# Improving GPT Penetration Testing Using Prompt Engineering Techniques

Daniel Lichtenberger
*Department of EECS*
*Texas A&M University-Kingsville*
Kingsville, USA
daniel.lichtenberger@students.tamuk.edu

Mengxiang Jiang
*Department of EECS*
*Texas A&M University-Kingsville*
Kingsville, USA
mengxiang.jiang@students.tamuk.edu

Samah Allahyani
*Department of EECS*
*Texas A&M University-Kingsville*
Kingsville, USA
samah.allahyani@students.tamuk.edu

***Abstract*—With the introduction of GPT4 in early 2023, many researchers discovered capabilities of the LLM that were in the past lacking. One of these capabilities is penetration testing in the field of cybersecurity. This allows security experts to automate a large part of the vulnerability exploration process since many of the tasks required are fairly routine. A framework for doing this called PentestGPT which was able to perform at the top 1% of users at the penetration test website HackTheBox. However, it still had difficulties tackling the more difficult servers on the site. In this paper we use some general purpose prompt engineering techniques to see if there are improvements to the penetration testing results.**

## I. Introduction

ChatGPT is a large language model (LLM) that generates automated responses that correlates with a response asked by users [1]. ChatGPT itself exploded in popularity that sparked greater curiosity in the field of artificial intelligence. The model itself has accumulated an information dataset that has expanded in more recent iterations with GPT-4 [2]. Apart from ChatGPT, LLMs are becoming an investment that could affect the lives of people depending on their use and accessibility. Cybersecurity is one field that is currently being explored with LLMs such as ChatGPT. Penetration testing is one such topic in the realm of cybersecurity that is being tested with ChatGPT and the results that the model produces.

## II. Proposed Approaches

In this section, we will cover the various prompt engineering techniques used to improve GPT4's penetration test performance.

### A. Bimodal Predictor

## References

[1] Max Engman. Evaluation of chatgpt as a cybersecurity tool: An experimental ctf based approach, 2023.
[2] Walid Hariri. Unlocking the potential of chatgpt: A comprehensive exploration of its applications, advantages, limitations, and future directions in natural language processing. *arXiv preprint arXiv:2304.02017*, 2023.