# Article Review #2: Investigating the Intersection of AI and Cybercrime: Risks, Trends, and Countermeasures

Cali J. Zuk

Department of Cybersecurity, Old Dominion University

CYSE 201S: Cybersecurity and the Social Sciences

Mr. Diwakar Yalpi

November 17th, 2024

## Overview

The recent media attention surrounding the development of generative artificial intelligence has spurred a wave of investment in these new products. These tools have also already been used for malicious purposes via deepfakes and frauds. The article being covered in this review focuses on the novel applications of generative AI tools for the purposes of committing cyber-crimes. The study utilizes a modification of the criminological theory called Cyber Routine Activities Theory; this theory takes the traditional framework of Routine Activity Theory and applies a cyber-security focus to it (Choi, 2008). Before proceeding further, it is worth describing what Routine Activity Theory is; this theory postulates that the daily routines of people drive the committing of crime, this is represented by three variables: suitable targets, lack of guardians, and a motivated offender (Cohen & Felson, 1979). A crime generally will not occur without those three criteria being met. In the case of cyber RAT, suitable targets are everywhere on the internet, but especially people with poor cyber hygiene, guardians can be peers, cyber professionals, or automated systems, and motivated offenders are people with both the ability to commit a cyber crime and the willingness to commit one (Shetty et al., 2024). Use of malicious AI tools acquired from the dark-web further lowers the ability required to commit a crime and should result in an eventual spike as more motivated offenders are able to take advantage of such a target rich environment. The researchers wanted to evaluate the effectiveness of the Cyber RAT framework in analyzing the growing field of AI cyber-crime. To that end, they proposed three questions to answer in this study: the first was to determine how information on malicious AI tools is being used and how the information is moving between the dark-web and the clear-net; the second is to determine how the media is contributing to the proliferation of these tools; and

finally how can society improve cyber-hygiene to better respond to the growing threats (Shetty et al., 2024).

## Method Review

The researchers split their process into two processes, one quantitative and one qualitative. The reason for this was to ensure that the study properly accounted for facets of the issue that the data may not yet show but that the experts can predict (Shetty et al., 2024). The quantitative process involved a detailed internet investigation into cyber-crime forums on both the dark-web and clear-net (Shetty et al., 2024). The investigation attempted to find examples of generative AI prompts that could work if deployed against a real target as well as gathering insight about the level of sophistication and technical expertise demonstrated by participants in these forums. They "… identified eight distinct forums that served as platforms for AI-generated prompts: FlowGPT, Respostas Oculus, Reddit, Dread, Legal RC, Hidden Answers, Dark Net Army, and YouTube."

For the qualitative process, the authors interviewed six industry professionals with expertise on both AI and cyber-crime. These experts were asked questions regarding AI's relationship with the media, what they believed strong cyber-hygiene is and how to improve it, and what implications the proliferation of AI tools both in the media and real world would have on policymaking. Their answers were then transcribed by an independent transcriber according to the Thematic Analysis Process. In this process, the answers are first broken up according to certain identified keywords and themes, and then these themes and keywords and then integrated to deliver insight into different domains (Naeem et al., 2023).

## Analysis

The researchers synthesized the expert insight from the qualitative section with the forum data from the quantitative section to determine first whether the Cyber RAT framework accurately matched generative AI cyber-crime. They determined that the data gathered did fit what the Cyber RAT framework predicted (Shetty et al., 2024). The qualitative portion strongly supported the idea that poor cyber-hygiene fulfills the suitable target requirement, and that good cyber guardianship can protect targets from themselves and thus the attackers. The researchers found that poor cyber hygiene is one of the largest predicators of cyber risk, especially with generative AI tools (Shetty et al., 2024). Furthermore, the prevalence of the AI tools further reduces the skill required to effectively no skill at all. The forums surveyed provided the necessary prompts to jailbreak generative AI tools and use them to attack information systems. The tools themselves, when properly jailbroken to remove their security safeguards are designed to be used by a completely unskilled user (Shetty et al., 2024). Traditionally, the lowest skill attacker was the script kiddie, but even a script kiddie has to find a script to copy and a target to deploy it against. The first portion of that process is now further automated. The forums themselves also provide encouragement through social networking and easy access to more skilled help, and this can push more individuals into the ranks of willing criminals by eliminating the various sources of hesitancy or overpowering them through peer pressure. The authors themselves mention the young and the old as two marginalized groups at highest risk as these groups are either too young to have been taught cyber-hygiene or too old for it to have been relevant in their careers and education (Shetty et al., 2024). Further investments into digital guardianship are advisable, as are increasing the accessibility and frequency of cyber training for all members of society. Finally, investment in defensive tools that utilize AI are also important,

as these tools will be essential to stopping the flow of easily automated attacks that malicious AI tools will enable (Shetty et al., 2024).

## Conclusion

Generative AI is a fascinating new tool that has many yet undiscovered applications in the lives of all those with access. It poses to be a transformational technology and potentially create new jobs that we cannot currently imagine. Much like all other novel technologies, we have not yet seen its most malicious applications. The investigation conducted in this study shines a much-needed light on the prospective illegal future of these tools. That these tools can be used for a bad purpose does not mean they should be outlawed or overregulated, a tool is a tool and whether it is used for benevolence or malevolence is entirely dependent on its wielder. The study also raised important considerations regarding the importance of good cyber security hygiene in stopping attacks. These tools only make launching attacks easier, they have not yet invented novel strategies of compromise. This means that at least in the present, being secure against them utilizes the same basic mentality that securing against non-AI powered attacks has so far required. The study also only focused on utilizing generative AI to launch traditional cyberattacks, it did not focus on the institutional damage that AI generated propaganda can do, nor the ease at which it has made the creation of fake news. Deepfakery is already being used to harass women and to spread political misinformation on social media. These tools will need regulation like all others and also constant vigilance from the general populace. With those in place, we can ideally be safe from their downfalls and use them in the promotion of a better future for humanity.

Works Cited

Choi, K. (2008). Computer Crime Victimization and Integrated Theory: An Empirical Assessment. *International Journal of Cyber Criminology.* https://www.semanticscholar.org/paper/Computer-Crime-Victimization-and-Integrated-Theory:-Choi/3c97233b35a7b7b537ff4d3c6db8aeb5e59911f4

Cohen, L. E., & Felson, M. (1979). Social Change and Crime Rate Trends: A Routine Activity Approach. *American Sociological Review, 44*(4), 588–608. https://doi.org/10.2307/2094589

Naeem, M., Ozuem, W., Howell, K., & Ranfagni, S. (2023). A Step-by-Step Process of Thematic Analysis to Develop a Conceptual Model in Qualitative Research. *International Journal of Qualitative Methods*, *22*, 16094069231205789. https://doi.org/10.1177/16094069231205789

Shetty, S., Choi, K.-S., & Park, I. (2024). Investigating the Intersection of AI and Cybercrime: Risks, Trends, and Countermeasures. *International Journal of Cybersecurity Intelligence & Cybercrime*, *7*(2). https://doi.org/10.52306/2578-3289.1187