Degree Project

UNIVERSITY OF SKÖVDE

Phishing Detection Challenges for Private and Organizational Users: A Comparative Study

Bachelor Degree Project in Informatics

First Cycle 30 credits

Spring Term 2023

Students: Johan Brandqvist & John Lieberth Nilsson

Supervisor: Johan Zaxmy

Examiner: Thomas Fischer

ACKNOWLEDGEMENTS

We would like to thank our supervisor Johan Zaxmy for the support and feedback throughout this project. He got us on the right track and put us back in the saddle when we banged our heads into the wall. We would also like to thank our examiner Thomas Fischer, who with very constructive and to some extent entertaining feedback made this study a much more complete work. Some of his comments brought a well-needed smile to our faces and will not be forgotten for the time being. Lastly, we would like to thank our friends and families who once again will have the privilege of being around us. They endured many absent nights and weekends, and it is now our turn to show them the gratitude that they deserve.

ABSTRACT

Email communication has become an indispensable aspect of modern life, enabling rapid and efficient information exchange for individuals and organizations worldwide. However, the rise of phishing attacks poses a significant threat to the security and privacy of email users, with attackers continuously refining their techniques to exploit unsuspecting victims. This systematic literature review (SLR) aims to examine the challenges faced in email phishing detection for both private and organizational users, highlighting the unique obstacles and requirements for each user group. By analyzing relevant and current research from the past three years, this study seeks to identify the key challenges in email phishing detection, including the evolving tactics used by attackers, the limitations of current detection methods, and the hurdles in user education and awareness. Furthermore, it investigates the differential impact of these challenges on private and organizational users, with a focus on the potential differences. The comparative analysis offers valuable insights into the distinct challenges private and organizational users face in their efforts to defend against email phishing attacks. Understanding these differences is crucial for developing targeted, effective solutions that can enhance the overall resilience and security of email communication systems. Ultimately, this SLR serves as a foundation for future research and development in the field of phishing detection, fostering a safer online environment for all users.

Keywords: **Phishing, Detection, Challenge, Comparison, Email**

Table of Contents

1 Introduction

Phishing attacks have remained a persistent threat in the realm of email security, despite advancements in email security technologies and the propagation of user education initiatives, the occurrence of phishing attacks remains unrelieved. The dynamic nature of these attacks, combined with the rapid evolution of the digital landscape, has imposed a continuous re-evaluation of the strategies and techniques employed in the battle against phishing. This SLR aims to provide a comprehensive and in-depth analysis of the current state of research and practice in the field of email phishing detection by using a thematic approach, focusing on the various challenges that have emerged for organizational users and private users respectively. Phishing is a problem of importance when it comes to network and system administrators due to the probability of being forced to handle such occurrences as a daily routine. In this context, our SLR seeks to delve deeper into the complexities of phishing detection, particularly focusing on the unique challenges faced by two distinct user groups: organizational users and private users. The comparative nature of this study will provide a broader perspective on the phishing landscape, highlighting the specific vulnerabilities and defense mechanisms inherent to each group.

Organizational users, typically network and system administrators, often find themselves at the forefront of this battle against phishing. They are tasked with the responsibility of safeguarding the digital assets of their organizations, a role that requires constant vigilance and adaptability given the evolving nature of phishing attacks. On the other hand, private users, while not as exposed, are still significantly impacted by phishing attacks. Their personal data and privacy are at stake, making it important to understand their unique challenges in phishing detection. By comparing the experiences and challenges of private and organizational users, we hope to uncover valuable insights that could inform future research and practice in phishing detection. Identifying commonalities and differences between these two user groups could lead to the development of more targeted and effective strategies for combating phishing.

2 Background

This chapter begins with an explanatory introduction to phishing and the different types of attacks. The chapter continues with an overview of recently gathered statistics of phishing attacks to see the extent of these attacks. Following this, is a brief explanation of the most common detection techniques when it comes to phishing detection. The last section of this chapter will show examples of the research background from which inspiration for this study was gathered.

2.1 Phishing

Phishing is a type of cyber-attack that involves the use of fraudulent emails, text messages, or websites to steal sensitive information, such as login credentials and financial information (Das et al., 2020). The increasing number of phishing attacks in recent years has made it imperative for organizations and individuals to develop effective methods for detecting and mitigating these threats. APWG (2022) mentions that the term phishing was first used in 1996 in a social engineering attack against AOL (America On-line) accounts by scammers online. The history behind the name phishing origins from one of the earliest forms of hacking against telephone networks, called “Phone Phreaking”. This resulted in “ph” becoming a common hacking character replacement for “f” (Khonji et al., 2013). Another term that can be associated with phishing is spoofing, such as domain spoofing, website spoofing, and email spoofing. As explained by Jalda et al. (2022), spoofing is when a perpetrator forges an email to mimic a legitimate email address. Since the email looks to come from a reliable source, it is often opened by the victim.

In Figure 1, a representation of the attack process of a common phishing attack is visualized. Abroshan et al. (2021) explain that the attack begins with the attacker sending an email with either a phishing link or a phishing attachment, some of these emails will be stopped by technical phishing prevention systems before it arrives in the user’s inbox. If the prevention systems fail, the user receives and then there is a chance that the user opens the attachment or clicks on the phishing link which is the first and second step in Figure 1. The attachment generally contains built-in malware (e.g. ransomware) or a non-legitimate document asking the victim to take urgent action (e.g. fake invoice). The third step in this process is when the victim gets tricked to submit sensitive information, or when the malware is not detected nor prevented by an endpoint security system (e.g. antivirus software).

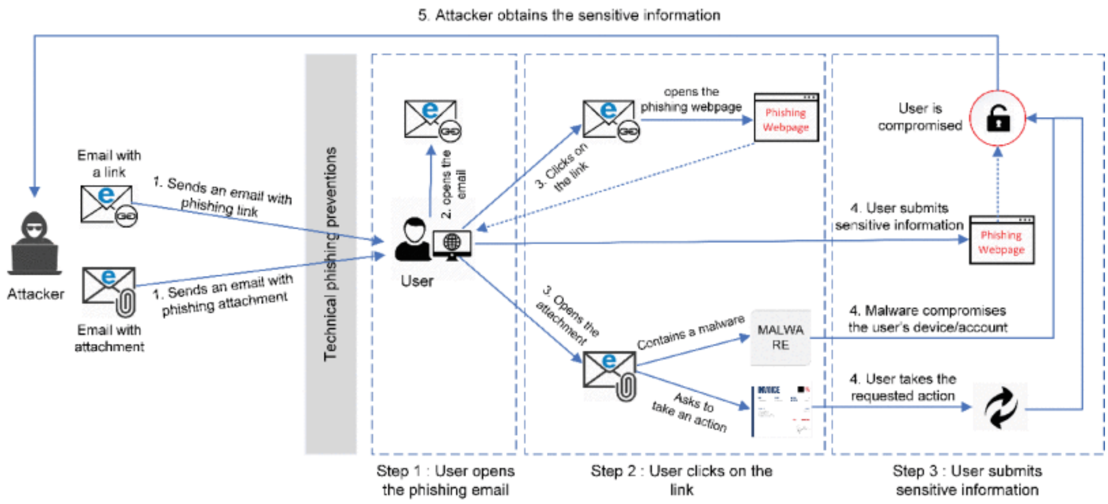

Figure 1. Attack process of a common phishing attack (Abroshan et al., 2021)

Dou et al. (2017) suggest that the phishing process can be divided into five steps: reconnaissance, weaponization, distribution, exploitation, and exfiltration.

In the reconnaissance phase, phishers look for famous web service brands with a broad customer base. They then design the phishing website and use social engineering on email spam in the weaponization phase. Phishers deliver emails to the victims in the distribution phase to exploit the weakness of humans to make them think the fraudulent websites are legitimate in the exploitation phase. Finally, in the exfiltration phase, phishers collect sensitive data from the phishing databases that they created (Dou et al., 2017).

Despite the availability of various anti-phishing tools and techniques, phishing remains one of the most successful and prevalent forms of cybercrime. The success of phishing attacks is largely due to their ability to mimic the appearance and behavior of legitimate sources, making them difficult to detect. This has led to a growing need for advanced phishing detection methods that can accurately identify and prevent these attacks in real-time. According to APWG (2022), the third quarter of 2022 saw a record-breaking 1,270,883 phishing attacks, making it the most severe quarter for phishing that the APWG has ever recorded. This is a consistent trend over recent years. Every quarterly report is a new record-breaker when it comes to the number of recorded phishing attacks according to APWG (2022). More statistics can be found in section 2.3.

Das et al. (2020) showed that there has been a significant increase in phishing research in the last decade. Continued, Das et al. (2020) show a wide range of techniques for detecting phishing, including machine learning (ML), deep learning (DL), computer vision, natural language processing (NLP), blacklisting/whitelisting, and behavioral analysis. However, most of these techniques suffer from limitations such as high false positive rates, low accuracy, and poor scalability according to the authors.

Phishing attacks usually seem credible and look sophisticated. Some common characteristics of phishing attacks can be examined to identify such attacks. The most common features of phishing attacks according to Madleňák & Kampová (2022) are discussed in the following list below.

- **Indirect Greeting**: Most legitimate organizations use a more personal approach with customers regarding communication and generally include the customers' names in some way, attackers usually use a more generic way of greeting their victims, without mentioning their names. These signs do not guarantee that it is a phishing attack since not all organizations use personal greetings when communicating with customers but are typical characteristics of a phishing attack.
- **Text Wording**: Attackers usually make grammatical errors and/or misspelled words in the text. This can usually be linked with a mistranslation because the message is translated from another language using publicly available translators.
- **Requesting Personal Information**: Phishing attackers usually act as a legitimate organization, for example, a bank, and the goal is to get the victim's sensitive information. Legitimate organizations never ask their customers to provide their sensitive information like login details or card numbers.
- **Urgency**: Another sign of phishing attacks can be the level of urgency shown by the attacker towards the victim. The attacker usually demands urgent actions from the victim, these actions must be performed immediately. Personal information is often the target, and account security or another security threat is often used to make the victim fall for this.
- **Irresistible Offers**: Attackers often make very lucrative offers that seem too good to be true. The offer might claim that the victim has won a price, funding, or other suspiciously lucrative offers.
- **Suspicious Domains.** Attackers are trying to act credible, to make this possible, the attacker usually uses a domain that is very similar to a legitimate organization. The attacker may use completely different domains or foreign domains, a more common example is to swap letters or symbols in a domain.
- **Suspicious Links and Attachments.** Attackers usually try to redirect the victims to websites that are not legitimate, or in other words, unsafe to make the victim enter personal details or make the victim download malware. Different forms of attachments, hyperlinks, and links to websites are used to achieve this. Knowing the characteristics of phishing attacks is crucial for stopping them and minimizing their success.

The typical phishing life cycle involves six distinct actors, as explained by Das et al. (2020) below:

- **Phishers**: These are individuals or organizations who carry out phishing attacks to obtain specific benefits, such as financial gain, fame, or identity hiding.

- **Web service providers**: These companies provide various services, such as email, social media, online shopping, and e-banking, via websites on the Internet.
- **Web service subscribers**: These are customers who subscribe to the web services provided by the web service providers. They are the primary targets of traditional phishing attacks.
- **Web hosting providers**: These companies offer website hosting services to web service providers.
- **Anti-phishing organizations**: These organizations support individuals and organizations in combating phishing attacks. They provide advice on anti-phishing controls, collect and analyze phishing data, and offer up-to-date phishing website blacklists, among other things.
- **Spear phishing targets**: These are specific individuals or companies targeted by phishers.

According to Das et al. (2020), each actor in the phishing process takes different actions and has varying reactions. Phishers employ sophisticated techniques to evade phishing detection approaches, while web service providers usually announce blacklists of phishing websites and recommend users to use strong authentication schemes. On the other hand, web service subscribers rely heavily on browser filters and third-party anti-phishing toolbars to detect and block phishing attempts.

The role of web hosting providers in the phishing process is rather ambiguous. Reputable providers usually enforce strict "Terms of Use" and avail certain anti-phishing solutions. However, free-to-use web hosting providers may not be able to afford to deploy good anti-phishing security measures, which leaves their customers vulnerable and attractive targets for phishing. Anti-phishing organizations such as APWG collect and analyze phishing data from various sources, such as users' reports via anti-phishing toolbars. They provide anti-phishing suggestions and solutions and may also cooperate with government agencies such as public security and law enforcement to detect and prevent cyber-attacks (Das et al., 2020).

2.2 Different Types of Phishing Attacks

The European Union Agency for Cybersecurity (ENISA) defines the term ransomware as *"a type of attack where threat actors take control of a target's assets and demand a ransom in exchange for the return of the asset's availability and confidentiality"* (ENISA, 2022). Ransomware has been one of the most common forms of cybercrime in recent years according to Agari (2023). The recent pandemic saw an increase in the number of ransomware attacks, and the number of ransomware victims has since declined as can be seen in Figure 2 based on data provided by APWG (2022).

The most common delivery method for ransomware attacks is phishing by email. Government institutions and companies that have been exposed to ransomware attacks are often reluctant to discuss how it happened, but many of them admit that successful attacks have been the result of employees falling for phishing emails (Agari, 2023). The latest report provided by ENISA (2022) mentions that phishing attacks are generally targeting a broad audience and thus used in a high amount of targets. This is different from other social engineering attacks that are customized to target specific employees. Social engineering is used to gain a technical foothold in the network to carry out further attacks using the employee's access inside an organization.

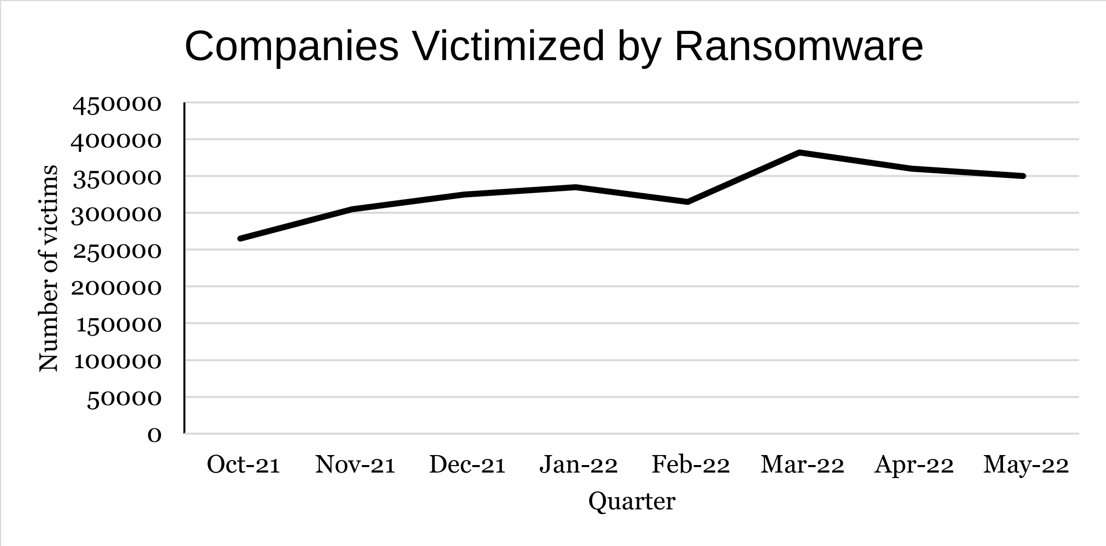

Figure 2. Companies victimized by ransomware, data from APWG (2022) (author's own)

Chiew et al. (2018) write about phishing techniques and suggest that they consist of three different components which are listed below:

1. **The mediums of phishing**: Three commonly used bases are voice, short messaging service (SMS), and the internet.
2. **The vectors of phishing which are used to conduct the attack**: Attackers use different vectors to launch phishing attacks, Email, and social networks are two examples of vectors that are available to the attackers via the internet.
3. **Technical approaches**: Different technical approaches are implemented by the attacker to improve the effectiveness of phishing. Spear phishing and whaling are two examples of technical approaches.

Heartfield & Loukas (2015) present a taxonomy of cyber-attacks, where they classify attacks based on several criteria, including the method of attack, the target of the attack, and the persistence of the attack. For instance, under the method of attack, they discuss various techniques such as email phishing, spear phishing, whaling, website phishing, vishing, and smishing. Each of these methods is explained in the next sections, highlighting their unique characteristics. Continued, they categorize the targets of attacks into individual users, organizations, and systems, emphasizing the varying strategies and implications for each target group. The persistence of attacks is also examined, distinguishing between one-off and ongoing attacks.

2.2.1 Spear Phishing

Spear phishing is a targeted phishing attack where attackers send fraudulent emails or texts to a specific individual or group. The term "spear phishing" is used to describe how attackers use personal information to tailor their attacks to a specific target. Unlike traditional phishing, which uses a more generalized approach, spear phishing is more personalized and can therefore be more difficult to detect (Karim et al., 2019).

2.2.2 Whaling

Aimed phishing attacks come in many forms, and one particularly sinister variation is called "whaling" or "CEO fraud". These attacks are specifically targeted at high-level officials within a company and are designed to appear as serious executive-level communications. Typically, the content of these attacks is crafted with a specific upper manager or senior executive in mind and may relate to an urgent executive issue that affects the entire company, or a customer complaint. These attacks are often accompanied by fake email addresses, designed to look like they come from a legitimate business establishment, or even from the CEO of the host company (Karim et al., 2019).

The risks and dangers associated with whaling attacks are similar to those of other forms of phishing and spoofing. It is important for individuals and organizations to remain vigilant and aware of the risks associated with these attacks, and to take appropriate steps to protect sensitive information and prevent data breaches. Implementing strong email security measures, such as two-factor authentication and email filtering, can help to mitigate the risks of phishing attacks and safeguard against the potentially devastating consequences of successful whaling attacks (Karim et al., 2019).

2.2.3 Smishing

Smishing, SMS phishing is a type of social engineering attack that targets individuals through their mobile devices, such as smartphones and tablets by sending text messages (Mishra & Soni, 2019). Smishing attacks are designed to trick individuals into giving away their personal and sensitive information by disguising themselves as trustworthy entities, such as banks or government organizations. The information collected through these attacks can be used for illegal activities such as identity theft and financial fraud.

Smishing attacks can be divided into three categories: Uniform Resource Locator (URL)-based, email-based, and phone number-based (Mishra & Soni, 2019). In URL-based smishing, the victim is directed to a malicious website by clicking on a link included in a message. This website will then prompt the victim to enter personal or financial information, often by offering incentives such as reward points or gifts. In some cases, this website may even result in downloading a harmful APK (Android Package Kit) file to the victim's device. In email and phone number-based smishing, the victim is encouraged to contact the attacker through the email address or phone number provided in the message. The attacker then tricks the victim into giving personal or financial information by posing as a trustworthy entity.

To avoid falling victim to smishing attacks, it is important to be vigilant when receiving unsolicited messages, especially those that ask for sensitive information or prompt a call or email. It is also recommended to verify the source of a message before responding, and never to provide personal information unless you have initiated the contact (Mishra & Soni, 2019).

2.2.4 Vishing

Vishing, voice phishing is a form of social engineering attack that combines the use of voice technology and human interaction to deceive targets into divulging sensitive information such as bank account numbers or social security numbers. The term "vishing" is derived from "voice" and "phishing", which is a similar type of attack that uses emails instead of phone calls (Griffin & Rackley, 2008).

Continued, the authors highlight the various techniques used in vishing attacks and the devastating consequences for targeted individuals and organizations. The study reveals that vishing attacks are becoming increasingly sophisticated, with attackers using advanced social engineering techniques to gain the trust of their targets and extract sensitive information.

Lastly, the authors mention that attackers often impersonate legitimate organizations such as banks, healthcare providers, or government agencies to gain the trust of their targets. They use voice technology to make the call appear to be coming from a trusted source and may use personal information obtained from previous data breaches or social media profiles to further deceive their targets.

2.2.5 Phishing Websites

Phishing attacks that target websites are a prevalent threat that is increasingly difficult to detect and mitigate. These attacks involve fraudulent websites that mimic legitimate ones, to trick users into disclosing sensitive information or performing malicious actions. To prevent website phishing attacks, various software-based detection techniques have been developed. These techniques typically involve analyzing website features such as URLs, content, and user interactions to identify suspicious behavior. ML-based approaches have also been used to improve the accuracy of these techniques. Despite the development of these detection techniques, website phishing attacks continue to be a significant threat, and attackers are continually developing new tactics to evade detection. According to Dou et al. (2017), researchers and security experts must remain alert and continue to develop new and improved techniques for detecting and mitigating website-based phishing attacks.

Clone phishing is a type of website phishing and refers to an attacker creating a replicate of a legitimate website, using a similar name. Whereas in regular website phishing, a fake website is created. Clone phishing is more of a targeted form of phishing since it targets users who usually visit legitimate websites, tricking the victims into that they are visiting a legitimate website. This might lead to the victims submitting personal information and login credentials as they believe they are on their usually safe website. It can also be a fake page residing on the official website, containing a link to fake resources, tricking the victim into that the link comes from a legit and trusted domain (Andryukhin, 2019).

2.2.6 Email Attacks

Phishing emails are usually sent in bulk and contain links to fraudulent websites or attachments that, when opened, install malware on the recipient's device (Abroshan et al., 2021). According to the latest report by APWG (2022), the number of phishing websites detected globally increased by 6% from 2018 to 2019, with a total of 342,000 unique phishing websites detected in 2019. The report provided by APWG (2022) also noted that phishing emails accounted for more than 90% of all phishing attacks, making it the most common method used by attackers.

One of the reasons for the increasing prevalence of phishing in emails is the ease with which attackers can automate the process of sending phishing emails. Additionally, attackers often use social engineering tactics, such as urgency and fear, to convince the recipient to click on a link or provide their information. For instance, the attacker may claim that the recipient's account has been compromised and that they need to take immediate action to secure it.

2.3 Statistical Overview

APWG is an organization that works with mapping out the phishing problem on a global scale. This work is done by reports from affected people and organizations (APWG, 2022). Figure 3 shows statistics about phishing attacks from the latest report posted by APWG. A total of 1,270,883 phishing attacks were observed in the third quarter of 2022, a new record and the worst quarter for phishing attacks that APWG has ever observed (APWG, 2022).

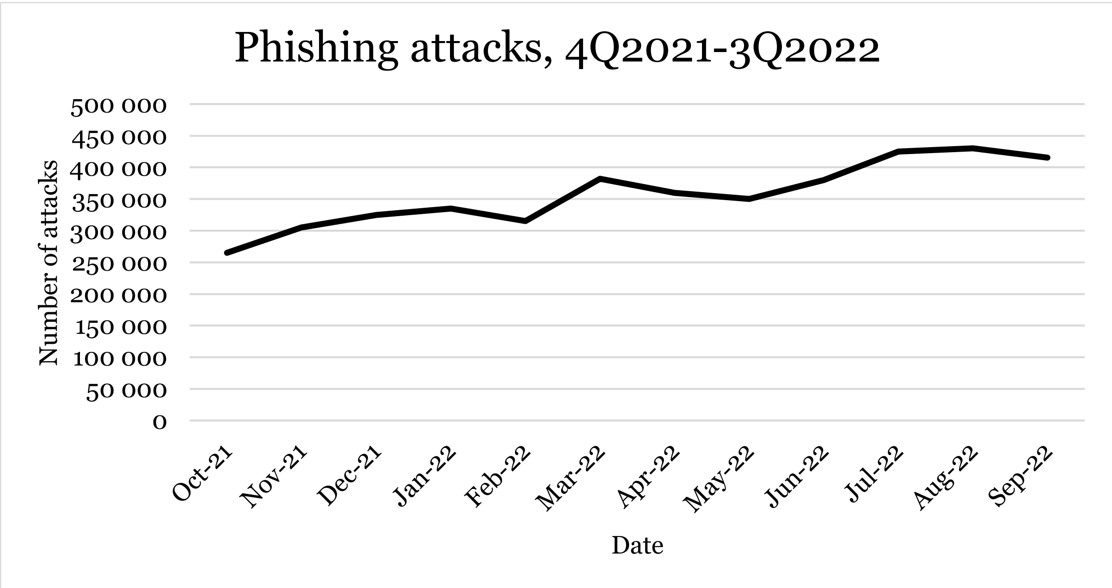

Figure 3. Phishing Attacks, data from APWG (2022) (author's own)

APWG (2022) also report statistics about targeted industries. Figure 4 shows a pie chart of the most targeted industries based on the APWG (2022) report. Phishing attacks against the financial sector remained the largest set of attacks. A huge increase in mobile phone-based fraud is being tracked, and vishing detection volumes have gone up more than three times than the reported results from the latest report from APWG (2022).

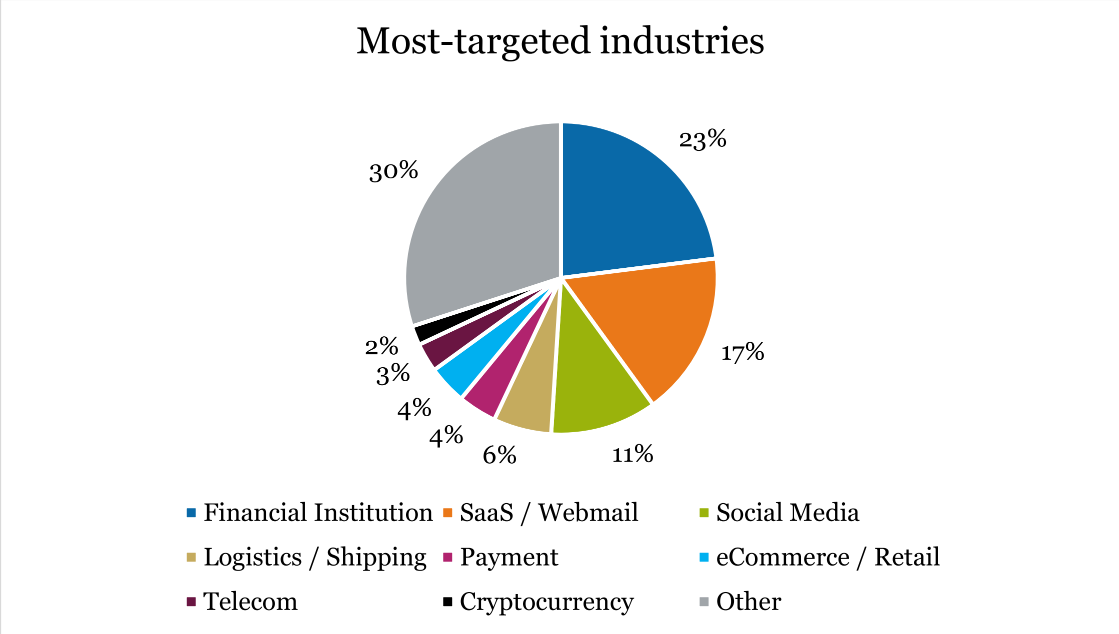

Figure 4. Most Targeted Industries, data from APWG (2022) (author's own)

The amount of phishing attacks in the world is more than ever and projections show that this trend will continue according to APWG (2022). This trend also results in an increase in research within this area so that phishing attacks can be understood better and mitigated in a better way. In Figure 5, a graph of the total number of hits of the search term "Phishing" is shown from the database IEEE, the number of hits is for each year respectively showing that as the phenomenon of phishing increases, so does the research.

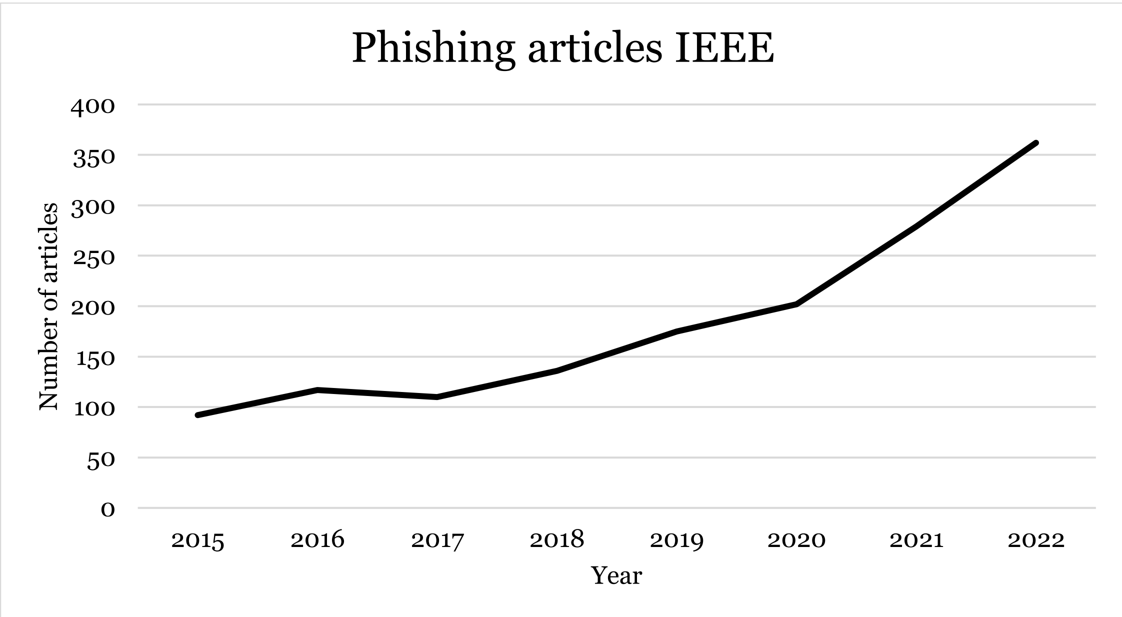

Figure 5. Phishing Articles IEEExplore (author's own)

According to SCB (2023), 2,960,700 individuals in Sweden were impacted by phishing attacks in 2019, representing 37% of the population aged 16-85.

2.4 Detection Techniques

To mitigate the risk of falling victim to phishing attacks, individuals and organizations must take steps to identify and avoid these emails. For instance, individuals should be cautious of emails that contain unusual requests or ask for personal information. Additionally, individuals should for example hover over the links in an email when this is possible, to view the destination URL before clicking on it (Waziri, 2015). This chapter presents and briefly describes different detection methods used to detect phishing attempts.

Phishing detection methods can be classified into two main approaches when it comes to organizations, as suggested by Bikov et al. (2019):

1. **User-dependent approach** - To raise and keep the awareness of employees, there are three main approaches that if properly and regularly executed are to decrease the potential risks of compromising an organization:
   - Regular communication with users regarding potential phishing threats, including their symptoms, methods, and appearance.
   - Regular user training, to provide the users with a user experience.
   - Exercising mandatory phishing tests to evaluate the users' awareness.
2. **Automated approach** - Mechanisms that if implemented are to increase security and decrease potential risks:
   - Blocklist/blacklist will block a specified type of data.
   - Heuristics analysis is used to analyze and evaluate data, based on multiple criteria.
   - Threat Intelligence is where information from various sources about current threat actors is collected, analyzed, and refined.
   - ML and AI technologies can for example detect phishing data, based on certain algorithms.

2.4.1 Machine Learning

ML has emerged as a defense mechanism against email spoofing, the AI mechanism is providing a different solution other than previously existing preventive measures. By leveraging learned experiences, ML algorithms can accurately detect spoofing attacks. As a type of AI, ML involves the use of pre-programmed algorithmic procedures to analyze input data and generate outputs. With the continuous input of new data, these algorithms can learn, analyze, and improve their performance over time. There are various types of ML algorithms, including supervised, semi-supervised, unsupervised, and reinforcement algorithms. With ML-based email protection, every email that arrives in a user's mailbox is subjected to thorough security checks based on constantly evolving analysis and feedback from users and administrators (Jalda et al., 2022). Subasi & Kremic (2020) evaluate the performance of ML methods by using different measures such as classification accuracy, F-measure, and area under the receiver operating characteristic (ROC) curves (AUC). These measures are used to assess the accuracy and effectiveness of the models.

2.4.2 Deep Learning

DL is a form of ML that relies on data-driven learning techniques. DL concentrates on comprehending the structure of data and creating models that can not only be utilized but also grasped by the general public. DL differs from traditional computational approaches. Instead of computers solving problems using unambiguous programmed instructions, DL enables the statistical analysis of input data to generate output within a specific range (Assefa & Katarya, 2022).

2.4.3 User Training

For the individual user, awareness, and training is important. After all, it does not matter how advanced and well-configured the detection systems are, users have to be aware of the different types of attacks that exist. ML methods do not always achieve perfect accuracy, which means that the user is still a vital component in detecting phishing attacks. According to Patil & Arra (2022), humans are accountable for 95% of security attacks. Due to humans being considered the weakest link, the importance of user awareness is of utter importance since the weakest link is what an attacker wants to try to penetrate. Based on the analysis of 49 studies, Das et al. (2019) highlight several key findings, including:

1. Users are often unable to distinguish between legitimate and fake websites, particularly when they are presented with realistic-looking phishing emails or websites.
2. People tend to rely on visual cues and superficial characteristics to identify phishing attacks, rather than carefully examining the content of the message or website.

3. User education and awareness programs can be effective in reducing the risk of phishing attacks, but they are often insufficient on their own.
4. Personal and situational factors, such as age, experience, and task complexity, can significantly impact how users respond to phishing attacks.

According to Baki & Verma (2022), background knowledge was found to be a marginal predictor of users' susceptibility to phishing, and training on websites was found to be more effective than training on emails. The study also found that warning techniques significantly improved users' performance.

Higashino et al. (2019) present an anti-phishing training system designed to increase users' awareness. The system involves a simulation game where users train and try to identify phishing emails and the users are then provided with feedback on their performance. The authors conducted a study to evaluate the system's effectiveness and found that it successfully improved users' ability to detect phishing emails.

2.5 Research Background

The research on phishing has increased over recent years according to Das et al. (2020), Patil & Arra (2022), and Salloum et al. (2022). In a literature study by Das et al. (2020), 300 articles between 2010 and 2017 were discussed, with a focus on phishing and spear phishing. Das et al. (2020) identified a lot of gaps in current research, such as under exploitation of deep learning techniques, lack of information about feature importance or selection in numerous studies, inadequate datasets and metrics for evaluation, and lack of diversity in user studies and long-term experiments.

Salloum et al. (2022) did a literature review with articles spanning from 2006 to 2022, covering a comprehensive analysis of phishing email detection. The study reviews 100 publications and elaborates on various phishing email detection methods such as misspelled domain names, poorly written emails, suspicious attachments or links, and phishing warning messages. Key observations include a lack of discussions on NLP techniques, and the availability of resources, especially those publicly accessible, and that while research has advanced in common languages like English, it has not generalized to other cultures or environments, especially non-English-speaking nations.

Khonji et al. (2013) discuss the complexities and challenges involved in phishing detection, mentioning that despite the common perception that educating end-users is a potential solution, such an approach is often insufficient. To address these challenges, the article proposes two potential solutions: improving user interfaces with active warnings and enhancing system behavior to automatically detect harmful messages. Finally, the article surveys various anti-phishing software techniques, including blacklists, rule-based heuristics, visual similarity, and ML-based classifiers. The authors conclude that ML-based detection techniques appear to be the most promising, as they have led to the most effective phishing classifiers to date.

When it comes to private users, Howe et al. (2012) did a study with a focus on the security of home computer users. The study does not only focus on phishing but contributes information about how and why the home user should be studied.

Tekeni & Botha (2022) did a more recent study with a focus on the home user, mentioning that many research studies use students as a data collection sample, which may not accurately represent the entire home user population. Additionally, the authors note that while some studies focus on organizational users, there are distinct differences between organizational and home users, such as the level of required network security knowledge. Thus, tailored security awareness campaigns for different age groups of home users are necessary to ensure the effectiveness of efforts in enhancing home security.

Görling (2006) discusses user education, and whether should users be interested in education when it is the IT security department's job for taking care of security. Görling (2006) mentions that if the systems are protected by user education, the question of security is in the hands of the user. The more relevant question is whether the users can act as educated users, user education is no general solution to security problems in information technology is not an argument not to inform and educate users.

Most research is focused on organizations, and the economic aspect is a major reason for this.

Görling (2006) mentions that security is not a thing for the sake of security, mentioning that investing money in computer security relies on the investor's understanding that it affects the bottom line positively. Organizations can provide funds for this type of research to help them mitigate phishing attacks in the future.

3 Problem

This chapter means to provide a detailed explanation, insight, and motivation for the problem areas on which this study is focused. The chapter will start with presenting the goals for this research, followed by the motivation for choosing those goals. Onwards, the chapter will explain and present the limitations that apply to this research. Lastly, this chapter provides insight into the expected results of the research. The main motivation for this study is the gap when it comes to comparing the current challenges for home users versus organizational users. With the COVID-19 outbreak, a lot of employees adopted working from home, and by doing so gave up the potential comfort of sitting at the office protected behind several security mechanisms. The pandemic raises interesting questions and concerns relevant to this study. It makes the questions raised in this research current; regular home users could be even more bombarded with phishing attempts now than before the outbreak. This study aims to find out what the major challenges are when it comes to phishing detection for organizational users and regular home users. One aim is to then compare these challenges. The resulting question that this research will try to answer is as follows:

- What are the challenges in phishing detection in emails for organizations and private users, respectively, and what are the major differences between them?

3.1 Research Goal

The first two goals are to investigate organizational users and private home users separately to compile current challenges when it comes to phishing detection. The two first goals will then be analyzed and compared to each other to reach the final goal. The final goal aims towards comparing the first two goals to find out how much it differs, and if it differs between private home users and organizational users.

3.2 Motivation

There are a lot of studies regarding phishing detection, most studies are aimed at organizations, with some that are aimed towards home users. There are however very few of these research papers that compare the regular home user and organizational users regarding phishing detection. However, the primary motivation for doing this study comes from the mutual interest in information technology, and while doing various activities over the internet, whether it is for work or fun, everyone would prefer to stay safe. By conducting a study on the subject, preferably helping both organizations and private users stay safe from phishing attacks on an everyday basis, this study can hopefully contribute to current and future research regarding phishing detection. Our research differs from previously mentioned research since the aim of this study is to compare challenges regarding phishing in email between organizations and private users.

3.3 Limitations

This study focuses mainly on phishing detection in emails since phishing emails are the biggest contributor to phishing attempts. The focus will to a certain extent also include studies related to URLs, since a phishing email often contains a link to a malicious website. This study will exclude studies related to phishing that focuses on SMS, voice, and social media.

3.4 Expected Result

When it comes to the expectations for this study, the expected result is a contribution to researchers wanting to do similar work and a contribution to users; home users, and organizational users. For researchers, the result will be a starting point for future, in-depth research on the topic. It is expected that this study can help users and researchers to see the difference between phishing detection in the work setting and their private life.

4 Method

This chapter will present a detailed explanation of the selected method, explaining why it was chosen, and other methods that were considered. This study aims to identify phishing detection challenges in emails, where the authors of this study will focus on two different approaches and later compare and discuss the findings, this is explained more in detail in section 3.1. When discussing different methodologies for this study, doing an SLR seemed a reasonable and safe approach. First, it was deemed as a reasonable method since there is existing research on the subject, which led to confidence in finding sufficient material. Second, both authors of this study felt that an SLR was the safest option due to inexperience, lack of social networks, and lack of social skills overall. Other research methods were considered, especially using a survey method but due to previously mentioned weaknesses and the time constraints of this study other methods were disregarded.

4.1 SLR

The method that best applies to this type of research is an SLR since this method can be used to see patterns in a specific area of research. Kitchenham (2004) explains that the meaning of a systematic literature review is identifying, evaluating, and interpreting all available research that is relevant to a particular research aim, or research question.

Kitchenham (2004) summarizes the different stages in a systematic review in three phases:

1. Planning the review
2. Conducting the review
3. Reporting the review

Continued, Kitchenham (2004) gives a detailed explanation of the stages in the mentioned phases and describes these phases. Figure 6 is based on the author’s explanation and represents an overview of the phases and stages.

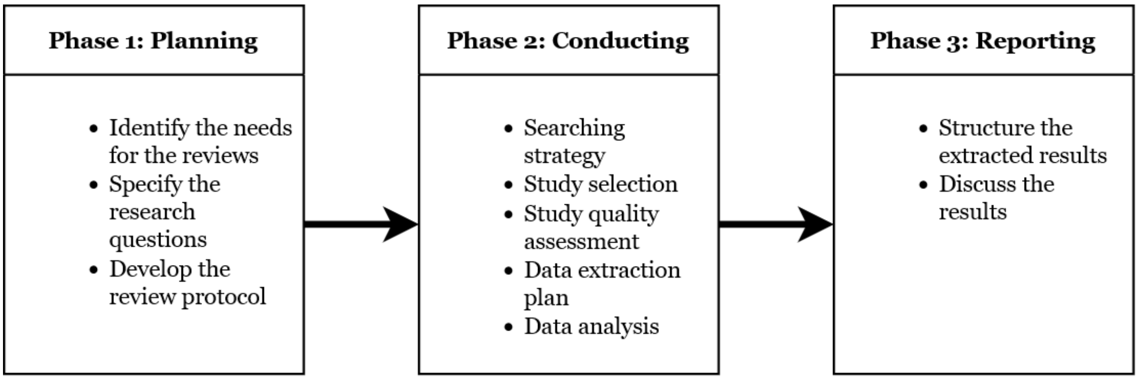

Figure 6. Phases and Stages, based on Kitchenham (2004) (author’s own)

Kitchenham (2004) mentions that a clear search strategy is important to correctly execute a systematic literature review. Readers need to be allowed to judge whether the applied strategy is complete and unbiased. To avoid biased article selection and to let the reader understand the conducted steps, it is important to let the search process be both transparent and replicable, in other words, the work must be documented sufficiently. Further, Kitchenham (2004) recommends using the following elements to achieve a valid level of documentation:

- Resources used (i.e., databases)
- Search terms used
- Inclusion and exclusion criteria
- Date of each search
- Total number of items for each search

Xiao & Watson (2019) mention that research questions drive the entire literature review process and Cronin et al. (2008) mention that a common mistake when doing this type of research is selecting too broad of a research question, it is better to narrow down and focus on the selected topic, and if necessary, broaden the scope of the review as the work progresses. The research question has a central role since it is difficult to produce reasonable search terms and article selection criteria when the research question or aim has not been defined. Since the article selection criteria are based on the research question it helps to limit the number of included studies because it establishes the used search terms and what type of data is searched for, hence limiting the amount of found matches.

When the research question, databases, search terms, and selection criteria have been defined, the next action is to initiate the article selection process. Meline (2006) illustrates a general framework for the article selection process with several steps which have been simplified for this report in Table 1 below.

Step	**Explanation**
1	Apply selection criteria to titles and abstracts
2	Eliminate studies that meet any exclusion criteria (i.e., one or more)
3	Retrieve the full text of the remaining studies after the previous step is completed
4	Assess remaining studies with the selection criteria
5	Include studies that meet all the inclusion criteria and none of the exclusion criteria.
6	Exclusion of studies within reason (i.e., incomplete, or ambiguous results)
7	The selection process ends at this stage, accepting studies for systematic review

Table 1. Article Selection Process, based on Meline (2006) (author's own)

4.1.1 Databases

Brereton et al. (2007) suggest that to find all relevant articles, it is necessary to conduct searches in multiple electronic resources, as no single database is likely to have everything. For this research, three databases have been chosen to increase the chances of finding relevant studies and to not miss out on critical data. Access to these databases has been granted via the university's web-based library service to ensure their applicability. Since most databases charge for access to stored articles, free access is essential for this project. The university provides free access to 81 databases in different fields of science, as of 2023. Based on the recommendations of Brereton et al. (2007), IEEE Xplore, ACM Digital Library, and ScienceDirect are considered relevant indexing services for software engineering and are included in this project. To further expand the potential of finding relevant studies, a fourth database was added to the list, Springer Link. This database was used by Salloum et al. (2022) in their study, a similar study to this, where a lot of inspiration was taken from. To summarize, the following databases have been selected for this study:

- IEEExplore
- ACM Digital Library
- ScienceDirect
- Springer Link

4.1.2 Search Terms

According to Xiao & Watson (2019), one of the most important aspects of conducting a successful and comprehensive SLR is the careful selection of suitable search terms. The outcome of a database query is strongly influenced by the choice of keywords that are being used. Further, the selection directly affects the review's thoroughness and quality. Failing to integrate relevant search terms can lead to procuring irrelevant articles, which highlights the significance of understanding and utilizing keywords specific to the topic to build a good collection of publications.

To further improve search terms, the correct usage of Boolean operators such as "AND" and "OR" can be used. By including Boolean operators, a search string will give a refined result of hits (Xiao &

Watson, 2019). Xiao & Watson (2019) highlight the importance of careful consideration when using Booleans since it will not only refine the search string but also a Boolean such as "OR" will yield a higher number of irrelevant hits. Boolean "OR" can be used to find more relevant publications and foremost used to find synonyms. However, as earlier mentioned, using synonyms comes with a compromise and should be considered before being implemented (Xiao & Watson, 2019). Getting too many irrelevant publications might lead to more work than originally needed.

When it comes to the chosen search strings used in this study, different keywords and search strings were tested to try and find usable search strings for all the databases. After conducting testing the following search string yielded the most promising result and was used:

- phishing AND detection AND (mail OR email OR e-mail)

The search string was used for covering both private- and organizational users, the difference in accepting articles between the authors' different approaches instead differed in the selection criteria, which are discussed in 4.1.3.

4.1.3 Selection Criteria

Wohlin et al. (2012) suggest that it is important to establish clear inclusion and exclusion criteria before selecting primary studies to minimize bias. These criteria should be developed in advance but may need to be revised during the selection process to account for factors that were not initially considered. The selection criteria chosen for this study can be seen in Table 2 and are explained below.

Inclusion criteria	IC1. Peer reviewed
	IC2. Written in English
	IC3. Published in journals or conference papers
	IC4. Published between 2020 and 2023
	IC5. Relevance to the subject
Exclusion criteria	EC1. Fails to meet inclusion criteria
	EC2. Behind paywall or login-requirement
	EC3. Identical to another chosen article
	EC4. Very specific study preventing reproduction
	EC5. The methodology, implementation, or results have complex technical content

Table 2. Selection criteria (author's own)

To ensure that the quality of this study is sufficient, IC1 was added to the inclusion criteria, meaning that every accepted article needs to be peer-reviewed. As explained by Spyns & Vidal (2015), peer reviewing is used for a quality assurance process. IC2 was added because most articles in information technology are written in English, and to exclude the potential of miss-translating a language other than English, only articles written in English are accepted. IC3 was added for further assurance of sufficient quality, studies must have been published in journals or conference papers. Articles in journals and conference papers need to have been peer-reviewed before being published. IC4 represents the span of years that this paper will focus on. Lastly, IC5 represents that the article is relevant to the chosen subject. Regarding IC5, the relevance depends on whether the article is used for organizations or private users.

Wohlin et al. (2012) mention that an SLR must not only have clear criteria for inclusion, but clear exclusion criteria to add further quality assurance to the study. EC1 was added so that if an article does not meet the inclusion criteria it is not accepted. Due to this being a student thesis without financial backing, EC2 was added so that articles behind paywalls are excluded. Additionally, this study excludes articles that require login credentials outside of the accessible databases provided by the university. EC3 was added to make sure articles that are identical or very similar to an already accepted article are excluded. EC4 excludes studies that are too specific or unique to be able to reproduce the result, a research paper exclusively aimed towards a small and specific population is an example that gets excluded because of the difficulties to replicate the study. Finally, EC5 is used to disqualify papers that are hard to understand or have an obscure description of the methodology, implementation, or results.

4.1.4 Reverse Snowballing

According to Kitchenham (2004), it is common for system literature reviews to rely on database searches that are executed with the help of prominent keywords. Kitchenham (2004) also mentions that complementary methods for systematic literature reviews should be invoked to gather as many relevant studies as possible. According to Wohlin (2014), most systematic literature studies do not succeed in making use of such additional methods. Database searches tend to involve considerable amounts of effort is pointed out as the key reason for that. The need to choose relevant search terms and databases along with considerable differences in how most databases' interfaces and their behavior keeps researchers from putting more effort into additional methods.

One applicable complementary method that can be used to prevent missing articles relevant to the study is reverse snowballing (Wohlin, 2014). Using reverse snowballing in a study indicates that the reference list of other studies is parsed to find more work of interest. According to Wohlin (2014), the identification of a reasonable start set of studies is the first step when reverse snowballing is used. It is also stated that there is no specified and easy way to establish the contents of a valid start set.

To deal with the implications mentioned in the previous two paragraphs, the following approach of reverse snowballing was selected. For this study, reverse snowballing was meant to be used as another layer on the previously mentioned selection process. The original plan was to use it if necessary, this is further explained later in this section. However, reverse snowballing was not utilized in this study due to time constraints, the method explained in this section did not get applied. This is further discussed in chapter 8.2. The reference lists of studies and articles that would have been accepted after going through the previously mentioned selection process are in that case used for reverse snowballing, meaning that the articles that are found in the reference list are used as the snowballing start set.

The snowballing start set is assessed with the previously mentioned selection criteria. The reverse snowballing is initiated in step 1 in Table 1. This results in the selection criteria being applied again to the start set. Studies that successfully pass these criteria are retrieved in full and evaluated against the selection criteria. The new studies get accepted into the study if they are not ruled out for meeting any exclusion criteria. The following limitations to the reverse snowballing are applied because of the restrictions of time for this study and are listed below:

1. Reverse snowballing will only be used if needed. (i.e., there are not enough articles to conduct a proper study). If used, it will be noted. However, this study means to find and compare the current challenges from studies posted between 2020 and now, 2023. Since reference lists refer to older studies, they would often become irrelevant to this study. Also, due to time constraints, reverse snowballing will not be utilized if it is not allowed in the schedule.
2. Articles that have been rejected in the initial selection process will not be assessed again and are instead ignored.
3. Reverse snowballing is only applied in one iteration. (i.e., studies that pass the reverse snowballing procedure and are accepted into this study will not have their reference list added to the reverse snowballing start set).

4.1.5 Analysis

After conducting the above approaches, the collected data needs to be analyzed. To do this, Wolfswinkel et al. (2013) have proposed a five-step iterative model that does this in a structured way. The model is intended for structured literature reviews and will be used in this study. The proposed model has been touched on in previous sections and is define, search, select, analyze, and present. Where in step four, the analysis step, accepted data from previous steps will be analyzed. This study will use the qualitative approach where thematic coding will be used to analyze the accepted data. Articles that get accepted for analysis are marked to then get coded. The coding then consists of finding relevant issues tied to the research questions as well as identifying patterns and categories. Continued, identified patterns and categories will be extracted. Finally, extractions and relations between extractions will be refined to meet the research goals. In step five, the results from the analysed data in step four will be presented in a structured way, this is found in chapter 5.

4.2 Validity

Berndtsson et al. (2008) explain that validity is the relation between the intended research compared to what was researched. Furthermore, Wohlin et al. (2012) mention that the validity of a study indicates the credibility of the results and whether the results are true and not biased by the researcher's perspective. Since validity is such an important aspect of a systematic literature review, ensuring good levels of validity is highly prioritized.

Several mitigation methods were utilized to ensure validity in this study and will be presented in this section. To ensure that articles are interpreted correctly, articles have been read in full to make sure that the presented results in this study are not taken out of context. A transparent process was used for both the literature search and selection to avoid selection bias. Four databases were searched with clear inclusion and exclusion criteria. The selection process was individually verified by both authors. The list below presents the summary of the identified validity threats in this study and how they were mitigated.

- **Construct Validity**: Ensured that the categorization of challenges and the search string used in the literature search accurately represent the intended concepts. This involved refining and expanding the search string.
- **Internal Validity**: Minimized biases in the literature search and selection process by following a structured and transparent methodology. Also, strive for objectivity in the interpretation of the results, for example, by having both authors independently analyze the data and then compare the findings.
- **External Validity**: Ensured that the selected literature adequately represents the broader field of research on phishing detection challenges. This involved expanding the literature search to include multiple databases and using a broader range for the search string.
- **Conclusion Validity**: Ensured that the conclusions drawn from the literature review are based on a comprehensive and unbiased analysis of the data. This involved cross-checking the findings with other sources, looking for patterns and consistencies across the different studies by using a thematic approach, as well as being cautious about making broad generalizations based on a limited number of studies.

Further, to ensure validity between the two authors of this study, several techniques have been utilized to avoid any data extraction bias. The following list of techniques has been iterated throughout the whole project.

1. **Clear communication and collaboration**: The authors have been communicating daily, mostly by being vocally connected but also by text messages. There has also been close collaboration throughout the whole writing process.
2. **Research background**: The project is strictly based on existing research with multiple sources to ensure relevance.
3. **Data verification**: Selection criteria from Table 2, or more specifically the inclusion criteria IC1 *(Peer reviewed)* and IC3 *(Published in journals or conference papers)* make sure that the data has been verified by other researchers to ensure its accuracy and reliability. Used data have been verified by both parties to ensure their accuracy and reliability.
4. **Peer review**: Feedback from peers and the supervisor of the project has been utilized to identify flaws or weaknesses.

To further ensure the validity between authors when conducting thematic coding in the analysis process, additional steps were taken.

1. To ensure consistent and accurate coding, clear coding criteria were established before

starting the coding process. The authors had a mutual agreement for identifying and categorizing themes.

2. Inter-rater reliability tests were conducted, meaning that periodically both authors independently coded the same literature, to then compare the results, as suggested by Wohlin et al. (2012).
3. A detailed audit trail was maintained of the coding process to ensure that the coding was transparent and reproducible.
4. Further peer reviewing was utilized to ensure the validity of coding.

4.3 Selection Procedure

This chapter will present the implementation of the techniques that were discussed earlier in Chapter 4, phase two as can be seen in Figure 6. The four chosen scientific databases mentioned in chapter 4.1.1 are probed with one search combination each. To ensure there were not too many hits, the search terms were applied to the title, abstract, and keywords on the databases. This amounts to a total of four search operations, one for each database. The applied search string is shown here as a reminder: Phishing AND detection AND (email OR mail OR e-mail)

All database searches were conducted on 2023-03-29. The selected databases have the option that addresses IC4 *(publication year)* and narrows down the number of hits to a total of 661 publications across the four databases. To determine what publications are of relevance to this study, actions were taken in different stages listed below:

1. On the databases for ACM, ScienceDirect, and Springer Link, every hit was downloaded for further assessment but on IEEExplore selection criteria were applied to the title and abstract and read directly on the website. This was done because IEEExplore only lets you download 10 articles at a time and at first glance, the number of irrelevant articles seemed high. The selection process was done by both authors simultaneously, starting from one end each to then meeting up in the middle. This yielded 97 and 87 publications respectively from the IEEExplore, downloading the publications for further assessment. To meet the selection criteria EC3 *(Identical to another chosen article)*, duplicates were checked against and removed if found using the research software Zotero. 229 articles were initially accepted from the total amount of 661 hits, the articles were uploaded to Zotero to make use of the duplicate items feature. There were no duplicates found amongst the hits that were uploaded when checked against each other. The number of publications remained at 229 after this stage.
2. Selection criteria are applied to titles, abstracts, and keywords on the remaining articles. In this step, the publications from IEEExplore mentioned in the previous step were switched between the authors to make sure the authors agreed. IC5 *(Relevance to the subject)*, was partly evaluated in this step when the abstracts were read. An evaluation of whether the article was aimed towards private, organizational users, or both were done. The articles got organized into folders in Zotero, one for organizations, one for private users, and one for both. 4 articles were disqualified in this step due to meeting EC2 *(Behind paywall or login-requirement)*. After evaluating the remaining 229 articles, 100 articles were removed, leaving 129 articles for continued inspection. To further narrow down the remaining candidates, publications that are difficult to understand or obscure were disqualified by applying EC5 *(The methodology, implementation, or results have complex technical content)*. After conducting EC5, 71 articles were qualified for further inspection.
3. The remaining 71 articles were retrieved in full text, applying selection criteria to whole articles. This resulted in EC2 being automatically evaluated since the articles could not be downloaded unless they were accessible. After all articles were retrieved, they were read in their entirety. If an article was not suited for organizations, the article got passed along to the author conducting the study on private users instead and vice versa. If the article was not suited for organizations or private users, it was removed. To make sure that IC5 was met, the articles were searched for information regarding phishing detection strategies and current challenges within the subject. 7 articles were disqualified after applying all selection criteria on the remaining full-text articles, meaning that the final bibliography contains 64 accepted articles, 48 for organizations, 13 for both, and 3 for private users.

5 Results

This chapter will present the result from the accepted articles, the full list of accepted articles can be seen in Appendix A. The information that has been extracted from the articles is focused on challenges in phishing detection and contains a summary of each article's study. A1 as an example, represents article one from the list of accepted articles, the articles are in alphabetical order by author names. So, articles one and two might not be presented next to each other, but rather in the relevant result section. This naming scheme is used to avoid using proper references for the authors to make the text more free-flowing and easier to read. Table 3 summarises the databases that were used, what the search terms were applied to, and the number of hits for each database before applying the selection criteria except IC4 *(publication year)*.

Database	**Applied on**	**Hits**
IEEExplore	Title, abstract & keywords	338
ACM Digital Library	Title & abstract	33
ScienceDirect	Title, abstract & keywords	25
Springer Link	Title, Abstract & keywords	265

Table 3. Database searches (author's own)

5.1 Accepted Articles

The results of the accepted articles for organizations begin in chapter 5.1.1 followed by the results of the accepted articles relevant to private users or both parties in chapter 5.1.2. When conducting the selection process, there appeared to be a deficit of articles exclusively aimed toward private users, compared to articles aimed toward organizations. So, articles that seemingly fit both groups are presented in the result section for private users.

5.1.1 Organizations

A1. Abdulraheem et al. (2022) present a study on the challenges of detecting phishing attacks in email communications for organizations and the application of ML methods to address this issue. The authors explore the effectiveness of various ML algorithms in classifying emails as phishing or non-phishing, aiming to help organizations better protect themselves against cyber-attacks. The article mentions that the challenges in phishing detection for organizations originate from the increasing sophistication of phishing attacks and the difficulty for users, even experienced ones, to identify malicious emails and websites.

A3. Adzhar et al. (2022) discuss the use of ML techniques for phishing email detection, and they highlight the challenges of each technique:

- Decision Trees (DT): Complexity due to numerous layers, which may lead to overfitting.
- Support Vector Machines (SVM): Achieving optimal performance may require careful selection of kernel functions and parameter tuning.
- Random Forest (RF): Ensuring that there is an actual signal in the features so that models built using those features perform better than random guessing.
- Achieving low correlation of predictions (and errors) made by individual trees.
- Naïve Bayes: The assumption of feature independence may not hold in real-world data, potentially reducing model accuracy.

A5. Al-Ghamdi & Alsubait (2022) aimed to develop an email multi-classification system using ML approaches to address the increasing need for digital forensics in dealing with cybercrimes committed via emails, such as phishing and fraud. The major challenge discussed in the article is the presence of imbalanced data, which requires the development of specialized algorithms and techniques to handle such cases effectively.

A6. Al-Haddad et al. (2021) investigate the challenges in detecting phishing and fraudulent emails in organizations and explores the use of ML techniques to address these issues. According to the authors, challenges in phishing detection originate from the need to differentiate between legitimate and fraudulent emails effectively. The authors highlight the main challenge being that hackers continuously improve their email formats and content, making it harder to identify phishing attempts.

A7. Al-Hamar et al. (2021) present a comprehensive study on challenges in detecting spear-phishing attacks in enterprise email and highlight the following challenges:

- Attackers often use a combination of techniques such as typosquatting and credential harvesting to bypass detection systems.
- Traditional email security systems may not be effective in detecting targeted spear-phishing attacks.
- Attackers use similar domain names to impersonate genuine organizations, making it difficult for users and security systems to differentiate between legitimate and malicious emails.
- Existing detection mechanisms, such as content analysis, may not be sufficient in identifying spear-phishing emails.

A8. Alhogail & Alsabih (2021) propose a novel solution to improve the performance and accuracy of phishing email detection using NLP. When it comes to challenges, they highlight:

- Rapidly changing phishing methods: Phishing techniques evolve quickly as cybercriminals adapt to new security measures. This makes it crucial to develop advanced solutions that can keep up with these evolving threats.
- Limited research on email body text: Most existing phishing detection research has focused on analyzing email headers, links, and other metadata. However, the email body text has received relatively little attention.
- Need for domain experts and handcrafted feature extraction: Many existing phishing detection methods rely on domain experts and handcrafted feature extraction to reduce data complexity and make patterns more visible to learning algorithms.
- Inefficiency in detecting zero-day phishing attacks: Many current phishing detection methods struggle to accurately identify zero-day phishing attacks, which are newly created phishing campaigns that exploit unknown vulnerabilities.
- Language limitations: Most existing phishing detection research focuses on English-language emails, potentially limiting the applicability of these methods to non-English email phishing attempts.

A9. Alotaibi et al. (2020) present a Convolutional Neural Network (CNN) based model called CNN Phish Detector (CNNPD) for detecting phishing emails. The model is trained and tested on two publicly available datasets, PhishingCorpus and SpamAssassin. The authors mention three major challenges in phishing detection in email:

- There are a limited number of email phishing datasets due to the confidential nature of emails, there are very few publicly available datasets for phishing email detection.
- The available datasets typically have an imbalance in the number of phishing and non-phishing emails, which can negatively affect the performance of the detection model.
- Emails often contain a mix of HTML tags, email header information, encoded images, files, text, and URLs, which need to be cleaned and pre-processed before being fed into the model.

A12. Bountakas & Xenakis (2023) propose a phishing email detection methodology named HELPHED that combines Ensemble Learning methods with hybrid features. The methods use two different ML algorithms to handle the hybrid features separately. Challenges in phishing detection that the authors mention are:

- The constant evolution of phishing emails makes it difficult for detection systems to adapt to new techniques and attack strategies.
- The imbalance between benign and phishing emails, which can affect the performance of detection algorithms.
- The lack of shared information to reproduce the results of existing works, hinders further research in the phishing email detection domain.

A13. Catal et al. (2022) conducted an SLR on the use of DL techniques for phishing detection, focusing on the challenges faced by organizations in detecting phishing in emails. The authors analyzed 43 articles to answer nine research questions and provide an overview of the current state of DL-based phishing detection models. The challenges mentioned in the article include the lack of interpretability of DL models, the need for fine-tuning hyperparameters, and the difficulty in detecting structural changes in URLs. The authors also point out the lack of labeled data, the existence of duplicate data points, and the short lifespan of suspicious websites.

A14. Cota & Zinca (2022) present the implementation and evaluation of five supervised ML algorithms for spam email detection. The challenges mentioned in the article, are primarily related to the difficulty in distinguishing certain types of emails from spam emails. This is because these emails often contain characteristics similar to spam, making it harder for ML algorithms to accurately classify them. Additionally, the size of the training dataset plays a significant role in the performance of the algorithms. Smaller training datasets result in lower accuracy, precision, recall, and F1 scores for the applied algorithms.

A16. Dey et al. (2021) present a comprehensive analysis of various classification algorithms used in ML for detecting phishing websites and emails. It highlights the challenges in email phishing detection and offers suggestions for improving the efficiency and accuracy of the model. They highlight the following challenges:

- Achieving high accuracy: Single ML algorithms may not provide the desired accuracy in detecting phishing emails, as they can become imprecise when categorizing new samples and might overfit.
- Training time: Some algorithms, such as the Random Forest Classifier, take a longer time to train the model, which could be a challenge in terms of efficiency.
- Text pre-processing: Proper pre-processing of text data, such as removing punctuations, inflectional endings, and unimportant words, is crucial to improve the model's accuracy. However, certain techniques like lemmatization may sometimes reduce the model's performance.

A20. Fayoumi et al. (2022) present a study on the application of ML methods to detect phishing emails. The authors discuss the challenge in detecting phishing emails due to the vast range of possible phishing scenarios and the constantly evolving nature of cyber-attacks. Further, they mention that ML methods require a significant amount of data for training to achieve high accuracy and it is challenging to find the optimal balance between achieving high classification performance and managing computational resources.

A21. Gallo et al. (2021) delve into the issue of detecting and mitigating phishing threats through ML techniques, with a particular emphasis on identifying malicious phishing emails. The study outlines various challenges and recommendations for enhancing phishing detection capabilities. They highlight four major challenges when it comes to phishing detection:

- Limited feasibility of examining all incoming emails: Due to intricate features and extensive computational demands, conducting an in-depth security assessment of all received emails is impractical. This constraint necessitates the exploration of feature reduction and unsupervised methodologies.
- Dependence on vigilant users: The proposed strategy relies on user reporting, making user participation essential for pre-filtering emails and selecting candidates for security evaluations. Cultivating user awareness is critical for fostering collective immunity.
- Insufficient defense against single-target attacks: The approach's limited visibility on isolated successful attacks results from analysts only intervening when an email is flagged as suspicious. This shortcoming underscores the significance of user education and the implementation of client-based tools.
- Drawbacks of supervised learning: Although the method addresses the issue of obtaining a labeled dataset, challenges persist, including class imbalance, misclassification of underrepresented attack variants, and the inability to detect novel phishing techniques.

A22. Gangavarapu et al. (2020) highlight the importance of feature engineering and ML in building effective intelligent systems. They detail various attempts made by researchers in mitigating the menace of unsolicited bulk emails (UBE) through the use of ML classifiers. The authors specify the lack of an effective strategy to handle security attacks on UBE filters.

A23. Giri et al. (2022) present two different models for detecting phishing emails and evaluate their performance using a dataset created by merging five smaller datasets. The article highlights the challenges in phishing detection, particularly when it comes to handling large datasets and the limitations of existing models.

A25. Gualberto et al. (2020) propose two methods for detecting phishing emails. The first method uses a bag-of-words approach with a Multi-Layer Perceptron classifier, while the second method extracts features using Latent Dirichlet Allocation and applies various classifiers. The authors mention that challenges and future work in phishing detection should focus on combining word embedding with techniques to improve the interpretability of results and better fit pre-trained models to new tasks.

A26. Gutierrez et al. (2020) present a study on phishing email detection using ML techniques and document embeddings. The authors acknowledge the need for more interpretable features and deeper insights into the differences between phishing and legitimate emails for improved detection. They also discuss the difficulties in interpreting features, the need for deeper insights into the characteristics of phishing and legitimate emails, and the exploration of alternative machine-learning techniques for improved accuracy.

A27. Hosseinalipour & Ghanbarzadeh (2022) present a novel approach to spam detection in emails using an optimization algorithm. They highlight one major challenge; detection techniques are prone to errors and limitations, and no single method has proven to be entirely effective.

A29. Jáñez-Martino et al. (2023) discuss the different strategies used by phishers and highlight three challenges in the area:

- Spammers use sophisticated strategies to bypass filters, such as poisoning text, obfuscated words, hidden text salting, and image-based spam.
- The changing nature of spam over time due to dataset shift and spammer strategies.
- Identifying the most relevant features of the data.

A30. Kovač et al. (2022)provide an overview of ML algorithms used for phishing detection and highlight the following challenges:

- Achieving perfect accuracy: Despite high accuracy rates achieved by ML algorithms, they still have not reached 100% accuracy in phishing and spam detection.
- Dataset size and ratio: The size of datasets used in the research varies, and the ratio of training to testing data is not consistent across studies, which might impact the algorithm's performance.
- Language and feature selection: Different studies use different features for detection, such as the email subject, body, and sender information. This inconsistency may affect the performance comparison between algorithms.
- Adapting to new phishing techniques: Phishing techniques continue to evolve, posing a challenge for ML algorithms to adapt and stay effective.

A31. Kumar et al. (2020) discuss a spam detection model for email filtering using ML algorithms. The model focuses on the content of the email to classify them as spam or non-spam, implementing various classifiers. According to the article, the main challenge is that their model is limited in testing emails based on a small corpus and that it can lead to the misclassification of some attributes.

A32. Lain et al. (2022) conducted a large-scale, long-term experiment to evaluate the effectiveness of various phishing prevention strategies in a large organization. The research aimed to understand the challenges in phishing detection and assess the effectiveness of phishing training, email warnings, and crowd-sourced phishing detection. The authors found that continuous phishing training based on simulated phishing emails was effective in reducing phishing susceptibility, with significant improvements observed in the first few months of training. However, they also found that embedded phishing training, as commonly deployed in the industry, could be ineffective and have negative side effects. This contradicts some prior literature and common industry practices. The study also highlights that certain demographic factors, such as age and computer use on the job, influenced susceptibility to phishing.

The research demonstrated the feasibility of using crowd-sourced phishing detection in an organization, as employees actively reported suspicious emails over a long period without experiencing significant “reporting fatigue”. The operational workload for processing reported emails remained acceptable, with the use of a secondary appliance triaging the reports.

Based on their findings, they caution against the deployment of methods like embedded phishing exercises and training, where the existing literature is less unanimous about their effectiveness, and suggest considering crowd-sourced phishing detection as a complementary way to improve the overall phishing prevention capabilities of an organization.

The authors identify the need for more research to better understand the effectiveness of phishing exercises and training, especially in the context of employees' normal working environment. They also call for further studies to explore the psychological effects of such training and how it may influence employees' future handling of real phishing emails.

A33. Lee et al. (2021) present D-Fence, a phishing email detection system that utilizes a multi-modular approach to cover a larger attack surface than existing solutions. Challenges mentioned in the article include:

- Adversarial attacks: Attackers can craft benign emails to resemble future phishing emails or bias the training set with a large number of phishing emails.
- Adaptive evasion: Attackers can create spear-phishing emails that are structurally similar to benign emails, use adversarial texts to deceive text analysis models, or hide malicious content behind legitimate-looking URLs.
- Text length limitation: The BERT-base transformer in the text module limits analysis to 512 tokens.
- Limitations of evaluations: The experimental results on the private EES 2020 dataset are not directly comparable with future studies. The cost-efficient and best-performing configurations may not be generally applicable to other environments and datasets, as learning models must be tuned for specific datasets and target costs.

A34. Li et al. (2022) present a phishing email detection method based on LSTM (Long Short-Term Memory) neural networks, which is designed to improve accuracy and reduce false negatives and false positives in detecting phishing emails. The proposed method outperforms traditional methods by effectively capturing induced statements in emails, allowing for improved detection rates. The authors mention a challenge in phishing detection which is the flexibility of phishing emails, allowing attackers to change the header features and evade detection systems based on ML.

A35. Li et al. (2020) present a phishing email detection method based on the persuasion principle, focusing on the human factor in cybersecurity. People are often considered the weakest link in security systems, and the persuasion principle provides valuable insights into how phishing emails exploit human vulnerabilities. The article highlights the challenges in phishing detection in emails, including the need for effective feature extraction and classification methods that capture the persuasive nature of phishing emails. Additionally, there is an ongoing need to monitor the evolution of phishing emails and refine feature extraction to enhance email classification accuracy.

A36. Magdy et al. (2022) propose an Artificial Neural Network (ANN) based email classifier for detecting spam, ham, and phishing emails. The study emphasizes the importance of careful feature selection and efficient classification. Misclassifying emails can result in losing essential information or exposure to attacks designed to steal user data.

A37. Mahdavifar & Ghorbani (2020) introduce DeNNeS, a deep-embedded neural network expert system designed for cyber threat detection. The authors highlight challenges include dealing with noisy or outlier data, managing missing feature values, and enhancing the expert system’s flexibility and robustness.

A38. Mukhopadhyay & Prajwal (2021) focus on developing an efficient and effective solution to detect and prevent phishing and trojan attacks in emails, proposing a system called E.D.I.T.H (Email Disintegration Intrusion-Detection of Trojan Hacktool). The authors acknowledge that there is room for improvement and future work. The main challenge lies in automatically obtaining emails without relying on third-party mail clients, such as Thunderbird.

A39. Muneer et al. (2021) present various phishing detection techniques and their effectiveness, emphasizing the challenges and limitations of each method. Some detection techniques, like the hybrid model, rely on the quality and size of the corpus, which may cause time-consuming analysis. Existing techniques may not consider graphical elements in HTML-formatted emails, limiting their effectiveness. There is no current list of blacklisted keywords, limiting the effectiveness of certain detection techniques. The article concludes that public awareness and preventative measures against phishing attacks are crucial, particularly focusing on areas with high internet usage.

A41. Nabeel et al. (2021) address the challenge of detecting unwanted emails (including phishing emails) in an enterprise setting, particularly in the context of end-to-end email encryption (E2EE) that prevents access to email content. The authors note that traditional approaches relying on email content or sender profiling features are ineffective in this scenario. The authors acknowledge that there are limitations to their approach, including the fact that header features can be manipulated by adversaries and that the structure of the enterprise graph may change over time, requiring regular retraining of the classifier. They also note that the performance of the classifier may be affected by using fresh senders or well-crafted emails that avoid header features.

A43. Nishikawa et al. (2020) propose a method to estimate the presence of manipulation methods in email content with Cialdini's principle. They also mention the challenges in their proposed method:

- Defining unique keywords for Cialdini's principles is not easy, and rule-based identification is difficult.
- Acquiring actual samples of malicious emails is extremely difficult, leading to the use of publicly available datasets like the Enron Email Dataset.
- Cialdini's principles can also be present in legitimate emails, making it hard to determine if an email is malicious based solely on these principles.
- The performance of the proposed models (baseline and neural network) is not adequately high (below 80% AUC).
- Different Cialdini's principles may require different models or feature information, leading to varying performance across principles.

A44. Ozker & Sahingoz (2020) present a study focused on developing a content-based phishing detection system by utilizing ML algorithms to analyze text and additional properties of web pages. By employing Python programming language, the authors processed 58 different features related to email and website content. The authors acknowledged that there are ongoing challenges in phishing detection for organizations, as attackers continually develop new tactics to bypass security measures.

A45. Paliath et al. (2020) propose an approach for phishing detection by using content-based feature extraction, which has proven to be very effective in phishing detection. Six different well-known classification algorithms were compared in this experiment, these algorithms were Rough Set Theory, SVM, Naïve Bayes, RF, Neural Networks, and Random Tree. The authors mention that there are remaining challenges in classification performance and error rates, which the authors plan to address in future research.

A46. Patil & Arra (2022) discuss the challenges of phishing detection in email, particularly for organizations. The article highlights the latest techniques for phishing detection, such as Visual selection features, ML, DL, and AI. The article further highlights that human beings are considered the weakest link. To minimize phishing attacks, a combination of user awareness training and technical detection methods is needed. The highlight of the challenges mentioned by the authors:

- Attackers constantly develop new methods and techniques to evade defensive mechanisms.
- Achieving high accuracy rates while maintaining low false-positive and false-negative rates in detection techniques.
- Humans are the weakest link, being the source of 95% of security attacks.

- The effectiveness of training is influenced by various factors, such as religious indicators and social peer pressure.
- The need to combine user awareness training and technical detection methods to minimize phishing attacks.

A47. Razak et al. (2022) evaluate phishing detection methods by comparing five ML classifiers. The study emphasizes that selecting the most efficient and accurate classifier is crucial for organizations to protect their confidential information from phishing attacks. The article highlights several challenges. Users often overlook the complete address of online pages when they receive URLs via emails, social media networks, or other sources, making it easier for phishing attacks to succeed. Moreover, the rapid pace of life and the constant influx of information contribute to users' carelessness, increasing the risk of falling for phishing scams. Another challenge is that anti-phishing solutions rely on blacklists, whitelists, or ML algorithms to block phishing attempts, which may not always be accurate or up to date.

A48. Rustam et al. (2023) present an ML-based approach for spam email detection, focusing on improving accuracy and addressing the challenges of data imbalance and feature extraction. The challenges mentioned in the article include the high computational time required for ML models that use the feature union approach and the small size of the dataset with an imbalanced target class ratio.

A49. Salahdine et al. (2021) propose a phishing attack detection technique based on ML comparing the performance of three classifiers called SVM, LR, and ANN. The dataset used for training and testing the models contained 4000 instances, with 2000 phishing emails and 2000 legitimate emails. The authors acknowledge the challenges come from the sophisticated techniques employed by cybercriminals, the large volume of incoming emails, and the need for timely detection.

A50. Salau et al. (2021) propose a blockchain-based solution for sharing phishing data among organizations, which aims to enhance the detection and prevention of phishing attacks. The authors mention the following challenges in phishing detection: Challenges in phishing detection in emails for organizations mentioned in the article include:

- Concurrency of transactions: Blockchains must be able to handle a high number of transactions in a short time, as phishing attacks can happen frequently.
- Transaction approval times: The approval times for phishing email reporting transactions depend on miner activity and other transactions running on the blockchain.
- Memory consumption: The ledger size grows with an increasing amount of reported phishing activity on the blockchain.

A51. Salloum et al. (2022) conducted an SLR and present several challenges in phishing detection:

- Inadequate feature selection: The article emphasizes the importance of selecting the most relevant features for effective phishing detection.
- Insufficient exploration of natural language processing (NLP) techniques: There is a need to focus more on NLP techniques in phishing detection research. The use of semantic analysis in the email content can help identify the sender's intention and better detect phishing emails.
- Limited availability of reference datasets: The article highlights the difficulty of obtaining reference datasets for phishing detection due to the ever-evolving nature of phishing attacks.
- Lack of standardized benchmarks: The absence of standard benchmarks makes it challenging to accurately compare different phishing detection techniques.
- Insufficient focus on semantic features: Research should shift from non-semantic feature analysis towards semantic analysis, which can better identify the sender's intent and help detect phishing emails more effectively.

A53. Shalini et al. (2022) address the challenges on the subject and propose an ML- and DL-based solution. The article highlights the difficulties organizations face, emphasizing the importance of understanding the characteristics of genuine and fraudulent emails to enhance cybersecurity and that attacks are difficult to detect due to the deceptive nature of fake emails, which often appear legitimate.

A54. Sheikhalishahi et al. (2020) present a methodology and framework for automatically clustering and classifying large sets of spam emails into labeled campaigns. The proposed approach utilizes ML and clustering algorithms. The authors highlight the importance of fast and effective analysis of spam emails in identifying cyber threats. The major challenge mentioned in the article is the difficulty in discerning between two campaigns whose emails have a similar structure, which may come from different attackers.

A55. Shi et al. (2021) present an approach to predict whether a development email contains feature requests by extracting semantic sequence patterns. The authors mention that the challenges in detecting feature requests in emails for organizations are related to balancing the cost and risk of analyzing a large number of emails and the extensive application of this approach to other textual resources, such as issue tracking systems, open forums, and project reviews.

A56. Shmalko et al. (2022) present a distributed phishing risk assessment tool called Profiler that combines the following three email profiling dimensions to detect phishing:

- Threat level
- Cognitive manipulation
- Email content type

This tool has been designed to address challenges with ML-based phishing detection algorithms that often struggle with concept drift, a phenomenon where phishing attackers modify the statistical characteristics of phishing attempts to bypass detection. The authors highlight the following challenges in phishing detection:

- ML models require large collections of data sets, containing both legitimate and phishing emails, to train on.
- Over time, ML models become less effective as phishing attack emails evolve rapidly.

A57. Sonowal (2020) presents a method for phishing detection in emails by identifying the best feature set with high accuracy and minimum features, the method involves three algorithms. The author highlights the challenges regarding phishing detection listed below:

- Fast detection is essential to minimize the impact of phishing emails.
- Identifying the best features is crucial for accurate detection.
- Trade-off between precision and recall, improving one can reduce the other.

A58. Sun et al. (2022) propose a decentralized phishing email detection system called Federated Phish Bowl using global word embedding and bidirectional long short-term memory. The authors highlight the challenge of adapting ML-based phishing email detection systems being limited to accessible training data, as privacy concerns restrict centralized learning for personal data-containing email contents.

A59. Venkatraman et al. (2020) present an approach to combat spam e-mails using an enhanced Naïve Bayesian algorithm that combines conceptual and semantic similarity approaches. The experiments were conducted on different data sets to evaluate the effectiveness of the proposed approach. Further, the authors also mention some challenges in spam e-mail classification; Training with more email documents may reduce the classifier accuracy through overfitting, however, training with fewer email documents may not be enough to classify spam emails accurately.

A60. Wang (2022) mentions that phishing emails can be very sophisticated and difficult to detect using traditional methods such as spam filters. This, along with attackers constantly adapting their techniques to evade detection, makes it challenging to keep up with new methods of attack. The author further mentions the challenge that the volume of emails that need to be analyzed is enormous, which can be time-consuming and resource intensive. Lastly, the author mentions that features used to classify emails as phishing or legitimate may change over time, requiring regular updates to the detection system.

A62. Zhao et al. (2021) discuss computer vision to tackle challenges in phishing detection. The major challenge they refer to is the lack of large, recent security datasets for benchmarking and comparing different solutions.

A64. Zhu (2020) presents an innovative phishing attack detection system based on advanced computational intelligence methods. The proposed intelligent system was trained using a complex large-scale dataset, resulting in successful identification of phishing attacks. The author mentions the main challenge is the high variability of phishing attacks, which makes it difficult to develop a single, static model that can detect all types of attacks. Another challenge mentioned is the imbalance between legitimate emails and phishing emails in data sets, which can lead to biased or inaccurate models. Additionally, the article mentions that phishing attacks are becoming more sophisticated and difficult to detect, which requires advanced computational intelligence methods to identify new attack patterns as well as the challenge that the high-dimensional nature of the data, which requires efficient algorithms to analyze large amounts of data quickly and accurately.

5.1.2 Private Users

A2. Adil et al. (2020) explore various techniques to prevent and combat phishing attacks, emphasizing the importance of installing Intrusion Detection Systems (IDS) and Intrusion Prevention Systems (IPS) in networks. Different experimental analyses were conducted to compare the results of these techniques which showed a significant improvement in phishing attack prevention. The authors mention that challenges in detecting phishing attacks in emails remain, the current awareness and education against phishing attacks are not sufficient, as many attacks still occur due to end-user negligence. Furthermore, advanced phishing attacks can still bypass some detection systems and tools, making it difficult to provide a comprehensive defense.

A4. Ahammad et al. (2022) propose an approach to phishing detection that analyses the email text body with ML algorithms. Tokenization, stemming, and module cloud visualization were used to create a corpus of 100 phishing-related words. A new dataset with each phishing word as a feature was created, and various ML models were applied, including LR, DT, SVM, Naïve Bayes, and KNN. They conclude that feature reduction and ML algorithms have a significant impact on phishing email detection. The authors talk about challenges and future work, mentioning that the constantly evolving tactics used by attackers are a challenge when it comes to phishing detection. The need to develop a user-friendly solution to help identify phishing emails effectively is needed.

A10. Bansal & Sidhu (2021) propose a hybrid method that combines Term Frequency-Inverse Document Frequency (TF-IDF) and ANN for detecting spam and phishing emails. The proposed hybrid ANN and OCR method aims to address these challenges by offering a more accurate and efficient way of detecting spam and phishing emails. It considers both text and HTML templates and shows improved performance compared to other methods like Naïve Bayes and XGBoost. Challenges for phishing detection in emails discussed by the authors are summarized below:

- The rapid growth of cybercrime and the increasing number of spam and phishing emails, make it a difficult subject to investigate.
- Traditional anti-phishing technologies are not always effective, as hackers continue to develop new tactics and techniques to bypass security measures.
- The need for a more comprehensive approach to detect spam and phishing emails, as many existing methods focus solely on text content and do not consider other formats like HTML templates.

A11. Bathala et al. (2021) present an approach to classify emails containing phishing URLs and spam images. The LazyPredict library is used for selecting the best-performing ML models. Even though there are improvements in accuracy from their approach, the authors discuss challenges for phishing detection in general, and are listed below:

- As phishing attacks evolve, constant improvements and updates in different detection models are required to maintain their effectiveness against phishing attacks.
- The validation accuracy for some of the models is lower than their training accuracy which indicates potential overfitting or a need for further investigation.

A15. Das et al. (2020) provide a comprehensive review of phishing and spear phishing detection techniques and user studies, with a focus on security challenges. The authors discuss the importance of user studies to evaluate how people respond to phishing attacks, where the need for further analysis of real-world studies with unrestricted participants. Challenges in phishing detection according to the authors are listed below:

- Limited research on spear phishing attacks, which are harder to detect and require a better understanding of users' behavior and targeted training techniques.
- Lack of proper datasets and evaluation metrics for better comparison and improvement of phishing detection techniques.

A17. Ding et al. (2021) address the challenge of detecting spear phishing emails by proposing an approach that combines stylometric, forwarding, and reputation features by improving the Synthetic Minority Oversampling Technique (SMOTE) algorithm to reduce the impact of unbalanced data, the new and improved algorithm is named KM-SMOTE. Logs or other historical email records are not needed for this method to work making it extensive and adaptable for organizations.

Experiments on a dataset containing 417 spear phishing emails and 13,916 non-spear phishing emails were conducted, comparing the performances of the SMOTE and KM-SMOTE algorithms. The results indicate that the new and improved KM-SMOTE algorithm outperforms the SMOTE algorithm. Using an enhancement ratio of 1.8, the authors achieved a maximum recall of 95.56%, a precision of 98.85%, and an F1-score of 97.16%. They also mention that the need for better features to distinguish spear phishing emails from phishing and benign emails and the impact of unbalanced data on detection performance is a challenge for spear phishing detection.

A18. Drury & Meyer (2020) analyze email account separation as a potential approach to prevent phishing attacks, arguing that attackers rarely obtain victims' email addresses directly from the services they impersonate, which makes it possible for users to recognize phishing attempts by using separate email addresses for different services. They mention that while email separation is found to be effective against general phishing, it is less effective against advanced types of phishing. The authors discuss challenges and open questions with their proposed approach listed below:

- Email Address Leakage: Email addresses are not always kept secret, and sometimes need to be shared for effective use of services. Once an email address leaks, it becomes necessary to contain the damage. Users might put too much trust in the separation system and become more susceptible to phishing emails.
- General Applicability: Email address separation is not applicable in all situations, such as organizations that do not allow users to choose their email addresses. Attackers might also shift to other methods like social media, Smishing, or Vishing, where accounts are harder to separate.
- Usability: To make email address separation more user-friendly, the process of creating different email addresses should be simplified, supported by email clients, and include tools for managing large numbers of addresses. Changing an email address after a data leak should also be convenient.
- Awareness: Email separation might increase user awareness, as they actively think about separating addresses. This could raise suspicion towards unknown senders and decrease the likelihood of falling for phishing attacks.

A19. Farahisya & Bachtiar (2022) present a study on spam email detection using Recurrent Neural Networks (RNN), experiments were conducted by adjusting parameters like the number of RNN layers, hidden size, optimizer, learning rate, and the number of epochs to get the best results. Their RNN approach was compared with other ML algorithms called Naïve Bayes and SVM. The suggested RNN approach outperformed Naïve Bayes and SVM, achieving 99.1% precision and 99% accuracy when using affect intensities. The authors mention that using N-grams based on phrases in the tokenization process would help improve the detection of spam emails and further enhance the performance of the RNN-based model. One challenge that the authors mention is the limited effect of basic emotional features used in the study, as emails might contain more complex emotions and text.

A24. Gogoi & Ahmed (2022) proposes a DL approach for phishing detection using pre-trained transformer models, specifically BERT and DistilBERT. No significant overfitting or underfitting was observed during the study. They further discuss challenges in phishing detection, including the constantly evolving nature of phishing attacks and the inefficiency of traditional feature extraction techniques in detecting phishing emails effectively. The proposed transformer-based approach successfully addresses these challenges by achieving high accuracy without relying on complicated feature extraction techniques.

A28. Jalda et al. (2022) present an approach for detecting phishing attacks and email spoofing using an ML algorithm called the stacking algorithm. The architecture of the proposed model consists of pre-processing, feature extraction, base model training, predictions, and meta-model training. They mention that the stacking algorithm is complex to understand but had good accuracy with high-speed processing data in terms of very quick results, having an accuracy score of 96%. The proposed model can improve the accuracy and speed of detecting phishing attacks and email spoofing, however, this method has not been implemented to verify the results and is listed as future work in the author's work. The challenges that the authors mention are the importance of intelligent pre-processing and feature extraction, as well as the use of a robust stacking algorithm.

A40. Muralidharan & Nissim (2023) present an ensemble DL framework for malicious email detection that outperforms state-of-the-art expert feature-based models. It utilizes the entire email, including header, body, and attachment, to generate features automatically, eliminating the need for human-defined features. The framework was evaluated on a dataset of 32,676 emails obtained from VirusTotal, demonstrating improved performance over existing methods.

The authors propose a framework that automatically generates features without human intervention, allowing it to adapt to changing trends in phishing attacks. They mention that the proposed framework outperforms state-of-the-art models with shorter development times, making it more practical for dealing with evolving threats by using the entire email to improve detection capabilities and reduce false positive rates. The authors also highlight that their proposed framework demonstrates language robustness, making it suitable for handling emails in various languages. Challenges in phishing detection in emails that the authors have mentioned are listed below:

- Malicious email campaigns are constantly changing, requiring adaptive solutions not dependent on fixed feature sets.
- Traditional methods that don't consider the entire email can have limitations, such as higher false positive rates for emails with attachments.

A42. Niroshan Atimorathanna et al. (2020) present NoFish, a novel heuristic solution to detect phishing attacks, which combines a browser extension and an email plugin. With an overall accuracy level of 97%, NoFish can detect phishing attacks in four domains: DNS phishing, URL phishing, email phishing, and phishing detection using image comparison. For email phishing detection, NoFish focuses on Microsoft Outlook and uses a Naïve Bayes Classifier to detect spam emails and faces challenges in detecting phishing websites without key features and needs to expand its email phishing protection to other email providers to be more effective for private users.

A52. Saraswat & Singh Solanki (2022) present a method for classifying emails as phished or ham using ML techniques. The dataset is pre-processed, and important features are extracted using regular expressions in Python. Supervised learning algorithms, including SVM, RF, LR, Naïve Bayes, and Voted Perceptron classifiers, are utilized to classify the emails. The results of the study show promising classification accuracy according to the authors, however, the dataset used may not represent real-world conditions, and the proposed method can be improved by expanding the dataset to include a variety of phished and ham emails. The article suggests developing an official framework that could be used across organizations and privately to protect users from phishing attacks.

A61. Wu & Guo (2022) introduces Holmes, a lightweight semantic-based anomalous email detector designed to effectively discover malicious emails in real-world cyber threat hunting. Holmes uses a combination of content analysis, correlation graph analysis, and ML models to identify and cluster anomalous emails. Holmes has been deployed in an enterprise environment and demonstrates its practical value in commercial competition by outperforming several well-known commercial detectors offered by security companies in VirusTotal.

The authors mention that malicious emails often manage to bypass traditional anti-spam gateways, making it necessary to have a behavioral anomaly detection system like Holmes for effective detection and that many phishing domains have short life spans, making it difficult for traditional email detectors to identify them. Holmes, on the other hand, is effective in detecting these domains.

Challenges in detecting phishing emails mentioned by the authors are listed below:

- Many phishing domains have short life spans, making it difficult for traditional email detectors to identify them.
- Malicious emails often manage to bypass traditional anti-spam gateways, making it necessary to have a behavioral anomaly detection system like Holmes for effective detection.
- As cyber threats continue to evolve, detection systems like Holmes must adapt by incorporating new techniques and data sources, and refining machine learning models for better performance.

A63. Zhu et al. (2022) present a lightweight phishing detection model that combines different mechanisms for efficient and accurate detection of phishing websites in email. The authors mainly discuss the challenge where attackers become more sophisticated in their approaches.

6 Analysis

This chapter begins with showing a taxonomy representation of themes and sub-themes that have emerged using thematic coding and can be seen in Figure 7, where the articles highlighted in blue are relevant for organizations, red for private users, and purple for both. The themes have been designed in collaboration between the authors, resulting in the same main themes and branching out to sub-themes to separate the result. The chapter continues with presenting the thematic coding in written form for organizations and private users respectively, the articles that are categorized as both are put together with the private users. Section 6.1 shows articles for organizations only, and section 6.2 shows articles for private users. Chapter 7 comes with an explanation of the themes and sub-themes and the results are analysed and compared with each other. The earlier mentioned name scheme in chapter 5 will be used in the analysis. The full list of accepted articles can be seen in Appendix A.

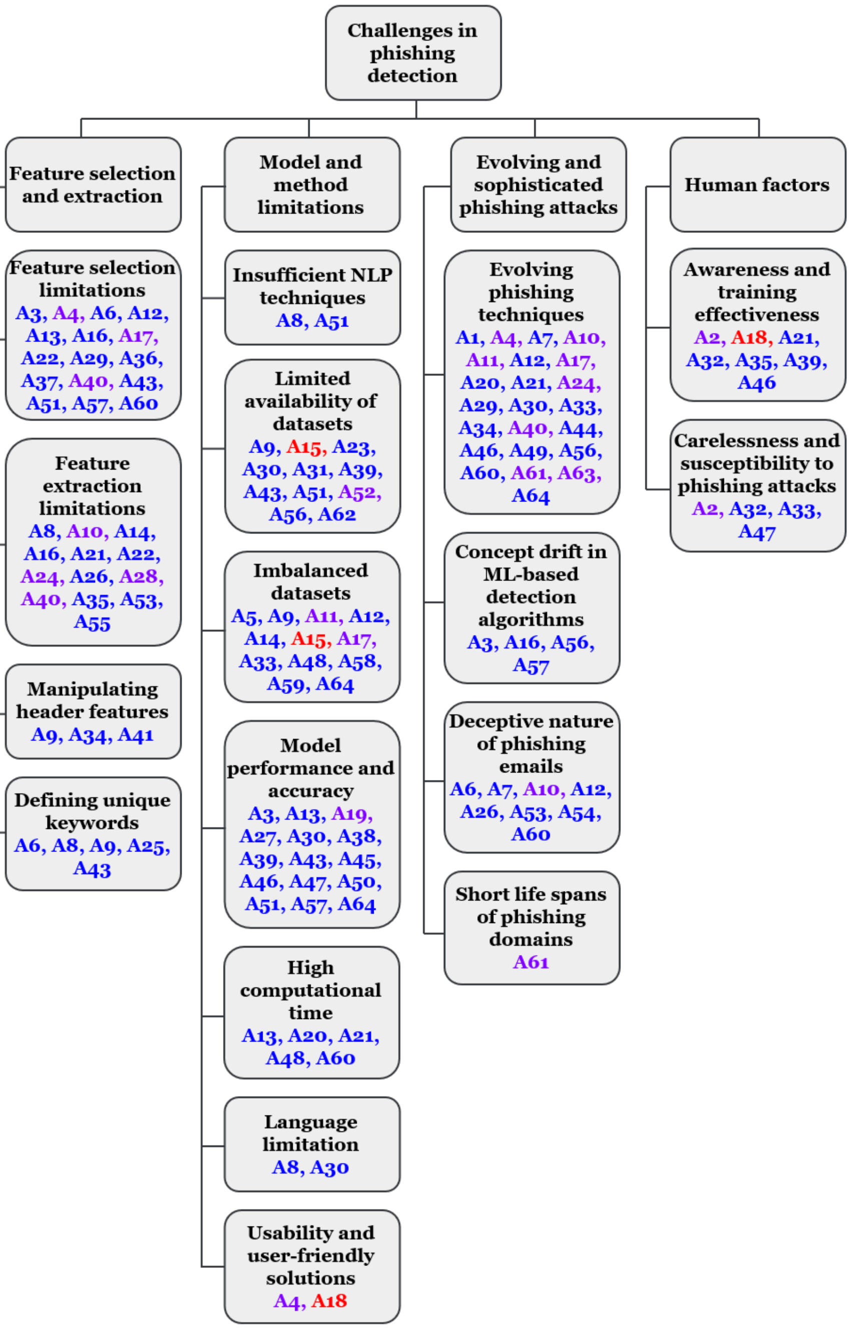

Figure 7. Taxonomy of challenges in phishing detection (author's own)

6.1 Organizations

After conducting a thematic analysis of the articles, focusing on challenges in phishing detection as the main theme, the following key sub-themes emerged for the organizational focus:

Theme 1: Feature selection and extraction

- Feature selection limitations (A3, A6, A12, A13, A16, A22, A29, A36, A37, A43, A51, A57, A60)
- Feature extraction limitations (A8, A14, A16, A21, A22, A26, A35, A53, A55)
- Manipulating header features (A9, A34, A41)
- Defining unique keywords (A6, A8, A9, A25, A43)

Theme 2: Model and method limitations

- Insufficient NLP techniques (A8, A51)
- Limited availability of datasets (A9, A23, A30, A31, A39, A43, A51, A56, A62)
- Imbalanced datasets (A5, A9, A12, A14, A33, A48, A58, A59, A64)
- Model performance and accuracy (A3, A13, A27, A30, A38, A39, A43, A45, A46, A47, A50, A51, A57, A64)
- High computational time (A13, A20, A21, A48, A60)
- Language limitation (A8, A30)

Theme 3: Evolving and sophisticated phishing attacks

- Evolving phishing techniques (A1, A7, A12, A20, A21, A29, A30, A33, A34, A44, A46, A49, A56, A60, A64)
- Concept drift in ML-based algorithms (A3, A16, A56, A57)
- Deceptive nature of phishing emails (A6, A7, A12, A26, A53, A54, A60)

Theme 4: Human factors

- Awareness and training effectiveness (A21, A32, A35, A39, A46)
- Carelessness and susceptibility to phishing attacks (A32, A33, A47)

6.2 Private Users

After conducting a thematic analysis of the articles aimed toward the regular user, the following themes emerge as challenges in phishing detection:

Theme 1: Feature Selection and Extraction

- Feature selection limitations (A4, A17, A40)
- Feature extraction limitations (A10, A24, A28, A40)

Theme 2: Model and method limitations

- Limited availability of datasets (A15, A52)
- Imbalanced datasets (A11, A15, A17)
- Model performance and accuracy (A19)
- Usability and user-friendly solutions (A4, A18)

Theme 3: Evolving and sophisticated phishing attacks

- Evolving phishing techniques (A4, A10, A11, A17, A24, A40, A61, A63)
- Deceptive nature of phishing emails (A10)
- Short life spans of phishing domains (A61)

Theme 4: Human factors

- Awareness and training effectiveness (A2, A18)
- Carelessness and susceptibility to phishing attacks (A2)

7 Comparative analysis

Here, a comparative analysis is presented, using the analysis from chapter 6 and comparing them to each other. Each section explains the emerged themes and sub-themes followed by each sub-theme. The sub-themes start by discussing the organization part, followed by the part for private users in paragraph two, just below. If a section only contains one paragraph, it will be stated to which perspective it belongs. A comparative analysis is made at the end of each 7.x section, so for example, the comparative analysis of theme 1 in 7.1 is made in 7.1.5.

7.1 Theme 1: Feature Selection and Extraction

Feature selection refers to what properties of an email should be analyzed by a machine learning algorithm, it can for example be words or phrases, metadata, or the structure of the email. Feature extraction means transforming the email's raw data into features or attributes used for learning a machine learning algorithm. The third sub-theme is manipulating header features in emails and the last sub-theme that emerged has to do with the difficulties in defining unique keywords when it comes to phishing detection. The keywords are used to search for phishing data.

7.1.1 Feature Selection Limitations

A lot of articles discuss the limitations in feature selection, A3, A6, A12, A29, and A60 explain very well that feature selection usability gets limited due to the constant changes in the attacker's methods. A3 and A16 mention that the changes lead to the need for re-learning and selecting new features as well as earlier important features instead of getting obsolete, which in the long run might lead to overfitting meaning that the model gets "trained too well" which instead hinders the models' performance. A13, A36, A37, A43, and A51 highlight the need for tuning parameters and careful selection procedures as major challenges.

A4 proposes an approach that analyses the email text body with ML algorithms. Feature reduction is highlighted to have a significant impact on phishing email detection. However, the approach has limitations as attackers are constantly changing their methods, and the technique works best when phishing words are present in the email body. A17 proposes an approach for detecting spear phishing emails by improving the SMOTE algorithm. A17 mentions that there is a need for better features to distinguish spear phishing emails from benign emails. A40 presents a DL framework for phishing email detection, by using the entire email to generate features automatically. A40 mentions that human-defined features can become obsolete, and constructing new features is a time-consuming process.

7.1.2 Feature Extraction Limitations

A8 highlights the need for domain experts when it comes to extracting the correct features, A14, A16, A21, A22, A26, A35, A53, A55, and A57 agree with this and further elaborate on the difficulties. Particularly in distinguishing legit emails from phishing emails and achieving high accuracy because of constant changes in attacks and the limited feasibility of examining high amounts of incoming emails. A57 also mentions the importance of identifying the most crucial features.

A10 proposes a hybrid method that considers both text and HTML templates and shows improved performance compared to other methods. A10 mentions that one challenge when it comes to phishing detection is that many existing methods focus solely on text content and do not even consider other formats. A24 proposes a DL approach for phishing detection and highlights the inefficiency of traditional feature extraction techniques as a challenge and limitation. A28 uses the stacking algorithm, which is an ML algorithm to detect phishing attacks. Both A28 and A40 argue that traditional methods that do not consider the entire email has limitations when it comes to feature extraction.

7.1.3 Manipulating Header Features

Organizations only: A9, A34, and A41 mention that emails need to be cleaned and pre-processed before being fed into a model. This gets difficult due to the flexibility in phishing emails with frequent manipulation of the email header which allows attackers to evade detection. This often leads to re-learning the model to keep up with the most current threats.

7.1.4 Defining Unique Keywords

Organizations only: A6, A8, A9, and A43 mention the challenges in establishing keywords due to more sophisticated, well-tailored, and continuously changing phishing emails. Continued, they emphasize the difficulty in determining whether an email is malicious or not. A25 suggests that future work in phishing detection needs to focus on combining word embedding with techniques to improve the interpretability of results and better fit pre-trained models to new tasks.

7.1.5 Comparison and Conclusion

The challenges regarding features are present from both viewpoints, and while both groups face issues with the limitations of traditional feature extraction techniques, organizations seem to grapple more with scalability, the need for expert involvement, and the importance of accuracy while private users are more focused on facing more issues around the inefficiency of current detection methods. The problem of manipulated header features and the difficulty of defining unique keywords seem to be of more relevance for organizations. However, both perspectives highlight the same fundamental limitation: the time-consuming nature of identifying and implementing new features due to the constant evolution of phishing methods. The necessity to continuously relearn and adapt to new threats is a common thread, emphasizing that phishing detection strategies need to be as dynamic and adaptable as the threats they aim to counter.

7.2 Theme 2: Model and method limitations

There are several NLP techniques to mitigate phishing attacks. However, a sub-theme that emerged during the analytical phase shows that there are insufficient NLP techniques for organizations. Datasets can be used for benchmarking different ML phishing detection systems amongst other things. The confidential nature of phishing emails and the constantly evolving nature of phishing attacks are two reasons why there is limited availability of datasets. Datasets can become imbalanced, negatively affecting the performance of phishing detection models. Model performance and accuracy are important when it comes to phishing detection, and it is a challenge to get it right. High computational time means that it takes too much time for an algorithm to run efficiently, making performance decline. Most existing phishing detection research focuses on the English language, potentially making it harder to detect non-English phishing emails. The need for usability and user-friendly solutions when it comes to phishing detection is a recurring challenge when analyzing articles.

7.2.1 Insufficient NLP Techniques

Organizations only: A8 and A51 agree that there is a need to focus more on NLP techniques since these techniques suffer from rapidly changing phishing methods. They further agree that using NLP techniques to analyze emails, can greatly improve the ability to detect phishing emails by identifying the intentions of the email.

7.2.2 Limited Availability of Datasets

Many articles mention the limited availability of datasets, A9 mentions that because of the confidential nature of emails, there are few publicly available datasets, A23, A30, A31, A39, A43, and A62 mention that a small number of datasets can lead to misclassification of some attributes. Further, A51 and A56 believe that there are few datasets due to the ever-evolving nature of phishing attacks. A58 adds that privacy concerns restrict centralized learning for personal data-containing email content.

A15 conducted a review of phishing detection techniques and user studies, highlighting that there is a lack of proper datasets for better comparison and improvement of phishing detection techniques for private users. A52 uses ML techniques to classify phishing emails, the dataset used is pre-processed and important features are extracted by using regular expressions in Python. A52 mentions that the dataset used may not represent real-world conditions and that the dataset can be expanded for better performance.

7.2.3 Imbalanced Datasets

A5, A9, A12, A14, A48, and A64 discuss the imbalance of datasets and agree that it often leads to the needing the development of specialized algorithms and techniques to handle imbalances effectively. Continued, an imbalance can negatively affect the performance of the detection model such as biased or inaccurate models. Additionally, the size of the training dataset plays a significant role in the performance of the algorithms. A33 mentions that learning models must be tuned for specific datasets. A59 and A64 state that training with more email documents may reduce the classifier accuracy through overfitting, however, training with fewer email documents may not be enough to classify spam emails accurately.

A11 proposes an approach to classify emails containing phishing URLs and spam images using ML models. The validation accuracy for some of the models is lower than their training accuracy, indicating potential overfitting due to imbalanced datasets. Both A15 and A17 mention that the impact of unbalanced data is a challenge when it comes to phishing detection.

7.2.4 Model Performance and Accuracy

A3 mentions that high performance may require careful selection of kernel functions and parameter tuning, and feature independence may not hold in real-world data potentially lowering model accurateness, A13 agrees and further adds the lack of understanding of DL models as well as the difficulty in detecting structural changes in URLs. A27 and A30 add that despite high accuracy rates achieved by ML algorithms, no model has reached 100% accuracy. A38 means that it is important and hard to achieve high accuracy without relying on third-party mail clients. A39 mentions that the existing techniques may not consider graphical elements in HTML-formatted emails, limiting their effectiveness. A43 proposed models where the performance of their proposed models is not adequately high. A45 mentions that there are challenges in classification performance and error rates. A46, A47, A50, A51, and A57 agree that achieving high accuracy rates while maintaining low false-positive and false-negative rates in detection techniques is a major challenge. A64 concludes the challenges well by stating that the main challenge is the high variability of phishing attacks, which makes it difficult to develop a single, static model that can detect all types of attacks.

A19 presents an approach using an ML algorithm called RNN to detect phishing emails, showing that their proposal outperforms other ML algorithms. A19 mentions that their model, like many other models, can be further enhanced to gain better performance.

7.2.5 High Computational Time

Organizations only: A13 highlights the need for fine-tuning hyperparameters, leading to high computational time for models. A20 adds to this by stating that it is challenging to find the optimal balance between achieving high classification performance and managing computational resources. A21 and A60 emphasize that the extensive computational demands when conducting an in-depth security assessment of all received emails are impractical. A48 mentions that high computational time is required for ML models that use certain approaches, the feature union approach is mentioned.

7.2.6 Language Limitation

Organizations only: A8 and A30 agree that most existing phishing detection research focuses on English-language emails, potentially limiting the applicability of these methods to non-English email phishing attempts.

7.2.7 Usability and User-friendly Solutions

Private users only: A4 concludes that the constantly evolving strategies by phishing attackers and highlights the need to develop a user-friendly solution are challenges within phishing detection. A18 proposes email account separation to prevent phishing attacks for private users. A18 mentions that this strategy is effective against phishing in general but less effective against more advanced types of phishing. A18 suggests that the process of creating different email addresses should be simplified, supported by email clients, and include tools for managing large numbers of addresses.

7.2.8 Comparison and Conclusion

Challenges in NLP, high computational time, and language limitations are problems that seem to only occur for organizations, which possibly is due to the many ML models discussed for organizations. While instead, a unique challenge for private users is the development of user-friendly solutions. Challenges concerning datasets have a common thread for both perspectives, often indicating or mentioning that the problem originates from the constant changes in attackers' approaches. The challenges for organizations when it comes to model performance appear to be more complex, involving both technical aspects of ML models and the characteristic variability of phishing attacks. For private users, the primary challenge seems to be achieving the best possible performance from the ML models.

7.3 Theme 3: Evolving and sophisticated phishing attacks

Many articles mention the difficult challenges regarding the constant and rapid evolution of phishing techniques. If one technique is detected, a small tweak from the attacker and the detection system might be vulnerable yet again, making it a clear theme. Concept drift refers to ML models depreciating over time, often because of earlier mentioned evolving attacks. Lastly, the deceptive nature of phishing emails emerged as a complementary sub-theme.

7.3.1 Evolving Phishing Techniques

A lot of articles mention that many other challenges originate from the constant evolution and adjustments in phishing techniques. A1, A29, A30, A34, A44, A46, A49, A56, and A64 agree that the increasing sophistication of phishing attacks makes it difficult for users and algorithms to identify mischievous emails and websites. A7 specifically highlights that the combination of techniques makes it increasingly difficult to protect against. A12, A20, and A60 agree that adapting to new techniques and attack strategies is a major challenge. A21 means that to detect new techniques, one must depend on aware and alert users that receive the email to report it as malicious. A33 state that the biggest threat is the adaptive techniques used by phishers.

A61 introduces a lightweight semantic-based anomalous email detector using a combination of content analysis, correlation graphs, and ML models to identify and cluster phishing emails. A63 presents a lightweight phishing detection model. Both A61 and A63 mention that as phishing attacks continue to get more sophisticated, detection systems must adapt by incorporating new techniques and refining ML models for better performance. A4, A10, A11, A24, and A40 all agree that the evolving nature of phishing attacks requires adaptive solutions to maintain effectiveness against phishing attacks.

7.3.2 Concept Drift in ML-based Algorithms

Organizations only: A3, A16, and A56 explain and agree that over time, ML models are becoming less effective because of phishing attacks' rapid and continued evolvement. A57 means that identifying the best features is vital for accurate detection and avoiding concept drift, yet hard to achieve.

7.3.3 Deceptive Nature of Phishing Emails

A6 state that the challenges in phishing detection originate from the need to separate between legitimate and deceptive emails successfully. A7 adds that the similarity between URLs further increases difficulties in emails. A26, A53, A54, and A60 agree that attacks are challenging to detect due to the deceptive nature of forged emails. Further, A12 adds a unique statement that the shortage of shared information among organizations hinders the research in detecting phishing emails.

A10 states that the rapid growth of cybercrime and the increasing number of phishing emails make it a difficult subject to investigate. Many existing methods focus solely on text content and do not consider other formats of the deceptive nature of phishing emails.

7.3.4 Short Life Spans of Phishing Domains

Private users only: A61 mention that malicious emails often manage to bypass traditional anti-spam gateways, one reason for this is that many phishing domains have short life spans, making it difficult for traditional email detectors to identify them. The presented approach by A61 is more effective in detecting these domains.

7.3.5 Comparison and Conclusion

Both organizational and private users are aware of the increasingly evolving nature of phishing attacks. The constant evolution and adaptation of these techniques pose significant challenges for both groups. Both organizational and private users understand the need for adaptive measures to counter phishing attacks. They recognize the necessity of refining strategies, upgrading detection systems, and improving machine learning models to stay ahead of attackers. Concept drift in ML-based algorithms is a challenge that is only mentioned in the organizational articles, probably because many of the ML models are discussed in the articles related to organizations. Both organizational and private users have trouble distinguishing between legitimate and phishing emails due to their deceptive nature. Organizations mention that the similarity between URLs and the shortage of shared information among organizations are two major challenges when it comes to the deceptive nature of phishing emails. Private users argue that existing methods solely focus on text content and do not consider other formats for phishing detection. Short life spans of phishing domains are a challenge only mentioned in one article by private users.

7.4 Theme 4: Human factors

Due to a few articles strongly suggesting that humans are the weakest link, as well as the fact that this study is partly aimed toward private users, human factors seemed to fit as a theme. This theme is divided into two parts, where awareness and training are the first challenges, followed by humans' carelessness and susceptibility to getting exploited.

7.4.1 Awareness and Training Effectiveness

A21 and A46 agree that some proposed strategies rely on vigilant users to report phishing attempts to effectively detect attacks. A32 instead cautions against implementing embedded phishing training and exercises, which contradicts a lot of other literature, undoubtedly making users and user training a challenge. A35 and A46 firmly state that humans are considered the weakest link that is often exploited. A39 states that public awareness and preventive measures against phishing attacks are fundamental challenges. Lastly, A46 mentions that the effectiveness of training can get influenced by factors such as religious indicators and social peer pressure.

A2 emphasizes the importance of using IDS and IPS in networks to help combat phishing attacks because the current awareness and education against phishing attacks not being sufficient, as many attacks still occur due to end-user negligence. A18 states that email separation might increase user awareness, as they actively think about separating addresses. However, awareness remains a challenge for phishing detection.

7.4.2 Carelessness and Susceptibility to Phishing Attacks

A32 states that humans are a challenging weakness that easily gets tricked to open email as long as no apparent warnings are residing in the email. A33 and A47 add the liability of humans opening malicious content hidden behind a legitimate-looking URL.

A2 mentions that awareness and education are lacking for end users in most cases, making them more susceptible to phishing attacks. A42 mentions that it is not practical to educate users about all phishing scenarios, stating that it is human nature to be deceived.

7.4.3 Comparison and Conclusion

For both groups, the human factor is a significant challenge, both agree that the user is the weak link due to the susceptibility of falling victim to a phishing attack. Both organizational and private users face similar challenges in terms of human vulnerability, lack of awareness, and the need for vigilance. However, for organizational users, the effectiveness of training can be influenced by various factors, such as religious indicators and social peer pressure, and there is disagreement about the efficiency of training.

8 Discussion

This chapter will present discussions about the method, implementation, and results of the conducted study.

8.1 Previous Research

Our SLR aligns in many aspects with the findings of previous research. Consistent with the research of Das et al. (2020), we identified a considerable number of gaps in the existing body of knowledge, such as the challenges of deep learning techniques and the lack of information regarding feature importance in phishing detection. This is an area that our SLR has extended, considering both the organizational and private sectors, where the issue of feature selection remains.

Our SLR also is in line with the insights of Salloum et al. (2022), who highlighted the need for research using NLP techniques in phishing detection. Our review further confirms this shortcoming, with the added factor of comparison between organizational and private users. Importantly, our SLR found that the evolution of phishing tactics is a major challenge that transcends both sectors. This evolution, often rapid, emphasizes the need for continual innovation in detection techniques, including the potential exploitation of advanced technologies such as deep learning and NLP.

Our study differs from the work of Khonji et al. (2013) in that we focus not only on the technical challenges of phishing detection but also on the comparative challenges between organizations and private users. Khonji et al. (2013) noted that user education often falls short as a solution to phishing, a finding that aligns with our review, where the challenge of user education was a common thread across both sectors.

The user-centric studies by Howe et al. (2012) and Tekeni & Botha (2022) provided valuable insights into the specific challenges faced by private users. While Tekeni & Botha (2022) argued that many studies inappropriately use students as representative samples for home users, our SLR expands on this observation, noting a similar lack of diversity in user studies within the organizational context. The need for tailored security awareness campaigns that cater to different demographic segments within both sectors emerged as insight from our review.

The economic perspective offered by Görling (2006) provided a crucial understanding of the focus on organizational security. Our SLR, while acknowledging the importance of economic considerations in shaping security strategies, also highlights the need to pay equal attention to the security challenges faced by private users. Our review suggests that there is a need for balanced investment and effort in both sectors, as the success of phishing attacks in either sector can have significant societal implications.

In conclusion, while our SLR aligns with many of the findings from previous studies, it also presents insights, notably the shared challenges of phishing detection in emails between organizational and private users. The rapid evolution of phishing tactics is a shared concern across both sectors, indicating the necessity for continuous research and the potential for deeper collaboration between academia and industry.

8.2 Methods, Implementation, and Results

This SLR followed a structured methodology, literature search, selection, and data extraction. Although this approach enabled a comprehensive analysis of the topic, it is essential to consider potential limitations or biases that may have influenced the results.

The literature search and selection process may have impacted the results by potentially omitting relevant studies that could have provided different perspectives on the challenges faced by private and organizational users. This limitation can arise from the use of specific keywords in search strings, or database restrictions during the search phase. Additionally, publication bias may have led to an overemphasis on studies that report significant findings, thereby skewing the overall picture of the challenges in phishing detection.

In chapter 4.1.4, reverse snowballing is explained, and the original intention was to use it in this study but was not used due to time limitations. Reverse snowballing should also have been applied to articles relevant for private users. This resulted in that there was a difference in relevant articles between organizations and private users. Only three articles were relevant for private users, although many articles overlapped between the two sectors. In hindsight, the choice of excluding articles that discuss phishing detection in SMS, voice, and social media was most likely the wrong decision. Chances are that a lot of phishing attempts toward private users happen on those channels, which essentially opens opportunities for future research.

The data extraction process relied on the coding and categorization of the identified challenges, which may have introduced subjectivity in the interpretation and synthesis of the results. Although efforts were made to maintain objectivity and consistency, it is possible that alternative interpretations of the data could yield different conclusions.

The study identifies the dynamic nature of phishing attacks as a significant challenge in feature selection and extraction. Attackers continually adapt their methods, forcing constant re-learning and selection of new features. This dynamic can lead to overfitting and requires careful parameter tuning and selection procedures as well as the need for experts when it comes to extracting the correct features. The manipulation of email headers by attackers to evade detection is another challenge, often requiring models to be re-learned again to keep up with current threats. The study points out the difficulty in establishing unique keywords due to the sophistication and constant evolution of phishing emails. There are limitations in current models and methods, including insufficient techniques with performance issues, and limited and imbalanced datasets due to the continuous evolution of phishing attacks. As a note from theme four regarding the human factor, humans are often considered as the weakest link in the security chain with issues such as lack of awareness, carelessness, and the general susceptibility to phishing attacks. Consequently, this study emphasizes the evolving and complicated nature of phishing attacks as a significant challenge. The rapid and non-stop evolution is the source of many, if not all challenges identified in this study.

The challenges faced by both organizational and private users are remarkably similar, indicating that phishing attacks pose a universal threat. Both groups struggle with the dynamic and evolving nature of phishing attacks, which employ increasingly clever obfuscation techniques, making detection difficult.

8.3 Validity of Results

Beyond the information described in chapter 4.2, a requirement to analyze without negatively impacting the integrity of the results, further research is needed going forward. This study is carried out using a structured, repeatable methodology that is transparent, allowing readers to comprehensively judge its authenticity. Consequently, the findings presented are expected to be unbiased and reliable, any questionable or insufficient validity is believed to be effectively mitigated.

We have continuously been working together and have noticed flaws with this study, some flaws were realized during the working process but too late to fix due to the time limit of this study, and some when the study was finished. One flaw is that both the result and analysis sections were written in different paragraphs to separate the articles and the sectors. This was an active choice to try and make the text easier to read and improve the trackability of the text, but it might be hard to read for some. Another flaw is the number of accepted articles between organizations and private users due to both the time limit for this study and the limitations that we set ourselves. SMS, vishing, and social media were excluded in the early stage during this study since the focus was on phishing detection in emails, this led to the number of articles for private users being much lower than the number of articles for organizations. There have been challenges with this study from start to finish, both of us had a good grasp about phishing in general. However, reading up on different types of phishing detection methods was a bit overwhelming at first, we have helped each other to understand different methods and avoid the more technical approaches, like how certain algorithms work, as it is beyond our knowledge.

8.4 Ethical and societal aspects

To avoid any potential disputes because of the work done in this study, ethical aspects are monitored closely. It is important to correctly cite and interpret others' work and use it in the correct context to avoid causing any harm. Another ethical aspect that has been considered is reusing figures, to make sure that is not a problem every figure has been approved for use or has been created by one of the authors of this study.

Considering the ethical aspects, first, the process of collecting data in this SLR respects the rights of the original authors. This involves that proper citation and acknowledgments of all sources have been properly implemented and that plagiarism, in any form, has been avoided.

Second, this SLR tries to avoid any form of bias in the selection and interpretation of the literature. This review has been conducted objectively, and the findings are presented without any prejudice. This is an important part due to the comparative aim of this study, to not compare biased findings.

Third, the findings of this SLR can potentially influence practices related to phishing detection, therefore the conclusions drawn from this study are drawn, to the best of our knowledge accurate, and based on the analysis of the literature.

Consequently, one aspect that must be considered when it comes to ethical issues is the fact that weaknesses and challenges in phishing detection are discussed in this study. Reading this study, people with malicious intent could potentially try to utilize the exposed weaknesses by targeting those weaknesses. The consequences can be severe, ranging from financial loss and identity theft to the compromise of sensitive data. These outcomes can lead to a loss of trust in digital systems, which is harmful to the progress of a digitally connected society.

The methods discussed to detect and prevent phishing attacks must respect user privacy and data protection rights. While it is necessary to monitor digital communications for potential threats, this should not violate an individual's right to privacy. Balancing these two aspects, security and privacy is a significant ethical challenge in the field of cybersecurity.

Phishing attacks pose a considerable threat to society, affecting both individual privacy and organizational security. For private users, falling victim to a phishing attack can lead to identity theft, financial loss, and emotional distress. For organizations, successful phishing attacks can result in the compromise of sensitive data, financial losses, and damage to reputation. This SLR can contribute to societal awareness about the frequency and sophistication of phishing attacks. It can guide educational initiatives, public awareness campaigns, and community outreach programs to help people recognize, avoid, and report phishing attempts. By understanding the existing challenges in phishing detection, organizations and individuals can enhance their cybersecurity practices. Continued, by improving security and trust in digital communications, the findings of this SLR could help bridge the digital gap. If individuals feel safer online, they may be more likely to adopt digital technology, thereby benefiting from the opportunities it offers. Furthermore, this SLR can be linked to the United Nations Sustainable Development Goals by contributing to several goals (SDG, 2015):

- **Goal 4. Quality Education**: This SLR can inform educational initiatives to improve digital literacy and cybersecurity awareness.
- **Goal 8. Decent Work and Economic Growth**: By identifying and addressing the challenges in phishing detection, organizations can reduce their economic losses associated with cyberattacks.
- **Goal 9. Industry, Innovation, and Infrastructure**: Improving cybersecurity through better phishing detection aligns with this goal, as it encourages building resilient IT infrastructure.
- **Goal 16. Peace, Justice, and Strong Institutions**: Improving phishing detection can contribute to building effective and accountable institutions on all levels. Cybersecurity can help ensure peace by reducing cybercrime.

9 Conclusion

This SLR set out to examine the current challenges in phishing detection in emails, as well as focus on the difference in these challenges for private users versus organizational users. We aimed to find the difficulties of phishing attacks, their detection mechanisms, and the hurdles faced in both the private and organizational spheres. To our surprise, we discovered a high degree of similarity in the challenges faced by both groups.

Our analysis uncovered that both private and organizational users grapple with the dynamic nature of phishing attacks, particularly the increasing sophistication, and adaptability of phishing strategies which seems to be the very foundation for a lot of other challenges. Phishing emails have evolved beyond simplistic scams to complex, context-aware schemes that are difficult to identify without advanced technical knowledge or specialized tools. Moreover, the reliance on end-user awareness and training to identify and avoid phishing attacks remains a persistent issue across both sectors. With contradicting opinions regarding user training, the question stands whether it is a contribution for the better. Despite regular training and awareness programs, the human element continues to be one of the weakest links in the fight against phishing.

Another notable finding was the insufficiency of traditional email filtering and detection systems in identifying advanced phishing attacks. These systems, which are mainly based on keyword and blacklisting, seem to be struggling to keep up with the rapidly evolving tactics used by cybercriminals. In both private and organizational contexts, there is a pressing need for more advanced and adaptive phishing detection systems.

Despite a comprehensive review process, our research was not without limitations. A potential imbalance in the selection of articles, as well as possibly flawed selection criteria, may have influenced our findings. While we attempted to present an unbiased and complete view of the current challenges, the conclusions drawn may have been affected by these limitations.

That said, our findings underline the universal nature of challenges in phishing detection in emails. It emphasizes that these challenges are not limited to any specific user group or organizational type. This insight could shape future efforts to tackle phishing, implying a need for universal solutions that can effectively serve both private and organizational users.

9.1 Future Work

For future work, an interesting idea would be a broader perspective, and more articles, e.g., user studies aimed towards private users. Further, the lack of substantial differences between private and organizational users suggests that unified approaches may be effective in addressing the common challenges faced by both user groups. Future research should focus on the development of adaptive and robust detection methods, improving user awareness through targeted education and training programs as well as adopting more precise and balanced selection criteria to avoid potential biases or imbalances. By understanding and addressing these shared challenges, researchers and practitioners can enhance the overall security and resilience of email communication systems for all users.

References

Abdulraheem, R., Odeh, A., Al Fayoumi, M., & Keshta, I. (2022). Efficient Email phishing detection using Machine learning. *2022 IEEE 12th Annual Computing and Communication Workshop and Conference (CCWC)*, 0354–0358. https://doi.org/10.1109/CCWC54503.2022.9720818

Abroshan, H., Devos, J., Poels, G., & Laermans, E. (2021). Phishing Happens Beyond Technology: The Effects of Human Behaviors and Demographics on Each Step of a Phishing Process. *IEEE Access*, *9*, 44928–44949. https://doi.org/10.1109/ACCESS.2021.3066383

Adil, M., Khan, R., & Nawaz Ul Ghani, M. A. (2020). Preventive Techniques of Phishing Attacks in Networks. *2020 3rd International Conference on Advancements in Computational Sciences (ICACS)*, 1–8. https://doi.org/10.1109/ICACS47775.2020.9055943

Adzhar, A. A., Mabni, Z., & Ibrahim, Z. (2022). A Comparative Study on Email Phishing Detection Using Machine Learning Techniques. *2022 IEEE International Conference on Computing (ICOCO)*, 96–101. https://doi.org/10.1109/ICOCO56118.2022.10031671

Agari, F. (2023, January 30). Ransomware Attacks: Why Email Is Still THE Most Common Delivery Method. *PhishLabs*. https://www.agari.com/blog/ransomware-attacks-why-email-still-most-common-delivery-method

Ahammad, S. M. M., Raviteja, T., Koushik, J., Dinesh, P. V., & Ashok, A. (2022). Machine Learning Approach Based Phishing Email Text Analysis (ML-PE-TA). *2022 Third International Conference on Intelligent Computing Instrumentation and Control Technologies (ICICICT)*, 1087–1092. https://doi.org/10.1109/ICICICT54557.2022.9917765

Al-Ghamdi, N., & Alsubait, T. (2022). Digital Forensics and Machine Learning to Fraudulent Email Prediction. *2022 Fifth National Conference of Saudi Computers Colleges (NCCC)*, 99–106. https://doi.org/10.1109/NCCC57165.2022.10067685

Al-Haddad, R., Sahwan, F., Aboalmakarem, A., Latif, G., & Alufaisan, Y. M. (2021). Email text analysis for fraud detection through machine learning techniques. *3rd Smart Cities Symposium (SCS 2020)*, 613–616. https://doi.org/10.1049/icp.2021.0909

Al-Hamar, Y., Kolivand, H., Tajdini, M., Saba, T., & Ramachandran, V. (2021). Enterprise Credential Spear-phishing attack detection. *Computers & Electrical Engineering*, *94*, 107363. https://doi.org/10.1016/j.compeleceng.2021.107363

Alhogail, A., & Alsabih, A. (2021). Applying machine learning and natural language processing to detect phishing email. *Computers & Security*, *110*, 102414. https://doi.org/10.1016/j.cose.2021.102414

Alotaibi, R., Al-Turaiki, I., & Alakeel, F. (2020). Mitigating Email Phishing Attacks using Convolutional Neural Networks. *2020 3rd International Conference on Computer Applications & Information Security (ICCAIS)*, 1–6. https://doi.org/10.1109/ICCAIS48893.2020.9096821

Andryukhin, A. A. (2019). Phishing Attacks and Preventions in Blockchain Based Projects. *2019 International Conference on Engineering Technologies and Computer Science (EnT)*, 15–19. https://doi.org/10.1109/EnT.2019.00008

APWG. (2022). *Anti-Phishing Working Group—Phishing Activity Trends Report 3rd Quarter 2022*. https://docs.apwg.org/reports/apwg_trends_report_q3_2022.pdf

Assefa, A., & Katarya, R. (2022). Intelligent Phishing Website Detection Using Deep Learning. *2022 8th International Conference on Advanced Computing and Communication Systems (ICACCS)*, *1*, 1741–1745. https://doi.org/10.1109/ICACCS54159.2022.9785003

Baki, S., & Verma, R. (2022). Sixteen Years of Phishing User Studies: What Have We Learned? *IEEE Transactions on Dependable and Secure Computing*, 1–1.

https://doi.org/10.1109/TDSC.2022.3151103

Bansal, C., & Sidhu, B. (2021). Machine Learning based Hybrid Approach for Email Spam Detection. *2021 9th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO)*, 1–4. https://doi.org/10.1109/ICRITO51393.2021.9596149

Bathala, H. V., Srihitha, P. V. N. P., Dodla, S. G. R., & Pasala, A. (2021). Zero-Day attack prevention Email Filter using Advanced Machine Learning. *2021 5th Conference on Information and Communication Technology (CICT)*, 1–6. https://doi.org/10.1109/CICT53865.2020.9672420

Berndtsson, M., Hansson, J., Olsson, B., & Lundell, B. (2008). *Thesis Projects: A Guide for Students in Computer Science and Information Systems*. Springer. https://doi.org/10.1007/978-1-84800-009-4

Bikov, T. D., Iliev, T. B., Mihaylov, Gr. Y., & Stoyanov, I. S. (2019). Phishing in Depth – Modern Methods of Detection and Risk Mitigation. *2019 42nd International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*, 447–450. https://doi.org/10.23919/MIPRO.2019.8757074

Bountakas, P., & Xenakis, C. (2023). HELPHED: Hybrid Ensemble Learning PHishing Email Detection. *Journal of Network and Computer Applications*, *210*, 103545. https://doi.org/10.1016/j.jnca.2022.103545

Brereton, P., Kitchenham, B. A., Budgen, D., Turner, M., & Khalil, M. (2007). Lessons from applying the systematic literature review process within the software engineering domain. *Journal of Systems and Software*, *80*(4), 571–583. https://doi.org/10.1016/j.jss.2006.07.009

Catal, C., Giray, G., Tekinerdogan, B., Kumar, S., & Shukla, S. (2022). Applications of deep learning for phishing detection: A systematic literature review. *Knowledge and Information Systems*, *64*(6), 1457–1500. https://doi.org/10.1007/s10115-022-01672-x

Chiew, K. L., Yong, K. S. C., & Tan, C. L. (2018). A survey of phishing attacks: Their types, vectors and technical approaches. *Expert Systems with Applications*, *106*, 1–20. https://doi.org/10.1016/j.eswa.2018.03.050

Cota, R. P., & Zinca, D. (2022). Comparative Results of Spam Email Detection Using Machine Learning Algorithms. *2022 14th International Conference on Communications (COMM)*, 1–5. https://doi.org/10.1109/COMM54429.2022.9817305

Cronin, P., Ryan, F., & Coughlan, M. (2008). Undertaking a literature review: A step-by-step approach. *British Journal of Nursing (Mark Allen Publishing)*, *17*, 38–43. https://doi.org/10.12968/bjon.2008.17.1.28059

Das, A., Baki, S., El Aassal, A., Verma, R., & Dunbar, A. (2020). SoK: A Comprehensive Reexamination of Phishing Research From the Security Perspective. *IEEE Communications Surveys & Tutorials*, *22*(1), 671–708. https://doi.org/10.1109/COMST.2019.2957750

Das, S., Kim, A., Tingle, Z., & Nippert-Eng, C. (2019). *All About Phishing: Exploring User Research through a Systematic Literature Review* (arXiv:1908.05897). arXiv. https://doi.org/10.48550/arXiv.1908.05897

Dey, N., Samhitha, S., Hariprasad, M., Anand, A., & Gadad, V. (2021). Analysis of Machine Learning Algorithms by Developing a Phishing Email and Website Detection Model. *2021 IEEE International Conference on Computation System and Information Technology for Sustainable Solutions (CSITSS)*, 1–7. https://doi.org/10.1109/CSITSS54238.2021.9683131

Ding, X., Liu, B., Jiang, Z., Wang, Q., & Xin, L. (2021). Spear Phishing Emails Detection Based on Machine Learning. *2021 IEEE 24th International Conference on Computer Supported Cooperative Work in Design (CSCWD)*, 354–359. https://doi.org/10.1109/CSCWD49262.2021.9437758

Dou, Z., Khalil, I., Khreishah, A., Al-Fuqaha, A., & Guizani, M. (2017). Systematization of Knowledge

(SoK): A Systematic Review of Software-Based Web Phishing Detection. *IEEE Communications Surveys & Tutorials*, *19*(4), 2797–2819. https://doi.org/10.1109/COMST.2017.2752087

Drury, V., & Meyer, U. (2020). No Phishing With the Wrong Bait: Reducing the Phishing Risk by Address Separation. *2020 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW)*, 646–652. https://doi.org/10.1109/EuroSPW51379.2020.00093

ENISA. (2022). *EU Agency for Cybersecurity*. https://www.enisa.europa.eu/publications/enisa-threat-landscape-for-ransomware-attacks

Farahisya, N. A., & Bachtiar, F. A. (2022). Spam Email Detection with Affect Intensities using Recurrent Neural Network Algorithm. *2022 2nd International Conference on Information Technology and Education (ICIT&E)*, 206–211. https://doi.org/10.1109/ICITE54466.2022.9759865

Fayoumi, M. A., Odeh, A., Keshta, I., Aboshgifa, A., AlHajahjeh, T., & Abdulraheem, R. (2022). Email phishing detection based on naïve Bayes, Random Forests, and SVM classifications: A comparative study. *2022 IEEE 12th Annual Computing and Communication Workshop and Conference (CCWC)*, 0007–0011. https://doi.org/10.1109/CCWC54503.2022.9720757

Gallo, L., Maiello, A., Botta, A., & Ventre, G. (2021). 2 Years in the anti-phishing group of a large company. *Computers & Security*, *105*, 102259. https://doi.org/10.1016/j.cose.2021.102259

Gangavarapu, T., Jaidhar, C. D., & Chanduka, B. (2020). Applicability of machine learning in spam and phishing email filtering: Review and approaches. *Artificial Intelligence Review*, *53*(7), 5019–5081. https://doi.org/10.1007/s10462-020-09814-9

Giri, S., Banerjee, S., Bag, K., & Maiti, D. (2022). Comparative Study of Content-Based Phishing Email Detection Using Global Vector (GloVe) and Bidirectional Encoder Representation from Transformer (BERT) Word Embedding Models. *2022 First International Conference on Electrical, Electronics, Information and Communication Technologies (ICEEICT)*, 01–06. https://doi.org/10.1109/ICEEICT53079.2022.9768612

Gogoi, B., & Ahmed, T. (2022). Phishing and Fraudulent Email Detection through Transfer Learning using pretrained transformer models. *2022 IEEE 19th India Council International Conference (INDICON)*, 1–6. https://doi.org/10.1109/INDICON56171.2022.10040097

Görling, S. (2006). *THE MYTH OF USER EDUCATION*. Virus Bulletin Conference, Montreal, QC. http://www.gorling.se/files/texts/StefanGorlingVB2006.pdf

Griffin, S. E., & Rackley, C. C. (2008). Vishing. *Proceedings of the 5th Annual Conference on Information Security Curriculum Development*, 33–35. https://doi.org/10.1145/1456625.1456635

Gualberto, E. S., De Sousa, R. T., De B. Vieira, T. P., Da Costa, J. P. C. L., & Duque, C. G. (2020). From Feature Engineering and Topics Models to Enhanced Prediction Rates in Phishing Detection. *IEEE Access*, *8*, 76368–76385. https://doi.org/10.1109/ACCESS.2020.2989126

Gutierrez, L. F., Abri, F., Armstrong, M., Namin, A. S., & Jones, K. S. (2020). Email Embeddings for Phishing Detection. *2020 IEEE International Conference on Big Data (Big Data)*, 2087–2092. https://doi.org/10.1109/BigData50022.2020.9377821

Heartfield, R., & Loukas, G. (2015). A Taxonomy of Attacks and a Survey of Defence Mechanisms for Semantic Social Engineering Attacks. *ACM Computing Surveys*, *48*(3), 37:1-37:39. https://doi.org/10.1145/2835375

Higashino, M., Kawato, T., Ohmori, M., & Kawamura, T. (2019). An Anti-phishing Training System for Security Awareness and Education Considering Prevention of Information Leakage. *2019 5th International Conference on Information Management (ICIM)*, 82–86. https://doi.org/10.1109/INFOMAN.2019.8714691

Hosseinalipour, A., & Ghanbarzadeh, R. (2022). A novel approach for spam detection using horse herd optimization algorithm. *Neural Computing and Applications*, *34*(15), 13091–13105.

https://doi.org/10.1007/s00521-022-07148-x

Howe, A. E., Ray, I., Roberts, M., Urbanska, M., & Byrne, Z. (2012). The Psychology of Security for the Home Computer User. *2012 IEEE Symposium on Security and Privacy*, 209–223. https://doi.org/10.1109/SP.2012.23

Jalda, C. S., Kumar Nanda, A., & Pitchai, R. (2022). Spoofing E-Mail Detection Using Stacking Algorithm. *2022 8th International Conference on Smart Structures and Systems (ICSSS)*, 01–04. https://doi.org/10.1109/ICSSS54381.2022.9782173

Jáñez-Martino, F., Alaiz-Rodríguez, R., González-Castro, V., Fidalgo, E., & Alegre, E. (2023). A review of spam email detection: Analysis of spammer strategies and the dataset shift problem. *Artificial Intelligence Review*, *56*(2), 1145–1173. https://doi.org/10.1007/s10462-022-10195-4

Karim, A., Azam, S., Shanmugam, B., Kannoorpatti, K., & Alazab, M. (2019). A Comprehensive Survey for Intelligent Spam Email Detection. *IEEE Access*, *7*, 168261–168295. https://doi.org/10.1109/ACCESS.2019.2954791

Khonji, M., Iraqi, Y., & Jones, A. (2013). Phishing Detection: A Literature Survey. *IEEE Communications Surveys & Tutorials*, *15*(4), 2091–2121. https://doi.org/10.1109/SURV.2013.032213.00009

Kitchenham, B. (2004). Procedures for Performing Systematic Reviews. *Keele, UK, Keele University*, *33(2004)*, 1–26.

Kovač, A., Dunder, I., & Seljan, S. (2022). An overview of machine learning algorithms for detecting phishing attacks on electronic messaging services. *2022 45th Jubilee International Convention on Information, Communication and Electronic Technology (MIPRO)*, 954–961. https://doi.org/10.23919/MIPRO55190.2022.9803517

Kumar, N., Sonowal, S., & Nishant. (2020). Email Spam Detection Using Machine Learning Algorithms. *2020 Second International Conference on Inventive Research in Computing Applications (ICIRCA)*, 108–113. https://doi.org/10.1109/ICIRCA48905.2020.9183098

Lain, D., Kostiainen, K., & Čapkun, S. (2022). Phishing in Organizations: Findings from a Large-Scale and Long-Term Study. *2022 IEEE Symposium on Security and Privacy (SP)*, 842–859. https://doi.org/10.1109/SP46214.2022.9833766

Lee, J., Tang, F., Ye, P., Abbasi, F., Hay, P., & Divakaran, D. M. (2021). D-Fence: A Flexible, Efficient, and Comprehensive Phishing Email Detection System. *2021 IEEE European Symposium on Security and Privacy (EuroS&P)*, 578–597. https://doi.org/10.1109/EuroSP51992.2021.00045

Li, Q., Cheng, M., Wang, J., & Sun, B. (2022). LSTM Based Phishing Detection for Big Email Data. *IEEE Transactions on Big Data*, *8*(1), 278–288. https://doi.org/10.1109/TBDATA.2020.2978915

Li, X., Zhang, D., & Wu, B. (2020). Detection method of phishing email based on persuasion principle. *2020 IEEE 4th Information Technology, Networking, Electronic and Automation Control Conference (ITNEC)*, 571–574. https://doi.org/10.1109/ITNEC48623.2020.9084766

Madleňák, M., & Kampová, K. (2022). Phishing as a Cyber Security Threat. *2022 20th International Conference on Emerging ELearning Technologies and Applications (ICETA)*, 392–396. https://doi.org/10.1109/ICETA57911.2022.9974817

Magdy, S., Abouelseoud, Y., & Mikhail, M. (2022). Efficient spam and phishing emails filtering based on deep learning. *Computer Networks*, *206*, 108826. https://doi.org/10.1016/j.comnet.2022.108826

Mahdavifar, S., & Ghorbani, A. A. (2020). DeNNeS: Deep embedded neural network expert system for detecting cyber attacks. *Neural Computing and Applications*, *32*(18), 14753–14780. https://doi.org/10.1007/s00521-020-04830-w

Meline, T. (2006). Selecting Studies for Systemic Review: Inclusion and Exclusion Criteria. *Contemporary Issues in Communication Science and Disorders*, *33*(Spring), 21–27. https://doi.org/10.1044/cicsd_33_S_21

Mishra, S., & Soni, D. (2019). SMS Phishing and Mitigation Approaches. *2019 Twelfth International Conference on Contemporary Computing (IC3)*, 1–5. https://doi.org/10.1109/IC3.2019.8844920

Mukhopadhyay, A., & Prajwal, A. (2021). EDITH - A Robust Framework for Prevention of Cyber Attacks in the Covid Era. *2021 2nd International Conference for Emerging Technology (INCET)*, 1–8. https://doi.org/10.1109/INCET51464.2021.9456186

Muneer, A., Ali, R. F., Al-Sharai, A. A., & Fati, S. M. (2021). A Survey on Phishing Emails Detection Techniques. *2021 International Conference on Innovative Computing (ICIC)*, 1–6. https://doi.org/10.1109/ICIC53490.2021.9692960

Muralidharan, T., & Nissim, N. (2023). Improving malicious email detection through novel designated deep-learning architectures utilizing entire email. *Neural Networks*, *157*, 257–279. https://doi.org/10.1016/j.neunet.2022.09.002

Nabeel, M., Altinisik, E., Sun, H., Khalil, I., Wang, H. (Wendy), & Yu, T. (2021). CADUE: Content-Agnostic Detection of Unwanted Emails for Enterprise Security. *Proceedings of the 24th International Symposium on Research in Attacks, Intrusions and Defenses*, 205–219. https://doi.org/10.1145/3471621.3471862

Niroshan Atimorathanna, D., Shehan Ranaweera, T., Devdunie Pabasara, R. A. H., Rukshila Perera, J., & Yapa Abeywardena, K. (2020). NoFish; Total Anti-Phishing Protection System. *2020 2nd International Conference on Advancements in Computing (ICAC)*, 470–475. https://doi.org/10.1109/ICAC51239.2020.9357145

Nishikawa, H., Yamamoto, T., Harsham, B., Wang, Y., Uehara, K., Hori, C., Iwasaki, A., Kawauchi, K., & Nishigaki, M. (2020). Analysis of Malicious Email Detection using Cialdini's Principles. *2020 15th Asia Joint Conference on Information Security (AsiaJCIS)*, 137–142. https://doi.org/10.1109/AsiaJCIS50894.2020.00032

Ozker, U., & Sahingoz, O. K. (2020). Content Based Phishing Detection with Machine Learning. *2020 International Conference on Electrical Engineering (ICEE)*, 1–6. https://doi.org/10.1109/ICEE49691.2020.9249892

Paliath, S., Qbeitah, M. A., & Aldwairi, M. (2020). PhishOut: Effective Phishing Detection Using Selected Features. *2020 27th International Conference on Telecommunications (ICT)*, 1–5. https://doi.org/10.1109/ICT49546.2020.9239589

Patil, K., & Arra, S. R. (2022). Detection of Phishing and User Awareness Training in Information Security: A Systematic Literature Review. *2022 2nd International Conference on Innovative Practices in Technology and Management (ICIPTM)*, *2*, 780–786. https://doi.org/10.1109/ICIPTM54933.2022.9753912

Razak, M. F. A., Jaya, M. I., Ernawan, F., Firdaus, A., & Nugroho, F. A. (2022). Comparative Analysis of Machine Learning Classifiers for Phishing Detection. *2022 6th International Conference on Informatics and Computational Sciences (ICICoS)*, 84–88. https://doi.org/10.1109/ICICoS56336.2022.9930531

Rustam, F., Saher, N., Mehmood, A., Lee, E., Washington, S., & Ashraf, I. (2023). Detecting ham and spam emails using feature union and supervised machine learning models. *Multimedia Tools and Applications*. https://doi.org/10.1007/s11042-023-14814-2

Salahdine, F., El Mrabet, Z., & Kaabouch, N. (2021). Phishing Attacks Detection A Machine Learning-Based Approach. *2021 IEEE 12th Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON)*, 0250–0255. https://doi.org/10.1109/UEMCON53757.2021.9666627

Salau, A., Dantu, R., & Upadhyay, K. (2021). Data Cooperatives for Neighborhood Watch. *2021 IEEE*

International Conference on Blockchain and Cryptocurrency (ICBC), 1–9. https://doi.org/10.1109/ICBC51069.2021.9461056

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

Saraswat, P., & Singh Solanki, M. (2022). Phishing Detection in E-mails using Machine Learning. *2022 2nd International Conference on Technological Advancements in Computational Sciences (ICTACS)*, 420–424. https://doi.org/10.1109/ICTACS56270.2022.9987839

SCB. (2023). *Statistikdatabasen*. https://www.statistikdatabasen.scb.se/pxweb/sv/ssd/

SDG. (2015). *United Nations Sustainable Development Goals*. https://sdgs.un.org/goals

Shalini, L., Manvi, S. S., Gowda, N. C., & Manasa, K. N. (2022). Detection of Phishing Emails using Machine Learning and Deep Learning. *2022 7th International Conference on Communication and Electronics Systems (ICCES)*, 1237–1243. https://doi.org/10.1109/ICCES54183.2022.9835846

Sheikhalishahi, M., Saracino, A., Martinelli, F., La Marra, A., Mejri, M., & Tawbi, N. (2020). Digital Waste Disposal: An automated framework for analysis of spam emails. *International Journal of Information Security*, *19*(5), 499–522. https://doi.org/10.1007/s10207-019-00470-x

Shi, L., Chen, C., Wang, Q., & Boehm, B. (2021). Automatically detecting feature requests from development emails by leveraging semantic sequence mining. *Requirements Engineering*, *26*(2), 255–271. https://doi.org/10.1007/s00766-020-00344-y

Shmalko, M., Abuadbba, A., Gaire, R., Wu, T., Paik, H.-Y., & Nepal, S. (2022). Profiler: Distributed Model to Detect Phishing. *2022 IEEE 42nd International Conference on Distributed Computing Systems (ICDCS)*, 1336–1337. https://doi.org/10.1109/ICDCS54860.2022.00152

Sonowal, G. (2020). Phishing Email Detection Based on Binary Search Feature Selection. *SN Computer Science*, *1*(4), 191. https://doi.org/10.1007/s42979-020-00194-z

Spyns, P., & Vidal, M.-E. (2015). *Scientific Peer Reviewing*. Springer International Publishing. https://doi.org/10.1007/978-3-319-25084-7

Subasi, A., & Kremic, E. (2020). Comparison of Adaboost with MultiBoosting for Phishing Website Detection. *Procedia Computer Science*, *168*, 272–278. https://doi.org/10.1016/j.procs.2020.02.251

Sun, Y., Chong, N., & Ochiai, H. (2022). Federated Phish Bowl: LSTM-Based Decentralized Phishing Email Detection. *2022 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, 20–25. https://doi.org/10.1109/SMC53654.2022.9945584

Tekeni, L., & Botha, R. A. (2022). A Systematic Literature Review into Security Education, Training and Awareness Aimed at Home Users. *2022 International Conference on Electrical, Computer and Energy Technologies (ICECET)*, 1–6. https://doi.org/10.1109/ICECET55527.2022.9873517

Venkatraman, S., Surendiran, B., & Arun Raj Kumar, P. (2020). Spam e-mail classification for the Internet of Things environment using semantic similarity approach. *The Journal of Supercomputing*, *76*(2), 756–776. https://doi.org/10.1007/s11227-019-02913-7

Wang, Y. (2022). A survey of phishing detection: From an intelligent countermeasures view. *2022 IEEE Conference on Telecommunications, Optics and Computer Science (TOCS)*, 761–769. https://doi.org/10.1109/TOCS56154.2022.10016193

Waziri, I. (2015). Website Forgery: Understanding Phishing Attacks and Nontechnical Countermeasures. *2015 IEEE 2nd International Conference on Cyber Security and Cloud Computing*, 445–450. https://doi.org/10.1109/CSCloud.2015.77

Wohlin, C. (2014). Guidelines for snowballing in systematic literature studies and a replication in software engineering. *Proceedings of the 18th International Conference on Evaluation and Assessment in Software Engineering*, 1–10. https://doi.org/10.1145/2601248.2601268

Wohlin, C., Runeson, P., Höst, M., Ohlsson, M. C., Regnell, B., & Wesslén, A. (2012). *Experimentation in Software Engineering*. Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-29044-2

Wolfswinkel, J., Furtmueller, E., & Wilderom, C. (2013). Using Grounded Theory as a Method for Rigorously Reviewing Literature. *European Journal of Information Systems*, *22*. https://doi.org/10.1057/ejis.2011.51

Wu, P., & Guo, H. (2022). Holmes: An Efficient and Lightweight Semantic Based Anomalous Email Detector. *2022 IEEE International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom)*, 1360–1367. https://doi.org/10.1109/TrustCom56396.2022.00192

Xiao, Y., & Watson, M. (2019). Guidance on Conducting a Systematic Literature Review. *Journal of Planning Education and Research*, *39*(1), 93–112. https://doi.org/10.1177/0739456X17723971

Zhao, J., Masood, R., & Seneviratne, S. (2021). A Review of Computer Vision Methods in Network Security. *IEEE Communications Surveys & Tutorials*, *23*(3), 1838–1878. https://doi.org/10.1109/COMST.2021.3086475

Zhu, E., Yuan, Q., Chen, Z., Li, X., & Fang, X. (2022). CCBLA: A Lightweight Phishing Detection Model Based on CNN, BiLSTM, and Attention Mechanism. *Cognitive Computation*. https://doi.org/10.1007/s12559-022-10024-4

Zhu, H. (2020). Online meta-learning firewall to prevent phishing attacks. *Neural Computing and Applications*, *32*(23), 17137–17147. https://doi.org/10.1007/s00521-020-05041-z

Appendix A – Accepted Articles (author's own)

Label	Relevance	Title	Authors & Publication Year
A1	Org.	Efficient Email phishing detection using Machine learning	Abdulraheem, R., Odeh, A., Al Fayoumi, M., & Keshta, I. (2022)
A2	Both	Preventive Techniques of Phishing Attacks in Networks	Adil, M., Khan, R., & Nawaz, M. (2020)
A3	Org.	A Comparative Study on Email Phishing Detection Using Machine Learning Techniques	Adzhar, A., Mabni, Z., & Ibrahim, Z., (2022)
A4	Both	Machine Learning Approach Based Phishing Email Text Analysis (ML-PE-TA)	Ahammad, S., Raviteja, T., Koushik, J., Dinesh, P., & Ashok, A. (2022)
A5	Org.	Digital Forensics and Machine Learning to Fraudulent Email Prediction	Al-Ghamdi, N. & Alsubait, T. (2022)
A6	Org.	Email Text Analysis for Fraud Detection through Machine Learning Techniques	Al-Haddad, R., Sahwan, F., Aboalmakarem, A., Latif, G., & Alufaisan, Y.M. (2021)
A7	Org.	Enterprise Credential Spear-phishing attack detection	Al-Hamar, Y., Kolivand, H., Tajdini, M., Saba, T., & Ramachandran, V. (2021)
A8	Org.	Applying machine learning and natural language processing to detect phishing email	Alhogail, A. & Alsabih, A. (2021)

A9	Org.	Mitigating Email Phishing Attacks using Convolutional Neural Networks	Alotaibi, R., Al-Turaiki, I., & Alakeel, F. (2020)
A10	Both	Machine Learning based Hybrid Approach for Email Spam Detection	Bansal, C. & Sidhu, B. (2021)
A11	Both	Zero-Day attack prevention Email Filter using Advanced Machine Learning	Bathala, H., Srihitha, P.V.N., Dodla, S., & Pasala, A. (2021)
A12	Org.	HELPHED: Hybrid Ensemble Learning PHishing Email Detection	Bountakas, P. & Xenakis, C. (2023)
A13	Org.	Applications of deep learning for phishing detection: a systematic literature review	Catal, C., Giray, G., Tekinerdogan, B., Kumar, S., & Shukla, S. (2022)
A14	Org.	Comparative Results of Spam Email Detection Using Machine Learning Algorithms	Cota, R. & Zinca, D. (2022)
A15	Priv.	SoK: A Comprehensive Reexamination of Phishing Research From the Security Perspective	Das, A., Baki, S., El Aassal, A., Verma, R., & Dunbar, A. (2020)
A16	Org.	Analysis of Machine Learning Algorithms by Developing a Phishing Email and Website Detection Model	Dey, N., Samhitha, S., Hariprasad, M., Anand, A., & Gadad, V. (2021)
A17	Both	Spear Phishing Emails Detection Based on Machine Learning	Ding, X., Liu, B., Jiang, Z., Wang, Q., & Xin, L (2021)
A18	Priv.	No Phishing With the Wrong Bait: Reducing the Phishing Risk by Address Separation	Drury, V. & Meyer, U. (2020)

A19	Both	Spam Email Detection with Affect Intensities using Recurrent Neural Network Algorithm	Farahisya, N. & Bachtiar, F. (2022)
A20	Org.	Email phishing detection based on naïve Bayes, Random Forests, and SVM classifications: A comparative study	Fayoumi, M., Odeh, A., Keshta, I., Aboshgifa, A., AlHajahjeh, T., & Abdulraheem, R. (2022)
A21	Org.	2 Years in the anti-phishing group of a large company	Gallo, L., Maiello, A., Botta, A., & Ventre, G. (2021)
A22	Org.	Applicability of machine learning in spam and phishing email filtering: review and approaches	Gangavarapu, T., Jaidhar, C., & Chanduka, B. (2020)
A23	Org.	Comparative Study of Content-Based Phishing Email Detection Using Global Vector (GloVe) and Bidirectional Encoder Representation from Transformer (BERT) Word Embedding Models	Giri, S., Banerjee, S., Bag, K., & Maiti, D. (2022)
A24	Both	Phishing and Fraudulent Email Detection through Transfer Learning using pretrained transformer models	Gogoi, B. & Ahmed, T. (2022)
A25	Org.	From Feature Engineering and Topics Models to Enhanced Prediction Rates in Phishing Detection	Gualberto, E., De Sousa, R., De B. Vieira, T., Da costa, J., & Duque, C. (2020)
A26	Org.	Email Embeddings for Phishing Detection	Gutierrez, L., Abri, F., Armstrong, M., Namin, A., & Jones, K. (2020)

A27	Org.	A novel approach for spam detection using horse herd optimization algorithm	Hosseinalipour, A. & Ghanbarzadeh, R. (2022)
A28	Both	Spoofing E-Mail Detection Using Stacking Algorithm	Jalda, C., Kumar, A., & Pitchai, R. (2022)
A29	Org.	Classifying spam emails using agglomerative hierarchical clustering and a topic-based approach	Jáñez-Martino, F., Alaiz-Rodríguez, R., Fidalgo, E., & Alegre, E. (2023)
A30	Org.	An overview of machine learning algorithms for detecting phishing attacks on electronic messaging services	Kovač, A., Dunder, I., & Seljan, S. (2022)
A31	Org.	Email Spam Detection Using Machine Learning Algorithms	Kumar, N., Sonowal, S., & Nishant (2020)
A32	Org.	Phishing in Organizations: Findings from a Large-Scale and Long-Term Study	Lain, D., Kostiainen, K., & Čapkun S. (2022)
A33	Org.	D-Fence: A Flexible, Efficient, and Comprehensive Phishing Email Detection System	Lee, J., Tang, F., Ye, P., Abbasi, F., Hay, P., & Divakaran, D. (2021)
A34	Org.	LSTM Based Phishing Detection for Big Email Data	Li, Q., Cheng, M., Wang, J., & Sun, B. (2022)
A35	Org.	Detection method of phishing email based on persuasion principle	Li, X., Zhang, D., & Wu, B. (2020)
A36	Org.	Efficient spam and phishing emails filtering based on deep learning	Magdy, S., Abouelseoud, Y., & Mikhail, M. (2022)

A37	Org.	DeNNeS: deep embedded neural network expert system for detecting cyber attacks	Mahdavifar, S. & Ghorbani, A. (2020)
A38	Org.	EDITH - A Robust Framework for Prevention of Cyber Attacks in the Covid Era	Mukhopadhyay, A. & Prajwal, A. (2021)
A39	Org.	A Survey on Phishing Emails Detection Techniques	Muneer, A., Ali, R., Al-Sharai, A., & Fati, S. (2021)
A40	Both	Improving malicious email detection through novel designated deep-learning architectures utilizing entire email	Muralidharan, T. & Nissim, N. (2023)
A41	Org.	CADUE: Content-Agnostic Detection of Unwanted Emails for Enterprise Security	Nabeel, M., Altinisik, E., Sun, H., Khalil, I., Wang, H., & Yu, T. (2021)
A42	Priv.	NoFish; Total Anti-Phishing Protection System	Niroshan, D., Shehan, T., Devdunie, R.A.H., Rukshila, J., & Yapa, K. (2020)
A43	Org.	Analysis of Malicious Email Detection using Cialdini's Principles	Nishikawa, H., Yamamoto, T., Harsham, B., Wang, Y., Uehara, K., Hori, C., Iwasaki, A., Kawauichi, K., & Nishigaki, M. (2020)
A44	Org.	Content Based Phishing Detection with Machine Learning	Ozker, U. & Sahingoz, O. (2020)
A45	Org.	PhishOut: Effective Phishing Detection Using Selected Features	Paliath, S., Qbeitah, M., & Aldwairi, M. (2020)

A46	Org.	Detection of Phishing and User Awareness Training in Information Security: A Systematic Literature Review	Patil, K. & Arra, S. (2022)
A47	Org.	Comparative Analysis of Machine Learning Classifiers for Phishing Detection	Razak, M., Jaya, M., Ernawan, F., Firdaus, A., & Nugroho, F. (2022)
A48	Org.	Detecting ham and spam emails using feature union and supervised machine learning models	Rustam, F., Saher, N., Mehmood, A., Lee, E., Washington, S. & Ashraf, I. (2023)
A49	Org.	Phishing Attacks Detection A Machine Learning-Based Approach	Salahdine, F., El Mrabet, Z., & Kaabouch, N. (2021)
A50	Org.	Data Cooperatives for Neighborhood Watch	Salau, A., Dantu, R., & Upadhyay, K. (2021)
A51	Org.	A Systematic Literature Review on Phishing Email Detection Using Natural Language Processing Techniques	Salloum, S., Gaber, T., Vadera, S., & Shaalan, K. (2022)
A52	Both	Phishing Detection in E-mails using Machine Learning	Saraswat, P. & Singh Solanki, M. (2022)
A53	Org.	Detection of Phishing Emails using Machine Learning and Deep Learning	Shalini, L., Manvi, S., Gowda, N., & Manasa, K. (2022)
A54	Org.	Digital Waste Disposal: an automated framework for analysis of spam emails	Sheikhalishahi, M., Saracino, A., Martinelli, F., La Marra, A., Mejri, M., & Tawbi, N. (2020)

A55	Org.	Automatically detecting feature requests from development emails by leveraging semantic sequence mining	Shi, L., Chen, C., Wang, Q., & Boehm, B. (2021)
A56	Org.	Profiler: Distributed Model to Detect Phishing	Shmalko, M., Abuadbba, A., Gaire, R., Wu, T., Paik, H., & Nepal, S. (2022)
A57	Org.	Phishing Email Detection Based on Binary Search Feature Selection	Sonowal, G. (2020)
A58	Org.	Federated Phish Bowl: LSTM-Based Decentralized Phishing Email Detection	Sun, Y., Chong, N., & Ochiai, H. (2022)
A59	Both	Spam e-mail classification for the Internet of Things environment using semantic similarity approach	Venkatraman, S., Surendiran, B., & Arun, P. (2020)
A60	Org.	A survey of phishing detection: from an intelligent countermeasures view	Wang, Y. (2022)
A61	Both	Holmes: An Efficient and Lightweight Semantic Based Anomalous Email Detector	Wu, P. & Guo, H. (2022)
A62	Org.	A review of spam email detection: analysis of spammer strategies and the dataset shift problem	Zhao, J., Masood, R., & Seneviratne, S. (2021)
A63	Both	CCBLA: a Lightweight Phishing Detection Model Based on CNN, BiLSTM, and Attention Mechanism	Zhu, E., Yuan, Q., Chen, Z., Li, X., & Fang, X. (2022)
A64	Org	Online meta-learning firewall to prevent phishing attacks	Zhu, H. (2020)