Isain Cortes, Jr.

4/18/2025

CYSE368

Reflection #3

Third 50 Hours Cybersecurity Internship Reflection

During the final 50 hours of my internship with CAE in Cybersecurity Community, contracting through Augusta University, our team successfully completed the transition from Google Gemini to Ollama in both the frontend and backend of the inherited codebase. After extensive testing, we settled on using the Llama3 model due to its balance between performance and efficiency, ideal for devices with moderate RAM. Using multithreading, we enabled our system to extract text from PDF's located in the current working directory, send that data to the LLM in chunks, and convert it into multiple choice questions.

In the last phase of our project, we had the opportunity to present our findings to Augusta's cyber department, as well as representatives from Booz Allen Hamilton and the Department of Defense. Below, my contributions focused on backend optimization, model selection, and security assessments. One of the challenges we faced was the performance limits of certain models which led me to suggest testing Deepseek, an alternative LLM. While Deepseek is based in China and poses potential privacy concerns when run through browser, I downloaded and tested the model locally through Ollama verifying network activity via netstat to ensure no external traffic. Performance-wise, Deepseek outpaced Llama3 in response time and token handling, offering more comprehensive answers with minimal lag.

This experience reinforced the importance of balancing efficiency and security in AI integration, especially in cybersecurity. It was rewarding to present meaningful work to professionals in the field and walk away with stronger technical, research, and presentation skills.

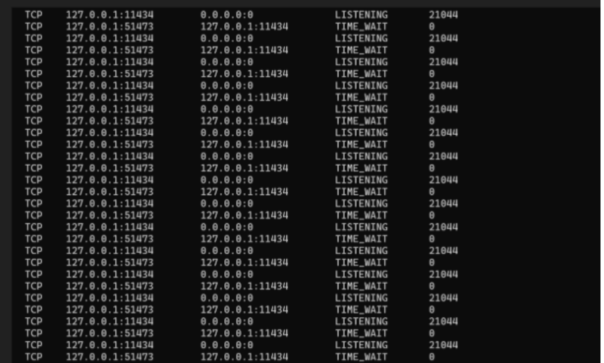
Isain - Deepseek Security Testing & Comparisons

Setup & Tools Used:

- DeepSeek running locally via Ollama & Docker
- Network monitoring with netstat -ano to track connections

Key Findings:

- No external connections detected
- Only localhost (127.0.0.1) traffic visible
- No communication with external servers or unknown IPs
- Secure operation when run through Docker



Isain - Deepseek Security Testing & Comparisons

DeepSeek Model:

- Faster response times despite handling more tokens.
- Efficient optimization allowing quicker processing of more complex tasks.
- More tokens used = detailed, comprehensive responses without significant lag.
- Optimized for speed with complex outputs.

Llama Models:

- Slower response times with fewer tokens.
- Less efficient when handling complex tasks in the same timeframe.