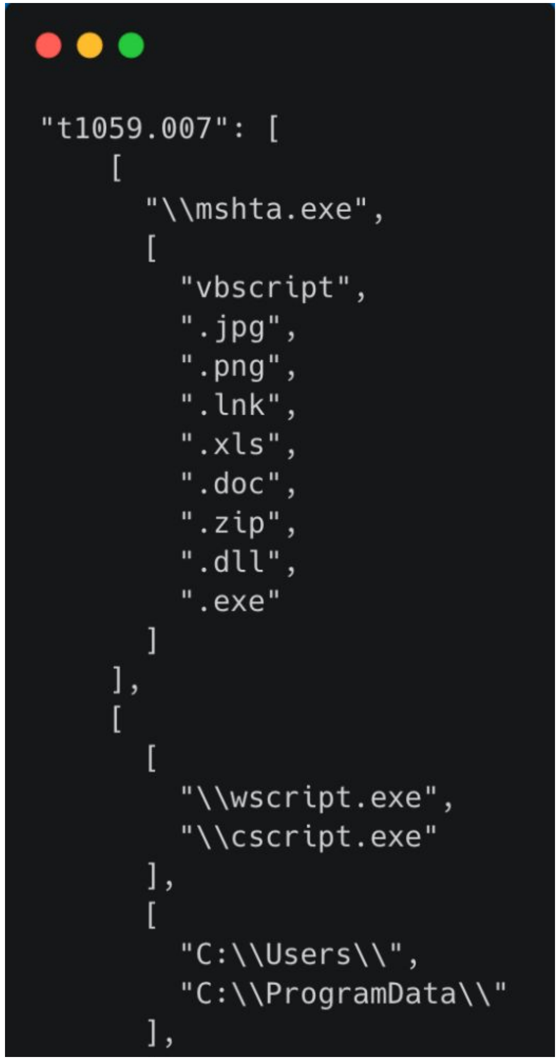


```
search: '| tstats `security_cc                                                        (_time)
  as lastTime from datamodel=E                                                    nt_process
  = "*ping*" Processes.parent_                                                    ses.parent_process="*&gt;*")
  OR (Processes.process = "*pi                                                    *"Processes.process="*&gt;*")
  by Processes.parent_process_
  Processes.original_file_name                                                    ess_guid
  Processes.user Processes.des                                                    tent_ctime(firstTime)`
  |`security_content_ctime(lastTime)` | `ping_sleep_batch_command_filter`'
```

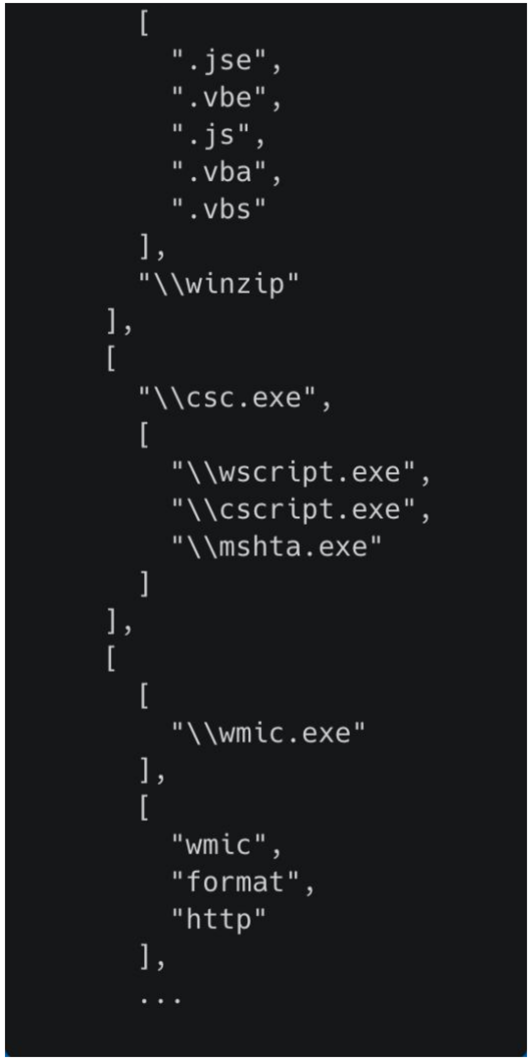
```
“ping“, “-n”, “Nul”, “&gt;“
```

Parsing Detection Rules

❑ T1059.007 - Command and Scripting Interpreter: JavaScript



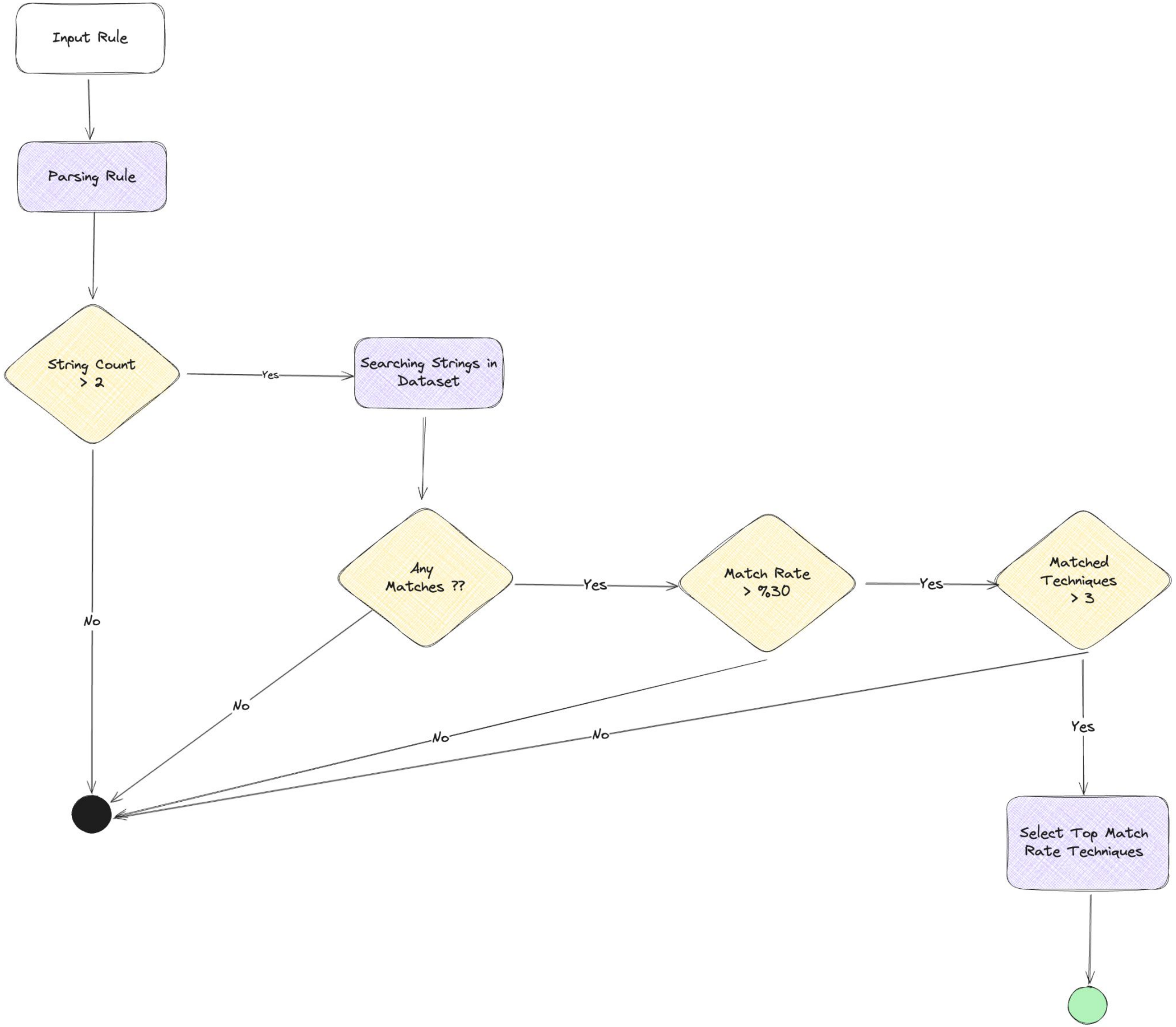
```
"t1059.007": [
    [
      "\\mshta.exe",
      [
        "vbscript",
        ".jpg",
        ".png",
        ".lnk",
        ".xls",
        ".doc",
        ".zip",
        ".dll",
        ".exe"
      ]
    ],
    [
      [
        "\\wscript.exe",
        "\\cscript.exe"
      ],
      [
        "C:\\Users\\",
        "C:\\ProgramData\\"
      ],
```



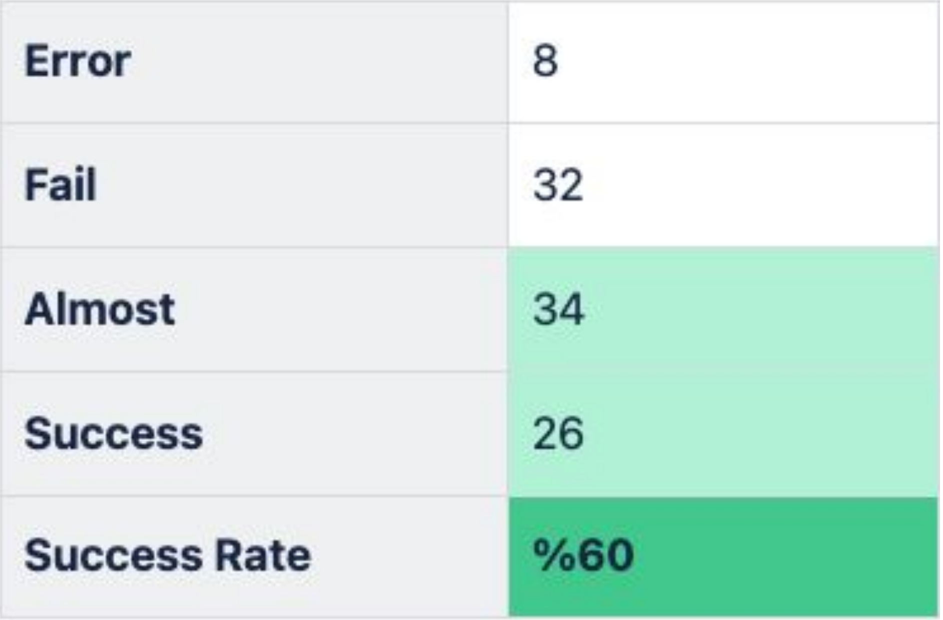
```
      [
        ".jse",
        ".vbe",
        ".js",
        ".vba",
        ".vbs"
      ],
      "\\winzip"
    ],
    [
      "\\csc.exe",
      [
        "\\wscript.exe",
        "\\cscript.exe",
        "\\mshta.exe"
      ]
    ],
    [
      [
        "\\wmic.exe"
      ],
      [
        "wmic",
        "format",
        "http"
      ],
      ...
```

Dataset Analysis

- ❑ **Total MITRE Techniques: 566**
- ❑ **Coverage Count**
  - ❑ 334
- ❑ **Not-Coverage Count**
  - ❑ 232
- ❑ **Coverage Percentage**
  - ❑ ~60%
- ❑ ~31000 Total Strings
  - ❑ after FP elimination: 23846 query strings



Results

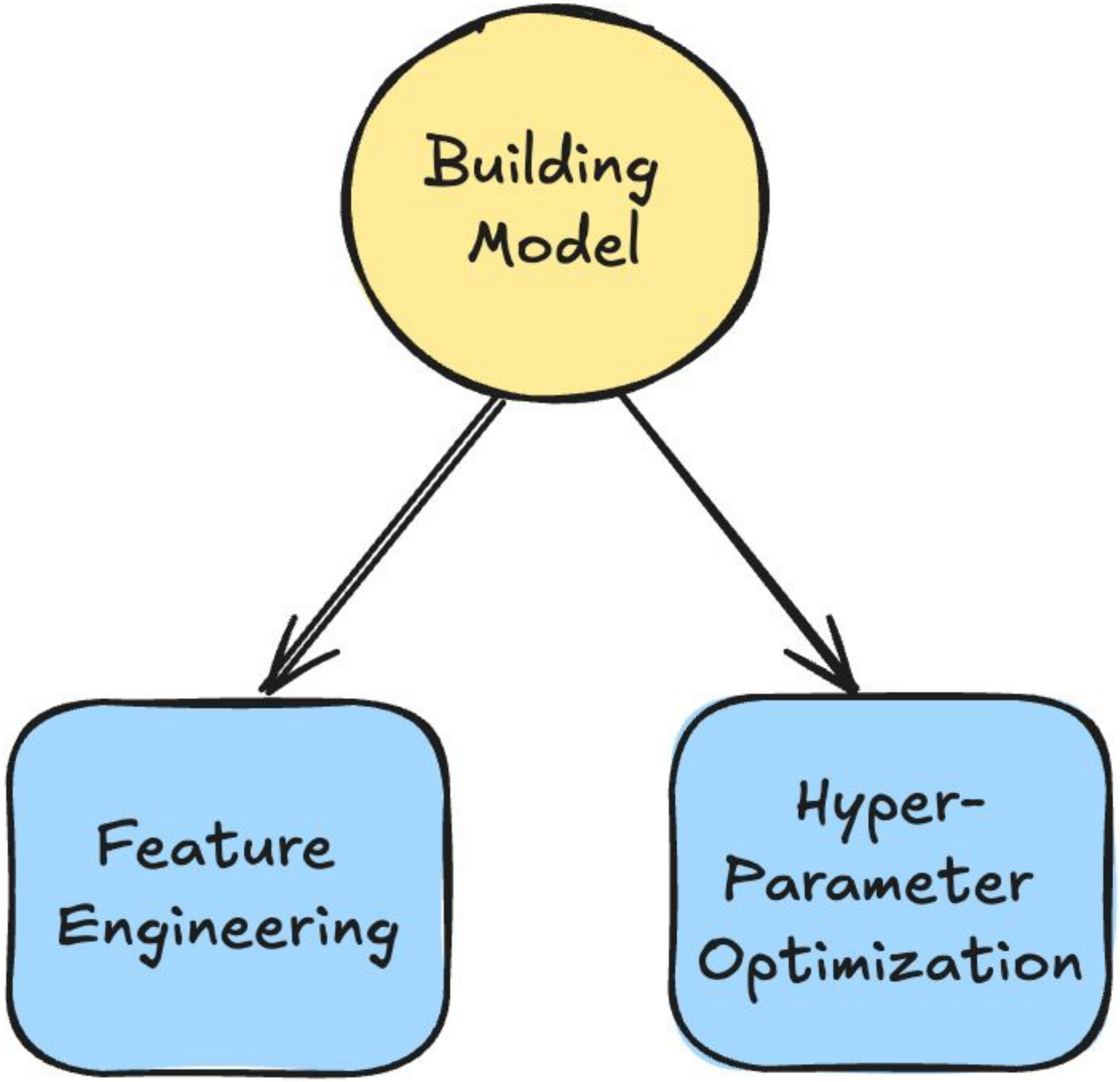


Error	8
Fail	32
Almost	34
Success	26
Success Rate	**%60**

Building MITRE AI Model

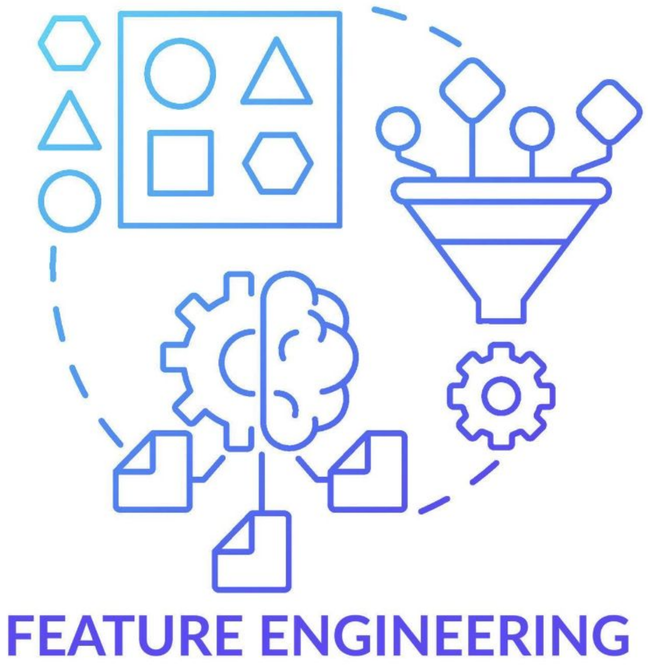
THE H@CK SUMMIT

ACADEMIC PARTNERS


Building Model
Feature Engineering
Hyper-Parameter Optimization

What’s Feature Engineering?

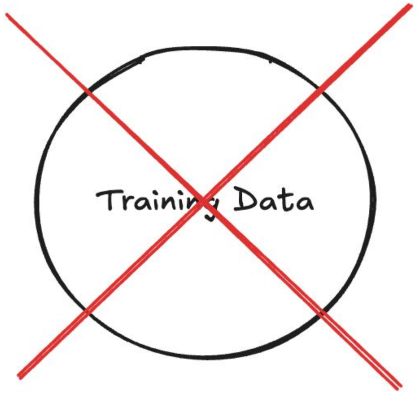
❑ Feature engineering is the process of selecting, transforming or creating input features used in machine learning.



https://www.vecteezy.com/vector-art/9725030-feature-engineering-blue-gradient-concept-icon-artificial-intelligence-problem-solving-in-machine-learning-abstract-idea-thin-line-illustration-isolated-outline-drawing

What’s Hyper-Parameter Optimization?

- Hyper-parameter optimization is the process of searching and selecting the best combinations of parameters that determine the behavior of a machine learning algorithm.

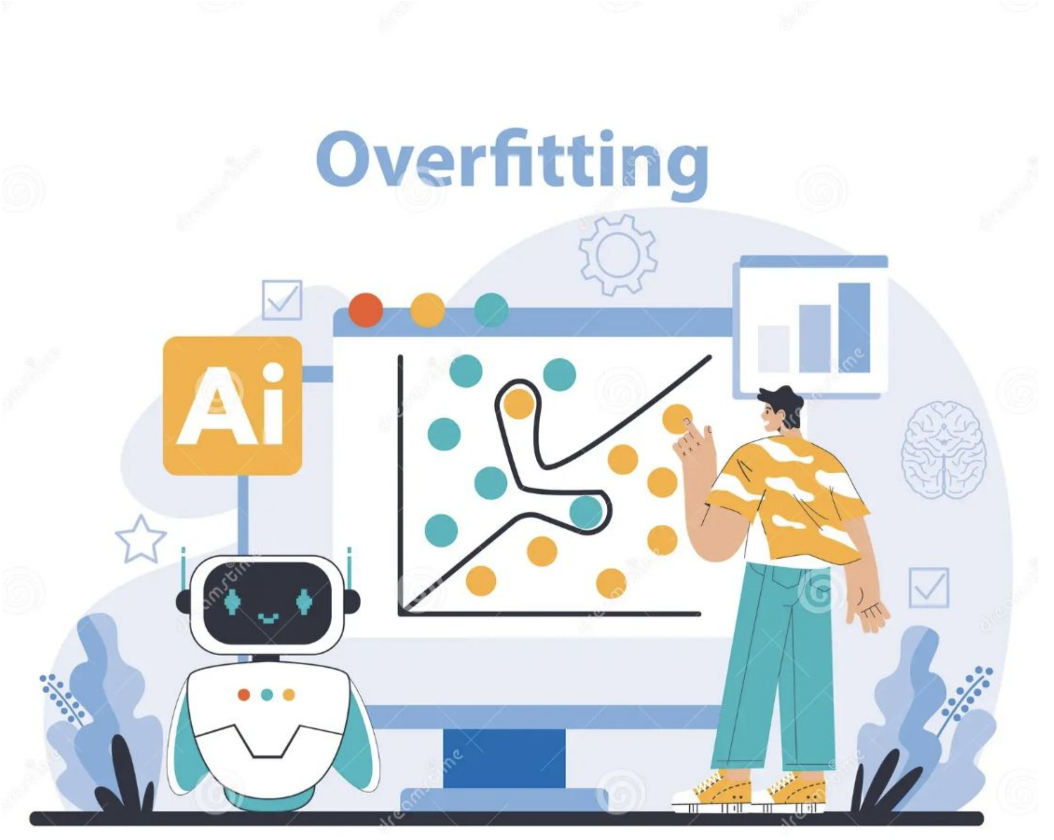


https://thenounproject.com/browse/icons/term/hyperparameter-tuning/?qv=6637415

Overfitting

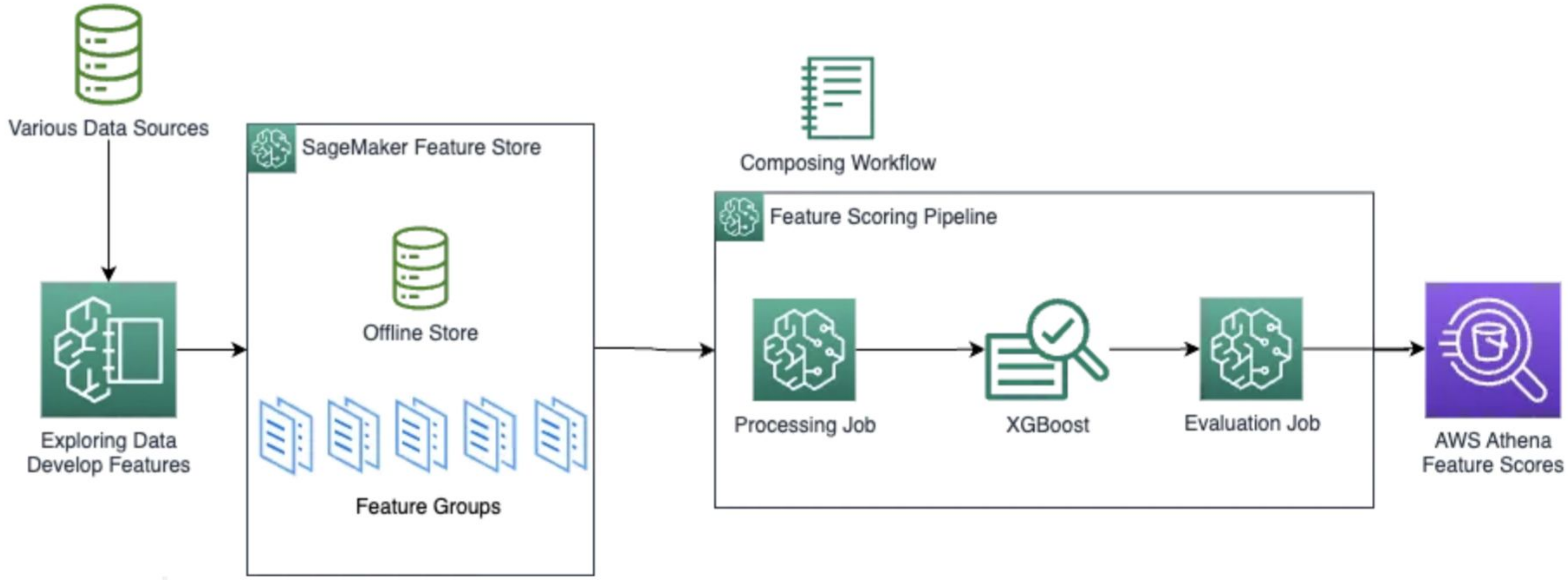
- To assess the quality of the features, an algorithm that can solve the problem is selected and trained with parameters prone to overfitting.

- The rationale behind using overfitting is as follows: If a model fits the training data extremely well, it may indicate that the features contain valuable information.



Feature Benchmarking Pipeline

Feature Benchmarking Pipeline



> Each feature group can have 2500 feature definitions. Features, which are

What’s SageMaker Feature Store?

❑ SageMaker Feature Store is a preferred method because it is cost-effective and compatible with companies' existing technology infrastructure.

❑ **Limitations**:

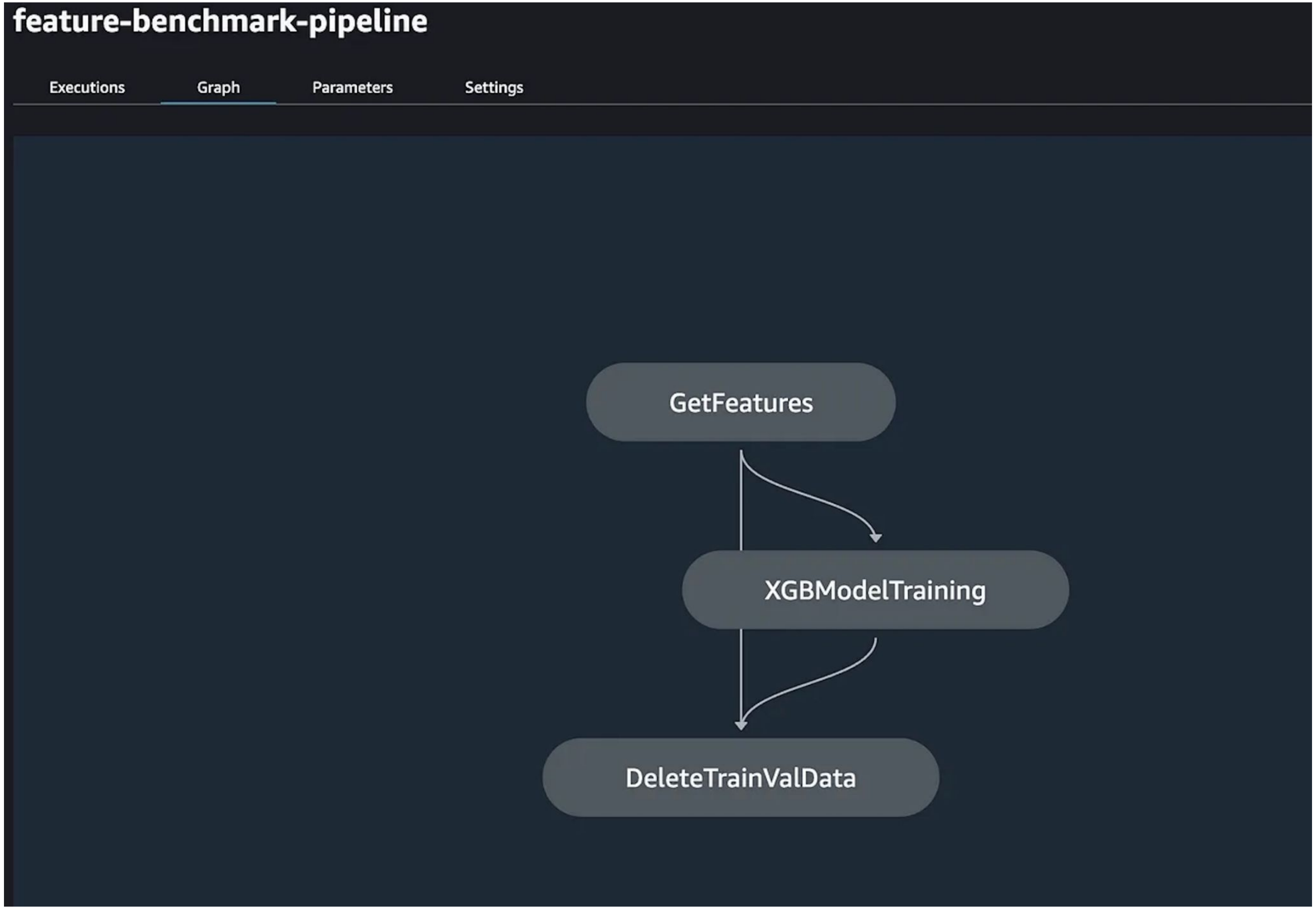
- A maximum of 2,500 features can be defined for each feature group.
- Features, especially transformer outputs from pre-trained models, can push this limit.
- In all feature groups, the timestamp (event time) must be in the following format:
  - yyyy-MM-dd’T’HH:mm:ssZ
  - yyyy-MM-dd’T’HH:mm:ss.SSSZ

What’s SageMaker Pipeline?

❑ The Amazon SageMaker Pipeline consists of interconnected steps for developing machine learning models. These steps are defined by a structure called a Directed Acyclic Graph (DAG)

Baseline Model: XGBoost

Feature Benchmarking Pipeline

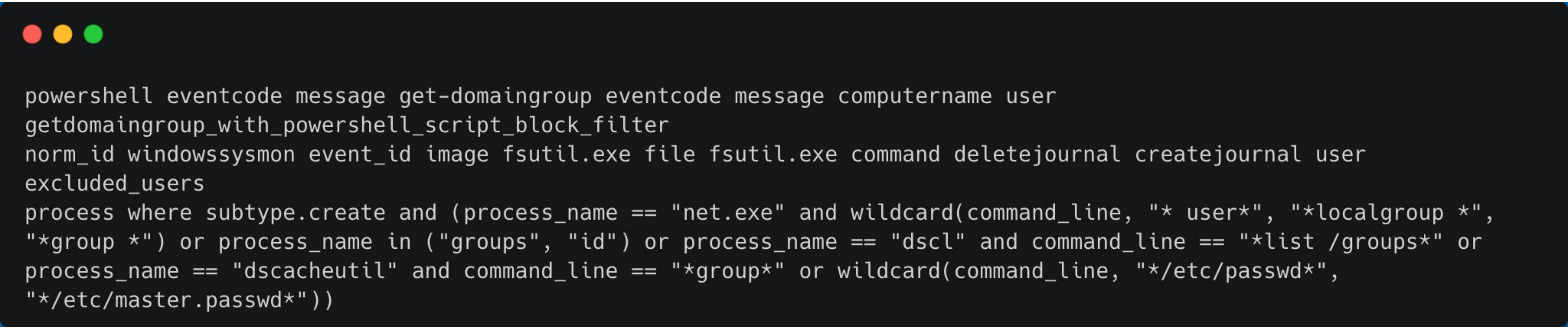


Experiment Results

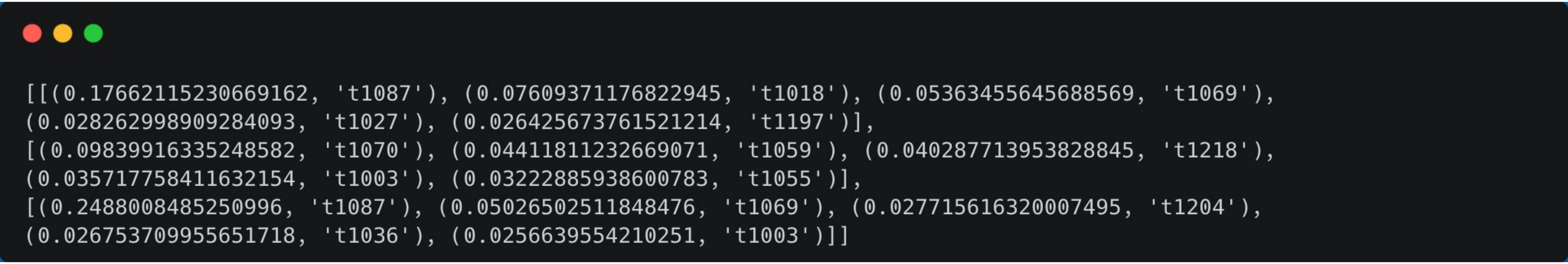
THE H@CK SUMMIT

ACADEMIC PARTNERS

First Results

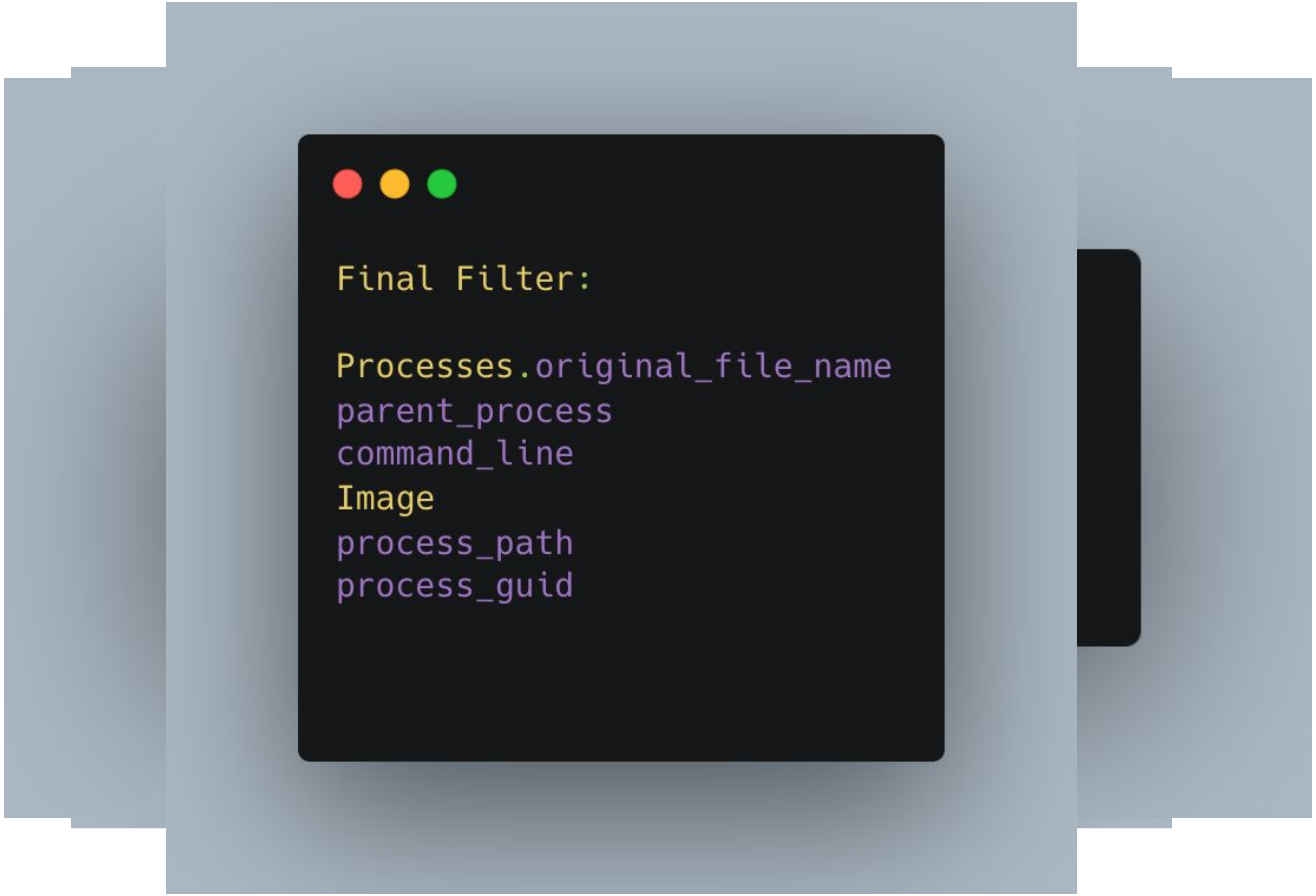


```
powershell eventcode message get-domaingroup eventcode message computername user
getdomaingroup_with_powershell_script_block_filter
norm_id windowssysmon event_id image fsutil.exe file fsutil.exe command deletejournal createjournal user
excluded_users
process where subtype.create and (process_name == "net.exe" and wildcard(command_line, "* user*", "*localgroup *",
"*group *") or process_name in ("groups", "id") or process_name == "dscl" and command_line == "*list /groups*" or
process_name == "dscacheutil" and command_line == "*group*" or wildcard(command_line, "*/etc/passwd*",
"*/etc/master.passwd*"))
```



```
[[(0.17662115230669162, 't1087'), (0.07609371176822945, 't1018'), (0.05363455645688569, 't1069'),
(0.028262998909284093, 't1027'), (0.026425673761521214, 't1197')],
[(0.09839916335248582, 't1070'), (0.04411811232669071, 't1059'), (0.040287713953828845, 't1218'),
(0.035717758411632154, 't1003'), (0.03222885938600783, 't1055')],
[(0.2488008485250996, 't1087'), (0.05026502511848476, 't1069'), (0.027715616320007495, 't1204'),
(0.026753709955651718, 't1036'), (0.0256639554210251, 't1003')]]
```

Experiment Results : Filters



```
Final Filter:

Processes.original_file_name
parent_process
command_line
Image
process_path
process_guid
```

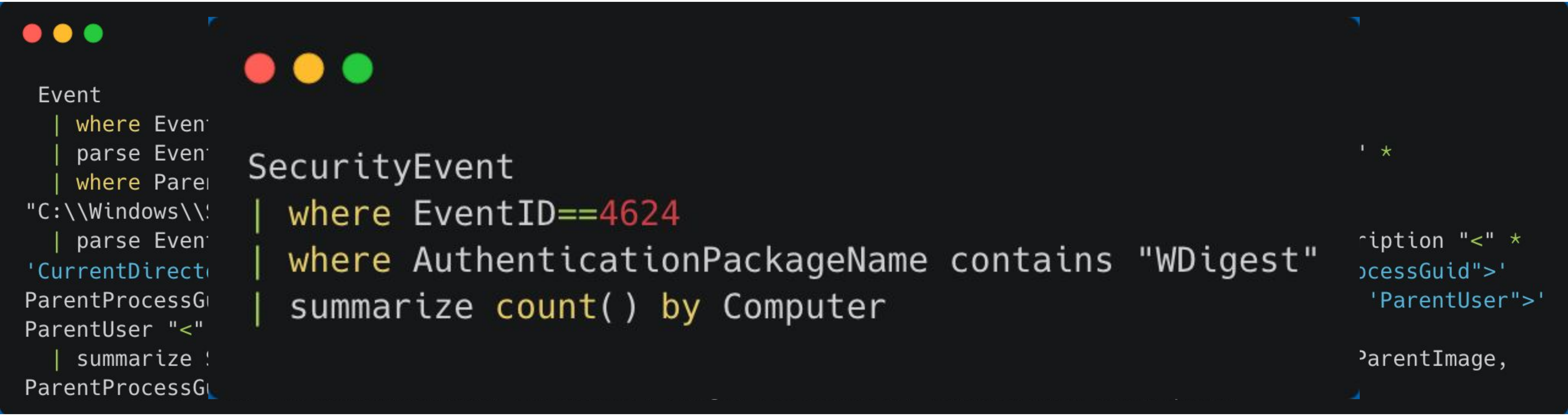
Testing Model Predictions for Specific Product

THE H@CK SUMMIT

ACADEMIC PARTNERS

MITRE For Microsoft Sentinel Query

Sample 2

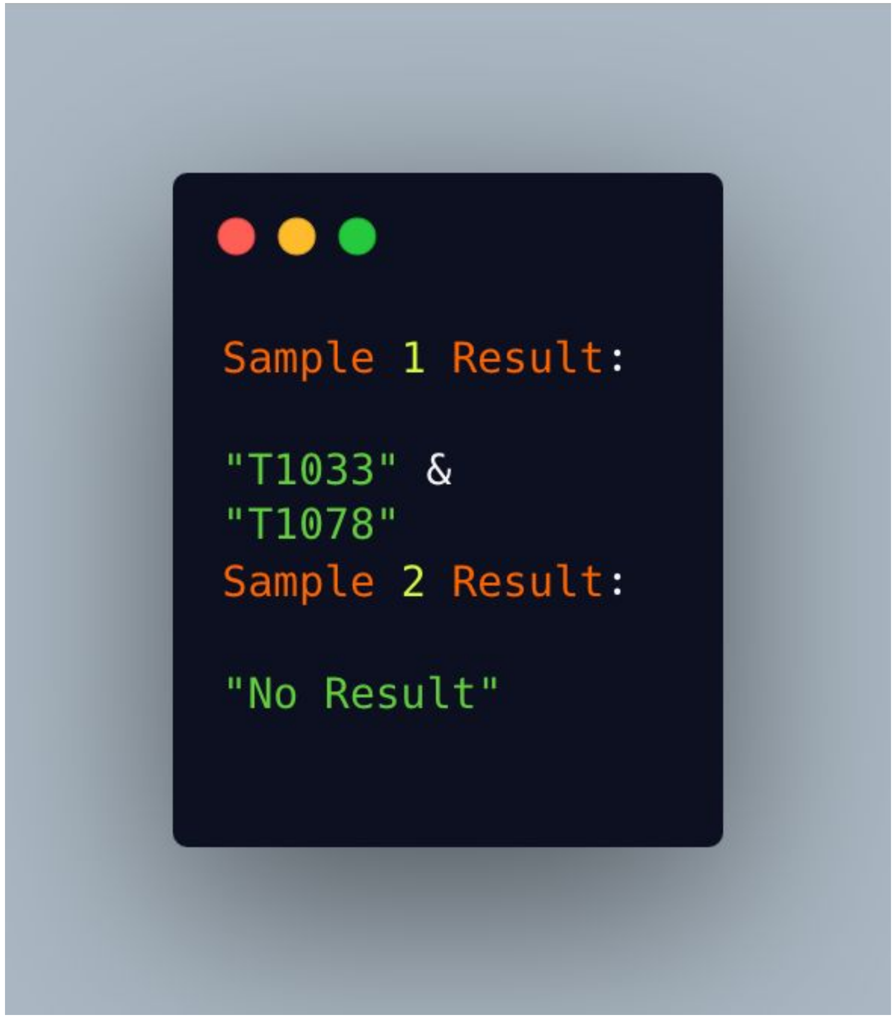


```
SecurityEvent
| where EventID==4624
| where AuthenticationPackageName contains "WDigest"
| summarize count() by Computer
```

```
 Event
   | where Even
   | parse Even                                            ' *
   | where Pare
"C:\\Windows\\
   | parse Even                                            ription "<" *
'CurrentDirect                                             ocessGuid">'
ParentProcessG                                              'ParentUser">'
ParentUser "<"
   | summarize                                             arentImage,
ParentProcessG
```

MITRE For Microsoft Sentinel Query

First Result:



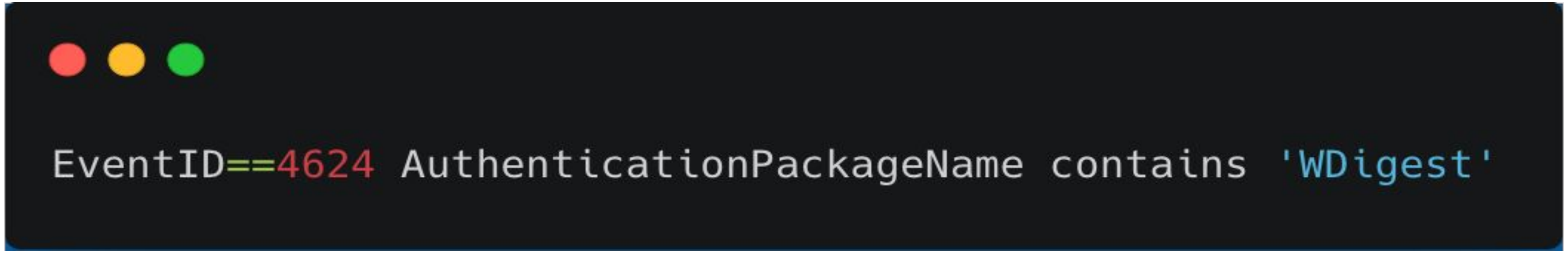
```
Sample 1 Result:

"T1033" &
"T1078"
Sample 2 Result:

"No Result"
```

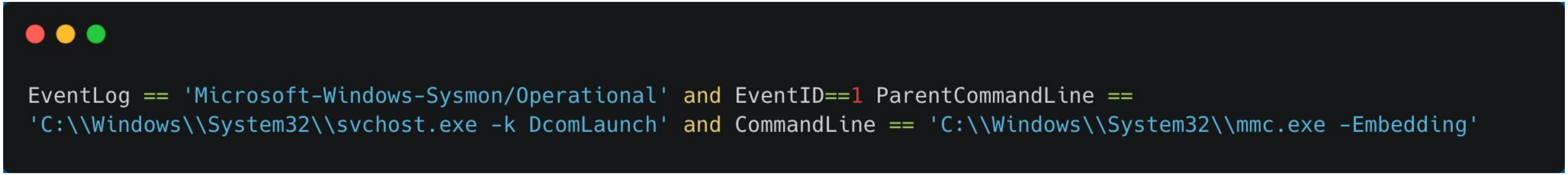
MITRE For Microsoft Sentinel Query

Sample 1



```
EventID==4624 AuthenticationPackageName contains 'WDigest'
```

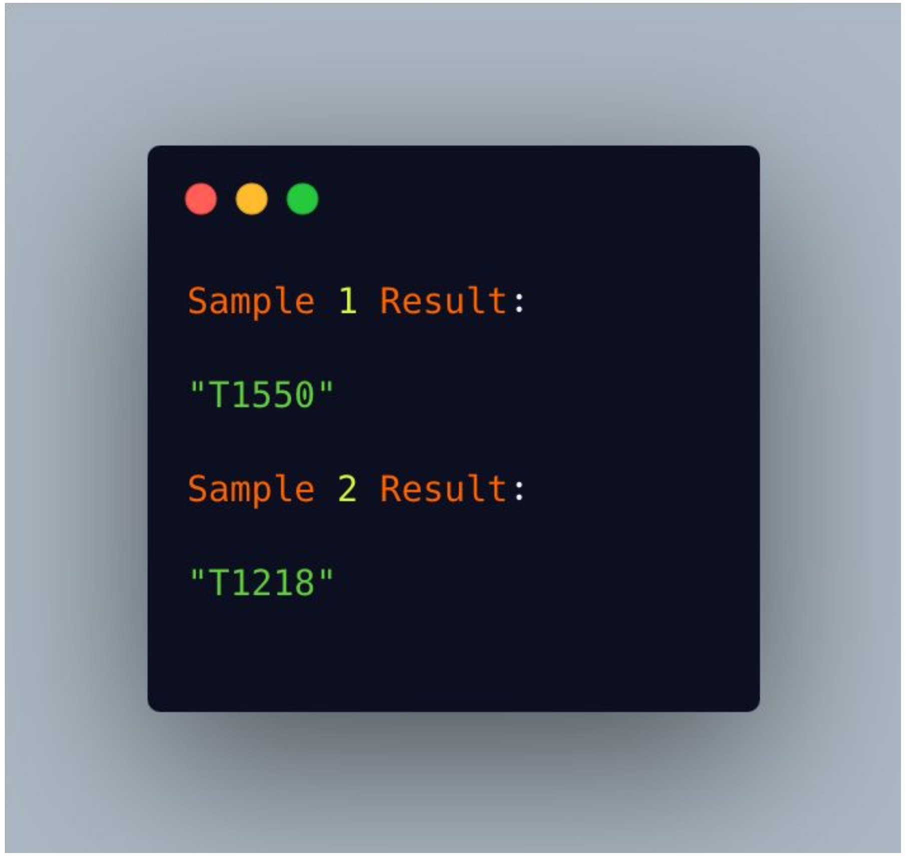
Sample 2



```
EventLog == 'Microsoft-Windows-Sysmon/Operational' and EventID==1 ParentCommandLine ==
'C:\\Windows\\System32\\svchost.exe -k DcomLaunch' and CommandLine == 'C:\\Windows\\System32\\mmc.exe -Embedding'
```

MITRE For Microsoft Sentinel Query

Final Result:



```
Sample 1 Result:

"T1550"

Sample 2 Result:

"T1218"
```

MITRE For Microsoft Sentinel Query

Phase	Match (%)	No Match (%)	Overall Score
Phase-1	55	45	55.56
Phase-2	70	30	70.37
Phase-3	81	19	80.65

Coverage Improvement

THE H@CK SUMMIT

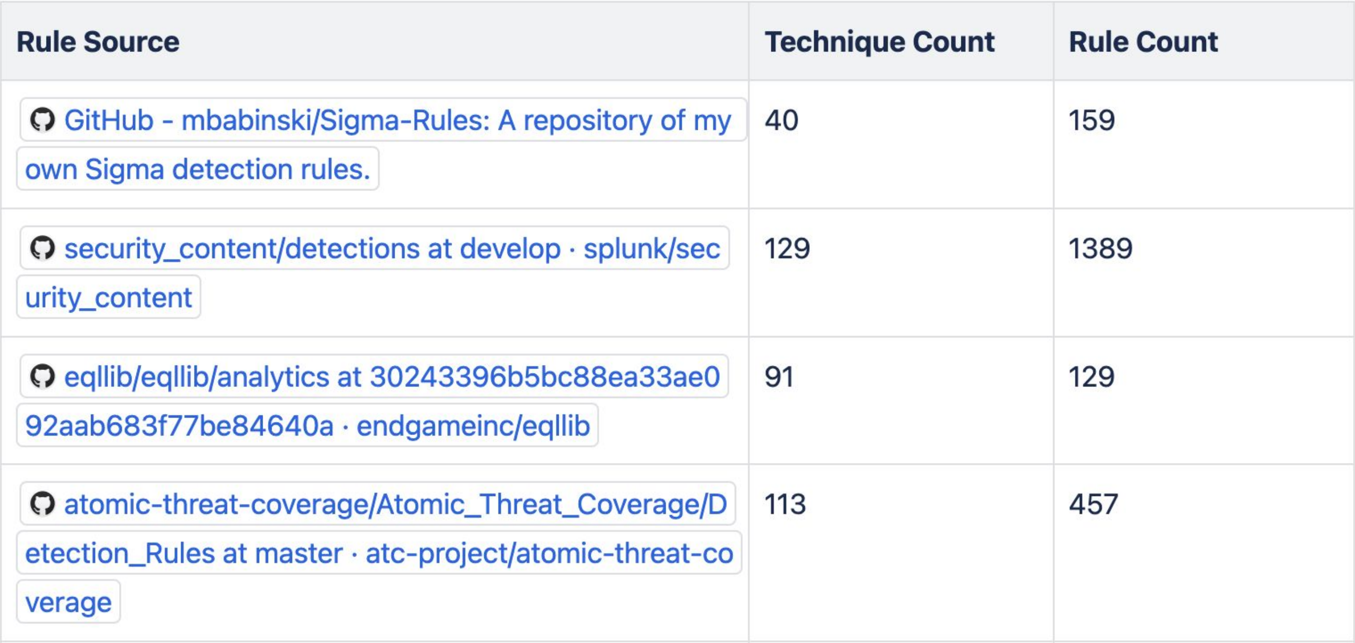
ACADEMIC PARTNERS

Problem Definitions

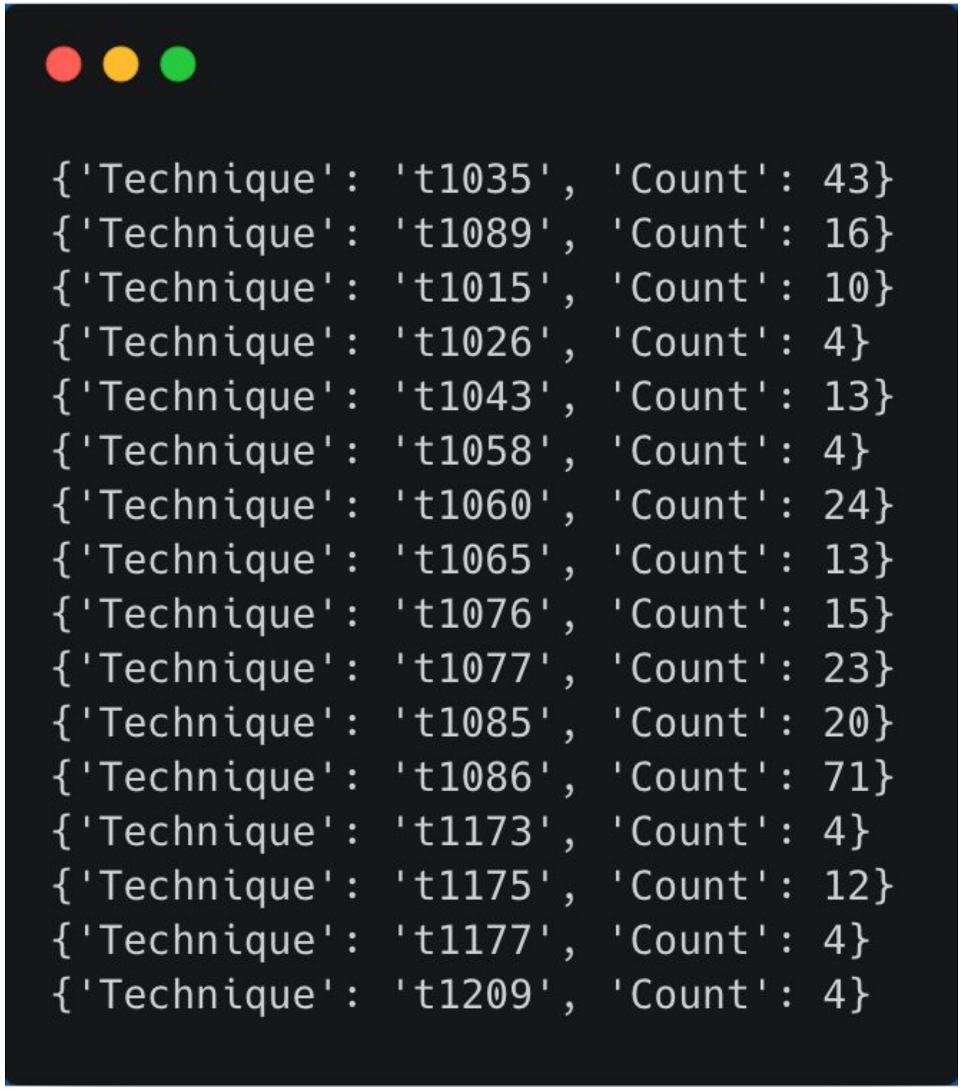
- ❑ Differences in Training Data
- ❑ Supported MITRE Techniques
- ❑ Outdated MITRE ATT&CK Version
- ❑ Lack of Dataset Enrichment
- ❑ Different Query Fields
- ❑ Focusing Certain Techniques
    - ❑ T1078: Valid Accounts
    - ❑ T1190: Exploit Public-Facing Application
    - ❑ T1110: Brute Force
    - ❑ T1562: Impair Defenses

Dataset Enrichment & Testing Model

- ❑ Open Source Count: More than 20
- ❑ Open Source Rule Count: More than 10k
- ❑ Each source helped us create examples for multiple techniques.



Rule Source	Technique Count	Rule Count
GitHub - mbabinski/Sigma-Rules: A repository of my own Sigma detection rules.	40	159
security_content/detections at develop · splunk/security_content	129	1389
eqllib/eqllib/analytics at 30243396b5bc88ea33ae092aab683f77be84640a · endgameinc/eqllib	91	129
atomic-threat-coverage/Atomic_Threat_Coverage/Detection_Rules at master · atc-project/atomic-threat-coverage	113	457



```
{'Technique': 't1035', 'Count': 43}
{'Technique': 't1089', 'Count': 16}
{'Technique': 't1015', 'Count': 10}
{'Technique': 't1026', 'Count': 4}
{'Technique': 't1043', 'Count': 13}
{'Technique': 't1058', 'Count': 4}
{'Technique': 't1060', 'Count': 24}
{'Technique': 't1065', 'Count': 13}
{'Technique': 't1076', 'Count': 15}
{'Technique': 't1077', 'Count': 23}
{'Technique': 't1085', 'Count': 20}
{'Technique': 't1086', 'Count': 71}
{'Technique': 't1173', 'Count': 4}
{'Technique': 't1175', 'Count': 12}
{'Technique': 't1177', 'Count': 4}
{'Technique': 't1209', 'Count': 4}
```

Product Based Results

- Splunk
  - Model v1: 637
  - Model v2: 463
  - -27.31%
- QRadar
  - Model v1: 612
  - Model v2: 436
  - -28.75%
- ArcSight
  - Model v1: 617
  - Model v2: 467
  - -24.31%

Sigma Rule Based Results

- ❏ **Splunk**
  - ❏ Model v1: 327
  - ❏ Model v2: 419
  - ❏ **+28.13%**
- ❏ **QRadar**
  - ❏ Model v1: 285
  - ❏ Model v2: 393
  - ❏ **+37.89%**
- ❏ **ArcSight**
  - ❏ Model v1: 254
  - ❏ Model v2: 429
  - ❏ **+24.31%**

Real-Time MITRE AI Service Implementation

THE H@CK SUMMIT

ACADEMIC PARTNERS

MLOps & Model Inference

- Machine learning enables data-driven decision making and automation in areas such as cybersecurity. However, deploying and managing models at scale presents challenges.
- The model inference phase enables trained models to become usable by applications. This process is the transformation of raw data inputs into meaningful predictions.

SageMaker Inference Types

- ❑ Real-time Inference
- ❑ Serverless Inference
- ❑ Batch Transform
- ❑ Asynchronous Inference

Amazon SageMaker

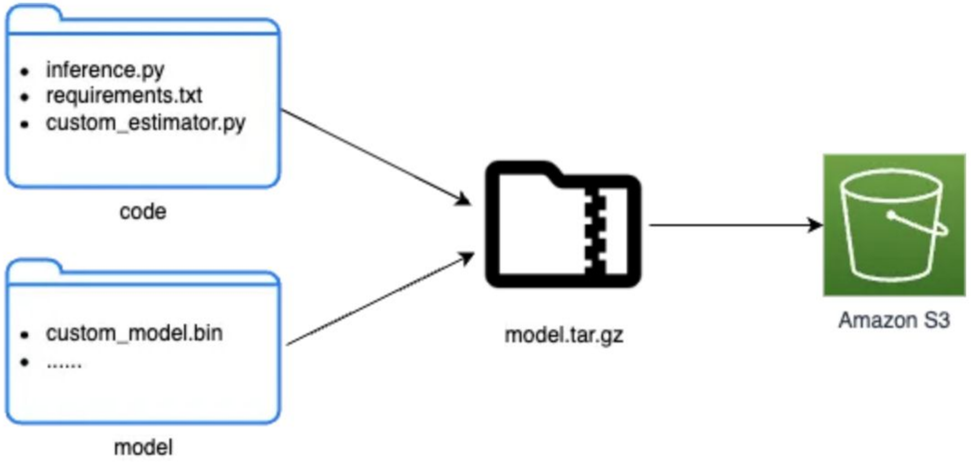
Advantages & Limitations of Serverless Inference

- ❑ **Advantages:**
  - Cost Efficiency
  - Auto Scaling
  - Simple Management
  - Flexibility
- ❑ **Limitations:**
  - Memory
  - The maximum data size
  - Processing time is limited
  - A cold start condition may occur

SageMaker Inference Process

Model Preparation and Packaging:

- Model outputs are stored on S3.
- Pre-processing and post-processing operations are defined with custom scripts.
- Pre-trained models such as HuggingFace can be packaged.



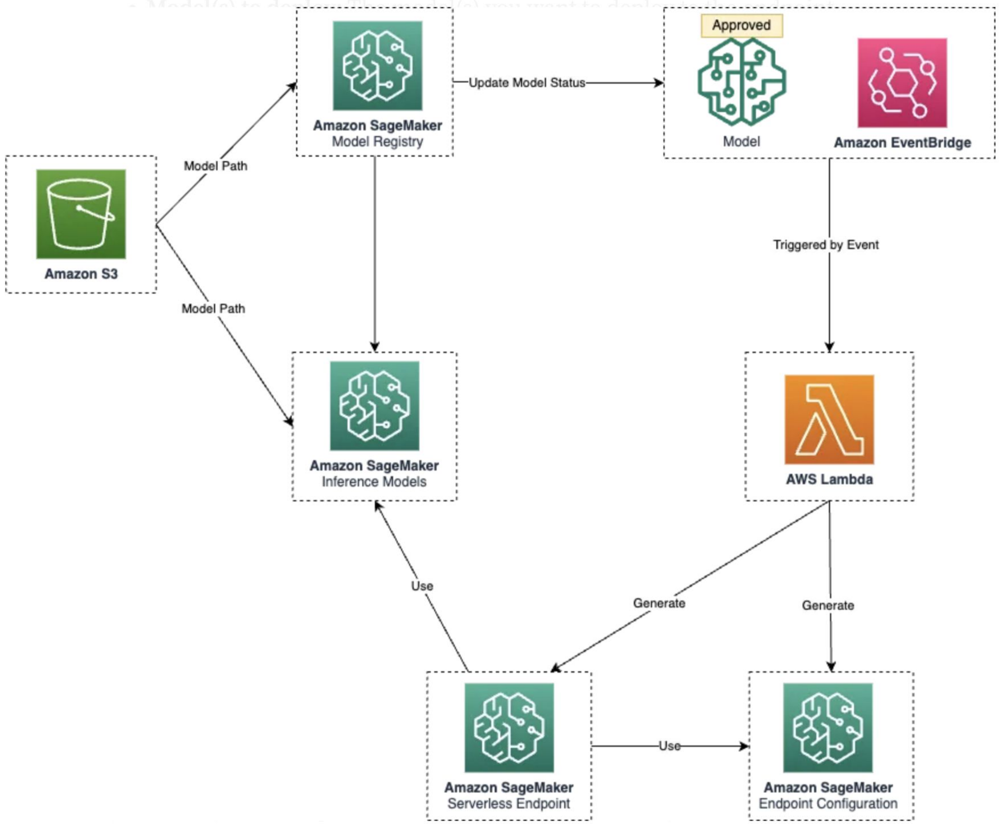
Writing a Custom Inference Script:

- A custom Python script is defined to process the incoming data, make predictions and format the output.

SageMaker Inference Process

Model Registration and Deployment:

- Models are versioned with the SageMaker Model Registry.
- Models can be served with custom Docker image.
- Automatic model deployment is provided using EventBridge and Lambda.

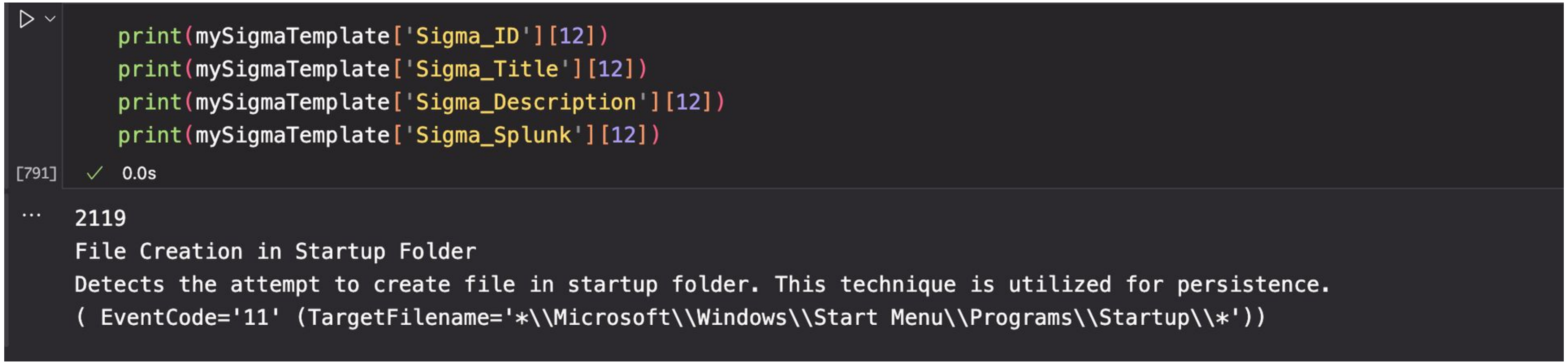


Use Cases

THE H@CK SUMMIT

ACADEMIC PARTNERS

Sigma Rule Integration

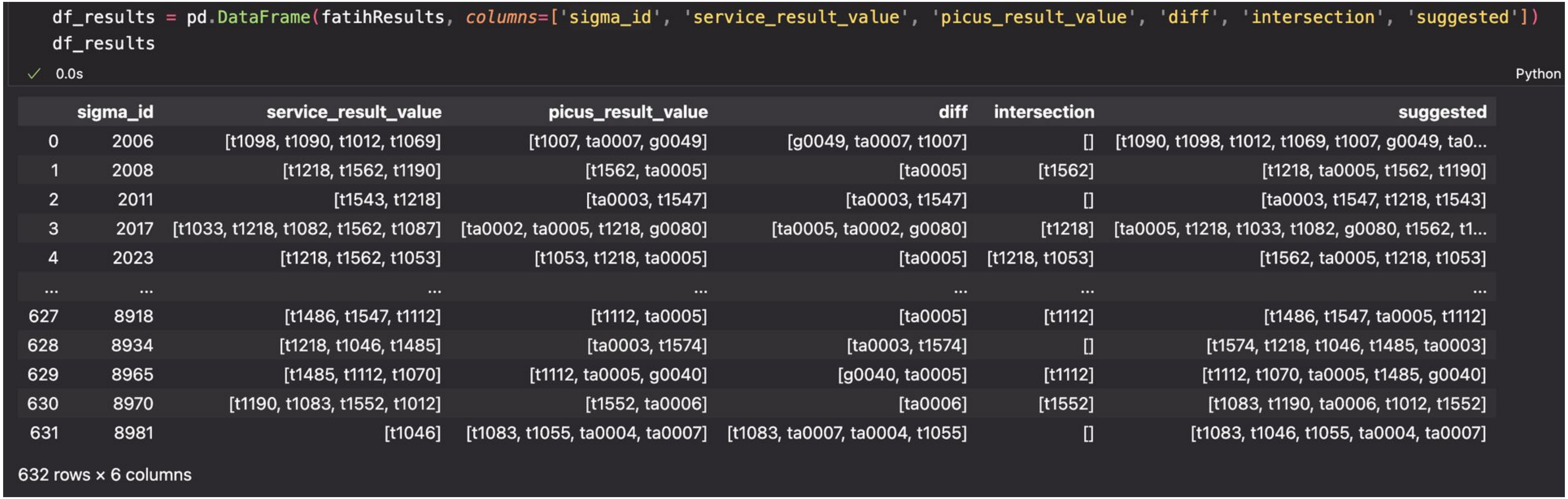


```
print(mySigmaTemplate['Sigma_ID'][12])
print(mySigmaTemplate['Sigma_Title'][12])
print(mySigmaTemplate['Sigma_Description'][12])
print(mySigmaTemplate['Sigma_Splunk'][12])
```

[791] ✓ 0.0s

```
2119
File Creation in Startup Folder
Detects the attempt to create file in startup folder. This technique is utilized for persistence.
( EventCode='11' (TargetFilename='*\\Microsoft\\Windows\\Start Menu\\Programs\\Startup\\*'))
```

Sigma Rule Integration



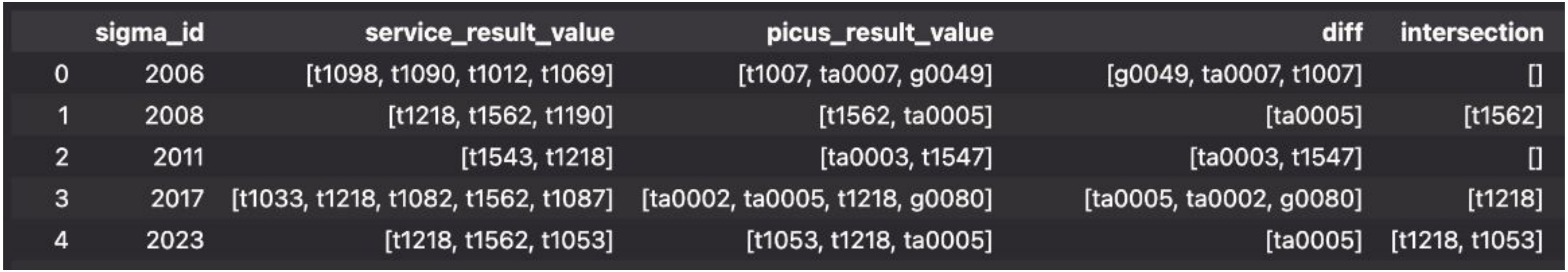
```python
df_results = pd.DataFrame(fatihResults, columns=['sigma_id', 'service_result_value', 'picus_result_value', 'diff', 'intersection', 'suggested'])
df_results
```

✓ 0.0s Python

	sigma_id	service_result_value	picus_result_value	diff	intersection	suggested
0	2006	[t1098, t1090, t1012, t1069]	[t1007, ta0007, g0049]	[g0049, ta0007, t1007]	[]	[t1090, t1098, t1012, t1069, t1007, g0049, ta0...
1	2008	[t1218, t1562, t1190]	[t1562, ta0005]	[ta0005]	[t1562]	[t1218, ta0005, t1562, t1190]
2	2011	[t1543, t1218]	[ta0003, t1547]	[ta0003, t1547]	[]	[ta0003, t1547, t1218, t1543]
3	2017	[t1033, t1218, t1082, t1562, t1087]	[ta0002, ta0005, t1218, g0080]	[ta0005, ta0002, g0080]	[t1218]	[ta0005, t1218, t1033, t1082, g0080, t1562, t1...
4	2023	[t1218, t1562, t1053]	[t1053, t1218, ta0005]	[ta0005]	[t1218, t1053]	[t1562, ta0005, t1218, t1053]
...	...	...	...	...	...	...
627	8918	[t1486, t1547, t1112]	[t1112, ta0005]	[ta0005]	[t1112]	[t1486, t1547, ta0005, t1112]
628	8934	[t1218, t1046, t1485]	[ta0003, t1574]	[ta0003, t1574]	[]	[t1574, t1218, t1046, t1485, ta0003]
629	8965	[t1485, t1112, t1070]	[t1112, ta0005, g0040]	[g0040, ta0005]	[t1112]	[t1112, t1070, ta0005, t1485, g0040]
630	8970	[t1190, t1083, t1552, t1012]	[t1552, ta0006]	[ta0006]	[t1552]	[t1083, t1190, ta0006, t1012, t1552]
631	8981	[t1046]	[t1083, t1055, ta0004, ta0007]	[t1083, ta0007, ta0004, t1055]	[]	[t1083, t1046, t1055, ta0004, ta0007]

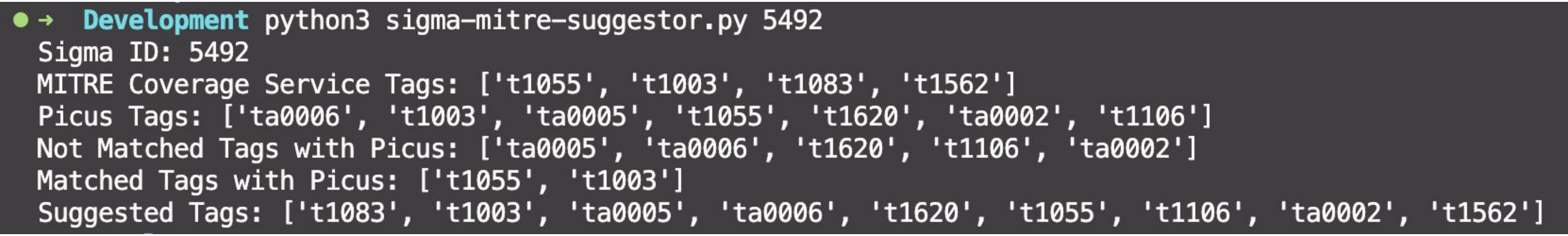
632 rows × 6 columns

Sigma Rule Integration



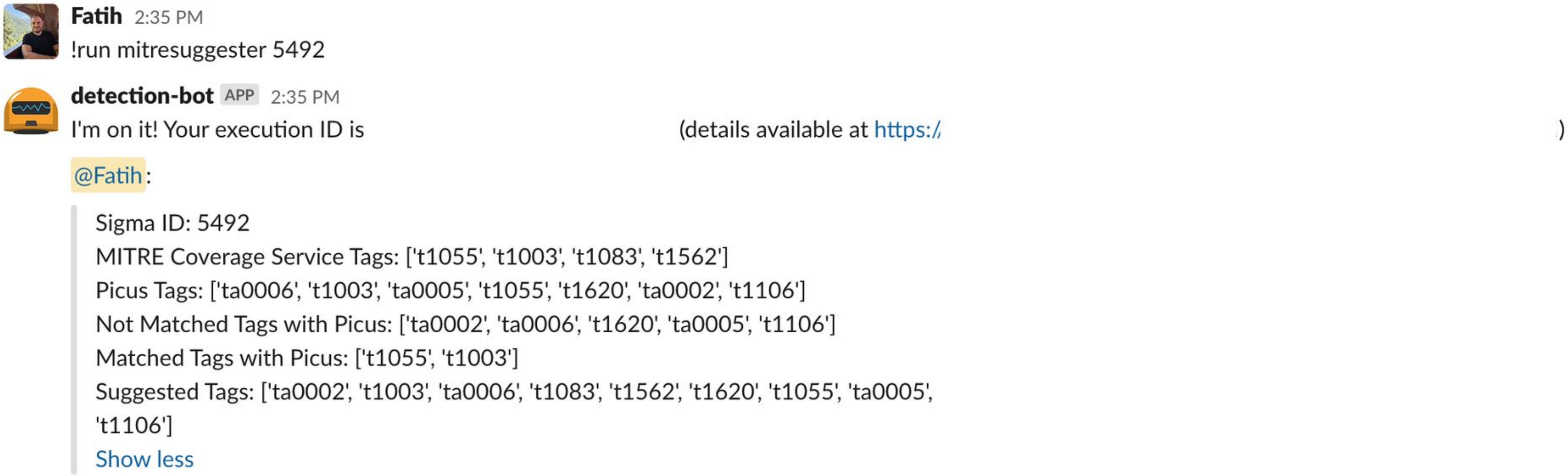
	sigma_id	service_result_value	picus_result_value	diff	intersection
0	2006	[t1098, t1090, t1012, t1069]	[t1007, ta0007, g0049]	[g0049, ta0007, t1007]	[]
1	2008	[t1218, t1562, t1190]	[t1562, ta0005]	[ta0005]	[t1562]
2	2011	[t1543, t1218]	[ta0003, t1547]	[ta0003, t1547]	[]
3	2017	[t1033, t1218, t1082, t1562, t1087]	[ta0002, ta0005, t1218, g0080]	[ta0005, ta0002, g0080]	[t1218]
4	2023	[t1218, t1562, t1053]	[t1053, t1218, ta0005]	[ta0005]	[t1218, t1053]

Technique Suggestions from Sigma Rule



```
● ➜  Development python3 sigma-mitre-suggestor.py 5492
  Sigma ID: 5492
  MITRE Coverage Service Tags: ['t1055', 't1003', 't1083', 't1562']
  Picus Tags: ['ta0006', 't1003', 'ta0005', 't1055', 't1620', 'ta0002', 't1106']
  Not Matched Tags with Picus: ['ta0005', 'ta0006', 't1620', 't1106', 'ta0002']
  Matched Tags with Picus: ['t1055', 't1003']
  Suggested Tags: ['t1083', 't1003', 'ta0005', 'ta0006', 't1620', 't1055', 't1106', 'ta0002', 't1562']
```

Technique Suggestions from Sigma Rule



Fatih 2:35 PM
!run mitresuggester 5492

detection-bot APP 2:35 PM
I'm on it! Your execution ID is (details available at https:/)

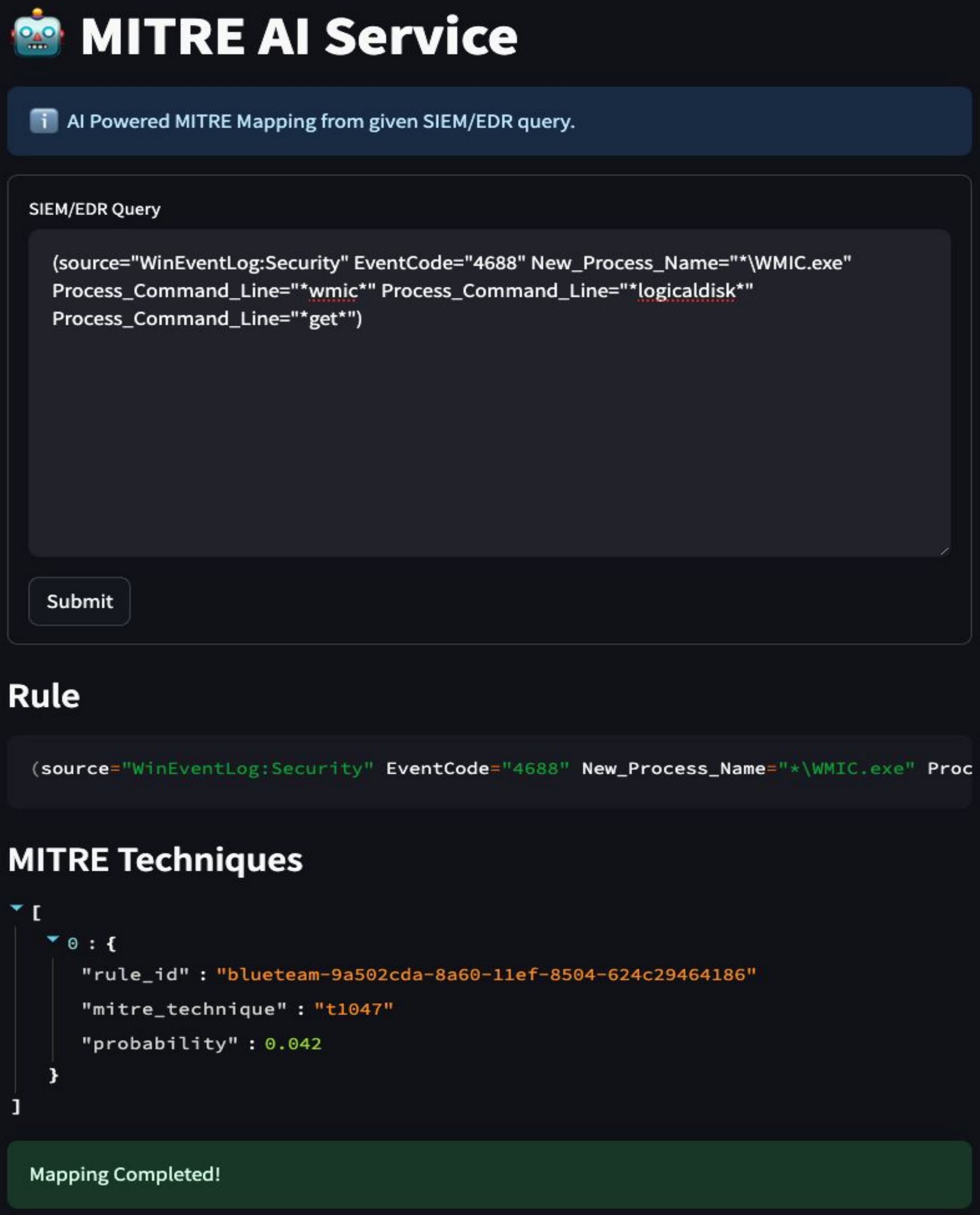
@Fatih:

Sigma ID: 5492
MITRE Coverage Service Tags: ['t1055', 't1003', 't1083', 't1562']
Picus Tags: ['ta0006', 't1003', 'ta0005', 't1055', 't1620', 'ta0002', 't1106']
Not Matched Tags with Picus: ['ta0002', 'ta0006', 't1620', 'ta0005', 't1106']
Matched Tags with Picus: ['t1055', 't1003']
Suggested Tags: ['ta0002', 't1003', 'ta0006', 't1083', 't1562', 't1620', 't1055', 'ta0005', 't1106']

MITRE AI Service

THE H@CK SUMMIT

ACADEMIC PARTNERS



MITRE AI Service

AI Powered MITRE Mapping from given SIEM/EDR query.

SIEM/EDR Query

```
(source="WinEventLog:Security" EventCode="4688" New_Process_Name="*\WMIC.exe"
Process_Command_Line="*wmic*" Process_Command_Line="*logicaldisk*"
Process_Command_Line="*get*")
```

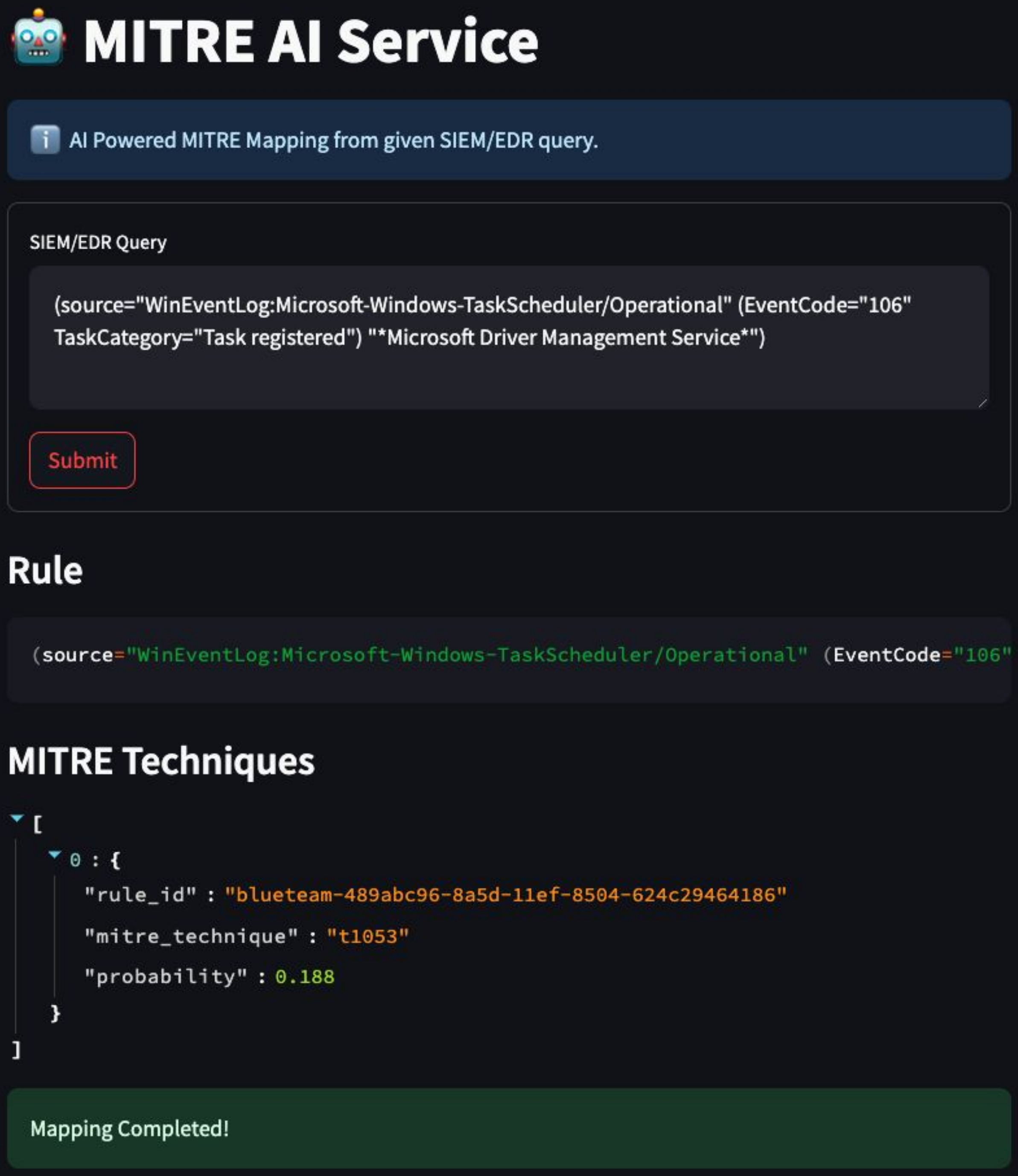
Rule

```
(source="WinEventLog:Security" EventCode="4688" New_Process_Name="*\WMIC.exe" Proc
```

MITRE Techniques

```
[
  0 : {
    "rule_id" : "blueteam-9a502cda-8a60-11ef-8504-624c29464186"
    "mitre_technique" : "t1047"
    "probability" : 0.042
  }
]
```

Mapping Completed!



MITRE AI Service

AI Powered MITRE Mapping from given SIEM/EDR query.

SIEM/EDR Query

```
(source="WinEventLog:Microsoft-Windows-TaskScheduler/Operational" (EventCode="106"
TaskCategory="Task registered") "*Microsoft Driver Management Service*")
```

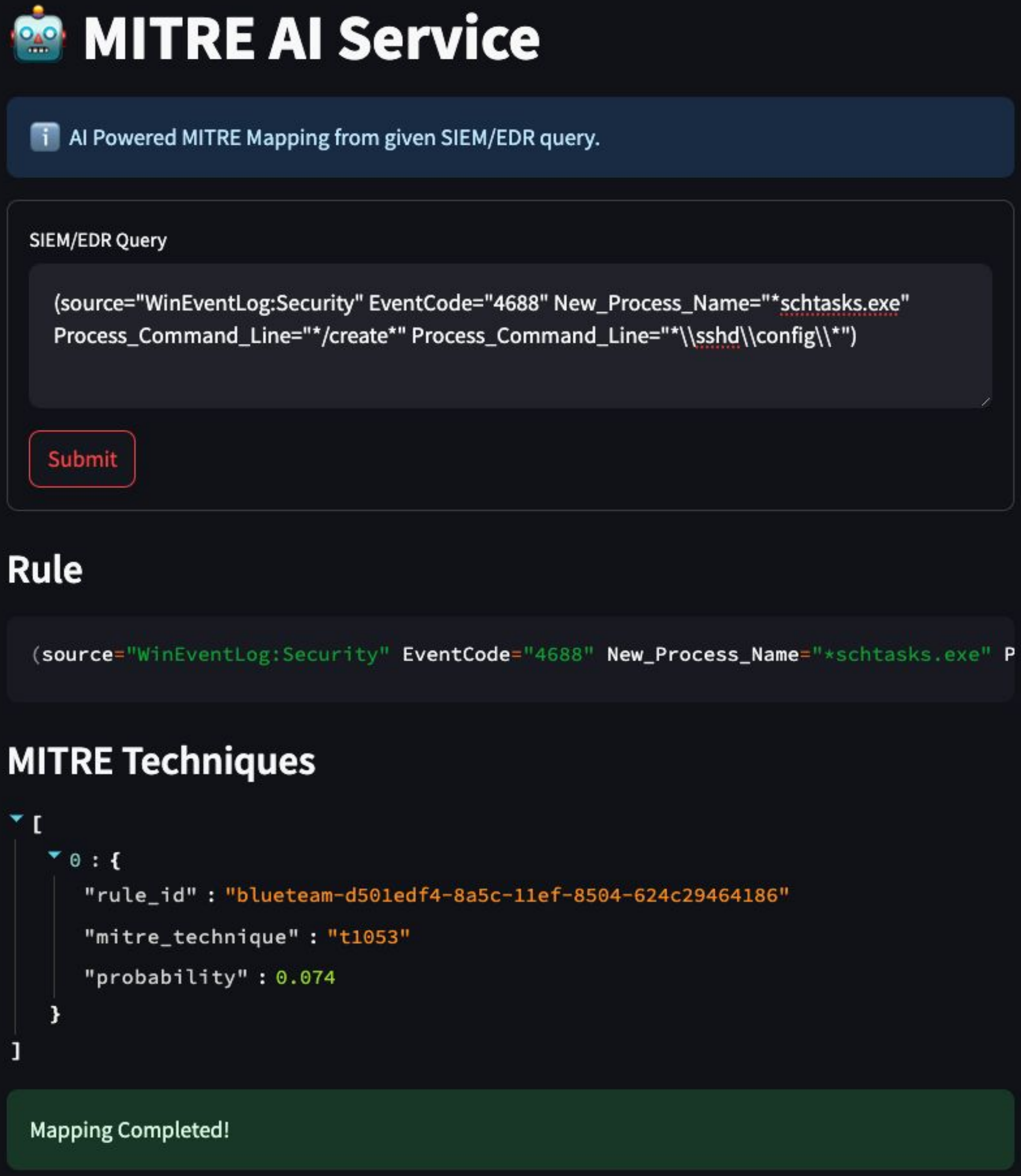
Rule

```
(source="WinEventLog:Microsoft-Windows-TaskScheduler/Operational" (EventCode="106"
```

MITRE Techniques

```
[
  0 : {
    "rule_id" : "blueteam-489abc96-8a5d-11ef-8504-624c29464186"
    "mitre_technique" : "t1053"
    "probability" : 0.188
  }
]
```

Mapping Completed!



MITRE AI Service

AI Powered MITRE Mapping from given SIEM/EDR query.

SIEM/EDR Query

```
(source="WinEventLog:Security" EventCode="4688" New_Process_Name="*schtasks.exe"
Process_Command_Line="*/create*" Process_Command_Line="*\\sshd\\config\\*")
```

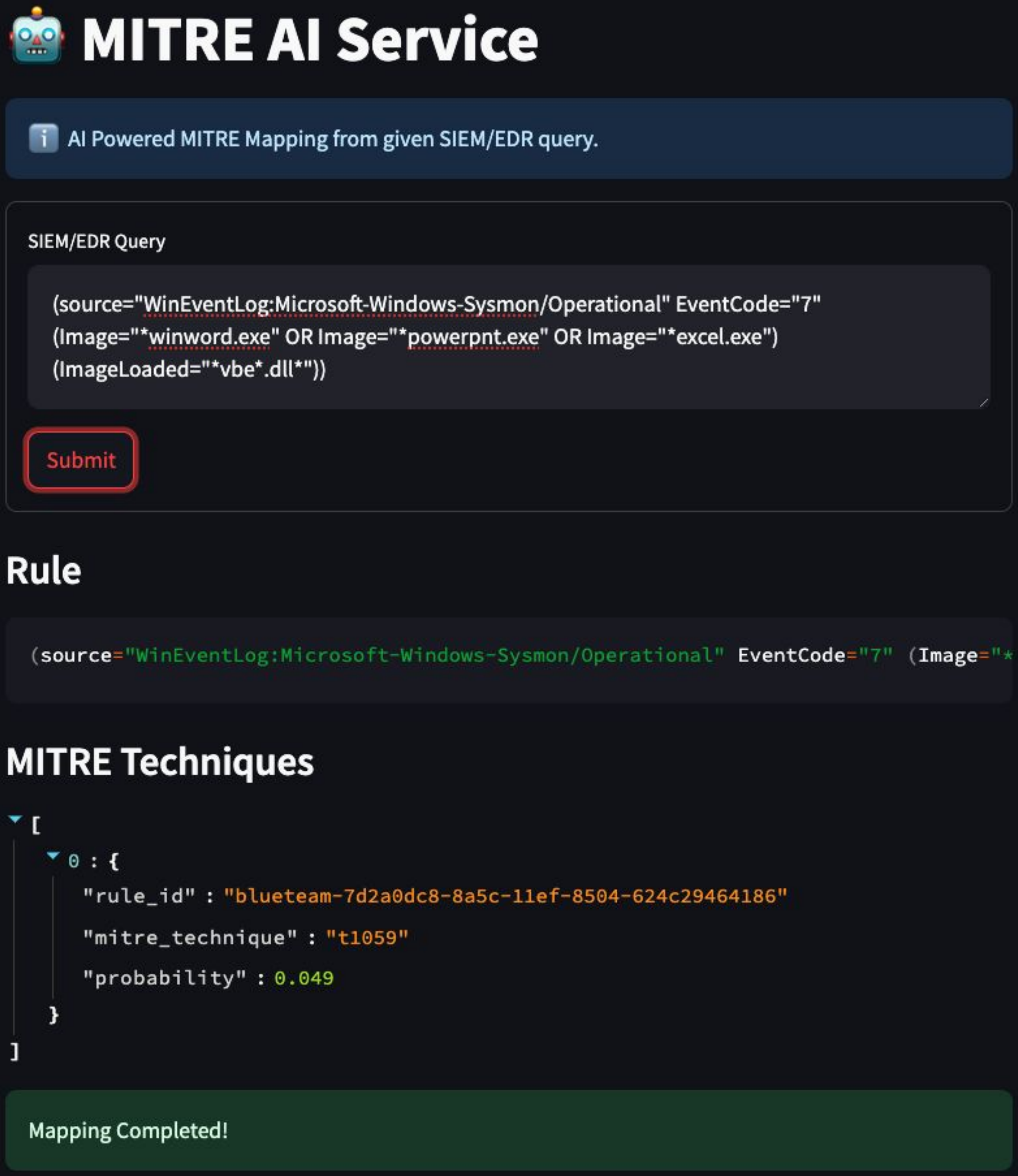
Rule

```
(source="WinEventLog:Security" EventCode="4688" New_Process_Name="*schtasks.exe" P
```

MITRE Techniques

```
[
  0 : {
    "rule_id" : "blueteam-d501edf4-8a5c-11ef-8504-624c29464186"
    "mitre_technique" : "t1053"
    "probability" : 0.074
  }
]
```

Mapping Completed!



MITRE AI Service

AI Powered MITRE Mapping from given SIEM/EDR query.

SIEM/EDR Query

```
(source="WinEventLog:Microsoft-Windows-Sysmon/Operational" EventCode="7"
(Image="*winword.exe" OR Image="*powerpnt.exe" OR Image="*excel.exe")
(ImageLoaded="*vbe*.dll*"))
```

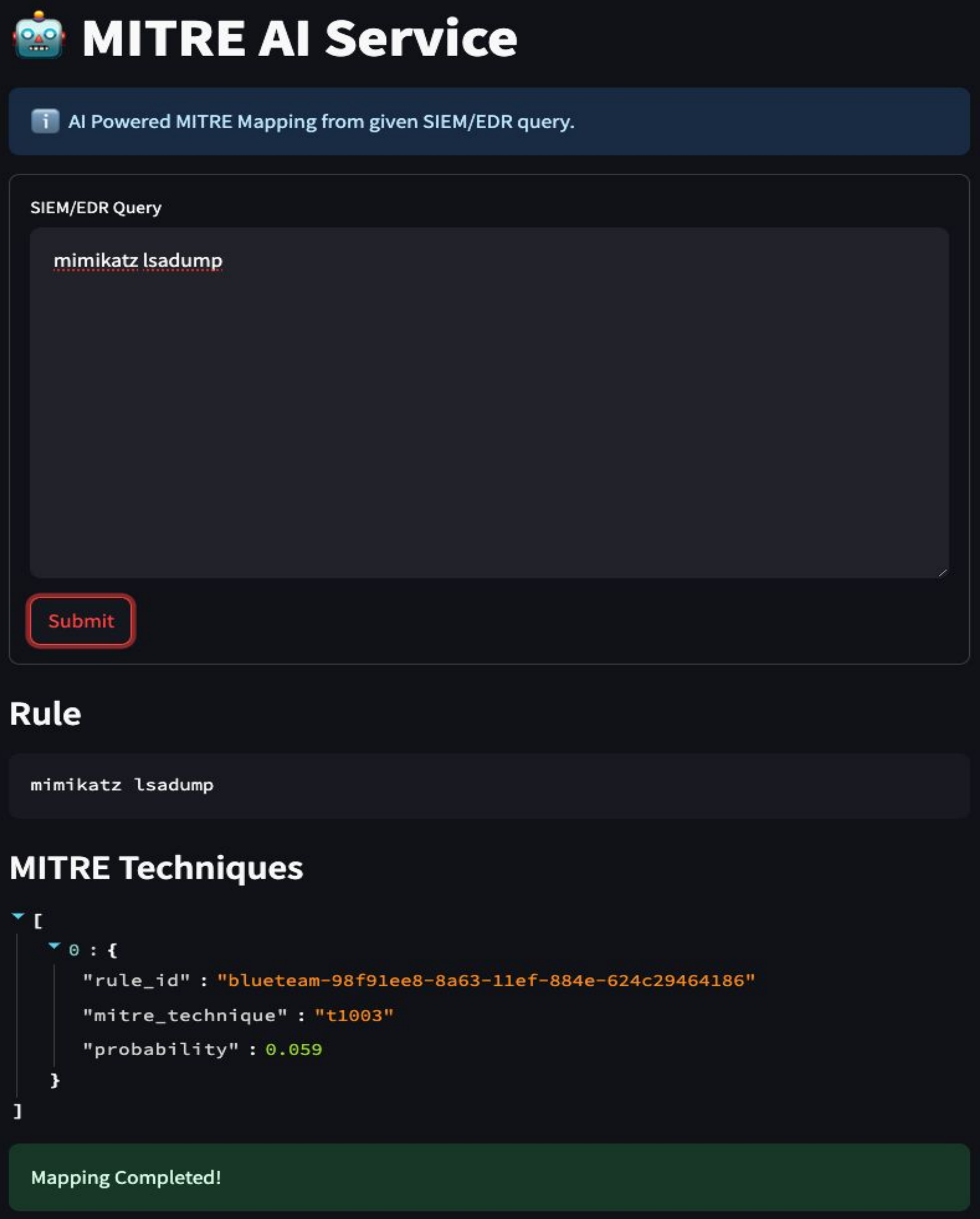
Rule

```
(source="WinEventLog:Microsoft-Windows-Sysmon/Operational" EventCode="7" (Image="*
```

MITRE Techniques

```
[
  0 : {
    "rule_id" : "blueteam-7d2a0dc8-8a5c-11ef-8504-624c29464186"
    "mitre_technique" : "t1059"
    "probability" : 0.049
  }
]
```

Mapping Completed!



MITRE AI Service

AI Powered MITRE Mapping from given SIEM/EDR query.

SIEM/EDR Query

mimikatz lsadump

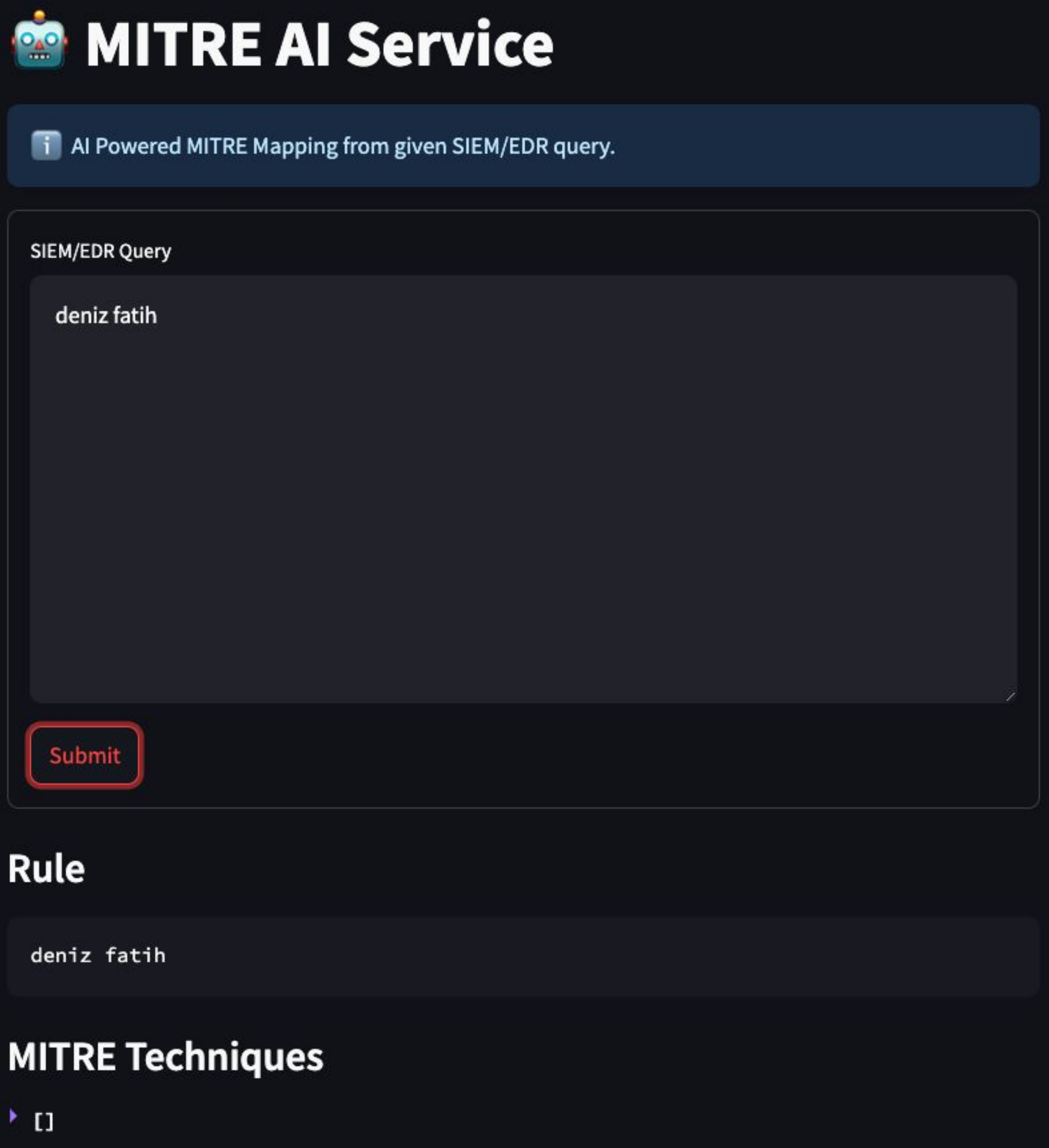
Rule

```
mimikatz lsadump
```

MITRE Techniques

```
[
  0 : {
    "rule_id" : "blueteam-98f91ee8-8a63-11ef-884e-624c29464186"
    "mitre_technique" : "t1003"
    "probability" : 0.059
  }
]
```

Mapping Completed!



MITRE AI Service

AI Powered MITRE Mapping from given SIEM/EDR query.

SIEM/EDR Query

deniz fatih

Rule

```
deniz fatih
```

MITRE Techniques

```
[]
```

FEEDBACK

AI-Powered MITRE ATT&CK Mapping: Enhancing Cyber Threat Detection

Fatih Erdogan
Deniz Sakli

https://thehacksummit.com/user.html#!/lecture/THS24-cc75/rate

Thank you for watching!

Deniz Saklı
Senior
Blue Team Engineer
PICUS
in denizsakli

Fatih Erdoğan
Senior
Blue Team Engineer
PICUS
in fatiherdogan1

THE H@CK SUMMIT

ACADEMIC PARTNERS