2024
FIRST
Cyber Threat Intelligence Conference

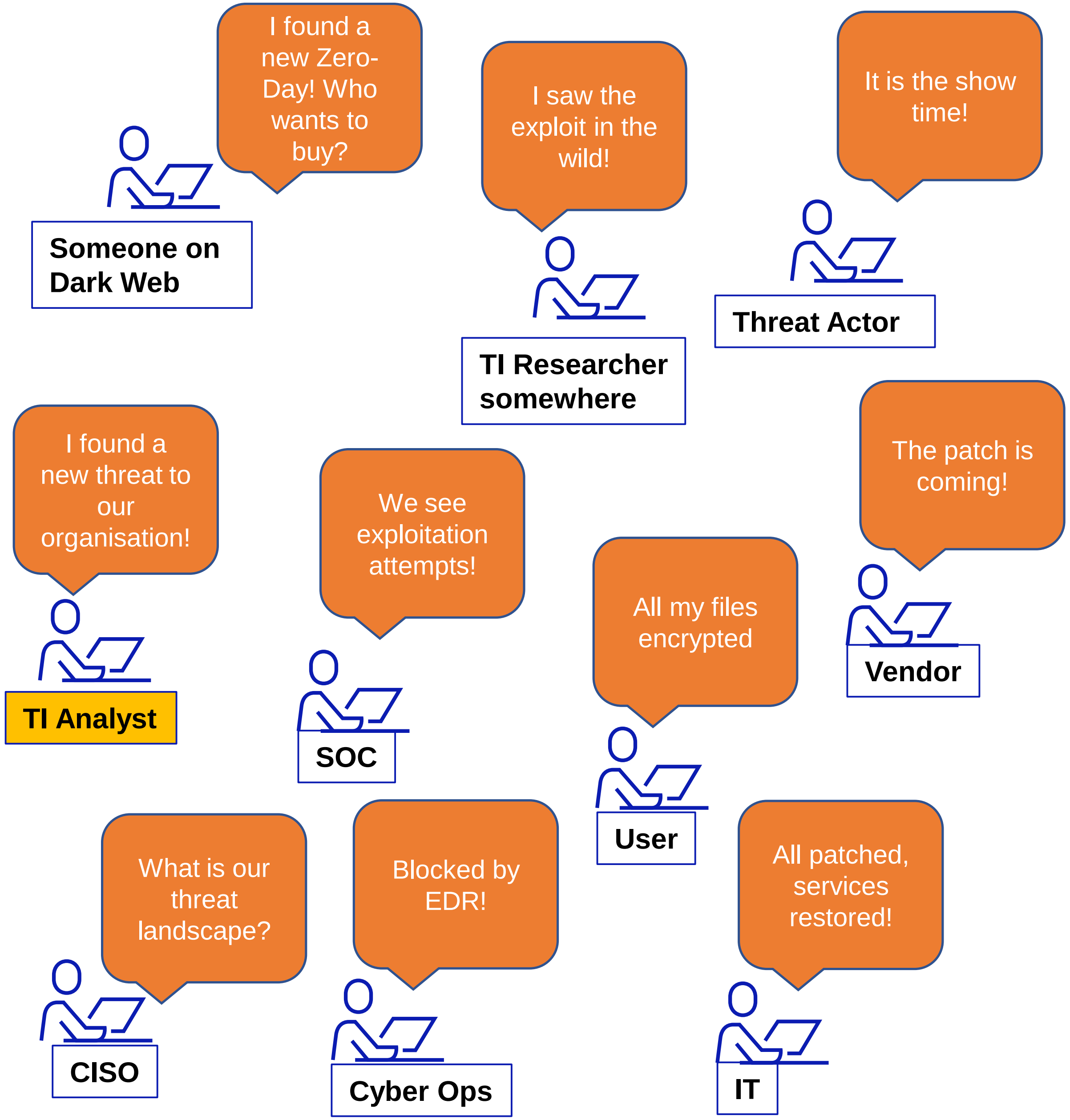
Berlin, Germany
April 15-17, 2024

Processing threat reports at scale using AI and ML: Expectations and Reality

Yury Sergeev

16.04.2024

- CTI analysts **read numerous reports every day**
- How can we **select only relevant news/reports** that will help us to focus on our PIR and SIR?
- Join industry communities?
- Hire more people to do the filtering?
- Delegate filtering to some other organisation?
- Develop some **tools to pre-screen reports** and filter out irrelevant ones?

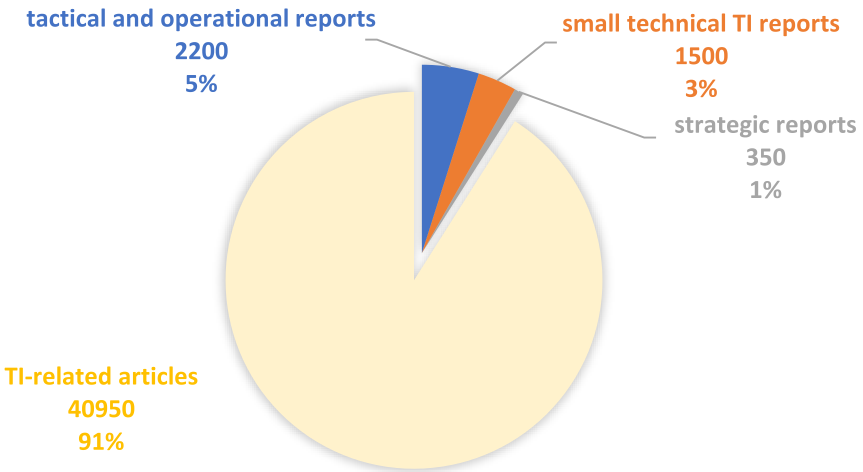


I found a new Zero-Day! Who wants to buy?
Someone on Dark Web
I saw the exploit in the wild!
TI Researcher somewhere
It is the show time!
Threat Actor
I found a new threat to our organisation!
TI Analyst
We see exploitation attempts!
SOC
All my files encrypted
User
The patch is coming!
Vendor
What is our threat landscape?
CISO
Blocked by EDR!
Cyber Ops
All patched, services restored!
IT

Can we keep up?

SANS Surveys show that reports and news have been at the top of sources for intelligence gathering for 3 consecutive years

STATISTICS FOR 2023*



Only X are relevant to you!?

To get what's relevant to you, an analyst needs:

~ sift through 180 articles a day

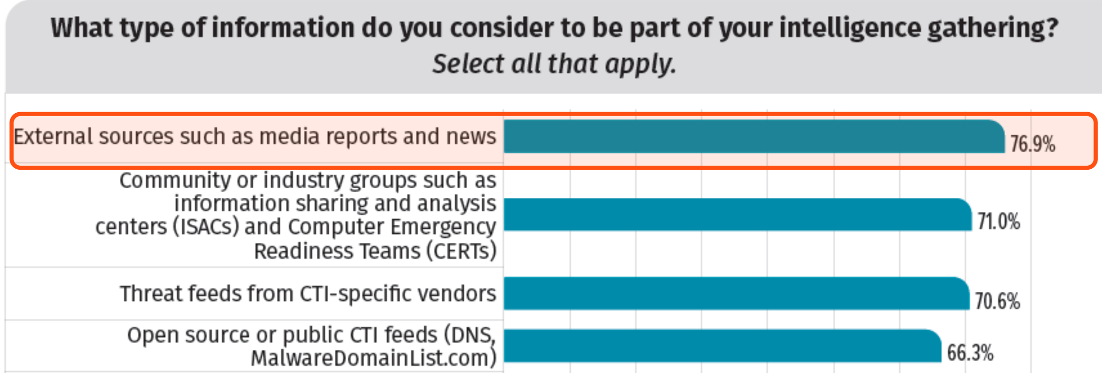
~ read 9 tactical/operational reports a day

~ read 6 atomic tech articles a day

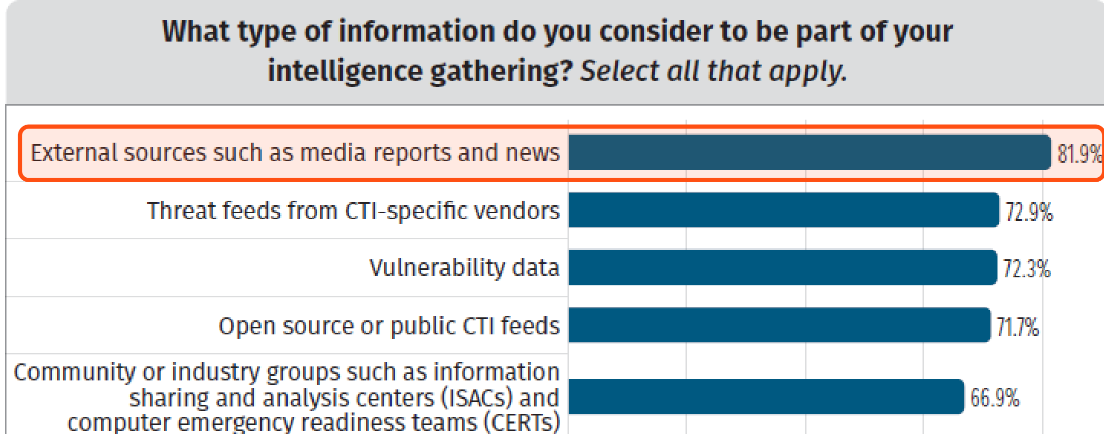
~ read 1.4 strategic reports a day

* - based on data from RST Cloud

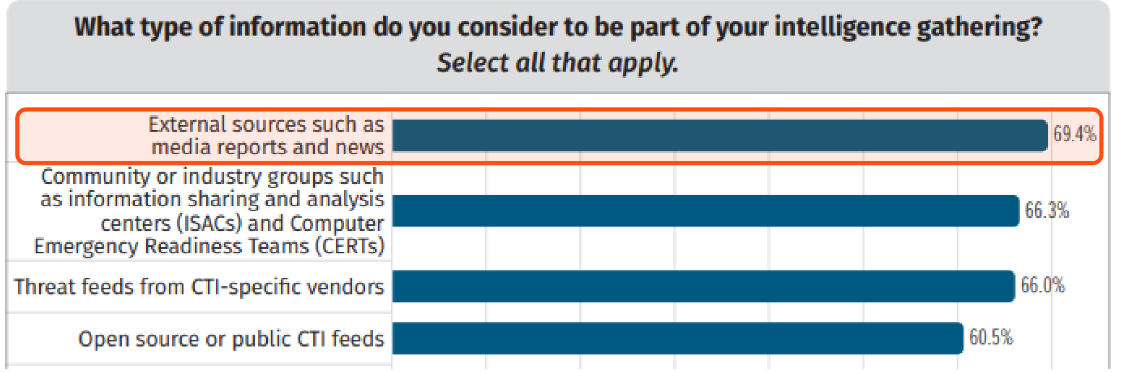
2021 SANS Cyber Threat Intelligence Survey



SANS 2022 Cyber Threat Intelligence Survey



SANS 2023 CTI Survey: Keeping Up with a Changing Threat Landscape



Sources of threat reports

Government agencies | ISACs | CERTs | TI providers | Security Vendors | Cyber communities | Individual researches

News / Ads | DFIR reports | Trends | Operational TI reports | Reverse Engineering Reports | Opinions | Strategic TI reports | Products | Tactical TI reports | Events

- How can we classify the incoming source data?
- What are the parameters of valuable sources?
- What is the way to extract data effectively?

Threat Actor Source type Diagrams

Exploits Lexicon Tone

Crypto Malware

Victims Motivation

TTPs SIGMA Languages

References

Code

Algorithms IOCs

Vulnerabilities Relations

Software Platforms Campaigns

Syscalls YARA

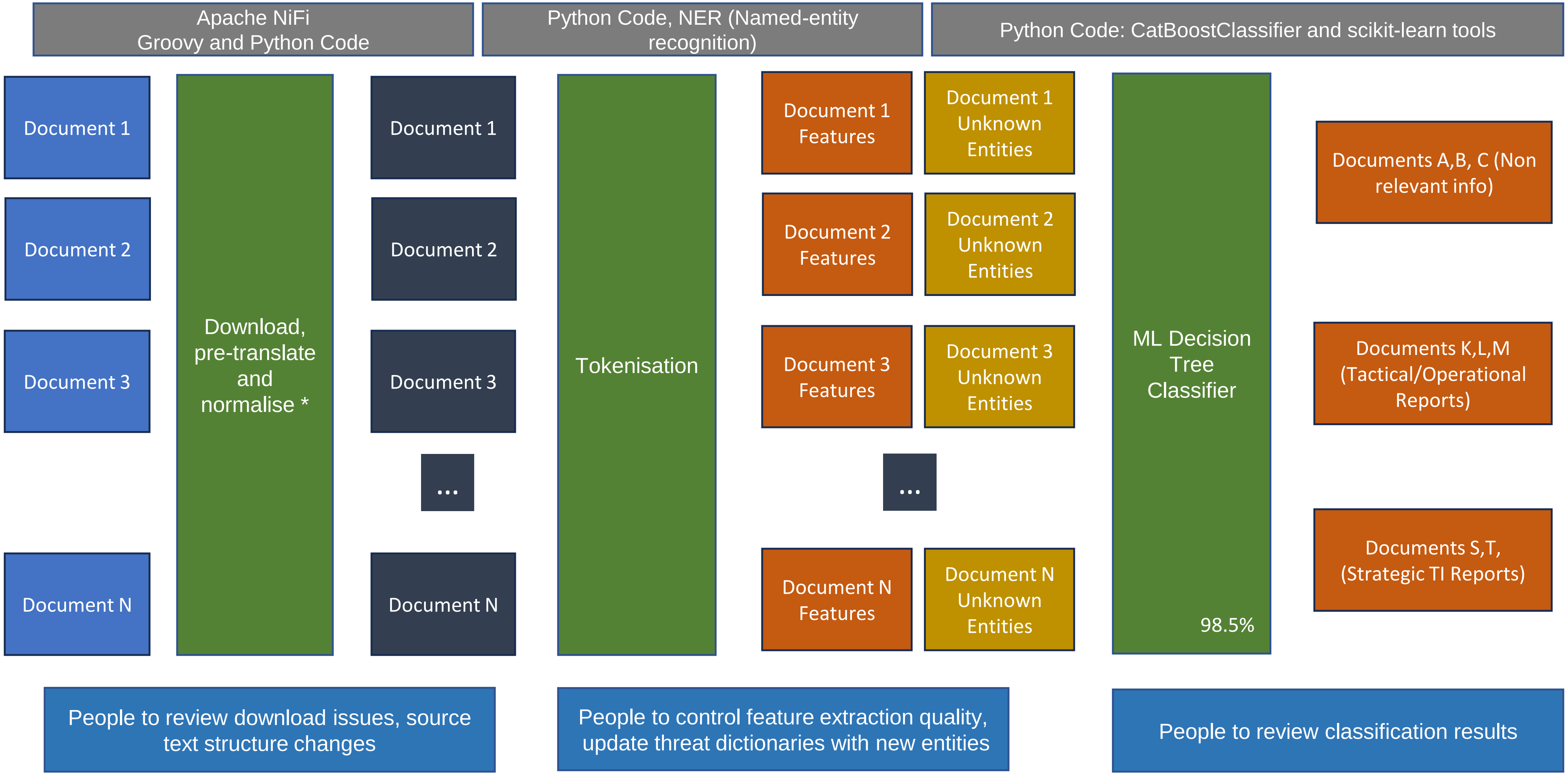
Windows Service Names Sectors

Windows Kernel32 Calls Commands Tools

TI Report processing: Agenda

Download, pre-translate and normalise	We will skip this part, however, translation, text recovery from images or image recognition is an interesting ML/AI topic to cover
Tokenisation	We will review NER and LLM approaches
Classification	We will review ML and LLM classification approaches
Filtering and deduplication	We will skip. Most of these processes are not related to ML or AI
Entity Relation Extraction	We will review ML and LLM approaches
Transformation	Will see where AI helps

NER and ML Approach to classification



Apache NiFi
Groovy and Python Code

Python Code, NER (Named-entity recognition)

Python Code: CatBoostClassifier and scikit-learn tools

Document 1

Document 2

Document 3

Document N

Download, pre-translate and normalise *

Document 1

Document 2

Document 3

...

Document N

Tokenisation

Document 1 Features

Document 1 Unknown Entities

Document 2 Features

Document 2 Unknown Entities

Document 3 Features

Document 3 Unknown Entities

...

Document N Features

Document N Unknown Entities

ML Decision Tree Classifier

98.5%

Documents A,B, C (Non relevant info)

Documents K,L,M (Tactical/Operational Reports)

Documents S,T, (Strategic TI Reports)

People to review download issues, source text structure changes

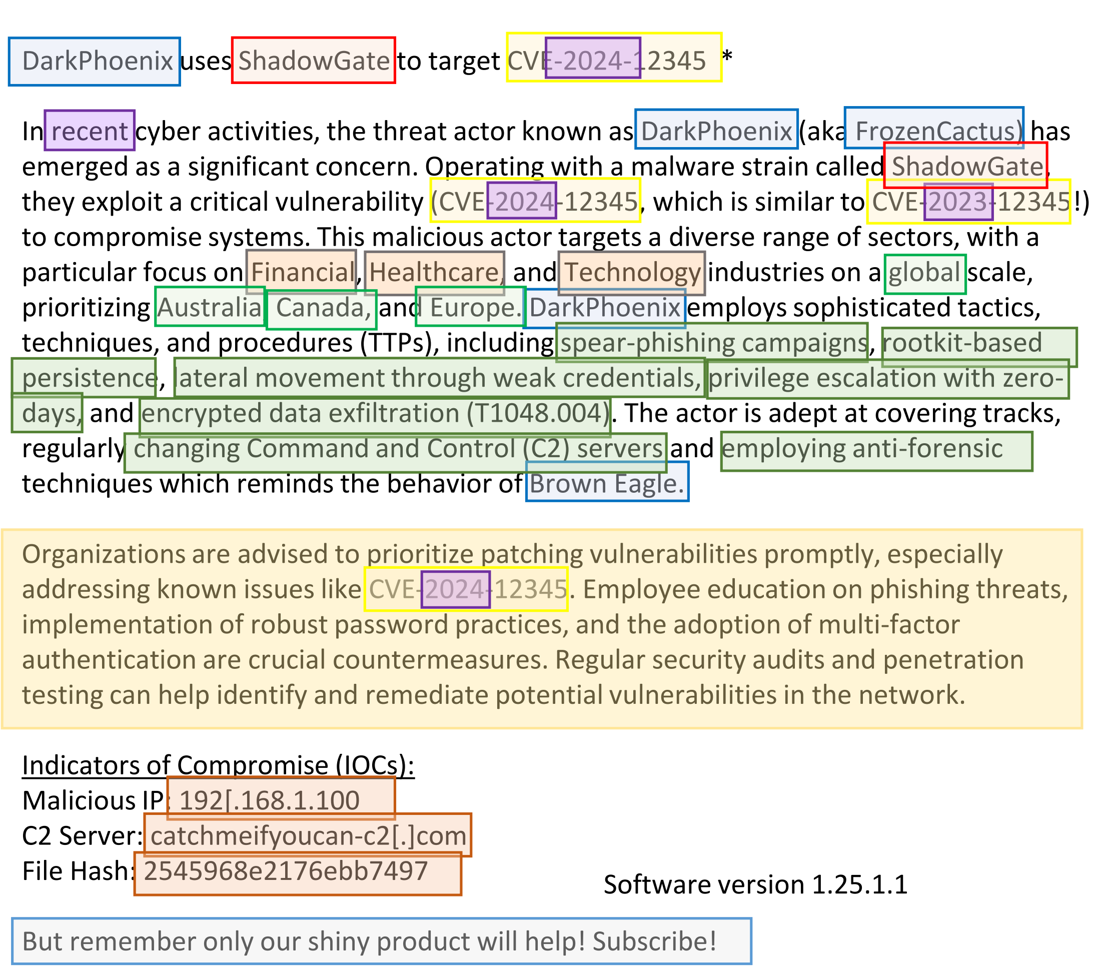
People to control feature extraction quality, update threat dictionaries with new entities

People to review classification results

* - think about onboarding logs to a SIEM: many little engineering difficulties. Skipped

https://catboost.ai/en/docs/concepts/python-reference_catboostclassifier

How does NER process work?



DarkPhoenix uses ShadowGate to target CVE-2024-12345 *

In recent cyber activities, the threat actor known as DarkPhoenix (aka FrozenCactus) has emerged as a significant concern. Operating with a malware strain called ShadowGate, they exploit a critical vulnerability (CVE-2024-12345, which is similar to CVE-2023-12345!) to compromise systems. This malicious actor targets a diverse range of sectors, with a particular focus on Financial, Healthcare, and Technology industries on a global scale, prioritizing Australia, Canada, and Europe. DarkPhoenix employs sophisticated tactics, techniques, and procedures (TTPs), including spear-phishing campaigns, rootkit-based persistence, lateral movement through weak credentials, privilege escalation with zero-days, and encrypted data exfiltration (T1048.004). The actor is adept at covering tracks, regularly changing Command and Control (C2) servers and employing anti-forensic techniques which reminds the behavior of Brown Eagle.

Organizations are advised to prioritize patching vulnerabilities promptly, especially addressing known issues like CVE-2024-12345. Employee education on phishing threats, implementation of robust password practices, and the adoption of multi-factor authentication are crucial countermeasures. Regular security audits and penetration testing can help identify and remediate potential vulnerabilities in the network.

Indicators of Compromise (IOCs):
Malicious IP: 192[.168.1.100
C2 Server: catchmeifyoucan-c2[.]com
File Hash: 2545968e2176ebb7497

Software version 1.25.1.1

But remember only our shiny product will help! Subscribe!

Threat Actor | Malware | Vulnerability | Time

Sector | Location | TTP | Course of Action

Indicators | Irrelevant info | Etc.

- OK! We have got all the tokens! Hooray. Done!

Well… But…

How can we find relations?

Is this report about Brown Eagle or about DarkPhoenix?

Who is targeted by whom?

Can we guess if BrownEagle targeting CVE-2023-12345 or not?

The C2 mentions is it ShadowGate, DarkPhoenix, or ShadowGate?

* This simulated threat intelligence report serves for educational purposes only and does not represent any actual threat or actor

Can ChatGPT do all of that?

Test Prompt (bad one): “You are a threat intel analyst. I will give you a text to analyse and extract key details. Give me a JSON output which will be including arrays of values for keys threat_actor, malware_name, vulnerability, time_indicator, sector, location, ttp (MITRE Attack Ds), course_of_action, indicator_of_compromise. The text is the following:”

Model 3.5 (free chat)

```
{  "threat_actor": ["DarkPhoenix", "FrozenCactus", "Brown Eagle"],
  "malware_name": ["ShadowGate"],
  "vulnerability": ["CVE-2024-12345", "CVE-2023-12345"],
  "time_indicator": "recent",
  "sector": ["Financial", "Healthcare", "Technology"],
  "location": ["Australia", "Canada", "Europe", "global"],
  "ttp": [    "spear-phishing campaigns",
   "rootkit-based persistence",
   "lateral movement through weak credentials",
   "privilege escalation with zero-days",
   "encrypted data exfiltration (T1048.004)",
   "changing Command and Control (C2) servers",
   "anti-forensic techniques"  ],
  "course_of_action": [   "prioritize patching vulnerabilities promptly",
   "employee education on phishing threats",
   "implementation of robust password practices",
   "adoption of multi-factor authentication",
   "regular security audits",  "penetration testing"  ],
  "indicator_of_compromise": [
   "Malicious IP: 192.168.1.100",
   "C2 Server: catchmeifyoucan-c2.com",
   "File Hash: 2545968e2176ebb7497"  ]}
```

API: Model 4.0-turbo-preview (1024 tokens, temp=1)

GPT4 Run 1

```
{  "threat_actor": ["DarkPhoenix", "FrozenCactus"],
  "malware_name": ["ShadowGate"],
  "vulnerability": ["CVE-2024-12345", "CVE-2023-12345"],
  "time_indicator": [],
  "sector": ["Financial", "Healthcare", "Technology"],
  "location": ["Australia", "Canada", "Europe"],
  "ttp": ["spear-phishing campaigns", "rootkit-based persistence", "lateral
movement through weak credentials", "privilege escalation with zero-days",
"encrypted data exfiltration", "T1048.004"]
  "course_of_action": ["prioritize patching vulnerabilities", "employee education
on phishing threats", "implementation of robust password practices",
"adoption of multi-factor authentication", "regular security audits and
penetration testing"],
  "indicator_of_compromise": [
     {"type": "IP", "value": "192.168.1.100"},
     {"type": "Domain", "value": "catchmeifyoucan-c2.com"},
     {"type": "File Hash", "value": "2545968e2176ebb7497"}  ]}
```

GPT4 Run 2

```
{  "threat_actor": ["DarkPhoenix", "FrozenCactus", "BrownEagle"],
 "malware_name": ["ShadowGate"],
  "vulnerability": ["CVE-2024-12345", "CVE-2023-12345"],
  "time_indicator": [],
  "sector": ["Financial", "Healthcare", "Technology"],
  "location": ["Australia", "Canada", "Europe"],
  "ttp": ["spear-phishing campaigns", "rootkit-based persistence", "lateral movement through weak credentials", "privilege escalation
with zero-days", "encrypted data exfiltration", "T1048.004"],
  "course_of_action": ["prioritize patching vulnerabilities", "employee education on phishing threats", "implementation of robust
password practices", "adoption of multi-factor authentication", "regular security audits and penetration testing"],
  "indicator_of_compromise": ["Malicious IP: 192.168.1.100", "C2 Server: catchmeifyoucan-c2.com", "File Hash:
2545968e2176ebb7497"]}
```

The bigger and more complicated the text is, and the higher cardinality of the entities is, the less deterministic answers we get

Techniques to improve answers:

- Prompts with NER specifics
- Model fine-tuning
- Use custom models
- RAG
- Splitting prompts to restrict the scope of task in each request
- “Conversation logic”
- Multiple runs

LLM classification

Tests using ChatGPT Model 3.5

Test prompt (bad one) ‘Classes: ["Tactical Threat Intel report", "Operational Threat Intel report", "Strategic Threat Intel report", "Other"] Classify the text into one of the above classes. Give a json formatted answer with a key report_class and the associated value:’

```
{
  "report_class": "Tactical Threat Intel report"
}
```

Medium quality   Easy   Cheap   10000 tokens a report -> cents per report

https://blog.talosintelligence.com/timbrestealer-campaign-targets-mexican-users/ (English, Tactical Threat Report)	"report_class": "Tactical Threat Intel report"	✓	
https://www.ctfiot.com/162025.html (Chinese, Tactical Threat Report)	"report_class": "Tactical Threat Intel report"	✓	
https://www.trendmicro.com/en_us/research/24/b/threat-actor-groups-including-black-basta-are-exploiting-recent-.html (English, Tactical Threat Report)	"report_class": "Operational Threat Intel report" Next Run: "report_class": “Tactical Threat Intel report"	✗ ✓	➔ "Tactical Threat Intel report“ ➔ (IoCs not in the report text but a link to them is given)
https://www.elastic.co/security-labs/introduction-to-hexrays-decompilation-internals (English, Malware Analysis)	"report_class": "Operational Threat Intel report"	✗	➔ Malware analysis article ➔ Helps understand the internal structures used in decompilation (IDA)
https://www.microsoft.com/en-us/security/blog/2024/02/20/navigating-nis2-requirements-with-microsoft-security-solutions/ (English, Solution Info)	"report_class": “Strategic Threat Intel report"	✗	➔ Talks about how MS helps to comply to NIS2
https://www.zscaler.com/blogs/product-insights/microsoft-midnight-blizzard-and-scourge-identity-attacks (English, Operational Threat Report)	"report_class": “Strategic Threat Intel report" Next Run: "report_class": “Tactical Threat Intel report"	✗	➔ Talks about high-level stuff, but still about one particular threat actor rather than a trend as a whole

LLM classification

Tests using ChatGPT Model 4.0

Test prompt (bad one) ‘Classes: ["Tactical Threat Intel report", "Operational Threat Intel report", "Strategic Threat Intel report", "Other"] Classify the text into one of the above classes. Give a json formatted answer with a key report_class and the associated value:’

```
{
  "report_class": "Tactical Threat Intel report"
}
```

Acceptable quality   Easy   Still cheap   15000 tokens a report -> cents per report

https://blog.talosintelligence.com/timbrestealer-campaign-targets-mexican-users/ (English, Tactical Threat Report)	"report_class": "Tactical Threat Intel report"	✓	
https://www.ctfiot.com/162025.html (Chinese, Tactical Threat Report)	"report_class": "Tactical Threat Intel report"	✓	
https://www.trendmicro.com/en_us/research/24/b/threat-actor-groups-including-black-basta-are-exploiting-recent-.html (English, Tactical Threat Report)	"report_class": “Tactical Threat Intel report"	✓	
https://www.elastic.co/security-labs/introduction-to-hexrays-decompilation-internals (English, Malware Analysis)	"report_class": “Other"	✓	
https://www.microsoft.com/en-us/security/blog/2024/02/20/navigating-nis2-requirements-with-microsoft-security-solutions/ (English, Solution Info)	"report_class": “Other"	✓	
https://www.zscaler.com/blogs/product-insights/microsoft-midnight-blizzard-and-scourge-identity-attacks (English, Operational Threat Report)	"report_class": “Strategic Threat Intel report" Next Run: "report_class": “Tactical Threat Intel report"	✗	➔ Talks about high-level stuff, but still about one particular threat actor rather than a trend as a whole

Classic NER/ML vs LLM

Feature	NER/ML	Public LLM	Private LLM
Data residency	Full control	Your data becomes the part of the public model	Full control
Cost	Cheap	Cheap	Expensive
Quality	High	Average	High
Effort to support	Average	Low	Very High
Qualification and skills required	Average	Low	Very High

Unknown entities

Let’s say you do not know these below.

What algorithm can you use to guess if it is a threat actor name?

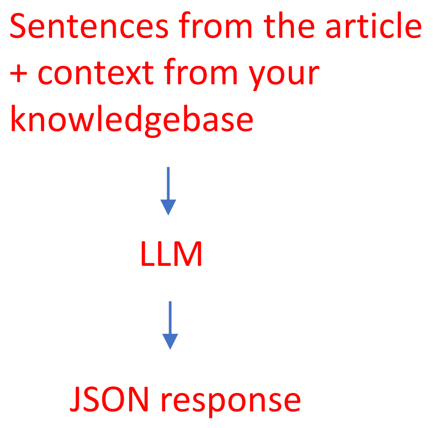
Try Regex

- Maverick Panda
- OceanLotus
- Charming Kitten
- Venomous Bear
- DarkPhoenix
- Brown Eagle
- APT-28
- APT-C-24

Try ML (for instance, RandomForestClassifier)

1. Length of the word: The number of characters in the word.
2. Presence of spaces or special characters: Check if the word contains spaces or special characters.
3. Capitalisation pattern: Determine if the word follows a specific capitalisation pattern (e.g., CamelCase, Title Case, all uppercase, all lowercase).
4. Presence of numbers: Check if the word contains numerical characters.
5. Presence of hyphens or other separators: Identify if the word includes hyphens or other separators.
6. Common acronyms or patterns: Look for common patterns like "APT-" or other specific substrings
7. Verbs that indicates an action: Look things that distinguish a subject from an object
8. etc

Try AI (e.g., OpenChat 3.5)



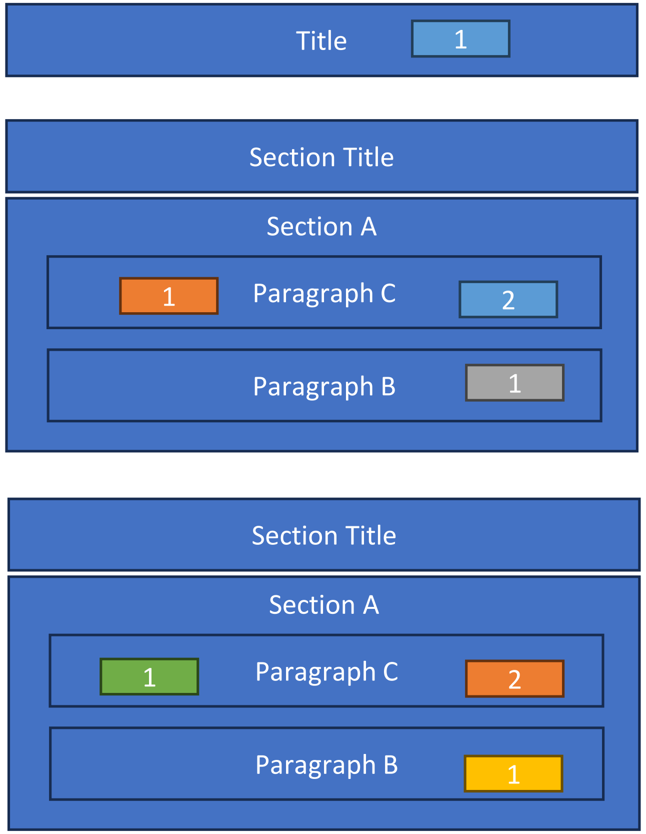
Sentences from the article + context from your knowledgebase

LLM

JSON response

Building a model of a report

A parsed report with its model split into chunks with extracted entities



How to extract the relations?

🟦 1	uses	🟧 1
🟧 1	attributed-to	▨ 1
🟨 1	indicates	🟦 1
🟧 1	originates-from	🟧 2
🟦 1	related-to	🟦 1

A serious journey starts with regex/keyword search to build the corpus of data and then to build the vocabularies of different objects

A simple one implies you rely on LLM to do it as is with mediocre quality and non-deterministic answers

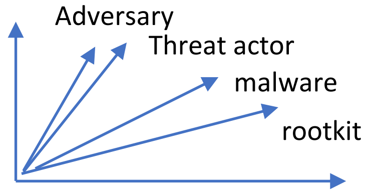
Basic / Error-prone / Many exceptions

Keyword/regex search

targets	targeting
originates	originating from
related	not related
version of	was version of
exploited by	not exploited by

Advanced: need large corpus and high-quality vocabs

Vector Search, Embeddings, ML



Similarity score from -1 to 1

Advanced: assistant, fine-tuning, custom models

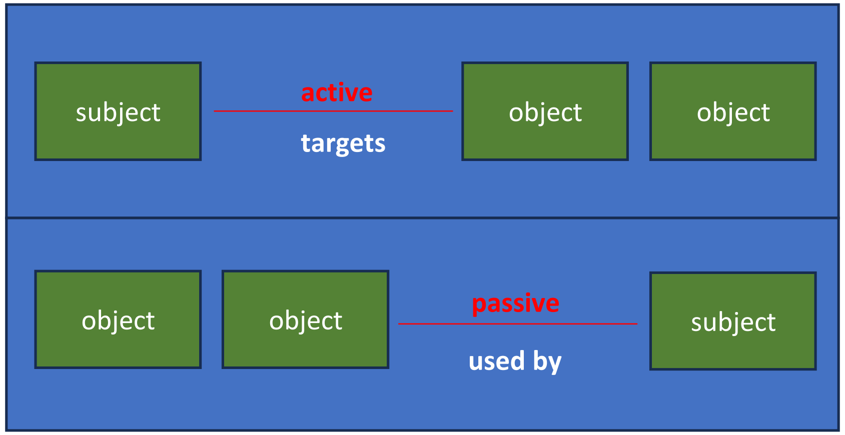
LLM

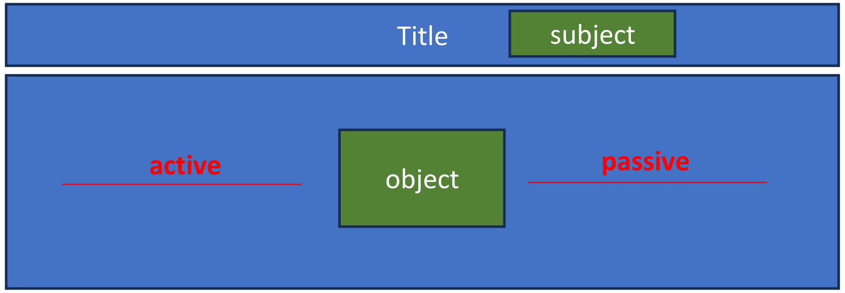
Give me relations between A-objects and B-objects!

Ok… let’s do fine-tuning and function calling…

Regex/keywords to extract relationships

1. Form a regex library
   - Relation 1: ‘regex_pattern1’, ‘regex_pattern2’, etc
   - Relation 2: ‘regex_pattern1’, ‘regex_pattern2’, etc
2. Tokenise and remove stop words
   - If not done on the previous steps, as this pre-processing could be already done for ML classification
   - This makes regex easier as reduced the variations of the words
3. Extract context around the entities
   - Search for patterns between the objects. Then if not found expand to sentence, paragraph, check titles
   - Could be “this threat actor” in the paragraph but the name of the entity in the title
4. Check the results manually
   - The process is prone to errors
   - Constant regex modifications



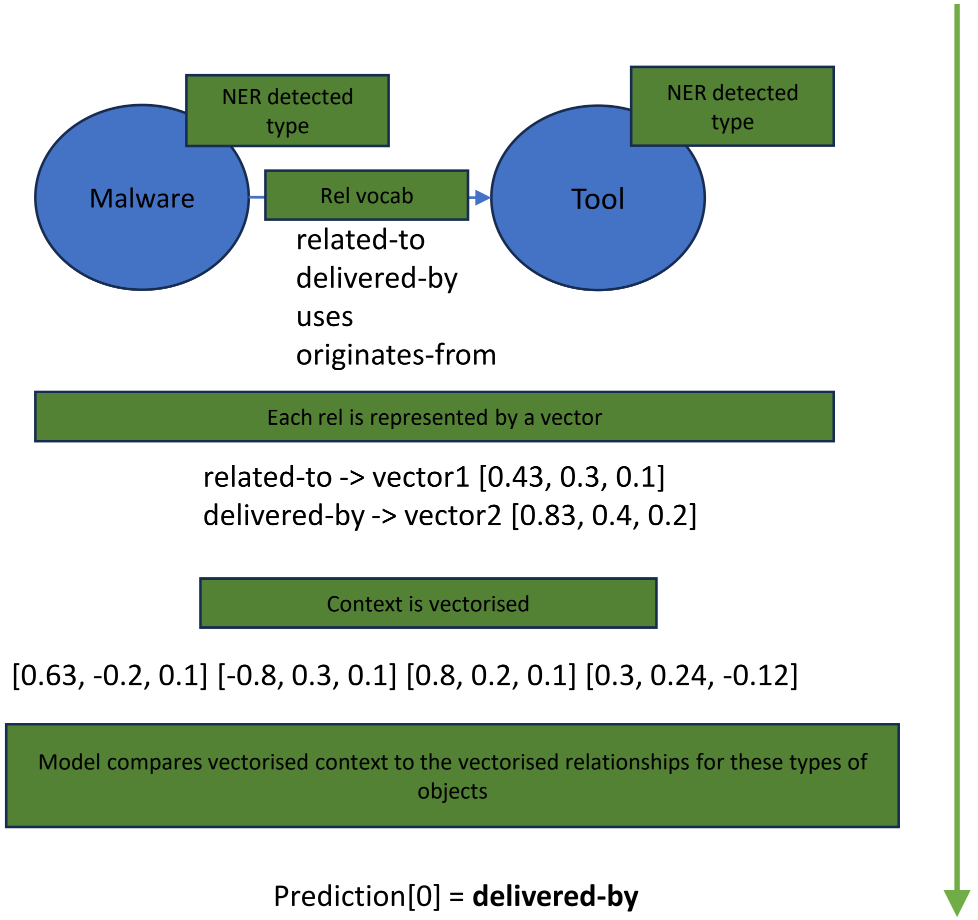


Building relationships using ML

1. Define Relationship Vocabulary
   - We are lucky to have STIX, but we are not limited by it
2. Extract context around the entities
   - How far? A couple of words? The boundary of the sentence? The paragraph? Consider titles? A combination of things?
3. Tokenise and remove stop words
   - If not done on the previous steps, as this pre-processing could be already done for ML classification
4. Feature extraction
   - Convert text to vectors (we need numbers)
5. Prepare a labelled dataset
   - Annotate relationships for the existing corpus of reports
6. Train a model
   - support vector machines (SVM), random forests, or neural networks
   - predict the relationship between pairs of objects
7. Fine-tune and apply the model

8. Continuous improvement

Simplified illustration of the method*



Malware

NER detected type

Rel vocab

Tool

NER detected type

related-to
delivered-by
uses
originates-from

Each rel is represented by a vector

related-to -> vector1 [0.43, 0.3, 0.1]
delivered-by -> vector2 [0.83, 0.4, 0.2]

Context is vectorised

[0.63, -0.2, 0.1] [-0.8, 0.3, 0.1] [0.8, 0.2, 0.1] [0.3, 0.24, -0.12]

Model compares vectorised context to the vectorised relationships for these types of objects

Prediction[0] = **delivered-by**

LLM: take it easy

1. Define Relationship Vocabulary

   In the prompt ask what you are looking for or use API to fetch. Use specific details about the format you expect

2. Define an assistant to set the right context

   Tell what the model should be an expert at

3. Add Function Calling

   Ask your API to give the names of the object of interest in the report

4. Post the whole thing; often no need to care finding context

   LLM ideally should find the context itself
   If you are sure that LLM will not miss anything, pass a certain section only
   A model has a limit on number of total tokens it consumes

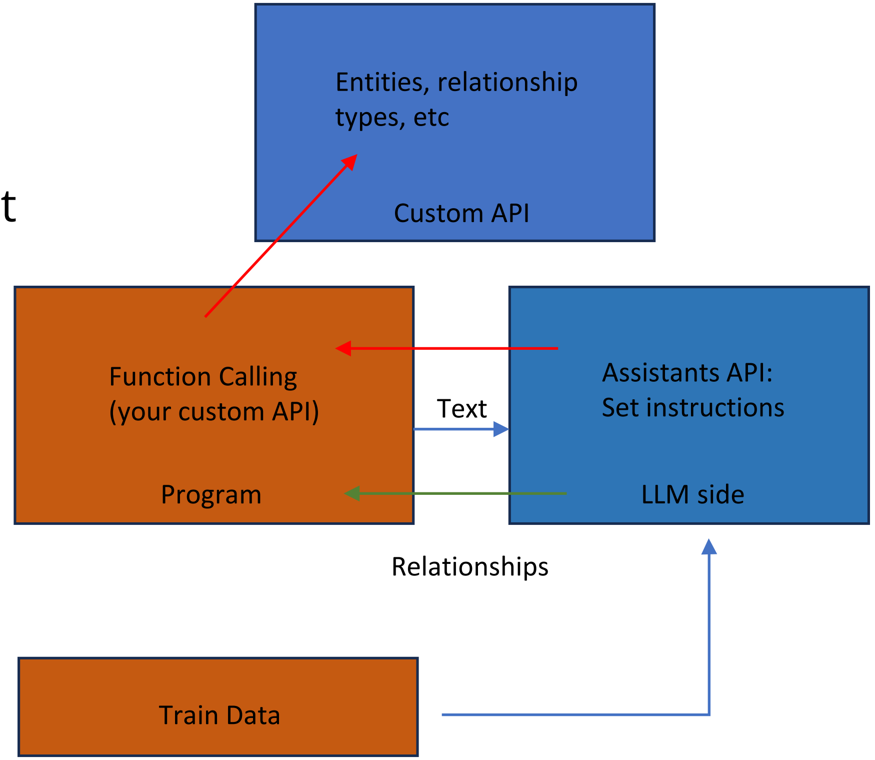
5. Error handling

   Wrong format
   Data quality questions if empty or expected more results
   LLM is not available / error during processing

6. Keep track of tokens consumed (input/output)

   No only billing but also to identify problems

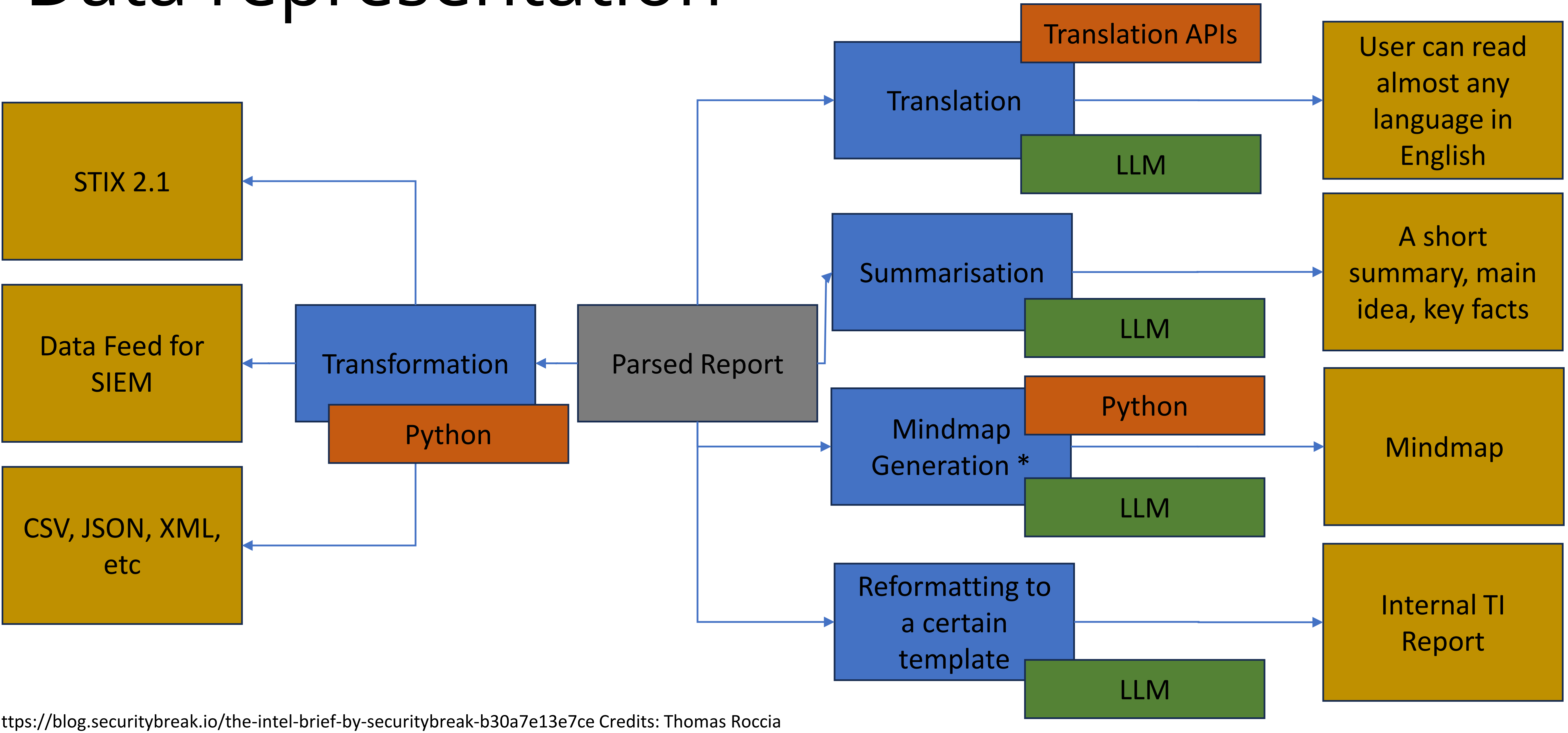
7. Fine tune by uploading training data


Entities, relationship types, etc
Custom API
Function Calling (your custom API)
Program
Text
Assistants API: Set instructions
LLM side
Relationships
Train Data

ML vs LLM for TI object relationship extraction

Feature	ML	Public LLM	Private LLM
Data residency	Full control	Your data becomes the part of the public model	Full control
Cost	Cheap	Cheap (extra cost if not just prompts)	Expensive
Quality	High	High	High
Effort to support	Average	Average	Very High
Qualification and skills required	High	Average	Very High

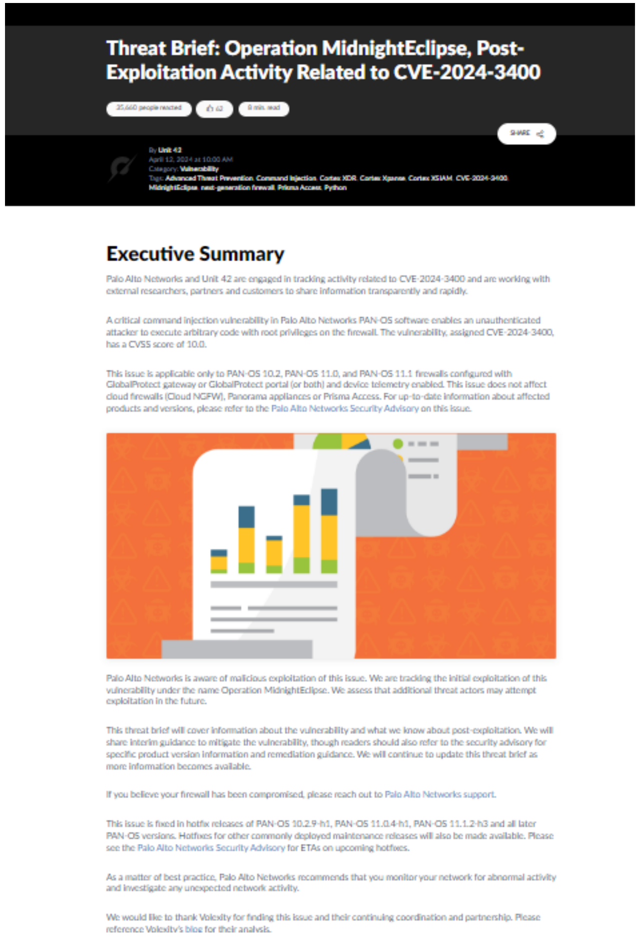
Data representation



* https://blog.securitybreak.io/the-intel-brief-by-securitybreak-b30a7e13e7ce Credits: Thomas Roccia

Input

Output



Threat Brief: Operation MidnightEclipse, Post-Exploitation Activity Related to CVE-2024-3400

35,660 people reacted · 62 · 8 min. read

SHARE

By Unit 42
April 12, 2024 at 10:00 AM
Category: Vulnerability
Tags: Advanced Threat Prevention, Command Injection, Cortex XDR, Cortex Xpanse, Cortex XSIAM, CVE-2024-3400, MidnightEclipse, next-generation firewall, Prisma Access, Python

Executive Summary

Palo Alto Networks and Unit 42 are engaged in tracking activity related to CVE-2024-3400 and are working with external researchers, partners and customers to share information transparently and rapidly.

A critical command injection vulnerability in Palo Alto Networks PAN-OS software enables an unauthenticated attacker to execute arbitrary code with root privileges on the firewall. The vulnerability, assigned CVE-2024-3400, has a CVSS score of 10.0.

This issue is applicable only to PAN-OS 10.2, PAN-OS 11.0, and PAN-OS 11.1 firewalls configured with GlobalProtect gateway or GlobalProtect portal (or both) and device telemetry enabled. This issue does not affect cloud firewalls (Cloud NGFW), Panorama appliances or Prisma Access. For up-to-date information about affected products and versions, please refer to the Palo Alto Networks Security Advisory on this issue.

Palo Alto Networks is aware of malicious exploitation of this issue. We are tracking the initial exploitation of this vulnerability under the name Operation MidnightEclipse. We assess that additional threat actors may attempt exploitation in the future.

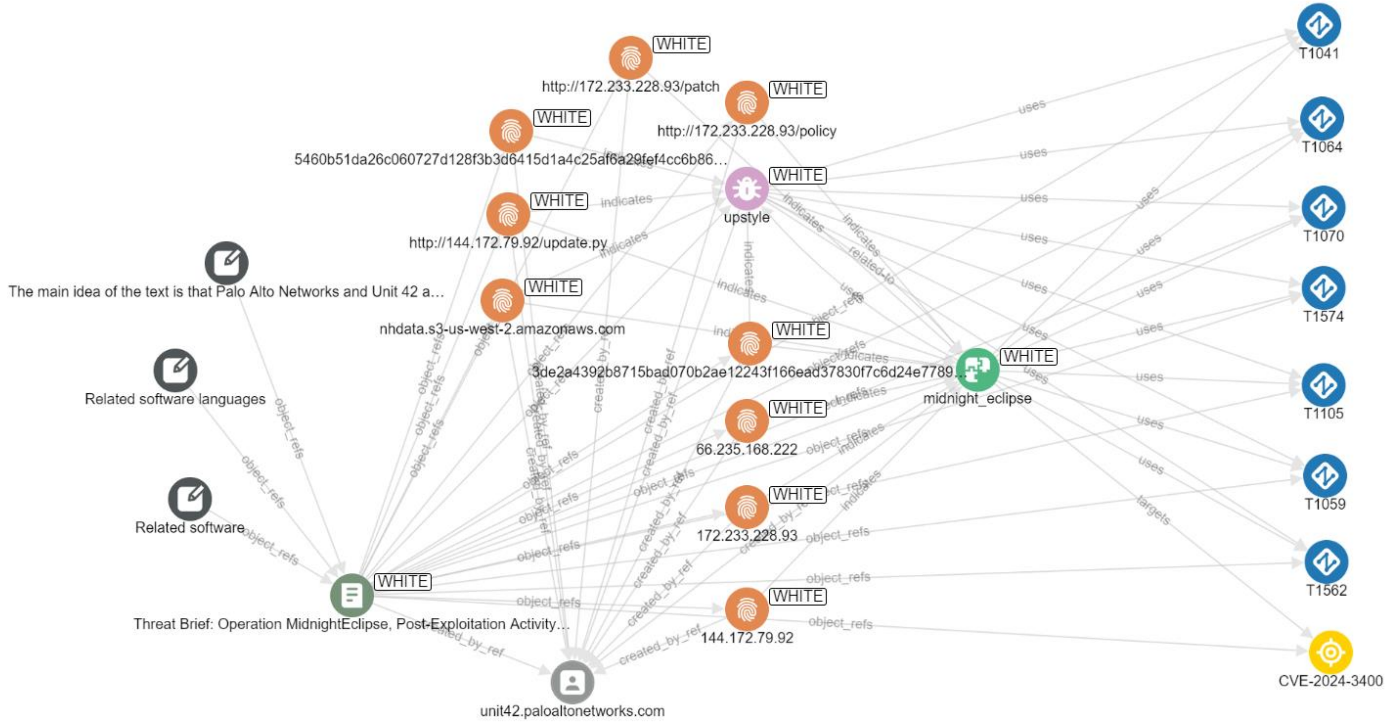
This threat brief will cover information about the vulnerability and what we know about post-exploitation. We will share interim guidance to mitigate the vulnerability, though readers should also refer to the security advisory for specific product version information and remediation guidance. We will continue to update this threat brief as more information becomes available.

If you believe your firewall has been compromised, please reach out to Palo Alto Networks support.

This issue is fixed in hotfix releases of PAN-OS 10.2.9-h1, PAN-OS 11.0.4-h1, PAN-OS 11.1.2-h3 and all later PAN-OS versions. Hotfixes for other commonly deployed maintenance releases will also be made available. Please see the Palo Alto Networks Security Advisory for ETAs on upcoming hotfixes.

As a matter of best practice, Palo Alto Networks recommends that you monitor your network for abnormal activity and investigate any unexpected network activity.

We would like to thank Volexity for finding this issue and their continuing coordination and partnership. Please reference Volexity's blog for their analysis.


WHITE
http://172.233.228.93/patch
http://172.233.228.93/policy
5460b51da26c060727d128f3b3d6415d1a4c25af6a29fef4cc6b86…
upstyle
http://144.172.79.92/update.py
The main idea of the text is that Palo Alto Networks and Unit 42 a…
nhdata.s3-us-west-2.amazonaws.com
3de2a4392b8715bad070b2ae12243f166ead37830f7c6d24e7789…
midnight_eclipse
Related software languages
66.235.168.222
Related software
172.233.228.93
Threat Brief: Operation MidnightEclipse, Post-Exploitation Activity…
144.172.79.92
unit42.paloaltonetworks.com
T1041
T1064
T1070
T1574
T1105
T1059
T1562
CVE-2024-3400
uses
indicates
related-to
targets
object_refs
created_by_ref

TI Report processing pipeline. Recap

Download, pre-translate and normalise	LLM can help with translation, image and text recovery, image recognition
Tokenisation	NER/ML is fine, but LLM helps
Classification	ML is fine and enough
Filtering and deduplication	ML is fine and enough
Entity Relation Extraction	Both approaches work, but I believe LLM will win
Transformation	LLM is handy

2024
FIRST
Cyber Threat
Intelligence
Conference

Berlin, Germany
April 15-17, 2024

Yury Sergeev
RST Cloud

ysergeev@rstcloud.net
https://www.rstcloud.com