Python and AI
for Modern Threat Intelligence

NSR 국가보안기술연구소
National Security Research Institute

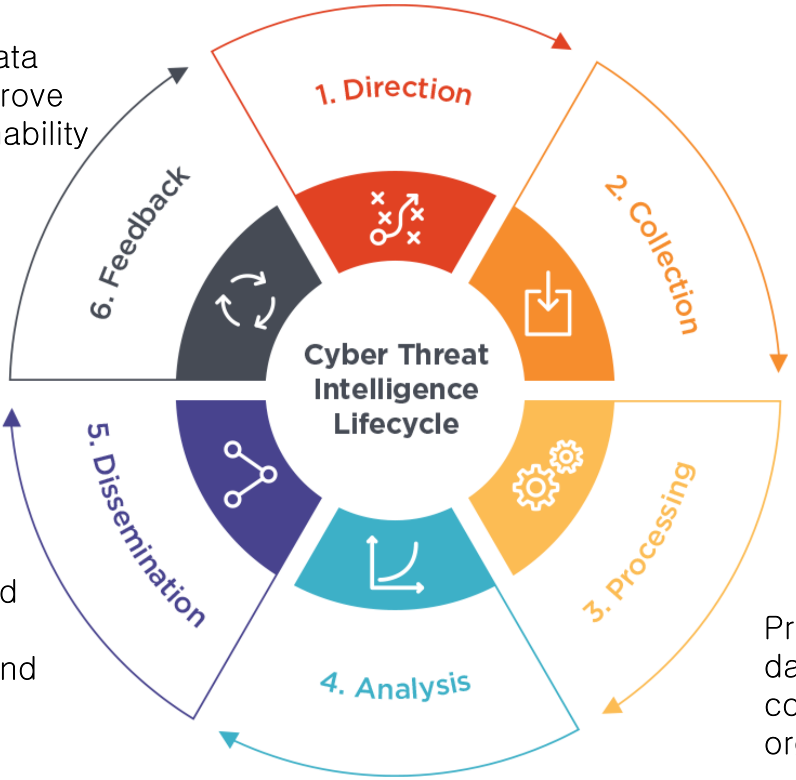
Cyber Threat Intelligence

Threat Intelligence Lifecycle

Direction involves identifying critical assets and processes requiring protection while determining specific intelligence needs for the organization.

Feedback gathers input from data consumers to continuously improve the value, accuracy, and actionability of the threat intelligence.

Collection accumulates relevant information from internal and externa sources to address key intelligence requirements.



Dissemination distributes analyzed intelligence to appropriate stakeholders in optimal formats and frequencies.

Processing transforms raw collected data into a format that can be easily consumed throughout the organization.

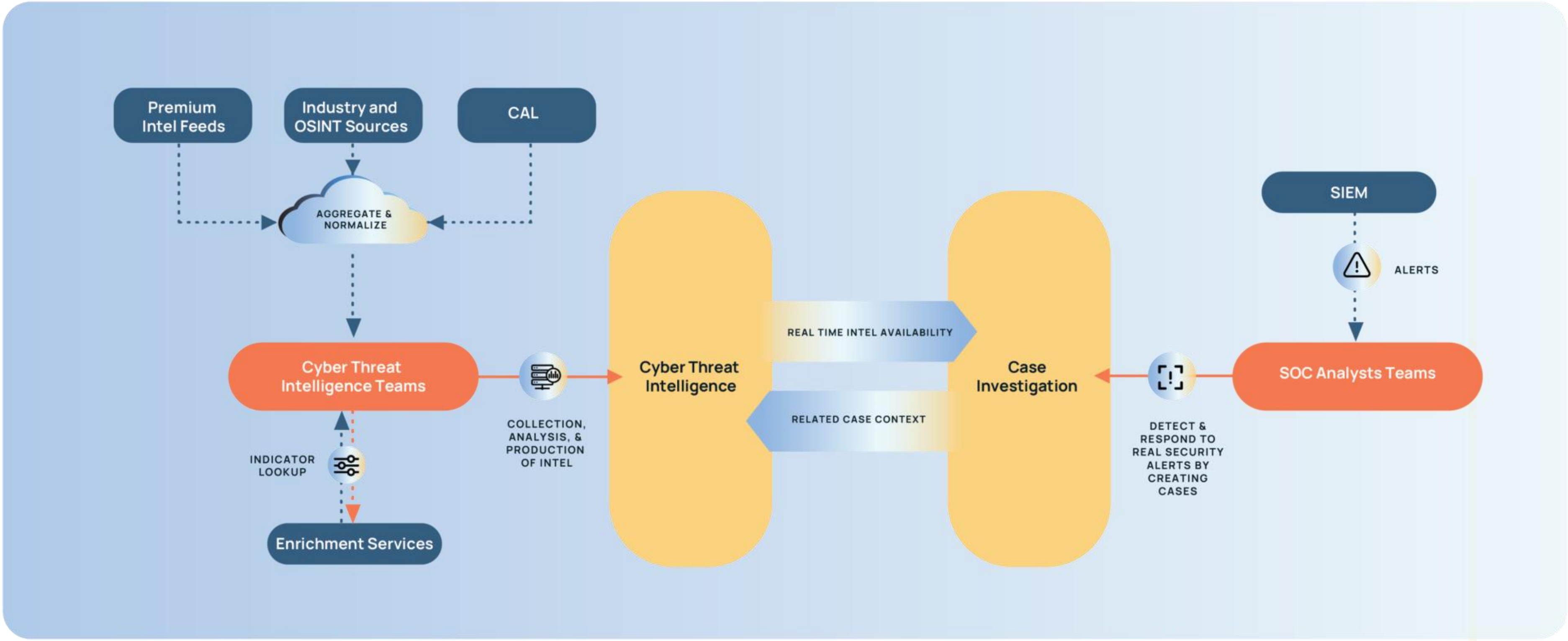
Analysis converts processed information into actionable intelligence for decision-making with clear findings and recommendations.

https://threat.media/definition/what-is-the-threat-intelligence-lifecycle/

Threat Intelligence with Security Teams

- Security Operation Center
  - Enhances detection and response with threat context
- Incident Response
  - Provides attack patterns and indicators for effective response
- Vulnerability Management
  - Prioritizes patching based on active exploit information
- Malware Analysis
  - Offers context on malware families and associated threat actors
- Business Operations
  - Informs risk-based decisions and resource allocation
- System Engineering & IT
  - Guides security control implementation against specific threats

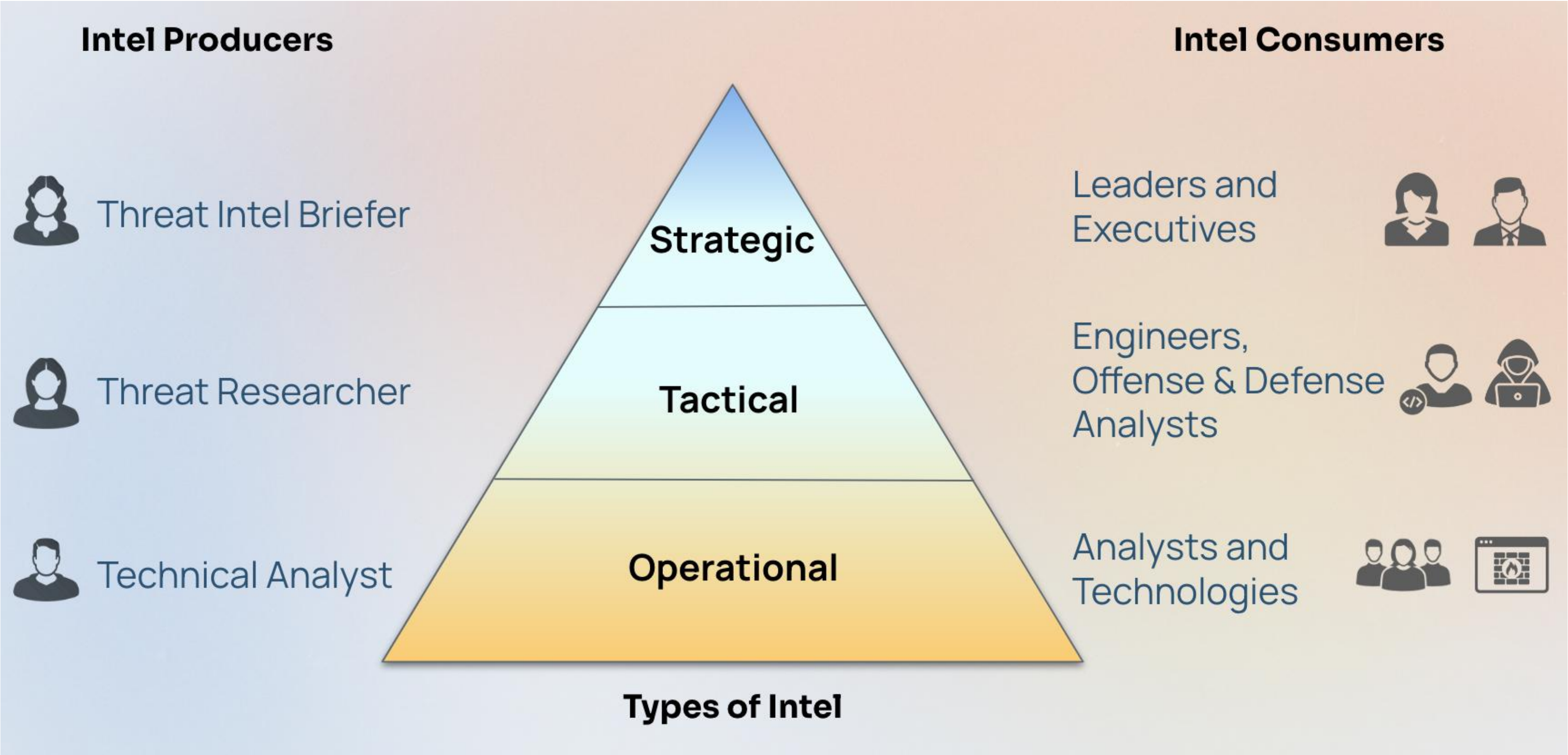
Threat Intelligence Use Case (SOC)



Premium Intel Feeds
Industry and OSINT Sources
CAL
AGGREGATE & NORMALIZE
Cyber Threat Intelligence Teams
INDICATOR LOOKUP
Enrichment Services
COLLECTION, ANALYSIS, & PRODUCTION OF INTEL
Cyber Threat Intelligence
REAL TIME INTEL AVAILABILITY
RELATED CASE CONTEXT
Case Investigation
DETECT & RESPOND TO REAL SECURITY ALERTS BY CREATING CASES
SIEM
ALERTS
SOC Analysts Teams

https://threatconnect.com/blog/improve-soc-efficiency-with-intelligence-powered-security-operations/

Type of Threat Intelligence

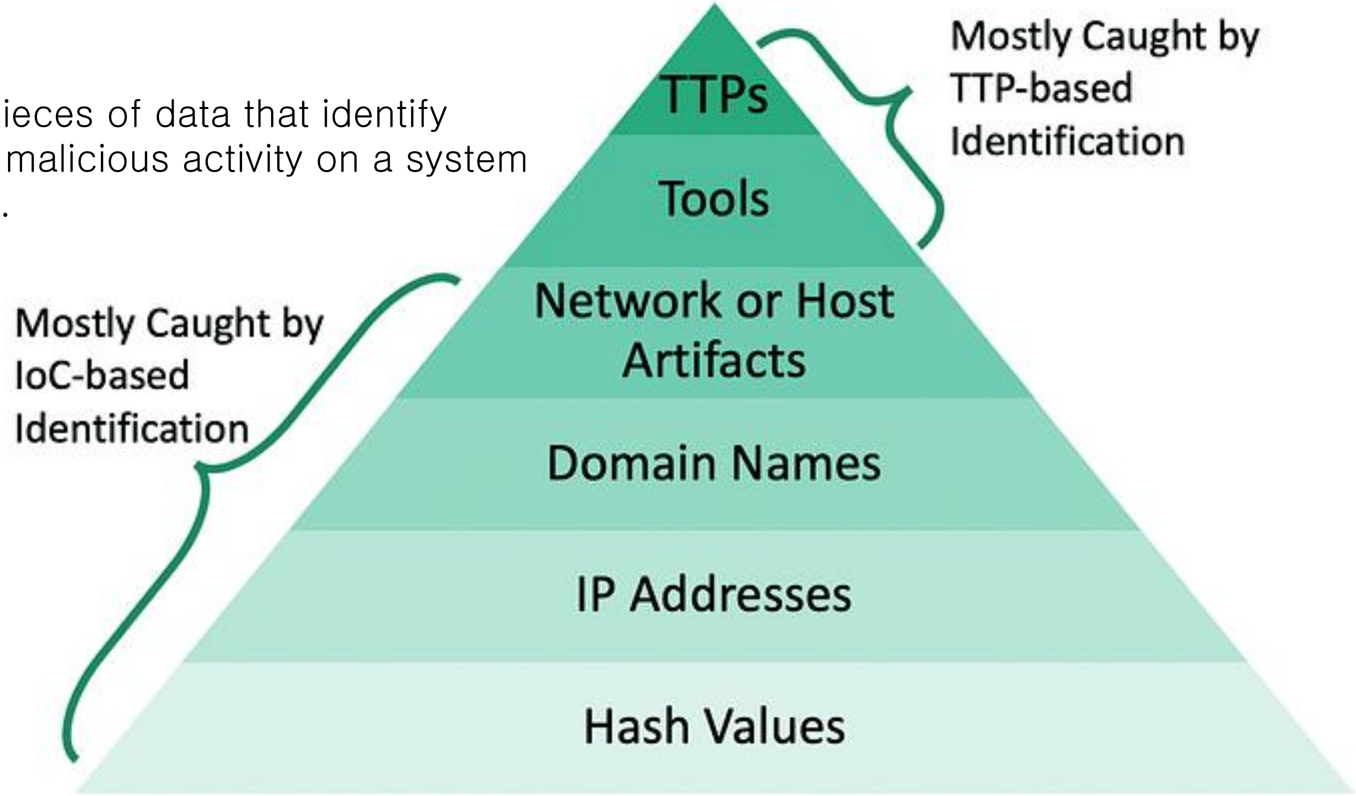


https://threatconnect.com/blog/the-7-tenets-of-threat-intelligence-operations-tenet-6-integrates-and-automates-threat-intel/

The Pyramid of Pain

TTPs describe how threat actors conduct their operations, offering a broader context than IOCs.

IOCs are pieces of data that identify potentially malicious activity on a system or network.


TTPs
Tools
Network or Host Artifacts
Domain Names
IP Addresses
Hash Values
Mostly Caught by TTP-based Identification
Mostly Caught by IoC-based Identification

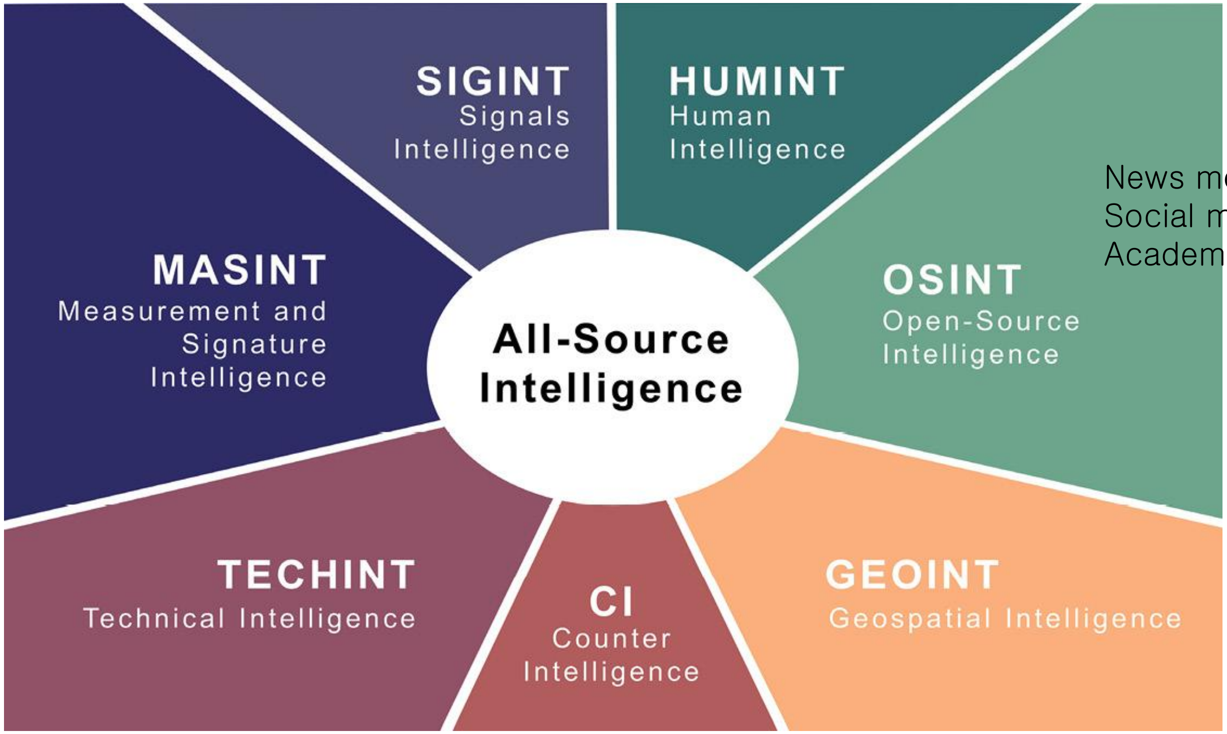
https://ibrahimakkdag.medium.com/tips-for-threat-hunters-comparison-of-indicators-of-compromise-iocs-and-tactics-techniques-and-47bc268fe7ff

All Source Intelligence

Communications intelligence,
Electronic intelligence, …

Interrogations, Agent networks, …

Radar signatures,
Nuclear intelligence,
Materials analysis, …



News media,
Social media,
Academic publications, …

Weapons systems analysis,
Military equipment capabilities,
Technical specifications, …

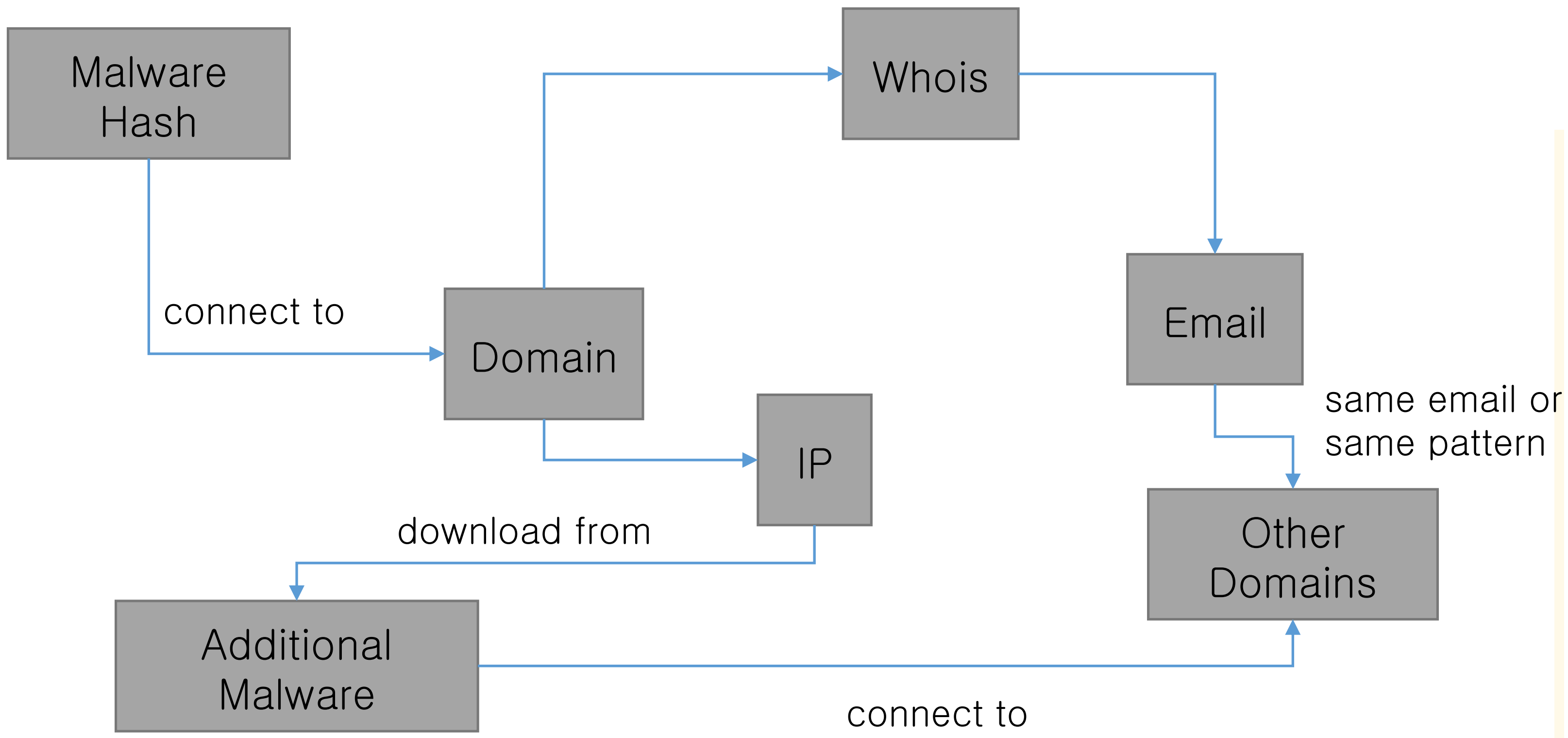
Double agent operations,
Surveillance detection,
Penetration testing, …

Satellite imagery,
Aerial photography,
Mapping/terrain data, …

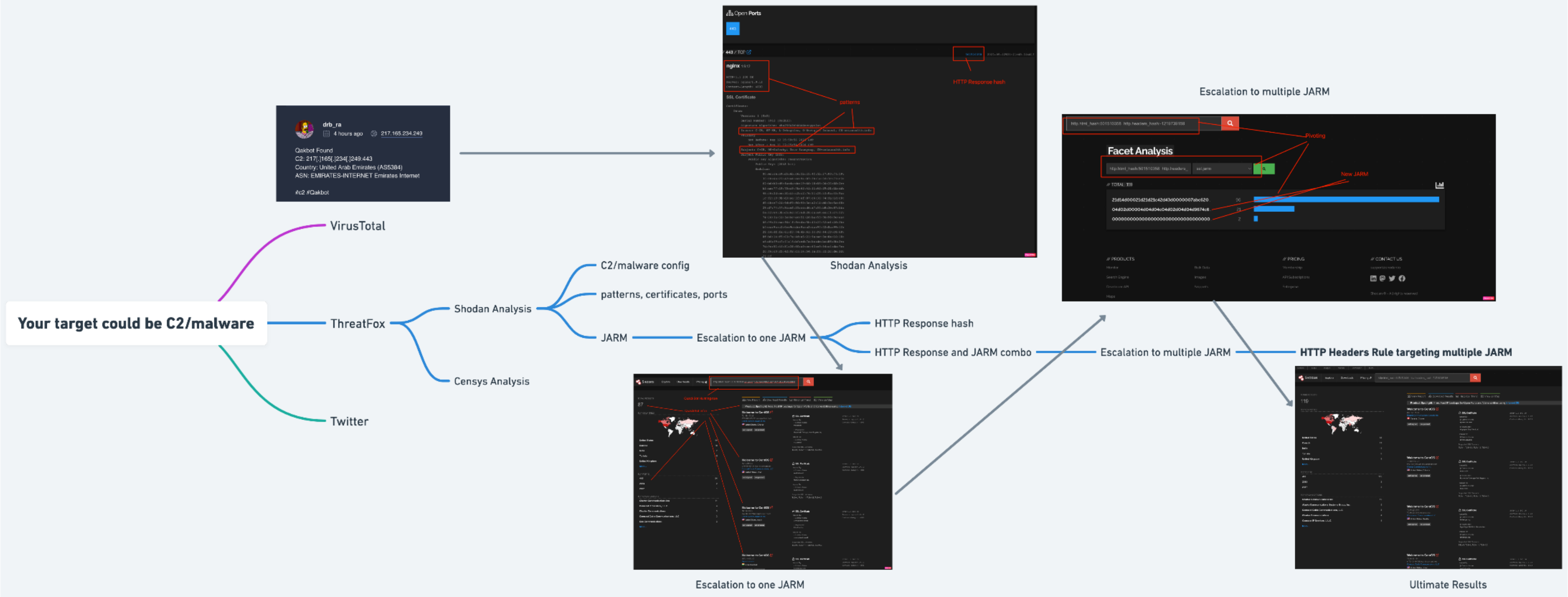
https://tjaglcs.army.mil/Periodicals/The-Army-Lawyer/tal-2022-issue-4/Post/4255/No-1-Distinguishing-Between-Operational-and-Intelligence-Activities

Pivoting

- Pivoting involves leveraging initial security discoveries to broaden investigative scope



Pivoting



drb_ra
4 hours ago 217.165.234.249

Qakbot Found
C2: 217[.]165[.]234[.]249:443
Country: United Arab Emirates (AS5384)
ASN: EMIRATES-INTERNET Emirates Internet

#c2 #Qakbot

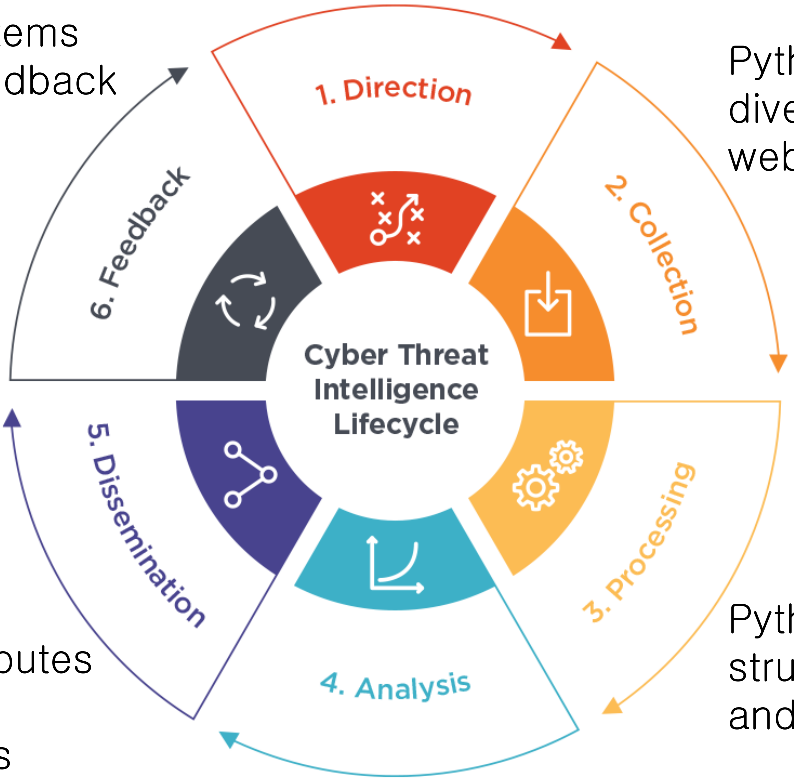
https://detect.fyi/hunting-malicious-infrastructure-using-jarm-and-http-response-bb4a039d4119

Python for Threat Intelligence

Python automates security needs assessment through scripted tools that identify critical assets

Python builds interactive systems for collecting stakeholder feedback on intelligence products

Python collects intelligence from diverse sources using APIs and web scraping libraries



Python transforms raw data into structured formats using Pandas and regex for analysis

Python generates and distributes visual intelligence reports through automated channels

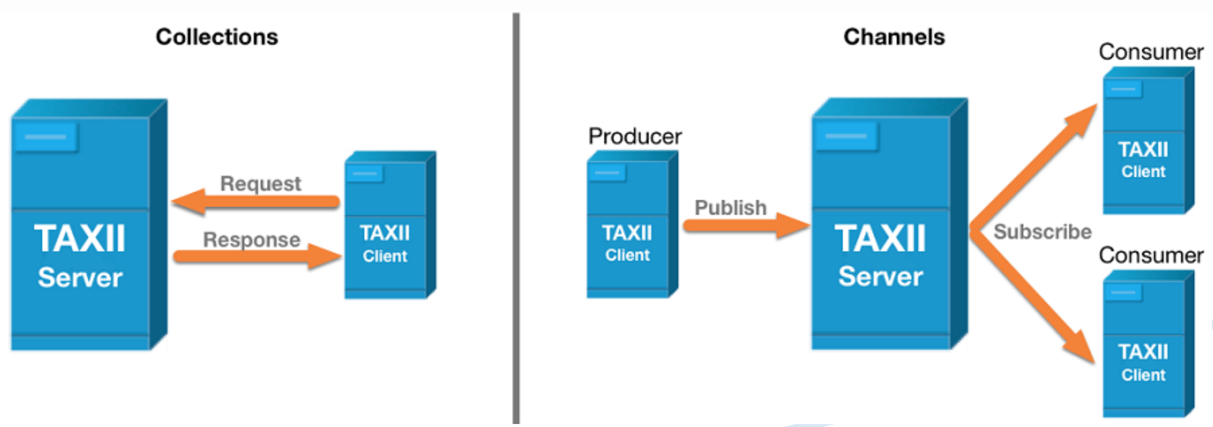
Python analyzes threats with data science libraries to uncover patterns and actionable insights

STIX & TAXII

- STIX (Structured Threat Information Expression)
  - Standardized language for cyber threat intelligence representation
  - Enables organizations to share structured threat information using objects, relationships, and properties
- TAXII (Trusted Automated Exchange of Intelligence Information)
  - Transport mechanism for sharing cyber threat intelligence
  - Defines APIs for STIX content exchange between producers and consumers of threat information

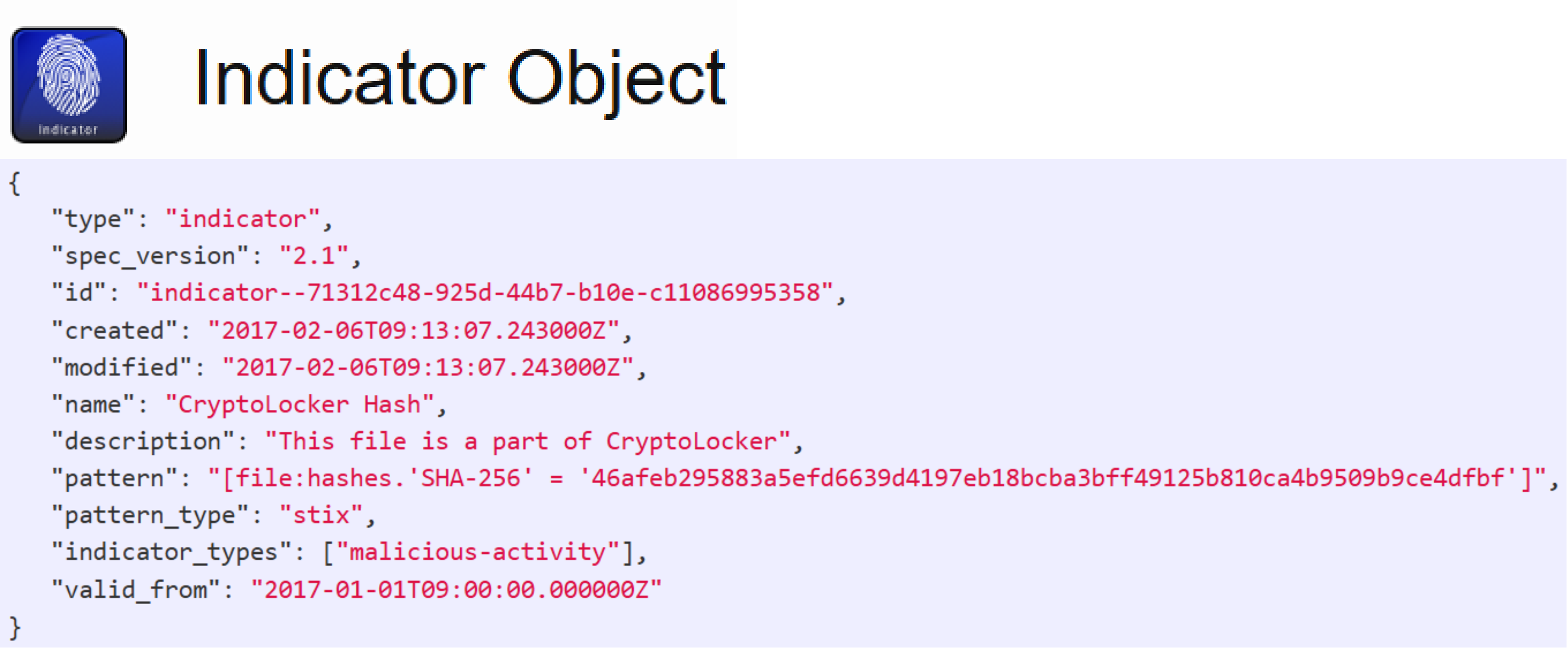
STIX Scenario

- STIX Producer: Company A
  - Creates Indicator SDO with CryptoLocker hash pattern and Malware SDO with malware details
  - Links objects with "indicates" Relationship SRO and shares bundle via TAXII server
- STIX Consumer: Company B
  - Receives Company A's intelligence and finds matching malware on their network
  - Creates Sighting SRO to report detection and publishes back to community via TAXII



STIX Scenario

- Contains detection info: name, pattern, type, validity timeframe
- Company A example: CryptoLocker SHA-256 hash marked as "malicious-activity"

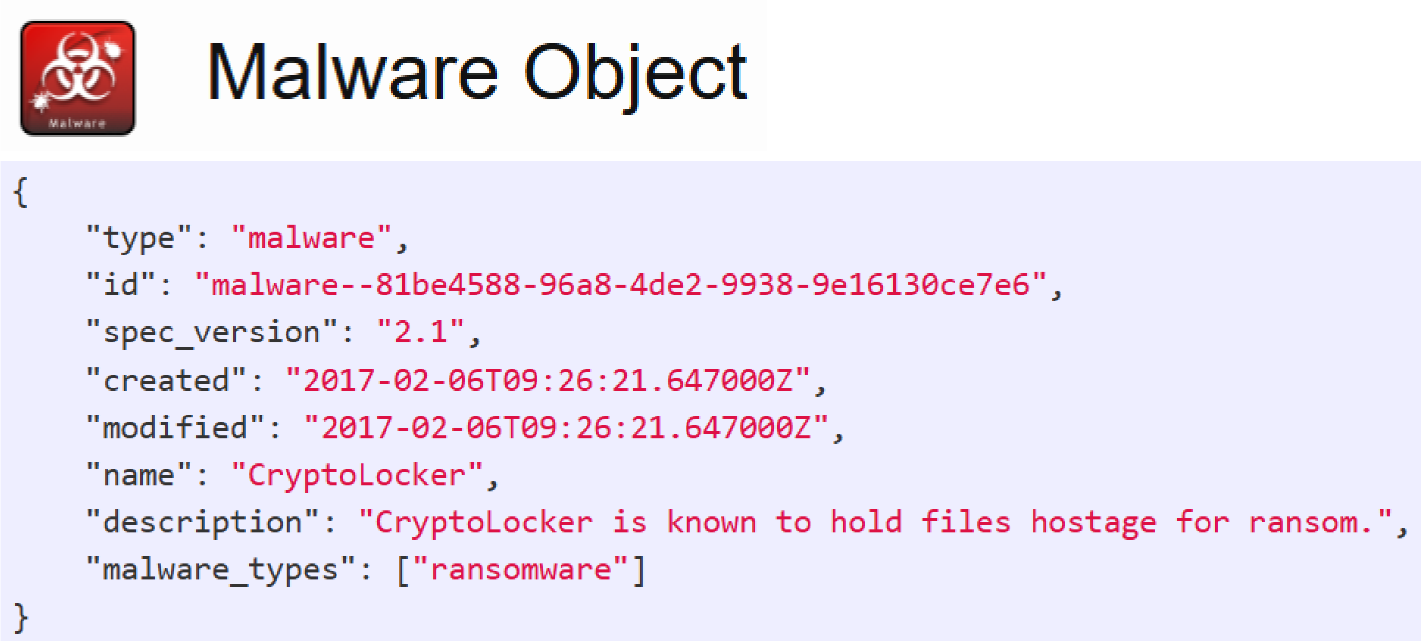


Indicator Object

```
{
   "type": "indicator",
   "spec_version": "2.1",
   "id": "indicator--71312c48-925d-44b7-b10e-c11086995358",
   "created": "2017-02-06T09:13:07.243000Z",
   "modified": "2017-02-06T09:13:07.243000Z",
   "name": "CryptoLocker Hash",
   "description": "This file is a part of CryptoLocker",
   "pattern": "[file:hashes.'SHA-256' = '46afeb295883a5efd6639d4197eb18bcba3bff49125b810ca4b9509b9ce4dfbf']",
   "pattern_type": "stix",
   "indicator_types": ["malicious-activity"],
   "valid_from": "2017-01-01T09:00:00.000000Z"
}
```

STIX Scenario

- Contains required properties: type, spec_version, id, created
- Provides descriptive context about the malware

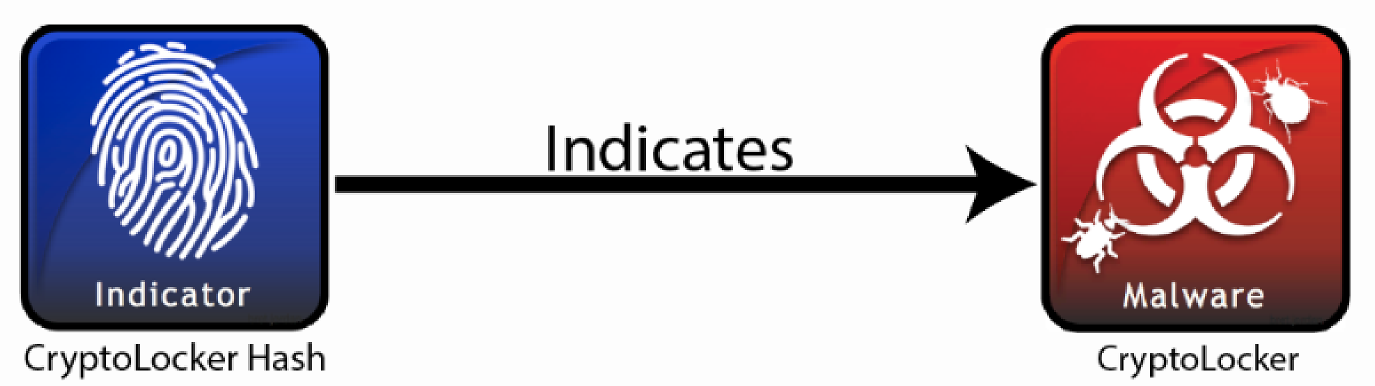


Malware Object

```
{
    "type": "malware",
    "id": "malware--81be4588-96a8-4de2-9938-9e16130ce7e6",
    "spec_version": "2.1",
    "created": "2017-02-06T09:26:21.647000Z",
    "modified": "2017-02-06T09:26:21.647000Z",
    "name": "CryptoLocker",
    "description": "CryptoLocker is known to hold files hostage for ransom.",
    "malware_types": ["ransomware"]
}
```

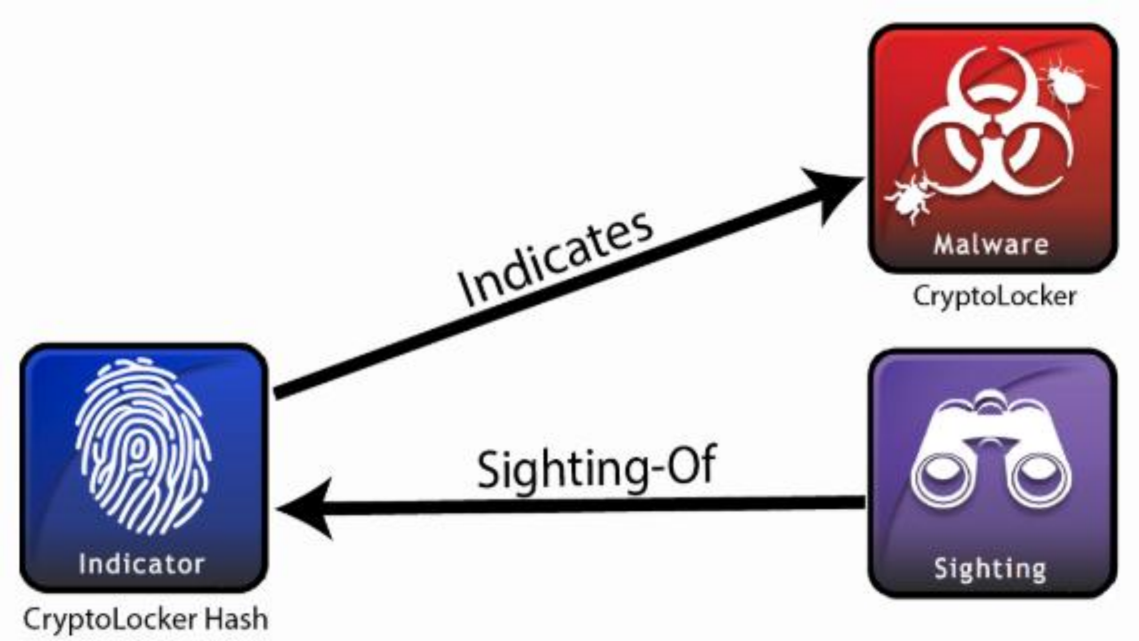
STIX Scenario

- Links Indicator to Malware SDOs using source_ref, target_ref, and relationship_type
- Company A example: Indicator "indicates" Malware, connecting detection pattern to CryptoLocker

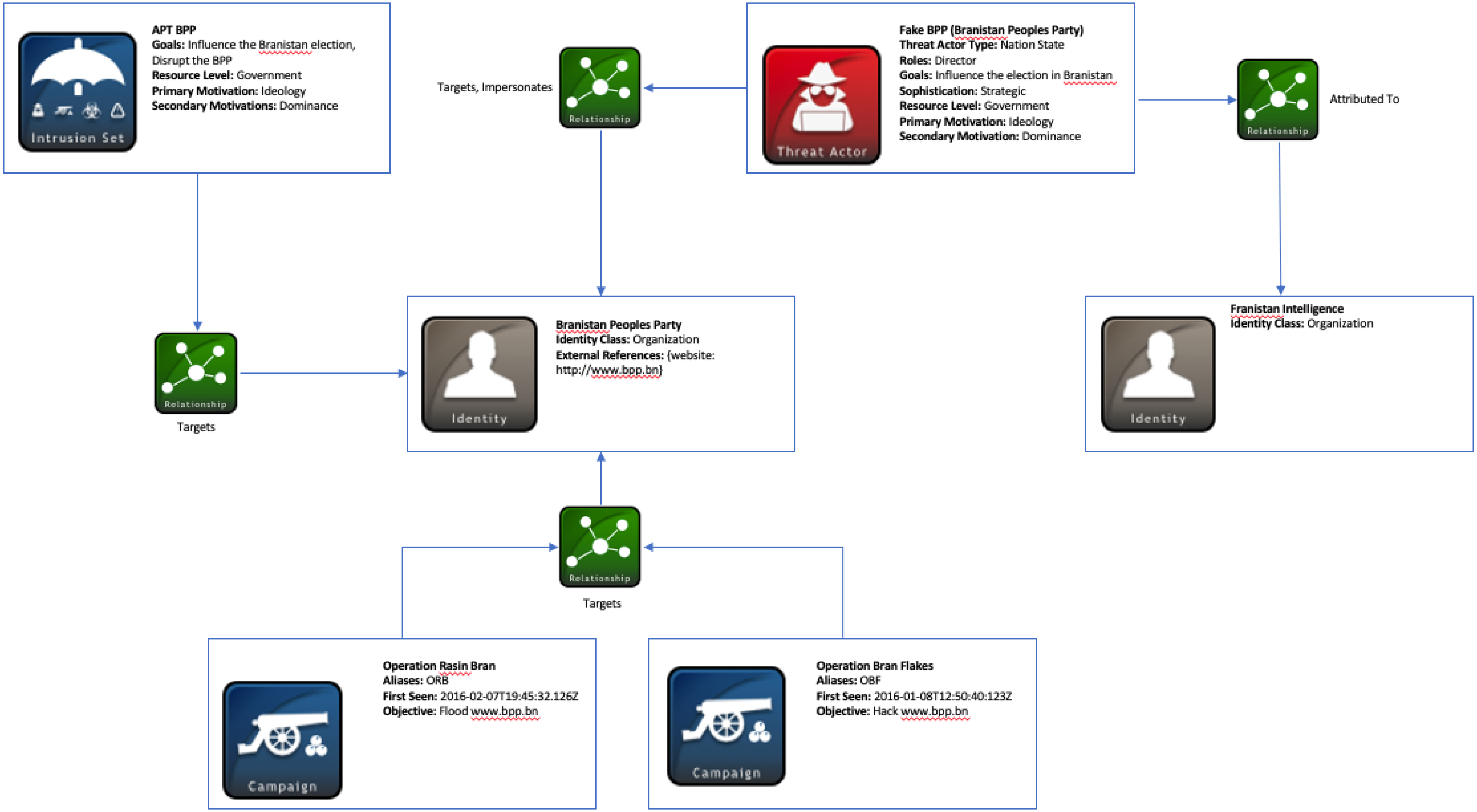


STIX Scenario

- Reports when an organization observes a previously shared threat indicator
- Company B example: Documents detection of Company A's CryptoLocker hash on their network



STIX Scenario



APT BPP
Goals: Influence the Branistan election, Disrupt the BPP
Resource Level: Government
Primary Motivation: Ideology
Secondary Motivations: Dominance

Fake BPP (Branistan Peoples Party)
Threat Actor Type: Nation State
Roles: Director
Goals: Influence the election in Branistan
Sophistication: Strategic
Resource Level: Government
Primary Motivation: Ideology
Secondary Motivation: Dominance

Branistan Peoples Party
Identity Class: Organization
External References: {website: http://www.bpp.bn}

Franistan Intelligence
Identity Class: Organization

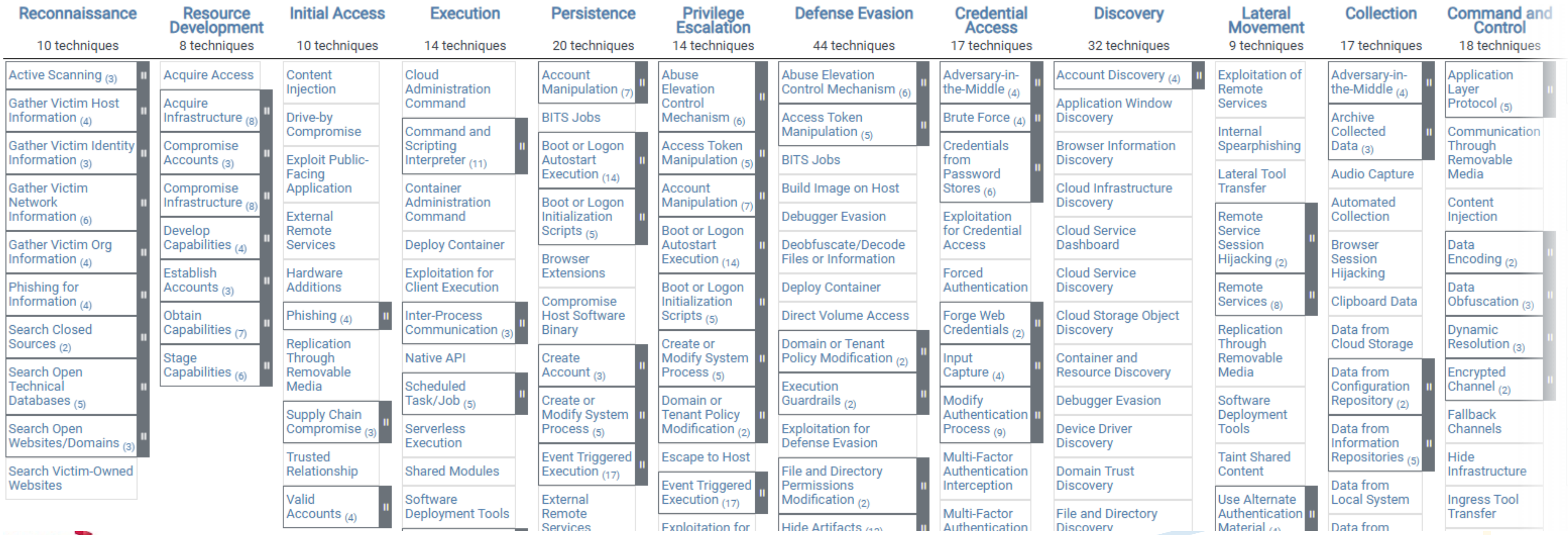
Operation Rasin Bran
Aliases: ORB
First Seen: 2016-02-07T19:45:32.126Z
Objective: Flood www.bpp.bn

Operation Bran Flakes
Aliases: OBF
First Seen: 2016-01-08T12:50:40:123Z
Objective: Hack www.bpp.bn

https://oasis-open.github.io/cti-documentation/examples/defining-campaign-ta-is

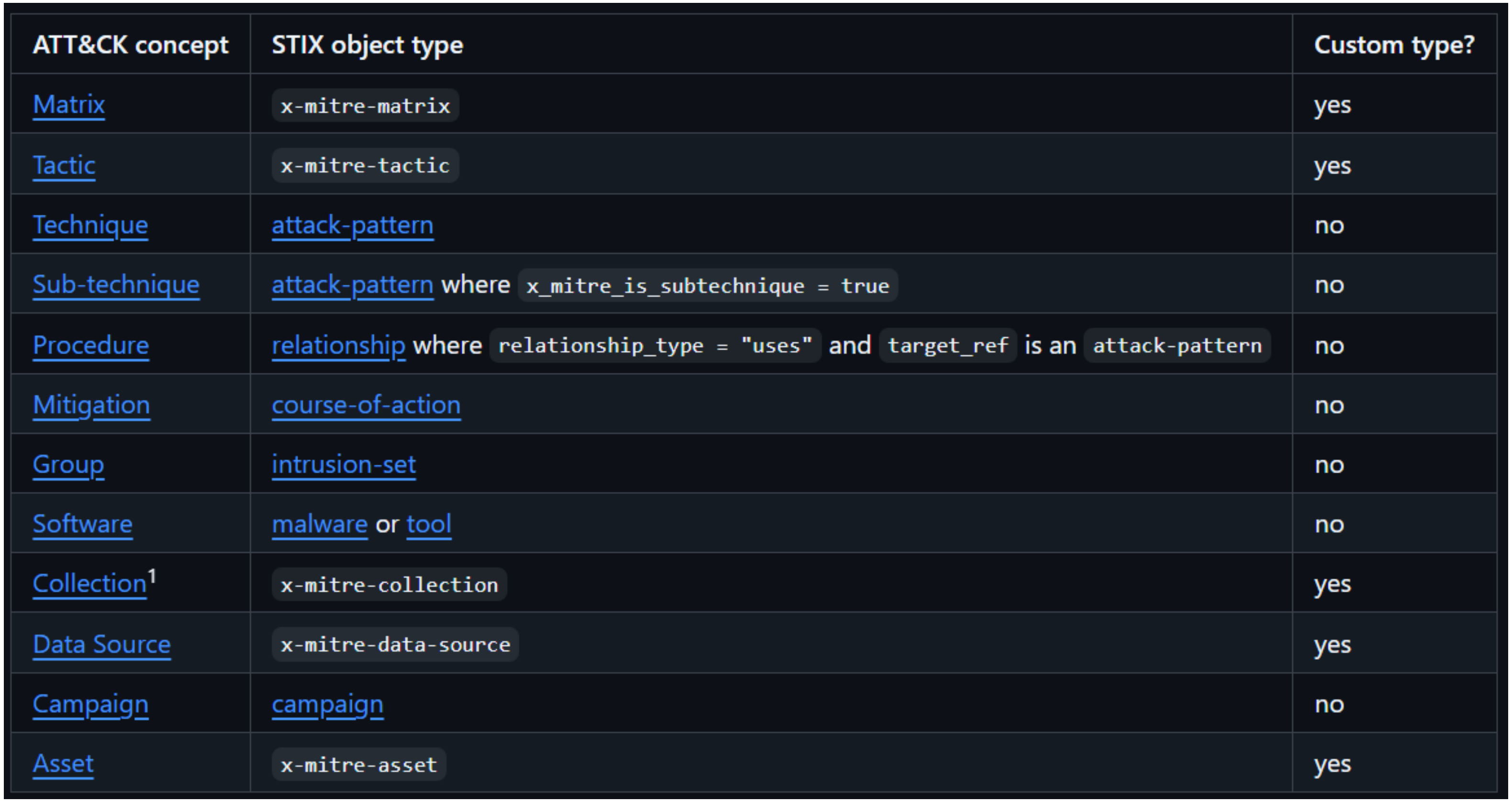
MITRE ATT&CK

- Free knowledge base of real-world adversary tactics and techniques
- Foundation for threat modeling across government and industry
- ATT&CK data is in STIX format for easy sharing and analysis



Reconnaissance 10 techniques	Resource Development 8 techniques	Initial Access 10 techniques	Execution 14 techniques	Persistence 20 techniques	Privilege Escalation 14 techniques	Defense Evasion 44 techniques	Credential Access 17 techniques	Discovery 32 techniques	Lateral Movement 9 techniques	Collection 17 techniques	Command and Control 18 techniques
Active Scanning (3)	Acquire Access	Content Injection	Cloud Administration Command	Account Manipulation (7)	Abuse Elevation Control Mechanism (6)	Abuse Elevation Control Mechanism (6)	Adversary-in-the-Middle (4)	Account Discovery (4)	Exploitation of Remote Services	Adversary-in-the-Middle (4)	Application Layer Protocol (5)
Gather Victim Host Information (4)	Acquire Infrastructure (8)	Drive-by Compromise	Command and Scripting Interpreter (11)	BITS Jobs	Access Token Manipulation (5)	Access Token Manipulation (5)	Brute Force (4)	Application Window Discovery	Internal Spearphishing	Archive Collected Data (3)	Communication Through Removable Media
Gather Victim Identity Information (3)	Compromise Accounts (3)	Exploit Public-Facing Application	Container Administration Command	Boot or Logon Autostart Execution (14)	Account Manipulation (7)	BITS Jobs	Credentials from Password Stores (6)	Browser Information Discovery	Lateral Tool Transfer	Audio Capture	Content Injection
Gather Victim Network Information (6)	Compromise Infrastructure (8)	External Remote Services	Deploy Container	Boot or Logon Initialization Scripts (5)	Boot or Logon Autostart Execution (14)	Build Image on Host	Exploitation for Credential Access	Cloud Infrastructure Discovery	Remote Service Session Hijacking (2)	Automated Collection	Data Encoding (2)
Gather Victim Org Information (4)	Develop Capabilities (4)	Hardware Additions	Exploitation for Client Execution	Browser Extensions	Boot or Logon Initialization Scripts (5)	Debugger Evasion	Forced Authentication	Cloud Service Dashboard	Remote Services (8)	Browser Session Hijacking	Data Obfuscation (3)
Phishing for Information (4)	Establish Accounts (3)	Phishing (4)	Inter-Process Communication (3)	Compromise Host Software Binary	Create or Modify System Process (5)	Deobfuscate/Decode Files or Information	Forge Web Credentials (2)	Cloud Service Discovery	Replication Through Removable Media	Clipboard Data	Dynamic Resolution (3)
Search Closed Sources (2)	Obtain Capabilities (7)	Replication Through Removable Media	Native API	Create Account (3)	Domain or Tenant Policy Modification (2)	Deploy Container	Input Capture (4)	Cloud Storage Object Discovery	Software Deployment Tools	Data from Cloud Storage	Encrypted Channel (2)
Search Open Technical Databases (5)	Stage Capabilities (6)	Supply Chain Compromise (3)	Scheduled Task/Job (5)	Create or Modify System Process (5)	Escape to Host	Direct Volume Access	Modify Authentication Process (9)	Container and Resource Discovery	Taint Shared Content	Data from Configuration Repository (2)	Fallback Channels
Search Open Websites/Domains (3)		Trusted Relationship	Serverless Execution	Event Triggered Execution (17)	Event Triggered Execution (17)	Domain or Tenant Policy Modification (2)	Multi-Factor Authentication Interception	Debugger Evasion	Use Alternate Authentication Material (4)	Data from Information Repositories (5)	Hide Infrastructure
Search Victim-Owned Websites		Valid Accounts (4)	Shared Modules	External Remote Services	Exploitation for	Execution Guardrails (2)	Multi-Factor Authentication	Device Driver Discovery		Data from Local System	Ingress Tool Transfer
			Software Deployment Tools			Exploitation for Defense Evasion		Domain Trust Discovery		Data from	
						File and Directory Permissions Modification (2)		File and Directory Discovery			
						Hide Artifacts (12)					

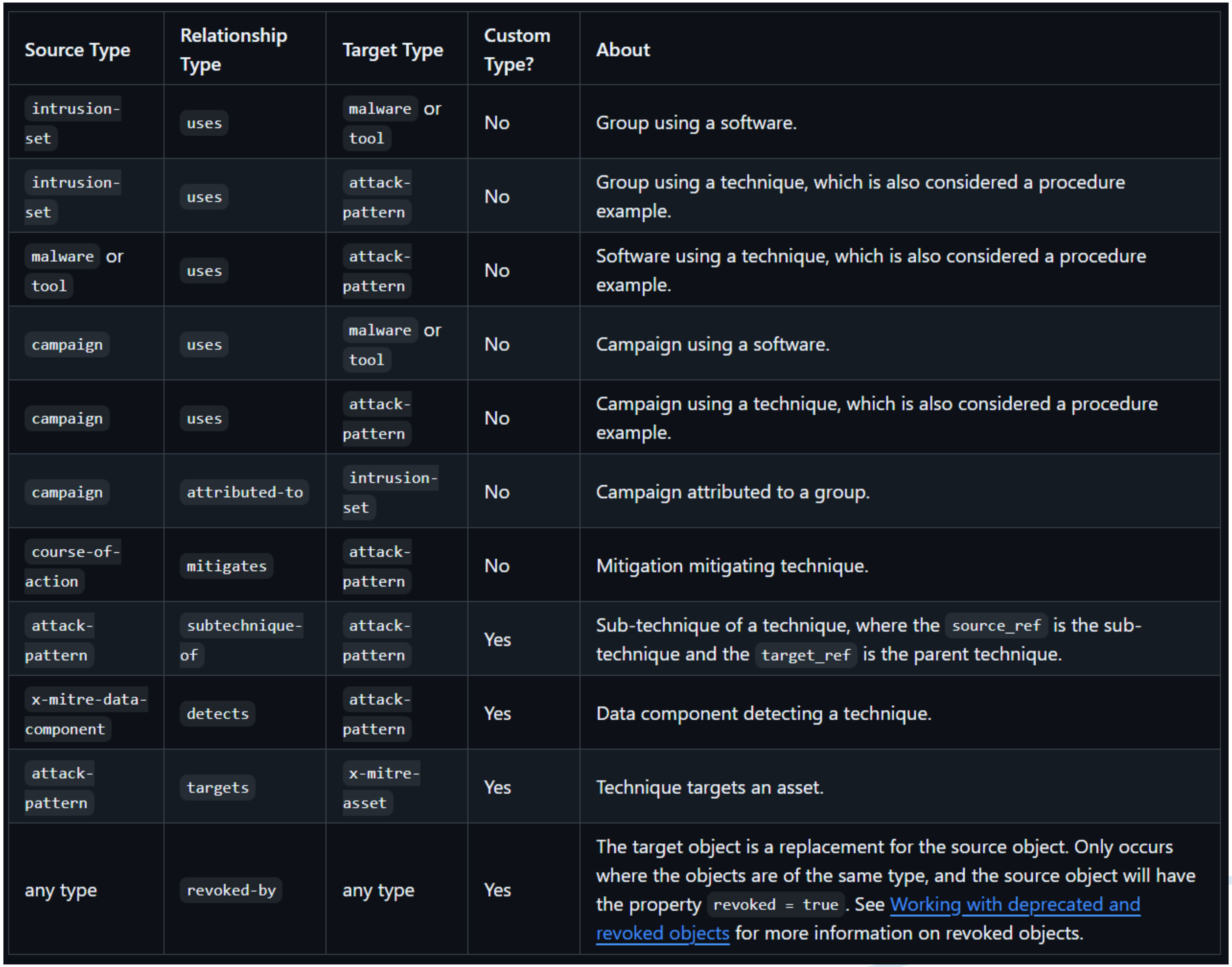
MITRE ATT&CK



ATT&CK concept	STIX object type	Custom type?
Matrix	`x-mitre-matrix`	yes
Tactic	`x-mitre-tactic`	yes
Technique	attack-pattern	no
Sub-technique	attack-pattern where `x_mitre_is_subtechnique = true`	no
Procedure	relationship where `relationship_type = "uses"` and `target_ref` is an `attack-pattern`	no
Mitigation	course-of-action	no
Group	intrusion-set	no
Software	malware or tool	no
Collection[1]	`x-mitre-collection`	yes
Data Source	`x-mitre-data-source`	yes
Campaign	campaign	no
Asset	`x-mitre-asset`	yes

https://github.com/mitre-attack/attack-stix-data/blob/master/USAGE.md

MITRE ATT&CK



Source Type	Relationship Type	Target Type	Custom Type?	About
`intrusion-set`	`uses`	`malware` or `tool`	No	Group using a software.
`intrusion-set`	`uses`	`attack-pattern`	No	Group using a technique, which is also considered a procedure example.
`malware` or `tool`	`uses`	`attack-pattern`	No	Software using a technique, which is also considered a procedure example.
`campaign`	`uses`	`malware` or `tool`	No	Campaign using a software.
`campaign`	`uses`	`attack-pattern`	No	Campaign using a technique, which is also considered a procedure example.
`campaign`	`attributed-to`	`intrusion-set`	No	Campaign attributed to a group.
`course-of-action`	`mitigates`	`attack-pattern`	No	Mitigation mitigating technique.
`attack-pattern`	`subtechnique-of`	`attack-pattern`	Yes	Sub-technique of a technique, where the `source_ref` is the sub-technique and the `target_ref` is the parent technique.
`x-mitre-data-component`	`detects`	`attack-pattern`	Yes	Data component detecting a technique.
`attack-pattern`	`targets`	`x-mitre-asset`	Yes	Technique targets an asset.
any type	`revoked-by`	any type	Yes	The target object is a replacement for the source object. Only occurs where the objects are of the same type, and the source object will have the property `revoked = true`. See Working with deprecated and revoked objects for more information on revoked objects.

https://github.com/mitre-attack/attack-stix-data/blob/master/USAGE.md

Data Collection

- API Querying
  - Uses Python scripts to retrieve structured data directly from providers
  - Gathers threat intelligence efficiently via platform APIs
- Web Scraping
  - Extracts unstructured data from websites using tools like BeautifulSoup
  - Collects IOCs from hacker forums and dark web sources
- Open Source Intelligence (OSINT)
  - Automates intelligence collection from feeds and unstructured sources
  - Monitors dark web activities and analyzes threat reports
- Offensive Threat Intelligence
  - Scans and exploits vulnerabilities in attackers' infrastructure
  - Analyzes malware commands to disable threats or infiltrate systems

API Providers

- Free/Community
  - MISP: Open-source threat intelligence platform for sharing, storing, and correlating indicators
  - PhishTank: Community-driven database of verified phishing websites
  - Validin: Historical DNS record lookup service for tracking domain changes
  - Malware Bazaar: Repository for sharing and analyzing malware samples
  - Pulsedive: Threat intelligence platform providing enriched IOC data
  - DNSQuery: DNS lookup tool for analyzing domain information
  - UrlScan: Website scanner that analyzes and detects suspicious websites
  - OTX: AlienVault's Open Threat Exchange for community threat intelligence sharing
- Enterprise
  - VirusTotal: Multi-engine malware scanning and file reputation service
  - Censys: Internet-wide scanning platform for attack surface management
  - Flare: Threat intelligence platform focused on dark web monitoring
  - Shodan: Search engine for internet-connected devices and vulnerabilities
  - MS Defender for TI: Microsoft's threat intelligence offering within Defender suite
  - Recorded Future: AI-powered threat intelligence platform with real-time risk assessment
  - Mandiant Threat Intelligence: Advanced threat intelligence with actor profiling and vulnerability research

Data Processing

- Structure Your Dataset
  - Integrate multiple source formats (CSV, JSON) into a unified, organized collection
  - Ensure field consistency through standardization processes for seamless analysis
- Language Conversion
  - Transform text between languages to enable global threat intelligence operations
  - Maintain consistent encoding protocols across multilingual data to preserve integrity
- Image Text Extraction
  - Convert visual information to text using OCR for documents and screenshots containing threat indicators
- Noise Elimination
  - Remove irrelevant data points that don't contribute meaningful intelligence value
  - Develop filtering parameters based on statistical methods and known benign patterns
- Pandas, NumPy, Elasticsearch, Apache Spark, SQLite, KQL, …

Data Analysis

- Statistics
  - Identify threat behavior patterns through correlation and regression analysis
  - Apply statistical testing to validate security event relationships
- Visualization
  - Create charts and diagrams to represent complex threat data clearly
  - Display geographical attack distributions and malware propagation timelines
- Enrich and Pivot
  - Add contextual information to raw security data from external intelligence sources
  - Examine multiple data attributes to discover hidden attack relationships
- Intelligence Identification
  - Distinguish meaningful threats from routine security alerts
  - Filter out false positives to focus analysis on genuine security risks
- Matplotlib, Bokeh, Pyvis, NetworkX, Seaborn, ggplot, Pygal, …

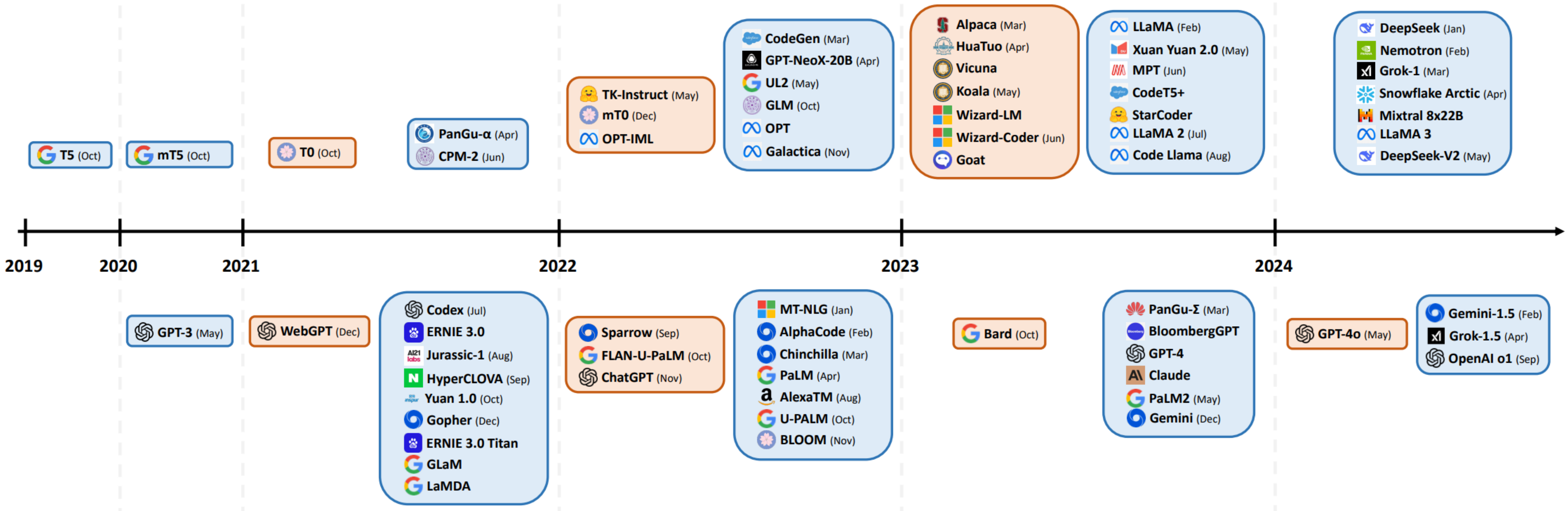
Disseminating Threat Intelligence

- Definition
  - The delivery of critical security insights to relevant decision-makers and teams.
- Tools
  - VThunting.py: Python tool for automated VirusTotal hunting and intelligence gathering
  - PyMISP: Python library for interacting with MISP threat intelligence platforms
  - PySTIX: Python implementation for working with STIX threat intelligence format
  - Discord: Messaging platform that can be leveraged for real-time threat intel distribution
  - Slack: Collaboration tool with channels and integrations for sharing security alerts
  - Jinja2: Template engine used to generate standardized threat intelligence reports
  - MS Teams: Microsoft's collaboration platform with channels for threat intelligence distribution

Generative AI and Threat Intelligence

LLM

- Definition
  - Large Language Models are advanced AI systems trained on vast amounts of text data to understand, generate, and manipulate human language across diverse tasks and contexts



T5 (Oct)

mT5 (Oct)

T0 (Oct)

PanGu-α (Apr)
CPM-2 (Jun)

TK-Instruct (May)
mT0 (Dec)
OPT-IML

CodeGen (Mar)
GPT-NeoX-20B (Apr)
UL2 (May)
GLM (Oct)
OPT
Galactica (Nov)

Alpaca (Mar)
HuaTuo (Apr)
Vicuna
Koala (May)
Wizard-LM
Wizard-Coder (Jun)
Goat

LLaMA (Feb)
Xuan Yuan 2.0 (May)
MPT (Jun)
CodeT5+
StarCoder
LLaMA 2 (Jul)
Code Llama (Aug)

DeepSeek (Jan)
Nemotron (Feb)
Grok-1 (Mar)
Snowflake Arctic (Apr)
Mixtral 8x22B
LLaMA 3
DeepSeek-V2 (May)

2019 2020 2021 2022 2023 2024

GPT-3 (May)

WebGPT (Dec)

Codex (Jul)
ERNIE 3.0
Jurassic-1 (Aug)
HyperCLOVA (Sep)
Yuan 1.0 (Oct)
Gopher (Dec)
ERNIE 3.0 Titan
GLaM
LaMDA

Sparrow (Sep)
FLAN-U-PaLM (Oct)
ChatGPT (Nov)

MT-NLG (Jan)
AlphaCode (Feb)
Chinchilla (Mar)
PaLM (Apr)
AlexaTM (Aug)
U-PALM (Oct)
BLOOM (Nov)

Bard (Oct)

PanGu-Σ (Mar)
BloombergGPT
GPT-4
Claude
PaLM2 (May)
Gemini (Dec)

GPT-4o (May)

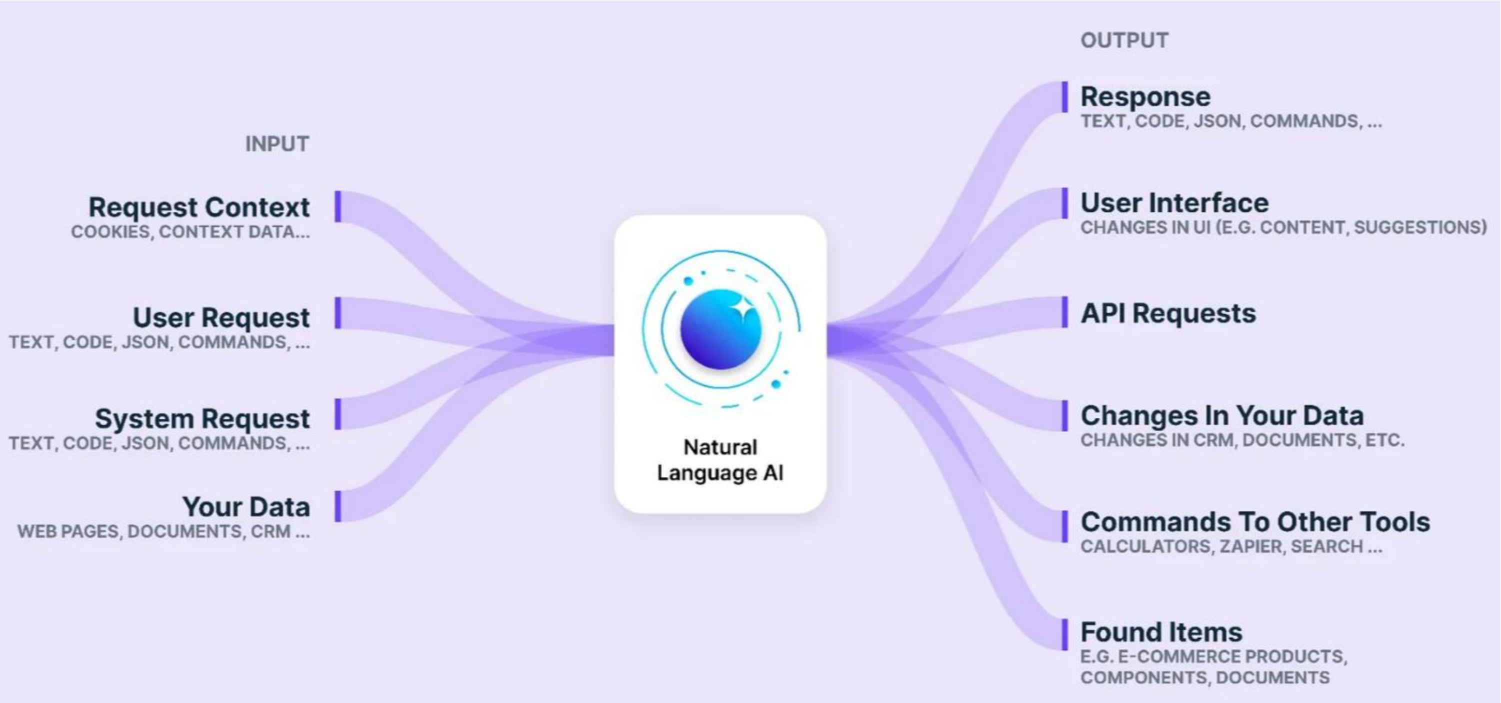
Gemini-1.5 (Feb)
Grok-1.5 (Apr)
OpenAI o1 (Sep)

- Blue = pre-trained, Orange = instruction-tuned
- Top = open-source, Bottom = closed-source

Naveed, Humza, et al. "A comprehensive overview of large language models." arXiv preprint arXiv:2307.06435 (2023).

LLM

- Text in, Text out!
  - LLMs were focused solely on processing and generating plain text, mainly used for tasks like summarization, translation, and classification

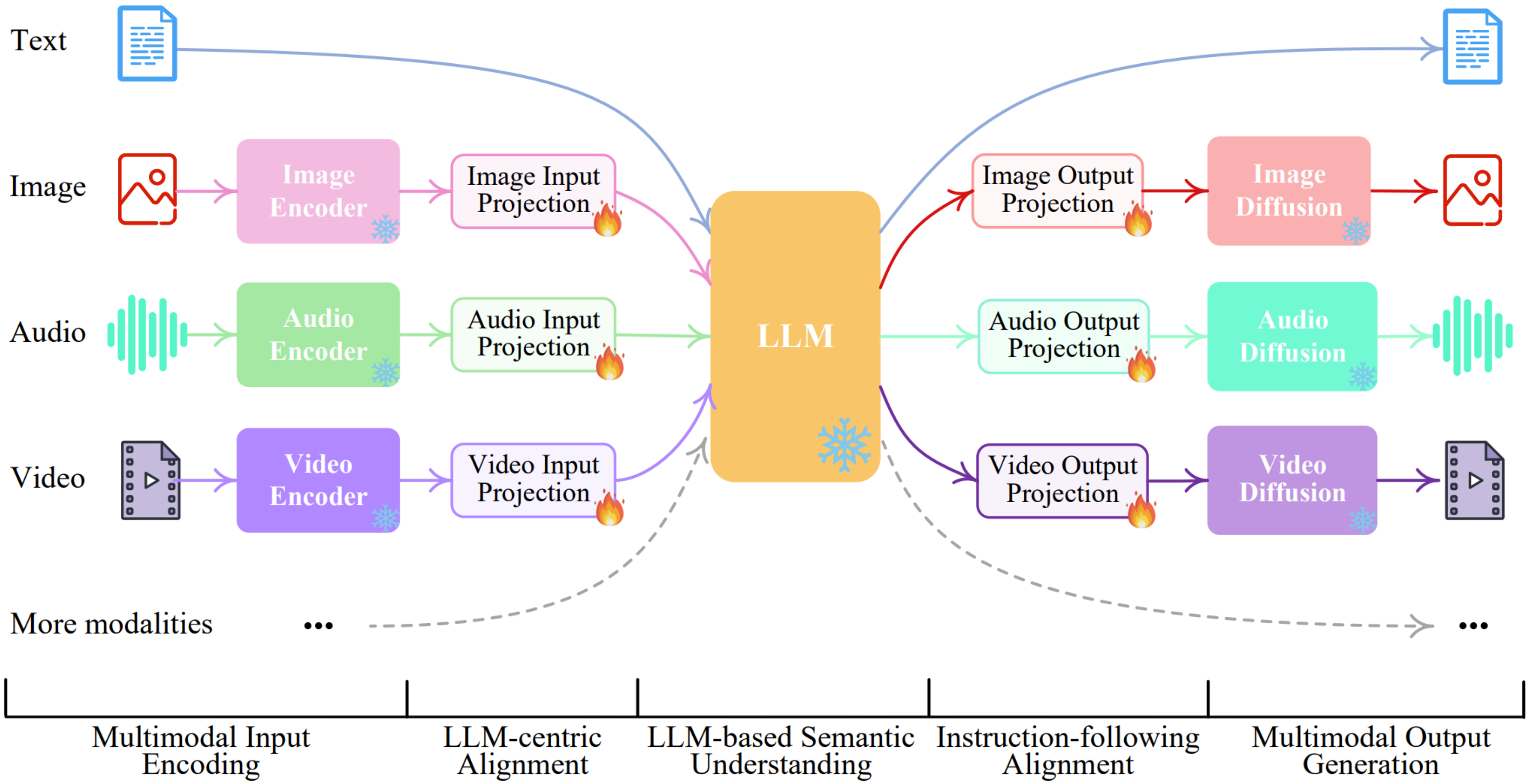


INPUT
Request Context
COOKIES, CONTEXT DATA...
User Request
TEXT, CODE, JSON, COMMANDS, ...
System Request
TEXT, CODE, JSON, COMMANDS, ...
Your Data
WEB PAGES, DOCUMENTS, CRM ...
Natural Language AI
OUTPUT
Response
TEXT, CODE, JSON, COMMANDS, ...
User Interface
CHANGES IN UI (E.G. CONTENT, SUGGESTIONS)
API Requests
Changes In Your Data
CHANGES IN CRM, DOCUMENTS, ETC.
Commands To Other Tools
CALCULATORS, ZAPIER, SEARCH ...
Found Items
E.G. E-COMMERCE PRODUCTS, COMPONENTS, DOCUMENTS

But with the advent of function calling, LLM now have the ability to interact with external tools and APIs—unlocking far more dynamic, real-world applications

LLM

- Anything in, Anything out!



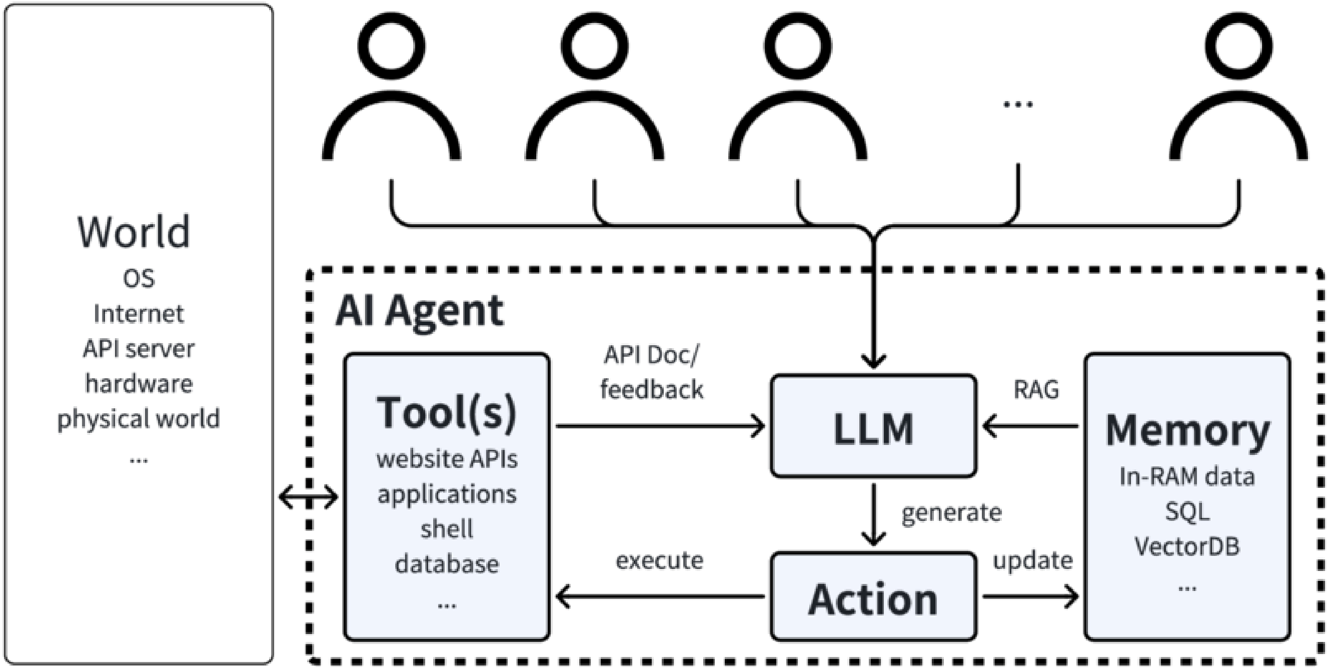
Wu, Shengqiong, et al. "Next-gpt: Any-to-any multimodal llm." Forty-first International Conference on Machine Learning. 2024.

LLM

- Limitations
  - Produces hallucinations
  - Lacks sufficient controllability
  - High training and inference costs
  - Difficulty maintaining context until maximum length
  - Struggles to update rapidly changing knowledge (temporal, regional, cultural)
  - Results difficult to interpret or verify without sources
  - Risk of sensitive information exposure (confidential data, PII)

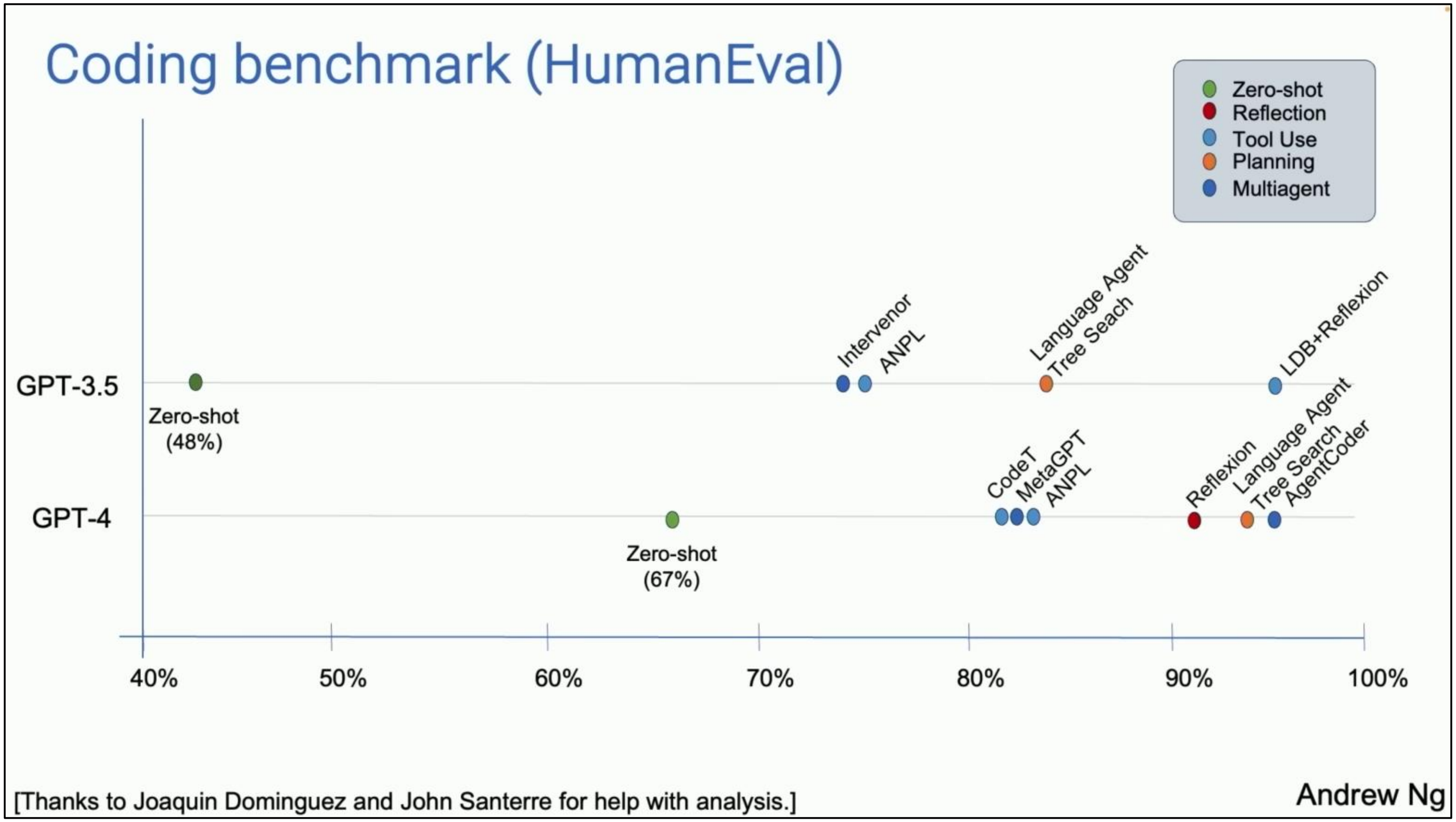
Agent

- Definition
  - LLM Agents are autonomous systems that use large language models as their reasoning engine to understand tasks, make decisions, and take actions through external tools to achieve user-defined goals



He, Yifeng, et al. "Security of ai agents." arXiv preprint arXiv:2406.08689 (2024).

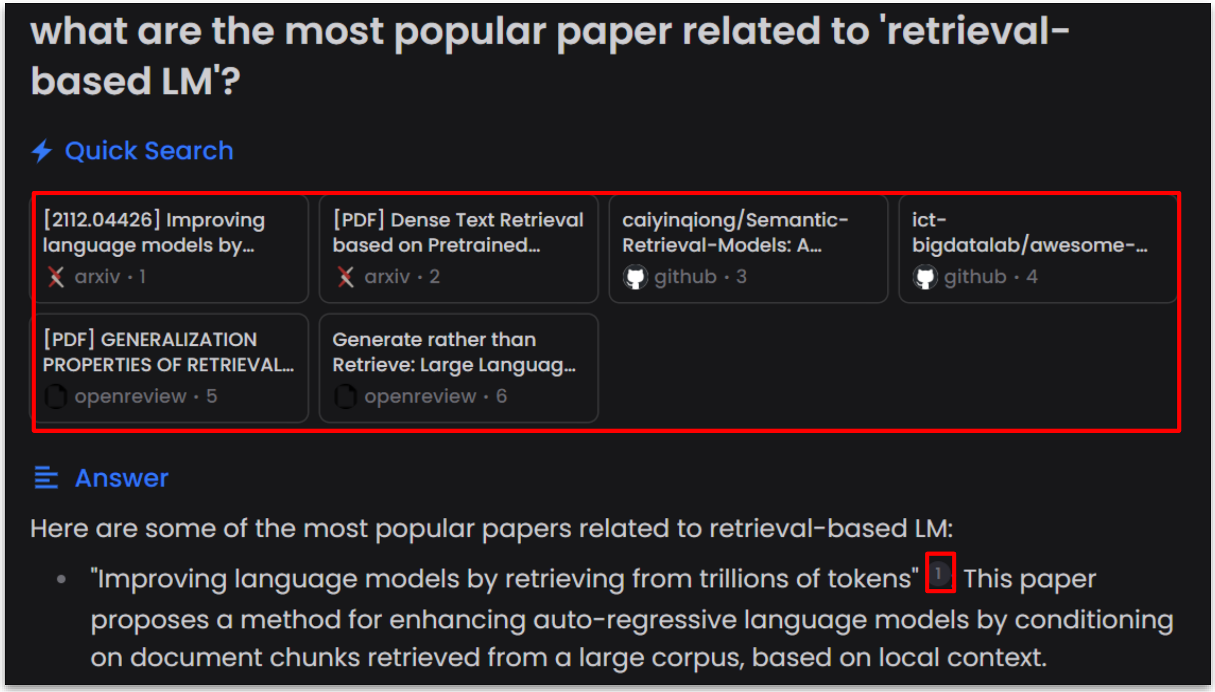
Agent



What's next for AI agentic workflows, https://youtu.be/sal78ACtGTc?si=zA7W9TnJekIBZsW2

RAG

- Definition
  - RAG combines external information retrieval with language model generation to produce more accurate and knowledge-grounded responses



what are the most popular paper related to 'retrieval-based LM'?

Quick Search

[2112.04426] Improving language models by... arxiv · 1

[PDF] Dense Text Retrieval based on Pretrained... arxiv · 2

caiyinqiong/Semantic-Retrieval-Models: A... github · 3

ict-bigdatalab/awesome-... github · 4

[PDF] GENERALIZATION PROPERTIES OF RETRIEVAL... openreview · 5

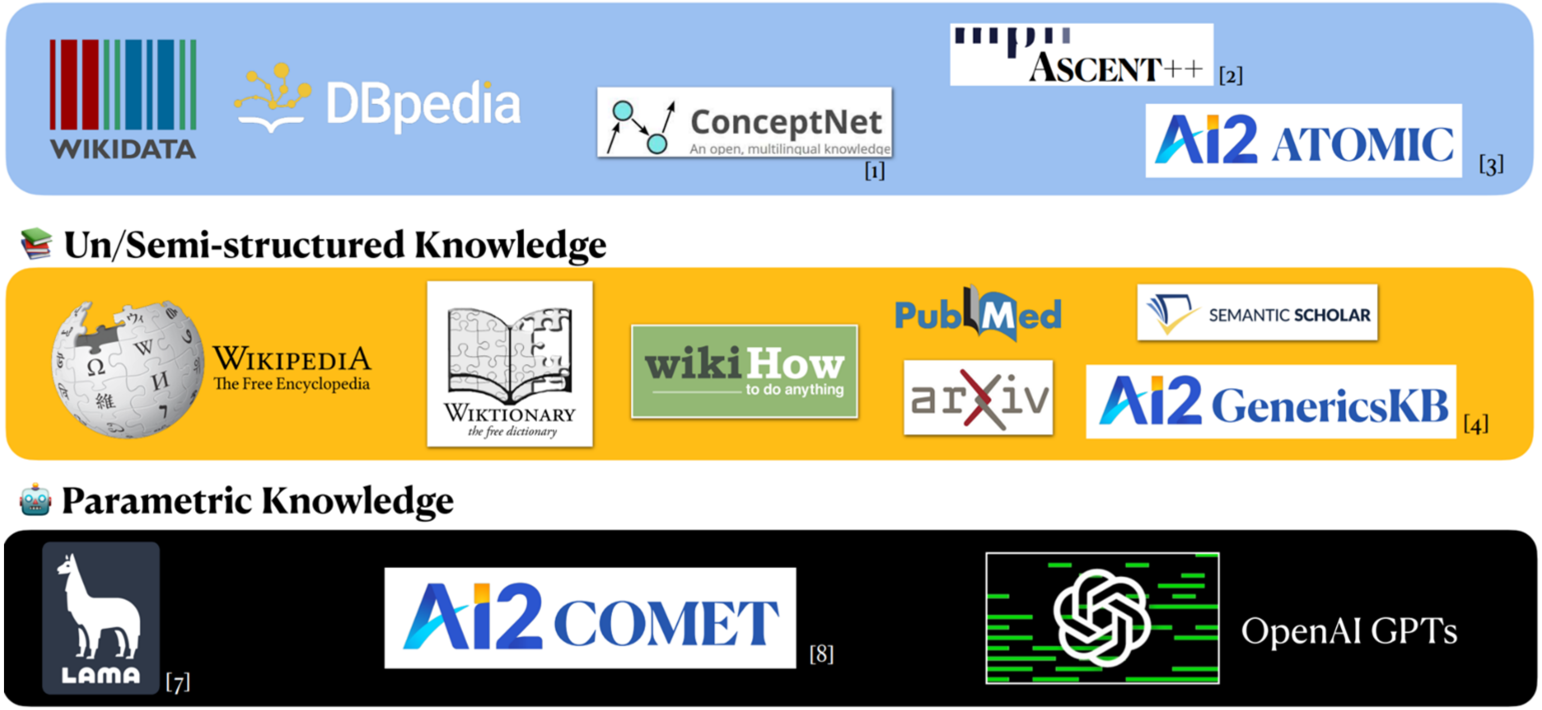
Generate rather than Retrieve: Large Languag... openreview · 6

Answer

Here are some of the most popular papers related to retrieval-based LM:

- "Improving language models by retrieving from trillions of tokens" 1 This paper proposes a method for enhancing auto-regressive language models by conditioning on document chunks retrieved from a large corpus, based on local context.

Knowledge Types



Structured Knowledge

Un/Semi-structured Knowledge

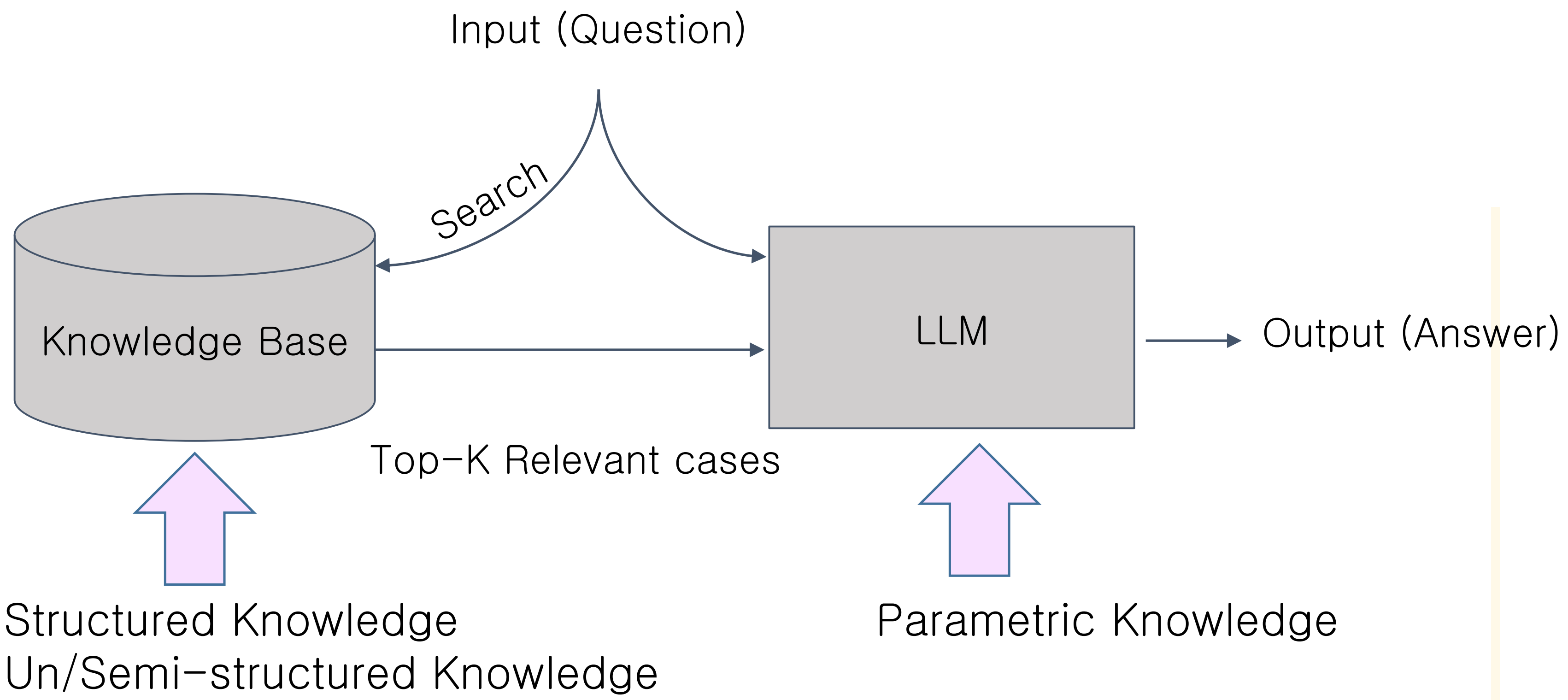
Parametric Knowledge

https://wenting-zhao.github.io/complex-reasoning-tutorial/

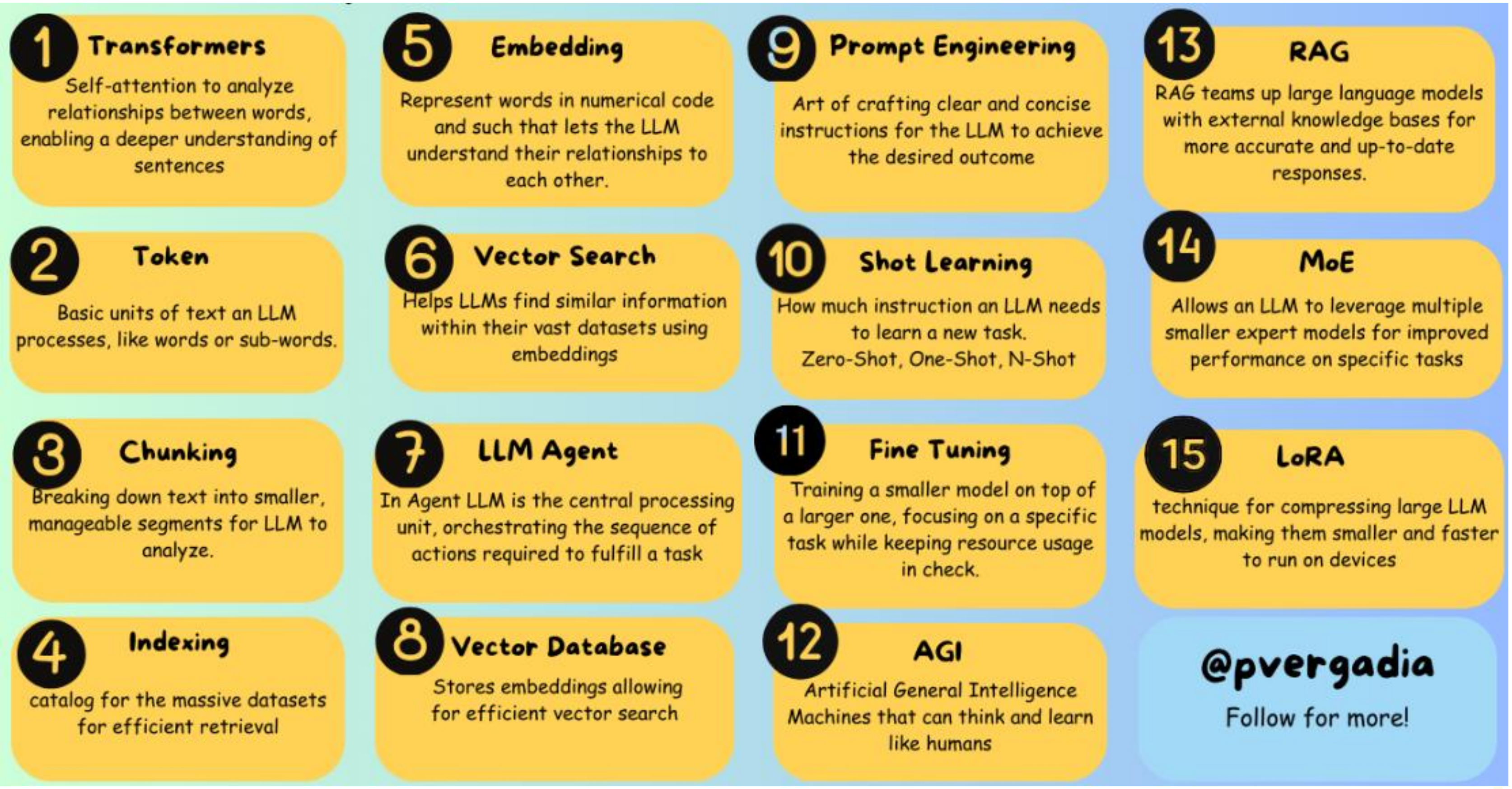
Knowledge Types

- Structured Knowledge
  - Vulnerability databases (CVE records with standardized fields)
  - MITRE ATT&CK framework tactics and techniques
  - Security event logs with defined fields and relationships
- Un/Semi-structured Knowledge
  - Threat intelligence reports and security blogs
  - Malware analysis narratives and researcher notes
  - Social media posts discussing emerging threats
- Parametric Knowledge
  - Microsoft Security Copilot
  - Google Sec-Gemini

RAG with Different Knowledge Types



Other Terms



1 Transformers

Self-attention to analyze relationships between words, enabling a deeper understanding of sentences

2 Token

Basic units of text an LLM processes, like words or sub-words.

3 Chunking

Breaking down text into smaller, manageable segments for LLM to analyze.

4 Indexing

catalog for the massive datasets for efficient retrieval

5 Embedding

Represent words in numerical code and such that lets the LLM understand their relationships to each other.

6 Vector Search

Helps LLMs find similar information within their vast datasets using embeddings

7 LLM Agent

In Agent LLM is the central processing unit, orchestrating the sequence of actions required to fulfill a task

8 Vector Database

Stores embeddings allowing for efficient vector search

9 Prompt Engineering

Art of crafting clear and concise instructions for the LLM to achieve the desired outcome

10 Shot Learning

How much instruction an LLM needs to learn a new task.
Zero-Shot, One-Shot, N-Shot

11 Fine Tuning

Training a smaller model on top of a larger one, focusing on a specific task while keeping resource usage in check.

12 AGI

Artificial General Intelligence
Machines that can think and learn like humans

13 RAG

RAG teams up large language models with external knowledge bases for more accurate and up-to-date responses.

14 MoE

Allows an LLM to leverage multiple smaller expert models for improved performance on specific tasks

15 LoRA

technique for compressing large LLM models, making them smaller and faster to run on devices

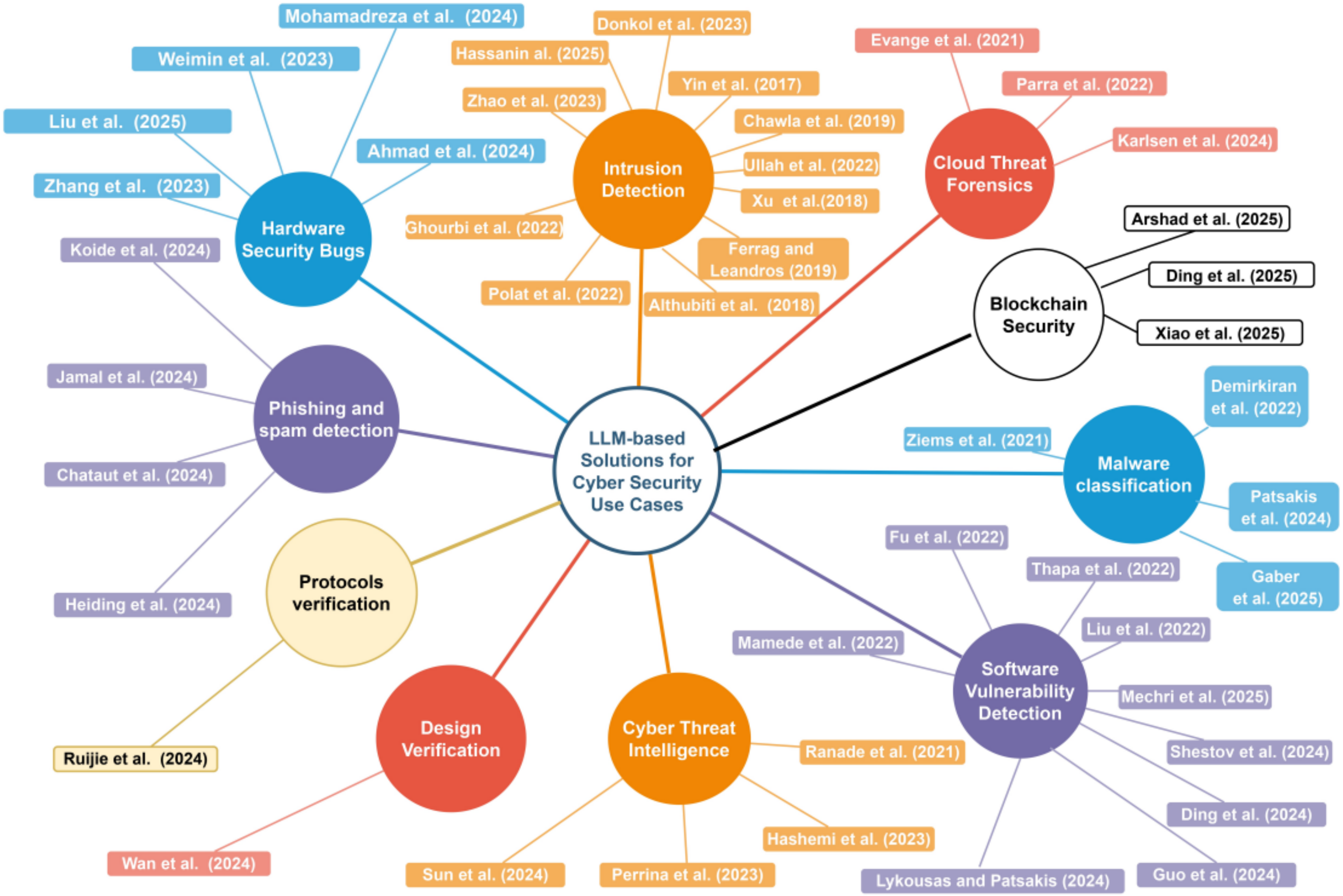
@pvergadia

Follow for more!

They are fascinating topics worth exploring in more depth, but for this class, just remember that modern LLMs can understand human instructions

https://www.thecloudgirl.dev/blog/top-15-llm-terms-you-need-to-know-in-2024

Generative AI in Cybersecurity



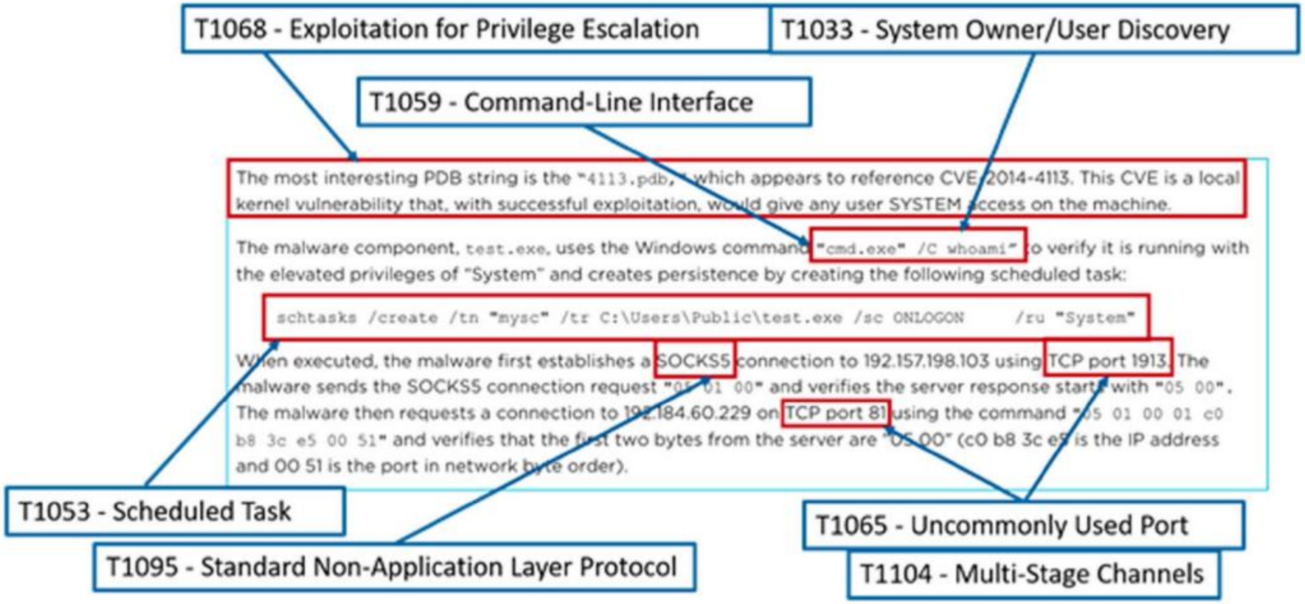
LLM-based Solutions for Cyber Security Use Cases

- Hardware Security Bugs: Mohamadreza et al. (2024), Weimin et al. (2023), Liu et al. (2025), Zhang et al. (2023), Ahmad et al. (2024)
- Intrusion Detection: Donkol et al. (2023), Hassanin al. (2025), Zhao et al. (2023), Yin et al. (2017), Chawla et al. (2019), Ullah et al. (2022), Xu et al.(2018), Ghourbi et al. (2022), Ferrag and Leandros (2019), Polat et al. (2022), Althubiti et al. (2018)
- Cloud Threat Forensics: Evange et al. (2021), Parra et al. (2022), Karlsen et al. (2024)
- Blockchain Security: Arshad et al. (2025), Ding et al. (2025), Xiao et al. (2025)
- Malware classification: Demirkiran et al. (2022), Ziems et al. (2021), Patsakis et al. (2024), Gaber et al. (2025)
- Software Vulnerability Detection: Fu et al. (2022), Thapa et al. (2022), Liu et al. (2022), Mamede et al. (2022), Mechri et al. (2025), Shestov et al. (2024), Ding et al. (2024), Guo et al. (2024), Lykousas and Patsakis (2024)
- Cyber Threat Intelligence: Ranade et al. (2021), Hashemi et al. (2023), Perrina et al. (2023), Sun et al. (2024)
- Design Verification: Wan et al. (2024)
- Protocols verification: Ruijie et al. (2024)
- Phishing and spam detection: Koide et al. (2024), Jamal et al. (2024), Chataut et al. (2024), Heiding et al. (2024)

Ferrag, Mohamed Amine, et al. "Generative AI in Cybersecurity: A Comprehensive Review of LLM Applications and Vulnerabilities." Internet of Things and Cyber-Physical Systems (2025).

Use Case 1: MITRE ATT&CK Mapping

- Objective
  - AI tool can automate threat report analysis by mapping human language descriptions to MITRE ATT&CK frameworks without STIX/TAXII expertise
  - Five-step workflow: summarize materials, parse behaviors, map to techniques, verify results, and generate reports



T1068 - Exploitation for Privilege Escalation

T1033 - System Owner/User Discovery

T1059 - Command-Line Interface

The most interesting PDB string is the "4113.pdb," which appears to reference CVE-2014-4113. This CVE is a local kernel vulnerability that, with successful exploitation, would give any user SYSTEM access on the machine.

The malware component, test.exe, uses the Windows command "cmd.exe" /C whoami" to verify it is running with the elevated privileges of "System" and creates persistence by creating the following scheduled task:

```
schtasks /create /tn "mysc" /tr C:\Users\Public\test.exe /sc ONLOGON     /ru "System"
```

When executed, the malware first establishes a SOCKS5 connection to 192.157.198.103 using TCP port 1913. The malware sends the SOCKS5 connection request "05 01 00" and verifies the server response starts with "05 00". The malware then requests a connection to 192.184.60.229 on TCP port 81 using the command "05 01 00 01 c0 b8 3c e5 00 51" and verifies that the first two bytes from the server are "05 00" (c0 b8 3c e5 is the IP address and 00 51 is the port in network byte order).

T1053 - Scheduled Task

T1095 - Standard Non-Application Layer Protocol

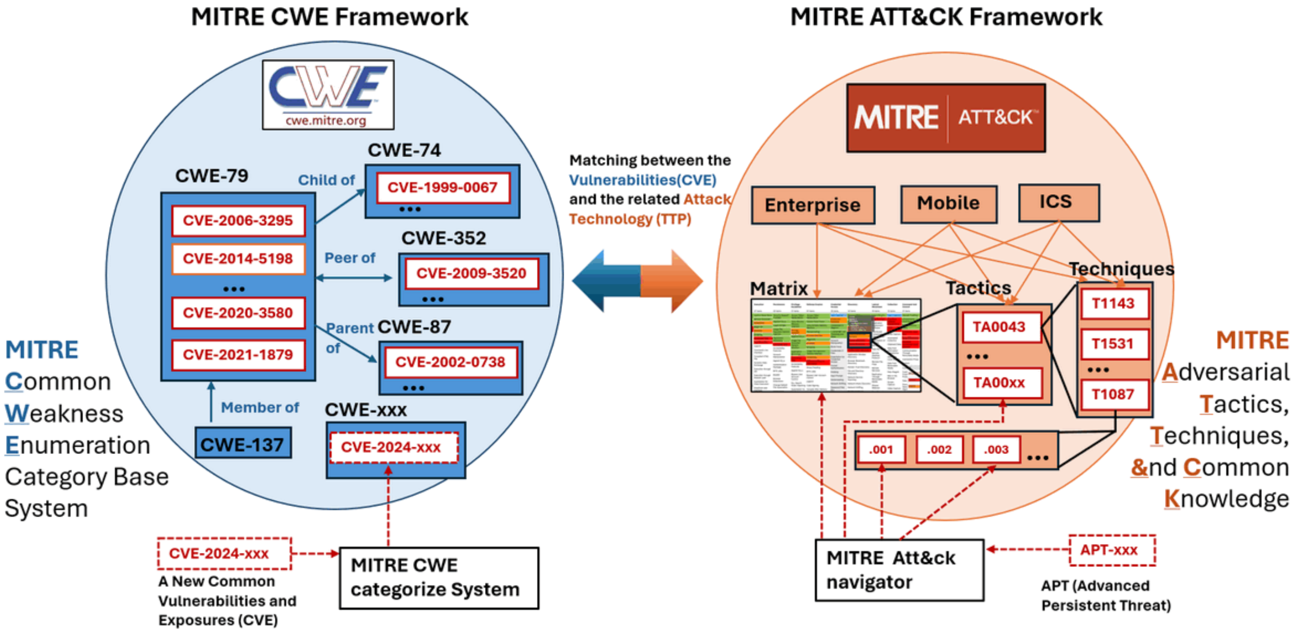
T1065 - Uncommonly Used Port

T1104 - Multi-Stage Channels

https://www.mitre.org/sites/default/files/2021-11/getting-started-with-attack-october-2019.pdf

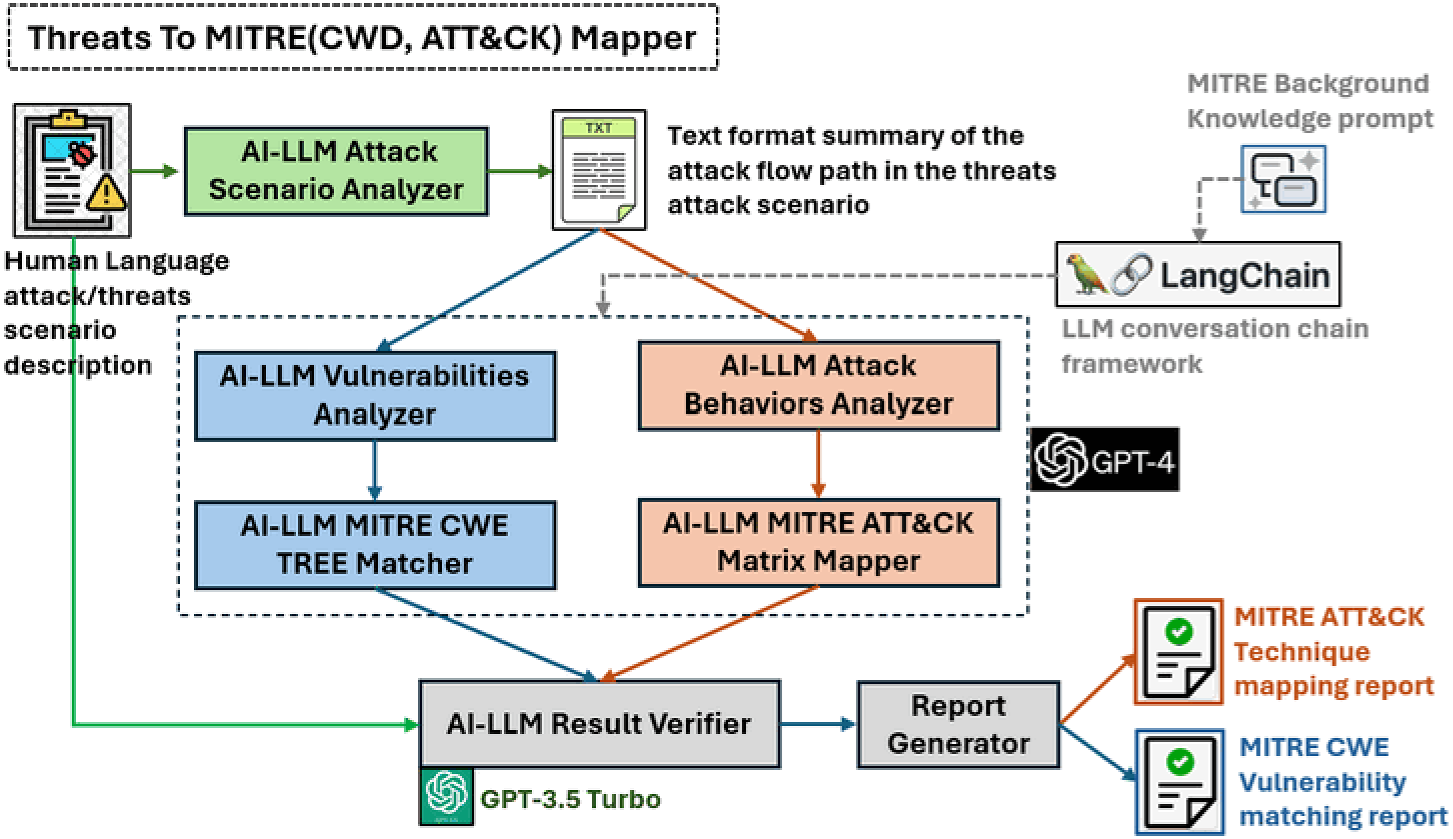
Use Case 1: MITRE ATT&CK Mapping

- MITRE CWE and ATT&CK
  - CWE is a community-developed list of common software and hardware weaknesses, faults, and vulnerabilities
  - ATT&CK is a knowledge base maintained by MITRE that documents the tactics, techniques, and procedures (TTPs) used by adversaries during cyberattacks



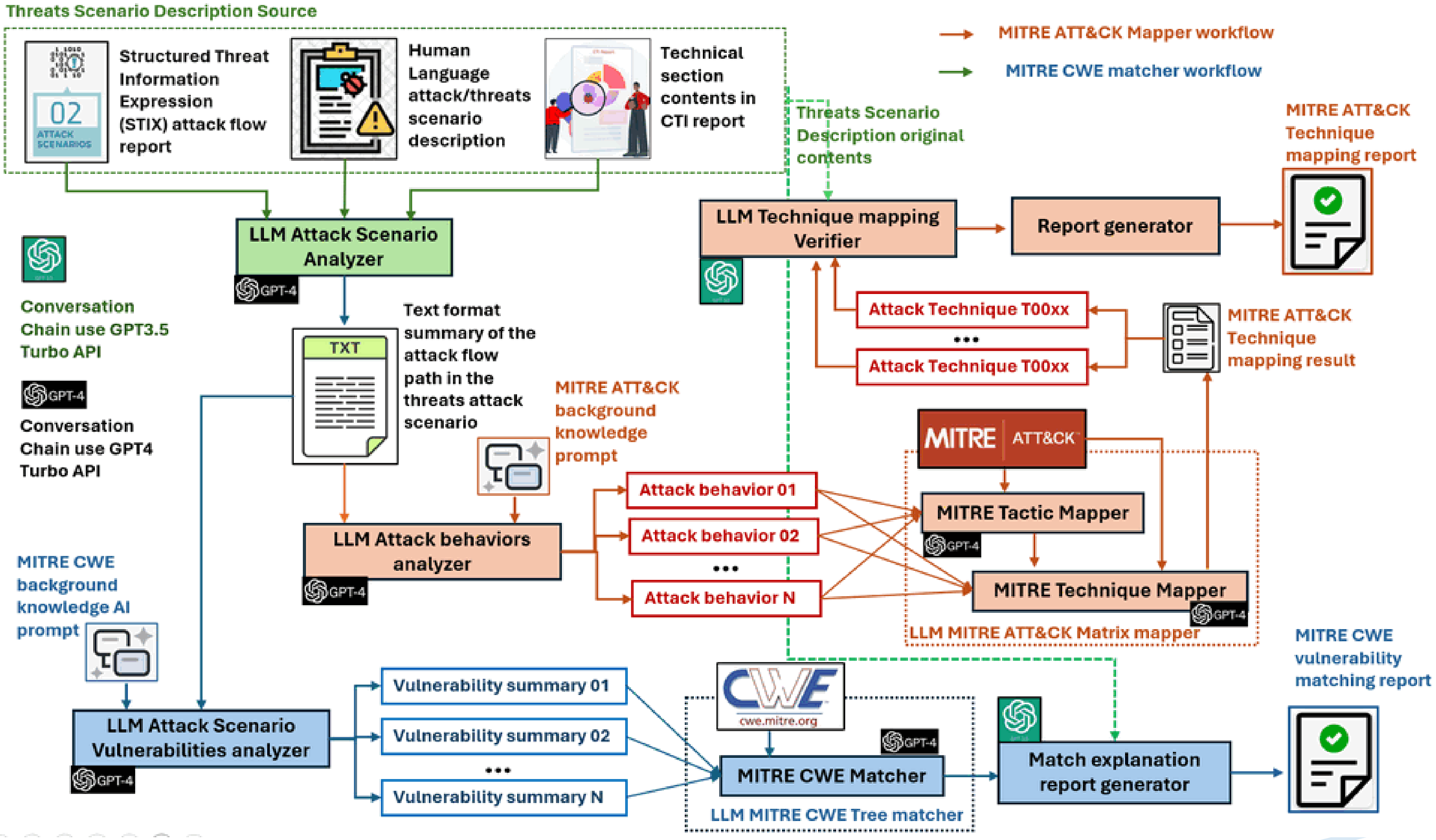
https://github.com/LiuYuancheng/Threats_2_MITRE_AI_Mapper

Use Case 1: MITRE ATT&CK Mapping


Threats To MITRE(CWD, ATT&CK) Mapper
MITRE Background Knowledge prompt
AI-LLM Attack Scenario Analyzer
TXT
Text format summary of the attack flow path in the threats attack scenario
Human Language attack/threats scenario description
LangChain
LLM conversation chain framework
AI-LLM Vulnerabilities Analyzer
AI-LLM Attack Behaviors Analyzer
GPT-4
AI-LLM MITRE CWE TREE Matcher
AI-LLM MITRE ATT&CK Matrix Mapper
MITRE ATT&CK Technique mapping report
AI-LLM Result Verifier
Report Generator
GPT-3.5 Turbo
MITRE CWE Vulnerability matching report

https://github.com/LiuYuancheng/Threats_2_MITRE_AI_Mapper

Use Case 1: MITRE ATT&CK Mapping



Threats Scenario Description Source

Structured Threat Information Expression (STIX) attack flow report

Human Language attack/threats scenario description

Technical section contents in CTI report

MITRE ATT&CK Mapper workflow

MITRE CWE matcher workflow

Conversation Chain use GPT3.5 Turbo API

Conversation Chain use GPT4 Turbo API

LLM Attack Scenario Analyzer

Text format summary of the attack flow path in the threats attack scenario

MITRE ATT&CK background knowledge prompt

LLM Attack behaviors analyzer

Attack behavior 01

Attack behavior 02

Attack behavior N

Threats Scenario Description original contents

LLM Technique mapping Verifier

Report generator

MITRE ATT&CK Technique mapping report

Attack Technique T00xx

Attack Technique T00xx

MITRE ATT&CK Technique mapping result

MITRE Tactic Mapper

MITRE Technique Mapper

LLM MITRE ATT&CK Matrix mapper

MITRE CWE background knowledge AI prompt

LLM Attack Scenario Vulnerabilities analyzer

Vulnerability summary 01

Vulnerability summary 02

Vulnerability summary N

MITRE CWE Matcher

LLM MITRE CWE Tree matcher

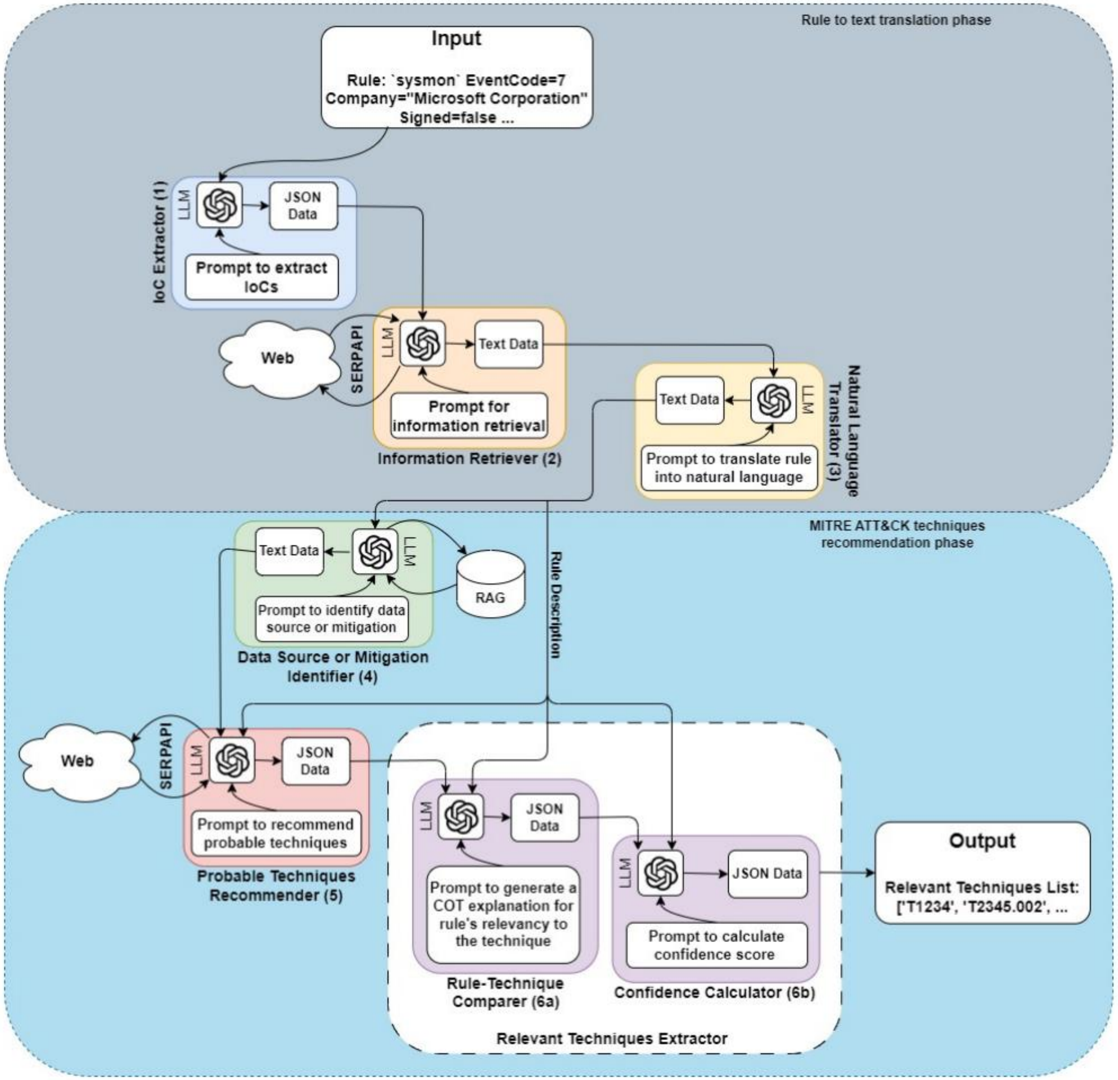
Match explanation report generator

MITRE CWE vulnerability matching report

NSR 국가보안기술연구소

https://github.com/LiuYuancheng/Threats_2_MITRE_AI_Mapper

Use Case 2: Mapping SIEM Rules to TTPs


Rule to text translation phase
Input
Rule: `sysmon` EventCode=7 Company="Microsoft Corporation" Signed=false ...
IoC Extractor (1)
LLM
JSON Data
Prompt to extract IoCs
Web
SERPAPI
LLM
Text Data
Prompt for information retrieval
Information Retriever (2)
Text Data
LLM
Prompt to translate rule into natural language
Natural Language Translator (3)
MITRE ATT&CK techniques recommendation phase
Text Data
LLM
RAG
Prompt to identify data source or mitigation
Data Source or Mitigation Identifier (4)
Rule Description
Web
SERPAPI
LLM
JSON Data
Prompt to recommend probable techniques
Probable Techniques Recommender (5)
LLM
JSON Data
Prompt to generate a COT explanation for rule's relevancy to the technique
Rule-Technique Comparer (6a)
LLM
JSON Data
Prompt to calculate confidence score
Confidence Calculator (6b)
Relevant Techniques Extractor
Output
Relevant Techniques List: ['T1234', 'T2345.002', ...

Wudali, Prasanna N., et al. "Rule-ATT&CK Mapper (RAM): Mapping SIEM Rules to TTPs Using LLMs." arXiv preprint arXiv:2502.02337 (2025).

(Hands-on 1) Data Collection and Processing

- Objectives
  - Learn how to collect and process threat intelligence data from various sources such as webpages, RSS feeds, and DNS records
  - Practice extracting indicators of compromise (IOCs) like IPs, domains, hashes, and filenames using web scraping and regular expressions
  - Explore feed parsing and DNS record resolution to enrich the context of security events
  - Gain experience with IP attribution by mapping IP addresses to autonomous systems using WHOIS data
- Python Packages
  - requests, beatifulsoup4, re, feedparser, dnspython, ipwhois

(Hands-on 2) Prompt Engineering for LLM

- Objectives
  - Learn how to integrate OpenAI's GPT models for automated classification of cyber attack techniques based on MITRE ATT&CK
  - Understand how to validate structured outputs using Pydantic models
  - Practice extracting and translating text from images using multi-modal prompts
- Python Packages
  - pydantic, openai

(Hands-on 3) Embedding for Similarity Search

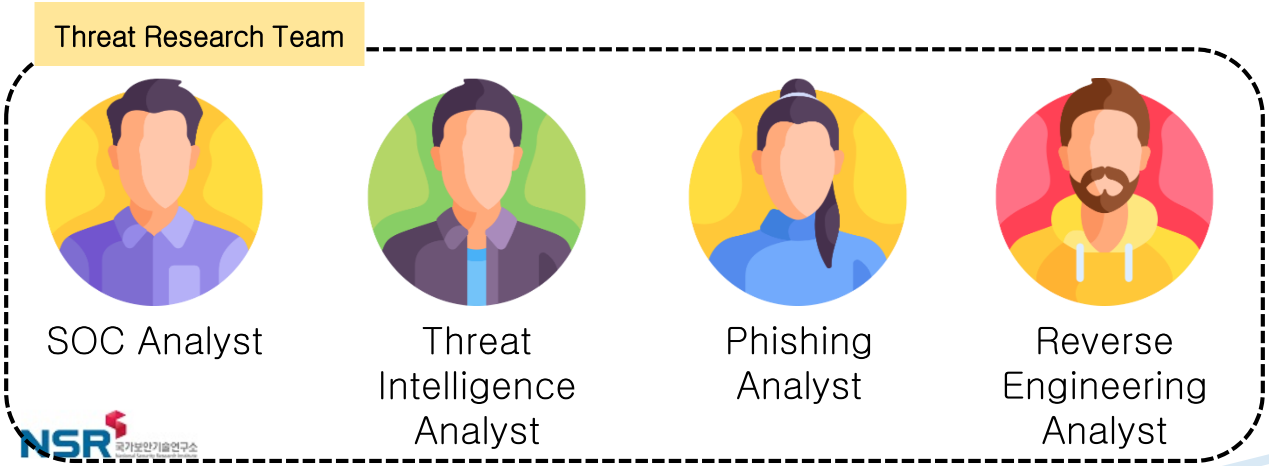
- Objectives
  - Learn to apply embedding models to transform textual threat group descriptions into numerical vectors
  - Explore dimensionality reduction and visualization techniques using PCA
  - Analyze and compare threat actor similarities using cosine similarity and query matching
- Python Packages
  - requests, sentence_transformers, stix2, pandas, matplotlib, sklearn

(Hands-on 4) Querying Data with LLM

- Objectives
  - Build a natural language interface for pandas DataFrames using LangChain agents
  - Generate SQL queries from natural language questions using SQLCoder, and explore how LLMs can automate structured data access
- Python Packages
  - kagglehub, langchain, transformers, sqlparse

(Hands-on 5) Multi Agents based Threat Research Team

- Objectives
  - Build an automated cybersecurity investigation system using multiple specialized AI agents working together to analyze security alerts
  - Demonstrate how different security roles (SOC Analyst, Threat Intelligence, Reverse Engineering, and Phishing Analysis) can collaborate through a coordinated workflow
- Python Packages
  - autogen



Threat Research Team

SOC Analyst

Threat Intelligence Analyst

Phishing Analyst

Reverse Engineering Analyst

Task

Investigate a suspicious login alert

Investigate a malware detection alert

Investigate a data exfiltration alert

Investigate a phishing campaign

Thank you ☺