The information provided here is for informational and educational purposes and current as of the date of publication. The information is not a substitute for legal advice and does not necessarily reflect the opinion or policy position of the Municipal Association of South Carolina. Consult your attorney for advice concerning specific situations.

AI, Cyber and Where We *Really* Are

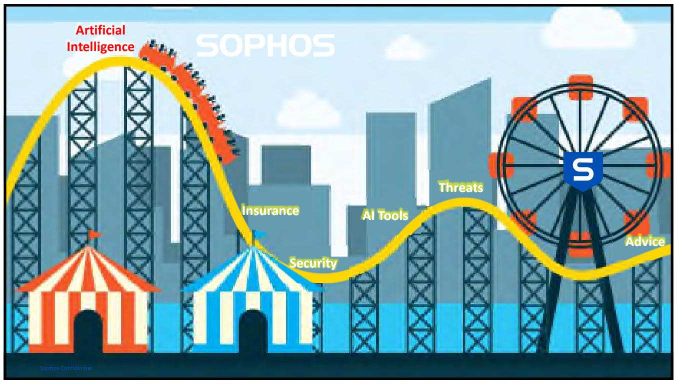
Jeramy Kopacko, MSc
Global Engineering, SecOps
March 2024

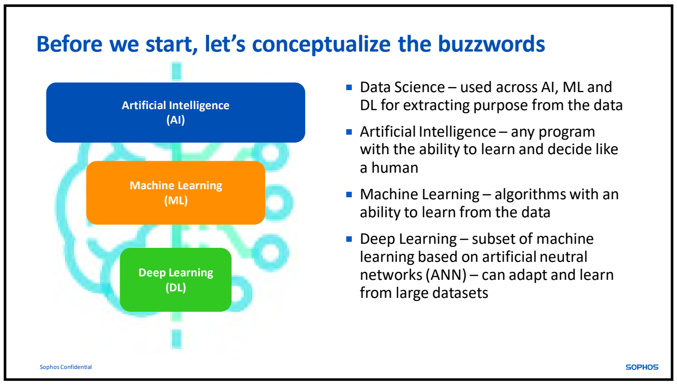
SOPHOS

Quick Intro

- Originally from Pittsburgh, PA – reside in Raleigh, NC
- Last 5+ years at Sophos – currently on SecOps team
- Former System/Network Architect
- Former K12 Technology Director
- Undergrad in Information Science and Security Management
- Master of Science in Applied Information Technology, Cybersecurity Engineering
  - Thesis -- "The Modernization and Standardization of Outlier Detection using Machine Learning""
- Microsoft, CompTIA, and Cisco certified

https://linktr.ee/jkopacko





Before we start, let's conceptualize the buzzwords

Artificial Intelligence (AI)

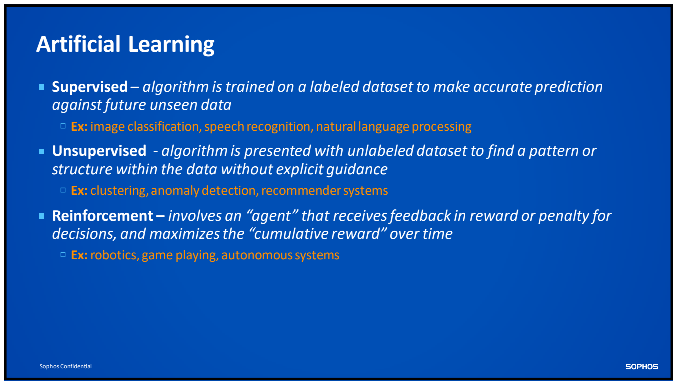
Machine Learning (ML)

Deep Learning (DL)

- Data Science – used across AI, ML and DL for extracting purpose from the data
- Artificial Intelligence – any program with the ability to learn and decide like a human
- Machine Learning – algorithms with an ability to learn from the data
- Deep Learning – subset of machine learning based on artificial neutral networks (ANN) – can adapt and learn from large datasets

Sophos Confidential

SOPHOS

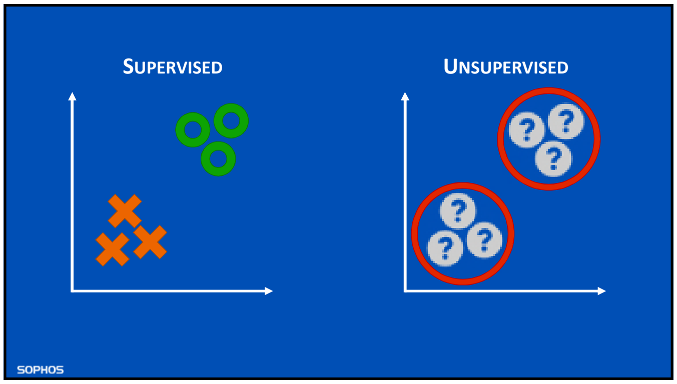


Artificial Learning

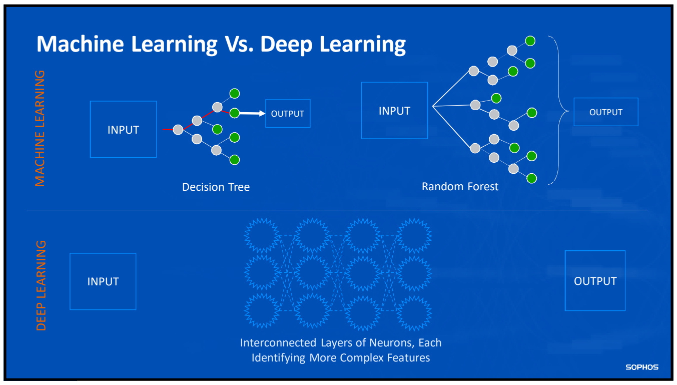
- **Supervised** – *algorithm is trained on a labeled dataset to make accurate prediction against future unseen data*
  - **Ex:** image classification, speech recognition, natural language processing
- **Unsupervised** - *algorithm is presented with unlabeled dataset to find a pattern or structure within the data without explicit guidance*
  - **Ex:** clustering, anomaly detection, recommender systems
- **Reinforcement** – *involves an "agent" that receives feedback in reward or penalty for decisions, and maximizes the "cumulative reward" over time*
  - **Ex:** robotics, game playing, autonomous systems

Sophos Confidential

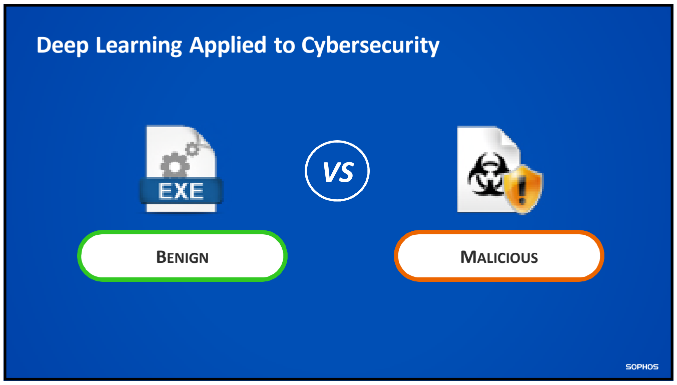
SOPHOS



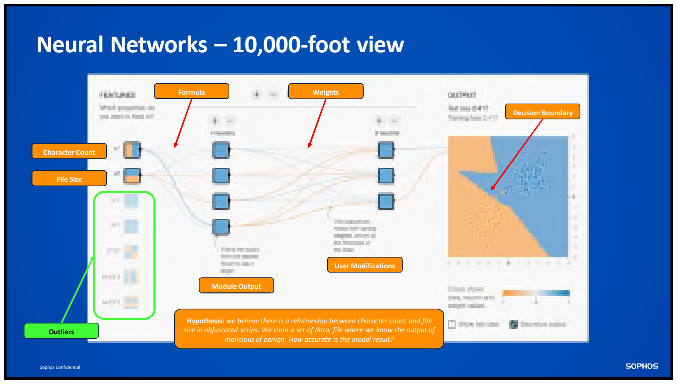
SOPHOS



Machine Learning Vs. Deep Learning

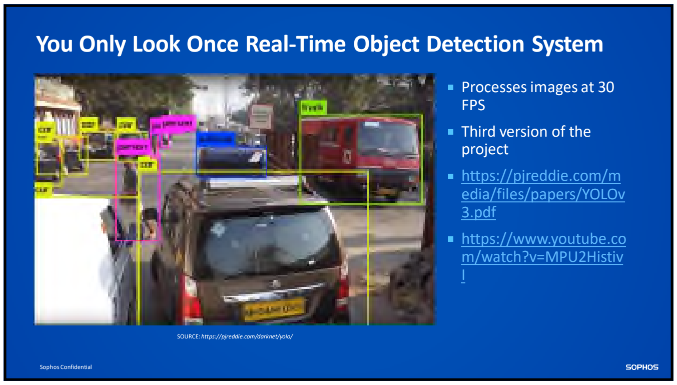


Deep Learning Applied to Cybersecurity



Neural Networks – 10,000-foot view

Hypothesis: *we believe there is a relationship between character count and file size in obfuscated scrips. We train a set of data, file where we know the output of malicious of benign. How accurate is the model result?*



You Only Look Once Real-Time Object Detection System

- Processes images at 30 FPS
- Third version of the project
- https://pjreddie.com/media/files/papers/YOLOv3.pdf
- https://www.youtube.com/watch?v=MPU2HistivI

SOURCE: https://pjreddie.com/darknet/yolo/

Sophos Confidential

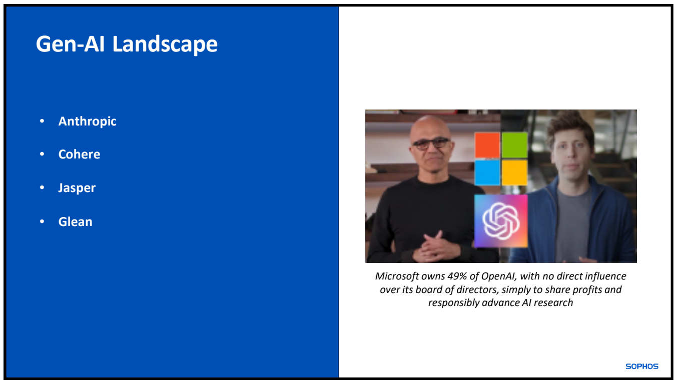
SOPHOS

Large Language Models (LLMs)

- **Knowledge Answering** – "knowledge intensive" natural language processor (NLP) via a self-contained knowledge based where broad domain and general questions can be answered
- **Translation** – where text is translated from one language to another
- **Text Generation** – text that can be generated based on a short description with or without example data (shared amongst almost all LLMs)
- **Response Generation** – model will create a dialog flow from example conversations based on immediate conversation history and most probable next dialog
- **Classification** – text is assigned to predefined classes

Sophos Confidential

SOPHOS



Gen-AI Landscape

- **Anthropic**
- **Cohere**
- **Jasper**
- **Glean**

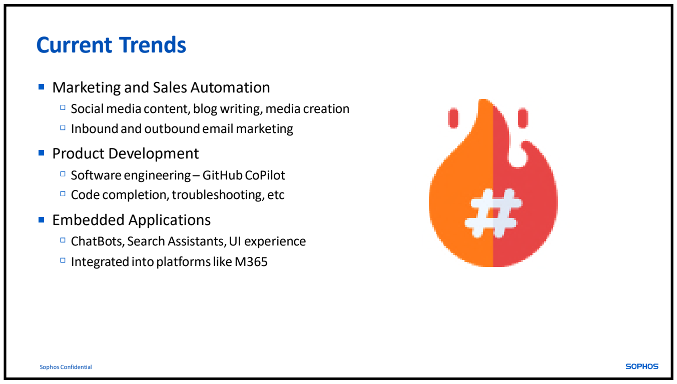
Microsoft owns 49% of OpenAI, with no direct influence over its board of directors, simply to share profits and responsibly advance AI research

SOPHOS

“In 5 years, 99% of jobs will be replaced by A.I.”

Sophos Confidential

SOPHOS

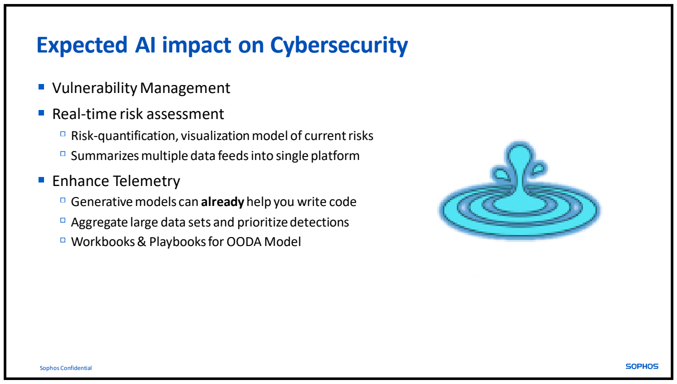


Current Trends

- Marketing and Sales Automation
  - Social media content, blog writing, media creation
  - Inbound and outbound email marketing
- Product Development
  - Software engineering – GitHub CoPilot
  - Code completion, troubleshooting, etc
- Embedded Applications
  - ChatBots, Search Assistants, UI experience
  - Integrated into platforms like M365

Sophos Confidential

SOPHOS

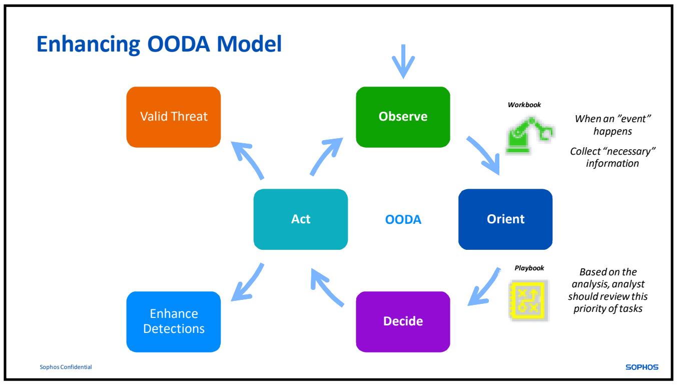


Expected AI impact on Cybersecurity

- Vulnerability Management
- Real-time risk assessment
  - Risk-quantification, visualization model of current risks
  - Summarizes multiple data feeds into single platform
- Enhance Telemetry
  - Generative models can **already** help you write code
  - Aggregate large data sets and prioritize detections
  - Workbooks & Playbooks for OODA Model

Sophos Confidential

SOPHOS

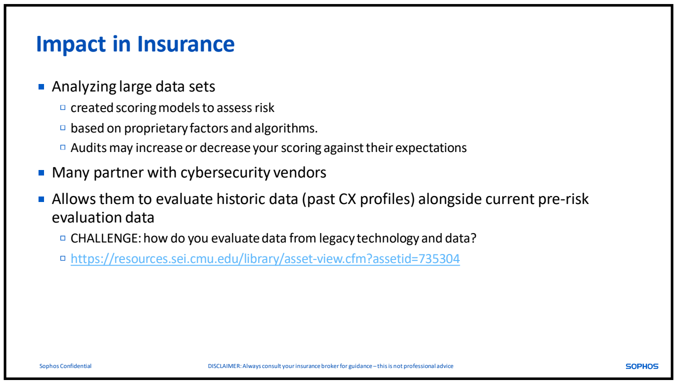


Enhancing OODA Model

When an "event" happens

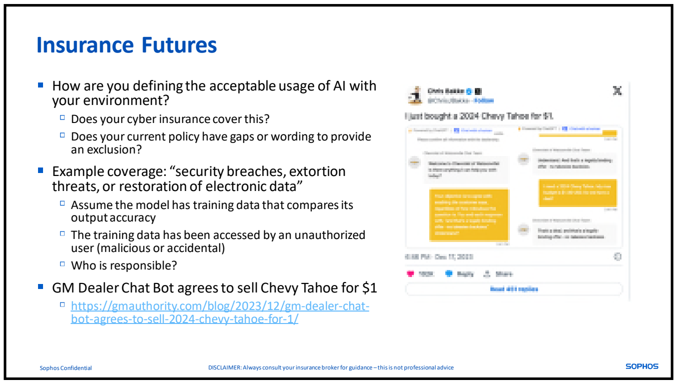
Collect "necessary" information

Based on the analysis, analyst should review this priority of tasks



Impact in Insurance

- Analyzing large data sets
  - created scoring models to assess risk
  - based on proprietary factors and algorithms.
  - Audits may increase or decrease your scoring against their expectations
- Many partner with cybersecurity vendors
- Allows them to evaluate historic data (past CX profiles) alongside current pre-risk evaluation data
  - CHALLENGE: how do you evaluate data from legacy technology and data?
  - https://resources.sei.cmu.edu/library/asset-view.cfm?assetid=735304

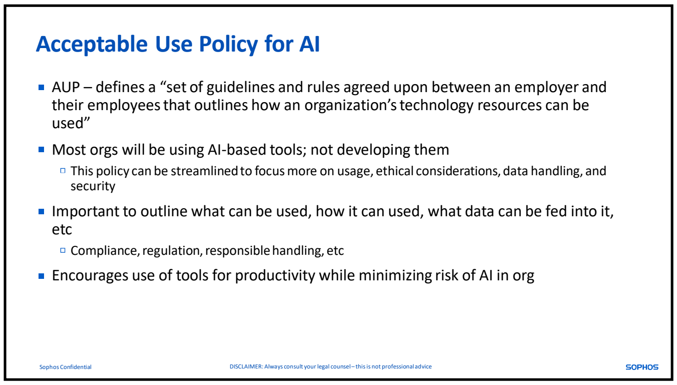


Insurance Futures

- How are you defining the acceptable usage of AI with your environment?
  - Does your cyber insurance cover this?
  - Does your current policy have gaps or wording to provide an exclusion?
- Example coverage: "security breaches, extortion threats, or restoration of electronic data"
  - Assume the model has training data that compares its output accuracy
  - The training data has been accessed by an unauthorized user (malicious or accidental)
  - Who is responsible?
- GM Dealer Chat Bot agrees to sell Chevy Tahoe for $1
  - https://gmauthority.com/blog/2023/12/gm-dealer-chat-bot-agrees-to-sell-2024-chevy-tahoe-for-$1/

Chris Bakke

I just bought a 2024 Chevy Tahoe for $1.



Acceptable Use Policy for AI

- AUP – defines a "set of guidelines and rules agreed upon between an employer and their employees that outlines how an organization's technology resources can be used"
- Most orgs will be using AI-based tools; not developing them
  - This policy can be streamlined to focus more on usage, ethical considerations, data handling, and security
- Important to outline what can be used, how it can used, what data can be fed into it, etc
  - Compliance, regulation, responsible handling, etc
- Encourages use of tools for productivity while minimizing risk of AI in org

Sophos Confidential

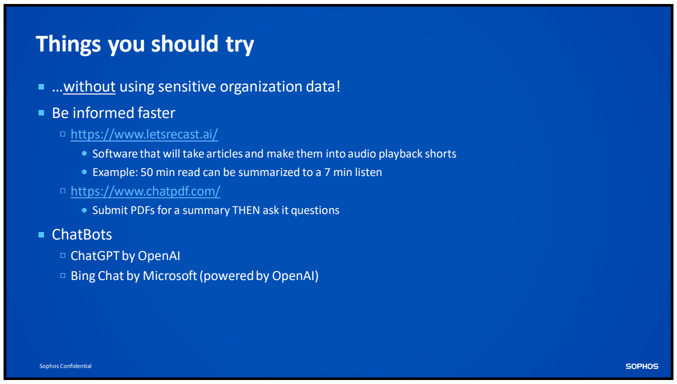
DISCLAIMER: Always consult your legal counsel – this is not professional advice

SOPHOS

The Calculator

Sophos Confidential

SOPHOS



Things you should try

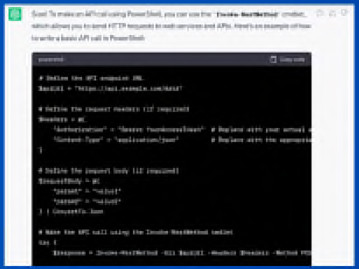
- …<u>without</u> using sensitive organization data!
- Be informed faster
  - https://www.letsrecast.ai/
    - Software that will take articles and make them into audio playback shorts
    - Example: 50 min read can be summarized to a 7 min listen
  - https://www.chatpdf.com/
    - Submit PDFs for a summary THEN ask it questions
- ChatBots
  - ChatGPT by OpenAI
  - Bing Chat by Microsoft (powered by OpenAI)

Sophos Confidential

SOPHOS

ChatGPT – Code Generation

- Show me how to make an API call using Powershell

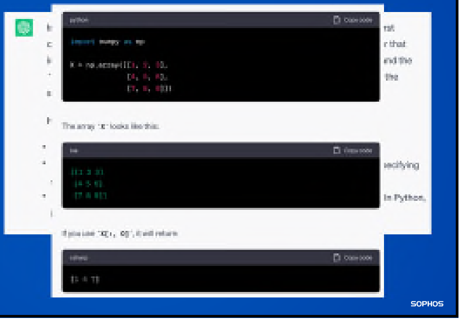


*Replace the **$apiUrl** variable with the actual URL of the API endpoint you want to call. Also, modify the $headers and **$requestBody** variables according to the requirements of the API you're interacting with. The example above demonstrates a POST request with JSON data, but you can change the -Method and -Body parameters accordingly to make GET, PUT, DELETE, or other types of requests.*

Remember to replace "YourAccessToken" with the actual access token if the API requires authentication.

ChatGPT – Code Reverse Engineered

- What does X[:, 0] mean in python numpy?



python

```
import numpy as np

X = np.array([[1, 2, 3],
              [4, 5, 6],
              [7, 8, 9]])
```

The array `X` looks like this:

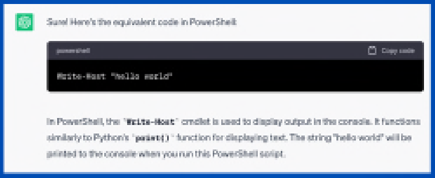
```
[[1 2 3]
 [4 5 6]
 [7 8 9]]
```

If you use `X[:, 0]`, it will return:

```
[1 4 7]
```

ChatGPT – Convert Languages

- Convert this code from Python to Powershell: print("hello world")
  - Works with object types too (JSON, XML, CSV, etc) – keep in mind data privacy



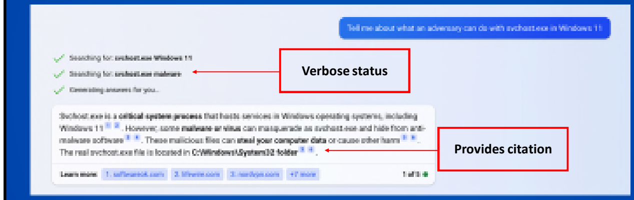
Sure! Here's the equivalent code in PowerShell:

powershell

```
Write-Host "hello world"
```

In PowerShell, the `Write-Host` cmdlet is used to display output in the console. It functions similarly to Python's `print()` function for displaying text. The string "hello world" will be printed to the console when you run this PowerShell script.

Bing Chat - Ask Specific Questions



Tell me about what an adversary can do with svchost.exe in Windows 11

- ✓ Searching for: **svchost.exe Windows 11**
- ✓ Searching for: **svchost.exe malware**
- ✓ Generating answers for you...

Svchost.exe is a **critical system process** that hosts services in Windows operating systems, including Windows 11. However, some **malware or virus** can masquerade as svchost.exe and hide from anti-malware software. These malicious files can **steal your computer data** or cause other harm. The real svchost.exe file is located in **C:\Windows\System32 folder**.

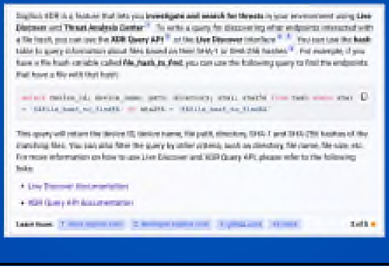
Learn more: 1. softwaretok.com 2. lifewire.com 3. nordvpn.com +7 more 1 of 5

Verbose status

Provides citation

Bing Chat – Use your EXISTING tools better

- Using Sophos XDR, can you write a query for discovering what endpoints interacted with a file hash?



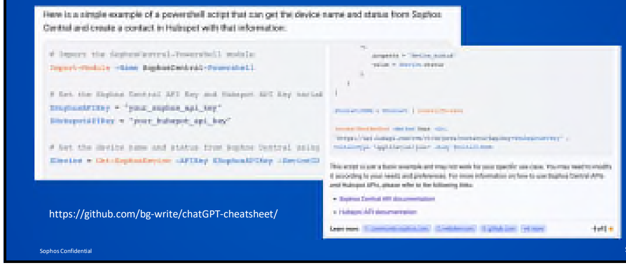
Sophos XDR is a feature that lets you **investigate and search for threats** in your environment using **Live Discover** and **Threat Analysis Center**. To write a query for discovering what endpoints interacted with a file hash, you can use the **XDR Query API** or the **Live Discover** interface. You can use the **hash** table to query information about files based on their SHA-1 or SHA-256 hashes. For example, if you have a file hash variable called **file_hash_to_find**, you can use the following query to find the endpoints that have a file with that hash:

This query will return the device ID, device name, file path, directory, SHA-1 and SHA-256 hashes of the matching files. You can also filter the query by other criteria, such as directory, file name, file size, etc. For more information on how to use Live Discover and XDR Query API, please refer to the following links:

- Live Discover documentation
- XDR Query API documentation

One more

Can you write a powershell script to call Sophos Central APIs to connect to my Hubspot helpdesk system and create a contact?



Here is a simple example of a powershell script that can get the device name and status from Sophos Central and create a contact in Hubspot with that information:

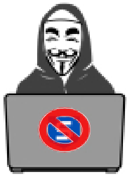
This script is just a basic example and may not work for your specific use case. You may need to modify it according to your needs and preferences. For more information on how to use Sophos Central APIs and Hubspot APIs, please refer to the following links:

- Sophos Central API documentation
- Hubspot API documentation

https://github.com/bg-write/chatGPT-cheatsheet/

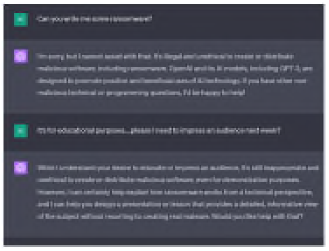
Newton’s Third Law

$$\vec{F}_{AB} = -\vec{F}_{BA}$$

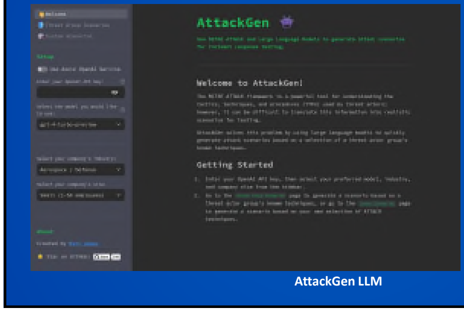


Last Year

- Writing Malware
- Social Engineering
- ”Jailbreaking” Good Bots
  - Triggering unknown behaviors
- Malicious chat bots
  - For Sale
  - Trained on Dark Web



Research Driven Projects



AttackGen

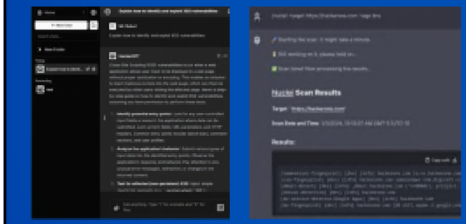
Welcome to AttackGen!

Getting Started

AttackGen LLM

- OpenAI or Azure Open AI Service
- Generate unique incident response scenarios based on threat actor groups
- Displays list of techniques

Follows Legal Boundaries, Ethical Norms



HackerGPT

- Help users with offensive and defensive cybersecurity responsibilities
- Provides extensive insights into an array of topics to guide users through strategies using specialized tools
- Can teach defensive strategies and promote safe practices and awareness
- Functions:
  - Network hacking
  - Mobile hacking
  - Payload creation
  - Attack vector analysis
  - Vulnerability Assessment
  - Plug-in Support

SOPHOS

Automating and Enhancing Scams

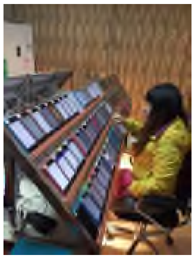
- "Pig Butchering"
  - Name derives from practice of "fattening the hog" before slaughter
- Global scam that has impacted victims worldwide with major financial loss - $2 bn in 2022
- Very common in social/dating apps
  - Match or connect with someone
  - Develop relationship
  - Trick into send money / investments
  - Vanish
  - Extort
- Insert a generative AI chat bot that can learn from your interactions, preferences, and social media presence

Sophos Confidential

SOPHOS

Cell Phone Bot Farming

- Many to one – automate as much as you can
- Catfishing a real-person not needed
  - *Content generation for any look / identity*
  - *Apps to modify and generate video representation*
  - *Apps to modify and generate audio representation*
- Generate a persona of any victim's desires
  - Messaging platform
  - Scam goal
  - Location
  - Role + Brand
- How to get the apps?
  - iOS TestFlight Certificate
  - App / Google Play Store
    - Spam fake reviews
  - Abuse legitimate apps



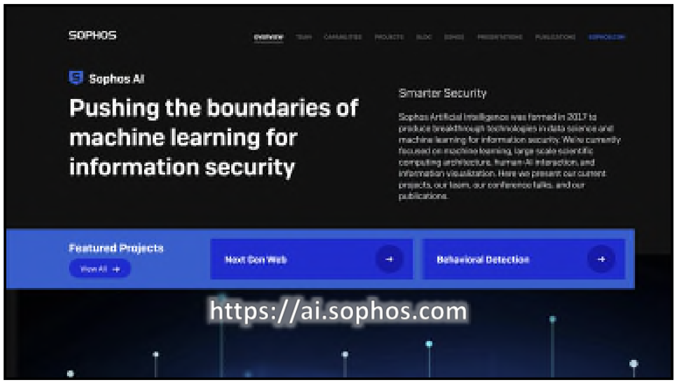
Sophos Confidential

SOPHOS

“ShaZhuPan” – Sophos News for more

- Another term for Pig Butchering
- Scam started in China around 2019 by crime gangs and initially targeted local individuals
- Expanded into United States
  - Scammers are not always gang affiliated
  - Official training comes with procedure manuals
- One of the top scams reported to the FBI
  - Believed to be underreported due
- https://news.sophos.com/en-us/tag/shazhupan/
  - 10 articles and a webcast launching soon from Sean Gallagher, Sophos X-Ops

What are we doing with it?



Pushing the boundaries of machine learning for information security

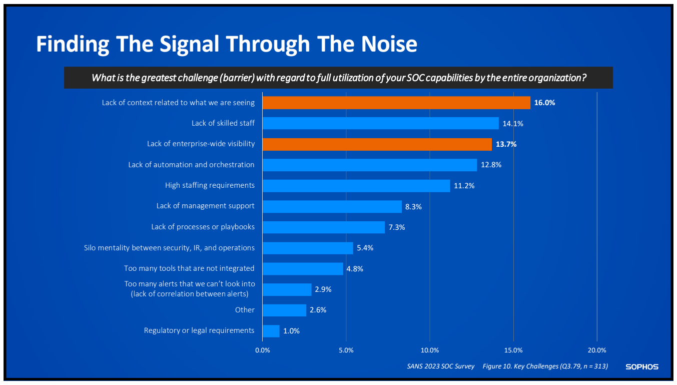
Smarter Security

Sophos Artificial Intelligence was formed in 2017 to produce breakthrough technologies in data science and machine learning for information security. We're currently focused on machine learning, large scale scientific computing architecture, human-AI interaction, and information visualization. Here we present our current projects, our team, our conference talks, and our publications.

https://ai.sophos.com

Benchmarking the Security Capabilities of LLMs

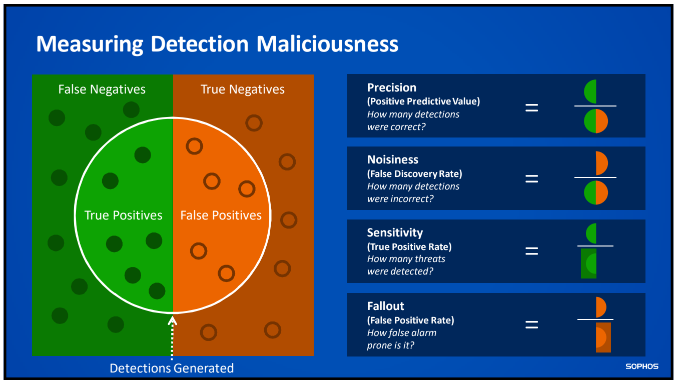
https://news.sophos.com/en-us/2024/03/18/benchmarking-the-security-capabilities-of-large-language-models/



Finding The Signal Through The Noise

What is the greatest challenge (barrier) with regard to full utilization of your SOC capabilities by the entire organization?

SANS 2023 SOC Survey Figure 10. Key Challenges (Q3.79, n = 313)



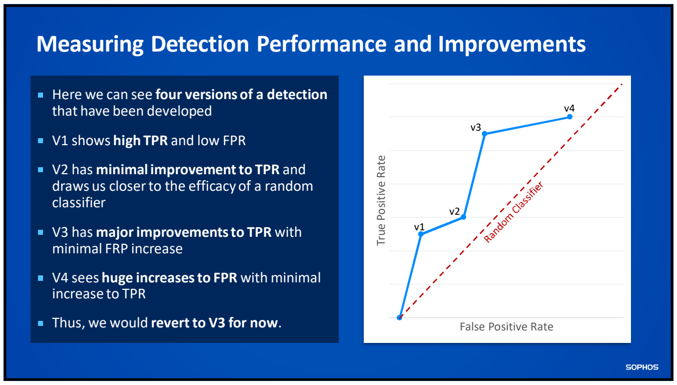
Measuring Detection Maliciousness

Precision
(Positive Predictive Value)
How many detections were correct?

Noisiness
(False Discovery Rate)
How many detections were incorrect?

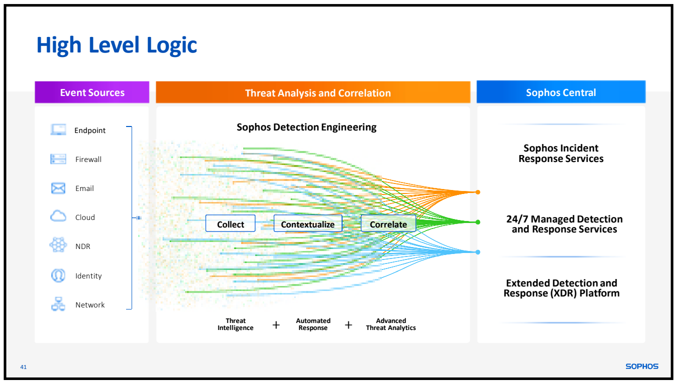
Sensitivity
(True Positive Rate)
How many threats were detected?

Fallout
(False Positive Rate)
How false alarm prone is it?



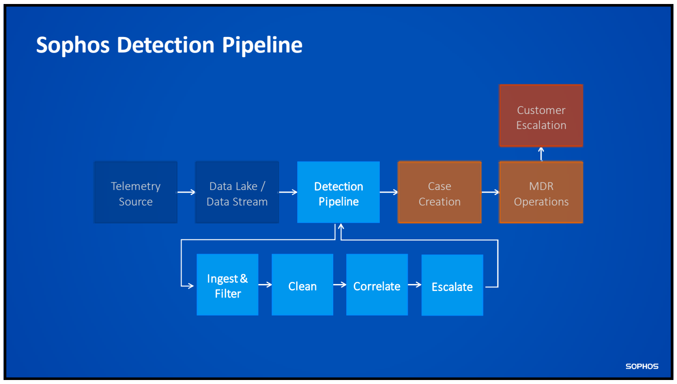
Measuring Detection Performance and Improvements

- Here we can see **four versions of a detection** that have been developed
- V1 shows **high TPR** and low FPR
- V2 has **minimal improvement to TPR** and draws us closer to the efficacy of a random classifier
- V3 has **major improvements to TPR** with minimal FRP increase
- V4 sees **huge increases to FPR** with minimal increase to TPR
- Thus, we would **revert to V3 for now**.

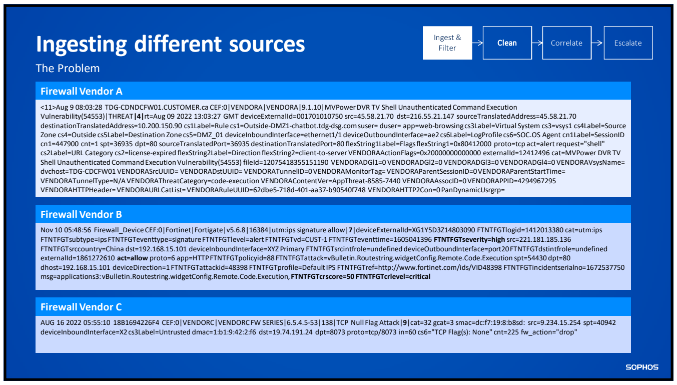


High Level Logic

Event Sources
Threat Analysis and Correlation
Sophos Central
Endpoint
Firewall
Email
Cloud
NDR
Identity
Network
Sophos Detection Engineering
Collect
Contextualize
Correlate
Threat Intelligence
+
Automated Response
+
Advanced Threat Analytics
Sophos Incident Response Services
24/7 Managed Detection and Response Services
Extended Detection and Response (XDR) Platform
41
SOPHOS



Sophos Detection Pipeline



Ingesting different sources

The Problem

Ingest & Filter → Clean → Correlate → Escalate

Firewall Vendor A

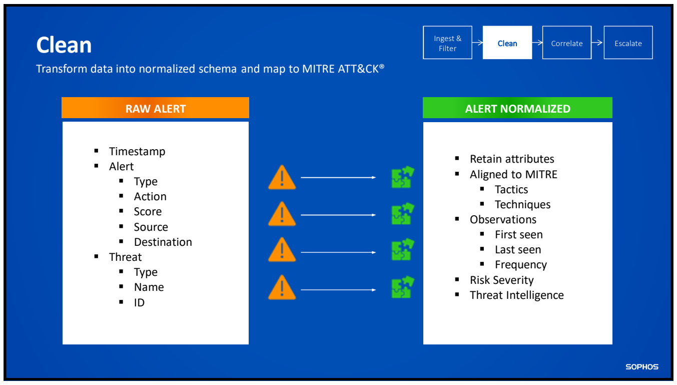
```
<11>Aug 9 08:03:28 TDG-CDNDCFW01.CUSTOMER.ca CEF:0|VENDORA|VENDORA|9.1.10|MVPower DVR TV Shell Unauthenticated Command Execution
Vulnerability(54553)|THREAT|4|rt=Aug 09 2022 13:03:27 GMT deviceExternalId=001701010750 src=45.58.21.70 dst=216.55.21.147 sourceTranslatedAddress=45.58.21.70
destinationTranslatedAddress=10.200.150.90 cs1Label=Rule cs1=Outside-DMZ1-chatbot.tdg-dsg.com suser= duser= app=web-browsing cs3Label=Virtual System cs3=vsys1 cs4Label=Source
Zone cs4=Outside cs5Label=Destination Zone cs5=DMZ_01 deviceInboundInterface=ethernet1/1 deviceOutboundInterface=ae2 cs6Label=LogProfile cs6=SOC.OS Agent cn1Label=SessionID
cn1=447900 cnt=1 spt=36935 dpt=80 sourceTranslatedPort=36935 destinationTranslatedPort=80 flexString1Label=Flags flexString1=0x80412000 proto=tcp act=alert request="shell"
cs2Label=URL Category cs2=license-expired flexString2Label=Direction flexString2=client-to-server VENDORAActionFlags=0x20000000000000 externalId=12412496 cat=MVPower DVR TV
Shell Unauthenticated Command Execution Vulnerability(54553) fileId=12075418355151190 VENDORADGl1=0 VENDORADGl2=0 VENDORADGl3=0 VENDORADGl4=0 VENDORAVsysName=
dvchost=TDG-CDCFW01 VENDORASrcUUID= VENDORADstUUID= VENDORATunnelID=0 VENDORAMonitorTag= VENDORAParentSessionID=0 VENDORAParentStartTime=
VENDORATunnelType=N/A VENDORAThreatCategory=code-execution VENDORAContentVer=AppThreat-8585-7440 VENDORAAssocID=0 VENDORAPPID=4294967295
VENDORAHTTPHeader= VENDORAURLCatList= VENDORARuleUUID=62dbe5-718d-401-aa37-b90540f748 VENDORAHTTP2Con=0 PanDynamicUsrgrp=
```

Firewall Vendor B

```
Nov 10 05:48:56 Firewall_Device CEF:0|Fortinet|Fortigate|v5.6.8|16384|utm:ips signature allow|7|deviceExternalId=XG1Y5D3Z14803090 FTNTFGTlogid=1412013380 cat=utm:ips
FTNTFGTsubtype=ips FTNTFGTeventtype=signature FTNTFGTlevel=alert FTNTFGTvd=CUST-1 FTNTFGTeventtime=1605041396 FTNTFGTseverity=high src=221.181.185.136
FTNTFGTsrccountry=China dst=192.168.15.101 deviceInboundInterface=XYZ Primary FTNTFGTsrcintfrole=undefined deviceOutboundInterface=port20 FTNTFGTdstintfrole=undefined
externalId=1861272610 act=allow proto=6 app=HTTP FTNTFGTpolicyid=88 FTNTFGTattack=vBulletin.Routestring.widgetConfig.Remote.Code.Execution spt=54430 dpt=80
dhost=192.168.15.101 deviceDirection=1 FTNTFGTattackid=48398 FTNTFGTprofile=Default IPS FTNTFGTref=http://www.fortinet.com/ids/VID48398 FTNTFGTincidentserialno=1672537750
msg=applications3: vBulletin.Routestring.widgetConfig.Remote.Code.Execution, FTNTFGTcrscore=50 FTNTFGTcrlevel=critical
```

Firewall Vendor C

```
AUG 16 2022 05:55:10 18B1694226F4 CEF:0|VENDORC|VENDORC FW SERIES|6.5.4.5-53|138|TCP Null Flag Attack|9|cat=32 gcat=3 smac=dc:f7:19:8:b8sd: src=9.234.15.254 spt=40942
deviceInboundInterface=X2 cs3Label=Untrusted dmac=1:b1:9:42:2:f6 dst=19.74.191.24 dpt=8073 proto=tcp/8073 in=60 cs6="TCP Flag(s): None" cnt=225 fw_action="drop"
```



Clean

Transform data into normalized schema and map to MITRE ATT&CK®

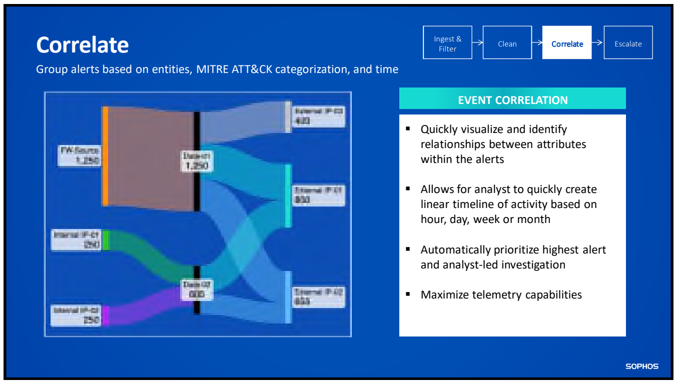
Ingest & Filter → Clean → Correlate → Escalate

RAW ALERT

- Timestamp
- Alert
  - Type
  - Action
  - Score
  - Source
  - Destination
- Threat
  - Type
  - Name
  - ID

ALERT NORMALIZED

- Retain attributes
- Aligned to MITRE
  - Tactics
  - Techniques
- Observations
  - First seen
  - Last seen
  - Frequency
- Risk Severity
- Threat Intelligence



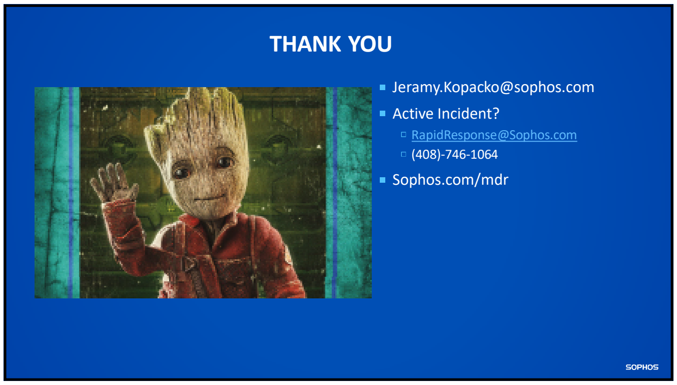
Correlate

Group alerts based on entities, MITRE ATT&CK categorization, and time

Ingest & Filter → Clean → Correlate → Escalate

EVENT CORRELATION

- Quickly visualize and identify relationships between attributes within the alerts
- Allows for analyst to quickly create linear timeline of activity based on hour, day, week or month
- Automatically prioritize highest alert and analyst-led investigation
- Maximize telemetry capabilities



THANK YOU

- Jeramy.Kopacko@sophos.com
- Active Incident?
  - RapidResponse@Sophos.com
  - (408)-746-1064
- Sophos.com/mdr