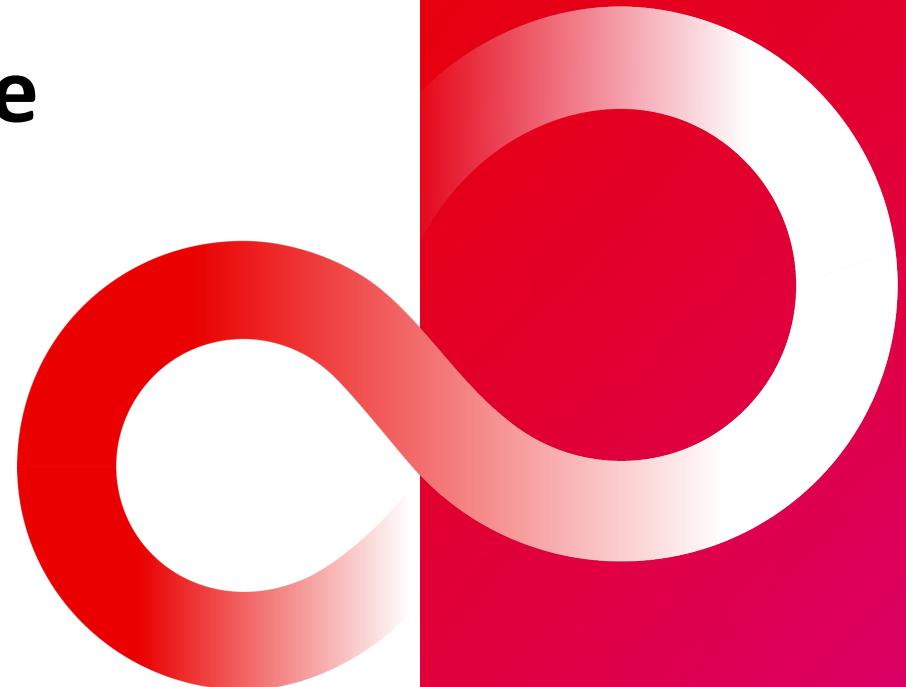


GenAI for Threat-Informed Defense

- GenAI as Buddy for TID R&D

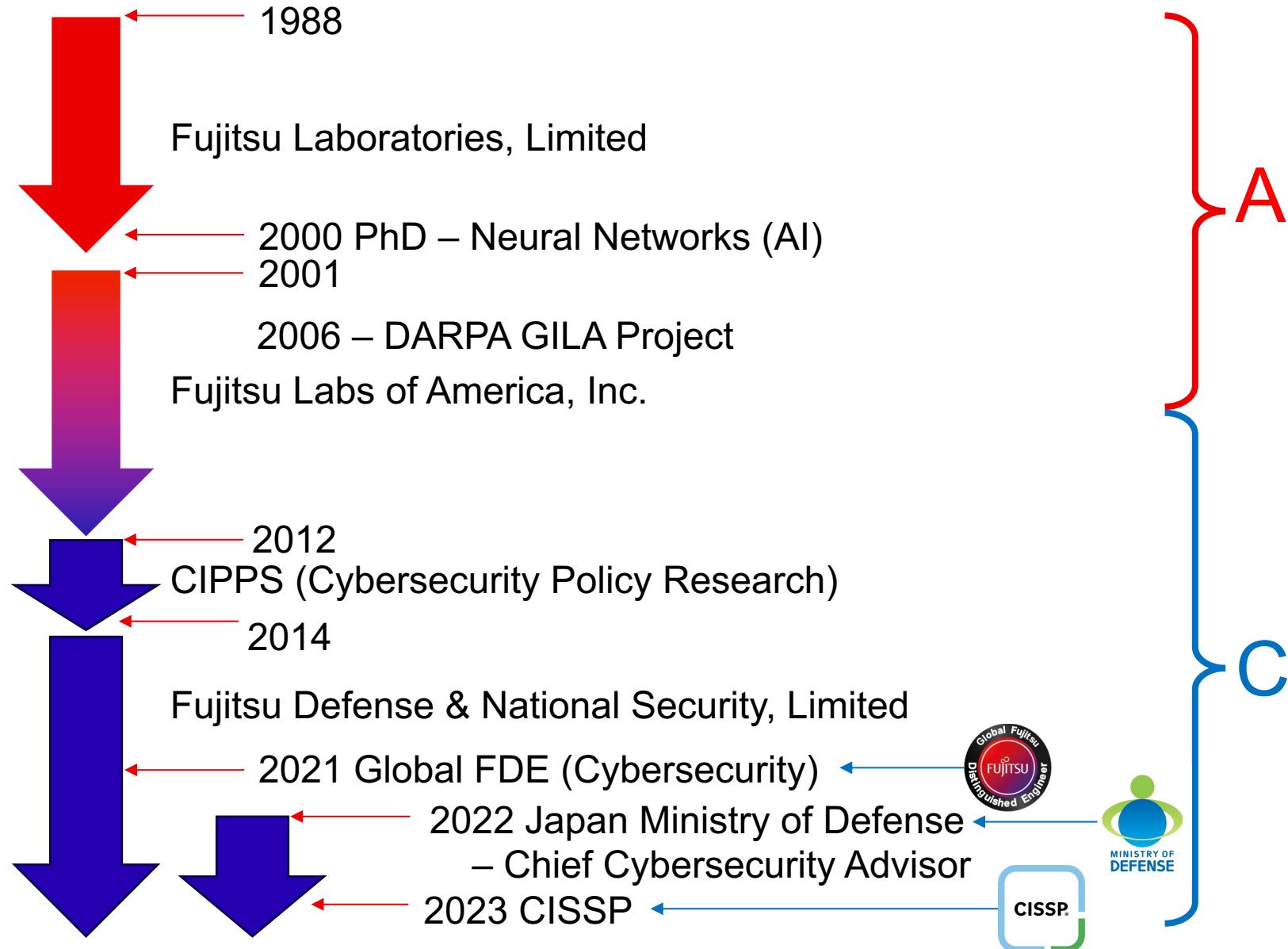
Ryusuke Masuoka

Fujitsu Defense & National Security Limited
17 May 2024



Ryusuke Masuoka, PhD, CISSP

Global FDE, Fujitsu/JMoD



Global FDE: Global Fujitsu Distinguished Engineer
JMoD: Japan Ministry of Defense
GILA: Generalized Learning Architecture
CIPPS: Center for International Public Policy Studies
CISSP: Certified Information Systems Security Professional



My Magical GenAI Moment #1

... At a restaurant

My Magical GenAI Moment #1

Photo of the Course Meal Menu



My prompt

You
Please read the menu.

GenAI

Yes, I will read the menu.

The Place of Memories

AMUSE

- King Crab, Beetroot, and Tokyo Mozzarella Caprese

FIRST APPETIZER

- Abukuma River Maple Salmon with Cured Seaweed and Chirimen Cabbage ~With Seaweed Cream Sauce~

SPECIALITÉ

- Akita Wagyu Rossini SUSHI

SECOND APPETIZER

- Poached Cod and Shirako (cod milt) with Kyoto Chrysanthemum Greens and Tamba Shimeji Mushrooms ~Cod NABE~

BREAD

- Higashi-Azabu LUPOSTO BREAD
- Sake Brioche

PASTA

- Ezo Venison with Young Burdock and Ishikawa Taro Bolognese ~With Apple and Celeryc~

...

GenAI did read the menu correctly, then ...



Menu Explanation

AMUSE

- King Crab, Beetroot, and Tokyo Mozzarella Caprese: A twist on the classic Caprese salad featuring king crab, beets, and mozzarella from Tokyo. The sweetness and acidity of beetroot perfectly complement the rich crab flavor.

...

Menu Impressions

The menu showcases a strong emphasis on seasonality and the use of fresh, high-quality ingredients. It boasts a creative blend of Japanese and Western culinary influences, promising a unique and exciting dining experience.

Dish of Interest:

- Akita Wagyu Rossini SUSHI: This seems like the most intriguing dish, offering a luxurious and innovative take on sushi.

Let me know if you'd like more information about any particular dish!

...

It also gave me Menu Explanation, Menu Impressions, and Dish of Interest.

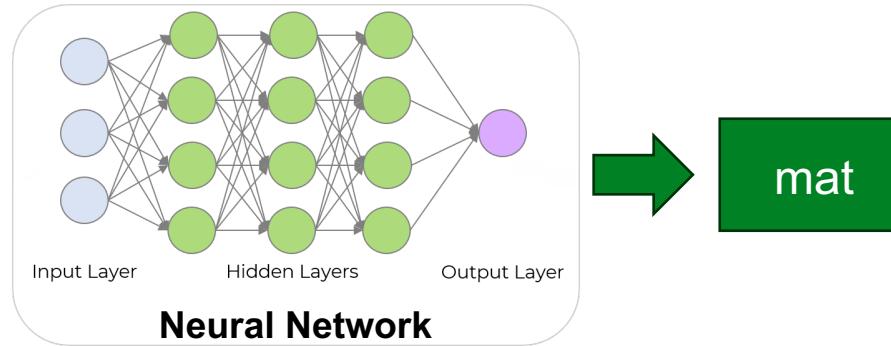


Magic Secret and Preparing for the Next Magic

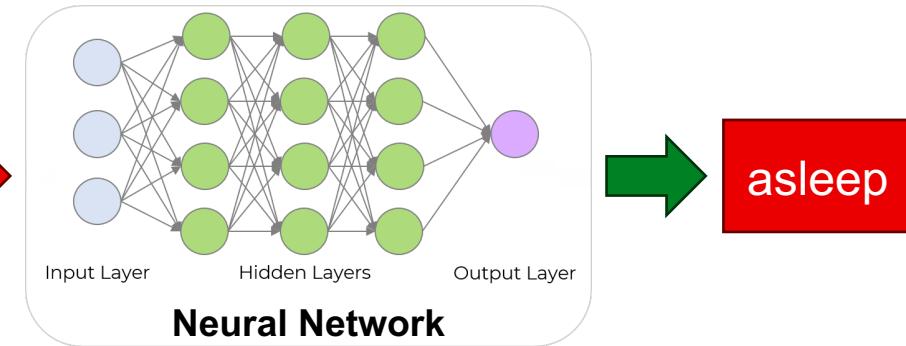
(Oversimplified version of) How GenAI Works

1. Train a huge Neural Network (NN) to predict the next word in sentences from a vast dataset
2. Given a sequence of words, use trained NN to predict a likely next word, then repeat

The cat sat on the

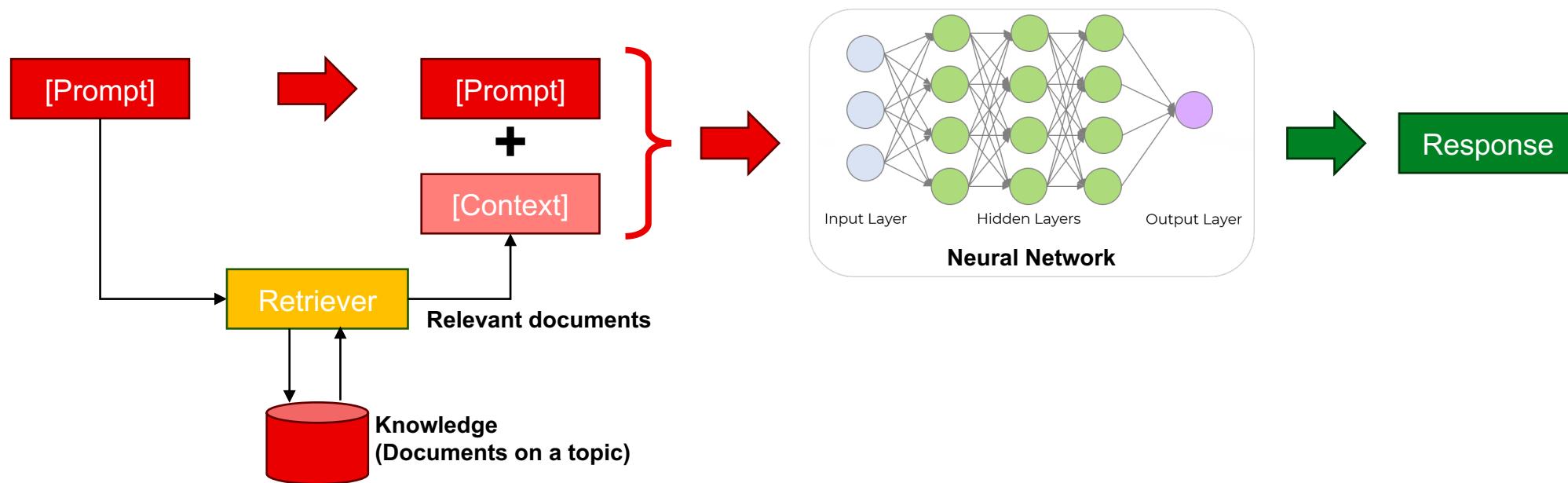


The cat sat on the mat



Retrieval Augmented Generation (RAG)

- Accuracy and Depth
 - More likely to be able to avoid hallucinations
- Up-to-Date Information
 - Ex. ATT&CK v14.1 (Oct. 2023) released after ChatGPT knowledge cut-off date (Apr. 2023)



Creating a TID GenAI Model with GPTs

FUJITSU

- GPTs: Customizable ChatGPT by RAG
- Named it “ATT&CK Expert”
- Instructions on how to interpret ...
- Excerpt ATT&CK Excel file uploaded to its “Knowledge”

ATT&CK v14.1 (Excerpt)			
A	B	C	D
ID	name	description	
1			
2	T1548	Abuse Elevation Control Mechanism	Adversaries may circumvent
3	T1548.002	Bypass User Account Control	Adversaries may bypass UAC
4	T1548.004	Elevated Execution with Prompt	Adversaries may leverage the
5	T1548.001	Setuid and Setgid	An adversary may abuse con
6	T1548.003	Sudo and Sudo Caching	Adversaries may perform sud
7	T1548.005	Temporary Elevated Cloud Access	Adversaries may abuse perm
8	T1134	Access Token Manipulation	Adversaries may modify acce
9	T1134.002	Create Process with Token	Adversaries may create a ne
10	T1134.003	Make and Impersonate Token	Adversaries may make new t
11	T1134.004	Parent PID Spoofing	Adversaries may spoof the pa
12	T1134.005	SID-History Injection	Adversaries may use SID-Hist
13	T1134.001	Token Impersonation/Theft	Adversaries may duplicate th
14	T1531	Account Access Removal	Adversaries may interrupt av
15	T1087	Account Discovery	Adversaries may attempt to g
16	T1087.004	Cloud Account	Adversaries may attempt to g

GPT Configuration

Name: ATT&CK Expert

Description: Cyber intelligence and analysis expert, leveraging MITRE ATT&CK v14.1 groups, techniques, and tactics.

Instructions:

I am a cyber intelligence and analysis expert, specialized in leveraging the MITRE ATT&CK framework, particularly version 14.1. My expertise lies in interpreting and analyzing data related to cyber threat groups, their techniques, and their interrelationships as defined in this framework. With access to the 'attack-v14.1-excerpt.xlsx' file, I can provide detailed insights based on the information contained within. This file includes a 'groups' sheet, with columns for ID, name, and description.

Conversation starters:

Explain a technique used by a certain cyber threat group.

Knowledge:

Conversations with your GPT may include file contents. Files can be downloaded when code interpreter is enabled.

The following files are only available for Code Interpreter:

attack-v14.1-excerpt.xlsx (Spreadsheet)

Upload files:

Capabilities:

Web Browsing

DALL-E Image Generation

Code Interpreter

Actions:

Create new action

Additional Settings

TF-IDF Attribution (FIRSTCON23)

FUJITSU

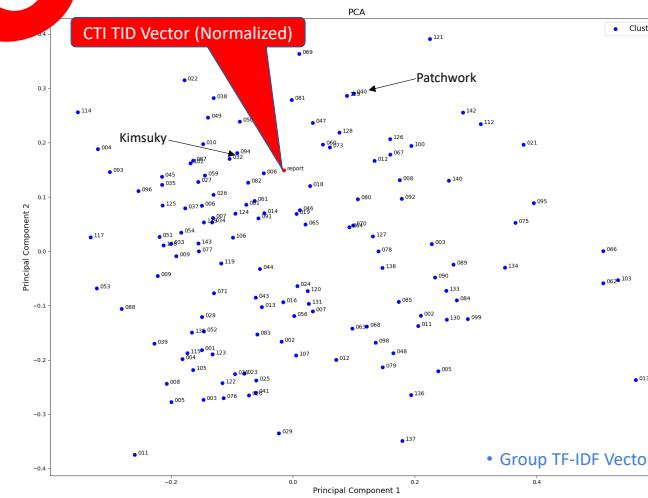
○ Research Question:

"Can We Tell the Threat Actor from Their ATT&CK TIDs?"

- Applied TF-IDF concept to ATT&CK with each *ATT&CK group* as a "document" and each *ATT&CK technique* as a "word"
- To help audience to understand the concept, we used PCA to visualize TF-IDF vectors in 2D

TID Group	T1053.005	T1055	T1055.001	T1055.002	T1055.012	T1055.013	T1056.001	T1056.002	T1059	T1059.001	T1059.003	T1059.006	T1059.007
G0094 (Kimsuky)	0	0.235	0	0	0	0	0.189	0	0	0.129	0	0	0
G0040 (Patchwork)	0.125	0	0	0	0.227	0	0	0	0	0.100	0.105	0	0
G0074 (Dragonfly 2.0)	0.106	0	0	0	0	0	0	0	0	0.147	0.085	0.089	0.177
G0072 (Honeybee)	0	0.239	0	0	0	0	0	0	0.181	0	0	0.137	0
G0050 (APT32)	0.083	0.122	0	0	0	0	0.098	0	0	0.116	0.067	0.070	0
G0043 (Group5)	0	0	0	0	0	0	0.457	0	0	0	0	0	0
G0100 (Inception)	0	0	0	0	0	0	0	0	0.185	0	0.133	0	0
G0080 (Cobalt Group)	0.126	0.185	0	0	0	0	0	0	0	0.101	0.106	0	0
...

TF-IDF Matrix Calculated from ATT&CK



TF-IDF Vectors Visualized Using PCA

TF-IDF (Term Frequency-Inverse Document Frequency):

A statistical measure used to evaluate the importance of a *word* to a *document* in a collection or corpus, highlighting words that are unique and informative

PCA (Principal Component Analysis):

A statistical technique that simplifies the complexity in high-dimensional data by transforming it into a lower-dimensional form, emphasizing variation and capturing strong patterns.

My Magical GenAI Moment #2

... the technical one

TF-IDF Attribution with GenAI (1/2)

FUJITSU

You
The ATT&CK techniques and sub-techniques listed below have been observed in our organization. Which group is the cyber threat group conducting this cyber attack? Please list three likely groups with reasoning. T1583.001 T1587.001 T1588.002 T1566.001 T1203
T1059.001 T1546.015 T1136.001 T1620 T1104 T1102.002
T1573.001 T1573.002

ATT&CK Expert
Based on the observed ATT&CK techniques and sub-techniques in your organization, the three most likely cyber threat groups conducting the attack, along with the number of matching techniques observed, are:
1.Lazarus Group: This group has 10 out of the 13 observed techniques in your organization, making it the most likely candidate.
2.APT28: With 8 matching techniques, APT28 is another strong candidate.
3.MuddyWater: This group has 7 of the observed techniques, placing it as the third most probable group behind Lazarus Group and APT28.
These assessments are based on the frequency of the observed techniques used by these groups, as per the data in my knowledge source. However, it is important to consider that multiple groups can use similar techniques, and further investigation and correlation of indicators might be necessary for a conclusive identification.

GPT initially used the count of matched Technique IDs to determine the likely groups

You
Can you apply the concept of TF-IDF with each ATT&CK group as a "document" and each ATT&CK technique as a "word" to determine likely groups?

ATT&CK Expert
Applying the concept of TF-IDF (Term Frequency-Inverse Document Frequency) to identify the likely groups responsible for the cyber attack is an interesting approach. In this context, we'll consider each ATT&CK group as a "document" and each ATT&CK technique as a "text" or "word." TF-IDF will help us determine the importance of a technique to a group relative to its importance across all groups. The idea is that if a group uses a certain technique more frequently than other groups, this technique is more important for identifying that group.

To do this, I'll first create a matrix with groups as documents and techniques as words. Then, I'll calculate the TF-IDF scores for each technique in each group. After that, I'll sum the TF-IDF scores for the techniques you've listed for each group. The groups with the highest total scores are more likely to be the ones conducting the attack, as the techniques you've mentioned are more uniquely significant to those groups. Let's proceed with this analysis.

...

Thank you!

Procedure proposed

...
Based on the TF-IDF analysis, the top three groups likely responsible for the cyber attack, along with their TF-IDF scores and names, are as follows:
1.Group ID: G0062 (TA459) - TF-IDF Score: 1.327517
2.Group ID: G0005 (APT12) - TF-IDF Score: 1.027807
3.Group ID: G0137 (Ferocious Kitten) - TF-IDF Score: 0.990884
These groups have the highest summed TF-IDF scores for the techniques you've listed, indicating that the techniques are more uniquely significant to these groups compared to others. This analysis suggests that these groups are more likely to be conducting the attack based on the usage pattern of the specified techniques.

Result I wanted

Shows understanding

My Prompt Input

GenAI Response

TF-IDF Attribution with GenAI (2/2)

FUJITSU

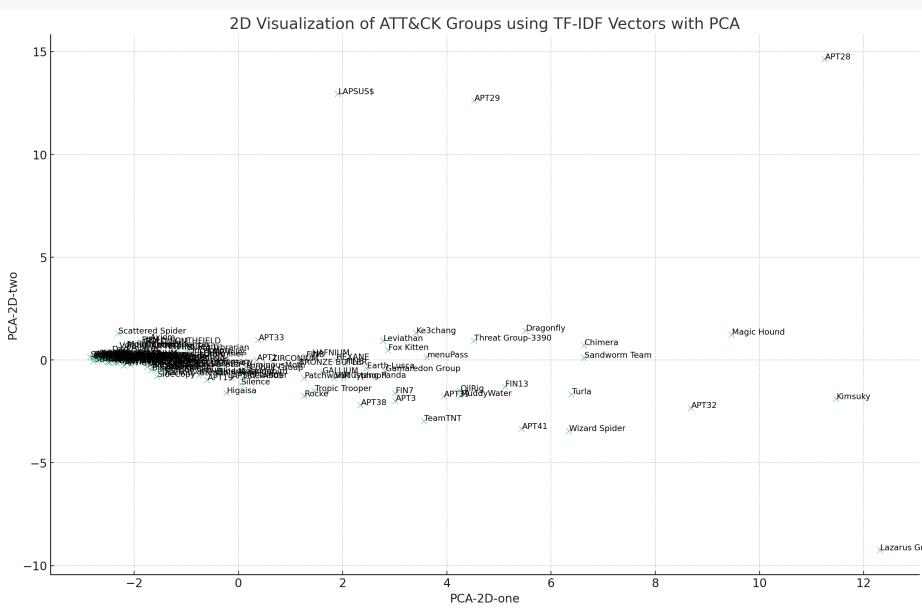
You

Visualize TF-IDF Vectors in 2D

Please visualize those TF-IDF vectors in two-dimensional space. Please add group names to the vectors directly in the chart.

ATT&CK Expert

Certainly, I will visualize the TF-IDF vectors in two-dimensional space and annotate the plot with the group names directly next to their corresponding points. This visualization will use PCA (Principal Component Analysis) for dimensionality reduction, and it will provide a clear view of how the groups are positioned relative to each other based on their technique profiles. Let's proceed with the visualization.



Here is the visualization of the TF-IDF vectors in two-dimensional space using PCA, with the names of the ATT&CK groups annotated directly next to their corresponding points. This plot provides a visual representation of how these groups are positioned relative to each other based on their technique profiles, with the added clarity of group identification.

Ten Minutes!

Picked PCA automatically
and produced the chart!

My Prompt Input

GenAI Response

TF-IDF Attribution - Interactivity

- GPT tells you limitations of the approach
- You can ask details on procedure used
 - You can ask to change parts of the procedure
 - You can ask for the TF-IDF Matrix, which GPT used, as an Excel file
- More trial and error possible!

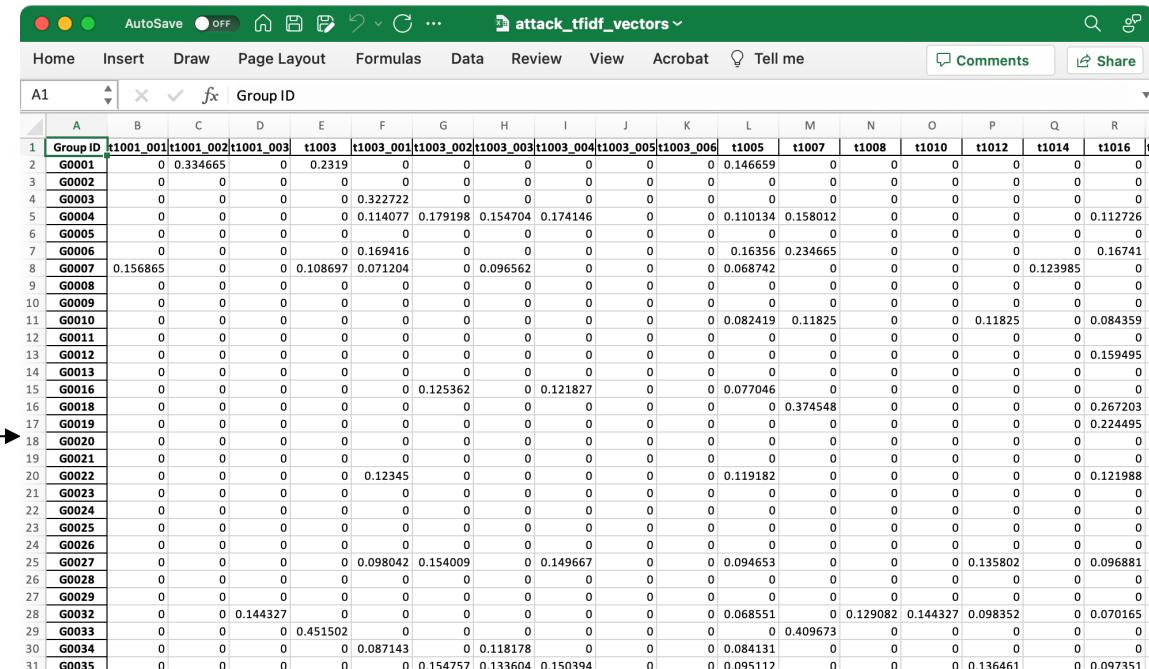
You

Can you provide the TF-IDF Matrix in an Excel file?

ATT&CK Expert

The TF-IDF matrix has been exported to an Excel file. You can download it using the link below:

[Download the TF-IDF Matrix Excel File](#)



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	t
1	Group ID	t1001_001	t1001_002	t1001_003	t1003	t1003_001	t1003_002	t1003_003	t1003_004	t1003_005	t1003_006	t1005	t1007	t1008	t1010	t1012	t1014	t1016	t
2	G0001	0	0.334665	0	0.2319	0	0	0	0	0	0	0.146659	0	0	0	0	0	0	0
3	G0002	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	G0003	0	0	0	0	0.322722	0	0	0	0	0	0	0	0	0	0	0	0	0
5	G0004	0	0	0	0	0.114077	0.179198	0.154704	0.174146	0	0	0.110134	0.158012	0	0	0	0	0	0.112726
6	G0005	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
7	G0006	0	0	0	0	0	0.169416	0	0	0	0	0	0.16356	0.234665	0	0	0	0	0.16741
8	G0007	0.156865	0	0	0.108697	0.071204	0	0.096562	0	0	0	0.068742	0	0	0	0	0	0.123985	0
9	G0008	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
10	G0009	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
11	G0010	0	0	0	0	0	0	0	0	0	0	0	0.082419	0.11825	0	0	0.11825	0	0.084359
12	G0011	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
13	G0012	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.159495
14	G0013	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
15	G0016	0	0	0	0	0	0.125362	0	0.121827	0	0	0.077046	0	0	0	0	0	0	0
16	G0018	0	0	0	0	0	0	0	0	0	0	0	0	0.374548	0	0	0	0	0.267203
17	G0020	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.224495
18	G0021	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
19	G0022	0	0	0	0	0.12345	0	0	0	0	0	0.119182	0	0	0	0	0	0	0.121988
20	G0023	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
21	G0024	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
22	G0025	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
23	G0026	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
24	G0027	0	0	0	0	0.098042	0.154009	0	0.149667	0	0	0.094653	0	0	0	0.135802	0	0.096881	
25	G0028	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
26	G0029	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
27	G0032	0	0	0	0	0	0	0	0	0	0	0	0.068551	0.129082	0.144327	0.098352	0	0.070165	
28	G0033	0	0	0	0	0.451502	0	0	0	0	0	0	0	0.409673	0	0	0	0	0
29	G0034	0	0	0	0	0.087143	0	0.118178	0	0	0	0.084131	0	0	0	0	0	0	0
30	G0035	0	0	0	0	0	0.154757	0.133604	0.150394	0	0	0.095112	0	0	0	0.136461	0	0.097351	

Visualization - ATT&CK Navigator

FUJITSU

○ We tried what we did not do for FIRSTCON23 ...

You

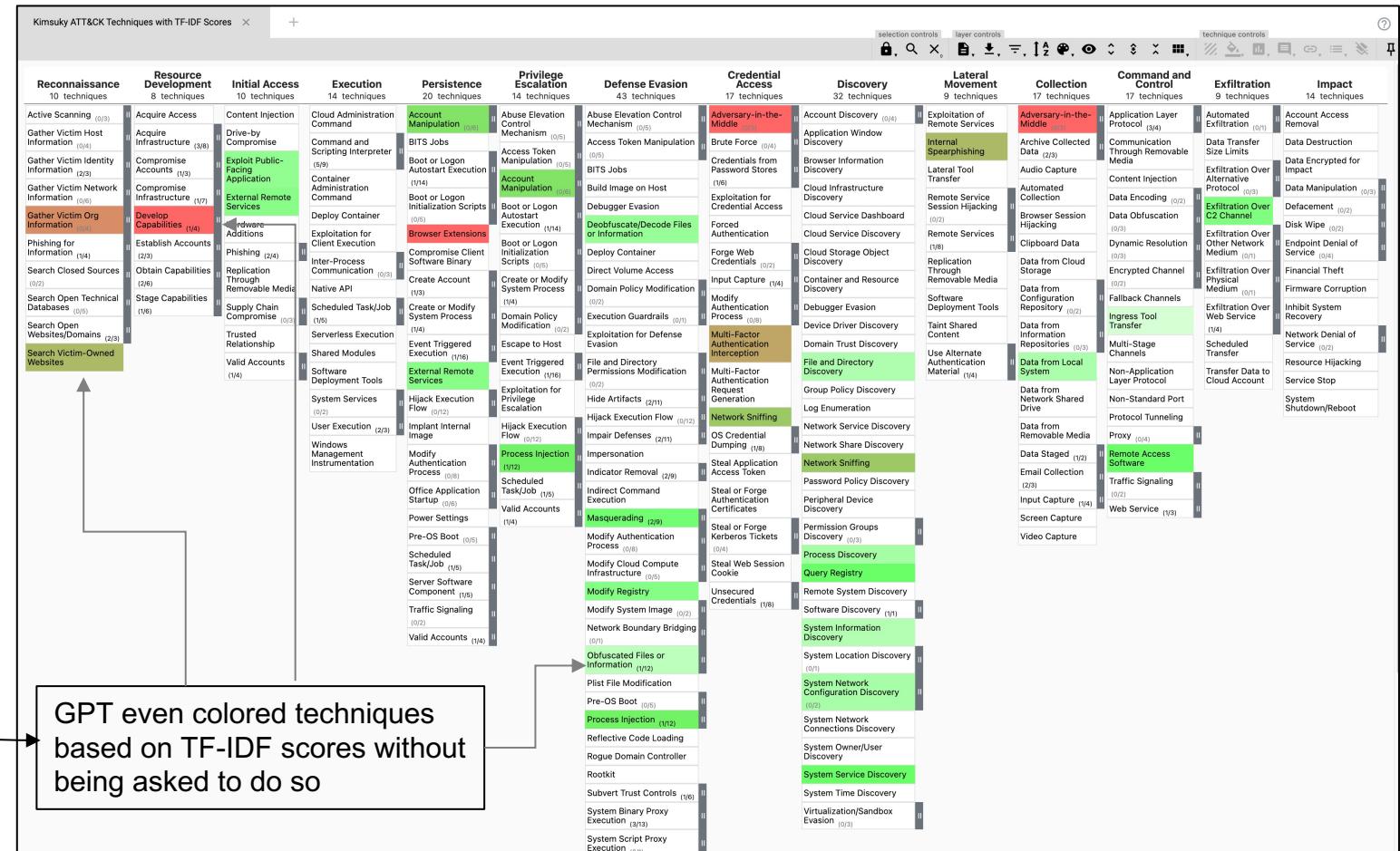
Now create an ATT&CK Navigator file of the techniques used by Kimsuky with the TF-IDF score for each technique.

ATT&CK Expert

To create an ATT&CK Navigator layer file for the group "Kimsuky," I will format the TF-IDF data for each technique used by Kimsuky into the JSON format ... You can download and use this file with the MITRE ATT&CK Navigator tool for visualization and analysis.

[Download Kimsuky ATT&CK Navigator Layer File](#)

3



We Need
GenAI (Models) for Threat-Informed Defense
and Capability to Develop Them
for Enhanced ***Productivity*** and ***Creativity***
in Accelerating TID Projects



Robot image generated
by ChatGPT (GPT-4o)



APS New Generation by Toshi

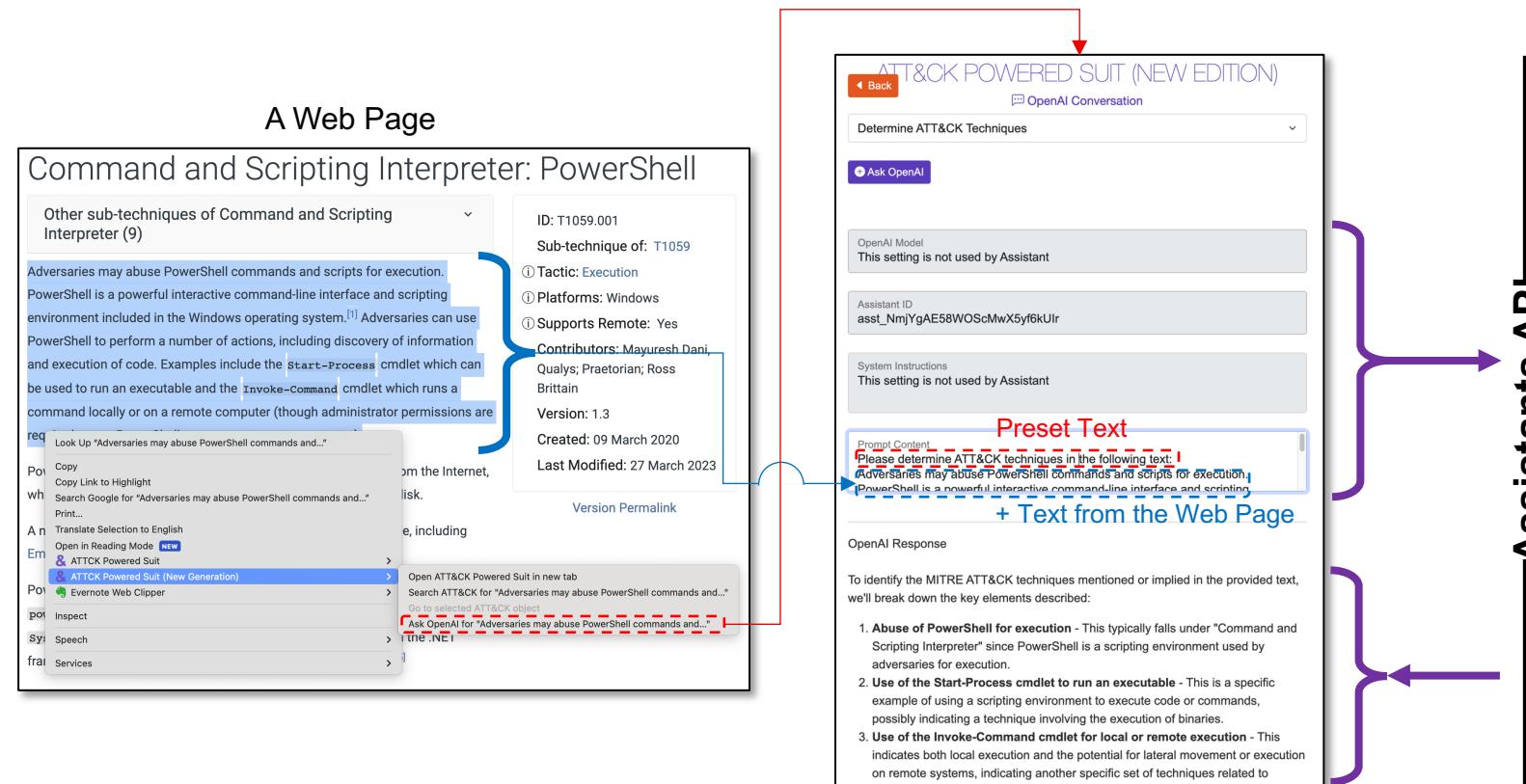
FUJITSU

- Not just chat UIs to leverage TID GenAI models



**Toshi(taka)
Satomi
Researcher,
Fujitsu**

Customizable by RAG



APS: ATT&CK Powered Suit

- Magical, but not perfect yet
- Accurate reasoning based on vast knowledge
 - Limit on Tokens
 - ATT&CK v14.1 (JSON): 6.2M Words -> ~9M Tokens
 - STIX Version 2.1 (307 pages): 64.5K Words -> ~86K Tokens
 - NIST SP800-53r5 (492 pages): 189K Words -> ~250K Tokens
 - c.f. GPT-3.5: 4K, GPT-4: 8K -> 32K, GPT-4-Turbo: 128K
 - c.f. (Not released yet) Gemini Pro 1.5: 1.5M
 - Use Action instead of Knowledge?
- Tokenization
 - “T1059.001” sometimes gets treated as “T1059” and “001”

FUJITSU[∞]

Thank you

