

Prepared By:

Nur Inqsyira Binti Zamri,

Bachelor of Computer Science
(Hons.) in Cybersecurity

2. PROBLEM STATEMENT

- Inaccessible scam detection tools.
- Ignorance of usability despite high accuracy.
- Lack of scam link awareness.
- Hard to interpret detail threat data.
- Impractical cybersecurity practices.

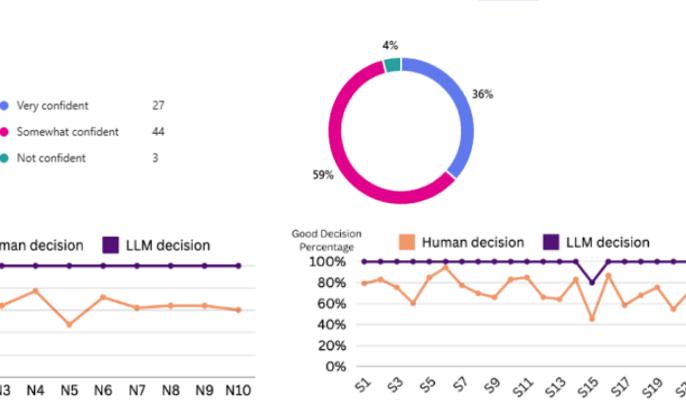
3. OBJECTIVE

- To explore how recent phishing detection research can be operationalized into accessible, real-time tools that support safe link practices in messaging platforms.
- To build an end-to-end scam link detection workflow that runs on open-source infrastructure with zero-cost requirements.
- To design an AI-powered mechanisms that converts technical threat data into user-friendly summaries.
- To evaluate the system's practical feasibility and responsiveness under real-world usage conditions.

4. LITERATURE REVIEW

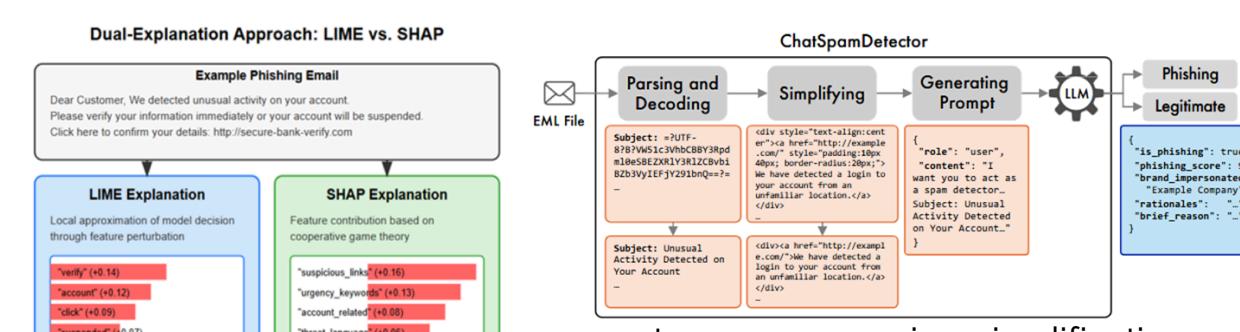
PHISHING DETECTION

- ML models (DNN, LSTM) boost accuracy but lack interpretability for end users.
- Chang and Aimeur (2024), LLMs (e.g. LLaMA-3, Gemma-2) excel in detecting nuanced, multilingual scams in chat environments; LLMs outperform humans in scam link recognition.

7. How confident are you in identifying scam links before clicking?


HUMAN-CENTERED EXPLAINABLE AI APPROACHES

- Users ignore alerts they don't understand—even when accurate (Wei et al., 2023).
- HITL offer contextual feedback, improve engagement and trust (Macke et al., 2023).
- Aljaraah et al. (2025) introduced EXPLICATE; translate model behavior into natural language explanations using SHAP/LIME + LLMs.



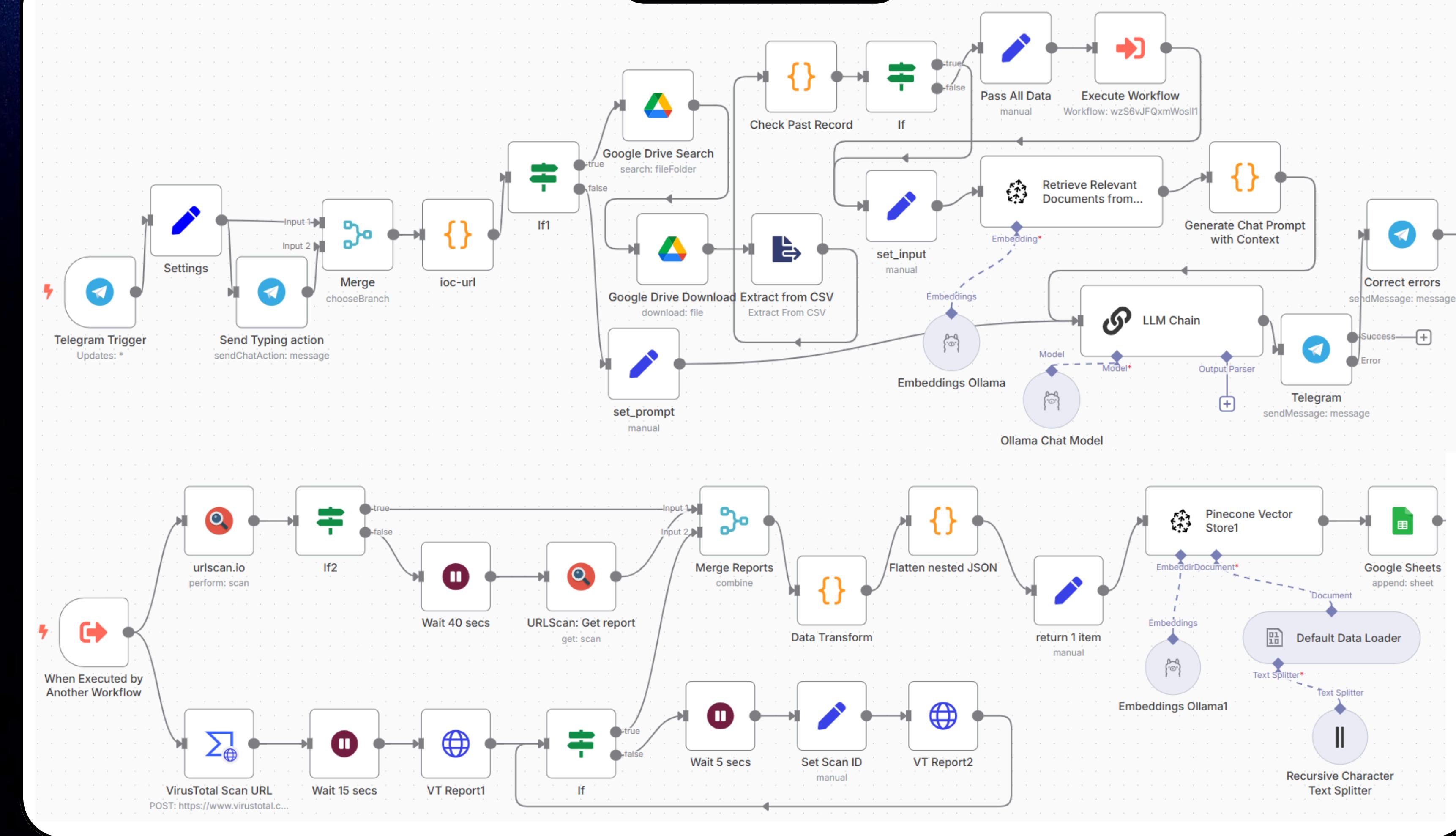
RAG FOR CYBERSECURITY

- Kurniawan et al. (2024) proposed CyKG-RAG; integrate structured knowledge graphs into the retrieval process; improve traceability and performance tuning.
- Kenneweg et al. (2024) introduced RAGVAL; an automatic benchmarking framework for accessing the quality of RAG.

1. ABSTRACT

Scam links are infiltrating platforms like Telegram with growing sophistication. InQueries is a Retrieval-Augmented Generation (RAG) chatbot that not only detects suspicious URLs, but also explains why—in simple language. Built on a fully zero-cost stack (n8n, Ollama, Pinecone), it scans URLs via VirusTotal and urlscan.io, classifies risk levels using custom heuristics, and retrieves relevant metadata for real-time explanation. Results show <25s latency for cached links and accurate threat similarity at cosine scores >0.70. This project proves explainable cybersecurity can be modular, low-cost, and publicly usable.

5. DESIGN



6. IMPLEMENTATION



This project supports the UN SDGs



7. TESTING

The screenshot shows the inQueries bot interface on a mobile device. The URL is https://koshiki-roguinamzn.mythic.com/L5sMnl. The interface displays a message: "High Risk with 100% confidence - 4 malicious (including trusted engines)". Below it, a warning message says: "Bad news, Inqsyira. After analyzing the URL using security engines, it appears to be a scam with 100% confidence. Please avoid clicking it—it could lead to credential theft, malware, or phishing traps." A Detection Summary table shows: Malicious: 4, Suspicious: 0, Harmless: 58, Undetected: 31. A note states: "What's concerning is that trusted engines like Fortinet and Sophos flagged it as malicious. That's a strong indicator of a real threat." A message from the user asks: "what is phishing traps?". The bot responds: "Imagine someone sends you an email that looks like it's from a bank or a friend, asking for some info. But, the email is actually fake and meant to steal your secrets! In this case, clicking on the link (the 'trap') would lead you into giving away sensitive information, which can put you at risk of identity theft, financial loss, or other problems." The timestamp is 12:02 AM.

Component	Specification
OS	Windows 11 64-bit
Processor	Intel Core i5-12450H (10th Gen, 8-core)
RAM	32 GB DDR4
GPU	NVIDIA GeForce RTX 3050 Laptop GPU (4 GB VRAM), used via gpu-nvidia Docker profile
LLM Models Tested	llama3.2:latest (via local Ollama), models/gemini-1.5-flash (cloud-based)
Embedding Model	mxbaie-embed-large:latest (1024-dimensional vectors)
Vector Store	Pinecone (free-tier, cloud-hosted vector database)
Automation Tool	n8n (running inside a Docker container)
API Services	VirusTotal, urlscan.io (free-tier access)
Internet Speed	~100 Mbps download / 8 Mbps upload, with ~30 ms ping, measured using Speedtest.net.

1. Avg. latency:

Cached URLs: 17.74s
New URLs: 63.3s

Text-only queries: 7.72s

2. Semantic Similarity Scoring:

Cosine similarity >0.70 yielded accurate matches in RAG retrieval.

8. CONCLUSION

Phishing detection can be made not only accurate but also accessible and educational. InQueries is a proof-of-concept project that demonstrates how cutting-edge research can be operationalized into tools that protect the public without sacrificing usability / interpretability.