Cyber Secure: Your Guide to Staying Safe in the Digital World

Anshul Anilkumar Mundakatil
San Jose State University
017416334

I. Introduction

In today's digital world, cybersecurity is of utmost importance. Public individuals and organizations face threats from malicious software, hackers, and viruses, making it critical to adopt safe online practices and remain vigilant against cyberattacks. To address these challenges, I propose **CyberSecure**, a virtual assistant tailored to raise cybersecurity awareness and provide actionable best practices.

This virtual assistant is designed for two primary user groups:

Public Users: Individuals seeking information on how to avoid scams, protect their personal information, and stay safe in the digital space.

Organizational Users: Employees and IT teams from organizations seeking guidance on internal security protocols, safe handling of digital assets, and identification of social engineering attacks.

By leveraging a Retrieval-Augmented Generation (RAG) pipeline, the virtual assistant will process curated cybersecurity documents to provide precise, context-driven responses. Integrating advanced large language models such as Gemma, Llama, and Qwen, and employing comprehensive evaluation metrics, CyberSecure aims to empower users with the knowledge to navigate cyberspace safely and confidently.

This proposal outlines the use cases, document datasets supporting the RAG application, the chosen LLM models, and the project's implementation and evaluation metrics.

II. Data

The documents in the corpus were selected based on their content. A wide range of cybersecurity topics was chosen, including malware, viruses, social engineering, organizational cybersecurity policies, and mitigation strategies.

Document List

- **An Introduction to Malware** by Robin Sharp
  - The document focuses on explaining various types of malware, including viruses, worms, and trojans and ways to mitigate them.
- **Cybersecurity Handbook** by ROF Network
  - A comprehensive guide covering best practices for individuals and organizations.
  - Includes tips on securing devices, data encryption, and safe internet browsing.
- **Cybersecurity for Small Business: Cybersecurity Basics**
  - A resource for small businesses to strengthen their security.
  - Addresses risks such as phishing attacks, data breaches, and employee training in small businesses.
- **Basic Cyber Security: A Guide for All to Manage Digital Security**
  - A foundational document offering practical steps for managing digital safety in everyday scenarios.
  - Covers password management, software updates, and securing online communications.

Document Properties

- **Corpus Size**: The dataset comprises over **34,896 words**, ensuring sufficient data to support at least **30 unique queries**.
- **Format**: All documents are provided in **PDF format**.

III. Architecture Diagram / Methodology

The architecture diagram provides a workflow of the Retrieval-Augmented Generation (RAG) system. Below is a detailed explanation of each module and its function:

Components of the System

- **AI Application**
  - User inputs their query.
  - The application sends the user's query to the system and receives the *model output*.
  - Acts as the interface for interaction with the RAG pipeline.
- **LLM (Large Language Model)**
  - Uses a combination of the query and the retrieved data to generate a response.
  - Ensures context-awareness by integrating relevant information from the retrieval module.
- **Chroma (Retrieval System)**
  - Responsible for fetching relevant information from a knowledge base or corpus.
  - *Retrieval*: Identifies and extracts the most pertinent data related to the user's query.
  - *Retrieved Data*: Provides structured and filtered data to the LLM for processing.

Workflow

1. The **AI Application** collects the user's query and forwards it to the **LLM**.
2. The **LLM** augments the query by interacting with the **Chroma** retrieval system.
3. **Chroma** retrieves relevant data from its corpus based on the input query.
4. The retrieved data is sent back to the **LLM**, where it is integrated with the original query.
5. The **LLM** processes the combined information and generates a contextually relevant response.
6. The final response is returned to the **AI Application** for user delivery.

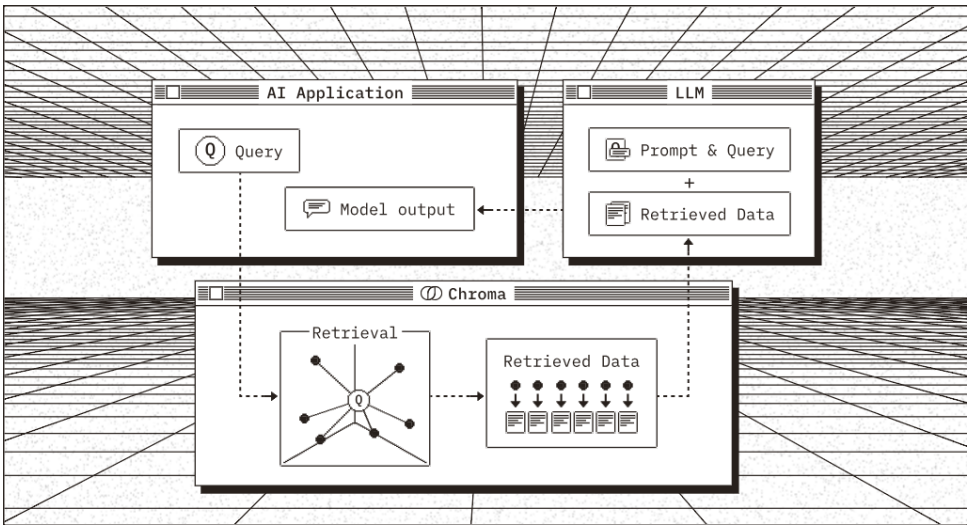

Figure 1: Architecture Diagram

IV. LLM Models Used

The Virtual Assistant (VA) for cybersecurity best practices and awareness utilizes the following Large Language Models (LLMs):

- **Gemma-2-2B-IT**
- **Qwen 2.5 1.5 B-Instruct**
- **Llama-3.2-3B-Instruct**

These models were selected based on their ability to handle a wide range of queries related to cybersecurity, providing accurate, context-aware responses in real-time.

V. Results

The performance of the RAG models (Gemma, Llama, and Qwen) was evaluated across three criteria: accuracy, groundedness, and speed. A total of 30 questions were posed to each model, and after receiving each response, users rated the accuracy on a scale of 1–5. Additionally, groundedness was calculated by measuring the cosine similarity between the model's response and the source content, while the response speed was measured in seconds.

Model	Average Accuracy	Average Groundedness	Average Speed (seconds)
Gemma	75.15	71.37	11.27
Llama	73.75	71.30	11.22
Qwen	86.67	76.65	12.11

Table 1: Performance of RAG models in terms of accuracy, groundedness, and speed.

Accuracy: The Qwen model demonstrated the highest average accuracy, scoring 86.67%. Gemma followed with an average accuracy of 75.15%, and Llama had an average of 73.75%.

Groundedness: In terms of groundedness, Qwen again showed the strongest performance, achieving an average of 76.65%. Gemma's groundedness score was 71.37%, and Llama scored slightly lower at 71.30%.

Speed: In terms of speed, Gemma was the fastest model with an average response time of 11.27 seconds. Llama recorded a response time of 11.22 seconds, while Qwen took 12.11 seconds on average. Overall, the Qwen model stood out for its higher accuracy and groundedness, but Gemma was the fastest in terms of response time.

VI. Conclusion

The aim of this project was to develop a robust cybersecurity virtual assistant that would help individuals as well as organizational users stay safe in the cyber world. The document corpus covered a wide range of topics in the field of cybersecurity, including hacking, phishing, viruses, best practices, and password protection. We evaluated three RAG models: Gemma, Llama, and Qwen across three metrics: accuracy, speed, and groundedness. The accuracy is based on user experience, while groundedness is measured using cosine similarity between the retrieved documents and the model's output. Speed is measured as the time the model takes to generate a response.

While all the models performed well, Qwen emerged as a standout. It achieved an impressive accuracy of 86.67% and a groundedness score of 76.65%, outperforming both Gemma and Llama. Although Gemma was not as strong in terms of accuracy and groundedness, it was the fastest model, which can be beneficial in applications requiring quick responses. Llama also performed admirably but lagged slightly behind Qwen in terms of accuracy and groundedness.

References

1. Sharp, R. (2017). *An introduction to malware.*

2. Resilience of Network. (2021). *Hellenic Republic Ministry of Digital Governance National Cybersecurity Authority.*

3. Federal Trade Commission. (n.d.). *Cybersecurity for small business: Fact sheets.* Retrieved from `https://www.ftc.gov/system/files/attachments/cybersecurity-small-business/cybersecuirty_sb_factsheets_all.pdf`

4. Global Interagency Security Forum. (2019). *Digital security guidelines.* Retrieved from `https://gisf.ngo/wp-content/uploads/2021/10/ACT_Digital_Security_Guidelines_2019.pdf`

5. Řehulka, Erik, and Marek Šuppa. "RAG Meets Detox: Enhancing Text Detoxification Using Open Large Language Models with Retrieval Augmented Generation." (2024).