I highly suggest going through README to get more insights. It was written to explain things more and put the project as clearly as possible

Also summary video only covers the questionnaire as it is quite slow now and took 4 minutes. I would suggest you skip some parts. For tools outputs are in the appendix.

Also to use any tools you need APIs or else you will get an error. API instructions for each tool in howToUse.txt

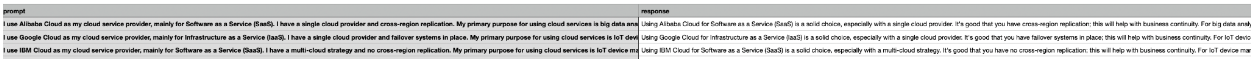
Project Name: Cybersecurity Toolkit and LLM-Powered Questionnaire

Avijit Prasad. Tutorial Group: F18B. Student Number: z5307473. Course: 6441

Description of the Project:
Project started as a web app idea. To provide a bunch of toolkits and a security questionnaire with a good UI to get people hands on experience on cybersecurity. Though, there were drastic changes in my questionnaire. I shifted more towards leveraging a language learning model (LLM) for the questionnaire. My idea is to let the user guide them with questions and generate a report with the technology we have in hand. With that I also went into a journey of compiling a dataset to finetune a cybersecurity bot, which will assist me in this task. All of this with a more technical toolkit, I thought it would be more viable to make a project with node servers.
Despite this shift, all JavaScript files were crafted with web integration in mind, featuring clear input structures and outputs in JSON and other web-friendly formats. This ensures a swift transition to a web app when server-side implementation commences.

Cybersecurity questionnaire -
Dataset: Compiled 1000 lines of cybersecurity prompts and responses using python scripts and external knowledge.
Prompts have various users like students, professionals etc. With fields like budget, cloud services used etc.
Responses has best practices, toolkits (free and paid), learning resources recommended

My dataset got a usability score of 5.35 in kaggle.



prompt	response
I use Alibaba Cloud as my cloud service provider, mainly for Software as a Service (SaaS). I have a single cloud provider and cross-region replication. My primary purpose for using cloud services is big data ana	Using Alibaba Cloud for Software as a Service (SaaS) is a solid choice, especially with a single cloud provider. It's good that you have cross-region replication; this will help with business continuity. For big data analy
I use Google Cloud as my cloud service provider, mainly for Infrastructure as a Service (IaaS). I have a single cloud provider and failover systems in place. My primary purpose for using cloud services is IoT devi	Using Google Cloud for Infrastructure as a Service (IaaS) is a solid choice, especially with a single cloud provider. It's good that you have failover systems in place; this will help with business continuity. For IoT devic
I use IBM Cloud as my cloud service provider, mainly for Software as a Service (SaaS). I have a multi-cloud strategy and no cross-region replication. My primary purpose for using cloud services is IoT device ma	Using IBM Cloud for Software as a Service (SaaS) is a solid choice, especially with a multi-cloud strategy. It's good that you have no cross-region replication; this will help with business continuity. For IoT device mar

Glimpse of dataset

Questionnaire: Series of generated questions through user responses that guides users through a journey to discover questions itself. Let them ask questions, type responses if they did not like the options and ask uptill 20 questions till they or LLM decide a report should be generated. More details in Readme and the video. I would suggest you read the initial_prompt.txt file in the questionnaire directory to get an insight into how the LLM is setting up.

Cybersecurity Toolkit Summaries
Password Brute Force Toolkit (beta version partially working)
Developed a toolkit that automates the process of password brute-forcing to test the strength of web application logins against common credentials.

Dependency Vulnerability Checker (beta version working)
Created a script to analyze project dependencies against known vulnerabilities, leveraging the GitHub Advisory Database to identify potential security risks.

Subdomain Discovery Tool (beta version working)
Implemented a DNS brute-forcing technique to discover subdomains, helping identify possible entry points for attackers or misconfigured services.

ZAP Active and Passive Scan Scripts (beta version working)
Integrated ZAP scanning tools into scripts for active and passive security scanning of web applications, through spider crawling to all endpoints, providing insights into potential vulnerabilities from headers to injection flaws.

Results:
The questionnaire successfully guides users through an interactive cybersecurity assessment, adapting questions based on user input. The toolkit comprises tools for password brute force checking, dependency vulnerability scanning, and subdomain discovery, each providing valuable insights into the security posture of a given web application.

What I Did:

- Developed a questionnaire, facilitating user interaction up to 20 questions.
- Diversified use of an LLM from a chatbot to a guided questionnaire structure. This is a completely original idea and the initial version is working perfectly. Incorporated the idea of ‘not everyone knows what to ask’. This and the dataset I generated were the hardest but most rewarding parts of my project.
- Able to generate a detailed report that will help the user to address cybersecurity needs.
- Created JavaScript tools that align with OWASP top 10 vulnerabilities, focusing on authentication flaws, security misconfigurations, and outdated components. (Go through Readme Part 2:Tools for deeper insights.)
- Maintained a consideration for future web app integration, ensuring outputs are ready for rendering in a web environment.
- A lot of error handling in code.

How I Was Challenged:

- The biggest issue was I was really motivated to do this project which I feel led me to increase the scope further and further. I can go into every tool and every step and how I was challenged, but it all came down to knowing what to do and what to leave. Deciding the important things first and implementing them and then trying to perfect them.
- Major technical setbacks were mainly as the LLM coding community is in its early stages, so a lot of debugging had to be done by me manually.
- The questionnaire right now is very slow as it uses GPT-4, I am planning to use my own fine tuned LLAMA2 for faster responses.
- In the subdomain tool, I first tried to get data from crt.sh through Axios, but there were too many errors to implement this completely so I shifted to brute-force.
- Different structure of package.json and requirements.txt with GithubAdvisory was a really challenging hurdle. Also there is a logic to filter properly on the basis of severity, which got me stuck for a long time.
- For the password brute force tool, I first tried brute force on a website. I got 404 errors. I tried to put a CSRF token which solved some issues but in a lot of websites issues persisted. A possible short time solution was to try different payloads but again the tool did require a proper api specification so I made a partial version only. A better approach is to use ZAP authentication but did not implement it.

- Asynchronous coding gets you especially when dealing with too many functions at a time. It took me hours to debug some tools.
- Still have to implement a cost effective solution ideally. I will be continuing working on this project after the submission to get the web app ready and put a cost effective fine-tuned LLM working to answer everything cybersecurity related, as the current script might incur heavy cost and is widely used.