

SigmaGen AI-Powered Sigma Rules Generation with MITRE ATT&CK for Enhanced Threat Detection

07 March 2025

Asia-Pacific ATT&CK Community Workshop, Singapore

WHO ARE WE?



Pham Truong, Giang
Cloud Security Engineer
Cyber Security Service (SAS) at FPT Software

A passionate Security Engineer with expertise in building and optimizing solutions to identify, analyze, and respond to security threats. Skilled in streamlining processes, enhancing threat visibility, and driving swift incident resolution to safeguard organizational assets.

 <https://www.linkedin.com/in/gianglucas/>



Nguyen Duc, Tung
Cloud Security Engineer
Cyber Security Service (SAS) at FPT Software

Security Engineer with a strong focus on leveraging automation to enhance efficiency and effectiveness. Experienced in designing and implementing automated workflows for threat detection, incident response, and security operations.

 <https://www.linkedin.com/in/tungdra/>

What is Sigma?

- **Rule-based detection format** designed for SIEM/EDR platforms.
- **Focus on log-based detections** across Windows, Linux, cloud, and other environments.

Advantages of Sigma rule:

- ✓ **Platform-Independent:** can be converted into different SIEM formats (e.g., Splunk, ElasticSearch, Sentinel).
- ✓ **Human-Readable & Flexible:** They allow **blue teams** and **SOC analysts** to define detection logic without deep programming knowledge.
- ✓ **Open-Source Standard:** Maintained by the community to keep up with **emerging threats**.

Sigma - Generic Signature Format for SIEM Systems



Welcome to the Sigma main rule repository. The place where detection engineers, threat hunters and all defensive security practitioners collaborate on detection rules. The repository offers more than 3000 detection rules of different type and aims to make reliable detections accessible to all at no cost.

Currently the repository offers three types of rules:

- [Generic Detection Rules](#) - Are threat agnostic, their aim is to detect a behavior or an implementation of a technique or procedure that was, can or will be used by a potential threat actor.
- [Threat Hunting Rules](#) - Are broader in scope and are meant to give the analyst a starting point to hunt for potential suspicious or malicious activity
- [Emerging Threat Rules](#) - Are rules that cover specific threats, that are timely and relevant for certain periods of time. These threats include specific APT campaigns, exploitation of Zero-Day vulnerabilities, specific malware used during an attack,...etc.

Captured from Sigma's Github repo - <https://github.com/SigmaHQ/sigma>

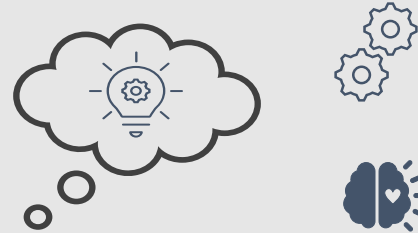
Sigma Rules at Scale The Challenge

Brief

Security teams rely on **Sigma rules** for **log-based threat detection** across SIEMs, but **manual rule creation is slow, inefficient, and does not scale.**

- **Threat intelligence is growing rapidly**, yet security teams struggle to convert insights into **actionable detection rules.**
- AI-based solutions **can assist**, but **current models lack precision**, leading to **errors in rule logic, syntax, and ATT&CK mappings.**

What are the challenges?



Can we automate the process to generate a sigma rules?

Harnessing the Power of AI?

Writing Sigma Rules Manually Is Slow and Inefficient

- **Threat intelligence is growing rapidly**, but manually converting it into Sigma rules is **too slow.**
- **Example:** A junior detection engineer takes **3 working days** to process a security blog and write **~20 Sigma rules.**
- **This approach doesn't scale**, leading to outdated detection rules.

AI Can Speed Up Rule Generation but Lacks Precision

- **LLMs can generate Sigma rules faster**, but they **often produce errors:**
 - **Incorrect MITRE ATT&CK mappings**
 - **Syntax and logic mistakes**
 - **High false positives and low reliability**
- This means **AI-generated rules still require heavy manual validation**, which defeats the purpose of automation.

5

How SigmaGen Fixes These Issues



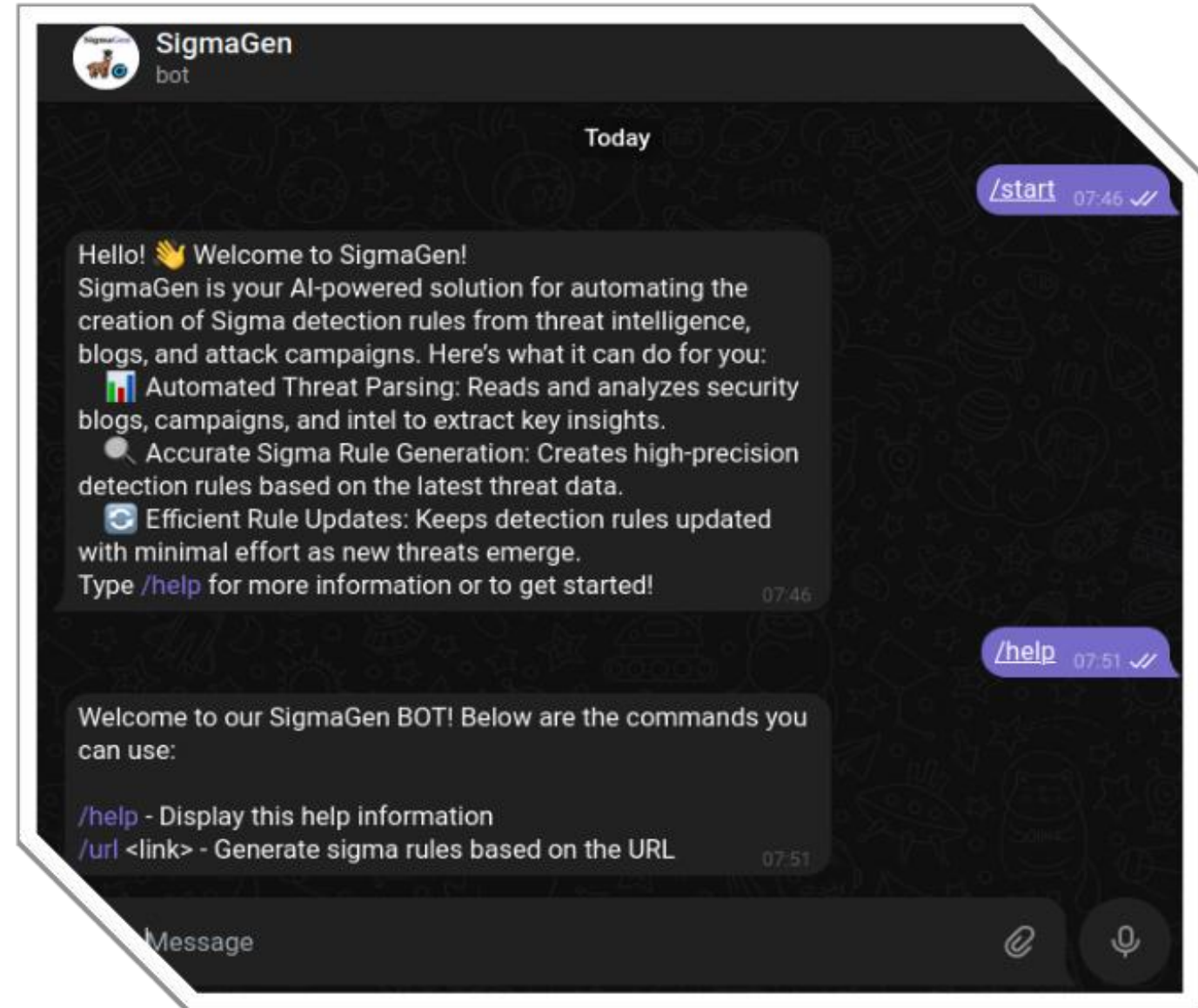
Now that we've framed the **problem and challenges**, we present **SigmaGen as the solution**.

◆ SigmaGen: AI-Powered Solution for Scalable Sigma Rule Generation

✓ **Automated Sigma Rule Generation** – AI extracts relevant attack patterns from security blogs and maps them to MITRE ATT&CK.

✓ **Fine-Tuned LLMs for Higher Accuracy** – Unlike generic AI models, SigmaGen has been trained on real-world Sigma rules.

✓ **Continuous Rule Updates** – AI ensures that new threats are detected faster by reducing the time required for rule creation.



Captured of SigmaGen chatbot

1&2. Multi-Source Data Processing

✚ Extracts data from:

- Security blogs, intelligence reports, malware analysis, and attack campaign documentation.

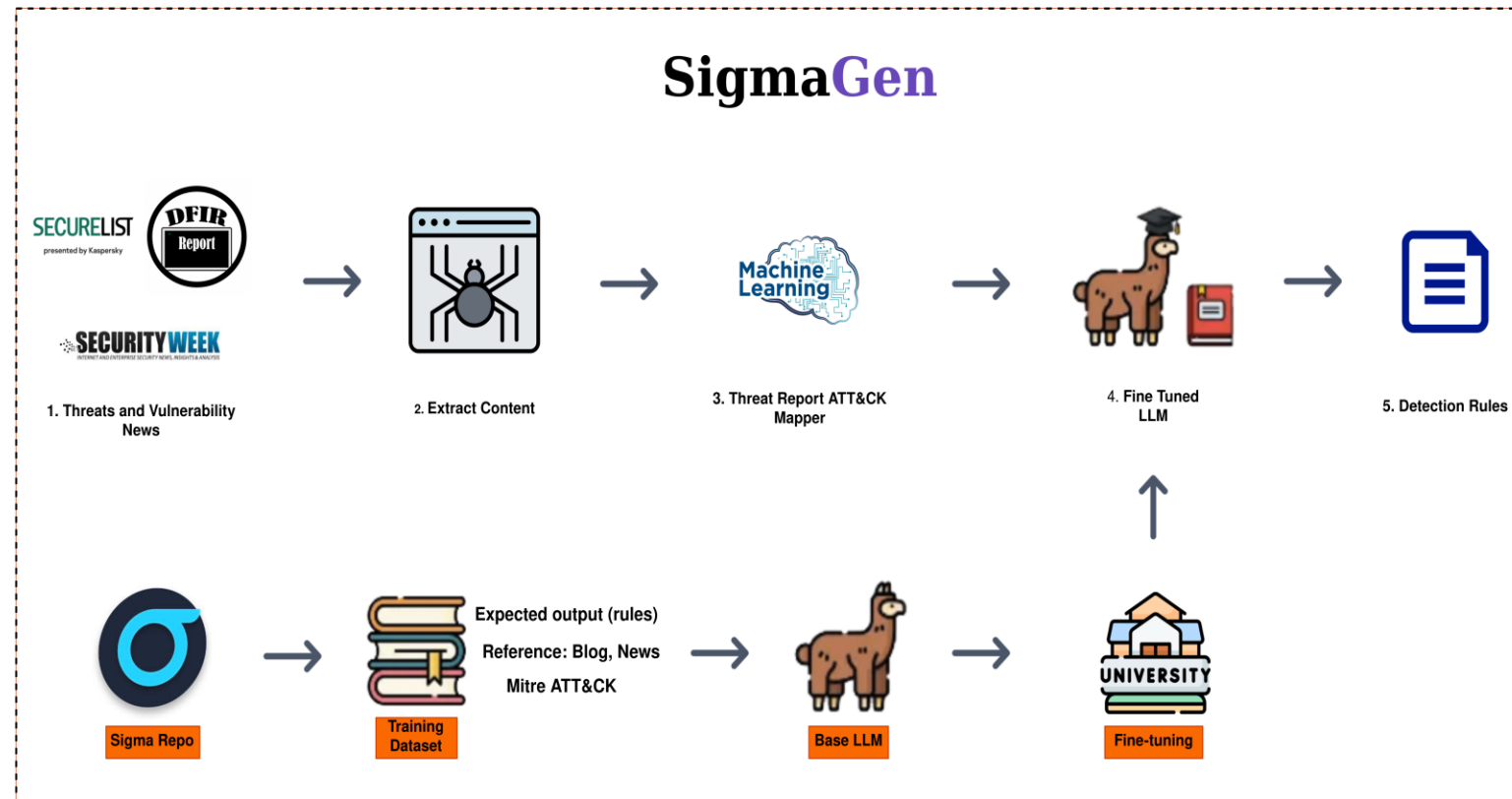
3. AI-Powered Threat Data Extraction & MITRE ATT&CK Mapping

- Use open-source platform **Threat Report ATT&CK Mapper (TRAM)**
- ✚ **How it Works:**
 - Uses **natural language processing (NLP)** and **transformer models** to extract key indicators (IoCs, behaviors).
 - **Applies machine learning models** to map extracted threats to **MITRE ATT&CK techniques**.

4. AI-Powered Sigma Rule Generation

✚ How it Works:

- Uses a **fine-tuned LLM model**, trained on real-world Sigma rules.
- Ensures that Sigma rules are **structured correctly**, preventing syntax and logic errors.



SigmaGen's High-level Design

TRAM Github repo: <https://github.com/center-for-threat-informed-defense/tram>

AI Fine-Tuning Approach



1. Curated Training Dataset

What We Did:

- Built a dataset of **200 high-quality, manually crafted Sigma rules** covering various MITRE ATT&CK techniques.
- Included **real-world detection rules** from **security blogs, DFIR reports, and open-source Sigma repositories**.
- Ensured **balanced representation** across **different attack techniques and platforms (Windows, Linux, Cloud, etc.)**.

2. AI-Guided MITRE ATT&CK Mapping

How It Works:

- Uses **natural language processing (NLP)** and **named entity recognition (NER)** to extract adversary behaviors from security blogs.
- Cross-references** extracted behaviors with the **MITRE ATT&CK knowledge base** using **vector similarity search**.
- Applies **confidence scoring** to reduce **misclassification of techniques** (one of the biggest issues with AI-generated rules).

```
{
  "messages": [
    {
      "role": "system",
      "content": "You are an expert in Sigma rule creation, following the official Sigma rule specification. You will receive summarized information from threat/vulnerability blogs, news, or reports as input, along with references to tactics and techniques from MITRE ATT&CK. Your task is to output a complete, syntactically correct Sigma rule in YAML format, adhering to the Sigma schema ({{ $schema: \"https://json-schema.org/draft/2020-12/schema#\", \"title\": \"Sigma rule specification V2.0.0 (2024-08-08)\", \"type\": \"object\", \"required\": [\"title\", \"logsources\", \"detection\"], \"properties\": {\"title\": {\"type\": \"string\", \"maxLength\": 256, \"description\": \"A brief title for the rule that should contain what the rules is supposed to detect\"}, \"id\": {\"type\": \"string\", \"description\": \"A globally unique identifier for the Sigma rule. This is recommended to be a UUID v4, but not mandatory.\"}, \"format\": \"uuid\"}, \"related\": {\"type\": \"array\", \"description\": \"A list of related Sigma rules to keep track of the relationships\"}"
    }
  ]
}
```

System Prompt: Assigned Role + Sigma Specification

```
"role": "user",
"content": "Instantly share code, notes, and snippets. ngentilkiwi/dcsync-dcshadow. splunkSecret\n\nsourcetype=XmlWinEventLog:Security AND EventCode=4662 AND NOT (SubjectUserSid=\\\"AUTORITE NT\\\\\\\\*\\\" OR SubjectDomainName=\\\"Window Manager\\\")\n\n\n(ObjectType=\\\"%{19195a5b-6da0-11d0-afd3-00c04fd930c9}\\\" OR ObjectType=\\\"domainDNS\\\")\n\n\nAND\n\n(Properties=\\\"*Replicating Directory Changes All*\\\" OR Properties=\\\"*{1131f6ad-9c07-11d1-f79f-00c04fc2dcd2}*\\\" OR Properties = \\\"*{9923a32a-3607-11d2-b9be-0000f87a36b2}*\\\" OR Properties = \\\"*{1131f6ac-9c07-11d1-f79f-00c04fc2dcd2}*\\\")\n\n\n)\n\n\nrename _time AS DStime, SubjectUserSid AS DSUserSid, SubjectDomainName AS DSDomainName, SubjectUserName AS DSUserName, SubjectLogonId AS DSLogonId, ObjectType AS DSObjectType, ObjectName AS DSObjectName, Properties AS DSProperties status AS DSStatus\n\n\n| join type=left Computer, DSLogonId\n\n\n[\\n\\nsearch sourcetype=XmlWinEventLog:Security AND EventCode=4624 NOT (TargetUserSid=\\\"AUTORITE NT\\\\\\\\*\\\" OR TargetDomainName=\\\"Window Manager\\\")\n\n\n| rename _time AS LogonTime,
```

User Prompt includes: Blog Extracted + MITRE Technique

```
205 {
206   "messages": [
207     {
208       "role": "system",
209       "content": "You are an expert in Sigma rule creation, followin
210     {
211       "role": "system",
212       "content": "You are an expert in Sigma rule creation, followin
213     {
214       "role": "system",
215       "content": "You are an expert in Sigma rule creation, followin
```

Over 200 prompts of real-world Sigma rules

```
"role": "assistant",
"content": "title: Possible DC Shadow Attack\nid: 32e19d25-4aed-4860-a55a-be99cb0bf7ed\nrelated:\n  - id: 611eab06-a145-4dfa-a295-3ccc5c20f59a\n  type: derived\nstatus: test\ndescription: Detects DCShadow via create new SPN\nreferences:\n  - https://twitter.com/gentilkiwi/status/1003236624925413376\n  - https://gist.github.com/gentilkiwi/dcc132457408cf11ad2061340dcb53c2\n  - https://web.archive.org/web/2018023014709/https://blog.alsid.eu/dcshadow-explained-4510f52fc19d?gi=c426ac876c48\nauthor: Ilyas Ochkov, oscd.community, Chakib Gzenayi (@Chak092), Hosni Mribah\nndate: 2019-10-25\nmodified: 2022-10-17\ntags:\n  - attack.credential-access\n  - attack.t1207\nlogsources:\n  product: windows\n  service: security\n  definition:
```

Assistant Output: Sigma Rule

Examples of SigmaGen-Generated Rules



Setup

Deployment *
Create new deployment
gpt-4o-mini-2024-07-18-ft-fd36cfabc0234177b7e9dcaf9447fcba (version:1)

System message Add your data PREVIEW Parameters

```
"role": "system",
"content": [
  {
    "type": "text",
    "text": "You are an expert in Sigma rule creation, following the official Sigma rule specification. You will receive summarized information from threat/vulnerability blogs, news, or reports as input, along with references to tactics and techniques from MITRE ATT&CK. Your task is to output a complete, syntactically correct Sigma rule in YAML format, adhering to the Sigma schema (https://json-schema.org/draft/2020-12/schema#). The rule must be in the following format: title: Sigma rule specification V2.0.0 (2024-08-08) type: object required: [title, logsource, detection, properties] title: A brief title for the rule that should contain what the rule is supposed to detect id: A globally unique identifier for the Sigma rule. This is recommended to be a UUID v4, but not mandatory. format: uuid related: A list of related Sigma rules to keep track of the relationships between detections. This can be used to indicate that a rule is derived from another rule, or that a rule has been obsoleted by another rule. items: A list of related Sigma rules to keep track of the relationships between detections. This can be used to indicate that a rule is derived from another rule, or that a rule has been obsoleted by another rule. id: A globally unique identifier for the Sigma rule. This is recommended to be a UUID v4, but not mandatory. format: uuid type: string oneOf: [const, derived, description] description: The rule was derived from the referred rule or rules, which may remain active. const: obsolete The rule obsoletes the referred rule or rules, which aren't used anymore. const: merged The rule was merged from the referred rules. title: A brief title for the rule that should contain what the rule is supposed to detect id: A globally unique identifier for the Sigma rule. This is recommended to be a UUID v4, but not mandatory. format: uuid related: A list of related Sigma rules to keep track of the relationships between detections. This can be used to indicate that a rule is derived from another rule, or that a rule has been obsoleted by another rule. items: A list of related Sigma rules to keep track of the relationships between detections. This can be used to indicate that a rule is derived from another rule, or that a rule has been obsoleted by another rule. id: A globally unique identifier for the Sigma rule. This is recommended to be a UUID v4, but not mandatory. format: uuid type: string oneOf: [const, derived, description] description: The rule was derived from the referred rule or rules, which may remain active. const: obsolete The rule obsoletes the referred rule or rules, which aren't used anymore. const: merged The rule was merged from the referred rules."
  ]
}
```

NG1.bat

This batch script contained three lines of ngrok configuration:

```
1 @echo off
2 ngrok config add-authtoken 2
3 ngrok tcp 3389
```

The ngrok config add-authtoken command saves the ngrok authtoken to the configuration file. While ngrok tcp can be used to start a TCP tunnel, which forwards all traffic on a public port to a local server (RDP in this case), it is also extremely useful for exposing services that run non-HTTP traffic (SSH, SIP, RDP, RTSP, GRPC, game servers, etc).

Extracted MITRE technique: T1090

title: Ngrok Command Line Execution
id: 8e0fbc04-6e4d-4c4b-9f88-d0bcfc5bc2e1
status: test
description: Detects execution of Ngrok command line (ngrok.exe) to create a TCP tunnel
references:
- https://thedfirreport.com/2024/02/06/real-intrusions-by-real-attackers-the-truth-behind-the-intrusion/
date: 2024-02-06
author: SigmaGen
tags:
- attack.execution
- attack.t1090
logsource:
product: windows
category: process_creation
detection:
selection:
Image|endswith: '\ngrok.exe'
CommandLine|contains: 'tcp '
condition: selection
falsepositives:
- Unknown
level: high

System Prompt

User Prompt includes:

1. Extracted Blog

2. MITRE Technique

Cost Breakdown: AI Processing & Hosting



Region: North Central US

Model: GPT 4o-mini

Training Costs:

- Tokens per prompt: ~96,000 tokens per prompt
- Uploaded prompts (200 prompts): 19,200,000 tokens = \$64
- Total dataset (1000 prompts): 96,000,000 tokens = \$320

Processing Assumptions:

- We assume that the system will process updates every week.
- Each update will require the hosting to be online for 4 hours, processing approximately 200 prompts per update.
- Over the course of a 4-week month, this results in the processing of 800 prompts per month.

Category	Token Usage/Prompt	Cost/1000 Tokens (\$)	Cost/Prompt (\$)	Estimate Prompts/Month	Cost/Month (\$)	Cost/Year (\$)
Input	13,359	0.000165	0.002204235	800	1.763	21.160
Output	370	0.00066	0.0002442	800	0.195	2.344
Hosting	N/A	N/A	N/A	N/A	27.2	326.4
Total					29\$	350\$

✓ AI-Powered Sigma Rule Generation Saves Time

- Traditional rule-writing is **slow & manual**; SigmaGen **automates** this process.
- AI reduces Sigma rule creation **from days to minutes**.

✓ Accuracy Matters: Fine-Tuned AI Ensures High-Quality Rules

- Generic LLMs **fail at MITRE ATT&CK mapping & syntax**.
- SigmaGen's **fine-tuned model** ensures **correct mappings & valid rule structures**.

✓ Continuous Rule Updates Keep Up with Evolving Threats

- Adversaries evolve **quickly**—static rules become ineffective.
- SigmaGen **automates frequent rule updates**, improving SOC agility.

✓ Cost-Effective & Scalable AI Processing

- AI **needs to be scalable and affordable** for **enterprise-wide deployment**.
- SigmaGen's optimized **\$29/month processing cost** ensures accessibility.

✓ The Future: Fully Automated Detection & Deployment

- **SigmaStream integration** will enable **end-to-end rule automation** from **generation to SIEM deployment**.
- AI-powered validation will further **reduce false positives and improve detection rates**.





THANK YOU

