Jefferson Morales and Jaydon Christen

Professor Bui

CSC 120

December 5, 2023

Assessing the Efficacy and Implications of Generative Cybersecurity AI: A Comprehensive Understanding of its Threat Detection Capabilities and Societal Impact

The defense from malicious attacks by hackers, spammers, and cybercriminals against internet-connected devices and services is known as cybersecurity. Organizations employ this procedure to guard against ransomware attacks, identity theft, phishing schemes, data breaches, and financial damages. A subset of artificial intelligence called "generative AI" is focused on producing new content or data and has special powers to support cybersecurity initiatives. Within this research, we seek to dive deeper into the environment of generative cybersecurity AI to better understand its efficacy by examining its threat detection capabilities and assessing its societal impact.

To begin examining generative cybersecurity AI, we must understand the world of cybersecurity and the defense methods used without generative AI. The article, “The Rise of Generative AI in Reinforcing Cyber Defense,” by Pooja Sharma, reveals that as of right now, “The cost of cybercrimes is equally on the rise and is projected to reach $10.5 trillion by 2025 from $6 trillion in 2022” (Sharma).[1] Cybersecurity is in high demand for consistent and effective methods of defense. Therefore, how is cybersecurity implemented? The article “Cybersecurity Is Critical for All Organizations – Large and Small” by Steve Ursillo and Christopher Arnold reveals how cybersecurity is implemented: determine - to manage cybersecurity risk to systems, people, assets, data, and capabilities, develop an organizational understanding. Protect - create and put into place the necessary safeguards to guarantee the provision of essential services. Detect - create and carry out the necessary actions to determine when a cybersecurity event occurs. React- create and carry out the necessary actions to address a cybersecurity incident that has been discovered. Recover - create and carry out the necessary actions to uphold resilience plans and restore any services or capabilities that were hampered by a cybersecurity incident.

[1] Sharma (July 2023) “The Rise of Generative AI in Reinforcing Cyber Defense”

Therefore, there are multiple important system tools and implements that are employed to lessen these harmful attacks. For example, Software and hardware firewalls are used to defend systems against intrusions by users connecting to the organization's networks from the outside as well as the inside. Web proxy and malware/spyware protection programs shield the computer from malicious software that could open pop-up windows or have more subtle intentions, like logging passwords and usernames for criminal purposes. Email inboxes are protected from being blocked by unwanted messages by anti-spam software. Users who visit websites that are intended to capture user information for illegal activities are protected by anti-phishing software.[2] Cybersecurity is pivotal to organizations and individuals in today's advanced technological world.

Now that we understand cybersecurity and how it is practiced without generative artificial intelligence, how does generative AI come into play? In Micheal Hill's article "If you don't already have a generative AI security policy, there's no time to lose," he says, "As business use cases skyrocket, the message for CISOs (chief information security officers) is clear: if you don't have a strong AI security policy specifically pertaining to generative AI you need to make one right away" (Hill). Hill calls generative AI the "new beast" as it is evolving quickly with fulfilling promises.[3] Why is this the case? As said before, generative AI is a subset of artificial intelligence that focuses on creating new data or content and has unique abilities to assist cybersecurity efforts. Organizations can proactively detect vulnerabilities, model potential attacks, create strong defenses, and react quickly to new threats by utilizing generative AI.[4] The research report "LEARNING FROM CYBERSECURITY, PREPARING FOR GENERATIVE AI" by the Atlantic Council reveals that "Generative AI refers to powerful algorithms that can

[2] Ursillo and Arnold (October 2023) "Cybersecurity Is Critical for all Organizations – Large and Small"
[3] Micheal Hill (July 2023) "If you don't already have a generative AI security policy, there's no time to lose"
[4] Sharma (July 2023) "The Rise of Generative AI in Reinforcing Cyber Defense"

produce or generate text, images, music, speech, code, or video" (Atlantic Council). Large language models (LLMs), which are composed of enormous artificial neural networks, are the foundation of these algorithms, and they are trained by consuming and analyzing massive volumes of data. Although not a novel technology, generative AI and LLMs gained widespread recognition after the tremendously successful release of ChatGPT and DALL-E at the end of 2022. Prominent tech giants like Microsoft, Google, and even more recent organizations like OpenAI and Anthropic are making significant investments in creating their own LLMs and related products for everyone to use. With generative AI's potential to completely transform society, innovators, investors, and governments have turned their attention back to these models and the products they drive as they become essential in the realm of cybersecurity and technology today.[5]

Following this information, we begin to understand threat detection abilities. From the last source used, "LEARNING FROM CYBERSECURITY, PREPARING FOR GENERATIVE AI" by the Atlantic Council, it is revealed that because generative AI reduces the time, money, and technical expertise needed to create large quantities of harmful, hyperrealistic content and possibly distribute it widely, it alters the nature of influence operations online and the moderation of illicit content. Content produced by automating the creation of spam, influence operations, fraudulent content, misinformation, and other illegal online activities through generative AI is more effective than previously produced misinformation. Through this research, we begin to understand its benefits and downsides or weaknesses. The research report evinces that generative AI can help mitigate and identify the harms they introduce within data curation, the intentional creation of data for the purpose of enabling AI and sophisticated data applications; model

[5] Atlantic Council (June 2023) "LEARNING FROM CYBERSECURITY, PREPARING FOR GENERATIVE AI": 1-12

training, the stage where practitioners attempt to minimize a loss function over the prediction range by fitting an optimal weight and bias combination to a machine learning algorithm; lastly, post-deployment, the collection of procedures, tasks, and actions carried out following an application's successful deployment to its intended environment.[6]

Generative AI has many potential uses in the field of cybersecurity. One important use is for the detection of anomalous behavior. One way this is done is through a type of generative AI model known as a Variational Autoencoder (VAE). Traditional autoencoders are neural networks that encode data in a compressed form and can reconstruct this data to its original form. VAEs are based on this concept but have the ability to generate new data to summarize the previous data in a new way.[7] VAEs are used to detect anomalies by seeing how the new data that is generated differs from the training data. This can help to identify any outliers in the data and alert analysts of potential security threats. This new technology is able to detect threats that may not be seen by previous cybersecurity systems. Furthermore, another use of generative AI in cybersecurity is for password security. AI models find weaknesses in patterns of passwords, and generative AI allows for the generation of recommendations for stronger, more secure passwords.[8]

One other application of generative AI in cybersecurity is for phishing detection. This is done by use of generative adversarial networks (GAN). GANs are another type of neural network that is applied for the detection of phishing attempts. These networks are composed of two parts: the generator and the discriminator. The generator produces false phishing data to train the network, which becomes increasingly similar to real phishing attempts. The second part, the

[6] Atlantic Council (June 2023): “LEARNING FROM CYBERSECURITY, PREPARING FOR GENERATIVE AI”: 1-12

[7] Rocca, “Understanding Variational Autoencoders”, Sept 2019.

[8] Stankovich, “Unlocking the Potential of Generative AI in Cybersecurity”, Oct 2023.

discriminator, is the part that detects phishing attempts. It is trained by the generator, which sends it fake data. The generator keeps sending more and more realistic data until it is indistinguishable from real data.[9]

Though there are many positive uses of generative AI in cybersecurity, there are also many ways it could be used for malicious purposes. One way generative AI is used maliciously is by generating phishing attacks that are undetectable. If the generator in generative adversarial networks can make fake phishing attempts that are indistinguishable from real ones, a similar generator could be used for real phishing attacks. Furthermore, attackers could use generative AI to spread disinformation and make fake news. One way this is done is through deep-faked videos. In these videos, generative AI is used to make it look like someone is saying or doing something that they did not actually do. This could be very destructive if it is a public figure, and if the people who see the video believe that it is real.

This technology is becoming increasingly problematic as it becomes more realistic. It used to be very easily detectable, but now some videos are almost indistinguishable. In, “LEARNING FROM CYBERSECURITY, PREPARING FOR GENERATIVE AI,” the Atlantic Council writes, “This increased volume of deep fakes not only risks flooding trust and safety systems with exponentially greater quantities of content that will need to be monitored but also injects greater quantities of hard(er)-to-detect forms of high quality (and potentially harmful) fake content into the system, too.”[10] This widespread use of misinformation has the potential to have detrimental effects on society. It will lead to distrust of the media, and it will be very difficult to know what is true and what is not.

[9] Wood, “Generative Adversarial Network”

[10] Atlantic Council (June 2023): “LEARNING FROM CYBERSECURITY, PREPARING FOR GENERATIVE AI”: 1-12

It is inevitable that generative AI in cybersecurity will have a large impact on society; as cybersecurity technology advances due to generative AI, security must be updated to keep up with possible illicit uses of this technology. All new capabilities that arise from generative AI in the field of cybersecurity are accompanied by threats of use for cyber attacks. Security systems that use generative AI technology should be mostly safe from these generative AI attacks for now while there is still a possible distinction between false generated data and real data. However, systems that are not using this new technology are very vulnerable to attacks. It seems the only solution to prevent generative AI attacks is to keep security systems at least as advanced as this technology since pre-generative AI security systems have no chance against generative AI models. Cutting-edge cybersecurity technology must be kept from people who wish to use it for malicious purposes to prevent attacks that will severely damage society through stolen money and information.

Reference Page

1.

Sharma, P. (n.d.). The Rise of Generative AI in Reinforcing Cyber Defense. Grazitti Interactive.https://www.grazitti.com/blog/the-rise-of-generative-ai-in-reinforcing-cyber-defense/

2.

Ursillo, S., & Arnold, C. (2023, October 23). *Cybersecurity is critical for all organizations – large and small*. IFAC. https://www.ifac.org/knowledge-gateway/preparing-future-ready-professionals/discussion/cybersecurity-critical-all-organizations-large-and-small

3.

Hill, M. (2023, July 17). *If you don't already have a generative AI security policy, there's no time to lose*. CSO Online. https://www.csoonline.com/article/646291/if-you-dont-already-have-a-generative-ai-security-policy-theres-no-time-to-lose.html

4. Sharma, P. (n.d.). The Rise of Generative AI in Reinforcing Cyber Defense. Grazitti Interactive.https://www.grazitti.com/blog/the-rise-of-generative-ai-in-reinforcing-cyber-defense/

5.

Atlantic Council. (2023). ANNEX 6: LEARNING FROM CYBERSECURITY, PREPARING FOR GENERATIVE AI. In *SCALING TRUST THE ON WEB* (pp. 1–12). Atlantic Council. http://www.jstor.org/stable/resrep51651.26

6.

Atlantic Council. (2023). ANNEX 6: LEARNING FROM CYBERSECURITY, PREPARING FOR GENERATIVE AI. In *SCALING TRUST THE ON WEB* (pp. 1–12). Atlantic Council. http://www.jstor.org/stable/resrep51651.26

7.

Rocca, Joseph. *Understanding Variational Autoencoders (VAEs).* https://towardsdatascience.com/understanding-variational-autoencoders-vaes-f70510919f73

8.

Stankovich, Miriam. *Unlocking the Potential of Generative AI in Cybersecurity: A Roadmap to Opportunities and Challenges.* https://dai-global-digital.com/unlocking-the-potential-of-generative-ai-in-cybersecurity-a-roadmap-to-opportunities-and-challenges

9.

Wood, Thomas. *Generative Adversarial Network.* https://deepai.org/machine-learning-glossary-and-terms/generative-adversarial-network