Sheldon Horrell

2/16/25

CYSE-201S

Article Review #1 - *Harnessing Large Language Models to Simulate Realistic Human Responses to Social Engineering Attacks: A Case Study*

Introduction

This article was chosen as it closely relates to something I deal with as a daily occurrence at work in the Telecommunications field, specifically in Sales orders and Customer Service. I will detail how closely the findings are related to real world examples I experience of the human behavior in response to phishing and social engineering attacks and highlight their correlation to the 7 principles.

Key Concepts

In this article review I will reference Open AI's ChatGPT, an artificial intelligence software available for public use.

This AI was trained on a Large Language Model, where it was fed text from human writing to learn the information and provide it back to users in a more human way.

Social Engineering and Phishing are a type of attack used by threat actors in attempt to trick or manipulate their target into providing sensitive information like account details or passwords.

Article Summary

This article is a case study on ChatGPT's 4.0 LLM and the results of using the Artificial Intelligence to simulate a human response to a real-world phishing email based on a specific data set of 20 types of human behavior. It requests a response to the email and is provided a name, one of the preselected behaviors, and the knowledge of user account with its login and password credentials.

Analysis of Results

This case study aimed to find the results of the AI's human like behavior when being prompted to respond to a phishing email. It measured the results of 20 types of specific human behaviors and their replies to the emails. It found and details the most likely types of behaviors to respond to the email with their account passwords. It clearly defines many of the principles. Relativism, Objectivity and Parsimony are clearly pronounced through this article. Determinism is also well

understood as they set out to collect a specific result. Skepticism is also displayed and can be found in the results and output from the AI in their responses, based on specific behaviors.

Implications for Cybersecurity

This presents the frightfully real capabilities of modern AI and its ability to mimic human behavior. The implications of threat actors using AI for nefarious purposes is an ever-changing landscape. Hackers or groups of hackers can already be found to have used AI for attacking in the wild. This study and more like it could help provide a better understanding for research and training of the effects of behavior types have on the social engineering attacks. For example, carelessness and naivety failed and provided their passwords every time while logical and cautious behaviors passed every time. It is possible through a better understanding of these known types can be used in training and understanding who may be more susceptible.

Conclusion

The case study on ChatGPT's ability to simulate human responses to phishing emails highlights both the strengths and risks of AI in cybersecurity. By analyzing how different behavioral types react to social engineering attacks, the study provides amazing insights into human elements of security. The findings emphasize the importance of cybersecurity awareness training, particularly for individuals who exhibit more vulnerable behavioral traits. Perhaps the most informative takeaway from this article is how accurate the results are. Working in Telecommunications in customer service I deal with an unbelievable amount of people daily who have fallen victim to either phishing or social engineering attacks and have had their accounts compromised. The scale at which it as grown in the just last 5 years is also unbelievable. Understanding these dynamics will help strengthen defenses against social engineering and the effects on targeted behavior types.

Citation

Asfour, M. & Murillo, J. C. (2023). Harnessing Large Language Models to Simulate Realistic Human
Responses to Social Engineering Attacks: A Case Study. International Journal of Cybersecurity Intelligence & Cybercrime: 6(2) , 21-49. DOI: https://doi.org/10.52306/2578-3289.1172
Available at: https://vc.bridgew.edu/ijcic/vol6/iss2/3