Benefits and Risks of Artificial Intelligence in Cybersecurity and Phishing Attacks

Mustafa Bešić

Master of Social Informatics
International Business and Information Academy
Tuzla, Bosnia and Herzegovina
besic.mustafa@yahoo.com

Abstract—With the development and expansion of artificial intelligence, there is a growing need for protection and security in the market. We can observe both positive and negative applications of artificial intelligence in phishing attacks, which are becoming increasingly complex and difficult to detect. Currently, artificial intelligence assists us in effectively and safely filtering incoming emails. The primary objective of this article is to raise public awareness about potential new, more sophisticated threats, as well as easier methods for detecting potential threats in cyberspace. Attackers utilizing artificial intelligence can more easily exploit audio and visual information to deceive users and gain access to desired information or data.

Keywords - artificial intelligence, phishing attacks, cyberspace, social engineering

I. INTRODUCTION

With the development of IT and the emergence of the first computer networks, criminals have moved from the streets to cyberspace. This trend has continued with the appearance of advanced software solutions, communication devices, smartphones, and laptops. Artificial intelligence is currently experiencing its full expansion and is spreading into all spheres of society. It is known that no development in the world is exclusively positive, and there is a justified fear that cybercriminals will start using artificial intelligence to achieve their goals.

Artificial intelligence did not just emerge recently, it has been developing for several years and has been mostly associated with specific business forms such as e-commerce, e-government, e-banking, and others. Currently, there is an opinion that artificial intelligence will take over all business processes in society. This claim is theoretically possible, but it also requires a significant amount of resources for the idea to have a chance to succeed in the real world.

Artificial intelligence engenders an immense amount of content every day that is increasingly difficult to differentiate from human-generated content. In this article, we focus on the „Phishing“ method of cyber attack, which is becoming more sophisticated the complex to detect, while on the other hand, artificial intelligence enables attackers to create and execute attacks on an organization or individual in a very short time.

II. THE CONCEPT OF ARTIFICIAL INTELLIGENCE (AI)

The paradigm of artificial intelligence was developed in the recent past when shows, movies, and series depicting the struggle between modern humans and advanced intelligence in the form of robots were aired. Today, artificial intelligence has experienced complete expansion, but what is less known is that it has been present in our lives for some time now. Artificial intelligence has been applied in areas such as e-banking, e-commerce, e-government, and others. The main task of artificial intelligence in this regard was to assist users in searches, mark user behavior, suggest information, and more. With the emergence of Chat GPT and the expansion of artificial intelligence, questions related to security, integrity, rights, ownership, management, and more have arisen. In this paper, our main goal is to examine how artificial intelligence affects phishing attacks in cyberspace. Why were phishing attacks chosen? The answer, in this case, is very clear – artificial intelligence can simulate the behavior of one user or person who wants to obtain important information from another, in this case, a real person. What is known from previous attacks is that human usually makes a mistake in hacking attacks, not the computer system or application. The deceived person believes in some software solution and allows it access to their computing device, after which the attack develops, information is collected, distributed, processed, sold, exported, and other forms of violence occur in cyberspace.

In this section, we can briefly describe artificial intelligence using the Turing test. The Turing test states that artificial intelligence is so intelligent that a user communicating with it via computer terminals cannot know whether they are communicating with a computer system or a real person. Alan Turing presented this theory in his classic article in 1950 (Turing, 1950). From the Turing test, we can obtain basic information to understand the primary goal of artificial intelligence and where it will move in the future. The answer is self-evident, the basic goal of artificial intelligence is to mimic human behavior, i.e., to make decisions, evaluate, suggest, and communicate with the other party in an open forum.

This section is of great interest to cybersecurity experts because it is extremely easy to manipulate users in this

way and obtain information that can be abused. Similarly, if we use artificial intelligence to detect attacks, we can alert users that malicious software or algorithms may be hiding behind certain content, which can cause harm to the organization or user.

III. AI AND CYBER ATTACKS

There are several cyber-attacks through which artificial intelligence can be used for illicit purposes. One such attack is phishing attacks.

A. Phishing Attacks

These attacks appeared with the first commercial appearance of the Internet and were a serious problem because they used social engineering to obtain client data. The main goal of these attacks is for the user to share their personal information with the attacker, believing the attacker's legitimacy[4]. In this regard, the attacker can use the brand of a reputable company, or an individual's brand, or persuade the user to obtain information. There are several types of phishing attacks, the most famous is:

- Email phishing attacks,
- HTTPS phishing attacks,
- Whaling attacks,
- Vishing,
- Smishing,
- Pharming[1].

1) Email Phishing Attacks

These are the most popular phishing attacks that organizations or individuals receive daily. Phishing attacks are well-crafted attacks that require obtaining adequate information. Attackers often gather information from the internet and falsely present themselves to increase their chances of success. The email contains information about someone and emphasizes the urgency of solving a problem. These attacks are extremely popular because they mostly use social engineering to successfully execute the attack [9].

2) HTTPS Phishing Attacks

Attacks involve sending a link in an email that the user is supposed to click on, downloading malicious software, or going to another website with compromised content. The attacker uses tools to shorten the link or to set up a link that appears to be secure and contains all the necessary security elements. Users can recognize these links by hovering their mouse over the link, and they should see a display of the link with suspicious content that contains several letters, characters, and symbols.

These attacks are highly popular, and there are several tools available that allow users to protect themselves successfully from these attacks[9].

3) Spear Phishing Attacks

Spear attacks represent more sophisticated attacks that require additional effort from the attacker. These attacks are executed deliberately and planned. The attacker collects information through social networks, the Internet, presentations, and seminars. After collecting the data, the attacker selects the target and, based on the collected information, falsely presents themselves and requests a service from the victim. It is not uncommon for the attacker to use things such as phone numbers, emails, pictures, or some other place for storing data.

The danger of these attacks is that they are aimed at individuals and smaller groups of users, so reporting the attack can go unnoticed[9].

4) Whaling Attacks or CEO Fraud

An attack that is also organized using publicly available information about organizations or individuals. The specificity of these attacks is that the attacker poses as the CEO of the organization and asks their subordinates (usually in finance) to transfer a certain amount of money to other accounts. The attacker creates the email to look as similar as possible in design and spelling to the real person and thus gains the victim's trust to perform the requested actions. The email usually also contains specific information such as job position, travel, and location information that is real.

An example of an attack that involves urgency and impersonation, is where the user is not given time to verify the information provided, but is asked to take action immediately[9].

```
Jim, i am currently stuck in a meeting, but we need to do a wire transfer as soon as possible for a payment Laura want us to get done today.

Can you get that done this morning? Let me know and I wil get you the infor you neeed. Thanks. David.
```

5) Vishing Attacks

Vishing attacks use voice effects to enhance the importance and urgency of the operation. The attacker prepares and, in this case, calls a responsible person requesting information or malicious actions based on publicly available information, such as a phone number. These attacks need to be identified now with special attention and control. Currently, there are artificial intelligence tools that can replicate the sound and create new content based on given audio information. In this way, there is a justifiable fear that attacks will become more concrete and targeted.

Attackers can use the voice of the organization's director, for example, who stated on a forum, TV, or radio station, as the source[9].

6) Smishing Attacks

Smishing attacks, as well as Vishing attacks, use publicly available information about individuals, groups, and organizations, and then attack their victims using SMS[9]. Here, prize games, discounts, purchases, and other ways of presenting themselves to the victim are popular. In exchange for a benefit, the victim provides their personal information or performs the requested action.

Several Smishing attacks currently exist in cyberspace, these are the most popular:

- Notification of delivered mail or package – the attacker falsely poses as a postal and package delivery service and requests users to confirm their order by entering confidential information into a form,
- Banking service – the attacker creates an SMS message where they pose as a bank and informs the user that there have been changes to their account and requests data entry from the user to complete or stop an action on their bank account[6],
- Winning a prize in a sweepstake – the attacker sends a message to the victim via SMS and requests that they confirm their identity as the winner of a sweepstake,
- Password management – the victim receives information that there has been a malicious takeover of their account (email, social media, website) and is asked to enter a two-factor authentication code. After gathering information about the victim, the attacker executes the „forgot password“ action and then asks the victim to send them their authentication code for account recovery in this way[6].

7) Pharming

An extremely sophisticated attack that unlines previous attacks, in addition to social engineering, must also contain technical knowledge. In these attacks, the attacker takes over DNS servers and, by controlling the server, redirects requests to malicious websites. Here, deviations between pages are possible, but sometimes the attacker hides grammatical errors, photos, font types, and other elements that may indicate that the website is malicious. After the user enters access data or assigns management rights, the attacker has completed his task.

IV. ANALYSIS OF THE CYBER SPACE

This section represents an analysis of the current state of cyberspace. The analysis contains information on the advantages and disadvantages of artificial intelligence in the current battle between attackers and cyber experts[2].

INTERPOL's report published for 2022 states that, on all continents, in addition to other criminal activities, cybercrime using phishing attacks is the biggest problem[2]. The attacker's goal is to collect information about the victim and later indirectly or directly gain access to financial resources. Phishing attacks have become very popular during the COVID-19 pandemic when the real world has moved into the digital realm. Attackers saw this transfer as an opportunity for additional earnings and further sophisticated and disguised their attacks. INTERPOL has recorded increased attacks on organizations, resulting in financial losses, data misuse, data leakage, and other negative activities.

A. Risks of AI in Cybersecurity and Phishing Attacks

According to The Guardian, the identification of phishing emails is currently more difficult because attackers can use popular chatbots to prepare emails and avoid detection by filters on SMTP (Simple mail transfer protocol) servers. Until now, it was possible to identify malicious phishing emails if they contained spelling or grammatical errors.

Using chatbots, attackers can create an email containing all the necessary information without any grammatical errors and in the specified language.

Cybersecurity experts used to be able to analyze email based on these parameters to determine if it was spam, but identification is currently difficult due to the content of the email[8].

Attackers can use artificial intelligence to create more comprehensive content and convince the victim that they know a lot about the topic they are writing about. Essentially, the content is mostly created by artificial intelligence for the attacker's tasks. Previously, attackers had to invest a lot of time to successfully create a phishing email, but now, using chatbots, it takes much less time and resources to create a successful phishing email. In the future, we can expect an increase in email phishing that will bypass controls on mail servers. Cybersecurity experts should conduct a more detailed analysis of which organizations they are working with and limit communication with unreliable partners or clients for organizations.

A new challenge that experts will face is voice cloning using artificial intelligence. This phenomenon could escalate because it allows attackers to prepare and execute an attack in an extremely short period. The attacker will use the voice of a person close to you and, using the Vishing method, will demand that you make a payment, purchase, or provide data. The additional concern is that artificial intelligence can mimic the thinking of the person it clones, their habits, vocabulary, and approach to information. Of course, these attacks would be much more difficult to carry out and would require a certain database of information about the victim or a person close to them. These attackers can become very dangerous if you have victims who are constantly exposed in public or have a large number of their video and audio content available online. In this

example, the attacker is only as strong as the information they have about victims.

The problems that are coming will be increasingly difficult to solve in traditional ways, so additional surveillance of systems, information, and material that can be compromised is needed.

B. Advantages Of Artificial Intelligence In Cybersecurity And Phishing Attacks

Artificial intelligence's advantage in digital system protection lies in the fact that it will be able to connect events and thus detect malicious software or spam emails. The first task where artificial intelligence will significantly advance is the detection of unwanted messages.

The detection of unwanted messages currently works in such a way that the administrator or responsible person implements a filter that, if they contain defined keywords or come from marked unwanted addresses, transfer that mail to spam. By using artificial intelligence in this step, it is possible to connect similar words or expressions that are not listed but are naturally related to the previously mentioned keywords and inform the user that it is possible to spam.

Detection of phishing attacks is a more dangerous and complex way of attack where the attacker tries to persuade the user to deliver their data to the attacker. Artificial intelligence has existed in this part for some time and is supported by popular email service providers. What we can expect from the development of artificial intelligence is a deeper analysis provided, brand or person linking with a real brand or person, code checking (HTML, Javascript, delivered files), and other email features.

Detection of other attacks will be carried out in more detail based on the current knowledge possessed by artificial intelligence. Email should make rapid progress in machine learning and recognizing unwanted mail because a large amount of data passes through mail servers every day and artificial intelligence could quickly detect and mark them as malicious through learning[4].

V. CONCLUSION

In this field, we can expect an expansion in the coming years. Artificial intelligence's development and application are increasingly permeating all spheres of modern human life. If modern humans do not adapt to the new circumstances that are becoming more certain, we can expect significant problems, such as changes in the economy, marketing, growing social inequality, and other negative aspects of technological development.

In the future, AI will create increasingly persuasive messages and more complex algorithms in conversations with humans.

The advantage of artificial intelligence in this domain compared to humans lies in its ability to process a large amount of high-quality data quickly, and most importantly, it easily adapts to changes and new information. Through the use of extensive knowledge, AI is moving towards developing effective manipulation algorithms to obtain desired information from users. Personalized attacks will become more frequent in the future, as attackers will use artificial intelligence to gather all publicly available information about a specific person and, through misuse, attempt to achieve financial or material gain.

Automatization of attacks is another negative aspect of artificial intelligence development. Attackers can use artificial intelligence to make attacks more convincing and automated, allowing them to collect data and reiterate their attacks for each user.

This part is particularly dangerous. The user may be unaware that they are currently under attack, unknowingly providing prompts to artificial intelligence about their private or social life. We can observe this in popular "Love" applications, where attackers exploit the victim's passions to obtain information or financial gain.

In the upcoming period, individuals and organizations must invest significant effort and knowledge to protect users on their applications, social networks, or business environments. Investing in security technologies should be a goal for all organizations to safeguard their operations and enable further development through the use of new technologies.

Security should focus on social engineering, specifically developing mechanisms for recognizing malicious emails.

Attackers often employ this method to obtain information or request financial services from individuals and organizations.

Checking web addresses should become our new reality if it hasn't been so far. Attackers clone addresses of popular brands by setting up prize games and actions to attract the attention of victims. After the initial interaction, they ask victims to enter their personal information and provide more details about themselves. If a user submits their data to the attacker, it doesn't necessarily mean that the attack will immediately follow upon receiving the information. Attacks are often prolonged for some time and then carried out massively toward multiple victims. In this way, attackers avoid constant pressure from protection agencies and do not expose themselves to risks.

After gathering and planning the attack, the attacker executes it within a defined timeframe, after which they destroy all the equipment they used to hide the evidence.

Research in this field will be highly interesting in the future. If we critically analyze the current situation, we can observe emerging trends in phishing attacks that we are currently experiencing.

The development of artificial intelligence has both positive and negative aspects, as is the case with any advancement in human history. However, it is crucial to conduct a comprehensive review of existing security systems and prepare users for potential new trends. While we cannot predict the exact direction of these trends, we can provide general information and guidelines to ensure the safety of everyone involved.

The security tools currently available in the market are adapting to new challenges, and it is necessary to constantly update the data to minimize the risk as much as possible.

Ahead of cybersecurity researchers lies an exciting period filled with challenges. The development of new methodologies is necessary to enhance existing knowledge and safeguard against future threats.

Security will become the most valuable resource in the future, as ensuring it for all users on a global scale will be challenging. Every individual should focus on personal growth and the establishment of ethical and responsible standards to effectively and positively utilize the advantages brought by artificial intelligence.

REFERENCES

[1] Duplico.io. Phishing napadi: koje sve vrste postoje i kako ih prepoznati?.https://duplico.io/phishing-napadi-koje-sve-vrste-postoje-i-kako-ih-prepoznati/. 2021.

[2] INTERPOL. 2022 INTERPOL global crime trend summary report. https://www.interpol.int/content/download/18212/304202/file/2022_Interpol_Global_Crime_Trends_Summary_Report.pdf. 2022

[3] J. T.Minkus and N. Memon. Leveraging Personalization to Facilitate Privacy.https://papers.ssrn.com/sol3/papers.cfm?abstract_id=244802. 2014.

[4] The Guardian. AI chatbots making it harder to spot phishing emails, say experts. https://www.theguardian.com/technology/2023/mar/29/ai-chatbots-making-it-harder-to-spot-phishing-emails-say-experts. 2023.

[5] Help net Security. Sophistication of phishing emails. https://www.helpnetsecurity.com/2023/03/08/sophistication-of-phishing-emails/. 2023.

[6] Terranova, M. Smishing Examples: What They Are and How to Stay Safe. TerraNova Security. https://terranovasecurity.com/smishing-examples/. 2022.

[7] F. Enos, S. Benus, R. L. Cautin, M. Graciarena, J. Hirschberg, and E. Shriberg. Personality Factors in Human Deception Detection: Comparing Human to Machine Performance. INTERSPEECH – ISLP, 2006.

[8] EUROPOL. Digital skimming. https://www.europol.europa.eu/crime-areas-and-statistics/crime-areas/cybercrime. 2019.

[9] IT Governance. The 5 most common types of phishing attack. https://www.itgovernance.eu/blog/en/the-5-most-common-types-of-phishing-attack#:~:text=Smishing%20and%20vishing&text=One%20of%20the%20most%20common,link%20to%20prevent%20further%20damage. 2019.