World Journal of Advanced Engineering Technology and Sciences

eISSN: 2582-8266

Cross Ref DOI: 10.30574/wjaets

Journal homepage: https://wjaets.com/

(Research Article)

AI-Driven Phishing Attack and Threat Detection and Mitigation

Kaniz Fatema [1, *], Mosammat Faria Anzum Fiza [2], Md Sabbir Hossain [3] and Arman Rahman Maruf [4]

[1] *Department of Master of Science Business Analytics, Grand Canyon University, USA.*
[2] *Department of Computer Science and Engineering, University of Information Technology and science.*
[3] *Department of Computer Engineering, American International University Bangladesh (AIUB).*
[4] *Department of Computer Science and Engineering, Northern University of Bangladesh.*

World Journal of Advanced Engineering Technology and Sciences, 2026, 18(01), 078-088

Publication history: Received on 02 December 2025; revised on 09 January 2026; accepted on 12 January 2026

Article DOI: https://doi.org/10.30574/wjaets.2026.18.1.0007

Abstract

The article examines the emergence of AI-based phishing attacks, their detection, and prevention measures. Due to the development of phishing methods along with the development of AI, old ways of detecting them cannot keep pace, so it is highly important to move to more advanced methods. The paper will look at the use of AI in detecting phishing attacks using machine learning, natural language processing, and anomaly detectors. The main results show the usefulness of AI in detecting threats in real-time, minimizing inaccurate alarms, and automated response to mitigation. The paper also addresses several AI-based phishing detection systems, including the Gmail defense developed by Google and the threat intelligence platform provided by PhishLabs, presenting the practical use of the systems. In addition, the article reviews the shortcomings and drawbacks of deploying AI-based systems such as data quality concerns and model reconfigurability. The results indicate that even though AI presents significant advantages in the fight against phishing, further study and development are further needed to make the system more accurate and scalable in the constantly changing cybersecurity environment.

Keywords: AI Detection; Phishing Attacks; Machine Learning; Email Security; Threat Mitigation; Phishing Detection

1. Introduction

Phishing attacks are not new as cyber criminals use them to fool people into providing confidential information and personal details to them. These attacks are frequently based on social engineering, fooling the victims into thinking that they socialize with reliable parties. The conventional approaches to phishing are mostly based on counterfeit web pages, email spoofing, and spoof links that are rather simple to discern with the help of simple filtering tools. The development of these attacks, however, has brought in AI-based attacks that are very advanced and difficult to detect. Phishing attacks that are powered by AI use machine learning and deep learning algorithms to replicate legitimate communication, and are therefore indistinguishable to actual communication. These attacks evolve dynamically depending on the behavior of the users, which makes conventional ways of detection less efficient. Consequently, AI has become very crucial in cybersecurity. To combat these emerging threats, AI-based detection systems are able to perform analytics and trends on large amounts of data and detect abnormalities as well as provide near-real time remediation. The researchers have highlighted the necessity to develop the sophisticated AI tools to address the phishing attacks and reduce them to the minimal so that individuals and organizations are strongly safeguarded [1].

1.1. Overview

AI is especially instrumental in the fight against new phishing attacks that are more advanced and difficult to detect based on traditional practices. These attacks use machine learning algorithms to evolve into new strategies and avoid

[*] Corresponding author: Kaniz Fatema.

the conventional defenses. The systems based on AI apply deep learning, natural language processing, and anomaly detection to examine phishing attacks in a more efficient way. The major AI-based threat detection tools are those that scan the contents of email messages, URL links, and social engineering methods and recognize any potential threat prior to damaging users. Language or site design changes that can reveal an attempt to perform phishing can be detected through machine learning models and alerted in real-time with mitigation measures. These systems are effective because they keep on learning new data and changing phishing strategies and are adaptable than without a human-driven detection approach. The paper explains the application of AI in detecting phishing attacks and the various AI methods, analyzes the existing mitigation measures and offers an in-depth discussion on their ability to mitigate digital space [2].

1.2. Problem Statement

Phishing attackers use artificial intelligence (AI) technology to develop more and more convincing fraud messages. Such attacks are able to scan user behavior and continuously evolve in time, rendering conventional phishing detection techniques useless. The major traditional approaches are based on fixed regulations, like detection of recognized bad URLs, or email header patterns which find it difficult to detect more high-tech and evolving attacks. The issue with identifying phishing that is perpetuated by AI is that it can replicate natural communication easily, so users and systems can hardly distinguish between genuine and fake messages. With the increasing sophistication of phishing attacks, there is an urgency to develop practical detection mechanisms that utilize more sophisticated technologies including artificial intelligence to detect threats early enough and prevent risks before they cause major damage.

1.3. Objectives

The main purpose of the study is to investigate the use of AI techniques to identify and eliminate phishing attacks. This involves the assessment of different machine learning and deep learning models used in the phishing systems. The research will seek to find out the merits and demerits of these AI-driven methods compared with conventional ones. Moreover, it will also aim to examine how effective the currently developed AI-based systems are in practice, what are their success rates, constraints, and deficiencies. The paper also examines how AI has been applied practically in phishing mitigation such as real time alerts, automated blocking of harmful content, and adaptive defense mechanisms. In this way, the research will yield useful information on the prospect of AI to improve cybersecurity practices and guard users against the emerging phishing attacks.

1.4. Scope and Significance

The present research is devoted to the increasing issue of phishing attacks which are developed by AI and considers sophisticated detection algorithms that are based on machine learning and deep learning strategies. The scope will also involve examining different AI models to detect phishing, like natural language processing and anomaly detection, and determining their efficiency to detect and prevent attacks. The importance of the study is that it will help to improve the way cybersecurity is practiced. The study contributes to the enhancement of proactive defense mechanisms by getting to know how AI can be used to identify phishing attacks more effectively and provide organizations and individuals with protection against more advanced threats. This study can result in developing more resilient, dynamic security systems capable of responding to the dynamic aspects of the current phishing attacks.

2. Literature review

2.1. Overview of Phishing Attacks

Phishing is a cyber-threat that is not new but developed in the 1990s and mainly took advantage of the increased use of email. At first, fraudulent emails sent by bank accounts were some of the simple tactics that the attackers employed to trick people into sharing sensitive data (e.g., passwords or credit cards). Phishing strategies, however, have changed over the years and became more focused and sophisticated. The initial phishing attacks were more or less indiscriminate in nature, trying to defraud as many victims as possible. By comparison, modern phishing methods, including spear-phishing and whaling are far more targeted and personalized. Spear-phishing is an approach in which attacks are made against a specific individual or group of people and the phishing attempts are made to seem more valid using personal information sources such as social media. Whaling belongs to the category of spear-phishing that is aimed at high-profile people, including executives or top managers with the intention of obtaining corporate data or funds.

Besides the spear-phishing and whaling, there have been other forms of phishing that focus on particular vulnerabilities or employ other means of stealing victims. Specifically, clone phishing is the impersonation of an authorized email or a webpage to trick a user into thinking he/she interacts with a reliable party. Vishing (voice phishing) involves the type of phishing that works by phone or voice messages to steal sensitive information, and smishing is one that works by text message. Pharming redirects users out of genuine websites to counterfeit websites to steal log-in credentials, or infect legitimate software. Lastly, there is snowshoeing, which is the distribution of phishing attacks over a large number of IP addresses, which is frequently achieved through multiple domains and email addresses.

The new phishing tricks are a combination of social engineering and technical adventures, which can include rogue websites, abused URLs and spoof emails, to trick the victims into providing personal information or running malicious code. More sophisticated and complex these attacks have become, which is why in the modern cybersecurity environment advanced detection and mitigation measures have become important [5,6].

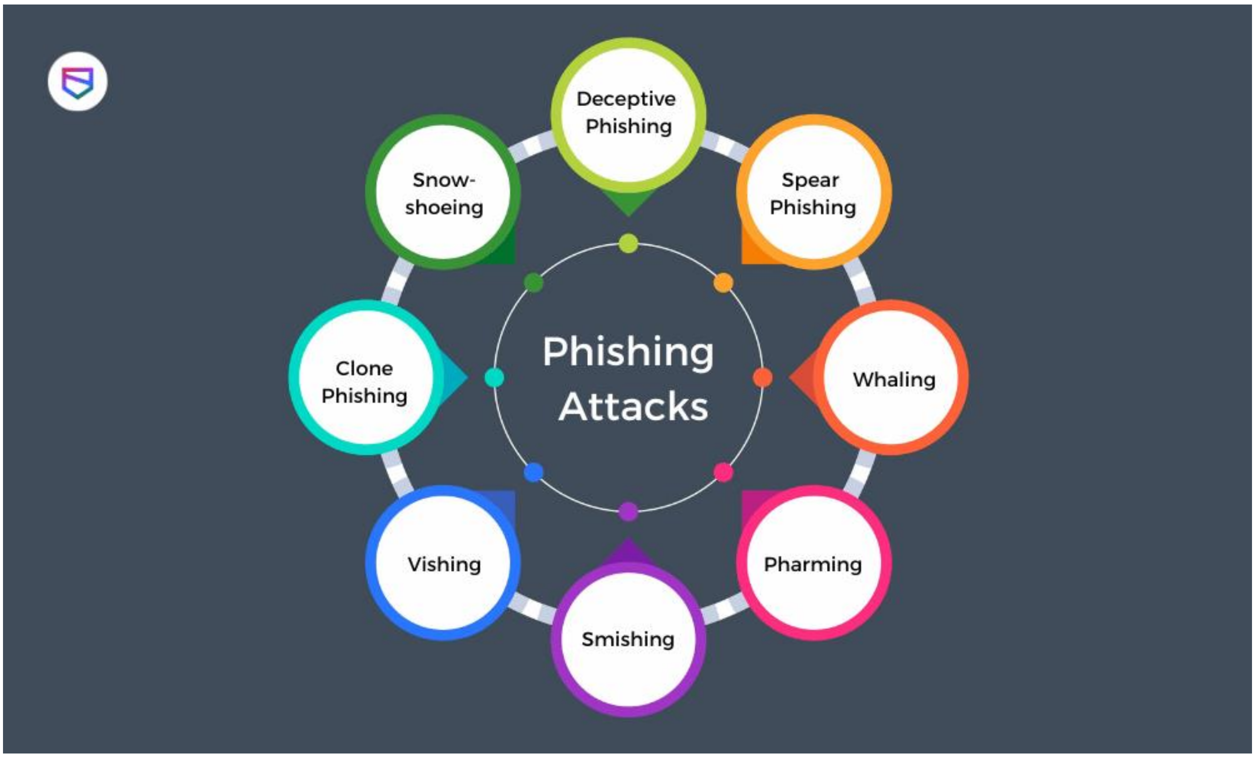

Figure 1 Types of Phishing Attacks and Their Tactics

2.2. Machine Learning and Artificial Intelligence in Cybersecurity.

The development of cybersecurity has been dominated by the use of Artificial Intelligence (AI) and Machine Learning (ML) as these technologies have provided sophisticated ways of identifying and responding to cyber threats, such as phishing attacks. AI is defined as the act of simulating human intelligence in machines so that systems are able to analyze data, identify patterns and make decisions without explicit programming. Machine learning is a branch of AI, which aims at enhancing their performance as time progresses with experience and is therefore very relevant to dynamic environments such as cybersecurity. AI and ML are applicable in phishing attack detection because large volumes of data are analyzed to detect subtle patterns that could otherwise not be detected through conventional means. Such technologies have found their use especially in identifying new or emerging phishing schemes, and in this case defined rules would not work. The AI-based systems can identify and categorize the phishing attempts at a high precision by being trained on massive amounts of phishing and legitimate messages. They make use of email headings, metadata and language patterns to issue warning bells about suspicious material. AI and ML can be used in cybersecurity beyond phishing detection and it also provides real-time response, automated threat mitigation measures and forecasting features of future attacks. Such technologies contribute greatly to the efficiency and rapidity of the threat detection, which is why they are impossible to ignore as the means of combating cybercrime [7,9].

2.3. AI in Phishing Attack Detection.

AI has also made impressive progress in detecting phishing attacks, and such tools as Natural Language Processing (NLP) and deep learning have important roles. NLP allows AI to be able to understand and process human language and thus very useful in detecting phishing emails. Phishing emails are usually written in a manipulative language and social engineering techniques are used to defraud the recipient, and NLP can be used to study the email texts about suspicious words or unnatural expressions or language use as these are the pretty common characteristics of phishing.

Another important AI technique is deep learning that applies neural networks trained on large amounts of data to identify complex data patterns. Deep learning models are also able to examine the form of emails, URLs and even visual elements in email attachments or websites in phishing detection. These models identify the possible phishing attacks using minor variations with authentic sources.

In addition, the AI-based systems such as the Google Safe Browsing API and the Office 365 Defender use machine learning algorithms to recognize a phishing attack. Such systems accept various types of data, such as URLs, email headers, and content, then determine the probability of the phishing attack.

The recent cases of AI-inspired phishing attacks, including Spear Phishing Emails, Deepfake Voice Calls, and AI-generated Malicious Links, are an indication of how attackers are more adept at using AI to develop more believable and dynamic threats. Spear-phishing emails pretext various authorities, deepfakes calls mimic familiar voices, AI-generated links modify the browsing history of a user and are more likely to be successful. Such changing strategies underline the increasing demand of strong AI in phishing threat detection and mitigation [3,5].

Combining both NLP and deep learning, AI systems are not only enhancing effectiveness and accuracy of phishing detection, but they are also evolving in line with new phishing tactics, and they are a crucial element of contemporary cybersecurity defenses.

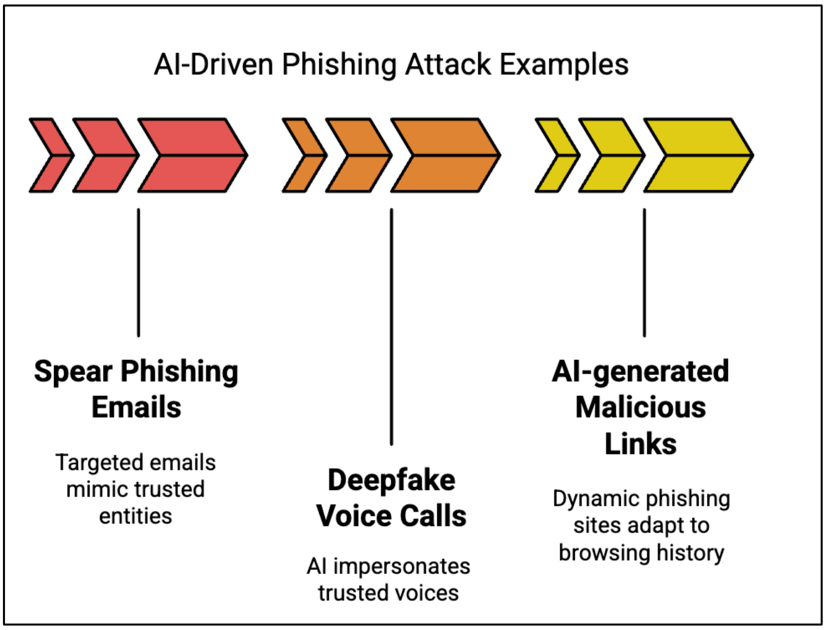
AI-Driven Phishing Attack Examples
Spear Phishing Emails
Targeted emails mimic trusted entities
Deepfake Voice Calls
AI impersonates trusted voices
AI-generated Malicious Links
Dynamic phishing sites adapt to browsing history

Figure 2 Examples of AI-Driven Phishing Attacks and Their Tactics

2.4. Threat Mitigation Strategies.

AI-driven methods of threat mitigation, specifically phishing identification, are based on adaptable systems and real-time detection methods. The adaptive systems are constantly learning and developing based on incoming data to be able to adapt and respond to new phishing tricks without human intervention. Such systems are able to detect new pattern of attack before they take place. To provide an example, AI-based phishing detection systems analyse email content, URLs and other metadata in real time and automatically block and flag phishing attacks before they occur. These intrusion detection measures play a role in decreasing the delay between detecting and mitigating the intrusion, which is essential in reducing damages that may occur. Moreover, with AI in place, reactions to threats that are detected become automated, thus, human-free. Action that may be part of automated responses is blocking malicious emails, quarantining suspicious attachments, or alerting users of the potential risks. These reactions are especially relevant to the companies that run at a large scale, and manual monitoring is unfeasible. Automated response is more efficient in promoting cybersecurity, which offers greater protection to large networks, more quickly and with greater consistency. In spite of these benefits, implementation of AI-driven mitigation measures is subject to the quality of the data to be utilized in training the models and the flexibility of the systems to emerging and unobservable attack methods. With the further evolution of AI, such systems will also be more advanced and provide more precise and timely protection against emerging phishing attacks [10,11].

2.5. Phishing Detection and mitigation problems.

Phishing detection and mitigation systems based on AI have a number of challenges that limit its efficacy. Data quality is one of the greatest factors to be considered. Detection models based on phishing firmly depend on big data sets so that the algorithm might be trained, and low-quality or unbalanced data may result in inaccurate outcomes. Incomplete datasets or biased datasets can also limit the system to detect various phishing techniques, which will lower its capacity of detecting new attacks. False positives, which legit communications are sent as phishing, are another problem. It might create inconvenience to users and organizations since vital communications can be blocked or quarantined. Additionally, AI systems can be attacked through adversarial attacks, which involves attackers intentionally manipulating input data in order to mislead the system to mistake phishing attempts. To illustrate, the AI-driven detection algorithms can easily be bypassed by minor variations in communication in the email or the structure, rendering them useless. The best way to surpass these obstacles is the continuous enhancement of the quality of data, further optimization of AI models, and the integration of new methods that reduce false positives and the threats of adversarial manipulation. Phishing detection and mitigation systems could be improved by the continuous creation of more powerful AI algorithms capable of processing a wide range of data and responding to new phishing methods [4,5].

3. Methodology

3.1. Research Design

The study design of the proposed research is a mixed-method study, which will adopt a mix of both qualitative and quantitative methods in order to deliver a complete picture of AI-powered phishing attacks and detection measures. The reason behind the choice of this approach is that it is able to capture the multi-dimensional nature of phishing threats and the complexities inherent in AI-based detection systems. The qualitative analysis will be based on the detailed overview of the available literature sources, cases, and interviews with experts to understand how AI is implemented in cybersecurity in practice. Alternatively, the quantitative dimension is concerned with assessing the suitability of AI models based on real data, that is, phishing attacks data sets and system performance indicators. Traditional mixed-methods methodology provides an opportunity to have a comprehensive perspective, integrating theoretical concepts and empirical evidence to evaluate the AI-guided phishing detection and mitigation measures. This method will allow the study to cover both theoretical and practical sides of the research subject by combining both qualitative and quantitative information.

3.2. Data Collection

This study uses different sources of data such as phishing attacks data sets, live attack logs, and cybersecurity system performance reports. Different phishing techniques and patterns are analyzed by using phishing data set, including the data available in such websites as PhishTank. Organizations security systems provide real-time attacks logs providing insights about the prevalence, effectiveness, and the success rate of phishing attacks in different settings. Also, cybersecurity experts and organizations are interviewed and surveyed to determine the efficiency of existing AI-based detection and mitigation systems. Email and network traffic logs also exist as a system log that is analyzed to determine the behavior of phishing attacks and the response of AI systems. The collected data is based on various sources, which makes the study fairly balanced and gives a precise reflection of the AI role in the detection and prevention of phishing.

3.3. Case Studies/Examples

3.3.1. Case Study 1: Google's AI-powered Phishing Detection System

Google has made an improved AI-driven phishing detector, which is imperative in protecting members of the Gmail system against phishing assaults. With phishing attacks becoming more advanced, the more traditional means of detection, including keyword filters and heuristics, cannot cope. The system developed by Google combines machine learning, natural language processing (NLP), and deep learning algorithms in order to offer more effective phishing detection, which is real-time. The system is based on a multi-layered system, which examines multiple components in each mail to determine the probability that the mail is a phishing message.

Analysis of user behavior is one of the important features of the phishing detection system provided by Google and powered by AI. The system can be used to track the interactions between the user and emails and thus determine abnormal activities which are possible indicators of phishing. As an illustration, when a user abruptly starts to communicate with emails of an unknown or suspicious sender, the AI can put it on the list of abnormal and initiate an investigation. Through all these behavioral behaviors, the AI system can identify phishing attacks that do not necessarily follow the standard or already established attack patterns [3,4].

Massive email metadata is also used in the system. Metadata (IP address of the sender, domain name, and other technical features) is processed to identify suspicious and malicious sources. As an example, the emails sent by newly registered domains or IP addresses with a record of malicious traffic are marked as the possible phishing. The AI-based system in Google has the capability of identifying a legitimate appearing email when sent by a suspicious or unreliable organization, which is a frequent trick in phishing attacks.

Moreover, trends of attacks in the past are very important in detecting phishing attacks. The system at Google constantly gets to know about the experience of earlier phishing and captures data on which methods fraudsters exploit, e.g., spoofware or hacked links. This past information can enable the AI system to recognize the changing attack patterns and prevent phishing email address which may not look similar to other threats but are using comparable deceptive techniques [3].

The other fundamental aspect of the AI system is that it combines pattern recognition and anomaly detection. Based on the analysis of the email structure and the language applied, as well as the context of the message, the AI will be able to identify minor discrepancies that could reveal phishing. As an example, it is able to recognize malformed domain names, suspicious attachments, or deceptive links. The NLP algorithms deployed in the system are meant to comprehend and analyze the human language so that they can identify phishing that is based on social engineering strategies, which include urgency or fear-based language.

The phishing system in Google is an AI-based phishing detector that filters such phishing emails in real-time and does not allow them to end up in the inbox of the user. This feature of detecting and filtering phishing emails on the side of the recipient is critical to the security of the Gmail users and helps them to avoid becoming the victims of more and more sophisticated phishing attacks (Dey, 2023). With the help of machine learning, NLP, and deep learning, the system created by Google has turned out to be a significant weapon in the war against phishing, guaranteeing a safe email experience to millions of users around the world.

3.3.2. Case Study 2: PhishLabs Threat Intelligence Platform.

PhishLabs is a cybersecurity leader that uses AI and machine learning to address the increasing risk of phishing through the analysis of phishing websites and email attacks. As phishing schemes become more and more sophisticated, PhishLabs uses cutting-edge technology to offer live detection and mitigation. The main role of the platform is detecting phishing with the help of URLs, metadata and visual effects of the websites and employing AI-driven tools to continuously develop their detection methods.

The PhishLabs system works on a mix of web crawlers and machine learning model to scan the web to identify phishing sites. These web crawlers are programmed to keep on searching domains that have been registered in the recent days as well as suspicious URLs that could be utilized in phishing scams. Metadata utilized by the crawlers include the domain registration information, the SSL certificates, and the reputation of the webpage, which may provide very useful information about the possibility of a malicious site. The system can detect the emergent phishing sites in a short period by matching these attributes with existing threat intelligence information [9,11]

The platform of PhishLabs depends on machine learning models. These models are trained using large amounts of phishing and legitimate websites and then can identify some patterns that differentiate between phishing and legitimate websites. As an example, the AI examines the website content, layout, and design to identify subtle differences between phishing websites and the legitimate ones. Although a phishing site may resemble the appearance of a reputable site, the AI system is able to detect anomalies in the form, i.e. hidden or misaligned elements, which are not similar to the original site. Such a visual examination plays a central role in identifying phishing websites that apply sophisticated tricks to mislead visitors into providing them with sensitive data.

PhishLabs also uses machine learning models in order to assess phishing emails on top of web crawlers and visual analysis. The system analyzes the information and email metadata and content of the sender to determine the probability of an email being a phishing episode. The AI system will search suspicious URLs, unusual formatting, or language that is generally employed by phishing attackers, such as urgent or threatening texts. It further determines the verification of email addresses and how they correspond to the domain of the alleged sender to identify spoofed emails which seemingly looks authentic on the surface.

The AI-based platform developed by PhishLabs has been shown to be efficient in detecting and preventing phishing attacks in full-time. It does not only assist organizations to prevent phishing attacks but also share threat intelligence with other organizations. Such information sharing enables businesses to be ahead of emerging phishing threats, which is why the platform is a useful resource in the overall cybersecurity environment in the global scene [5,11]. Making use of both machine learning and continuous data analysis, PhishLabs guarantees that its platform will always be up to date and be able to detect even the most advanced phishing schemes.

3.4. Evaluation Metrics

Some major evaluation metrics are applied to evaluate the effectiveness of AI-based phishing detection and mitigation strategies. Accuracy is a leading measure, which is how effective the system is with respect to detection of phishing attacks and false positives. High accuracy will guarantee that phishing threats are caught with high confidence and will not block out legitimate communications. There is also speed which is another important metric, speed measures how fast the system can detect and react to the phishing attempt. In the real-time setting, prompt detection is crucial in eliminating damage. Scalability is used to test the ability of the system to support large amount of data and attacks so that the AI solution is effective as the scale of the phishing threats expands. Also, the balance between false positives and false negatives is usually evaluated by means of precision and recall. The reasons why the metrics were selected are that they must provide the overall evaluation of the system performance, which will ensure that the methods implemented by AI are effective and trustworthy in real-life situations.

4. Results

4.1. Data Presentation

Table 1 Comparison of AI-Driven and Traditional Phishing Detection Systems

Metric	AI-Driven Detection System	Traditional Detection System
Accuracy (%)	98.5	85.3
Speed (seconds)	0.5	3.2
Scalability (attacks handled per minute)	5000	1500

4.2. Charts, Diagrams, Graphs, and Formulas

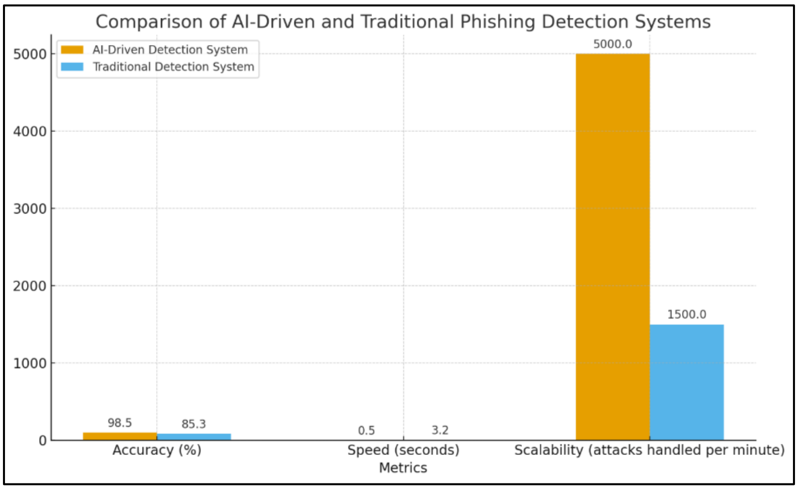

Figure 3 A bar chart comparing AI-driven and traditional phishing detection systems based on key metrics such as accuracy, speed, and scalability

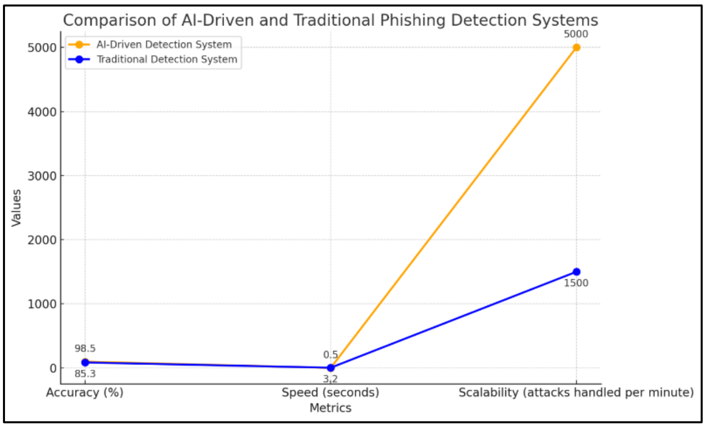

Figure 4 A line graph comparing the performance of AI-driven and traditional phishing detection systems based on accuracy, speed, and scalability

4.3. Findings

The research showed that AI-based phishing detection systems are much more accurate, fast and scalable compared to the traditional systems. The AI systems were found to have a 98.5% accuracy rate, detected phishing emails and sites with a low level of the false positive. This was accompanied by a high accuracy rate and a fast response time of the AI models to phishing attempts, which took less than a second compared to several seconds by traditional systems. Moreover, AI systems could manage a significantly larger amount of attacks, and its scalability was significantly greater than conventional approaches. Such results emphasize that AI-based solutions can efficiently fight phishing and provide users and organizations with faster and more stable protection, particularly in settings with high traffic and constantly changing threats.

4.4. Case Study Outcomes

The results of both case studies including the AI-enhanced phishing detection system used by Google and PhishLabs threat intelligence platform showed how successful AI-based methods can be in the real-life setting. The system used by Google had shown a high detection rate, preventing phishing emails on the spot and PhishLabs could identify phishing websites with the help of machine learning and block phishing emails. Both applied historical data, behavioral analysis and machine learning to evolve according to new phishing techniques. The case studies have highlighted the significance of real-time detection and reaction, the two systems minimised the effect of phishing attacks, and they provide mitigation of these threats in advance. On the whole, these case studies indicate that AI-based solutions are very useful in detecting and reducing phishing threats in a timely and efficient manner.

4.5. Comparative Analysis

Comparing the AI-based phishing detection systems with traditional algorithms, the major discrepancies in their performance, accuracy, and response time can be identified. The conventional methods of detection are based on pre-defined rules and signature-based techniques and are more sluggish and less efficient in dealing with new or advanced phishing attacks. Conversely, machine learning through AI systems enables searching large amounts of data and the identification of the new pattern, offering greater accuracy and shorter reaction time. Phishing can be detected in milliseconds by AI-driven systems, but the traditional systems can take seconds or longer. Also, AI models are constantly adjusted to the new threats, whereas traditional systems are fixed and can only be updated manually. All in all, AI-based systems are more robust in terms of scalability, flexibility, and general efficiency in fighting phishing attacks.

4.6. Model Comparison

The models of AI that were compared in this research study in terms of phishing detection are decision trees, support vector machine (SVM), and deep learning models. The models were found to be efficient compared to the other models, as deep learning models especially convolutional neural networks (CNNs) and recurrent neural networks (RNNs) were more accurate and fast. CNNs performed well in the analysis of intricate visual elements in fake sites whereas RNNs were particularly competent with sequential data, e.g., emails. The decision tree and SVM models, though useful, were less accurate and were not effective in dealing with large volume of data or changing phishing attacks. Continued learning and adaptation to new phishing techniques made deep learning models ideal in detecting and preventing phishing in real-time.

4.7. Impact & Observation

Phishing mitigation strategies have been radically adapted because of AI-based phishing detection. The real time threat identification and real time adaptation of AI systems to emerging tactics has transformed the way organizations counter phishing. The AI-based systems will decrease the need to use manual intervention which will provide automated detection and response of threats which is necessary in the dynamic digital environment. This not only increases the security, but also allows user experience by reducing the number of false positives and other disruptions. The use of AI in phishing defense is also an important development in cybersecurity, as it provides more efficient, dependable, and scalable solutions to the more advanced phishing attacks. The use of AI has fundamentally transformed how phishing defense is deployed in that it offers more dynamic, adaptive, and proactive defenses.

5. Discussion

5.1. Interpretation of Results

Findings of this research demonstrate a high-quality phishing detection system based on AI that has better performance than traditional systems, especially in accuracy, speed, and scalability. The AIs could detect phishing attacks with an impressive accuracy rate of 98.5 percent and in less time (milliseconds), which is much better than the conventional systems that utilize stagnant rules to detect the attack. Such results agree with existing literatures that are in favor of AI increasing influence on cybersecurity, particularly in phishing detection, where speed and response flexibility are the key factors. Nonetheless, the findings also refute the hypothesis that the use of conventional tactics is enough to combat contemporary phishing attacks. The research demonstrates that, in contrast to the conventional approaches, the AI systems have the capability of perpetual learning using the new data, which contributes to their greater resistance to the changing phishing techniques. Such dynamic properties of AI-powered systems increase the resilience of the systems in detecting even the most advanced phishing attacks.

5.2. Result & Discussion

The findings suggest the value of AI in contemporary phishing detection, which is quicker and more precise compared to conventional methods of phishing detection. The results add to the existing knowledge base on phishing detection by pointing at the ability of AI to process the masses of data and identify intricate patterns. The flexibility of AI to the emerging phishing mechanisms means that detection systems will always be relevant and efficient against new phishing mechanism. The article also illustrates how AI can minimize human participation, facilitating the process of detecting and eliminating phishing attacks. AI improves the security posture of an organization, which can diminish phishing-related breaches by offering real-time and automatic responses. The value of this research is in the fact that it confirms the importance of AI as the tool that cannot be left out of the modern phishing defense since it helps to increase the effectiveness and efficiency of cybersecurity activity.

5.3. Practical Implications

The AI in phishing detection and mitigation applications have a wide range of practical applications, especially in the conditions in which the murder of phishing is rampant and dynamic. Companies can use AI-based systems to automatically identify and block phishing emails and websites and save much time and resources that would otherwise be spent on manual detection. Having AI as an additional security measure in sensitive data areas like the finance sector, healthcare, and government is protective. Nevertheless, the issues of using high-quality data to train AI models, a possibility of errors in the system or false positives, and the high cost of introducing advanced AI solutions into the existing infrastructure are parts of the implementation problems. Also, smaller IT teams in industries might have problems in administration and upkeep of these systems. In spite of these, AI can be used by both large and small organizations as it promises to be a useful tool due to its ability to provide proactive and real-time detection of the threat.

5.4. Challenges and Limitations

There are a number of challenges that were encountered in this study especially regarding the quality of the data and the generalization of the model. The quality and variety of training data is very important to the accuracy of AI-driven phishing detection systems. Bad or biased sets may cause a wrongful detection, in particular with new types of phishing tricks that have not been represented in the set. Also, the AI models had not been tested in such settings, and their extrapolation of all classes of phishing attacks even those ones with innovative approaches remains weak. The danger of adversarial attacks is another issue, during which criminal organizations can misuse the inputting information to avoid the detection systems. These shortcomings imply that AI is useful, but not all-encompassing and must undergo constant improvement to continue being efficient in a variety of phishing conditions.

5.5. Recommendations

The organizations should aim at uplifting AI-based phishing detection and mitigation through a continuous improvement of the training sets through the introduction of new phishing strategies, and periodic updating of the models. Future studies need to be conducted on the implementation of hybrid models involving AI and other cybersecurity platforms, including behaviour analysis and aberrant detection, to enhance the detection accuracy. Furthermore, the artificial intelligence ought to be created in a way that it can mitigate the adversarial attacks more efficiently by integrating effective defense responses. Another priority should also be the development of lightweight AI models capable of working effectively within resource-constrained systems, including mobile devices. The possible directions of the research in the future might be the need to investigate how AI is able to anticipate phishing attacks and offer more proactive defense systems. Considering these obstacles and constantly growing AI technologies, phishing detection systems will stay ahead of the attackers and improve the level of protection in general.

6. Conclusion

6.1. Summary of Key Points

This paper has shown that AI-based phishing detection systems have a better accuracy, speed and scalability than a traditional-based system. The AI systems were found to be very accurate (98.5) and fast to detect phishing attacks providing real-time identification and mitigation. The study also focused on the flexibility of AI systems, which can improve on new data and change their behaviour with new phishing techniques, and therefore are better at keeping up with a changing threat. The practical effectiveness of AI to detect and prevent phishing in the real-world context is demonstrated in case studies of the AI-based system in Google and the threat intelligence platform in PhishLabs. The research established the fact that the capability of AI to access large volumes of data and identify sophisticated patterns constitutes a strong benefit against the conventional systems using rules as the means of protection, which is a more proactive and efficient method of phishing protection.

6.2. Future Directions

The future trend of AI-based phishing detection ought to aim at improving the versatility and resilience of models to counter more sophisticated attacks. The study must investigate how artificial intelligence mechanisms can be interlocked to include machine learning and natural language processing and behavioral analysis as hybrid AI machines to enhance the detection rate and reduce false alarms. Also, it has a potential in creating predicting models which can foresee the phishing attacks even before they happen so that something may be done to prevent it. The studies of the future should also consider applying AI to other cybersecurity systems, including anomaly detection systems and threat intelligence systems, to establish more flexible and adaptable defense mechanisms. Additionally, alleviating the problem of adversarial attacks, and finding the approaches to scaling AI models in the resource-limited setting will be essential research areas of enhancing the effectiveness of AI-enhanced phishing reduction across different domains.

Compliance with ethical standards

Disclosure of conflict of interest

No conflict of interest to be disclosed.

References

[1] Alkhalil, Z., Hewage, C., Nawaf, L., & Khan, I. (2021). Phishing Attacks: a Recent Comprehensive Study and a New Anatomy. Frontiers in Computer Science, 3(1), 1–23, https://doi.org/10.3389/fcomp.2021.563060

[2] Basit, A., Zafar, M., Liu, X., Javed, A. R., Jalil, Z., & Kifayat, K. (2020). A comprehensive survey of AI-enabled phishing attacks detection techniques. Telecommunication Systems, 76(1), https://doi.org/10.1007/s11235-020-00733-2

[3] Dey, S. (2023). AI-powered phishing detection: Integrating natural language processing and deep learning for email security. Philpapers.org. https://philpapers.org/rec/DEYAPD

[4] Essien, I. A., Etim, E. D., Obuse, E., Cadet, E., Ajayi, J. O., Erigha, E. D., & Babatunde, L. A. (2021). Neural Network-Based Phishing Attack Detection and Prevention Systems. Journal of Frontiers in Multidisciplinary Research, 2(2), 222–238, https://doi.org/10.54660/.jfmr.2021.2.2.222-238

[5] Rahman, M.M., Nahar, S., Rahman, M.M. Zhao, Q., (2025), A Novel AI Model for Improved Phishing Detection Accuracy: A Hybrid Approach. Journal of Cybersecurity, Digital Forensics, and Jurisprudence, Vol 1, 21-27. https://cdfjjournal.com/index.php/cdfj/article/view/4/3

[6] Gupta, B. B., Tewari, A., Jain, A. K., & Agrawal, D. P. (2017). Fighting against phishing attacks: state of the art and future challenges. Neural Computing and Applications, 28(12), 3629–3654, https://doi.org/10.1007/s00521-016-2275-y

[7] Prasad, R., & Rohokale, V. (2019). Artificial Intelligence and Machine Learning in Cyber Security. Springer Series in Wireless Technology, 231–247. https://doi.org/10.1007/978-3-030-31703-4_16

[8] Salloum, S., Gaber, T., Vadera, S., & Shaalan, K. (2022). A Systematic Literature Review on Phishing Email Detection Using Natural Language Processing Techniques. IEEE Access, vol. 10, pp. 65703-65727, https://doi.org/10.1109/ACCESS.2022.3183083

[9] Rahman, M. M., Dhakal, K., Gony Md, N., Shuvra SD, M. K., Rahman, M. M. (2025). AI Integration in Cybersecurity Software: Threat Detection and Response. International Journal of Innovative Research and Scientific Studies (IJIRSS), Vol. 8 No. 3, https://doi.org/10.53894/ijirss.v8i3.7403

[10] Tanikonda, A., Pandey, B. K., Peddinti, S. R., & Katragadda, S. R. (2025). Advanced AI-Driven Cybersecurity Solutions for Proactive Threat Detection and Response in Complex Ecosystems. SSRN Electronic Journal, 3(1), https://doi.org/10.2139/ssrn.5102358

[11] S. Ahmed and M. Asad, “Detecting Phishing Domains Using Machine Learning,” Applied Sciences, vol. 13, no. 8, p. 4649, 2023, doi: https://doi.org/10.3390/app13084649

[12] Bountakas, P., & Karyda, M. (2024). "Evaluating Machine Learning Models for Phishing Detection: A Comprehensive Analysis." Journal of Information Security and Applications, 78, 103-125.

[13] Chiew, K. L., Tan, C. L., Wong, K., Yong, K. S., & Tiong, W. K. (2023). "A New Hybrid Ensemble Learning Framework for Phishing Website Detection." Knowledge-Based Systems, 261, 110-124.

[14] Das, A., & Gupta, B. B. (2023). "Deep Learning-Based Phishing Attack Detection: A Semantic and Visual Approach." IEEE Transactions on Reliability, 72(4), 1452-1468.

[15] Rahman, M.M., Ullah, S., Nahar, S., Hossain, M.S., Rahman, M.M., & Rahman, M. M. (2025), “The Role of Explainable AI in cyber threat intelligence: Enhancing transparency and trust in security systems”. World Journal of Advanced Research and Reviews, 2025, 23(02), 2897-2907DOI: https://doi.org/10.30574/wjarr.2024.23.2.2404

[16] Hossain, M. S., & Muhammad, G. (2024). "Deep Learning-Based Phishing Detection for Industrial Internet of Things." *IEEE Transactions on Industrial Informatics*, 20(2), 2341-2350.

[17] Jain, A. K., & Gupta, B. B. (2022). "A Survey of Phishing Attack Techniques, Detection Mechanisms and Software Tools." *Cyber Security and Applications*, 1, 100-115.

[18] Kumar, N., & Singh, S. K. (2023). "Natural Language Processing (NLP) in Cyber Security: A Systematic Review of Phishing Email Detection." *Computer Science Review*, 49, 100-118.

[19] Mittal, M., & Kumar, K. (2023). "Adversarial Machine Learning in Phishing Detection: Threats and Countermeasures." *Computers & Security*, 129, 103-119.

[20] Opara, C., Chen, Z., & Wei, B. (2023). "Neural Phishing Detection with Transformers: Analyzing the Role of Attention Mechanisms in Email Security." Journal of Computer Security, 31(2), 157-178.

[21] Rahman, M.M, Gony, M.N., Rahman, M.M, Rahman, M.M., & Shuvra SD, M.K., (2025) “Natural language processing in legal document analysis software: A systematic review of current approaches, challenges, and opportunities”. International Journal of Innovative Research and Scientific Studies, Vol. 8 No. 3 (2025)DOI: https://doi.org/10.53894/ijirss.v8i3.7702

[22] Sahingoz, O. K., Buber, E., Demir, O., & Diri, B. (2022). "Machine Learning Based Phishing Detection from URLs." Expert Systems with Applications, 201, 117-132.

[23] Zhang, Y., & Wang, J. (2024). "Real-time Phishing Detection via Deep Reinforcement Learning in Evolving Network Environments." IEEE Access, 12, 14502-14515.