Volume 9, Issue 12, December – 2024 International Journal of Innovative Science and Research Technology
ISSN No:-2456-2165 https://doi.org/10.5281/zenodo.14565177

The Role of Artificial Intelligence Detecting and Preventing Phishing email

Kennelyn S. Araneta[1]; Nurfraida A. Julasbi[2]; Syeddin Nadvi A. Masbud[3]; Fathar A. Mohammad[4]; Juljamar A. Mohammad[5]; Giner A. Nur[6]; Haniza S. Sajiron[7]; Ericka A. Salahuddin[8]; Omar S. Tiamwatt[9]; Shernahar K. Tahil[10]
Mindanao State University-Sulu, Philippines.

Abstract:- The research study utilizes a structural basis characterized by a multi-method approach, including the use of documents and interviewing the experts and practitioners of cybersecurity. This is a dual technique that allows for the dissection of AI applications that are being used to catch phishing violation. The study also reveals that the primary AI methods are machine learning, which is the process of applying mathematical and statistical algorithms to the pattern of email data to classify the incoming mail as legitimate or malicious. This is accomplished through NLP, which is the capability to look at the words and phrases to determine if there are any suspicious actions, in email traffic, to detect if there were any changes. Recharge shows that artificial intelligence plays a crucial role in boosting the accuracy and efficiency of phishing detection system to a great extent. In the case of machine learning models, for example, the bot can be trained on a huge dataset to help identify low-levels signs of phishing attacks, which will eventually reduce the time needed of recognizing and reacting. Furthermore, NLP algorithms allow the practitioners a more profound examination of language used in phishing emails; thereby, systems are able to identify not only the common templates but also the novel attacks that may not have patterns. On the other hand, the research also sees the obstacle faced in the application of AI in phishing detection. A key worry is the flexibility of cybercriminals, who are perpetually coming up with new methods to get around automated security systems. The hide-and-seek nature includes the requirement of perpetual training and the updating of AI models, ensuring that the models are indeed resistant to attacks from new forms of threats. Moreover, the study finds that, although AI can do the majority of the work in phishing detection, it is not infallible. There may be these false positives and a few false negatives on some of them, which could clearly lead to disruptions in legitimate communications or help the criminals get on board unchecked. Ethical considerations also emerge as a significant theme in the research. The reliance on AI for cybersecurity raises questions about privacy, data security, and the potential for bias in algorithmic decision-making. The study emphasizes the importance of transparency in AI systems and the need for organizations to maintain a balance between automated solutions and human oversight.

Keywords:- *Phishing Attack, Artificial Intelligence and Cybersecurity*

I. INTRODUCTION

The emergence of technologies that enable digital communication has transformed the manner in which people and corporations communicate with each other, but it has also created serious cybersecurity issues. One of the most widespread and damaging types of cybercrime that emerged is phishing email. Fraudulent electronics messages take advantage of human weaknesses and often culminate in unauthorized purchase of private assets, loss of funds, or even an organization's standing. It is worth noting that cybercriminals have mastered the art of penetration; therefore, traditional security systems have been rendered obsolete. As a result of the increasing menace posed by phishing, organizations are adopting artificial intelligence technology to secure emails. In this case, VIPN can utilize Al technologies, including machine learning (ML) and deep learning, to combat phishing attacks effectively. The purpose of this paper is to discuss the potential of artificial intelligence as a means of identifying and preventing phishing email use, problems, and outlook. This can help organizations shield themselves and their stakeholders against any cyber threats by understanding how AI can be used against phishing. Phishing has always been an issues, but the solutions are AI-powered. AI does this by scanning for similarities by using different methods such as behavioral analysis, natural language processing, and anomaly detection. Phishing detection techniques such as analyzing email headers and contents. For example, such systems are able to identify harmful URLs, attachments that may be suspicious, and sender behavior that is not typical (Eze, C.S., & Shamir, L.2024).

II. UNDERSTANDING PHISHING EMAILS

A. Definitions and Characteristics

Phishing is a form of cybercrime that makes use of fake websites and emails, all with the aim of telling users to disclose sensitive information. Such e-mails are characterized by certain elements, as depicted by Jahankhani et al. (2020), such as urgent requests, spoofed sender addresses, and links to malicious websites.

- *Sender Impersonation:*

Phishers emails employ a method of copying real addresses in order to try and trick recipients. For instance, an email could be made to support@bank-example.com rather than true support@bank.com.

ISSN No:-2456-2165 https://doi.org/10.5281/zenodo.14565177

➢ *Urgent Language:*

Phishers also use emails that make it seem as though things are urgent or require fast response. The use of phrases such as "Your account has been stolen!" or "You must act now!" Tend to be threats that are realizable.

➢ *Requests for Sensitive Information:*

No legitimate organization would send an email and remind you that you owe them personal information. Phishing emails usually ask users to submit personal information or click on some links to verify their accounts.

➢ *Poor Grammar and Spelling:*

Phishing emails will contain spelling or grammatical errors that if properly looked for would make one discredit the whole message. While more well-organized attacks are sufficiently persuasive, many attacks still show such attributes.

➢ *Generic Greetings:*

Someone would address an email to someone else but use a common phrase 'Dear Customer' instead of 'Dear Joshua' where Joshua is the customer being addressed.

➢ *Spoofed Sender Addresses:*

'From' address works – no need to fwd norephishing@xyz.com – spoofed meaning forward email received as fencing means many times an email appears to be from a different.

B. Common Phishing Techniques

Popular techniques include spear phishing, whaling, and clone phishing. Defending on which technique they use, attackers tailor their methods for individual users or organizations, sometimes deploying personalized information to enhance their credibility (Bertino & Islam, 2018).

➢ *Phishing Attackers come in many forms, but some of the most common Techniques are:*

- *Spear Phishing:*

This particular technique targets specific individuals or organizations, often using personal information to create a sense of trust. Attackers might study their targets on social media to create convincing communications.

- *Whaling:*

Whaling is essentially a high-end type of phishing that targets the big fish inside a company, such as top executives or important officials. This kind of operation typically relies heavily on phishing lures, but more often than not, these lures are highly customized messages that look perfectly authentic.

- *Clone Phishing:*

In this strategy, a fraudulent copy of an email that was sent legitimately in the past is produced, having swapped out the original attachment with one that contains malware. The con's success relies on unsuspecting users extending their trust originally placed in generous communications they have received.

- *Vishing and Smashing:*

Vishing refers to voice phishing, while smashing is phishing over SMS; these are the two variations where attackers use phone calls or text messages to trick victims. Hackers may take on the identities of authentic organizations; they are not to extract sensitive information from the unsuspecting.

C. Artificial Intelligence Basics

The term artificial intelligence means the simulation of human intelligence by machines programmed to think and learn like humans. AI encompasses a wide variety of technologies and methodologies like machine learning, natural language processing, and robotics. Thus, at its core, artificial intelligence intends to enable computers to perform activities that otherwise required human intelligence; for example, understanding languages, recognizing patterns in them, solving problems, and making decisions.

➢ *Machine Learning*

Machine learning (ML) is a subset of AI that enables systems learn from data and improve over time. In phishing detection, ML algorithms analyze patterns in email characteristics to identify potential threats (Sharma et al., 2021). Machine learning (ML), being a part of artificial intelligence, allows computer systems to autonomously learn and grow their performance by themselves with the utilization of past data without any guidance. Phishing prevention context refers to ML algorithms analyzing large volumes of email data to uncover and detect the patterns and irregularities that can be matched to phishing activities. The ability to classify new emails received as either phishing or legitimate ones can be developed by the ML model through the use of a data set that is labeled and has samples of both phishing and non-phishing emails.

➢ *Deep Learning*

Deep learning is the utilization of neutral networks along with some others that include classification, regression, and representation learning, which are the different deep tasks that it can perform. The area is inspired by biological neuroscience and is mainly based on composing artificial neurons into layers and training them to handle inputs. The term 'deep' refers to the large number of layers (from three to some thousands) used in the network. There can be two approaches to the training of models: supervised and semi-supervised. Deep learning is an even more advanced version of machine learning that utilizes neural networks with many layers to operate on huge sets of data. This method is very effective in finding the complex patterns that may not be visible in the traditional methods of machine learning. When dealing with phishing detection, deep learning can be used to break down the email message into their components, such as text, image, and links, to be able to identify sophisticated phishing that could avoid being deleted by common methods. Deep learning (DL), a more advanced form ML, utilized neural networks to process complex data. DL models can enhance phishing detection by understanding intricate patterns in email content and structure (Zhang et al.2020).

ISSN No:-2456-2165 https://doi.org/10.5281/zenodo.14565177

D. Application of Artificial Intelligence in Email Security

➢ *Email Filtering*

AI-driven email filtering system utilize ML algorithms to automatically categorize incoming emails, flagging potential phishing attempts based on learned characteristics (Sarker et al., 2021). These systems use algorithms that analyze different email properties like sender reputation, content analysis, and link verification. Phishing attacks are still the most common and dangerous threats to cybersecurity, mainly through email, which is used to fool people and organizations. Such attacks are usually carried out by faking trustworthy entities that convince the users to give out vital information or make the wrong move. The traditional email filters, which are strict rule-based, have become weak against the advanced strategies of modern cybercriminals. Here Artificial Intelligence (AI) comes into play by altering the rules of the game to identify phishing emails and thus stopping them from going through. These systems analyze and process vast amounts of data, which includes email, sender information, and user behavior to differentiate secure emails from fraudulent ones. AI, on the contrary, learns from new threats over time, thus increasing its precision and the ability of it to change. Through the use of AI, the detecting process is made more efficient by way of a reduction in the number of human mistake and the provision of a more ingenious shield against the phishing danger. With the intelligence that AI has, it can recognize slight hints such as uncommon wordings or suspicious sender actions, which makes it a key instrument in the digital environment that is evolving. With the fact that email is a very essential communication tool, the use of AI to protect against phishing attacks.

➢ *Behavior Analysis*

Among cyberattacks, phishing emails are usually skilled and occur quite often. They mislead individuals into disclosing sensitive data or making risky decisions. Leveraging the methods of AI and monitoring digitals behavior has becomes one of the main tactics to thwart such incidents since companies are making them more complex and are concocting new ways of evading traditional control. An irregular user action that could be a phishing menace can be figured out by AI. The program monitors the normal activity of users and can differentiate suspicious conditions that need to be further examined. The AI tools will flag, for instance, if a user randomly visits uncommon sites or begins to receive emails from unknown senders.

E. Challenges and Limitations of AI in Phishing Detection

Despite its advantages, AI in phishing detection faces challenges, including data privacy concerns, the need for large datasets for training, and the potential for adversarial attacks that can deceive Al system (Wang et al., 2020).

➢ *Data Privacy Concerns:*

Are basically the questions regarding the main points of data collection, storage, and usage of personal data. Since Al systems need a large amount of data work properly, consequently, there is a probability that personal information might be inappropriately used or poorly secured and maybe even accessed by people who are not authorized.

➢ *The Need for Large Datasets for Training:*

Al models, particularly those that leverage machine learning, need a vast amount of data to detect trends and supply the right forecasts. The amount and the quality of the data used for training are paramount to Al's performance. Nevertheless, collecting all there data set can be cumbersome, especially when, in a worse scenario, personal information would get compromised.

➢ *Potential for Adversarial Attacks:*

An adversarial attack is technique in which the input data is tampered with to make Al systems give wrong predictions or classifications. A case in point is highlighting a small difference in an image that may cause an Al model to identify the image incorrectly. It is a typical scenario that show that Al system are not 100% secure and reliable, and therefore, it becomes a concern of the safety and security of these systems, especially in the areas of facial recognition and self-driving automobile.

➢ *Interpretability of Models:*

Most AI models, especially deep learning ones, considered "black boxes." The reasoning behind deducting these to be is abstruse, however. Depending on how the models are classified, the users and security people might be able to clarify the preferred way of classification or not, thus it leads to people wondering whether or not they can trust these systems.

➢ *Resource Intensive:*

The phony email filtering AI system introduction may be a very slow and costly project that requires large costs of time and money. Large data sets and powerful processors are required to train the AI. Nevertheless, small firms may get in trouble while completing the project due to the resources needed for the development and maintenance of AI systems.

F. Future Trends and Innovations

Future trends in AI for phishing detection include the integration of real-time threat intelligence, improve natural language processing capabilities, and the development of more sophisticated algorithms that adapt to emerging phishing techniques (Alzahrani & Alhassan, 2023).

➢ *Integration of Real-Time Threat Intelligence:*

That is, to utilize the latest information retrieved about phishing fingerprints to recognize scams as they appear to become.

➢ *Improved comprehension of human language:*

We refer to the technology that helps computers better understand and analyze human language, making it easier for them to detect phishing emails that use deceptive language.

➢ *Creating Smarter Algorithms:*

That is to create more intelligent programs that, with time, smartly discover and recognize different new phishing attack types as they arise.

ISSN No:-2456-2165 https://doi.org/10.5281/zenodo.14565177

G. Ethical and Privacy Consideration

The integration of AI in email security systems should not only be a testimony of the productivity but also the ethical side of it must also be given due consideration. Thus, the companies have to look into the following ethical and privacy issues:

- *User Consent:*

Users need to be provided with the privilege of deciding whether to share information that is relevant to them or not before any data collection and analysis takes place. Besides, the company should very clearly articulate the data they will collect and the purpose of its use in phishing detection techniques.

- *Bias in AI Models:*

AI models are often programmed to depict the bias embedded in the training data of AI models. To do this, organizations should effectively make sure that the models they use are being trained on more diverse datasets, as it will reduce bias and promote the fair treatment of users of different types.

- *Transparency and Accountability:*

Organizations should opt for transparency in their AI systems, providing users with complete information about their data handling and decision-making. Building accountability systems is one of the measures by which AI-based solutions can gain the trust of the users.

III. CONCLUSIONS

Artificial intelligence has an essential role in the detection and blockage of phishing emails, giving organizations advanced tools against this constant threat. AI-based solutions can help organizations improve their email security and thus lower phishing attacks and the possible outcomes of these attacks. Nevertheless, it is important to be aware of the ethical issues and problems that come with AI technologies. The deployment of AI in phishing detection raises ethical concerns, particularly regarding user privacy and data handling. Organizations must ensure compliance with regulations such as GDPR while maintaining effective security measures (Cavoukian, 2019).

REFERENCES

[1]. Eze, C.S., & Shamir, L. (2024). “Analysis and prevention of AI-Based Phishing Email Attacks.” Electronics. 13(10), 1839.
[2]. Jahankhani, M., Mavridis, P., & Alazab, M. (2020). Phishing Attacks: A Survey of the current Trends and Future Directions Journal of Cybersecurity and Privacy, 1 (3), 152-171.
[3]. Bertino, E., & Islam, N. (2018). Botnets and Internet of Things Security. Computer Security, 76, 77-89.
[4]. Sharma, A., & Gupta, R. (2021). Machine Learning Techniques for Phishing Detection: A Review. International Journal of Information Security, 20(3), 329-346.
[5]. Zhang, Y., & Wang, T. (2020). Deep Learning for Phishing Detection: A Survey. Computers & Security, 92, 101737.
[6]. Sarker, I.H., & Sultana, N. (2021). A Survey on Phishing Detection Techniques: Current Status and Future Directions. Journal of Network and Computer Applications, 177, 102924.
[7]. Wang, Y., & Li, J. (2020). Challenges and Opportunities in Phishing Detection Using Machine Learning. IEEE Access, 8, 124123-124134.
[8]. Alzahrani, A., & Alhassan, I. (2023). The Role of Artificial Intelligence in Detecting and Preventing Cyber and Phishing Attacks. ResearchGate.
[9]. Cavoukian, A., (2019). Privacy by Design: The 7 Foundational Principles. Information and Privacy Commissioner of Ontario.