Project 2 (20 pts)

Due: Friday, April 12 at 11:59pm on Canvas.

Description

In this project you will be tasked with testing and reporting on software vulnerabilities found in various source codes files using OpenAI's GPT 3.5 LLM. You are provided three Java source code files that may contain SQL injection vulnerabilities. Your task is to use GPT 3.5 to assist you in determining and finding the vulnerabilities, if any exist in the code.

Testing the Source Code

Decide on a collection of prompts that inform the LLM of what its tasks are. The source code will need to be fed to the LLM as prompts as well. The key process is to observe the LLM's output and determine if the output is correct or meaningful. If not, provide additional prompts to drive the LLM into answering the questions correctly. Things to consider include:

1. You may want to ask the LLM to output the information in a specific format.
2. You can tell LLM to focus on SQL injection vulnerability.
3. You will need to understand the code yourself so you can determine if the LLM's output is correct, and to provide additional prompts to guide it towards the desired behaviors.

Writing the Report

Write a report with the following.

1. Provide a detailed list of the prompts you used along with the responses from LLM.

2. Analysis of the LLM's results, together with the source code, as to whether the LLM's output is correct

3. Explain your reasoning for adding the prompts to improve LLM's output for the problem at hand.

Your report and outcomes will be evaluated based on the accuracy of results, and the demonstration of the ability of using prompt engineering to "train" an LLM for this task. We will evaluate your report for both accuracy and depth of thoughts demonstrated.

Submission

Submit a report in PDF with the content described above.