Reinventing the Complete Cybersecurity Stack With AI Language Models

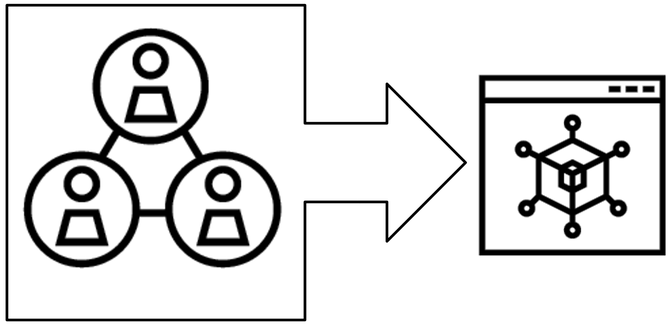
Bartley Richardson, PhD
Director, Cybersecurity Engineering | LLM Developer Day 2023

Agenda

- Cybersecurity As Language Problems
- Security-Specific Copilots
- So What? How We Can Deeply Integrate Language Models
- Three Layers of the Stack Where Language Models Work
- What to Take Away and Next Steps

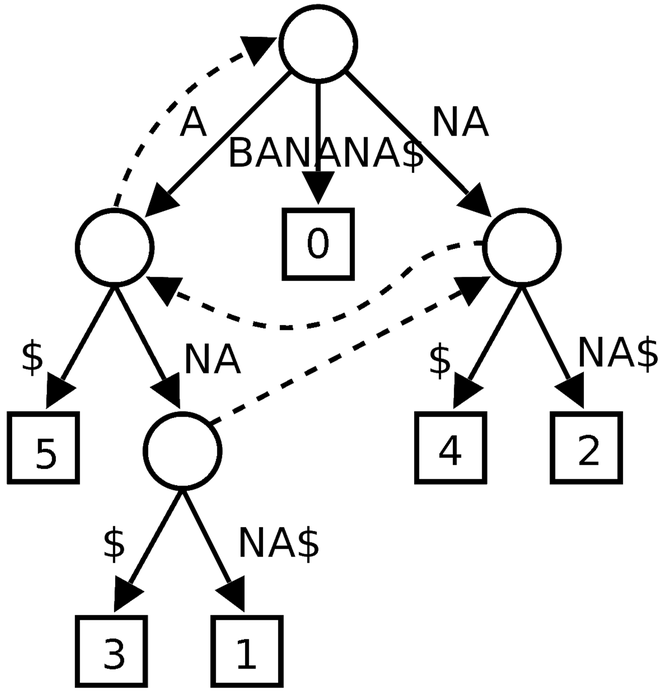
Representing Cyber Problems As Language Problems

Not a new idea



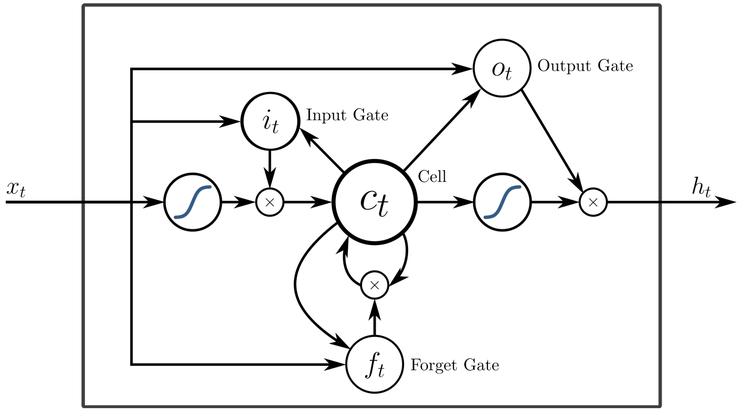
Agent Based Modeling

Only could represent simplistic behaviors



Probabilistic Suffix Tree

Suffered from vocabulary explosion for available memory



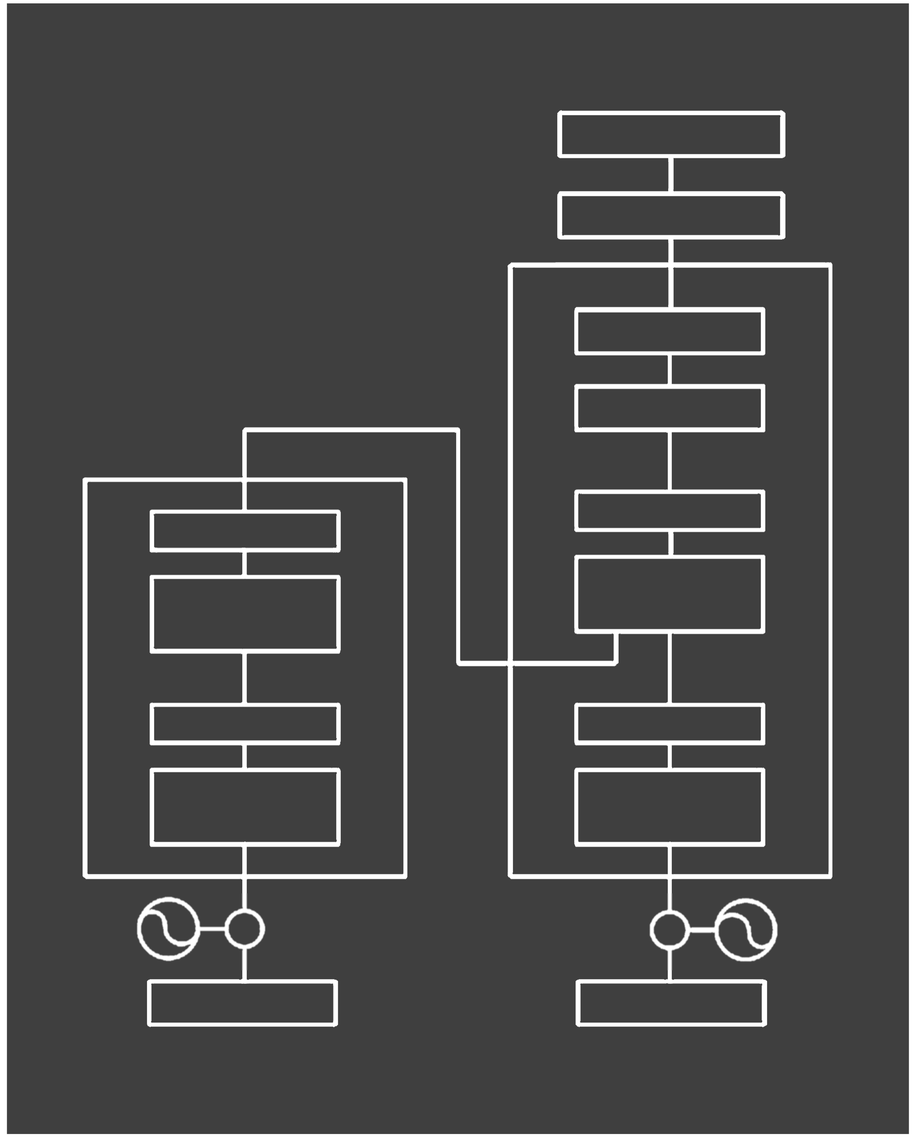
LSTM

Lacked enough context for complex associations

Language Models

Powered by the transformer architecture

- Transformer architecture allows sufficient context window (with attention) to learn simple tasks and small corpus
- High utility and success when:
  - Target application/domain is small and limited in scope
  - Small context window is adequate to give salient results
  - Relative number of tokens is small or can be constrained
- Example applications for cybersecurity include:
  - Log parsing (cyBERT)
  - Root cause analysis
  - Basic anomaly detection
- Architecture fails when even more additional context/tokens needed or when domain becomes large



Copilots

Copilots for Cybersecurity

Address three critical issues and are the easy button of LLM integration

Address the Staffing Shortage

Staffing is already low, and copilots allow analysts to be more efficient and effective

Democratize Expert Service

Customers can obtain near real-time answers and guidance to complex deployment scenarios

Natural Interface for Customers/Users

No need to struggle with complex query languages and esoteric languages – natural language is the ultimate interface

Key Challenges for Security Copilots

Enable ease-of-use and fast pivoting while ensuring information is correct and verifiable

Maintain Topical Responses

Answer questions about specific products/services without devolving into a “know everything” chatbot

Mitigate Hallucinations

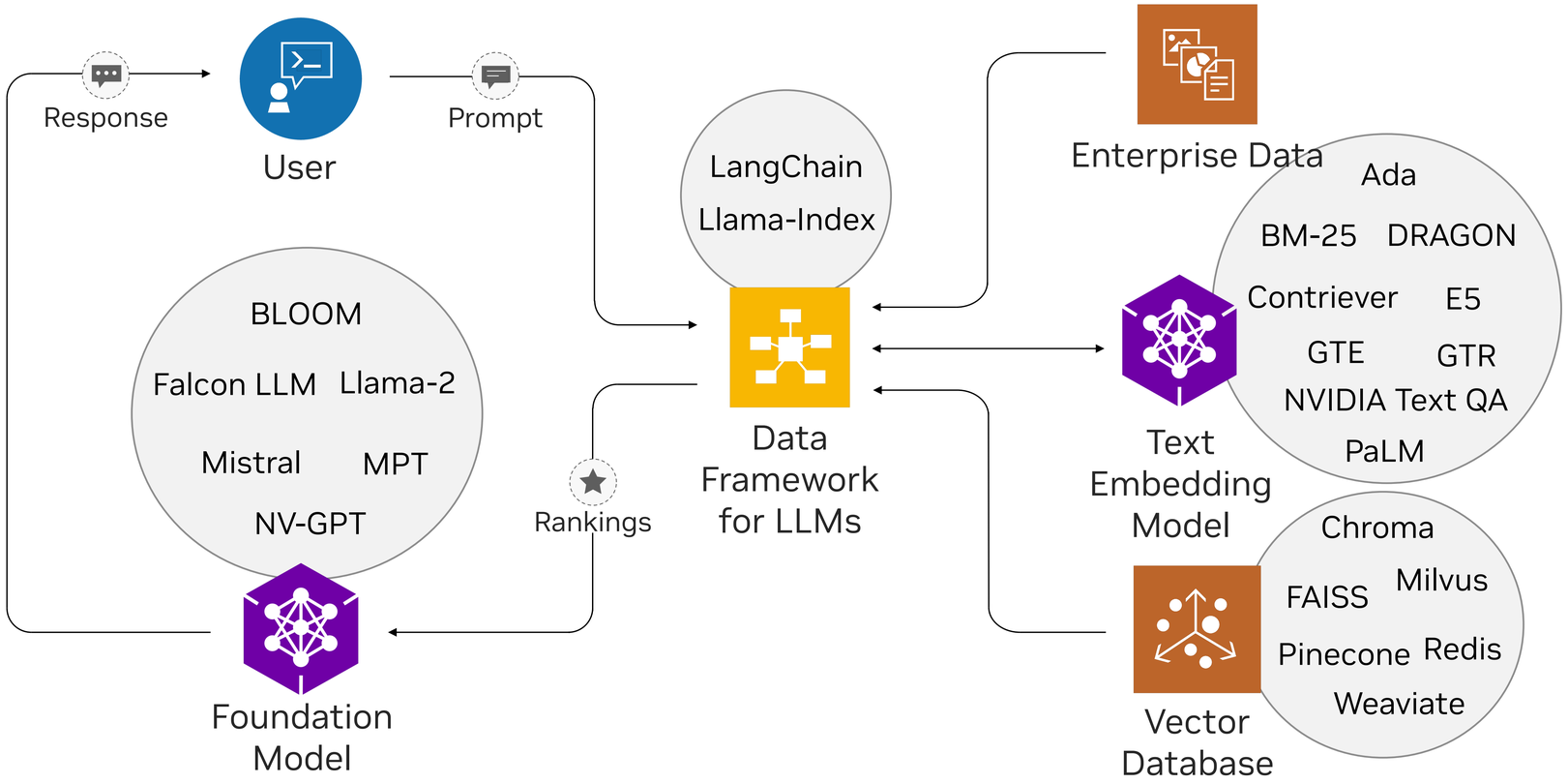
Ensure information is correct and be clear when an answer is unknown or unknowable

Maintain Up-to-Date Information

Utilize recent information in context with a LLM without retraining or fine-tuning frequently

Generative AI Knowledge Base Chatbot

The easy button for KB-based chatbots with Retrieval Augmented Generation (RAG)



Accurate Responses

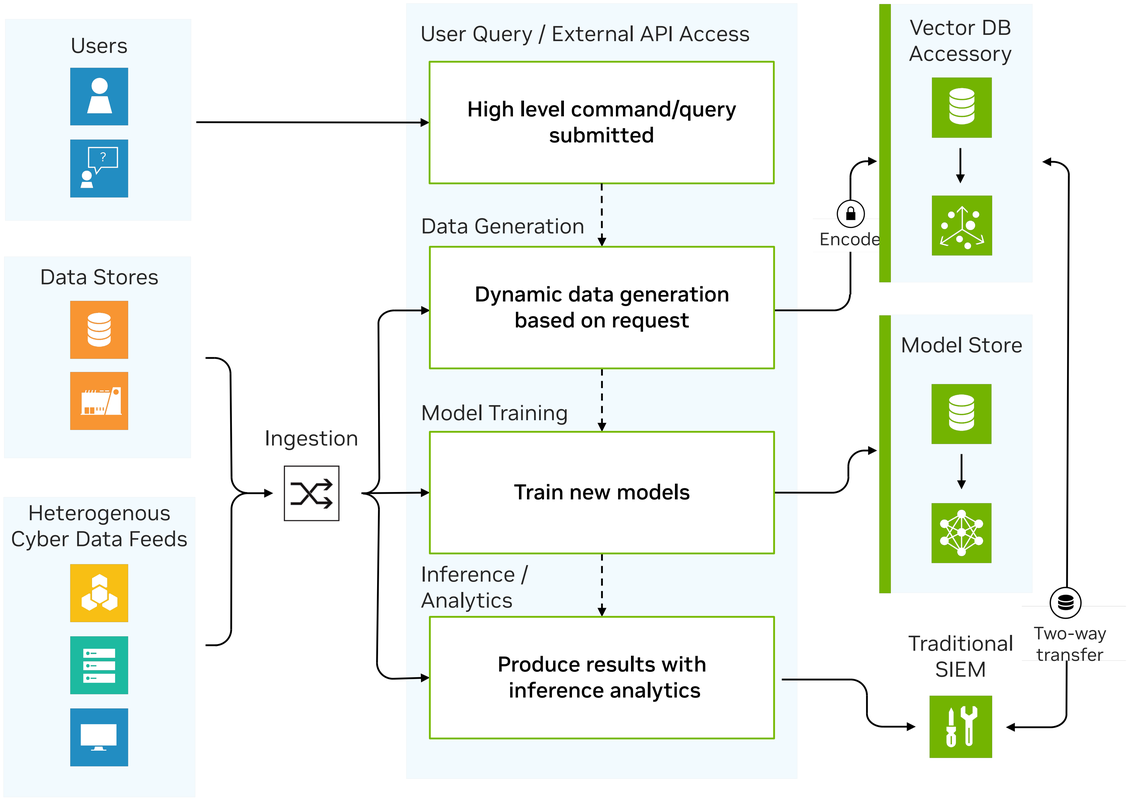
Generate Answers at Scale

Deploy Anywhere

Integration of Language Models Deeper in the Cyber Stack

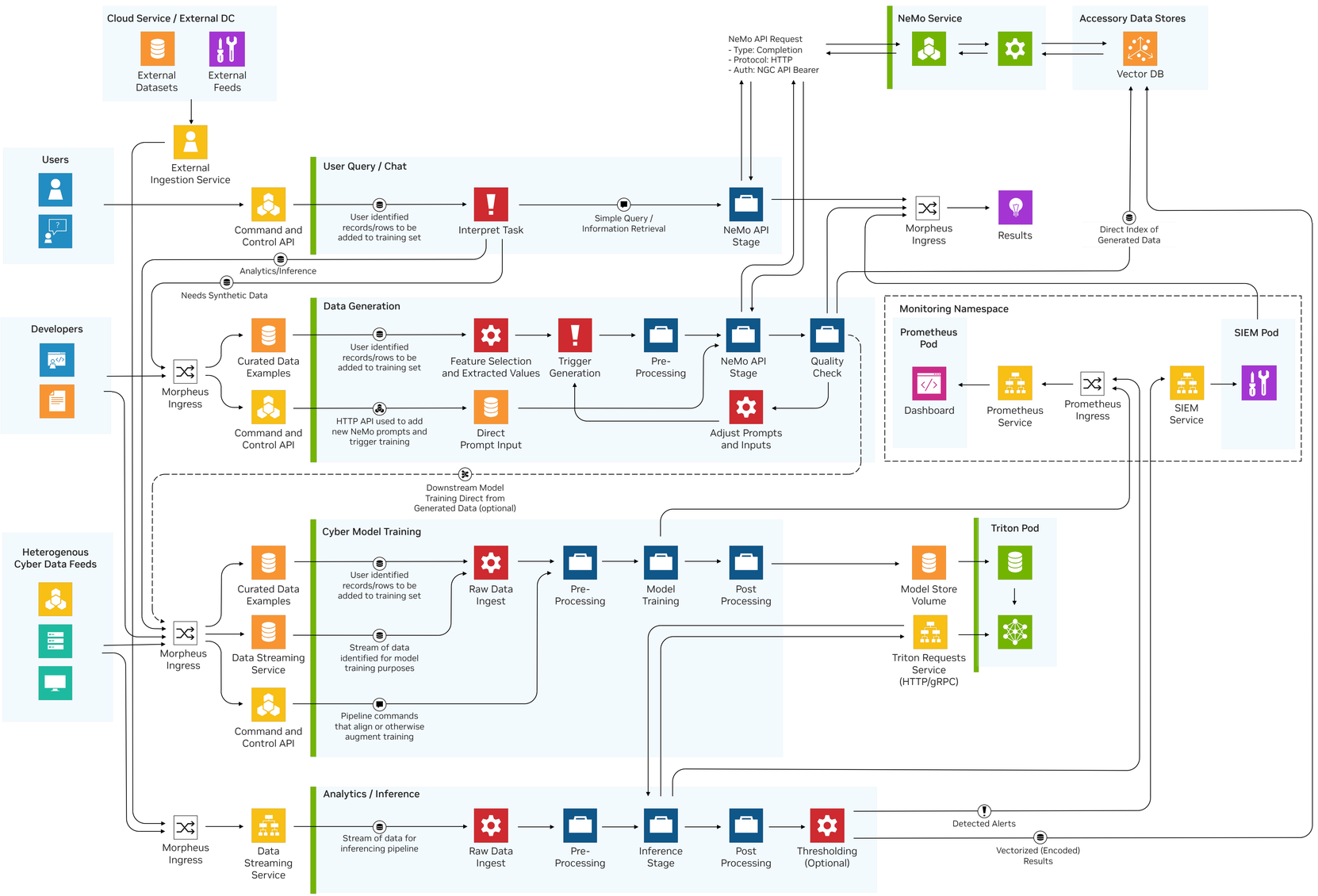
Copilots are just the beginning

- Generative models provide effective ways to address data gaps, especially for rare events
- Large language models can be used to generate massive amounts of high-quality, on-topic synthetic data used to train lightweight detection models
- While powerful, LLMs greatly increase inference time if placed inline with cybersecurity pipelines
- Fundamental work is being done to optimize generative AI and LLM applications specifically for high-velocity cybersecurity data feeds
- This enabling technology will only be fully realized in cyber when integrated throughout existing pipelines and frameworks
- Let's talk about three concrete examples



Direct Integration with Real-Time Security Feeds

Generic Cyber+LLM Stack Architecture



CVE Exploitability and Vulnerability Analysis

Dramatically improve vulnerability defense with generative AI

High Accuracy

When the LLM agent is confident in its conclusion regarding exploitability, it matches human analysts with **85%** accuracy

Faster Time to Assess

Allows analysts to investigate individual CVEs an average of **4 times** faster (vs a human)

Reduce CVE Noise

Prioritized patches to **protect your specific environment** with LLM thought process, observations, and results. Repeat until conclusion is reached.

“

This is awesome! It really speeds up the triage of CVEs from specific container scans.

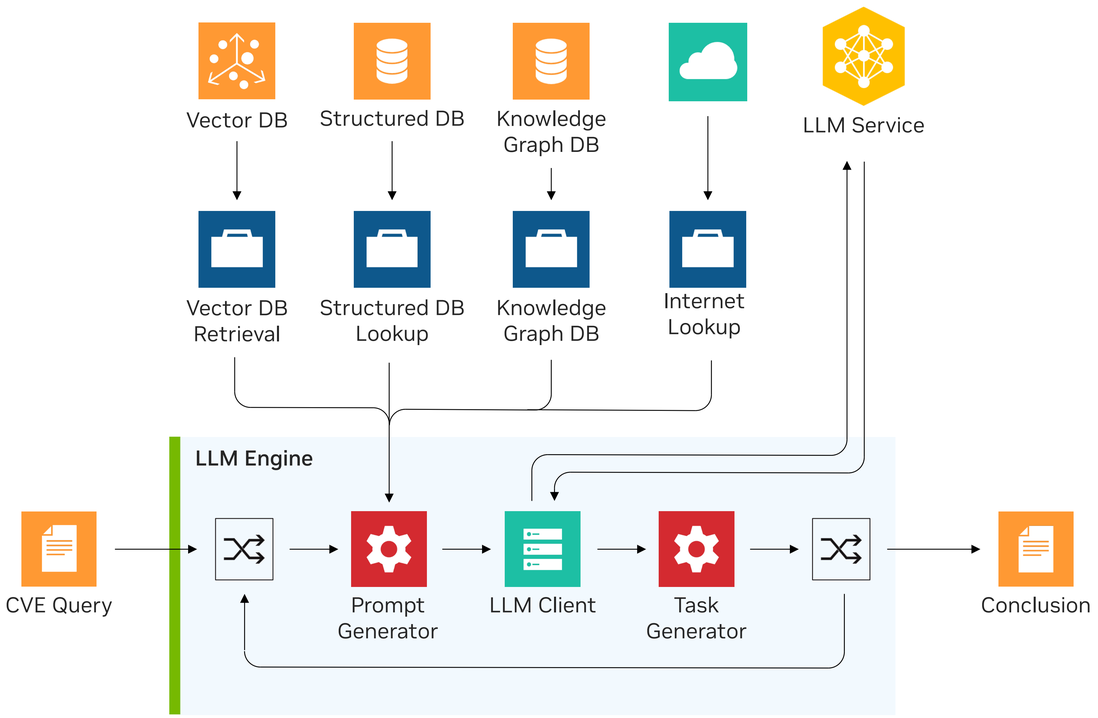
— NVIDIA container security analyst

”

CVE Exploitability Using the LLM Engine

LLM Engine supports model generated RAG tasks and multiple loops

CVE Query: Is CVE-2022-42890 exploitable in morpheus:v23.07.00a?

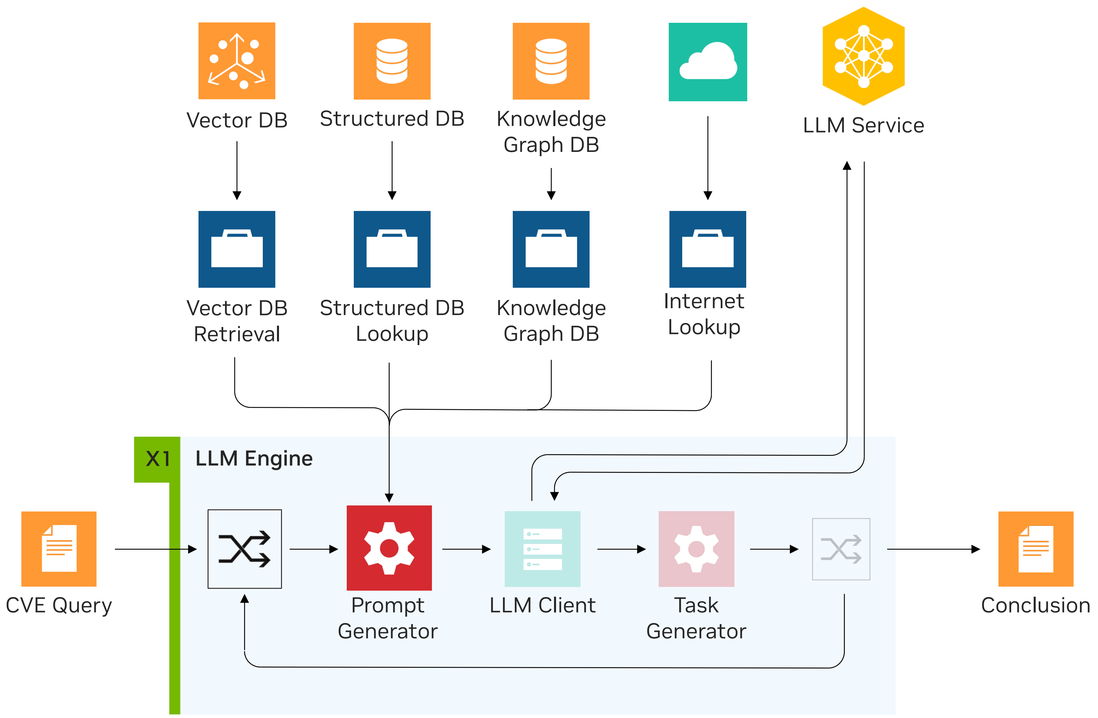


CVE Exploitability Using the LLM Engine

LLM Engine supports model generated RAG tasks and multiple loops

CVE Query: Is CVE-2022-42890 exploitable in morpheus:v23.07.00a?

Prompt Generator: << Adds up-to-date CVE info and checklist items >>
LLM Client: Checklist item 1: Is the component in product? Verify container running Apache Batik + determine version. Vulnerability affects v1.0 to 1.9.1
Action: Software Bill of Materials QA System
Task Generator: Search for Apache Batik in the SBOM QA System

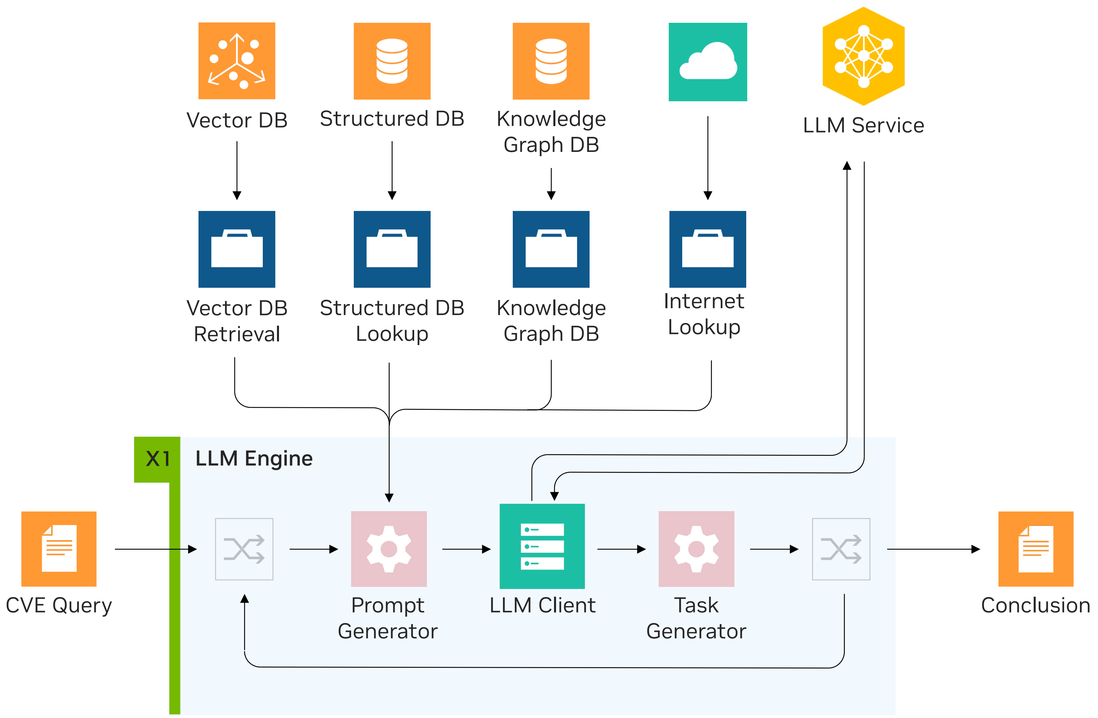


CVE Exploitability Using the LLM Engine

LLM Engine supports model generated RAG tasks and multiple loops

CVE Query: Is CVE-2022-42890 exploitable in morpheus:v23.07.00a?

Prompt Generator: << Adds up-to-date CVE info and checklist items >>
LLM Client: Checklist item 1: Is the component in product? Verify container running Apache Batik + determine version. Vulnerability affects v1.0 to 1.9.1
Action: Software Bill of Materials QA System
Task Generator: Search for Apache Batik in the SBOM QA System

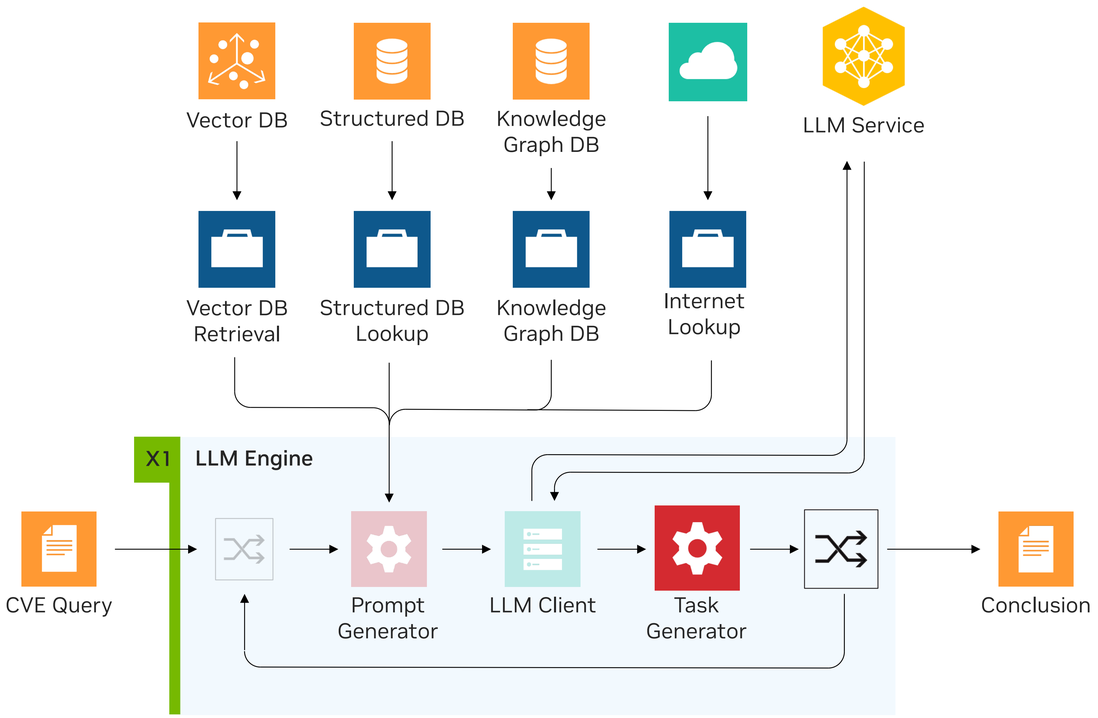


CVE Exploitability Using the LLM Engine

LLM Engine supports model generated RAG tasks and multiple loops

CVE Query: Is CVE-2022-42890 exploitable in morpheus:v23.07.00a?

Prompt Generator: << Adds up-to-date CVE info and checklist items >>
LLM Client: Checklist item 1: Is the component in product? Verify container running Apache Batik + determine version. Vulnerability affects v1.0 to 1.9.1
Action: Software Bill of Materials QA System
Task Generator: Search for Apache Batik in the SBOM QA System

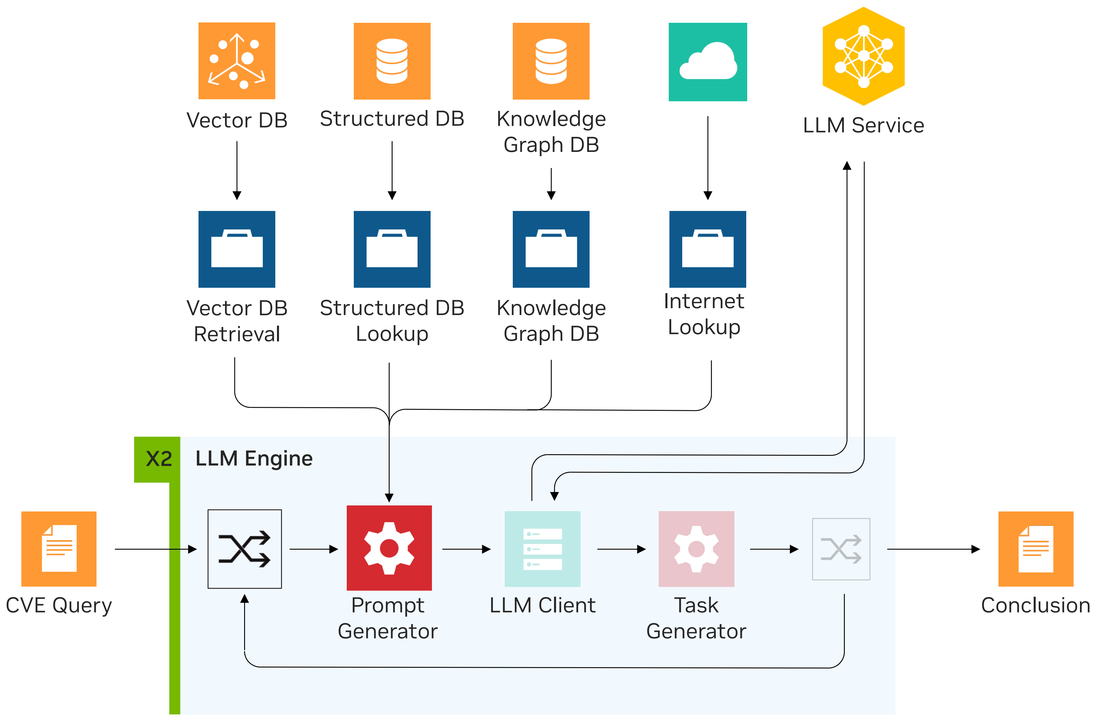


CVE Exploitability Using the LLM Engine

LLM Engine supports model generated RAG tasks and multiple loops

CVE Query: Is CVE-2022-42890 exploitable in morpheus:v23.07.00a?

Prompt Generator: <<Returns relevant SBOM information>>
LLM Client: Observation: Apache Batik 1.0 is listed in the container's software BOM + the container is vulnerable.
Checklist item 2: Can the vulnerable component be executed? Find the requirements to run Apache Batik. Verify the container is running these requirements.
Action: Internet Lookup
Task Generator: Search for Apache Batik requirements using the internet lookup

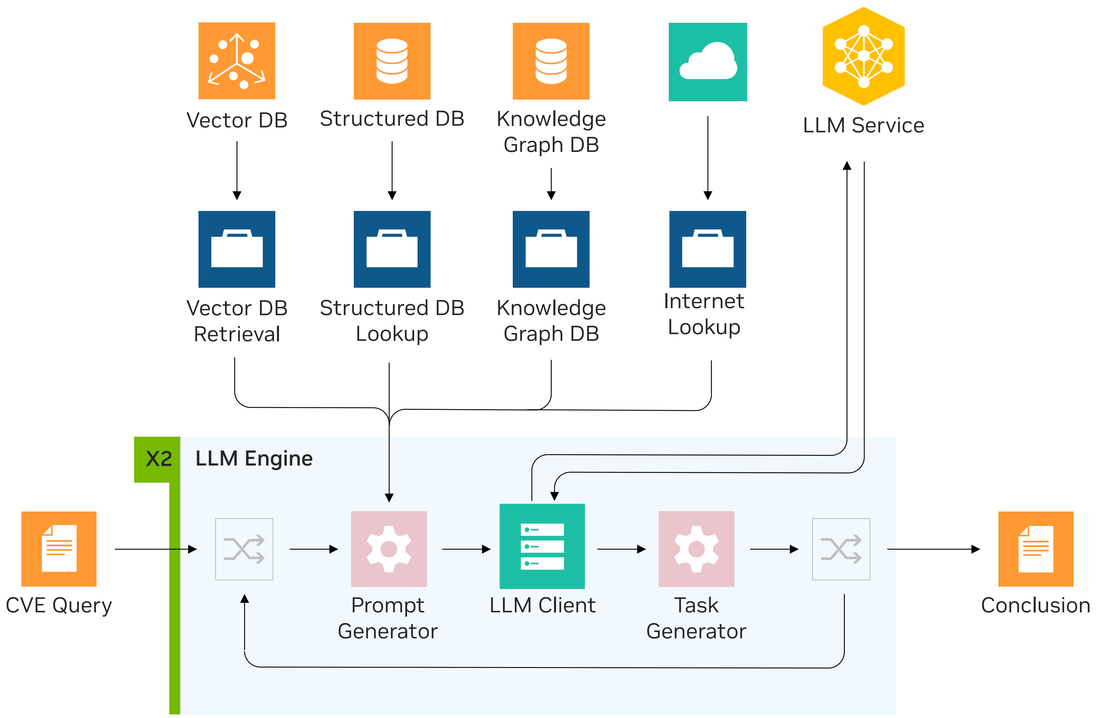


CVE Exploitability Using the LLM Engine

LLM Engine supports model generated RAG tasks and multiple loops

CVE Query: Is CVE-2022-42890 exploitable in morpheus:v23.07.00a?

Prompt Generator: <<Returns relevant SBOM information>>
LLM Client: Observation: Apache Batik 1.0 is listed in the container's software BOM + the container is vulnerable.
Checklist item 2: Can the vulnerable component be executed? Find the requirements to run Apache Batik. Verify the container is running these requirements.
Action: Internet Lookup
Task Generator: Search for Apache Batik requirements using the internet lookup

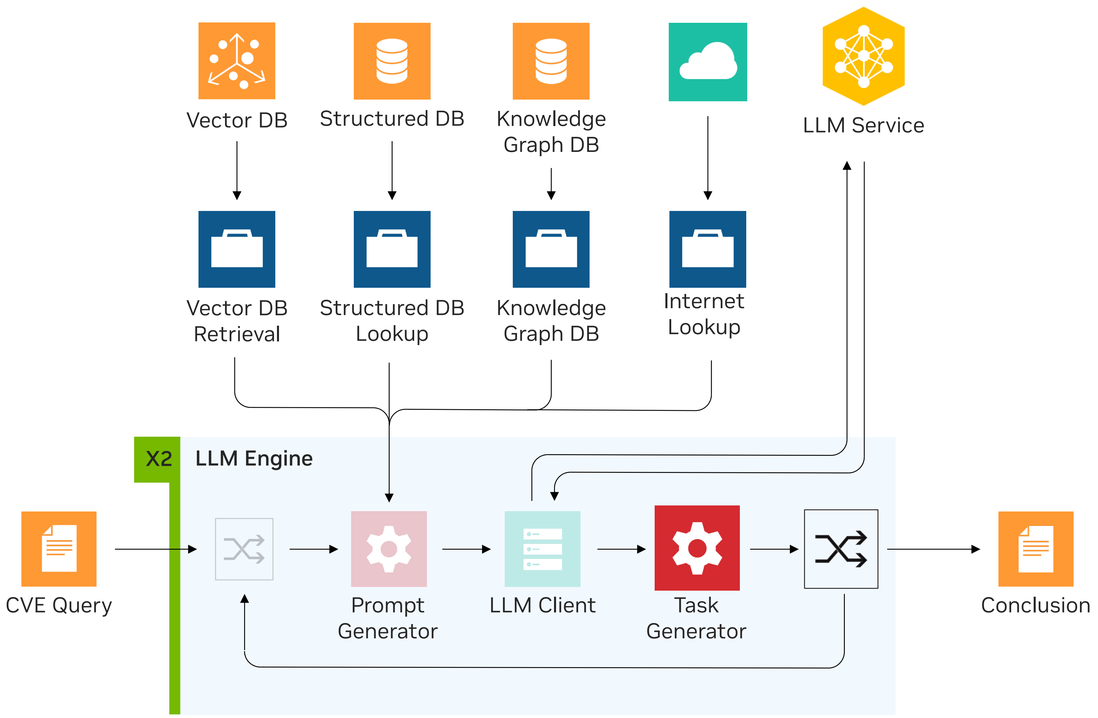


CVE Exploitability Using the LLM Engine

LLM Engine supports model generated RAG tasks and multiple loops

CVE Query: Is CVE-2022-42890 exploitable in morpheus:v23.07.00a?

Prompt Generator: <<Returns relevant SBOM information>>
LLM Client: Observation: Apache Batik 1.0 is listed in the container's software BOM + the container is vulnerable.
Checklist item 2: Can the vulnerable component be executed? Find the requirements to run Apache Batik. Verify the container is running these requirements.
Action: Internet Lookup
Task Generator: Search for Apache Batik requirements using the internet lookup

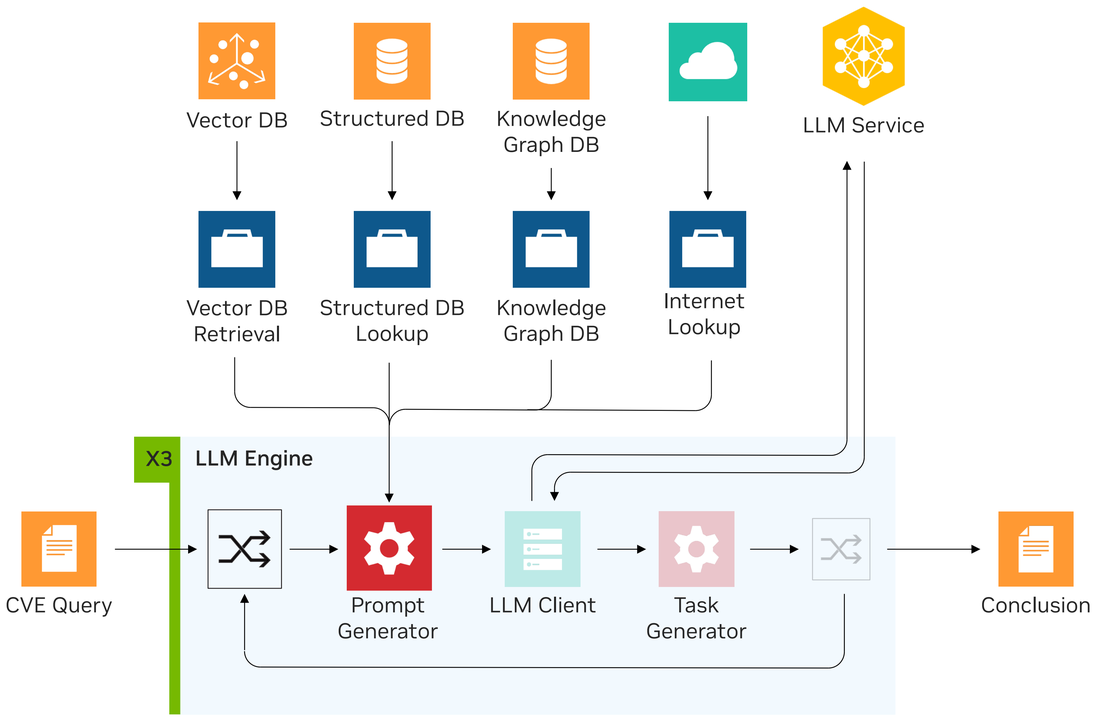


CVE Exploitability Using the LLM Engine

LLM Engine supports model generated RAG tasks and multiple loops

CVE Query: Is CVE-2022-42890 exploitable in morpheus:v23.07.00a?

Prompt Generator: <<Returns relevant environment requirements for vulnerable library>>
LLM Client: Observation: Apache Batik requires Java Runtime Environment 1.4+. Need to determine if container contains the requirement.
Checklist item 2: Can the vulnerable component be executed? Find the requirements to run Apache Batik. Verify the container is running these requirements.
Action: Software Bill of Materials QA System
Task Generator: Search Java Runtime Environment using SBOM QA System

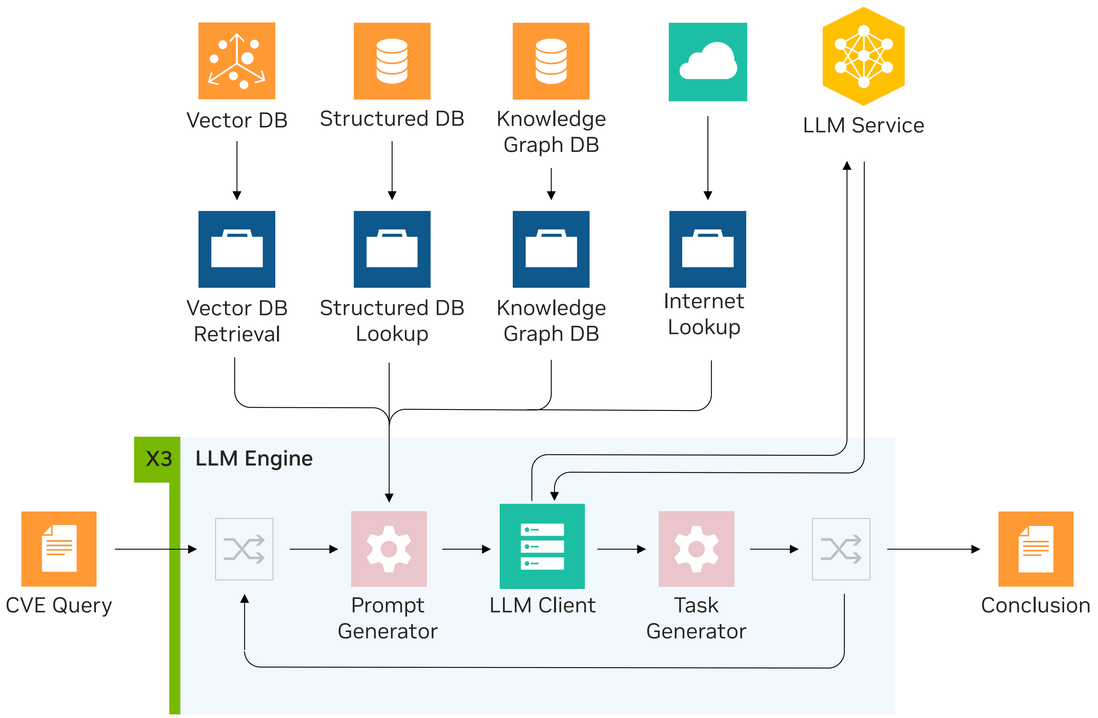


CVE Exploitability Using the LLM Engine

LLM Engine supports model generated RAG tasks and multiple loops

CVE Query: Is CVE-2022-42890 exploitable in morpheus:v23.07.00a?

Prompt Generator: <<Returns relevant environment requirements for vulnerable library>>
LLM Client: Observation: Apache Batik requires Java Runtime Environment 1.4+. Need to determine if container contains the requirement.
Checklist item 2: Can the vulnerable component be executed? Find the requirements to run Apache Batik. Verify the container is running these requirements.
Action: Software Bill of Materials QA System
Task Generator: Search Java Runtime Environment using SBOM QA System

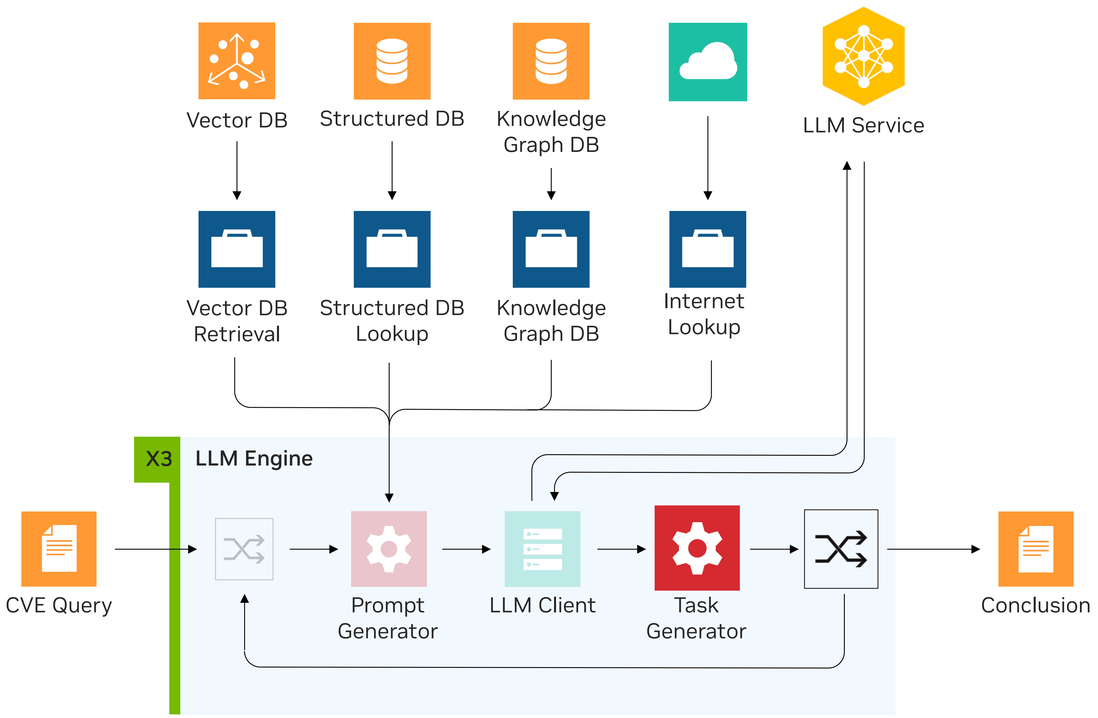


CVE Exploitability Using the LLM Engine

LLM Engine supports model generated RAG tasks and multiple loops

CVE Query: Is CVE-2022-42890 exploitable in morpheus:v23.07.00a?

Prompt Generator: <<Returns relevant environment requirements for vulnerable library>>
LLM Client: Observation: Apache Batik requires Java Runtime Environment 1.4+. Need to determine if container contains the requirement.
Checklist item 2: Can the vulnerable component be executed? Find the requirements to run Apache Batik. Verify the container is running these requirements.
Action: Software Bill of Materials QA System
Task Generator: Search Java Runtime Environment using SBOM QA System

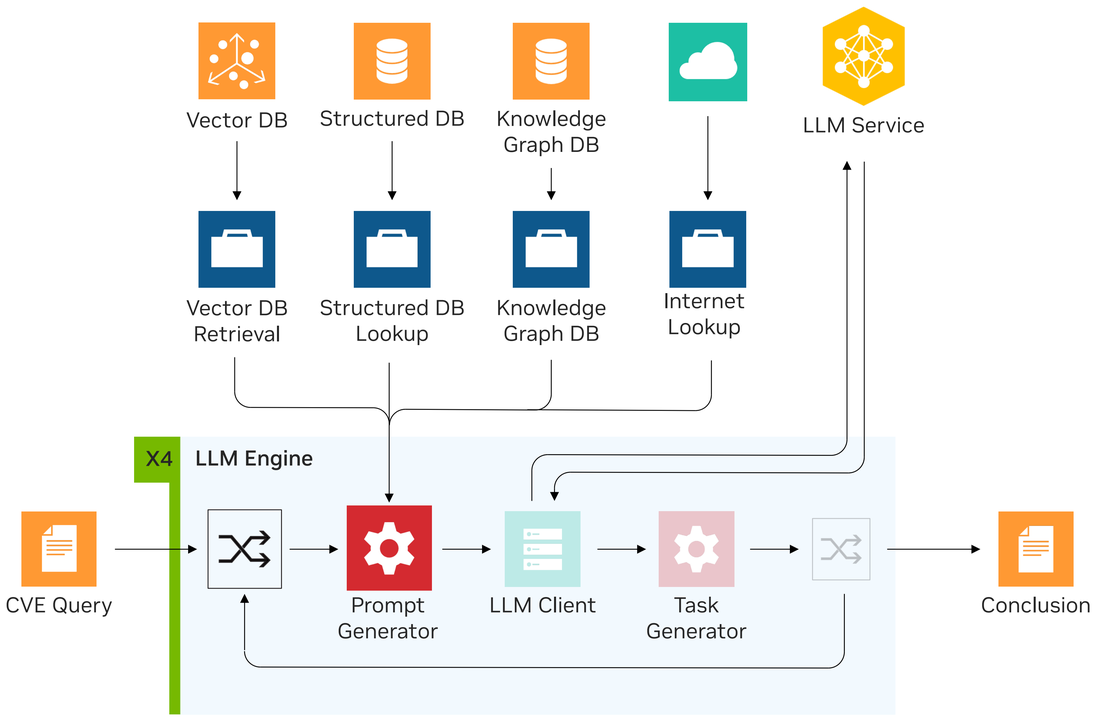


CVE Exploitability Using the LLM Engine

LLM Engine supports model generated RAG tasks and multiple loops

CVE Query: Is CVE-2022-42890 exploitable in morpheus:v23.07.00a?

Prompt Generator: <<Returns relevant SBOM information>>
LLM Client: Observation: No, there is no Java Runtime Environment in the container's software bill of materials.
Final Thought: The vulnerable component cannot be executed in the container without installing Java Runtime Environment.

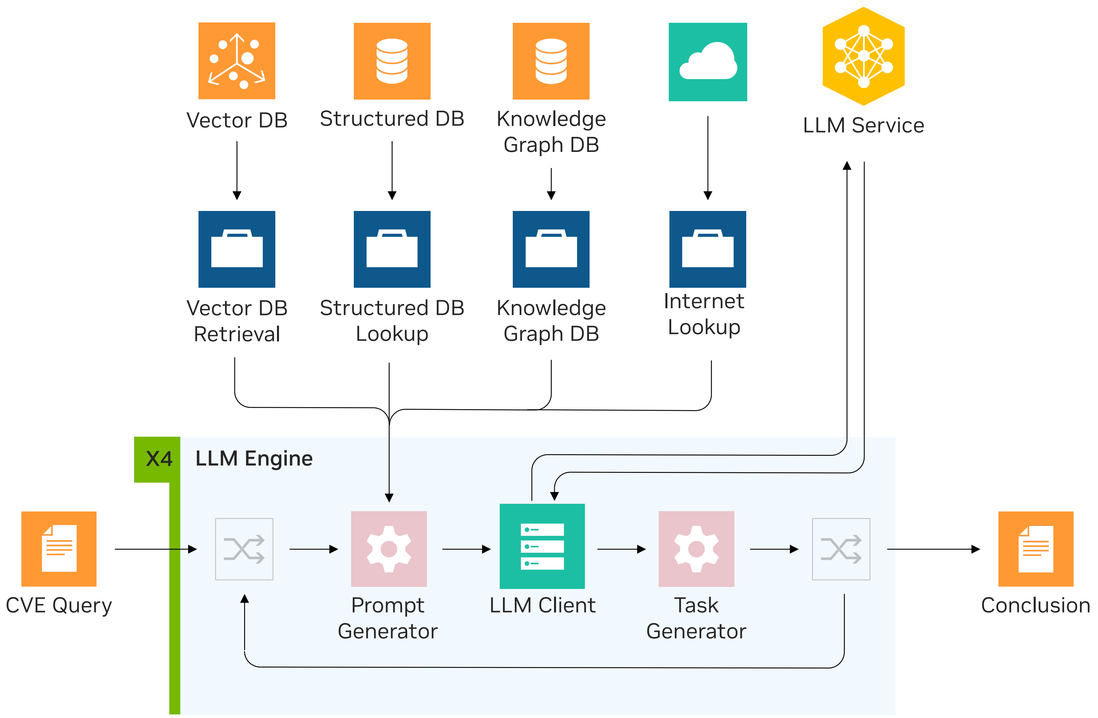


CVE Exploitability Using the LLM Engine

LLM Engine supports model generated RAG tasks and multiple loops

CVE Query: Is CVE-2022-42890 exploitable in morpheus:v23.07.00a?

Prompt Generator: <<Returns relevant SBOM information>>
LLM Client: Observation: No, there is no Java Runtime Environment in the container's software bill of materials.
Final Thought: The vulnerable component cannot be executed in the container without installing Java Runtime Environment.

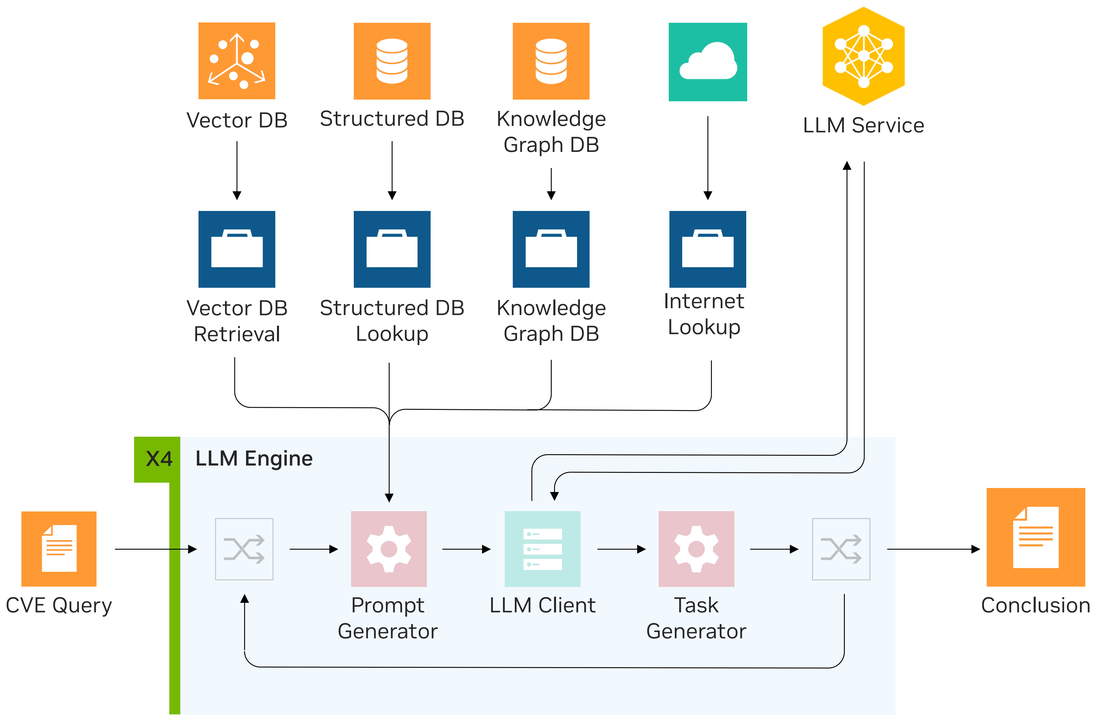


CVE Exploitability Using the LLM Engine

LLM Engine supports model generated RAG tasks and multiple loops

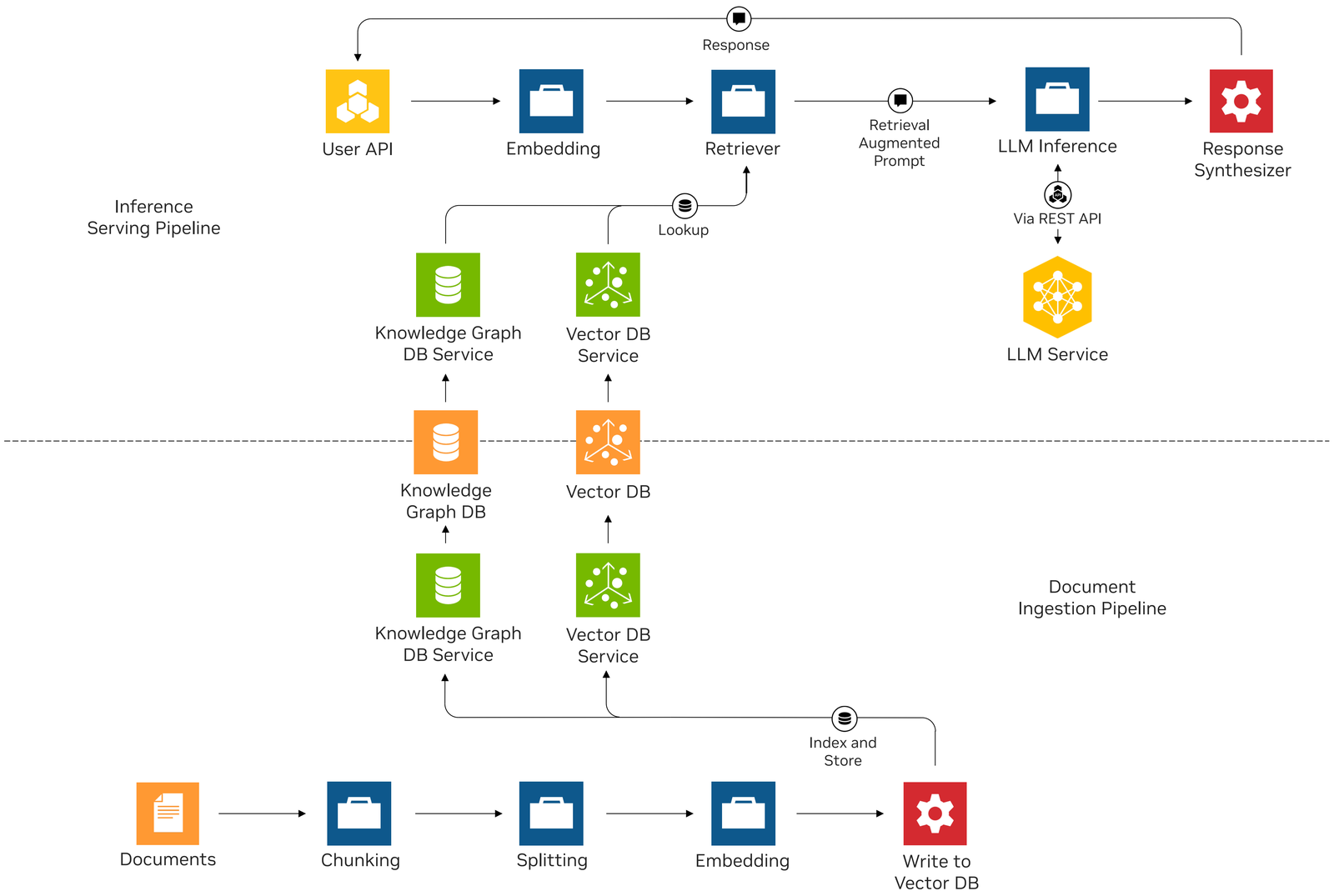
CVE Query: Is CVE-2022-42890 exploitable in morpheus:v23.07.00a?

Conclusion: The CVE is not exploitable in this container because Apache Batik, the vulnerable component, requires Java Runtime Environment which is not present.



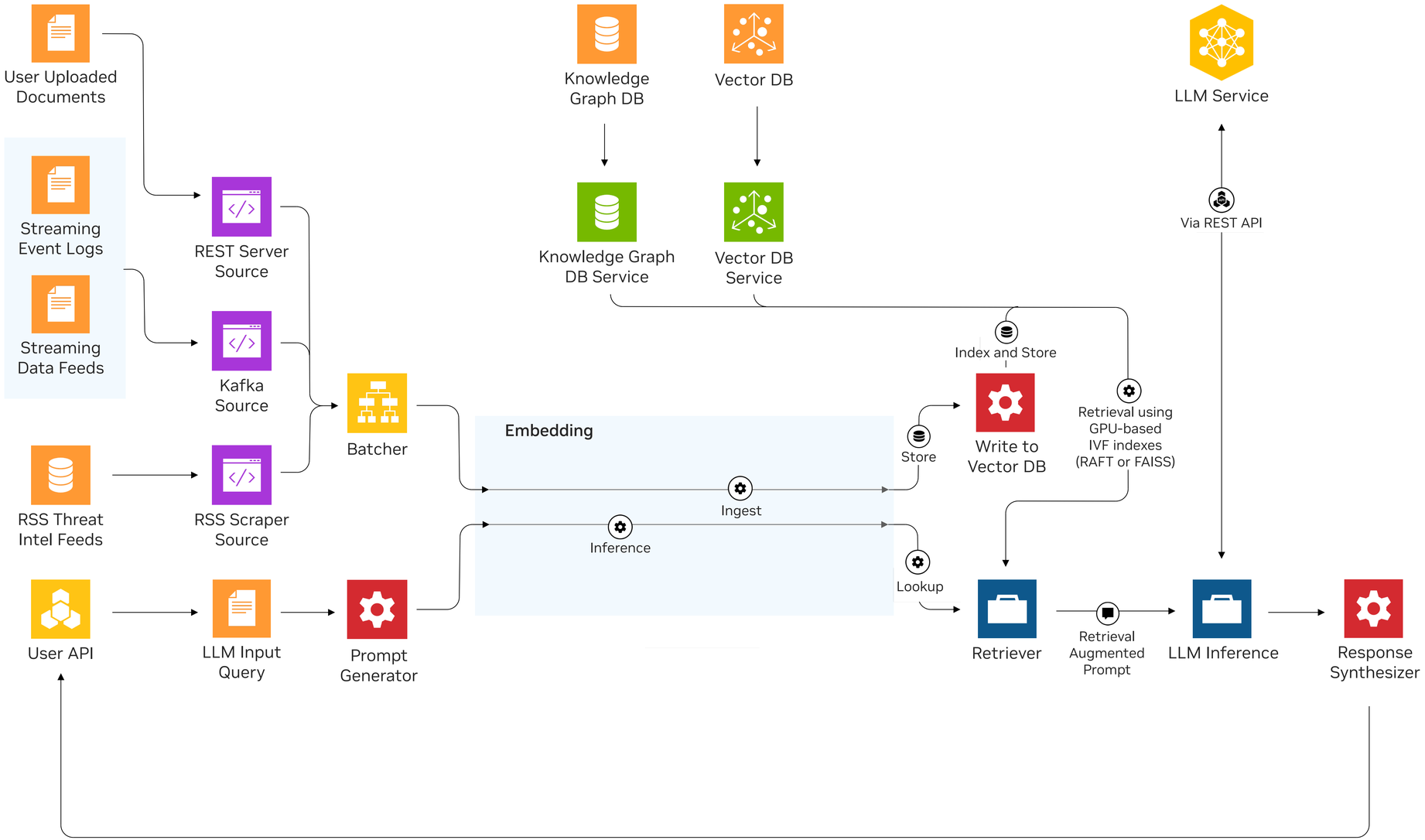
Traditional RAG Architecture

Document ingest doesn't need to be always on



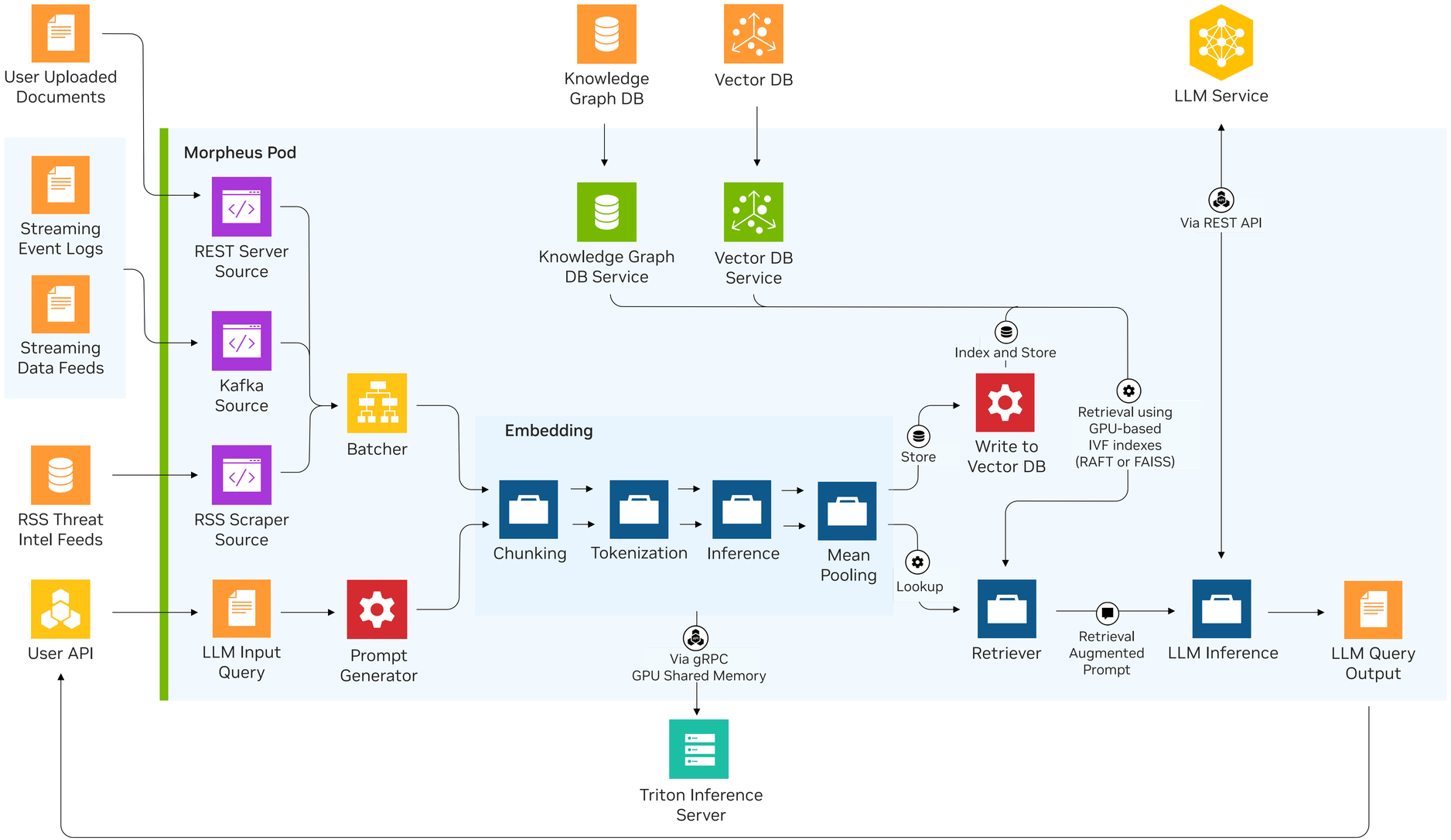
Cybersecurity Specific RAG Architecture

Need an always on service and as many reusable components as possible



Cybersecurity Specific RAG Architecture

Need an always on service and as many reusable components as possible



Foundation Models for Cybersecurity

Foundation Models for Cybersecurity

Enable next-generation models and detection via three key modalities

Address the Data Gap

Rare events mean there is limited data available to train models, and cross-org visibility is challenging

Perform “What If” Scenarios

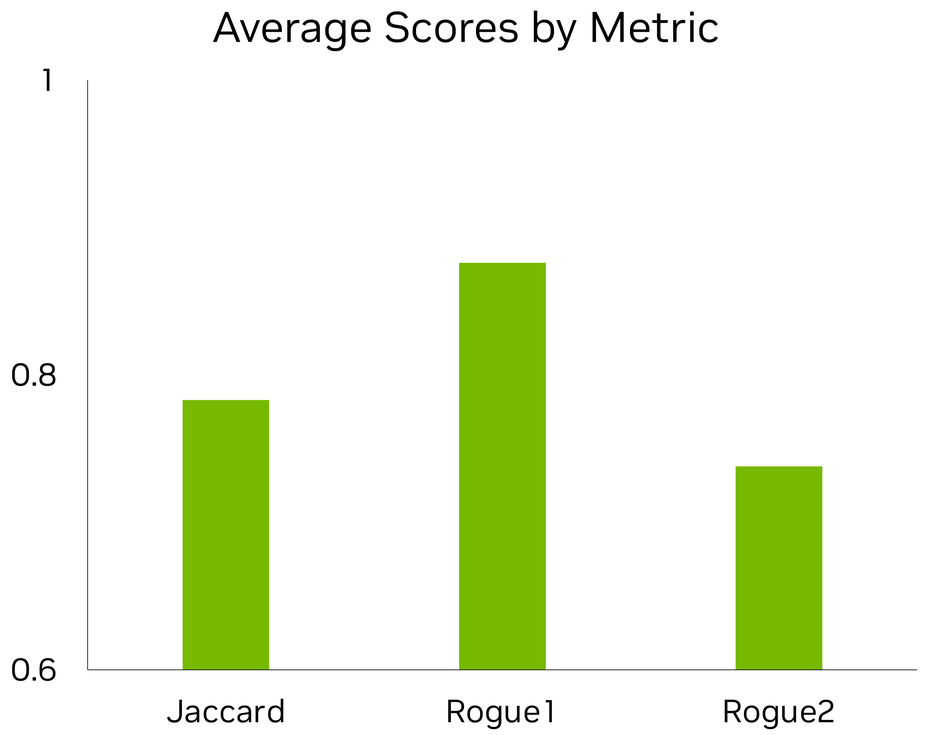
Afford threat hunters and red teams the ability to create data that is representative of an attack that didn’t happen

Feed Downstream Anomaly Detectors

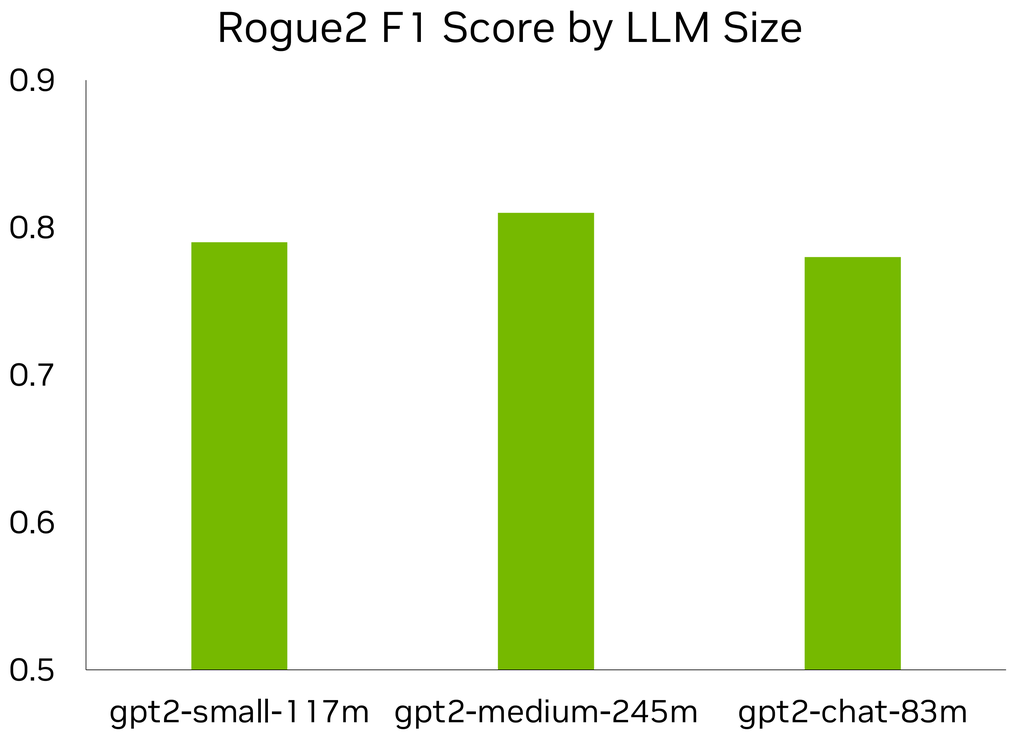
Utilize recent information in context with a LLM without retraining or fine-tuning frequently

CyberGPT Similarity Metrics

Measuring how well the synthetic data matches real-world data



Average similarity scores of logs generated with a character-level tokenizer compared to real logs

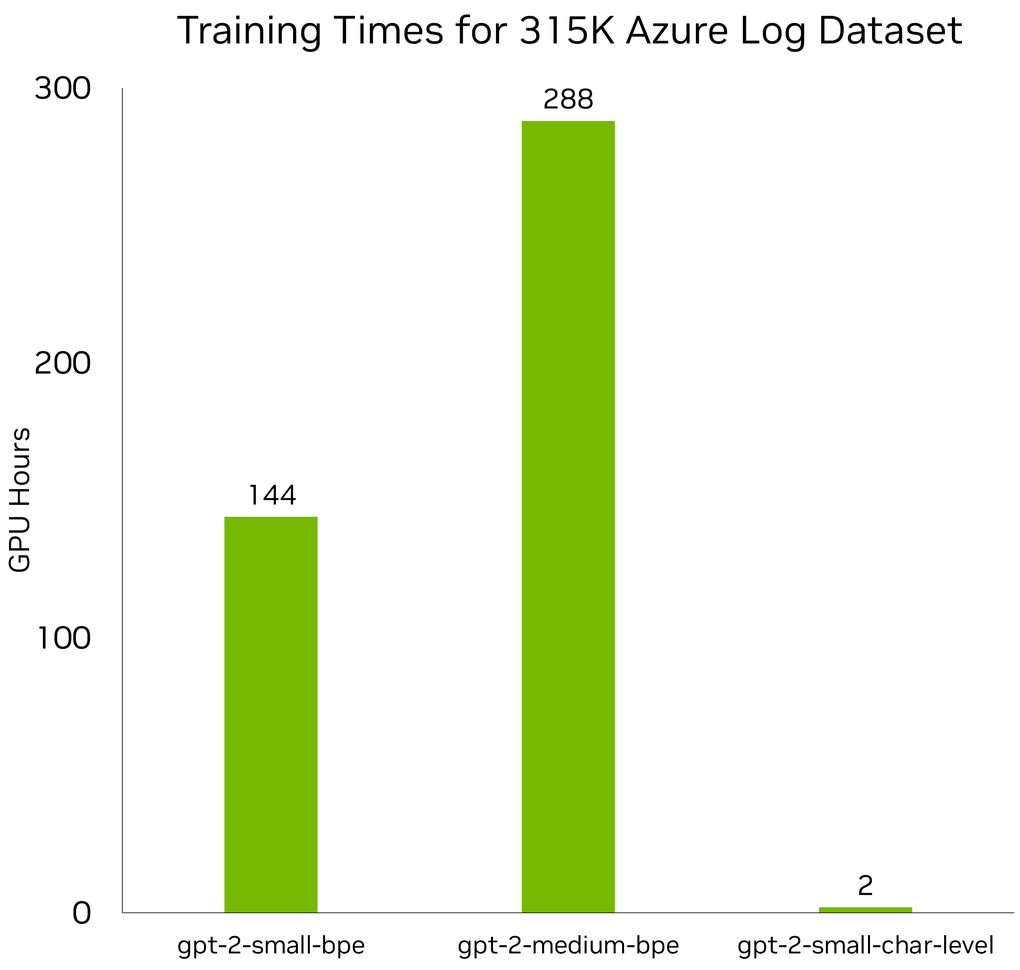


Comparison of average similarity scores between generated and real logs for various models

CyberGPT Training Times

A supercomputer is not needed to realize quality results

- Training time for user-specific Azure AD log generation is accomplishable in well under a week
  - As few as 2 GPU hours for GPT-2-small model with character-level tokenization
  - 5000 iterations for both BPE models
  - 3000 iterations for character-level models
  - Trained on 315k rows of data
- Training times using multi-user Azure, SharePoint, Confluence, and Jira logs was accomplished in under a day
  - 12 GPU hours using a GPT-2-small model with character-level tokenization
  - Trained on 2.3 million rows of 100 user logs with 1000 iterations
- Training times for RMA (Return Material Authorization) generating using summaries of raw log tarballs from DGX boxes also accomplished in under a day
  - 17 GPU hours using a GPT-2-small model with character-level tokenization
  - Trained on 973K rows from 331 different sets of logs with 1300 iterations



Choices for Tokenizer

Train your own or use an off-the-shelf model

Character Level Tokenizer

- Does not need tokenizer training
- No need to build and maintain a vocabulary of sub-words
- Can naturally handle unseen words
- Generated logs display lower similarity to actual logs compared to models trained with BPE tokenizer
- Can be used with smaller datasets and models

Off-the-Shelf Tokenizer

- Tends to split field names into multiple entries
- Extensive pre-existing vocabulary (50K)
- Can handle fields with long English text better
- Primarily oriented towards well-formatted text
- Faster than a Hugging Face tokenizer trained from scratch
- Unfamiliarity with domain-specific terminology and log syntax

Train Custom Tokenizer

- Custom tokenizer can handle log files' specific syntax and vocabulary
- Can effectively process log files' unique symbols and abbreviations
- Reduced tokenization errors
  - Trained not to split camel case field names
- Domain-specific terminology
- More efficient use of resources due to custom vocabulary

CyberGPT Demo

```
 Desktop clients", "userAgent": "Mozilla/5.0 (iPad; CPU iPad OS 9_3_6 like Mac OS X) AppleWebKit/536.1 (KHTML, like Gecko) CriOS/42.0.865.0 Mobile/77O124 Safari/536.1 Edge/99.14477", "deviceDetail": {"deviceId": "a44625d
c-6f81-449a-9799-8005f7209b42", "displayName": "ATTACKTARGET-LT", "operatingSystem": "Windows 10", "browser": "Edge 99.14477", "trustType": "Azure AD registered"}, "location": {"city": "New John", "state": "Smithfort", "
countryOrRegion": "XR", "geoCoordinates": {"latitude": -89.144242, "longitude": 168.119022}}, "correlationId": "fca5e7d1-51b0-4f8c-8d09-320267af2e83", "conditionalAccessStatus": "success", "appliedConditionalAccessPolici
es": [], "authenticationContextClassReferences": [], "originalRequestId": "8d4429ac-5131-4611-a5d8-d0ba6394fed3", "isInteractive": false, "tokenIssuerName": "", "tokenIssuerType": "AzureAD", "authenticationProcessingDeta
ils": [], "networkLocationDetails": [], "clientCredentialType": "none", "processingTimeInMilliseconds": 187, "riskDetail": "none", "riskLevelAggregated": "none", "riskLevelDuringSignIn": "none", "riskState": "none", "ris
kEventTypes": [], "riskEventTypes_v2": [], "resourceDisplayName": "Adobe Identity Management Service", "resourceId": "80abac89-d2c0-4b6c-8ffc-5023bd7efc5e", "resourceTenantId": "d3e5a967-5657-4a42-afcc-6106b6c3c299", "ho
meTenantId": "d3e5a967-5657-4a42-afcc-6106b6c3c299", "authenticationDetails": [], "authenticationRequirementPolicies": [], "authenticationRequirement": "singleFactorAuthentication", "servicePrincipalId": "", "userType":
"Member", "flaggedForReview": false, "isTenantRestricted": false, "autonomousSystemNumber": 34974, "crossTenantAccessType": "none", "privateLinkDetails": {}, "ssoExtensionVersion": "", "uniqueTokenIdentifier": "c0ISzRKIb
8kGXywGLXT1MvLPdAQcc2CXMf13L20MmBvpIv2", "incomingTokenType": "primaryRefreshToken", "authenticationProtocol": "none", "appServicePr[redacted]null, "resourceServicePrincipalId": null}}, {"time": "2022-08-08T07:52:01.11
7249Z", "resourceId": "/tenants/d3e5a967-5657-4a42-afcc-6106b6c3c299/providers/Microsoft.aadiam", "operationName": "Sign-in activi[redacted]rsion": "1.0", "category": "NonInteractiveUserSignInLogs", "tenantId": "d
3e5a967-5657-4a42-afcc-6106b6c3c299", "resultType": "0", "resultSignature": "None", "durationMs": 0, "callerIpAddress": "13.113.[redacted]Id": "029c45d0-237e-4dca-abec-59634f6dc8a2", "identity": "Attack Target
", "Level": 4, "location": "XR", "properties": {"id": "c95265b3-8565-485d-9c63-18c21fbc6300", "createdDateTime": "2022-08-08T07[redacted]1.808249Z", [redacted]isplayName": "Attack Target", "userPrincipalName": "attacktarget@domai
n.com", "userId": "d735e84b-dcca-404d-9f7d-700f360f41a6", "appId": "9a7e67c7-6f05-42a3-b226-97c7ec3e9696", "appDisplayName": "[redacted] Identity Manag[redacted]", "ipAddress": "13.113.40.157", "status": {"errorCode": 0}, "clientA
ppUsed": "Mobile Apps and Desktop clients", "userAgent": "Mozilla/5.0 (iPad; CPU iPad OS 9_3_6 like Mac OS X) AppleWebKit/536.[redacted]TML, like Gecko)[redacted]S/42.0.865.0 Mobile/77O124 Safari/536.1 Edge/99.14477", "deviceDetai
l": {"deviceId": "a44625dc-6f81-449a-9799-8005f7209b42", "displayName": "ATTACKTARGET-LT", "operatingSystem": "Windows 10", "br[redacted]": "Edge 99.144[redacted]"trustType": "Azure AD registered"}, "location": {"city": "New John"
, "state": "Smithfort", "countryOrRegion": "XR", "geoCoordinates": {"latitude": -89.144242, "longitude": 168.119022}}, [redacted]orrela[redacted]": "5a490380-[redacted]4232-bed6-0e95ab1954e0", "conditionalAccessStatus": "success", "appli
edConditionalAccessPolicies": [], "authenticationContextClassReferences": [], "originalRequestId": "e88d2a9d-2b5f-4[redacted]dd-dcd30[redacted], "isInte[redacted]ve": false, "tokenIssuerName": "", "tokenIssuerType": "AzureAD", "auth
enticationProcessingDetails": [], "networkLocationDetails": [], "processingTimeInMilliseconds": 102, "riskDetail"[redacted]e", "riskLev[redacted]ne", "riskLevelDuringSignIn": "none", "riskState": "none", "riskEventTy
pes": [], "riskEventTypes_v2": [], "resourceDisplayName": "Adobe Identity Management Service", "resourceId": "7[redacted]7-8b82-49e9-90e2[redacted], "resourceTenantId": "d3e5a967-5657-4a42-afcc-6106b6c3c299", "homeTenant
Id": "d3e5a967-5657-4a42-afcc-6106b6c3c299", "authenticationDetails": [], "authenticationRequirementPolicies"[redacted]authent[redacted]tionRequirem[redacted]: "singleFactorAuthentication", "servicePrincipalId": "", "userType": "Member"
, "flaggedForReview": false, "isTenantRestricted": false, "autonomousSystemNumber": 34974, "crossTenantAcces[redacted]"none[redacted]ivateLinkDetails": {}, "ssoExtensionVersion": "", "uniqueTokenIdentifier": "kEZXLwgcXK12JNOoD
XulmvdQLBybb3RjlSk77qYLv1vHXMbA", "incomingTokenType": "primaryRefreshToken", "authenticationProtocol": "none", [redacted]incipalId": null, "resourceServicePrincipalId": "537d9419-7e2b-4778-9102-62256bcfcf45", "rngcSt
atus": 0}}, {"time": "2022-08-08T08:04:31.446503Z", "resourceId": "/tenants/d3e5a967-5657-4a42-afcc-6106b6c3c299/[redacted]Microsof[redacted]am", "operationName": "Sign-in activity", "operationVersion": "1.0", "category": "N
onInteractiveUserSignInLogs", "tenantId": "d3e5a967-5657-4a4
---------------
 "0", "resultSignature": "None", "durationMs": 0, "callerIpAddress": "156.206.181.119", "correla[redacted]d-8813[redacted]2c51cf044c9", "identity": "Cynthia Perry", "Level": 4, "location": "XY", "properties": {"
id": "9a4e5780-0772-478f-b2d6-19abac955e00", "createdDateTime": "2022-08-06T19:29:25.4392592Z", [redacted]Display[redacted]"Cynthi[redacted], "userPrincipalName": "cperry@domain.com", "userId": "ac5fa380-9d5b-470e-beb2-5ba10aca48e6
", "appId": "9de6e985-97a5-4d82-9c45-aedaa9c25ef9", "appDisplayName": "AWS Single-Account Ac[redacted]"ipAddress[redacted].206.181[redacted]9", "status": {"errorCode": 0}, "clientAppUsed": "Browser", "deviceDetail": {"deviceId": "",
"operatingSystem": "Windows"}, "location": {"city": "Hannaville", "state": "Thomashaven", [redacted]ryOrRegion": "X[redacted]eoCoordinates": {"latitude": 1.1452695, "longitude": 128.249211}}, "correlationId": "aadce4b0-52ff-4dc8
-b7f8-c24cf4b885b4", "conditionalAccessStatus": "success", "appliedConditionalAccessPolici[redacted]"authenticati[redacted]textClassReferences": [], "originalRequestId": "9682756d-2d87-43ad-95e1-1982697404e1", "isInteractive":
 false, "tokenIssuerName": "", "tokenIssuerType": "AzureAD", "authenticationProcessingDetails[redacted]"networkLocat[redacted]etails": [], "clientCredentialType": "none", "processingTimeInMilliseconds": 108, "riskDetail": "none",
 "riskLevelAggregated": "none", "riskLevelDuringSignIn": "none", "riskState": "none[redacted]"riskEven[redacted]: [], "ris[redacted]tTypes_v2": [], "resourceDisplayName": "AWS Single-Account Access Service", "resourceId": "9a84f55e-3a4
0-492c-9476-e069f64ed7a6", "resourceTenantId": "d3e5a967-5657-4a42-afcc-6106b6c3c[redacted]omeTenan[redacted]d3e5a9[redacted]7-4a42-afcc-6106b6c3c299", "authenticationDetails": [], "authenticationRequirementPolicies": [], "authen
ticationRequirement": "singleFactorAuthentication", "servicePrincipalId": "", "user[redacted]"Member", [redacted]d[redacted]iew": false, "isTenantRestricted": false, "autonomousSystemNumber": 279165, "crossTenantAccessType": "none
", "privateLinkDetails": {}, "ssoExtensionVersion": "", "uniqueTokenIdentifi[redacted]"fyK[redacted]Zx064b3Q[redacted]r105vZsw8IlOMeEQrz", "incomingTokenType": "none", "authenticationProtocol": "none", "appServicePrincipalId":
 null, "resourceServicePrincipalId": "37ee873b-f901-4e25-be69-010243106d0f"[redacted]time": [redacted]8-06T19:[redacted]452902", "resourceId": "/tenants/d3e5a967-5657-4a42-afcc-6106b6c3c299/providers/Microsoft.aadiam", "operationN
ame": "Sign-in activity", "operationVersion": "1.0", "category": "NonInterac[redacted]erSignIn[redacted]"tenantId": "d3e5a967-5657-4a42-afcc-6106b6c3c299", "resultType": "0", "resultSignature": "None", "durationMs": 0, "callerI
pAddress": "156.206.181.119", "correlationId": "52d99b7f-d0cb-47f2-b3a2-1fdf8[redacted]6", "iden[redacted]"Cynthia Perry", "Level": 4, "location": "XY", "properties": {"id": "3f81cc81-5415-4518-bdd8-7d53b3bc6200", "createdDateT
ime": "2022-08-06T19:32:41.1312902", "userDisplayName": "Cynthia Perry", "user[redacted]ipalName": [redacted]domain.com", "userId": "ac5fa380-9d5b-470e-beb2-5ba10aca48e6", "appId": "9de6e985-97a5-4d82-9c45-aedaa9c25ef9", "appDi
splayName": "AWS Single-Account Access", "ipAddress": "156.206.181[redacted]119", "statu[redacted]errorCode": [redacted]"clientAppUsed": "Browser", "deviceDetail": {"deviceId": "", "operatingSystem": "Windows"}, "location": {"city": "Hanna
ville", "state": "Thomashaven", "countryOrRegion": "XY", "geoCo[redacted]": {"lati[redacted]": 1.1452695, "longitude": 128.249211}}, "correlationId": "c585b228-efba-4b57-a82a-ff25512e536a", "conditionalAccessStatus": "success",
 "appliedConditionalAccessPolicies": [], "authenticationContextC[redacted]es": [[redacted]riginalRequestId": "2435bad9-fdfe-4251-b076-d87f574d968d", "isInteractive": false, "tokenIssuerName": "", "tokenIssuerType": "AzureAD
", "authenticationProcessingDetails": [], "networkLocationDetails": [[redacted]red[redacted]lType": "none", "processingTimeInMilliseconds": 92, "riskDetail": "none", "riskLevelAggregated": "none", "riskLevelDuringSignIn": "n
one", "riskState": "none", "riskEventTypes": [], "riskEventTypes_v2": [], [redacted]yName": "AWS Single-Account Access Service", "resourceId": "9a84f55e-3a40-492c-9476-e069f64ed7a6", "resourceTenantId": "d3e5a967-56
57-4a42-afcc-6106b6c3c299", "homeTenantId": "d3e5a967-5657-4a42-afcc-6106b6c3c[redacted]enticationDetails": [], "authenticationRequirementPolicies": [], "authenticationRequirement": "singleFactorAuthentication", "servic
ePrincipalId": "", "userType": "Member", "flaggedForReview": false, "isTenantRestri[redacted]: false, "autonomousSystemNumber": 279165, "crossTenantAccessType": "none", "privateLinkDetails": {}, "ssoExtensionVersion": "", "un
iqueTokenIdentifier": "JM2NyqLt1JGIlIZXlWYx1vUBxuINggWGjNFgNPNISUb1RaNP", "incomingTokenType": "none", "authenticationProtocol": "none", "appServicePrincipalId": null, "resourceServicePrincipalId": "37ee873b-f901-4e25-be
69-010243106d0f", "rngcStatus": 0}}, {"time": "2022-08-06T19:42:19.512421Z", "resourceId": "/tenants/d3e5a967-5657-4a42-afcc-6106b6c3c299/providers/Microsoft.aadiam", "operationName": "Sign-in activity", "operationVersio
n": "1.0", "category": "NonInteractiveUserSignInLogs", "tenantId": "d3e5a967-5657-4a42-afcc-6106b6c3c299", "resultType": "0", "resultSignature": "None", "durationMs": 0, "callerIpAddress": "13.113.40.157", "correlationId
": "96f79d07-b1b6-4ba8-9e47-89f7dc4fa2ed", "identity": "Attack Target", "Level": 4, "location": "XR", "properties": {"id": "d12422cd-6c99-4688-9c71-64453606300", "createdDateTime": "2022-08-06T19:42:19.8874212Z", "userDis
playName": "Attack Target", "userPrincipalName": "attacktarget@domain.com", "userId": "d735e84b-dcca-404d-9f7d-700f360f41a6", "appId": "9a7e67c7-6f05-42a3-b226-97c7ec3e9696", "appDisplayName": "Adobe Identity Management"
, "ipAddress": "13.113.40.157", "status": {"errorCode": 0}, "clientAppUsed": "Mobile Apps and Desktop clients", "userAgent": "Mozilla/5.0 (iPad; CPU iPad OS 9_3_6 like Mac OS X) AppleWebKit/536.1 (KHTML, like Gecko) CriO
S/42.0.865.0 Mobile/77O124 Safari/536.1 Edge/99.14477", "dev
---------------
```


VIDEO

Example Data Generated

"callerIpAddress": "205.33.24.92", "correlationId": "65e13716-16ae-45df-9b01-95ca34682f12", "identity": "Aaron Cole", "Level": 4, "location": "XD", "properties": {"id": "0c4001e3-9786-4a85-85d9-0c8dd8628500", "createdDateTime": "2022-08-12T13:56:01.692513Z", "userDisplayName": "Aaron Cole", "userPrincipalName": "acole@domain.com", "userId": "5d557969-1645-4ba4-be83-b6fd943659f7", "appId": "af50c7b0-ad5f-40ec-840e-8d89be454ce4", "appDisplayName": "Citrix ShareFile", "ipAddress": "205.33.24.92", "status": {"errorCode": 0}, "clientAppUsed": "Browser", "deviceDetail": {"deviceId": "", "operatingSystem": "Windows10", "browser": "Rich Client 4.36.1.0"}, "location": {"city": "Kristinmouth", "state": "Norrisbury", "countryOrRegion": "XD", "geoCoordinates": {"latitude": 26.278784, "longitude": -92.042573}}, "correlationId": "01786521-b9fe-48c9-b4d5-8b1aac1451f7", "conditionalAccessStatus": "success", "appliedConditionalAccessPolicies": [], "authenticationContextClassReferences": [], "originalRequestId": "2ab20a27-bc3d-4595-a084-4662d3f5225", "isInteractive": false, "tokenIssuerName": "", "tokenIssuerType": "AzureAD", "authenticationProcessingDetails": [], "networkLocationDetails": [], "clientCredentialType": "none", "processingTimeInMilliseconds": 80, "riskDetail": "none", "riskLevelAggregated": "none", "riskLevelDuringSignIn": "none", "riskState": "none", "riskEventTypes": [], "riskEventTypes_v2": [], "resourceDisplayName": "Citrix ShareFile Online", "resourceId": "24578515-1951-4064-967d-0d2acaf0aa70", "resourceTenantId": "d3e5a967-5657-4a42-afcc-6106b6c3c299", "homeTenantId": "d3e5a967-5657-4a42-afcc-6106b6c3c299", "authenticationDetails": [], "authenticationRequirementPolicies": [], "authenticationRequirement": "singleFactorAuthentication", "servicePrincipalId": "", "userType": "Member", "flaggedForReview": false, "isTenantRestricted": false, "autonomousSystemNumber": 214655, "crossTenantAccessType": "none", "privateLinkDetails": {}, "ssoExtensionVersion": "", "uniqueTokenIdentifier": "cZWoA12L8kyHE3IhpR97xXmX4VxDpry8dYSntsVNuJUUve0", "incomingTokenType": "none", "authenticationProtocol": "none", "appServicePrincipalId": null, "resourceServicePrincipalId": "0e68752b-7dfa-400b-a6f5-1b54af90696a", "rngcStatus": 0}}, {"time": "2022-08-12T13:54:11.810994Z", "resourceId": "/tenants/d3e5a967-5657-4a42-afcc-6106b6c3c299/providers/Microsoft.aadiam", "operationName": "Sign-in activity", "operationVersion": "1.0", "category": "NonInteractiveUserSignInLogs", "tenantId": "d3e5a967-5657-4a42-afcc-6106b6c3c299", "resultType": "0", "resultSignature": "None", "durationMs": 0, "callerIpAddress": "13.113.40.157", "correlationId": "92afe834-f63c-4c27-b042-b15538ad86d7", "identity": "Attack Target", "Level": 4, "location": "XR", "properties": {"id": "8e4e2bcd-f052-4897-bda3-4a72f0da6800", "createdDateTime": "2022-08-12T13:54:12.506994Z", "userDisplayName": "Attack Target", "userPrincipalName": "attacktarget@domain.com", "userId": "d735e84b-dcca-404d-9f7d-700f360f41a6", "appId": "9c5b7fe3-0ad2-4ea6-94e5-9e0001f367e3", "appDisplayName": "Articulate 360", "ipAddress": "13.113.40.157", "status": {"errorCode": 0}, "clientAppUsed": "Mobile Apps and Desktop clients", "userAgent": "Mozilla/5.0 (iPad; CPU iPad OS 9_3_6 like Mac OS X) AppleWebKit/536.1 (KHTML, like Gecko) CriOS/42.0.865.0 Mobile/77O124 Safari/536.1 Edge/99.14477", "deviceDetail": {"deviceId": "a44625dc-6f81-449a-9799-8005f7209b42", "displayName": "ATTACKTARGET-LT", "operatingSystem": "Windows 10", "browser": "Edge 99.14477", "trustType": "Azure AD registered"}, "location": {"city": "Smithfort", "state": "Smithfort", "countryOrRegion": "XR", "geoCoordinates": {"latitude": 3.7564095, "longitude": -121.574606}}, "correlationId": "5686e723-f969-4cee-b7ea-1c3050a65204", "conditionalAccessStatus": "success", "appliedConditionalAccessPolicies": [], "authenticationContextClassReferences": [], "originalRequestId": "bb69a40d-1688-4816-88ff-dadc99798427", "isInteractive": false, "tokenIssuerName": "", "tokenIssuerType": "AzureAD", "authenticationProcessingDetails": [], "networkLocationDetails": [], "clientCredentialType": "none", "processingTimeInMilliseconds": 253, "riskDetail": "none", "riskLevelAggregated": "none", "riskLevelDuringSignIn": "none", "riskState": "none", "riskEventTypes": [], "riskEventTypes_v2": [], "resourceDisplayName": "Articulate 360 Online", "resourceId": "436260d5-3791-4aea-a4da-aa698ec35f2f", "resourceTenantId": "d3e5a967-5657-4a42-afcc-6106b6c3c299", "homeTenantId": "d3e5a967-5657-4a42-afcc-6106b6c3c299", "authenticationDetails": [], "authenticationRequirementPolicies": [], "authenticationRequirement": "singleFactorAuthentication", "servicePrincipalId": "", "userType": "Member", "flaggedForReview": false, "isTenantRestricted": false, "autonomousSystemNumber": 34974, "crossTenantAccessType": "none", "privateLinkDetails": {}, "ssoExtensionVersion": "", "uniqueTokenIdentifier": "H4wrwcKTXDprRUSop3fI0s3FHaErLZhOXd1hMe85ah9UJGE", "incomingTokenType": "primaryRefreshToken", "authenticationProtocol": "none", "appServicePrincipalId": null, "resourceServicePrincipalId": "94b9f3ba-5f4b-4edc-99dc-a420b76960f9", "rngcStatus": 0}}, {"time": "2022-08-12T13:04:18.566764Z", "resourceId": "/tenants/d3e5a967-5657-4a42-afcc-6106b6c3c299/providers/Microsoft.aadiam", "operationName": "Sign-in activity", "operationVersion": "1.0", "category": "NonInteractiveUserSignInLogs", "tenantId": "d3e5a967-5657-4a42-afcc-6106b6c3c299", "resultType": "0", "resultSignature": "None", "durationMs": 0, "callerIpAddress": "13.113.40.157", "correlationId": "22f6c158-de01-4441-820d-53c87596e6f1", "identity": "Attack Target", "Level": 4, "location": "XR", "properties": {"id": "93f00e17-5696-4ee2-a777-7961678c6700", "createdDateTime": "2022-08-12T13:04:18.621764Z", "userDisplayName": "Attack Target", "userPrincipalName": "attacktarget@domain.com", "userId": "d735e84b-dcca-404d-9f7d-700f360f41a6", "appId": "9a7e67c7-6f05-42a3-b226-97c7ec3e9696", "appDisplayName": "Adobe Identity Man
---
"none", "riskLevelAggregated": "none", "riskLevelDuringSignIn": "none", "riskState": "none", "riskEventTypes": [], "riskEventTypes_v2": [], "resourceDisplayName": "Adobe Identity Management Service", "resourceId": "30663ca7-f8a9-43ab-8ca1-7906bdbc1485", "resourceTenantId": "d3e5a967-5657-4a42-afcc-6106b6c3c299", "homeTenantId": "d3e5a967-5657-4a42-afcc-6106b6c3c299", "authenticationDetails": [], "authenticationRequirementPolicies": [], "authenticationRequirement": "singleFactorAuthentication", "servicePrincipalId": "", "userType": "Member", "flaggedForReview": false, "isTenantRestricted": false, "autonomousSystemNumber": 230297, "crossTenantAccessType": "none", "privateLinkDetails": {}, "ssoExtensionVersion": "", "uniqueTokenIdentifier": "6F2AZCfNgvm8mjBoJI26SPY0qheWRAIp47yeWAIVoDSdEgjW", "incomingTokenType": "primaryRefreshToken", "authenticationProtocol": "none", "appServicePrincipalId": null, "resourceServicePrincipalId": "a6c259e5-7f16-48b2-a9f3-75becd6daa9b"}}, {"time": "2022-08-17T18:07:36.993929Z", "resourceId": "/tenants/d3e5a967-5657-4a42-afcc-6106b6c3c299/providers/Microsoft.aadiam", "operationName": "Sign-in activity", "operationVersion": "1.0", "category": "NonInteractiveUserSignInLogs", "tenantId": "d3e5a967-5657-4a42-afcc-6106b6c3c299", "resultType": "0", "resultSignature": "None", "durationMs": 0, "callerIpAddress": "13.113.40.157", "correlationId": "a8c05930-a44c-4906-b691-9362743ec925", "identity": "Attack Target", "Level": 4, "location": "XR", "properties": {"id": "cece7f81-9649-4ab1-86ee-ff7c642d9500", "createdDateTime": "2022-08-17T18:07:36.124929Z", "userDisplayName": "Attack Target", "userPrincipalName": "

Example Data Generated

"callerIpAddress": "205.33.24.92",
"correlationId": "65e13716-16ae-45df-9b01-95ca34682f12", "identity": "Aaron Cole", "Level": 4, "location": "XD",
"properties": {"id": "0c4001e3-9786-4a85-85d9-0c8dd8628500",
"createdDateTime": "2022-08-12T13:56:01.692513Z",
"userDisplayName": "Aaron Cole",
"userPrincipalName":
"acole@domain.com", "userId":
"5d557969-1645-4ba4-be83-b6fd943659f7", "appId": "af50c7b0-ad5f-40ec-840e-8d89be454ce4",
"appDisplayName": "Citrix ShareFile",
"ipAddress": "205.33.24.92", "status":
{"errorCode": 0}, "clientAppUsed":
"Browser", "deviceDetail":
{"deviceId": "", "operatingSystem":
"Windows10", "browser": "Rich Client 4.36.1.0"}, "location": {"city":
"Kristinmouth", "state": "Norrisbury",
"countryOrRegion": "XD",
"geoCoordinates": {"latitude":
26.278784, "longitude": -92.042573}},
"correlationId": "01786521-b9fe-48c9-b4d5-8b1aac1451f7",
"conditionalAccessStatus": "success",
"appliedConditionalAccessPolicies":
[],
"authenticationContextClassReferences"
: [], "originalRequestId": "2ab20a27-bc3d-4595-a084-4662d3f5225",
"isInteractive": false,
"tokenIssuerName": "",
"tokenIssuerType": "AzureAD",
"authenticationProcessingDetails": [],
"networkLocationDetails":
"authenticationRequirement

"callerIpAddress": "205.33.24.92", "correlationId": "65e13716-16ae-45df-9b01-95ca34682f12", "identity": "Aaron Cole", "Level": 4, "location": "XD",
"properties": {"id": "0c4001e3-9786-4a85-85d9-0c8dd8628500", "createdDateTime":
"2022-08-12T13:56:01.692513Z", "userDisplayName": "Aaron Cole",
"userPrincipalName": "acole@domain.com", "userId": "5d557969-1645-4ba4-be83-b6fd943659f7", "appId": "af50c7b0-ad5f-40ec-840e-8d89be454ce4", "appDisplayName":
"Citrix ShareFile", "ipAddress": "205.33.24.92", "status": {"errorCode": 0},
"clientAppUsed": "Browser", "deviceDetail": {"deviceId": "", "operatingSystem":
"Windows10", "browser": "Rich Client 4.36.1.0"}, "location": {"city":
"Kristinmouth", "state": "Norrisbury", "countryOrRegion": "XD", "geoCoordinates":
{"latitude": 26.278784, "longitude": -92.042573}}, "correlationId": "01786521-b9fe-48c9-b4d5-8b1aac1451f7", "conditionalAccessStatus": "success",
"appliedConditionalAccessPolicies": [], "authenticationContextClassReferences":
[], "originalRequestId": "2ab20a27-bc3d-4595-a084-4662d3f5225", "isInteractive":
false, "tokenIssuerName": "", "tokenIssuerType": "AzureAD",
"authenticationProcessingDetails": [], "networkLocationDetails":
"authenticationRequirement": "singleFactorAuthentication", "servicePrincipalId":
"", "userType": "Member", "flaggedForReview": false, "isTenantRestricted": false,
"incomingTokenType": "none", "authenticationProtocol": "none",
"appServicePrincipalId": null, "resourceServicePrincipalId": "0e68752b-7dfa-400b-a6f5-1b54af90696a", "rngcStatus": 0}}, {"time": "2022-08-12T13:54:11.810994Z",
"resourceId": "/tenants/d3e5a967-5657-4a42-afcc-6106b6c3c299/providers/Microsoft.aadiam", "operationName": "Sign-in activity",
"operationVersion": "1.0", "category": "NonInteractiveUserSignInLogs", "tenantId":
"d3e5a967-5657-4a42-afcc-6106b6c3c299", "resultType": "0", "resultSignature":
"None", "durationMs": 0, "callerIpAddress": "13.113.40.157", "correlationId":
"92afe834-f63c-4c27-b042-b15538ad86d7", "identity": "Attack Target", "Level": 4,
"location": "XR", "properties": {"id": "8e4e2bcd-f052-4897-bda3-4a72f0da6800",
"createdDateTime": "2022-08-12T13:54:12.506994Z", "userDisplayName": "Attack Target", "userPrincipalName": "attacktarget@domain.com", "userId": "d735e84b-dcca-404d-9f7d-700f360f41a6", "appId": "9c5b7fe3-0ad2-4ea6-94e5-9e0001f367e3",
"appDisplayName": "Articulate 360", "ipAddress": "13.113.40.157", "status":
{"errorCode": 0}, "clientAppUsed": "Mobile Apps and Desktop clients", "userAgent":
"Mozilla/5.0 (iPad; CPU iPad OS 9_3_6 like Mac OS X) AppleWebKit/536.1 (KHTML, like Gecko) CriOS/42.0.865.0 Mobile/77O124 Safari/536.1 Edge/99.14477",
"deviceDetail": {"deviceId": "a44625dc-6f81-449a-9799-8005f7209b42",
"displayName": "ATTACKTARGET-LT", "operatingSystem": "Windows 10", "browser":
"Edge 99.14477", "trustType": "Azure AD registered"}, "location": {"city":
"Smithfort", "state":

Application Specific Synthetic Data Generation and Model Training

Generative AI for Synthetic Data Generation

Multiple modalities that address the lack of available training data

Create Content at Scale

Classification of artificial intelligence that can produce various types of content, including text, images, and audio. Recent advances allow generation based on prompts and existing content.

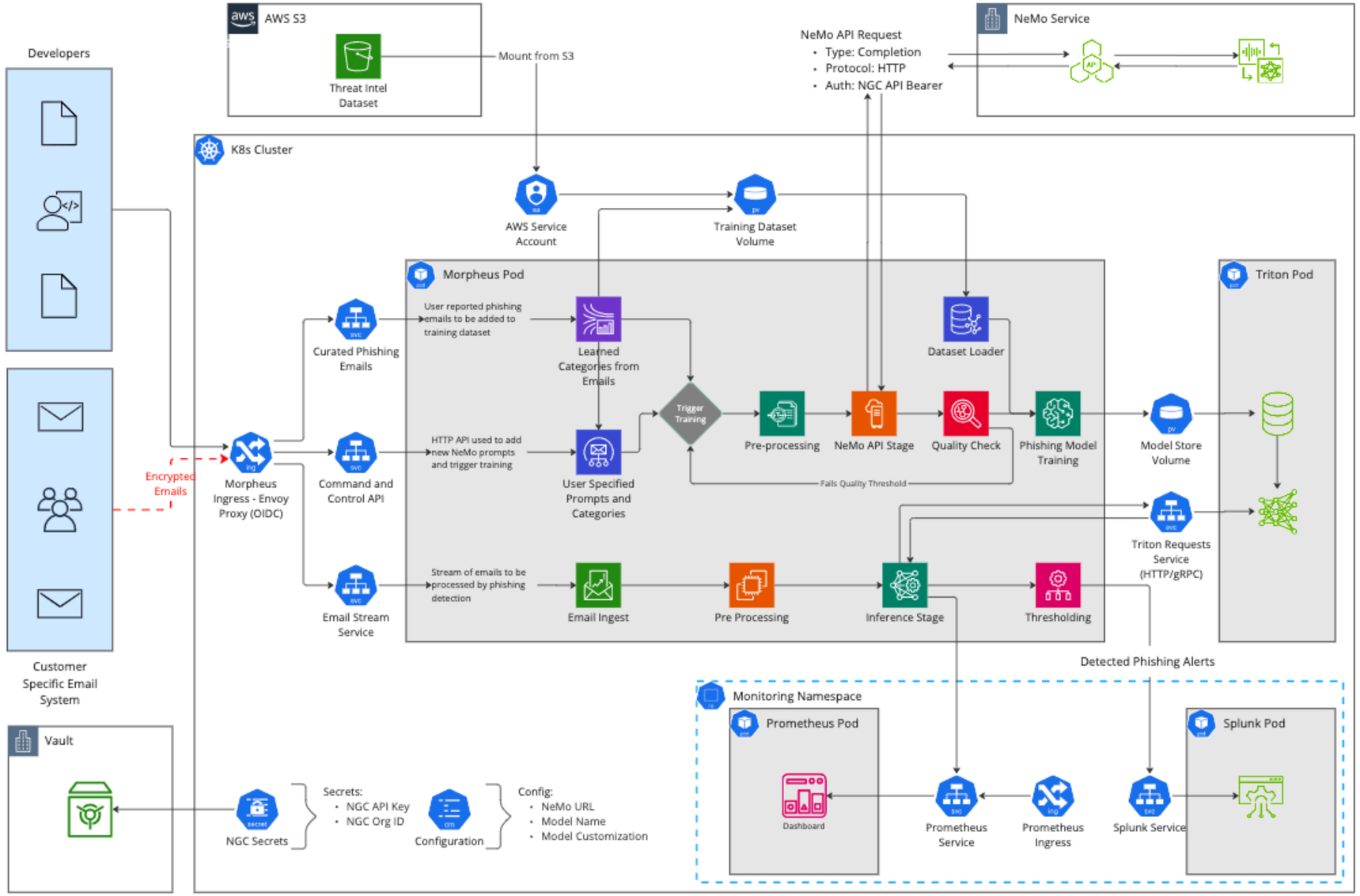
Multiple Types of Models and Architectures

Includes GANs (Generative Adversarial Networks), lighter weight language models, and larger models with 100+ billion parameters (e.g., GPT-3).

Harness for Increased Data Visibility and Availability

Because these techniques generate content like existing examples, what are ways we can utilize generative models to help address the lack of available training data for specific cybersecurity scenarios?

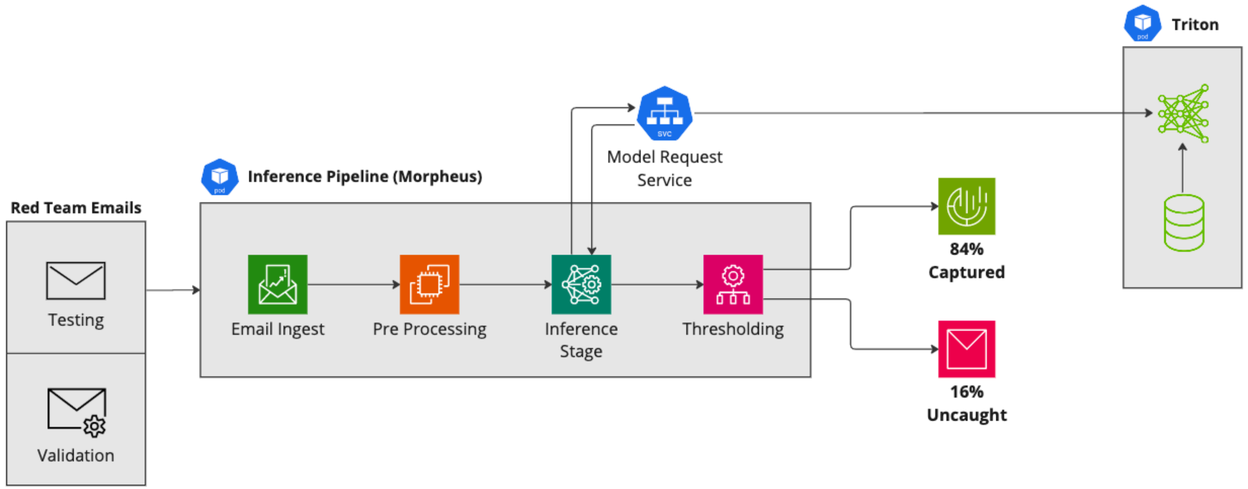
Spear Phishing Architecture



Synthetic Data Generation for Spear Phishing

Default models trained for off-the-shelf intents

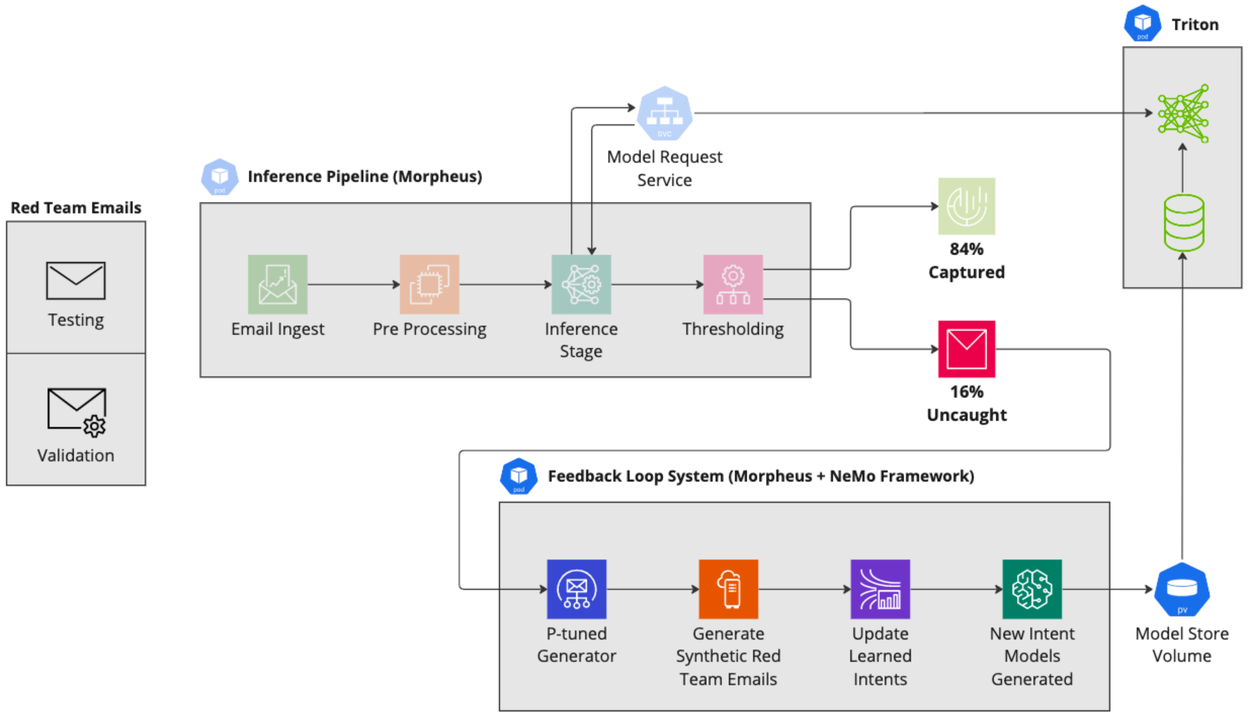
- Spear phishing detection pipeline built using synthetically generated spear phishing emails that correspond to specific behavioral intents
- Detection was 84% accurate on tests of NVIDIA internal red team emails
- Of these emails, 16% were not caught using the off-the-shelf models



Synthetic Data Generation for Spear Phishing

Feedback loop to learn from uncaught emails results in higher detection accuracy

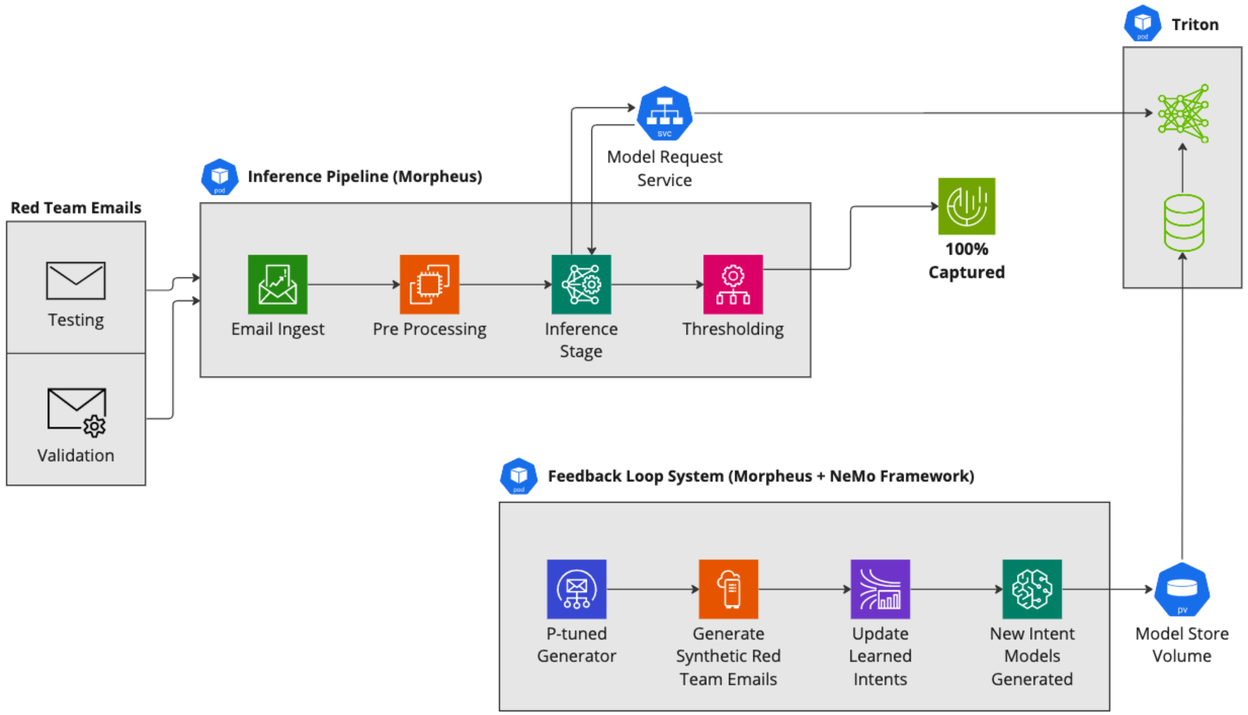
- Uncaught emails (16%, ~600 red team emails) were used to create a new synthetic dataset from the p-tuned generator
- New intent models were learned from the large corpus of synthetic emails and integrated into the spear phishing detection pipeline



Synthetic Data Generation for Spear Phishing

After feedback loop, achieved 100% detection

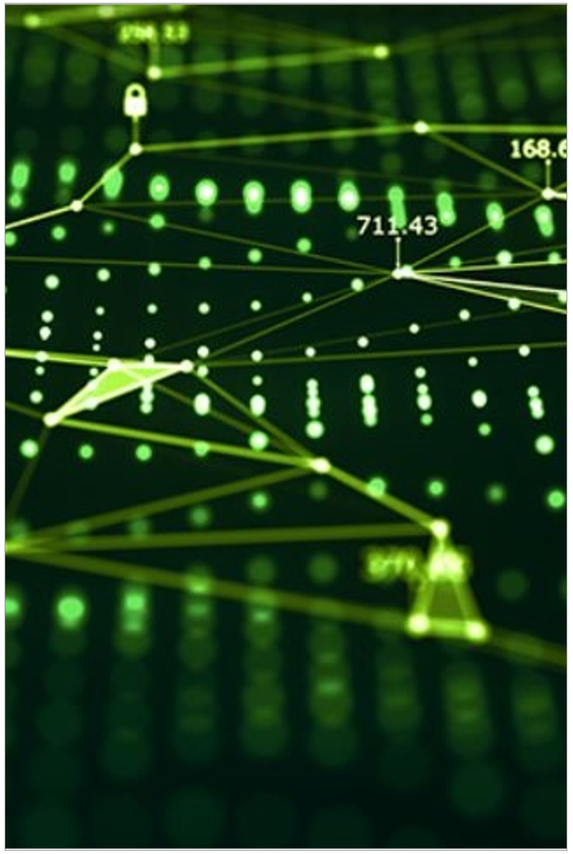
- New intent models used in ensemble with existing models in the original inference pipeline
- Results in 100% red team email detection



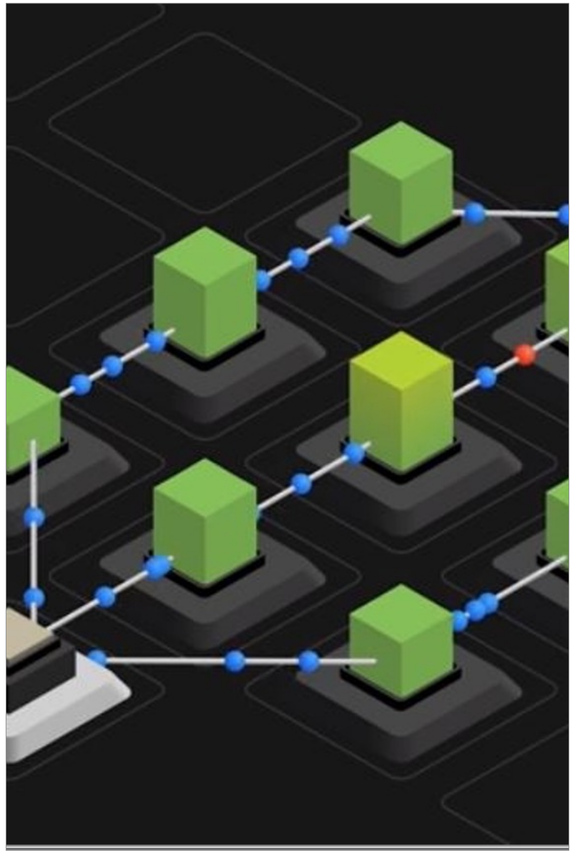
Now What

More Information and Code/Models to Get Started

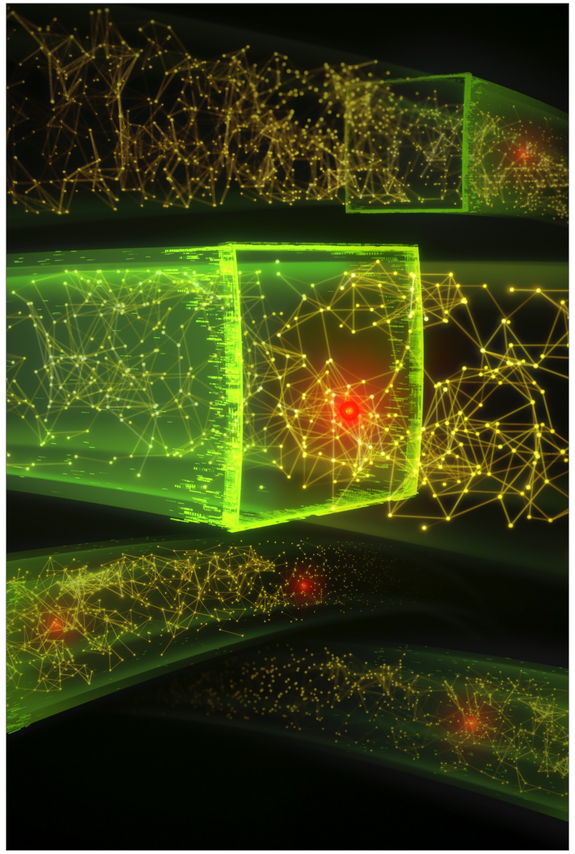
Workflows, use cases, and models available now



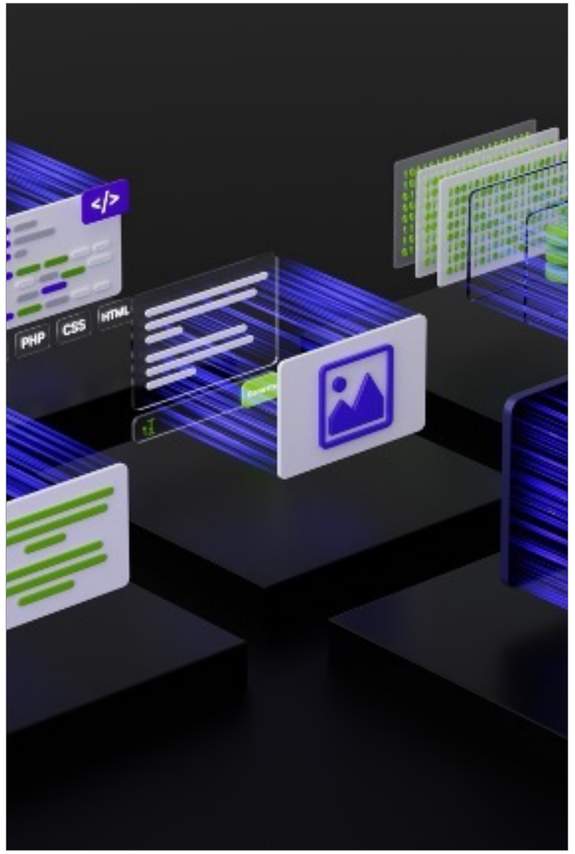
Spear Phishing Workflow
Available now on NVIDIA NGC



LLM Engine Integration
Coming in the November release of Morpheus



NVIDIA Morpheus
Available on NGC and GitHub, easy way to get started with accelerated compute for cyber



NVIDIA NeMo Framework
Easy way to get started with generative AI workloads and model training

Bartley Richardson, PhD
brichardson@nvidia.com

ENGINEERING AND PRODUCT TEAMS

Rachel Allen, PhD
Chris Arangio
Görkem Batmaz
Justin Boitano
Hsin Chin
Chau Dang
Shawn Davis
Michael Demoret
Eli Fajardo
David Gardner
Chris Harris
Ben Jarmak
Pete MacKinnon
Marco Murgia
Ryan Olson
Devin Robison
Nicola Sessions
Killian Sexsmith
Bhargav Suryadevara
Tad ZeMicheal, PhD

SECURITY ADVISORS

Rama Akkiraju
Travis Carroll
Sonu Nayyar
Jey Nivethan
David Reber, Jr.
Jason Recla
Daniel Rohrer
Gurpreet Singh
Pradeep Thalasta
Hemakamakshi Vivekanandan