CyberScienceLab

University of Guelph

Canada Cyber Foundry

Developing a RAG to Mitigate LLM Hallucination in CVEs

Introduction

This project implements a Retrieval-Augmented Generation (RAG) model aimed at validating the use of Common Vulnerabilities and Exposures (CVEs) in security reports. By leveraging Meta's Llama3 and retrieving correct CVE descriptions, the system ensures that CVEs are accurately referenced and contextually applied within threat intelligence reports.

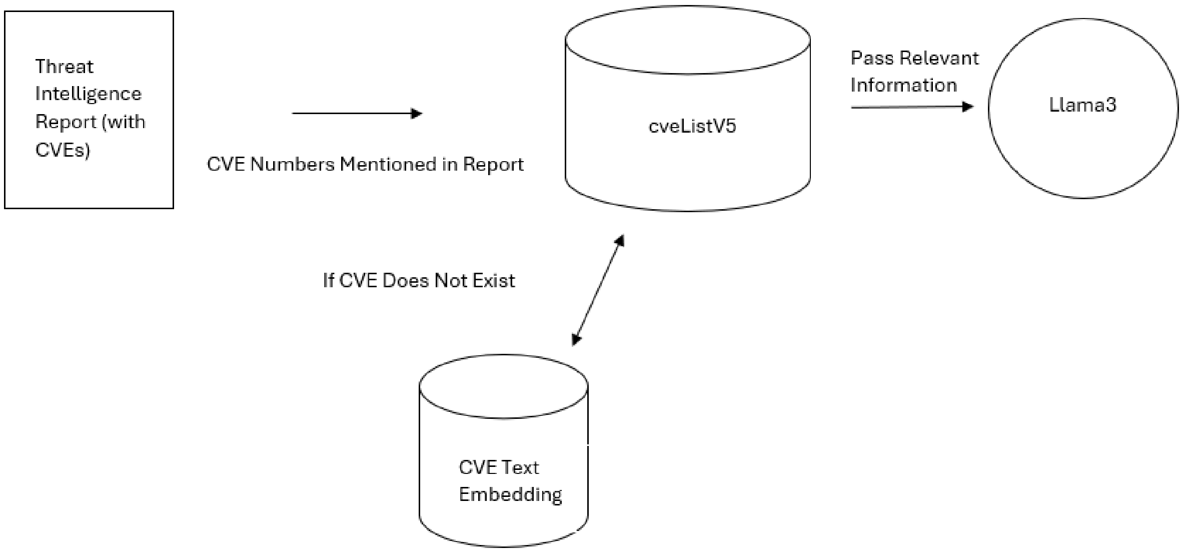

Threat Intelligence Report (with CVEs)
CVE Numbers Mentioned in Report
cveListV5
Pass Relevant Information
Llama3
If CVE Does Not Exist
CVE Text Embedding

Diagram of Retrieval Augmented Generation with CVEs

After processing a threat intelligence report, Llama3 will identify and extract all mentioned CVEs. These CVEs will then be cross-referenced with the cvelistV5 database to retrieve detailed information. If a CVE is not found—whether due to an outdated database or an incorrect entry—Llama3 will pass the unknown CVE into a vector embedding containing textual representations of known CVEs. This embedding will help suggest the most similar existing CVE as a replacement. Once all relevant information is gathered, Llama3 will analyze the use of the CVEs based on the data provided.

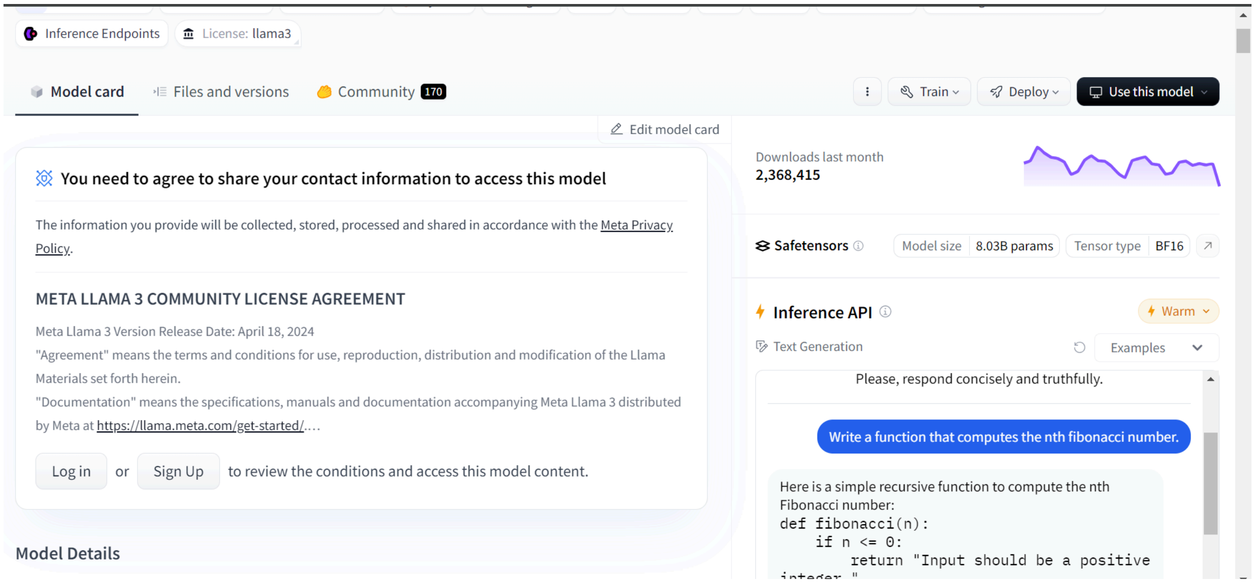
Package Requirements:

Install the following using pip/conda:

- Python Version 3.12
- Accelerate
- Bitsandbytes
- Langchain
- Sentence-transformers
- Transformers
- Tqdm
- pytorch
- tokenizers
- torchaudio
- Torchvision
- Huggingface_hub
- Pandas

Access Llama3 HuggingFace Token:

[HuggingFace Website]



You need to agree to share your contact information to access this model

The information you provide will be collected, stored, processed and shared in accordance with the Meta Privacy Policy.

META LLAMA 3 COMMUNITY LICENSE AGREEMENT

Meta Llama 3 Version Release Date: April 18, 2024

"Agreement" means the terms and conditions for use, reproduction, distribution and modification of the Llama Materials set forth herein.

"Documentation" means the specifications, manuals and documentation accompanying Meta Llama 3 distributed by Meta at https://llama.meta.com/get-started/.…

Log in or Sign Up to review the conditions and access this model content.

Model Details

Please, respond concisely and truthfully.

Write a function that computes the nth fibonacci number.

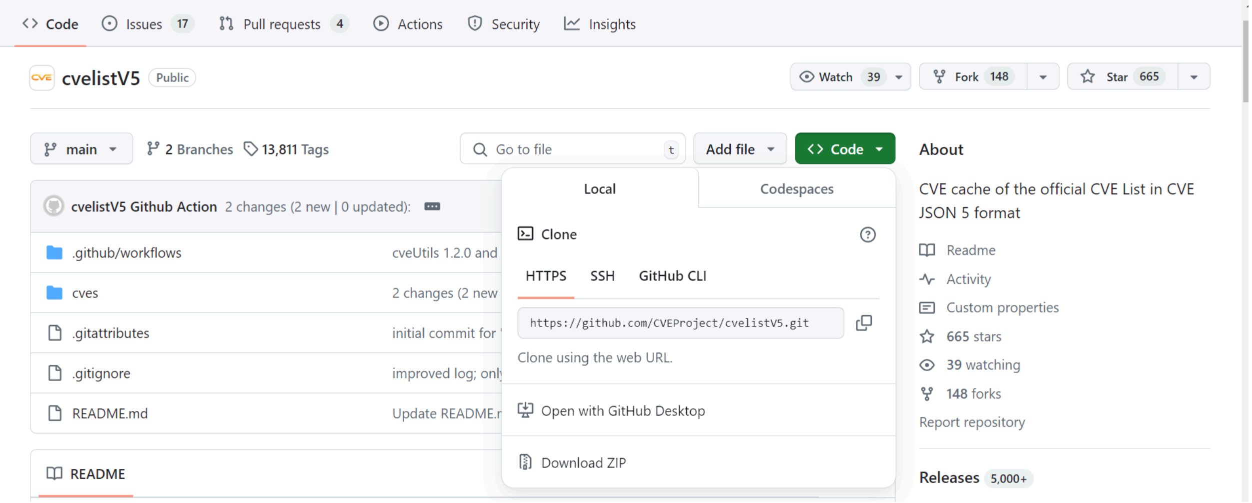
Here is a simple recursive function to compute the nth Fibonacci number:

```
def fibonacci(n):
    if n <= 0:
        return "Input should be a positive
integer."
```

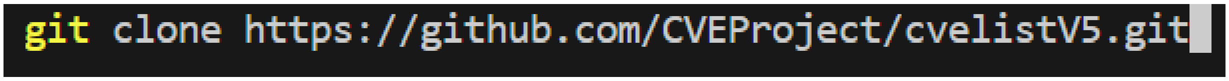
1. Create/Log into your Huggingface account
2. Get access to Llama3's Token
3. Export the token in your python environment by typing the command: export HF_TOKEN=<YOUR TOKEN>

Running The Program with the Provided Vector Embedding

1. Clone the CVE JSON Dataset from the CVElist GitHub.

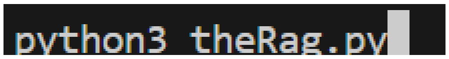


```
https://github.com/CVEProject/cvelistV5.git
```



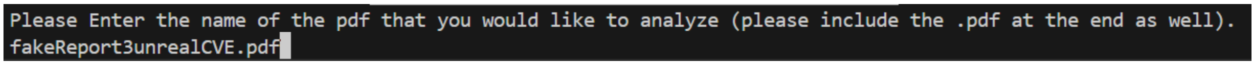
```
git clone https://github.com/CVEProject/cvelistV5.git
```

2. Run the Program: python3 theRag.py



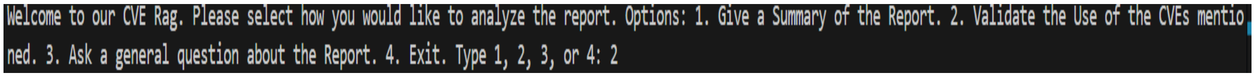
```
python3 theRag.py
```

3. When prompted, enter the report name (including .pdf).



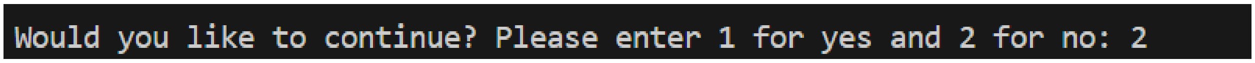
```
Please Enter the name of the pdf that you would like to analyze (please include the .pdf at the end as well).
fakeReport3unrealCVE.pdf
```

4. Main Menu: Type “2” to analyze the report



```
Welcome to our CVE Rag. Please select how you would like to analyze the report. Options: 1. Give a Summary of the Report. 2. Validate the Use of the CVEs mentio
ned. 3. Ask a general question about the Report. 4. Exit. Type 1, 2, 3, or 4: 2
```

5. After Llama3 provides the output, you can continue or exit: Type "1" to continue or "2" to exit.



```
Would you like to continue? Please enter 1 for yes and 2 for no: 2
```

Output

Correct Report:

Here are the verifications for the CVEs mentioned in the report:

1. CVE-2024-3400:
    * Correct CVE Description: A command injection as a result of arbitrary file creation vulnerability in the GlobalProtect feature of Palo Alto Networks PAN-OS software for specific PAN-OS versions and distinct feature configurations may enable an unauthenticated attacker to execute arbitrary code with root privileges on the firewall.
    * Report Excerpt: "On April 10, 2024, Volexity identified a critical zero-day exploitation in the GlobalProtect feature of Palo Alto Networks PAN-OS, tracked as CVE-2024-3400, leading to unauthorized remote code execution."
    * Verification: Correct Usage. The report accurately describes the vulnerability and its impact, and provides mitigation recommendations.

2. No other CVEs are mentioned in the report.

Note that the report does not mention any other CVEs besides CVE-2024-3400. Therefore, there is no need to verify any other CVEs.

Fake Report:

Here are the verifications for each CVE mentioned in the report:

1. **CVE-2024-100000000: Buffer Overflow in XYZ Web Server**

* Correct CVE Description: Not found (this CVE does not exist)
* Report Excerpt: "A buffer overflow vulnerability has been identified in XYZ Web Server version 3.4.1, which allows remote attackers to execute arbitrary code on the affected server."
* Verification: Incorrect Usage
* Explanation: The report mentions a fictional CVE (CVE-2024-100000000) and describes a buffer overflow vulnerability in XYZ Web Server, which does not match any of the correct CVE descriptions provided.

2. **CVE-2024-0008: SQL Injection in ABC Database Management System**

* Correct CVE Description: CVE-2024-0008, Vendor: Palo Alto Networks, Product: PAN-OS, Description: Web sessions in the management interface in Palo Alto Networks PAN-OS software do not expire in certain situations, making it susceptible to unauthorized access.
* Report Excerpt: "An SQL injection vulnerability exists in ABC Database Management System version 2.8.0, which allows attackers to execute arbitrary SQL commands."
* Verification: Incorrect Usage
* Explanation: The report describes an SQL injection vulnerability in ABC Database Management System, which does not match the correct CVE description of a web session expiration vulnerability in Palo Alto Networks PAN-OS software.

3. **CVE-2024-0009: Remote Code Execution in GHI Software**

* Correct CVE Description: CVE-2024-0009, Vendor: Palo Alto Networks, Product: PAN-OS, Description: An improper verification vulnerability in the GlobalProtect gateway feature of Palo Alto Networks PAN-OS software enables a malicious user with stolen credentials to establish a VPN connection from an unauthorized IP address.
* Report Excerpt: "A remote code execution vulnerability exists in GHI Software version 4.2.1, allowing attackers to execute arbitrary commands on the affected system."
* Verification: Incorrect Usage
* Explanation: The report describes a remote code execution vulnerability in GHI Software, which does not match the correct CVE description of an improper verification vulnerability in Palo Alto Networks PAN-OS software.

4. **CVE-2024-0010: Denial of Service in JKL Network Appliance**

* Correct CVE Description: CVE-2024-0010, Vendor: Palo Alto Networks, Product: PAN-OS, Description: A reflected cross-site scripting (XSS) vulnerability in the GlobalProtect portal feature of Palo Alto Networks PAN-OS software enables execution of malicious JavaScript (in the context of a user's browser) if a user clicks on a malicious link, allowing phishing attacks that could lead to credential theft.
* Report Excerpt: "A denial of service vulnerability has been identified in JKL Network Appliance version 5.6.3, which can be exploited to crash the system or degrade its performance."
* Verification: Incorrect Usage
* Explanation: The report describes a denial of service vulnerability in JKL Network Appliance, which does not match the correct CVE description of a reflected cross-site scripting (XSS) vulnerability in Palo Alto Networks PAN-OS software.

In summary, none of the CVEs mentioned in the report are used correctly, as they do not match the correct CVE descriptions provided.
Based on the relevant information provided, I recommend the CVE that most closely resembles the chunk of text as CVE-2024-29506. Here's why:

* The description of CVE-2024-29506 mentions a stack-based buffer overflow in the pdfi_apply_filter() function via a long PDF filter name, which allows an attacker to potentially execute arbitrary code.
* The vulnerability affects Artifex Ghostscript before 10.03.0, which is a software that processes PDF files.
* The description does not mention remote exploitation, but it does not rule it out either. The vulnerability could potentially be exploited remotely if an attacker can trick a user into opening a malicious PDF file.
* The mitigation involves upgrading to a newer version of Ghostscript, which is similar to the mitigation suggested for CVE-2024-1000000, which involves upgrading to XYZ Web Server version 3.4.2.

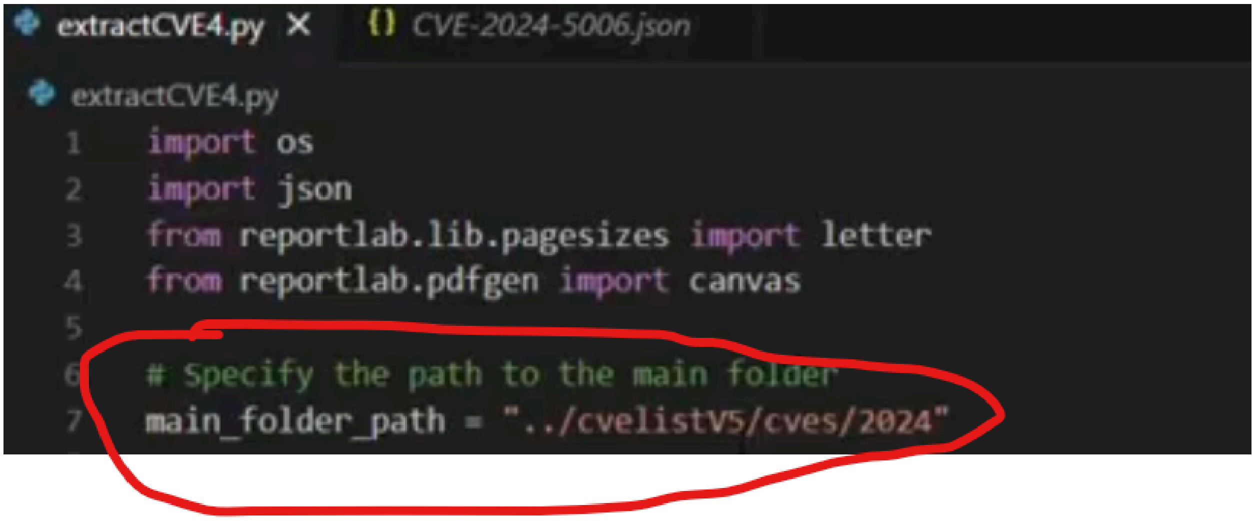
While there are some differences between the two vulnerabilities, CVE-2024-29506 is the most similar based on the information provided.

Updated The CVE Text Embedding (Optional):

The following embedding is conducted using the Sentence Transformer Library, all-mpnet-base-v2

1. Open extractCVE4.py

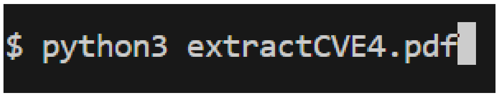
2. Update file path if needed to access the CVE github



```python
import os
import json
from reportlab.lib.pagesizes import letter
from reportlab.pdfgen import canvas

# Specify the path to the main folder
main_folder_path = "../cvelistV5/cves/2024"
```

3. In the terminal, type “python3 extractCVE4.py”

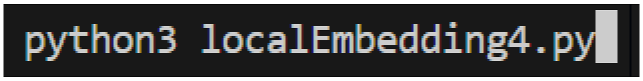


```
$ python3 extractCVE4.pdf
```

4. Export the text file into a PDF using a third-party resource (e.g., Word).

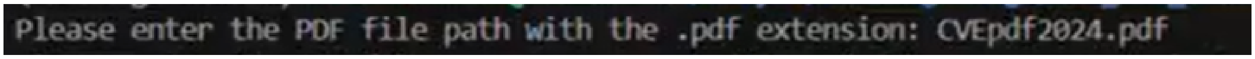
5. Export the new pdf into the main repository

6. Embed the PDF using the command: “python3 localEmbedding4.py”



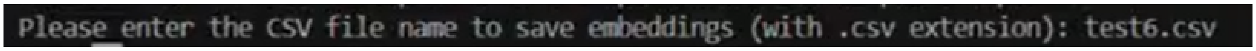
```
python3 localEmbedding4.py
```

7. You will be prompted for the pdf name, provide it include the .pdf



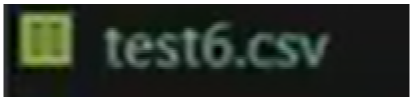
```
Please enter the PDF file path with the .pdf extension: CVEpdf2024.pdf
```

8. You will be asked to name the embedding file, name it with .csv



```
Please enter the CSV file name to save embeddings (with .csv extension): test6.csv
```

9. The embedding file has been created



test6.csv

UNIVERSITY of GUELPH

CANADA CYBER FOUNDRY

CYBERSCIENCELAB

Secure the future with us

cybersciencelab.com