Article #2 Review:

Examining the Impact of Personality Traits on Susceptibility to Social Engineering Attacks Using Large Language Models

Bailey Williams

Department of Cybersecurity, Old Dominion University

CYSE 201s: Cybersecurity and the Social Sciences

Prof. Diwakar Yalpi

November 17, 2024

This study used Large Language Models (LLMs) to simulate different human responses to phishing emails based on their primary personality traits. The case study used the GPT-4 version of ChatGPT in OpenAI's Playground (Asfour & Murillo, 2023, pp. 21, 25). The model temperature, which influences how creative or predictable its responses will be, was set to zero, meaning the responses were as predictable as possible (*LLM Temperature*, n.d.).

Research Question and Hypothesis

The research question was "How does the simulated behavior of human targets, based on the Big Five personality traits, responds to social engineering attacks?" (Asfour & Murillo, 2023, p. 22). This question aims to identify differences in how different personality traits respond to social engineering attacks so that cybersecurity training can be tailored to address weaknesses specific to each trait. The hypothesis for the experiment is that personas that exhibit different Big Five personality traits will respond to social engineering differently.

Research Methods

This study consisted of five stages: experimental design, simulation tool setup, target personae generation, attacks simulation, and response analysis. In the experimental design phase, the researchers determined the scenario that would be completed in the following stages (Asfour & Murillo, 2023, p. 24). In the simulation tool setup, the LLM model was configured in the OpenAI Playground. Next, in the target personae generation stage, the research team used GPT-4 to create twenty different personas, each possessing a unique personality trait associated with one of the Big Five personality traits. After the simulation and personae were generated, the next stage was to simulate the social engineering attacks. Each persona was attacked three times with the same phishing email and was prompted to respond to it while exhibiting its designated personality trait (Asfour & Murillo, 2023, pp. 25-26). The phishing email was designed to mimic a real one and posed as an authentic security alert from Apple, which requested the recipient to respond with their current password for verification (Asfour & Murillo, 2023, p. 25).

Data Collection and Analysis

The final stage of the experiment was the response analysis. The responses for each phishing attempt (60 total, 3 for each persona), were collected and used as the basis of the analysis. The success rate of attacks against each persona was compared and used to determine which of the Big Five personality traits was most susceptible to social engineering attacks (Asfour & Murillo, 2023, p. 26). Personae with high agreeableness traits and those with high conscientiousness traits were shown to be 99% susceptible to phishing attacks. Personae with high neuroticism were 66.6% susceptible and personae with High Openness to Experience and Extraversion showed 0% susceptibility (Asfour & Murillo, 2023, p. 27).

Relation to the Social Sciences

Multiple social science principles are present in the research paper, including objectivity, determinism, and parsimony. The research team displayed objectivity by not assuming any personality traits are more susceptible to social engineering attacks. By giving equal opportunity to each Big Five personality trait, the researchers remained objective in their experiment. Determinism was presented in the background of the case study in one of the primary purposes of the study: the idea that susceptibility to social engineering attacks is influenced by one's primary personality trait (Asfour & Murillo, 2023, p. 23). The nomothetic model of determinism is also present in the discussion portion of the paper, where it is suggested that the successfulness of cyber incidents caused by social engineering is the result of different primary personality traits (Asfour & Murillo, 2023, p. 28). Finally, parsimony is present in how the researchers summarize the background information in a way that is easy to understand for the average reader (Asfour & Murillo, 2023, p. 23).

Connections to Course Concepts

Class concepts from multiple modules were included in the research paper, including the Big Five model, victim precipitation, the education social force, and symbolic interactionism. The Big Five Personality Traits model was covered in Module Five as part of the personality theories and cyber offending section (Yalpi, n.d.-b, p. 10). The Big Five model relates to the research paper because it is the foundation for the personality trait each persona is given in the experiment. Victim precipitation is discussed in Module Four during the examination of how psychological factors may increase the risk of victimization (Yalpi, n.d.-d, p. 12). Victim precipitation relates to the research paper as the study aims to identify which personality traits increase a person's susceptibility to social engineering attacks (Asfour & Murillo, 2023, p. 23). The education social force was a topic in Module Eight, which examined how different social forces influence an individual's cybersecurity practices. Cybersecurity education as a social force suggests that an individual's cybersecurity awareness and exposure to cybersecurity education programs reduce their victimization risk (Yalpi, n.d.-c, p. 18). In the analysis of the experiment, the researchers suggest that the study shows that a "one-size-fits-all approach to cybersecurity may be insufficient" and suggests tailoring security training to different personality types (Asfour & Murillo, 2023, p. 29). Finally, symbolic interactionism, also from Module 8, studies "how individuals interact in the digital world" (Yalpi, n.d.-c, p. 7). In the paper, symbolic interactionism is featured as a core aspect of the findings, as the experiment revealed the importance of personality traits in how individuals engage with phishing emails (Asfour & Murillo, 2023, p. 30).

Applications for Marginalized Groups

This study is related to multiple concerns of marginalized groups. The study suggests tailoring cybersecurity training to different personality types, requiring employees to undergo personality testing as part of their onboarding process (Asfour & Murillo, 2023, p. 29). Personality testing may cause employees to be more mindful of how they present themselves at work, leading to a more respectful environment for minority employees. Another way the study relates to the challenges of marginalized groups is by encouraging future studies of how personality relates to the successfulness of cybersecurity attacks, which may encourage more women to pursue cybersecurity by treating cybersecurity like a social science (Asfour & Murillo, 2023, p. 30; Yalpi, n.d.-a, p. 26).

Societal Applications

The study identified two contributions to society in the discussion portion of the paper. The results showed that organizations should consider individual personalities when conducting cybersecurity risk assessments. By considering individual personalities, organizations can better predict how well their cybersecurity standards and policies will be followed (Asfour & Murillo, 2023, p. 28). Another societal application is that the results can help design more precise cybersecurity systems, such as ones that send additional warnings to individuals with high-risk personality traits (Asfour & Murillo, 2023, p. 29).

Conclusion

In conclusion, this study utilized LLMs to determine that an individual's dominant personality trait may influence how they respond to social engineering attacks. This study was related to the social science principles of objectivity, determinism, and parsimony and the class concepts of the Big Five model, victim precipitation, education as a social force, and symbolic interactionism. Future applications of this research may make the cybersecurity workplace more respectful towards minorities and encourage more women to research cybersecurity. Finally, the study can be applied to society by predicting how well employees will follow standards and policies and making cybersecurity training more effective.

References

Asfour, M., & Murillo, J. (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. https://doi.org/10.52306/2578-3289.1172

LLM Temperature. (n.d.). Hopsworks. Retrieved November 15, 2024, from https://www.hopsworks.ai/dictionary/llm-temperature

Yalpi, D. (n.d.-a). *CYSE 201S (Module 3): Strategies to Study Cybersecurity through an Interdisciplinary Social Sciences Lens*.

Yalpi, D. (n.d.-b). *CYSE 201S (Module 5): Applying Psychological Principles of Cyber Offending, Victimization, and Professionals*.

Yalpi, D. (n.d.-c). *CYSE 201S (Module 8): Social Dynamics, Social Structures, and Cybersecurity*.

Yalpi, D. (n.d.-d). *CYSE 201S (Module /Week 4): Cybersecurity and Human Factors*.

Link to Social Science Cybersecurity Study: https://doi.org/10.52306/2578-3289.1172