AI in Social Engineering: The Next Generation of Offensive Cyber Operations

Henry Collier
Marshall University, Huntington, USA

Authorh@marshall.edu

Abstract: Few can argue that Social Engineering is the most effective way to gain access to a system. From an offensive cyber perspective, Social Engineering reduces the need to identify a new vulnerability within a system to gain access by praying on psychological factors like emotion and fear. Social engineering has an effective rate of ranges between 80% and 90%. When AI is added to the picture, the likelihood of a successful social engineering scheme increases because AI can add a layer of realism and personalization that gets past a person's barriers. When AI-influenced social engineering attacks are added to an offensive cyber operation, the attack surface grows increasingly more significant. The larger the attack surface, the higher the likelihood of a successful cyber operation. These targets can attack governmental agencies, the military, critical infrastructure, the healthcare system, and more. Understanding how AI in Social Engineering is impacting Offensive Cyber Operations is important. This case study looks at how AI is impacting Social Engineering and how AI influenced Social Engineering is being used by Nation State Threat Actors in Offensive Cyber Operations. The U.S. Department of Defence

Keywords: Offensive cyber operations, Social engineering, Artificial Intelligence

1. Introduction

The cyber warfare domain is relatively new. The first was identified by the Economist in 2010 when they declared that warfare had entered the fifth domain- cyberspace (Shoaib, 2016). In 2011, the United States Department of Defence officially included cyberspace in its battle domains, making cyberspace the fifth domain (Shoaib, 2016). At the time when cyberspace became the next domain in warfare, artificial intelligence (AI) was in its infancy and was not necessarily considered as the cyber domain that was being established. With the advancements in AI, we are starting to see a continuous cross-integration and penetration of artificial intelligence across multiple domains (Zeng, 2022). AI has begun to reframe cybersecurity from both the attack and defence perspectives (Zeng, 2022). As these attacks are becoming even more prolific, it is easy to see that AI spans the entire spectrum of cyber-attacks. One area that is of great concern is the human firewall approach (Author, 2024).

Historically, the most successful social engineering schemes required a great deal of preparation and work in order to appropriately target someone and compromise them. In order for any form of manipulation to work, you must know your target. This is certainly true with spear phishing, which is a very targeted approach to social engineering. With the introduction of AI, these attacks have become almost effortless (Author, 2024)(Zeng, 2022). AI takes the process of data gathering and completes it in a fraction of the time it takes a person to do. Furthermore, once the data is collected AI can then develop effective social engineering schemes based on the data (Zeng, 2022). One of the reasons why AI is so successful at developing these social engineering attacks is that today's population has grown accustomed to putting their entire lives on social media platforms for everyone to see. The vast amounts of personal information available creates the perfect storm when it comes to developing social engineering attacks because the data from social media can be used to identify the buttons that the attack needs to push in order for it to be successful (Arif, et al., 2024). A great example of this is the grandparent's attack, whereby the threat actor would call an elderly person and pretend to be their grandchild. Historically, the target could detect deception because the caller doesn't sound like their grandchild. However, it is possible to take snippets from TiKTok and create a deepfake voice that can be used in the call. This significantly increases the probability of success.

When one considers the possibilities that AI brings to the cyber attack table, it is easy to make the leap that AI will be used, especially with social engineering, as a means of offensive cyber operations. The four main nation-state threat actors, China, Russia, Iran, and North Korea, invest heavily in their cyber operations and are trying to dominate the environment. Security researchers are seeing more complex and advanced attacks, and indicators are showing these attacks are coming from all four of the main nation-state threat actors

(Federal Bureau of Investigation, 2024). Furthermore, it is clear that these nation-state threat actors also want to manipulate the democratic process (Pira, 2023). They do this through massive disinformation schemes targeting voters in countries like the United States and Great Britain (Pira, 2023). This form of attack does not

look to many as a social engineering attack, but when you peel away the layers, it is clear it is (Balaban, 2023). Instead of gaining access to a system or taking someone's money, this scheme's goal is to manipulate the election process to either ensure their preferred candidate gets elected or to simply sow distrust about the democratic process among the population.

Failing to understand the depth of this problem is a security risk for any nation. This paper analyzes the current environment and assesses how the target nations defend themselves.

2. Behavioral Psychology and AI in Offensive Cyber

To best understand how AI is being used as a component of offensive cyber operations with social engineering being a targeted approach, we need to first break down social engineering. Social engineering is simply manipulating someone to do something they otherwise would not do (Author, 2021). In order to affect this manipulation in a successful manner, the threat actor needs to understand the target's behavioral tendencies and then find one they can poke to get a response (Author & Author, 2020). Humans are multifaceted beings, filled with emotion and culture, influencing behavior (Author & Author, 2020) (Author et al., 2023). These emotions and cultural influences impact the way a person makes security-minded decisions, often resulting in compromise (Author & Author, 2020) (Author et al., 2023). Most cyber professionals think there is a techncal solution to this problem, but there is not (Author, 2021). Humans will circumvent any technical solution if they feel the solution is getting in the way of what they want to do. Theat actors know this, and work very hard to develop effective methods of attack. For example, if someone is a pet lover, the threat actor could send a message on social media about some poor animals in distress and provide a link for the user to support recovery efforts. This could lead to credit card fraud but also may lead to a system compromise if the website the user is being sent to has malware embedded into it through a drive-by attack (Provos, et al., 2007). The limits that a threat actor will go to in order to compromise a system are limitless, and no system is off limits. This is supported by how China has been working to compromise the critical infrastructure systems within the United States (Cybersecurity & Infrastructure Security Agency, 2024). This should be very concerning to the population because these critical infrastructure systems are the heartbeat of a nation. Gas pipelines ensure we have access to natural gas to heat our homes, the electrical grid provides the power that we rely on so much in our lives, and the water systems ensure we have clean, reliable water to drink. These are just a few examples of critical infrastructure, and if any of these or the others are compromised by a nation-state threat actor, then it puts the population in the hands of that threat actor. When a nation is going to invade another, a significant aspect of the attack is to destroy or degrade the services that the people rely upon to survive. This is done to both instill fear, and to cause distraction within the targeted nation.

3. Social Engineering/Cyber Threats Created by AI

Now that we have a better understanding of the target of these attacks, we need to delve into the different forms of attack that are best propagated by AI. Several of these attacks are multi-domain attacks, meaning they combine more than one form of attack. For example, using AI to convince someone to click on a link that takes them to a malicious website is a form of social engineering. At the same time, the malware that was generated by AI, which is going to be downloaded through a drive-by download attack, is not social engineering. When combined, this type of attack generates a more effective response. The malware that is created by AI frequently has attributes that human-developed malware has. AI can generate Malware that is dynamic in nature, which makes detecting it with conventional antivirus protection more difficult (Arif, et al., 2024). Furthermore, the natural language process used by AI when generating phishing attempts produces incredibly convincing emails and messages that give the appearance of being legitimate (Arif, et al., 2024). In addition to the new complexity of AI-driven social engineering schemes and AI-developed malware, we also need to look at the efficiency of how these attacks are conducted and how AI is supporting these efforts. AI has the power to automate these attacks on an entirely new level. Historically, social engineers would have to spend hours sending out phishing emails. With AI, this process takes seconds. AI can send out thousands of phishing emails, increasing the probability of success (Arif, et al., 2024) (Malatji & Tolah, 2024). Let's not forget how AI is able to generate convincing deepfakes to support their attacks (Malatji & Tolah, 2024). AI can further be used as part of a distributed denial of service attack by generating hundreds of botnets, which can then be directed to attack a specific victim. An example of this type of attack, although it was not done by AI, is the attack on the Domain Name System provider DYN in 2016, where the attacks came from internet-connected devices like printers, IP cameras, thermostats, and baby monitors. Now, imagine if AI was conducting this attack and ask yourself what the likelihood of an organization being attacked in this manner is able to recover. The DYN attack lasted three days. What if China attacked the United States, Great Britain, or any other nation

within NATO with millions of botnets? How likely is it that the attack will be repulsed? Small-scale attacks that have historically been conducted by threat actors can now be magnified through the use of AI.

4. Impacts of AI-Induced Offensive Cyber Operations on the World

Cyberwarfare, the fifth domain of warfare, is a gray space that every nation operates in. If the battlespace were land, sea, or air, it would be easy to determine when the actions of a nation-state threat actor equated to an act of war and required some form of kinetic response. As cyber is the gray space that it is, there is no hard line to determine an act of war. In fact, most nations tend to err on the side of caution versus pursuing what many would consider a nuclear response, one that is disproportional. Historically, cyber-attacks fell into one of three primary categories—politically motivated, socio-culturally motivated, and economically motivated (Gandhi, et al., 2011). Although these categories can be qualified as they are related to offensive cyber warfare, there is more to it than these categories. Nation-state threat actors have very specific reasons to conduct warfare, including cyber warfare. For example, China is looking to gain global economic dominance, which is evident in how it conducted its early attacks, where the goal was to steal intellectual property from the United States war machine (Jinghua, 2019). On the other hand, Iran uses Islam as the basis of their dislike of the West (Larison, 202).

China, Russia, North Korea, and Iran are constantly working to gain the upper hand over the West. China has infiltrated many of the critical infrastructure systems within the United States (Cybersecurity & Infrastructure Security Agency, 2024). North Korea has targeted companies like Sony after Sony produced a movie that did not positively reflect North Korea's supreme leader. Russia, on the other hand, has worked with criminals to attack the United States on multiple fronts. Hospitals, financial institutions, and the government have all been probed and, in many cases, breached by Russia or one of its adversaries. For example, Russia has ties to the ransomware group Black Basta, which targets healthcare institutions. The Intelligence community believes that groups like Black Basta are attacking the United States at the request of Russia (McKeon, 2024). This approach differs from China, Iran, and North Korea, which tend to conduct cyber-attacks independently as each country has invested heavily in its cyber capabilities. China, Russia, and Iran worked to influence the 2024 United States Election (Cybersecurity & Infrastructure Security Agency, 2024).

It is clear that the four main nation-state threat actors are using AI as a component of their offensive cyber operations. Their goals are clear: they want to erode trust in digital systems, obtain global economic dominance, erode confidence in the democratic election process, and set the stage so they would win any future kinetic wars that might come with time. They do all of this while skirting the boundary of what would be considered an escalation event that could lead to kinetic war. This is a constant battle, and this battle is not going to get any easier. One could predict that with threat actors using AI to attack and defenders using AI to mitigate and stop risk, someday, the AI battle could make the global Internet unusable. With how much today's society relies on its connectivity, this is something that needs to be prevented.

5. Future Work

Further study needs to be conducted to better understand the full extent of the problem. This case study only scratches the surface. However, much of the information that would be needed to further prove how AI is being used in offensive cyber warfare is classified and not available to the research community. Even those in the research community with clearances and access to this information have no pathway to conduct this type of research. The challenges to overcome this problem are great, but with time and effort, it is possible to conduct further research into this topic and develop a better understanding of what is happening in cyberspace.

6. Conclusion

The cyber warfare domain is the newest addition to the battlespace, joining Land, Sea, Air, and Space. With AI advancing daily, researchers are starting to see attacks that are either conducted by AI or heavily influenced by AI. As the attacks become more sophisticated and prolific, it is easy to see the entire spectrum of cyber-attacks. Humans, the most vulnerable part of a network, are at the greatest risk to national security. Humans are so vulnerable because they are multifaceted beings full of emotion and behavior that is contradictory to logic. Even the most logical-minded individual still has feelings, biases, and behavioral tendencies that could lead to a cyber system being breached. It is well known that AI can process large amounts of data quickly, and with end users posting their entire lives on social networking sites, threat actors are using AI to mine this data to develop attacks. Nation-state threat actors use this to develop social engineering attacks as part of their

offensive cyber strategy. With AI being able to generate thousands of attacks in a fraction of the time it takes a person to do it, the threat actors fully embrace AI. With cyberspace being a gray area, nation-state threat actors push the boundaries because nobody wants to go to war over a cyber incident. The question of whether or not nation-state threat actors are using AI and social engineering as part of their offensive strategy is not up for debate. The evidence is clear that they are doing so in order to achieve their agenda. The question becomes how we defend against these attacks and keep the Internet up and running in the process. Perhaps the Luddites have it right, and we should eliminate technology. Short of doing this, it is unknown what the future will bring.

References

Arif, A., Khan, M. I. & Khan , A. R., 2024. An overview of cyber threats generated by AI. International Journal of Multidisciplinary Sciences and Arts, 4(3).

Author, H. & Author, A., 2020. The Port Z3ro Effect:. Ciooebgageb, AIRCC Publishing Corporation.

Author, H. D., 2020. Social Media: A Social Engineer's Goldmine. Larnaca, s.n.

Author, H. D., 2021. Enhancing Information Security By Identifying and Embracing Executive Functioning and the Human Behaviors Related to Susceptibility. Colorado Springs: s.n.

Author, H., 2022. Including Human Behaviors into IA Training Assessment: A Better Way Forward!. Chester, European Conference on Cyber Warfare and Security (.

Author, H., 2024. AI: The Future of Social Engineering. Johanesburg, Academic Conferences International.

Author, H., Morton , C., Altharhi, D. & Kliener, J., 2023. Cultural Influences on Information Security. Athens , European Conference on Cyber Warfare and Security.

Balaban, D., 2023. Social Engineering And The Disinformation Threat In Cybersecurity, s.l.: Forbes.

Cybersecurity & Infrastructure Security Agency, 2024. CISA.gov. [Online] Available at: https://www.cisa.gov/topics/cyber-threats-and-advisories/nation-state-cyber-actors/china#:~:text=Nation%2DState%20Cyber%20Actors,-China%20Cyber%20Threat&text=CISA%2C%20the%20National%20Security%20Agency,of%20multiple%20critical%20infrastructure%20organiza
[Accessed 20 01 2025].

Cybersecurity & Infrastructure Security Agency, 2024. Nation-State Cybe Actors, s.l.: CISA.

Federal Bureau of Investigation, 2024. China Cyber Trheat: Chinese Military Hackers Target U.S. Business, s.l.: s.n.

Gandhi, R. A., Laplante, P. A. & Sousan, W., 2011. Dimensionso f Cyber-Attacks: Cultural, social, Economic, and Political. IEEE Technology and Society Magazine, February.

Jinghua, L., 2019. What Are China's Cyber Capabilities and Intentions?, Washington : IPI Global Observatory.

Larison, D., 202. Why the US and Iran hate each other!, Cambrige: MIT Center for International Studies.

Malatji, M. & Tolah, A., 2024. Artificial intelligence (AI) cybersecurity dimensions: a comprehensive. AI and Ethics.

McKeon, J., 2024. xtelligent Healttec Security. [Online] Available at: https://www.techtarget.com/healthtechsecurity/news/366615458/US-calls-out-Russia-for-enabling-healthcare-cyberattacks#:~:text=commitments%2C%20Neuberger%20said%2C%20Russia%20continues,charges%2C%20is%20a%20Russian%20national.[Accessed 20 January 2025].

Pira, F., 2023. Disinformation a problem for democracy: profiling and risks of consensus manipulation. Frontieres in Sociology, Volume 8.

Provos, N. et al., 2007. The Ghost In The Browser Analysis of Web-based Malware, s.l.: Google.com.

Shoaib, M., 2016. AI-Enabled Cyber Weapons and Implications for Cybersecurity. Journal of Strategic Affairs.

Zeng, Y., 2022. AI Empowers Security Threats and Strategies for Cybder Attacks. Procedia Comptuer Science, Volume 208, pp. 170-175.