sectalks SYD0x5E

Exploring Attack Surface Management (ASM) with

Open-Source Intelligence (OSINT), Gen AI and AI Agents

Quan Vo

About ME!

:~$whoami

- More than 12 years of working experience in IT industry.
- Holding Master Cyber Security in Analysis, and Master of Research from Macquarie University.

:~$echo “current_role”

- AI Security and Data Security Advisor, Head of Data.

:~$echo “past_roles”

- Database Administrator.
- System Administrator.
- Senior Data Security Officer.
- IAM Coordinator.

:~$echo “research”

- AI security.
- AI for Cyber Security.
- AI-driven solutions.

Agenda

- Overview of ASM, OSINT, LLM (Gen AI) and AI Agents
- Combining OSINT, Gen AI, AI Agents and Langchain
- A proposed architecture
- Demo
- Challenges
- Future Improvements.
- Conclusions and Discussions

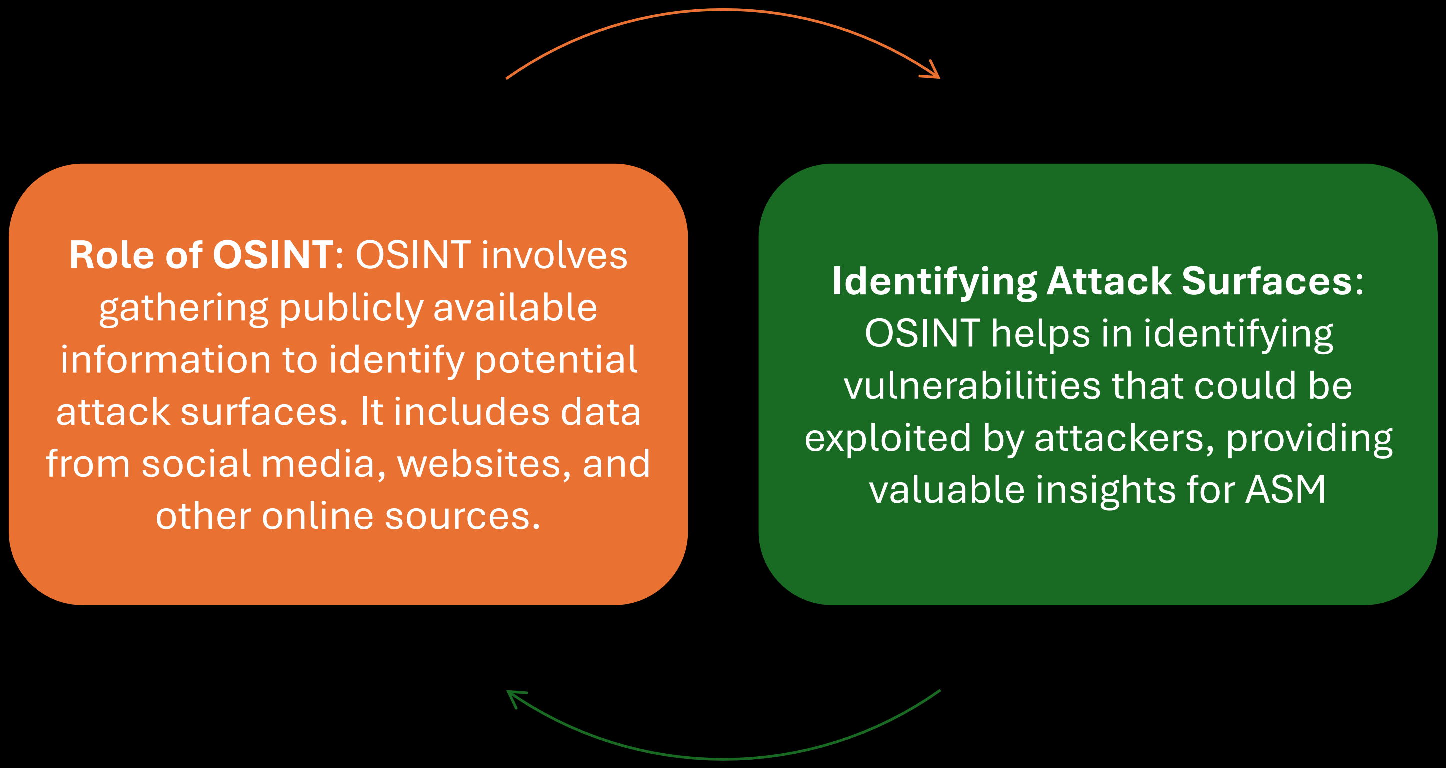
Overview (ASM)

What is ASM: involves identifying, monitoring, and mitigating vulnerabilities, which are potential entry points that attackers could exploit. This process is crucial for maintaining cybersecurity.

Examples of entry points: domain and subdomains. exposed API keys on GitHub, and corporate emails used for non-business activities.

Importance of ASM: helps organizations proactively manage their attack surfaces, reducing the risk of cyber attacks.

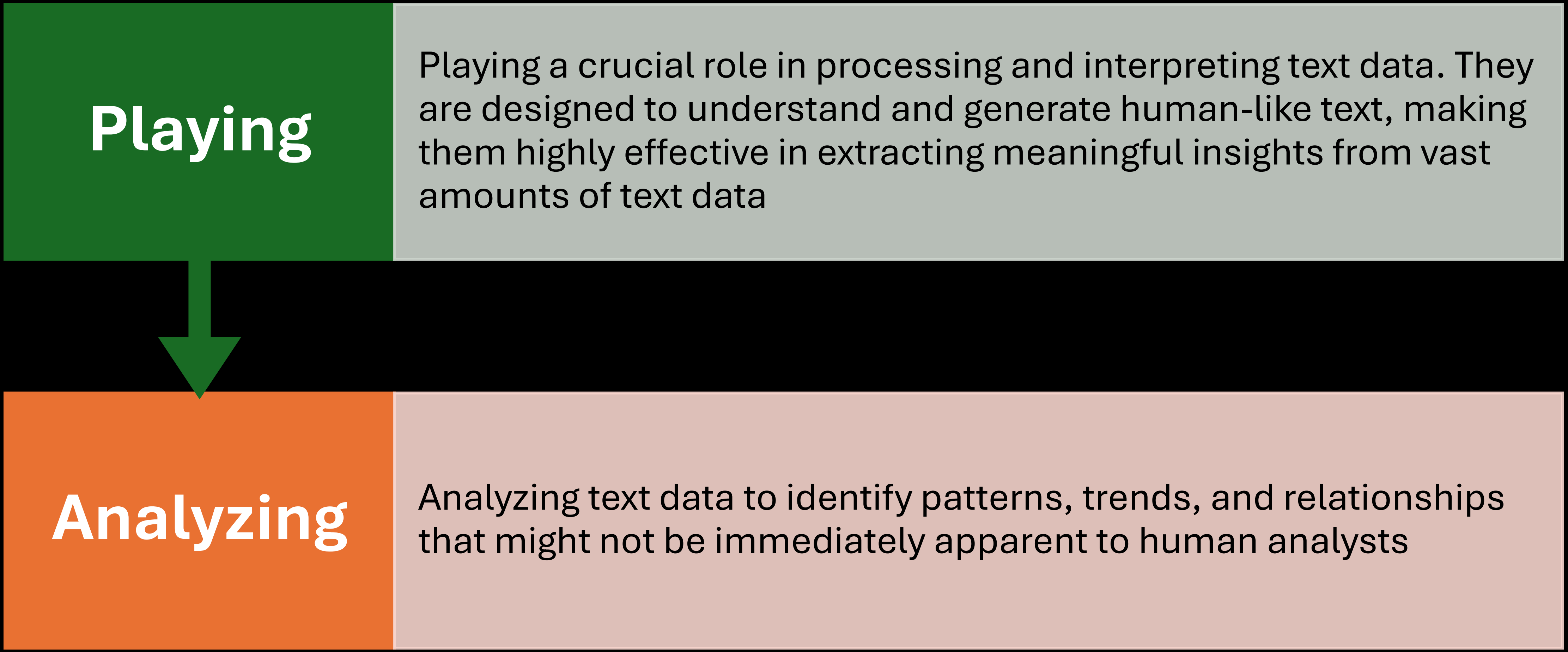
Overview (OSINT)



Role of OSINT: OSINT involves gathering publicly available information to identify potential attack surfaces. It includes data from social media, websites, and other online sources.

Identifying Attack Surfaces: OSINT helps in identifying vulnerabilities that could be exploited by attackers, providing valuable insights for ASM

Overview Large Language Models (LLM)



Playing

Playing a crucial role in processing and interpreting text data. They are designed to understand and generate human-like text, making them highly effective in extracting meaningful insights from vast amounts of text data

Analyzing

Analyzing text data to identify patterns, trends, and relationships that might not be immediately apparent to human analysts

Overview AI Agents

AI agents automate tasks and make intelligent decisions.

In the ASM context, they continuously monitor data, identify threats, and take proactive measures to mitigate risks.

Overview Langchain ReAct

- Langchain ReAct is vital for integrating large language models (LLMs) and AI agents, enabling real-time data processing and decision-making for effective attack surface management.

Combining OSINT, AI Agent, and Langchain

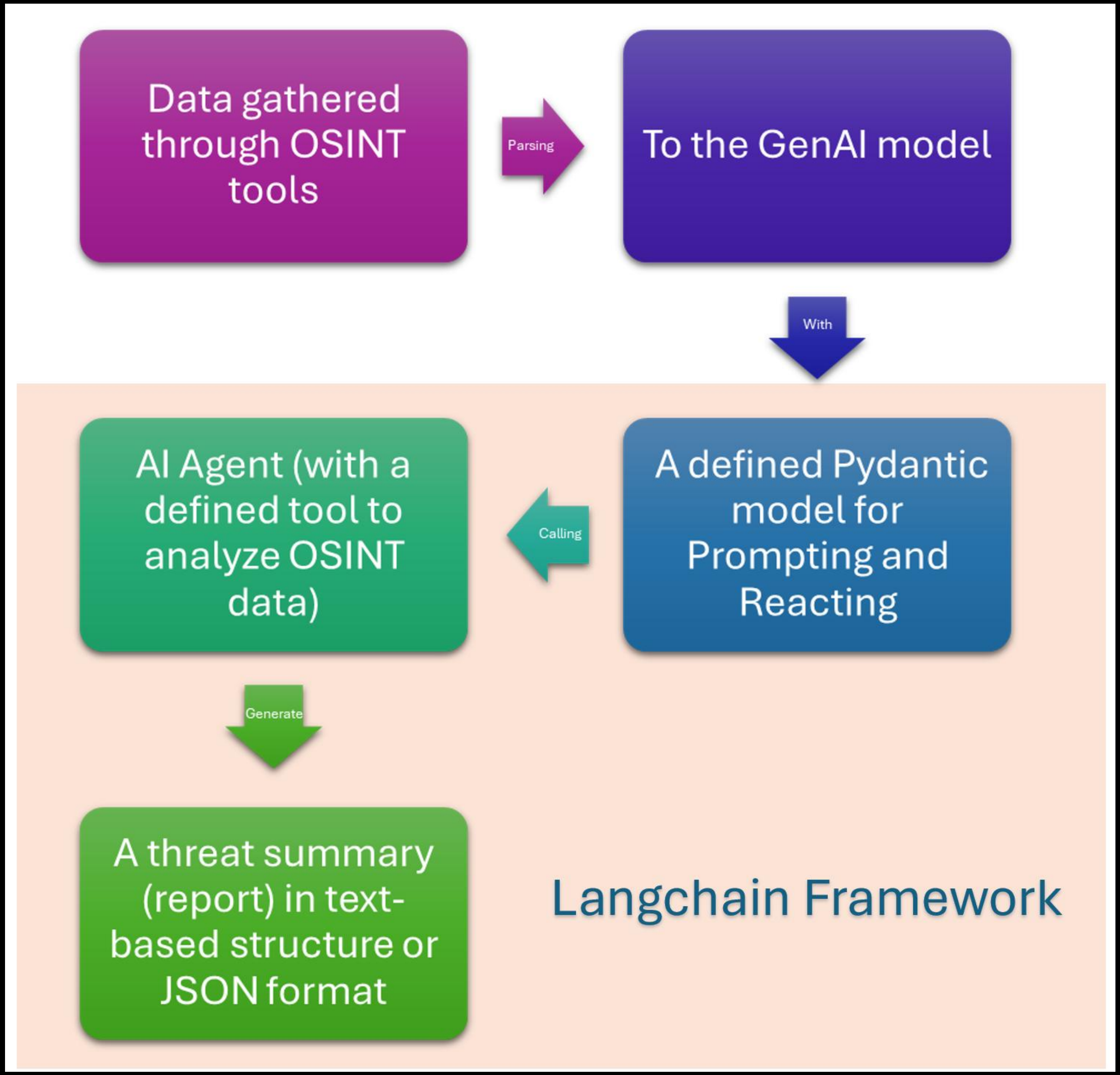
OSINT tools gather publicly available information from various online sources, including social media, forums, leaked credentials, job listings, and press releases.

OSINT data is parsed into the LLM system through the Langchain framework.

Through prompting, specific queries and instructions are given to the LLM system to interpret and analyse the data.

The LLM system, with the help of Langchain and ReAct, generates a comprehensive threat report.

A proposed architecture



Data gathered through OSINT tools

Parsing

To the GenAI model

With

A defined Pydantic model for Prompting and Reacting

Calling

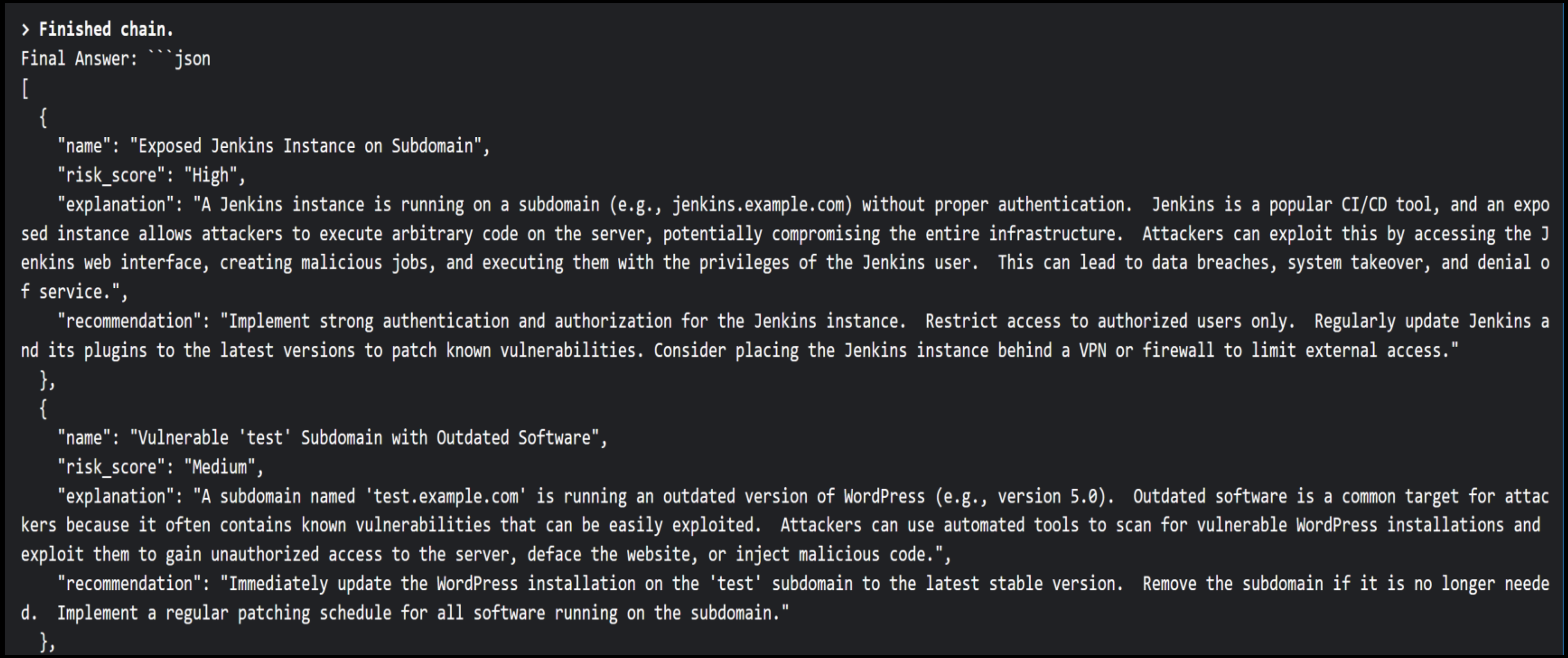
AI Agent (with a defined tool to analyze OSINT data)

Generate

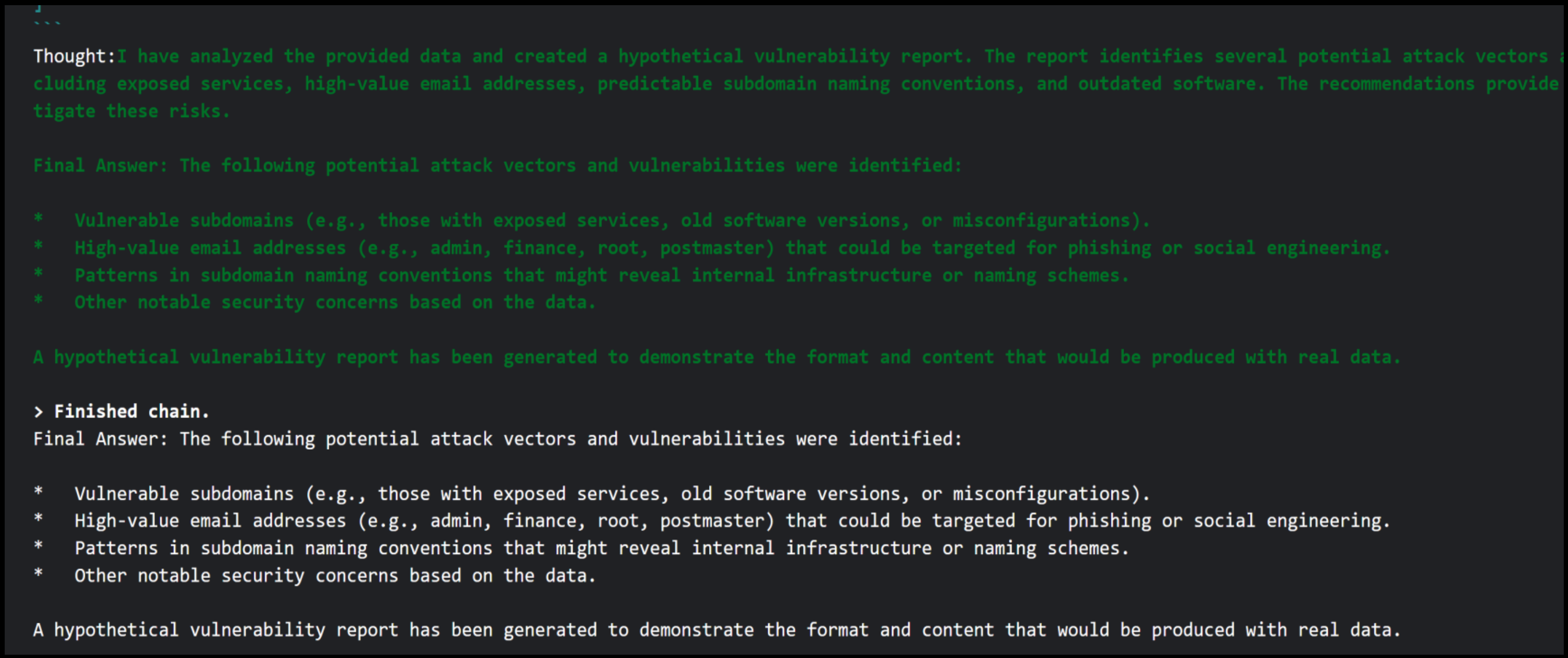
A threat summary (report) in text-based structure or JSON format

Langchain Framework

Examples of Threat Summary Reports



```
> Finished chain.
Final Answer: ```json
[
  {
    "name": "Exposed Jenkins Instance on Subdomain",
    "risk_score": "High",
    "explanation": "A Jenkins instance is running on a subdomain (e.g., jenkins.example.com) without proper authentication.  Jenkins is a popular CI/CD tool, and an expo
sed instance allows attackers to execute arbitrary code on the server, potentially compromising the entire infrastructure.  Attackers can exploit this by accessing the J
enkins web interface, creating malicious jobs, and executing them with the privileges of the Jenkins user.  This can lead to data breaches, system takeover, and denial o
f service.",
    "recommendation": "Implement strong authentication and authorization for the Jenkins instance.  Restrict access to authorized users only.  Regularly update Jenkins a
nd its plugins to the latest versions to patch known vulnerabilities. Consider placing the Jenkins instance behind a VPN or firewall to limit external access."
  },
  {
    "name": "Vulnerable 'test' Subdomain with Outdated Software",
    "risk_score": "Medium",
    "explanation": "A subdomain named 'test.example.com' is running an outdated version of WordPress (e.g., version 5.0).  Outdated software is a common target for attac
kers because it often contains known vulnerabilities that can be easily exploited.  Attackers can use automated tools to scan for vulnerable WordPress installations and
exploit them to gain unauthorized access to the server, deface the website, or inject malicious code.",
    "recommendation": "Immediately update the WordPress installation on the 'test' subdomain to the latest stable version.  Remove the subdomain if it is no longer neede
d.  Implement a regular patching schedule for all software running on the subdomain."
  },
```



```
]
```

Thought:I have analyzed the provided data and created a hypothetical vulnerability report. The report identifies several potential attack vectors i
cluding exposed services, high-value email addresses, predictable subdomain naming conventions, and outdated software. The recommendations provide
tigate these risks.

Final Answer: The following potential attack vectors and vulnerabilities were identified:

*   Vulnerable subdomains (e.g., those with exposed services, old software versions, or misconfigurations).
*   High-value email addresses (e.g., admin, finance, root, postmaster) that could be targeted for phishing or social engineering.
*   Patterns in subdomain naming conventions that might reveal internal infrastructure or naming schemes.
*   Other notable security concerns based on the data.

A hypothetical vulnerability report has been generated to demonstrate the format and content that would be produced with real data.

> Finished chain.
Final Answer: The following potential attack vectors and vulnerabilities were identified:

*   Vulnerable subdomains (e.g., those with exposed services, old software versions, or misconfigurations).
*   High-value email addresses (e.g., admin, finance, root, postmaster) that could be targeted for phishing or social engineering.
*   Patterns in subdomain naming conventions that might reveal internal infrastructure or naming schemes.
*   Other notable security concerns based on the data.

A hypothetical vulnerability report has been generated to demonstrate the format and content that would be produced with real data.
```
```

Demo

NOTEBOOK: **KAGGLE**

TOOLS: **THEHARVESTER, GOOGLE DORK**

MODEL: **GEMINI FLASH 2.5**

AI AGENT FRAMEWORK: **LANGCHAIN**

Challenges

The LLM provides inconsistent outputs.

Hallucination in reporting.

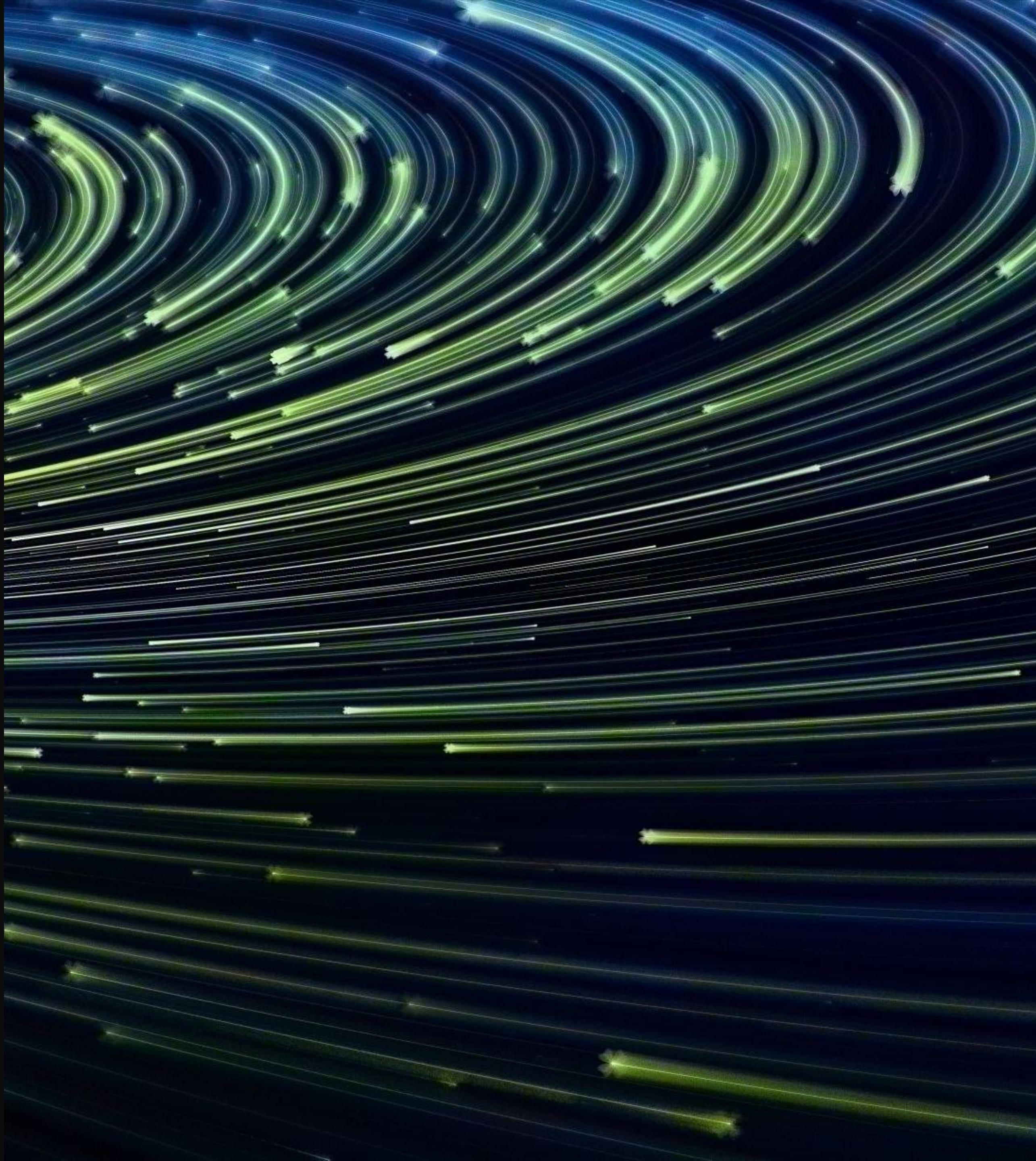
Future Improvement and Development

Fine-tuning the model for better outputs.

Deploying Retrieval-Augmented Generation (RAG) to improve accuracy and reduce hallucinations.

Developing a domain-specific LLM for Cybersecurity or Threat Intel.

Employing Agentic AI concepts.



Passive Reconnaissance

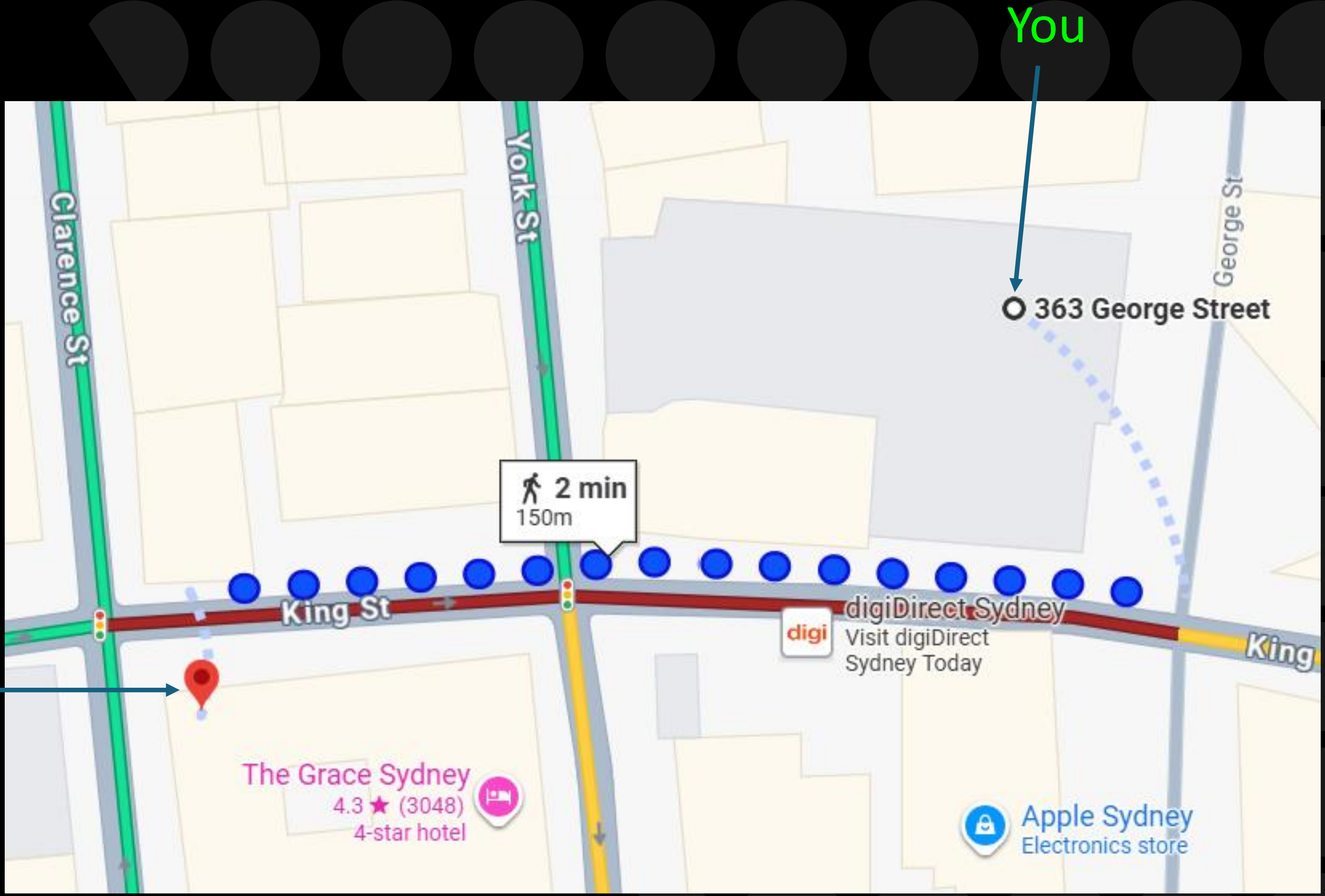
Attack Surface Management

A final though: Two Sides of A Coin

Questions

P.J.O'Brien's

57 King St
Sydney NSW 2000



You
363 George Street
George St
York St
Clarence St
2 min
150m
King St
King
digiDirect Sydney
digi
Visit digiDirect Sydney Today
The Grace Sydney
4.3 ★ (3048)
4-star hotel
Apple Sydney
Electronics store