Learning from phishing emails

Creating new metrics to measure the effect of anti-phishing training in a large company

Anne-Kee Doing 4453972

This page was intentionally left blank

Learning from phishing emails

Creating new metrics to measure the effect of anti-phishing training in a large company

by

Anne-Kee Doing 4453972

to obtain the degree of Master of Science
at the Delft University of Technology,
to be defended publicly on Monday August 28, 2023 at 15:00 AM.

Project duration	February 6, 2023 – August 28, 2023		
Thesis committee	Chair	Prof.dr.ir. P. van Gelder	Safety and Security Science
	1^{st} supervisor	Dr. S. Parkin	Organisation and Management
	2^{nd} supervisor	Dr. Y. Zhauniarovich	Organisation and Management
	External supervisor	Dr. E. Bárbaro	ING Security Analytics
	External supervisor	F. van der Roest	ING IT security consultant

Cover image from Shutterstock.com (Vector, n.d.)

TU Delft

Acknowledgements

Dear reader,

In front of you is my thesis report, the result of six months of dedicated work and research. I owe a debt of gratitude to several individuals who played a significant role in shaping this report. I want to thank my thesis committee for their valuable insights and guidance throughout the process. Our discussions about various possibilities and options provided the flexibility and inspiration that led to the final project.

I would like to thank my parents for their unwavering support and enthusiasm for listening to a subject quite distant from their own interests. To my fellow students, I'm grateful for sharing the journey of graduation, and to my friends and family, I appreciate the timely distractions that kept me balanced during demanding moments. Your roles have been essential, and I'm truly thankful for your presence.

Happy reading!

Anne-Kee Doing,
Delft University of Technology,
August 14, 2023

Summary

Phishing attacks are a growing cause of cybersecurity incidents such as data breaches. With these attacks, malicious actors try to gain access to systems by exploiting the vulnerability of employees. Particularly, intruders use different tricks to create convincing phishing emails to force people to behave less securely. Unfortunately, technological applications are insufficient in detecting all possible phishing attempts. As the consequences of a successful breach can be disastrous for a company, especially in the banking sector, pressure is put on employees to recognise and report potential phishing emails. To aid them in this task, phishing training is provided by employers. Employees are taught what phishing emails look like and which actions they should take if an email appears malicious. However, the effectiveness of these training sessions is difficult to evaluate.

To understand the interaction of employees with phishing emails without influencing their behaviour, we study the emails employees report as suspicious to characterise the security culture at a bank. A better understanding of the behaviour provides grounds for recommendations to improve anti-phishing training and create a safer environment. The newfound metrics can provide an alternative to current methods.

This research uses Exploratory Data Analysis (EDA) to evaluate the email reporting behaviour of employees at a bank to answer the Research Question *How can email reporting patterns in a large organisation measure the relationship between phishing training, reported emails and employee behaviour?* With this case study, we apply EDA to a large dataset containing bank employees' reported emails over 16 months. We analysed the reported emails and related them to the provided phishing training events. Moreover, we did a text analysis of the emails' content using the Term Frequency - Inverse Document Frequency (TF-IDF) method. Additionally, we extract the dominant topics of the emails using topic modelling. Lastly, with the help of interviews, the results are tied to employees' experiences to understand their behaviour.

The major findings of the research are, firstly, the new metrics we identified to measure the security culture of a company. These metrics were found from both the analysis of employee behaviour over time, as well as the analysis of the email content. From the analysis of the reporting behaviour over time, new metrics include the unique reporters in relation to the total reported emails over time. Besides the unique reporters, unique reported emails can uncover the presence of campaigns. For example, a single email can explain the increase from 50 to 350 daily benign reports. The difference between the total reports and unique reports uncovered this. Secondly, topic analysis and content comparison show similarities between benign and malicious reported emails, indicating an increased vigilance of employees on these attributes.

A second finding originates from the analysis of the email components. One of the components was used in all simulation emails, while it was not present in all the benign and malicious reported emails. This shows that the simulation emails can be extended to include different scenarios. Therefore, we recommend the company to extend the phishing simulation emails to contain varied phishing tactics to expose employees to other types of attacks and incorporate all aspects taught during the E-learning.

Lastly, the analysis shows no concrete relation between the number of reported emails and the timing of the simulation wave. Although the reported p-value of benign emails after the simulation is 0.03, this significance can also be explained by external factors.

With these results, we can measure employees' security culture and awareness in real-world circumstances without influencing the employees' behaviour, providing a new approach to investigating phishing behaviour. Adding to the research of Steves et al. (2020), the click rates can be explained by more than solely the employees' awareness levels, and new explanations come forward to handle phishing threats. Moreover, the absence of a required test environment for the analysis created a solution for existing gaps. For example, as seen in (Hillman et al., 2023).

A limitation of exploratory data analysis is that results are often ambiguous and mainly provide possible directions for future research. Furthermore, external factors influencing the behaviour could provide alternative reasons for the discussed interactions.

To conclude, by using the reported emails to measure security behaviour related to phishing, we found new metrics which do not influence the employees in their daily behaviour while still providing insights to improve the tactics of a company in combating phishing attacks. Reporting behaviour can be used to analyse the current anti-phishing tactics of a company and provide suggestions for improvements.

Future research should explore the differences in applying the method in other companies and across sectors. Overlap and differences can create an understanding of the diversity in security culture and the effect of external factors. Combining these results with a comprehensive understanding of a company's operations can expose directions for improvement in the security approach. Additionally, the effect of the recommendations can be analysed using the metrics we proposed. This can be done with a follow-up analysis of the behaviour to see whether the desired effect can be observed.

Contents

List of Figures

List of Tables

Abbreviations

AI Artificial Intelligence

BERT Bidirectional Encoder Representations from Transformers

EDA Exploratory Data Analysis

ISA Information Security Awareness

KPI Key Performance Indicators

LDA Latent Dirichlet Allocation

ML Machine Learning

NLP Natural Language Processing

OSINT Open Source Intelligence

RoI Return on Investment

SAT Security Awareness Training

TF-IDF Term Frequency - Inverse Document Frequency

XAI eXplainable Artificial Intelligence

1 Introduction

1.1. Background

Cybersecurity is a developing domain where innovations and best practices follow each other in rapid succession (Tulane University, n.d.). In this domain, it is important to consider that cybersecurity differs from conventional information security in three areas. Firstly, there is more focus on the external effects of breaches on the physical world; secondly, human and organisational factors play a larger role; and thirdly, the emphasis lies on managing residual risks rather than excluding risks (P. van Gelder, personal communication, September 9, 2022). This calls for a different approach to the vulnerabilities surrounding cybersecurity. Furthermore, employees are often seen as one of the biggest liabilities when looking at cybersecurity, and criminals increasingly use phishing and social engineering to gain access to systems. (Khusainov et al., 2022; Ortiz et al., 2016; Xu et al., 2023).

Phishing emails are responsible for approximately 90% of all data breaches (CISCO, 2022). The phishing emails imitate normal email traffic but are damaging and trick people into clicking on malicious links or giving away their login credentials. This allows attackers to bypass security measures designed to prevent cyber attacks (Tyler & Davey, 2020). A commonly used tactic by attackers is to send phishing emails in bulk, targeting large audiences (IBM, 2023). With this method, even a small success rate can be profitable. Another tactic is to personalise phishing emails by gathering information about a victim to write convincing emails and stating facts about the intended target (Violino, 2023). When a phishing email is specifically designed for an individual like this, it is called spear phishing, and a shift can be seen towards these kinds of attacks (Arshad et al., 2021). Furthermore, once an attacker successfully deceives a victim, it takes, on average, 295 days to identify and contain the ensuing breach. (Security, 2022). In conclusion, the rising effectiveness and growing frequency of phishing emails, coupled with the significant time required to identify successful breaches, underscore the challenges in overcoming this cyber threat.

During the Covid-19 pandemic, the time people spend online has increased, which malicious actors exploit. (Bhardwaj et al., 2020; Schulze, 2020). Even now, while most measures have been lifted, working from home and online meetings have become the new standard. Security measures are less effective in these situations as cybersecurity measures are generally implemented better in the workplace, increasing the likelihood of successful phishing attacks (Bispham et al., 2022; Din & Soare, 2023). As a counterattack, email spam filters and phishing detection methods are improving (Kaddoura et al., 2022), but they cannot catch all malicious emails before arriving in an employee's inbox. Therefore, the pressure on employees to understand cybersecurity and behave securely is increasing.

Staff behavioural aspects can in- or decrease companies' cybersecurity drastically (Kam et al., 2020). One of the approaches to increase security awareness among staff is to provide security training for employees (Al-Alawi & Al-Bassam, 2019). There are multiple ways to provide this training, such as sending fake phishing emails or a lecture setting where an instructor informs staff how to recognise phishing emails. Several studies suggest such training can help increase companies' security (Feeley et al., 2022; Kanwal et al., 2022), but they are still recent and limited. Generally, training effectiveness is measured with clicking rates, the percentage of people who click on a link in the simulation email, where low rates are seen as more secure. However, as stated by K. Greene et al. (2018), a clicking rate of zero is impossible. Considering this and the lack of consensus in the literature for the best anti-phishing training (Jampen et al., 2020; Lain et al., 2022), the correct way to deal with phishing vulnerability is difficult to determine.

For several banks, phishing is found to be the most common way for attackers to gain access to their systems or networks (De Nederlandsche Bank N.V., 2022; ING group N.V., 2022). Vital sectors such as the banking sector are highly regulated, as a security breach in these sectors could have disastrous consequences for the economy and safety of citizens (Kuchumov et al., 2021). To set a base level of security between banks, governments have introduced regulations all banks must adhere to (European Banking Authority, 2018), and compliance is thoroughly checked (Dutch Government, 2023). To keep additional measures feasible on top of the regulations, a trade-off has to be made between security, time allocation and costs to provide adequate security measures. However, making this trade-off is ambiguous. To make it easier for the banks, they cooperate. For example, in the Netherlands, information is shared between the different banks to increase the effect of lessons learned and improve decisions between trade-offs and which tools to invest in (T. de Haan, personal communication, October 10, 2022).

Due to the economic outlook on these cybersecurity investments, adequate investment in the most effective resources is essential. The effectiveness of different approaches in preventing data breaches has been studied extensively. For instance, Chakraborty and Nisha (2022) looks into the principal vulnerabilities of companies due to cyber-attacks. Secure systems are crucial for companies to protect against cyber attacks and safeguard sensitive information. Adequate investment in resources can reduce vulnerability and mitigate potential security threats. Frameworks are developed to accomplish this, such as the principles of security by design, which can help design systems more securely to anticipate risk and prevent breaches from happening (Sveikauskas, 2021).

As described, companies already take extensive action to prevent and mitigate the consequences of successful phishing attacks. Despite this, ’easy’ phishing emails still effectively deceive employees (Griffiths, 2023). Additionally, the most sophisticated spear phishing emails are unlikely to be better detected by a short training session. It cannot be expected that the average employee has the knowledge and capability to recognise these emails if experts struggle with them. Some cues can be taught, but no amount of training will bring the clicks to zero (Bhardwaj et al., 2020), and the cost-benefit balance must also be considered here. Therefore, additional research is necessary to understand the best approach.

1.2. Recent Developments

1.2.1. Behavioural developments

Until now, research and development focus often lay on technical applications, while cybersecurity is a socio-technical system. As a result, behavioural aspects are understudied. Nonetheless, an upcoming, relatively new area within cybersecurity is the training of employees on cybersecurity-related issues with Security Awareness Training (SAT). Anti-phishing training teaches participants to recognise phishing emails and prevent data breaches. For example, the SLAM method, which stands for Sender, Link, Attachment, and Message, can teach people to pay extra attention to these components of an email (Kelvas, 2023). Additional cues, such as time pressure and unexpected attachments, can also be present in emails. Several studies have examined the effectiveness of training staff on cybersecurity-related flaws, such as Chowdhury et al. (2022) and Osterman Research (2019). Osterman Research (2019) focuses on the increase in safety within cybersecurity, comparing the absence of training and introducing the training. The training content is also researched, whereby Chowdhury et al. (2022) has studied how people learn and how training can be personalised.

Financial institutions organise various security training and awareness activities (ING, 2021). This is done to increase understanding among staff about potential weaknesses, how to mitigate the risk of their exploitation, and how to behave if an incident occurs. In addition, training is required to meet regulatory and compliance needs set by governmental institutions.

Cybercriminals personalise their attacks to increase the plausibility of their emails (Bhardwaj et al., 2020). This includes extracting job descriptions and personal information from social media such as LinkedIn. Steves et al. (2020) already shows that premise alignment with an employee's job can make a phishing email more difficult to identify correctly. This calls for personalising the training and including these types of attacks (Marin et al., 2023). The downsides of personalised training are the increased costs and ambiguous signals for employees, as different cues have to be taught in training for employees with other job descriptions. Further understanding employee behaviour and the critical factors and components in emails to identify suspicious emails will increase a company's security. To achieve this, it is necessary to gain a deeper understanding of employee behaviour in relation to phishing.

1.2.2. Technological developments

The increased quality and quantity of phishing emails make it difficult for technological tools to protect organisations against these attacks. With the recent rise of Natural Language Processing (NLP) applications such as Chat GPT, it is becoming easier for attackers to write convincing phishing emails. Besides the developments in training, the content of phishing emails and the approach to analysing the content is also developing. A driving factor for this development is the progress of NLP, a branch of Artificial Intelligence (AI). With NLP, machines can understand text data and generate texts based on human queries. Additionally, NLP can analyse text documents and extract different aspects, such as the main topics, and subjective qualities, such as the emotion of a text. With the aid of NLP, text documents can be generated by malicious actors, increasing the quality of the phishing emails and thus making them harder to detect. On the other side, defenders can use the same techniques to analyse incoming emails and improve their phishing filters with the gained information, stopping more attacks.

An additional method to improve the phishing filters and map the threats at the company is the possibility for employees to report suspicious emails. Employees can report suspicious emails for evaluation to determine if they are benign or potentially malicious. With this information, a company can learn about new types of attacks and deal with these threats accordingly.

1.3. Research gaps

Based on the existing literature, several research gaps can be identified in the studies that research the effectiveness of phishing training and the most suitable measure.

1.3.1. Employee behaviour

First, how employees engage with training courses and how seriously they take the assignments remains under-researched. The literature argues the need for more research to increase the understanding of cybersecurity vulnerabilities caused by human behaviour within organisations (Morgan et al., 2020). This also includes thoroughly understanding the threats employees face and providing them with the right tools to deal with and respond to them. Without this knowledge, the training is insufficient, and employees cannot rely on the training to help them correctly identify suspicious emails.

1.3.2. Real-world environment

When researching phishing susceptibility, a case study is a frequently used approach, either by sending counterfeit phishing emails to unknowing employees or letting participants complete a survey (Georgiadou et al., 2022; Goel et al., 2017; Reeves et al., 2020). By contrast, real phishing emails are rarely used to study employee behaviour. By researching interaction with actual emails in real-world situations, the organisational aspects of susceptibility can be investigated instead of solely the individual's susceptibility (Hillman et al., 2023). However, understanding the best way to measure in real-world circumstances is difficult because multiple factors influence behaviour. Compared to an experimental setup, where isolating the factors you want to study is possible, a non-controlled setting is more challenging. Namely, external factors cannot be excluded in a non-controlled setting where participants are observed and studied within their everyday environments. Additionally, finding the proper measures that do not influence the behaviour is challenging.

1.3.3. Reported emails as indicator

A commonly used tool to measure employees' security behaviour regarding phishing is a phishing simulation, where companies send counterfeit phishing emails to their employees to measure how they interact with them, for example, by Hillman et al. (2023) and Lain et al. (2022). These phishing emails imitate real malicious emails to evoke the realistic behaviour of employees to measure employees' responses to potential phishing emails. Correspondingly, some research focuses on the reporting rates of emails by creating a test environment, for example, Lain et al. (2022) and Parsons et al. (2014). Other approaches include sending simulation emails whereby the participant is in a real-world environment (Vishwanath et al., 2011). Or, the participant has to answer a questionnaire about their behaviour in real-world circumstances (Al-Shanfari et al., 2021).

Until now, reporting rates in phishing-related research provided answers to the reporting rates of simulation emails. However, the similarity or difference in employee behaviour between these simulation emails and regular email traffic lacks comprehensive understanding. Looking at actual behaviour concerning real phishing emails and an accurate comparison with malicious emails is missing. An explanation is that the absence of enough data is a limiting factor to perform this kind of research. It is often left as future research, for example, by Hillman et al. (2023). Additionally, Marin et al. (2023) look into the theories and factors that influence the reporting behaviour of employees. The factors are researched through a questionnaire, but again, real-world analysis is missing. Therefore, the relevancy of the studies and the reporting rates depend on the similarity between simulation and actual reporting rates. As there is limited information about unreported malicious emails, the similarity of reported emails with simulation emails can indicate the accuracy of the reporting rates.

As research conflicts about which factors influence behaviour (Zhuo et al., 2022), it is challenging to create a test environment which reflects the required and essential circumstances. Furthermore, observing the past behaviour of employees in their natural environment can only be done by looking into already recorded measures that don't interfere with the behaviour because, ideally, there is minimum user interaction (Foroughi & Luksch, 2018). Fortunately, in the context of phishing, there lies an opportunity in the reported emails. Employees can often report suspicious emails for examination and have them evaluated by an external party. Because these reports are often kept, they provide a dataset that can show why employees behave a certain way. The reported emails relate to the True and False positives in Table 1.1. Please note that false negatives may include emails employees deleted without interacting with them, so they cannot all be considered successful phishing attempts.

Table 1.1: Confusion matrix of the email type and employee decision when interacting with an email.

		Employee Decision	
		Engage	Report
Email type	Benign	True Negative	False Positive
	Malicious	False Negative	True Positive

Current literature has not been able to compare reported phishing statistics with the behaviour surrounding phishing training. Measuring reporting rates of phishing emails and comparing their content to the simulation wave can provide insight into the security culture. Thus, the reporting behaviour of employees provides an opportunity for observing employees' normal behaviour with real phishing emails. Furthermore, the components discussed in the training can be linked to real phishing emails. Moreover, language-based models are evolving quickly and can provide insight into large datasets. Sifting through thousands of emails by hand is time-consuming, but these models can overcome this challenge.

Exploring the possibilities of large-scale data analysis on the reported emails and the limitations in this approach can indicate the direction the sector can go. Without exploratory studies, it is guessing which directions could lead to improved security.

1.3.4. Interventions

It is unknown what the similarities and differences are between malicious and benign reported emails. If they are very similar, this can be a sign that people within the company should be more careful with the emails they send. If the opposite is true, and malicious and benign reported emails are very different, something else makes employees report benign emails. The difficulty with finding additional interventions that target how emails are written is that attackers can incorporate them into their emails without developing difficult technological devices or algorithms. With the assistance of applications like ChatGPT, writing more convincing phishing emails without spelling mistakes and the right tone will only become easier. Interventions can aim to improve the security culture and utilise methods only employees can use. Attackers do not have easy access to several things company employees have. Among others, these are legitimate company email addresses to write an email from. Because employees are more suspicious of unfamiliar senders (Williams et al., 2018), employees who only interact with internal contacts will be more suspicious of external senders. This makes departments with external interaction a relatively easier target, as they communicate more with external email addresses, which are easier to fake by attackers. These employees have to base their decisions more on subjective feelings than the traditional cues as they do not always apply to them (Williams et al., 2018).

There is, however, a gap in the literature addressing real-world measurements and a suitable measure not to influence the expected behaviour of employees. These gaps provide an opportunity to devise new methods leading to recommendations the company can implement in combating phishing. The gap in knowledge until now motivates the research questions at the end of this chapter. Within this work, we aim to close the stated research gaps. In particular, our goal is to address the following research question:

1.4. Research Questions

An essential and ongoing challenge in phishing research is to devise metrics that accurately measure the security culture and the effect of training without influencing the system itself. Existing research primarily focuses on metrics related to simulation emails or questionnaires, which do not always reliably reflect the real-world behaviour of employees. The new metrics should aim to be measurable without changing the current system and without influencing the behaviour of the actors. This is where reported emails can provide an outcome, as their presence is the desired behaviour, and employees are not influenced if reported emails are investigated. To find these metrics in a real-world environment, the reported emails by employees must be gathered and combined with external factors to understand their presence and provide insight into the relationship between training and behaviour. This research investigates which lessons can be learned from the reported emails in a bank by conducting exploratory research on a case study. This leads to the main research question of this research.

RQ. How can email reporting patterns in a large organisation measure the relationship between phishing training, reported emails and employee behaviour?

To answer this question, several sub-questions (SQ) are constructed to give direction to the research. First, the reported emails must be gathered, and it is necessary to understand which factors are measured. In combination with the literature, we can then determine which factors are worth investigating further.

With the knowledge of the components, we can examine the overlap between the components and the training the employees have received, which results in Sub-Question 1.2.

It’s essential to examine the progression of the reported emails over time, as outlined in Sub-Question 1.3. Answering this question provides the context necessary to answer the main Research Question.

With the trends in the reported emails, the influence of E-learning and the simulation can be investigated. As correlation and causation are easily confused, not only is the entire batch of emails analysed but influences in subgroups are analysed to determine whether effects are seen throughout the entire company or in smaller entities. Additionally, differences in job tasks can potentially explain the contents of reported emails, resulting in Sub-Question 1.4.

Additional input from the employees can provide an understanding of the company's current situation, raising the need to answer Sub-Question 1.5.

SQ1. What measurable factors are present in the reported emails?

SQ2. How do the components present in simulation emails relate to the training and the reported emails?

SQ3. Which trends can be discovered in the reporting behaviour of employees?

SQ4. Which differences can be found in the email content for different classifications and business lines?

SQ5. How do employees perceive the current anti-phishing training?

By addressing these (sub)questions, insights can be gained, enabling the formulation of practical recommendations to enhance the company's anti-phishing approach. These insights will play a role in assisting the company in implementing necessary changes to bolster its anti-phishing training and improve its simulation emails.

1.4.1. Characterisation of the system

The nature of this research is a characterisation to understand the system's phishing interactions better. As cybersecurity is a wicked problem and part of a complex system, overarching solutions are hard to find (Salloum et al., 2022). Phishing is essential to this system, making understanding each subdomain important. Furthermore, the study focuses on understanding the opportunities in the system and describing the phenomenon of reporting in context. It is a case study that explains in depth what the situation at the company can accommodate and what is being recorded. This also means that not every aspect of the problem can be sifted through in detail. Exploratory Data Analysis (EDA) provides the tools to create a characterisation of the current system and can guide future research.

1.5. Research approach

This research studies the reporting behaviour of employees at a bank. The large-scale analysis in this research is a significant contribution to existing literature, as this approach has not been adopted before. Using reported emails to investigate phishing behaviour addresses the research gap of studies looking into employee behaviour without influencing this behaviour with the chosen measurement method. A period of 16 months is analysed in which the employees were not influenced, as the research was constructed after the observed period.

With the reported emails in a bank, a characterisation of the behaviour of employees is constructed. We identify trends by visualising reported email counts and several attributes over time. Furthermore, with NLP methods, the emails can be compared based on the sent in the body of the email. The comparison between the content of the simulation emails and the reported emails can provide valuable insight into the effect of the simulation on the employees and their reporting behaviour. Additionally, extracting the topics creates an understanding of attackers' tactics and suspicious-looking benign emails. With the NLP technique Term Frequency - Inverse Document Frequency (TF-IDF), the content of the emails is compared, and the dominant topics are found. To further compare the emails, the components of interest have to be identified to give direction to the comparison. With a select number of interviews, the results are placed in the context of the company in this case study. These steps provide recommendations to the company and a direction for future research.

1.6. Outline

The problem will be defined in Chapter 2 to create an overview of the system under observation. Subsequently, the current training methods and their shortcomings will be introduced, after which the external effects and assumptions will be stated. Chapter 3 provides the methodology for answering the research questions and outlines the possibilities for data analysis and employee input. The results are indicated in Chapter 4 and show the development of reports over time and the content comparison between emails. Next, the results are discussed in Chapter 5, and recommendations are made for the company to develop their ant-phishing tactics. The limitations of the research are presented in Chapter 6. The study ends with the conclusions in Chapter 7.

2 Problem Definition

In this chapter, we addressed several core concepts and elaborated upon them to provide the necessary knowledge as the background of the thesis. Some of the concepts include the current literature, which explores concepts in the field of cybersecurity, specifically related to phishing emails. Furthermore, the behaviour of employees and the cues in emails to determine the classification of an email provide an understanding of the current phishing practices. Additionally, for this research, the opportunity arose to perform an analysis of reporting behaviour of employees at a bank. Because the data used in the analysis originates from a bank, the study's results are linked to the company. Therefore, the background and specifications of the company are outlined to sketch the restrictions and unique circumstances.

2.1. Phishing tactics

Phishing comes in various forms, and attackers utilise different mediums and content to deceive their victims. For this research, we chose only to study phishing emails and omit other mediums attackers can use.

There is an important difference between spam and phishing emails. Spam emails try to get you to interact with the email to use their service. The email portrays the content the sender wants you to interact with. Phishing, on the contrary, tries to breach the security of your company or computer by deceiving you with fake content, either by obtaining credentials or installing software on your computer. The computers provided by the company have preinstalled security software to fight these attacks. Due to these preemptive measures, many attacks are halted before they can do damage. However, not all attacks can be stopped, and secure behaviour by employees is needed to decrease successful attacks.

Depending on the time and effort attackers put into the attack, a phishing email can be easier or harder to recognise. One technique to make an email more difficult is that attackers use personal information to make their emails more convincing. With Open Source Intelligence (OSINT), an attacker can gather information about the target person by collecting and analysing publicly available data. For example, from social media, your job description can be extracted. If a phishing email is designed for a specific individual, it is called a spear phishing attack. Another tactic is to send out mass emails and see what sticks. In this case, the goal is to reach as many people as possible, hoping a small percentage is convinced of the genuine nature of the email. Understanding which tactics the attackers focus on can aid in detecting the attacks sooner and more efficiently.

2.1.1. Email campaigns

As stated, attackers can launch large email campaigns to target multiple people at once. For a company, it is beneficial to detect the presence of these campaigns to aid employees in detecting these emails. If the emails in the campaign are the same, this makes it easier for the automated detection and handling of the emails. In this case, only one of the emails has to be evaluated, after which all the duplicate emails are handled the same. For example, all similar emails are blocked before arriving in other inboxes. However, the difficulty with these campaigns is that attackers slightly change their specifics, making it harder to detect all emails which are part of the campaign only with technological measures. Therefore, detecting campaigns in different ways could benefit security.

One option is that phishing emails not detected by technological solutions can be identified through reports of employees (Caputo et al., 2014; Lain et al., 2022). In this scenario, management can view employees as an additional barrier to malicious emails rather than a risk.

2.2. Phishing training

The increased understanding of phishing emails led to a surge in the types of available phishing-related training. A summary of these techniques and their Key Performance Indicators (KPI) can be found in Table 2.1 and will be elaborated on further.

Table 2.1: Different types of phishing training and the popular Key Performance Indicators to measure their effect.

Type	KPIs
Embedded Training	1. Report rates 2. Click rates 3. Compromise rates
Mandatory E-learning	1. Completion rate 2. Time to completion 3. Correct answers (in case of quiz)
Classroom-based Training	1. Attendance
Email feedback	1. Follow-up reporting

2.2.1. Embedded training

With embedded training, counterfeit simulated phishing emails are sent to the employees to test their responses. These phishing emails are often constructed within the company and aim to mimic real phishing emails. With this test, it is possible to see how employees behave when they receive a phishing email. The KPIs of this type of training are often the number of people who reported the email, those who clicked on a malicious link, and those who entered their credentials. Simulated phishing emails are mainly a measure of behaviour but are often viewed as training as well. However, it is important to note that only those employees who click on links or are compromised receive additional e-learning information through a web page or automatic subscription to additional online learning. Additionally, as found by Lain et al. (2022), embedded training can make employees more susceptible to phishing instead of less.

Using simulation emails as a measure of behaviour makes it possible to evaluate the employee in a real-world environment where stress and distraction might affect them. However, the emails are often sent out in batches, creating a deviated signal. Furthermore, the emails are not genuine but fabricated to test certain types of attacks and can miss specific types of phishing tactics.

2.2.2. Mandatory E-learning

Another training option is to have mandatory e-learning for all employees. This type of training can take different forms, for example, providing reading material the employee has to go through to learn about phishing tactics and negation methods. The KPI of this type of training is whether an employee completed the E-Learning. Additionally, the time it took for the employee to complete the learning can be measured. Alternatively, the training can contain quizzes the employee has to pass before completing the training. In this case, the number of correct answers can function as additional KPI.

Gamification

A sub-type of E-learning is emerging in gamified E-learning, where participants learn through a game-like setup. Such interactive game formats can effectively educate employees about phishing attacks

(Sheng et al., 2007).

2.2.3. Classroom-based training

Training can also be given in a classroom-like setup. This can be in the form of a lecture or workshop where attendance is recorded. The workshop format can also be used to gather input from employees and engage them in how phishing training is provided.

2.2.4. Email feedback

Feedback after reporting an email is also a type of training (Jampen et al., 2020). This relates only to the employees who have reported an email and carefully read the feedback about the email's classification. With this information, they can evaluate new emails better.

2.3. Reporting rates

Ideally, training exercises would result in desired behaviour by employees. Regarding phishing, the desired behaviour is to report malicious and suspicious-looking emails. As the company registers all reported emails over time and their attributes, the development of these reports can be investigated.

Only the reported emails constitute the data analysed within this study. Employees have already identified these emails as potential phishing attempts. Using the reported emails to make policy for unreported emails can lead to survivor bias, as we concentrate on the entities that passed the selection process of the employees and exclude those that did not (Nikolopoulou, 2022). They do not provide information about the emails that succeeded in their attack and obtained sensitive information. Rather, it would be ideal to include information on phishing emails that no one reports, but this information is unknown. As there is currently no approach where unreported real phishing emails can be analysed, the reported emails will be studied with this limitation in mind, as awareness about this potential bias can reduce it. Moreover, the results have to be concentrated on the current reporting culture and avoid assumptions about unreported emails.

In addition to the reported emails, the company performs a phishing simulation. With this simulation, there is information about unreported simulation emails. Although the same statistics cannot be assumed directly for attacker emails, they indicate employees' security behaviour for these specific emails. Emails similar to the training exercises are likely to evoke similar responses.

2.4. Costs

When looking into cybersecurity measures, limiting the costs is an important aspect, as profit and loss are drivers for companies when deciding on investments (Gordon & Loeb, 2002). Costs related to phishing do not only correspond to the cost of a breach. If this were the case, companies would be more vigilant, adopt all possible security measures, and evaluate each received email carefully. Besides the costs of a breach, there are also costs related to the time employees have to spend dealing with phishing emails. People behave more securely if the costs of behaving as such outweigh the costs of the risk. In other words, the Return on Investment (RoI) should be positive. But, this factor is often not calculated (Moore et al., 2016), as estimating this risk is difficult (Cavusoglu et al., 2004).

It is possible to have all email traffic professionally evaluated. In this decision, several types of costs play a role. Predominantly, the consideration between the costs of a breach and the costs of paying an independent external company to evaluate the emails. Additionally, sending all internal email traffic to a third party is not the most secure option, as the emails can contain sensitive information, which is then shared with a third party.

Reporting an email generates additional transaction costs for employees. If an email is reported, it is transferred to the deleted mailbox. If an email turns out benign, the employee would have to return to their bin folder to retrieve the email. If the benign email contained an action the employee should perform, there are more related costs because the action is postponed. Additionally, the costs of losing employees' trust can be found in the increased resistance towards online communication and safety measures (Vishwanath et al., 2011). To conclude, if the number of false positive emails decreases, there are fewer costs for the company, both in costs for email evaluation and costs related to the time allocation of employees.

2.5. Email components

Phishing attacks are a form of social engineering where people's vulnerabilities are exploited. As found by Kamruzzaman et al. (2023), attackers do not have to be technologically skilled to carry out an attack. Instead, they can use psychological tricks to gather information, a technique used throughout history during wars and political schemes. Applying these tactics is becoming more accessible with the internet, as the possibility to reach people anonymously has increased and sending out mass (email) messages is one of these additional developments.

Every email has several components, such as the metadata and the body of the email. Based on these components, an employee or computer model can determine whether an email is phishing. We outline several of the components and tactics in emails which can indicate an email is phishing.

2.5.1. SLAM

The SLAM (Sender, Link, Attachment, and Message) method can be introduced within training (Kelvas, 2023). This method teaches employees to check whether the sender is trustworthy. When a sender is from outside of the company, a banner appears at the top of the email to alert the employee to this. Additionally, checking whether the link address is legitimate is important if an email contains a link. This can go further than the link in the email itself. Attackers can use techniques where an obscured link redirects a victim to an insecure address with URL shortening services (Lev, 2010). While the original link can seem trustworthy, the redirect address is not. These counterfeit addresses can persuade employees to share their credentials or download malware. Alternatively, phishing emails can contain an infected attachment, which can cause harm if macros are activated. Lastly, the message itself can be evaluated by its content. Whether the email is expected, in the correct style, or contains spelling errors. These are the factors most tested in training exercises as they are easy to incorporate.

2.5.2. Emotions

Incentives employees experience can cause them to behave in unwanted ways, decreasing the company's security (Asghari et al., 2016). Attackers can exploit this by evoking certain emotions, making people less focused on the task at hand (Zhuo et al., 2022). For example, topics related to presents or payments evoke a happy feeling and make people more perceptible (Williams et al., 2018). Furthermore, Hillman et al. (2023) found that employees are likelier to interact with the constructed simulation emails if they use personalised phrasing, which induces a familiar and safe feeling. As a result of these emotions, employees are more likely to overlook the authenticity cues taught in the training (Williams et al., 2018). Furthermore, time pressure and fatigue can decrease an employee's efficacy in reporting phishing emails. Companies can improve their security by introducing policies that address these factors and components.

Goel et al. (2017) present a case study where students are tempted with prizes or urged to act by implying they could lose something. The results show these types of phishing emails can be very effective and that the content of an email is important when examining an employee's behaviour. In addition to salary, Blythe et al. (2011) discovered that security issues are also a persuasive topic in phishing emails.

2.5.3. Cues

Emails contain cues that evoke a response from the readers. The presence of an URL in an email can provide a warning to an employee and is often mentioned in training.

Additionally, urgency cues in an email effectively persuade employees to interact with a phishing email. Vishwanath et al. (2011) found that attention to urgency cues and email titles is likely to trigger a response from people examining a phishing email. When a person has to focus on urgency cues in a text, there is less space to deal with other cues and thought processes. Williams et al. (2018) found that, besides urgency cues, authority cues are another effective tactic to increase the likelihood a person will click on a link.

2.5.4. Relevant Topics

Employees are less on guard if an email is similar to daily email traffic or if they have been expecting a specific email. Attackers have to guess which topic is most relevant when sending the email. Using words and phrases often present in company emails can also give an employee a sense of familiarity, one of the principles of social engineering attacks (Kamruzzaman et al., 2023).

2.6. Frameworks

Several models and frameworks have been developed to structure the factors and components of influence during a phishing attack. These models often either categorise the factors that influence the behaviour of employees or present a way to detect the difficulty of a phishing email. Using frameworks not only aids in structuring the factors but also protects the vulnerabilities within a company. By using proxies for factors, attackers cannot use the findings to explore weaknesses. We elaborate on several frameworks used in current literature.

Zhuo et al. (2022) presents the Phishing Susceptibility Model, which categorises variables into three stages: long-term stable, situational and in-the-moment variables. With this framework, the factors to include in research can be more systematically defined. Identifying which category a variable falls into can also help to establish recommendations to support employees better.

The NIST phish scale aims to identify the difficulty of a phishing email (Steves et al., 2020). This can help to explain high click rates and provide an alternative explanation for behaviour besides the characteristics of employees. For example, a benign email may look suspicious and be reported by an employee. Rating emails based on characteristics instead of intent makes it clearer where malicious and benign emails overlap and differ. The paper of Steves et al. (2020) creates a Phish Scale with two parts. One part includes a rating system for the observable characteristics of the phishing email. The second part considers the email's alignment with the target audience.

Sutter et al. (2022) builds upon the Phish scale and presents a Machine Learning (ML) model to predict the number of people who would fall for phish email based on the two proposed Phish Scale components. The results show a ML model that indicates the alignment between the employee's workplace and email content is a key factor for the difficulty of a phishing email. Therefore, as the difficulty of an email is related to the click rates, this research indicates a possibility for machine learning to predict the clicking rates of employees. Furthermore, combining the difficulty of an email with other types of information, such as job specifications, could contribute to developing a more general model.

Current research only determines the difficulty rating of constructed as part of the study, not real phishing emails.

2.7. Behaviour of employees

The objective of phishing training is to boost cyber awareness, enabling employees to spot, avoid and report phishing threats as this safeguards the client and company data which relates to the values of the stakeholders. The employees should be wary of the possibility that an email is phishing. Once they find an email suspicious, this should trigger them to think back to their training about dealing with this kind of email. The desired behaviour of employees is to report phishing emails. Contrary, engaging with phishing emails is naturally the worst behaviour. Alternatively, an email could be ignored or put in the trash. Although this is safe behaviour, a company could learn more from reported emails. They can signal campaigns targeting employees now or help to identify future threats. Regular, non-phishing emails should warrant engaging interactions. The factors influencing an employee when deciding on a suspicious email are shown in Figure 2.1. There is both agreement and disagreement among scholars regarding the factors that affect employees.

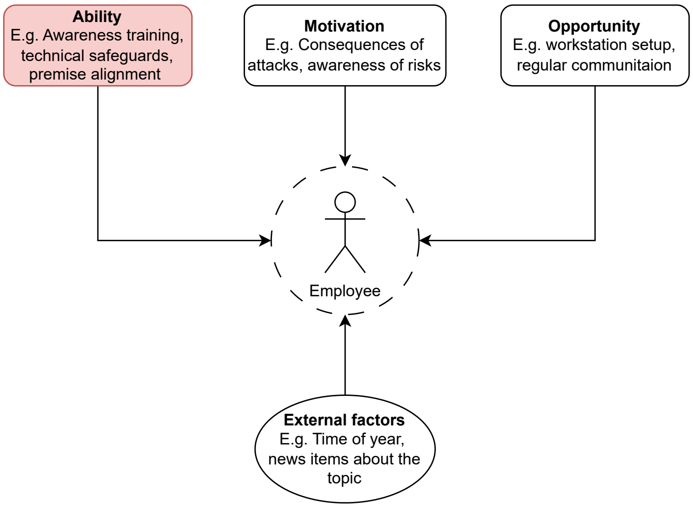

Ability
E.g. Awareness training, technical safeguards, premise alignment
Motivation
E.g. Consequences of attacks, awareness of risks
Opportunity
E.g. workstation setup, regular communitaion
Employee
External factors
E.g. Time of year, news items about the topic

Figure 2.1: The MOA model is used to describe factors that influence the phishing behaviour of employees. Phishing training generally focuses on the ability of an employee, marked in red.

2.7.1. Consensus

A characteristic that influences an employee's interaction with an email related to their ability is the difficulty of the phishing email (Steves et al., 2020). Certain emails are much more difficult for an employee to distinguish; someone who deals with invoices all day is naturally less suspicious of another invoice from an unknown email address. This person can be more suspicious about an email asking for employee credentials and correctly identifying the email as phishing without any other change in the situation. Research using this insight is limited, and the difficulty of an email for an individual employee is hard to determine as you need detailed knowledge about the employee.

Factors such as someone's mood or workload are more personal and situation dependent. Existing literature shows that during times of stress, an individual behaves differently and is more susceptible to phishing emails (Zhuo et al., 2022). There is less space to think thoroughly about your decisions. It takes time to carefully evaluate whether an email is legitimate, and during stressful situations, there is an increased likelihood of errors (Reason, 1990). Email load and deadlines can worsen this. An assumption is that the more time employees spend with an email, the more accurately they identify it (Tyler & Davey, 2020). Thus, if someone is busy or distracted, they will likely click on a malicious email. Understanding the behaviour of employees over time shows the main risks and indicates which technological or sociological developments should take precedence.

2.7.2. Conflicts

Several papers find conflicting results regarding whether a factor is important when predicting if an employee will fall for a phishing email. For example, age and gender are conflicting attributes. Some studies find women are more susceptible to phishing (Halevi et al., 2015; Iuga et al., 2016; Jagatic et al., 2007), while others contradict this (Rocha Flores et al., 2014; Zhuo et al., 2022). Furthermore, higher phishing-related knowledge is generally agreed as a factor that can reduce phishing susceptibility (Zhuo et al., 2022). However, Georgiadou et al. (2022) used the period during the COVID-19 crisis to test participants' attitudes, competency and actual behaviour. There seemed to be a mismatch between the competency tested with a questionnaire and the behaviour of identifying phishing emails. This indicates that increased general cybersecurity knowledge does not directly improve phishing-related behaviour.

Overall, problems seem to arise with the causality of results and the conflicts in the literature signal that results can easily be misinterpreted.

2.8. Setup at company

Companies can have different email and phishing attack management systems in place. To comprehend the research's scope and limitations, it is essential to have a basic knowledge of the setup of the company being studied. This information allows the research recommendations to be specified and tailored to the selected case.

2.8.1. Bank characteristics

Banks are familiar with phishing threats, as attackers often pose as bank employees to scam unsuspecting customers to provide credentials to gain access to their bank accounts (Nederlandse Vereniging van Banken, 2023) With a successful attack in the Netherlands, the bank is partially responsible for paying back victims (Slachtoffer Hulp Nederland, 2023). A bank does not only experience threats to their customers, as employees themselves can also be victims of phishing attacks. Even more so, from the 2022 annual reports of several banks in the Netherlands, phishing is the most common way for attackers to gain access to the bank's system (De Nederlandsche Bank N.V., 2022; ING group N.V., 2022). This research focuses on these last forms of attacks aimed at employees and the threats related to breaches on this side.

2.8.2. Stakeholders

We must identify the stakeholders to understand who our recommendations can aim for and which stakes we must consider. From a stakeholder analysis, we identify several actors: customers, shareholders, employees, attackers and society. They all have different priorities when it comes to this topic. For customers, using the provided services safely is most important. Alternatively, shareholders want to secure the company's profit, while employees want to be able to perform their tasks without delay or blame for their behaviour. Furthermore, attackers sending phishing emails serve as threat actors in this system and try to gain access. Lastly, society is actively involved as a bank performs a public service. Would something happen to the bank, the consequences could be widespread and negative for the public.

Previous research found that bank employees had better Information Security Awareness (ISA) than the general public (Reeves et al., 2020). This suggests their response to emails is also different from the general public's, creating a subgroup of potential victims with their own security profile. This is a further argument that the conclusions from this report cannot be generalised to other sectors without additional considerations.

Overall, in case of a breach, multiple entities are affected. Our recommendations during this research will focus on the actions the company and its stakeholders can take to create a safer environment. This includes the steps that employees who are responsible for providing training to their colleagues can take. We will only include changes the company can make, excluding external aid the company deploys.

2.8.3. Email flow

When an email is sent, there are multiple stages it can go through to determine whether the email is safe. Figure 2.2 provides an overview of the steps and checks an email undergoes.

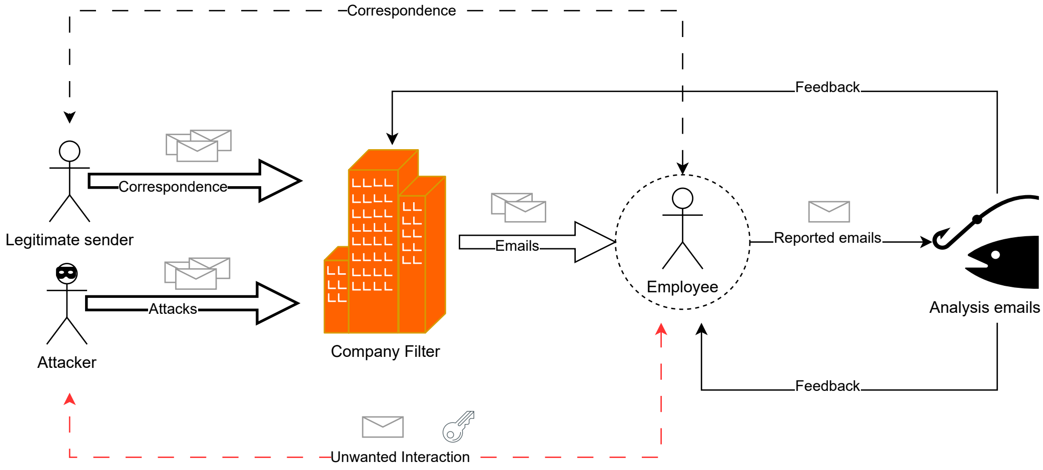


Figure 2.2: An overview of the email flow within the company. Starting on the left with the email's sender, who can be legitimate or an attacker. This sender can be from in- or outside of the company. These emails pass through a filter before arriving in an employee's inbox. This employee can interact with the email or report it as suspicious. If the email is reported, it is evaluated, and feedback is sent to the employee.

The company filter is dependent on the email service providers. On the one hand, attackers can challenge and test this filter by creating an account on the same carrier and sending themselves emails to investigate which ones manage to bypass the filter. On the other hand, these filters change over time and adapt based on the type and content of known phishing emails to block attackers better.

Once an email arrives in an employee's inbox, they decide how they interact with the email. Most commonly, the email is part of normal email traffic, and the employee responds to the email. In some cases, however, the email appears suspicious to the employee, and they can report the email. They can do this from within their email carrier by forwarding the email to the correct email address or clicking a 'report phish' button in their inbox. In the last case, the email is automatically forwarded to the correct email address. Additionally, the email is immediately moved to the bin folder of the employee's email.

If an employee reports an email, an external company evaluates the email and determines whether it was a phishing attempt or a legitimate one. The final classification is then sent back to the employee. Additionally, if the email is classified as malicious, this is communicated with the department responsible for the email filters, after which they can adapt their anti-phishing filters.

2.8.4. Training in the company

Each employee in the company receives various types of training to achieve different objectives, such as creating a safe workspace and promoting responsible banking. All employees also receive anti-phishing training to make employees more resilient against phishing emails. To help employees identify and report suspicious emails, they receive training organised within the company. This is a mandatory e-learning training every employee has to complete upon joining the bank. This knowledge learned through training does have to be maintained. According to Reinheimer et al. (2020), a reminder must occur every six months to keep the knowledge current. Within the company, the mandatory E-learning training is repeated for all employees periodically, approximately once every year.

The second type of training adopted in the company is sending simulated phishing emails in a wave to all employees. With this approach, employees get an email with a link, and if they click on the link and submit their password, they are sent to a page with information about the fact that the email was part of a test and how they could have recognised it. The cues for identifying the emails as suspicious are taught in the E-learning. The simulation wave is both a tool to measure the behaviour of the employees and to provide additional selective training to employees who did not recognise the phishing attempt. Not all employees receive these emails, and during the most recent wave, about 50%

of the employees were randomly selected. In addition, some departments within the company have embedded simulation training throughout the year, targeting all their employees. The company has the ambition to extend this to all locations, but this approach will be implemented after the completion of this research and is, therefore, outside the research scope. However, when making recommendations for this type of training, it's crucial to consider these changes.

The feedback on the reported emails is another instance for employees to learn about phishing. If an email is indeed phishing, they receive the feedback they correctly identified the email. If the email was benign, they also get this feedback, signalling that the email was legitimate. Employees can apply this knowledge to the subsequent emails they are unsure about.

2.8.5. Data contents

The company records a lot of phishing data, among which the emails employees have reported as potential phishing. We can access these emails reported from January 2022 to April 2023 to perform the study.

During this year, employees have received e-learning training, and there has been a company-wide simulation wave where counterfeit phishing emails were sent to 50 % of the employees. We know the content of the emails and the click, compromise and reporting rates of the simulation wave. Additionally, the reported simulation emails are part of the larger dataset with all reported emails. For those, we have all the attributes the regular reported emails have as well.

We elaborate further on the content of the datasets in Section 3.2.2.

2.9. External effects

As we show in Figure 2.1, besides the identified factors in the data, numerous external events or circumstances can influence the reporting behaviour of employees. Partially due to these external effects, direct causality cannot be shown because many factors can differ from person to person.

If someone has a bad day, their behaviour could be different. In addition to personal and psychological factors, external factors such as your surroundings can also impact your behaviour. Concerning phishing, this can be global news events which influence employees. Additionally, newsletters from the company are sent out every month containing information about recent threats and best practices that should be considered with extra attention.

These kinds of external effects cannot be accounted for within this research but are mitigated by looking at general statistics instead of individual behaviour. Furthermore, it is possible to show a correlation between the factors in the data without implying causality, but the correlation can still be significant. Additionally, some external events are known and mapped. These are based on previous research.

2.9.1. Considered external factors

The following measures will be taken to address some concerns and external influences.

1. Some events can influence reporting behaviour. A timeline will be created to provide an overview of events that can influence the reporting numbers. This timeline contains the intranet messages addressing phishing, phishing-related events in the news, and the general holiday period.

2. There is fluctuation in the number of phishing emails malicious actors send. A CISCO report analysis about cybersecurity in 2020 shows that the number of phishing emails fluctuates, with an outlier in December (CISCO, 2022). This will be considered during the analysis and added to the timeline.

3. If there are more employees, the reported emails will also increase. The number of reported emails will be corrected for the number of people within the company and/or business role.

There are several factors we cannot correct for. Therefore, the conclusions that can be drawn are limited as there can be alternative explanations for several of the findings.

2.9.2. Assumptions

During this research, assumptions are made concerning several external factors.

1. The total number of emails an employee gets stays consistent throughout the year.
2. An external company classifies the emails employees have reported as suspicious. The assumption is that this classification is accurate for all emails.
3. All employees receive the same training. An employee can have missed the E-learning during the observed period, but this is a low number of employees. Additionally, new employees must pass the training at the start of their job here. Due to these two reasons, we assume all employees have completed the phishing e-learning training.
4. People use the mail inbox on their laptop, and/or the used interface does not change the reporting behaviour.
5. The email filter of the email service provider is consistent for the duration of the research.

3 Methodology

The correct method must be identified to answer the stated research questions. With the aid of EDA and interviews among staff, the patterns in reporting behaviour can be examined. We observe the current system and the reporting rates over time. Additional research will be done on the content of the reported emails.

3.1. Exploratory Research

This research is data-driven, made possible by the bank's data availability. Furthermore, the nature of the research is exploratory. This brings along several limitations and challenges and steers the focus of the results.

With EDA, the aim is to analyse datasets and summarise the main characteristics to increase the understanding of the system. With input from the company, the current system can be mapped. Additionally, we can provide insight into the possible next steps the company can take to improve the current system. Furthermore, the results and uncovered relations can serve as a guide for future research.

One of the drawbacks of conducting EDA is the lack of definitive conclusions that can be drawn. Furthermore, the obtained results may be biased or subjective (Chatfield, 1995, p. 25). This often makes them non-generalisable. Considering these properties of the chosen method, the primary value of this research will be the method and metrics developed to measure the influence of simulation waves, as the technique is broader applicable and not affected by the outlined limitations.

3.2. Datasets

The first step of the analysis is to uncover and understand which data is accessible, providing an answer to Sub-Question 1.1. Within this research, several datasets are combined to make the analysis possible.

First, the information about the current system will partly be gained from data representing the behaviour of employees surrounding phishing reports. We use the emails reported in the last year for this. For each email, it is known who reported it, who sent it, what the classification is and several attributes of the email itself. A list of all factors present in the dataset about the reported emails can be found in Appendix A. Privacy regulations prohibit using individual characteristics such as age.

Second, one dataset contains information about the mandatory E-learning employees followed. This data contains the course, the moment an employee finished the training and the employee's job description. This job description can be aggregated to the business line an employee belongs to. Due to privacy regulations, the E-learning cannot be traced back to the original employee who finished the training. However, the vast majority of employees finished the training before data for this research was collected. Therefore, as it is impossible to link the E-learning completion to the individual employee, only the aggregated counts are used.

Third, the information about individual employees is contained in another dataset. For each employee, we know their email address and their job description. The lowest collected level of information is to which department an employee belongs.

Fourth and last, information about the counterfeit emails from the simulation wave is known. Within the wave, ten simulation emails were used. We know the content of these emails, compromise rates, click rates and number of correctly reported emails.

By combining these datasets, it is possible to determine when the training took place and the changes in reports over time. Once the data is collected and merged accordingly, it must be cleaned.

3.2.1. Data Cleaning

After collection, data is formatted to be analysed. We examined the data quality and identified several issues.

First, there are some missing values in the employees' job descriptions. Additional information for incomplete entries was provided based on additional data, as described in (Chatfield, 1995, p. 37). By combining additional sources of information and personal communications with company experts, gaps in employees' business lines have been filled whenever possible. Unfortunately, this was not the case for all entries.

Second, we discovered missing information about the employees who have left the company since they reported an email. Where possible, additional data is used to assign the correct attributes to these employees. Where this was not possible, the information was left blank. Consequently, if the missing information was vital for a particular analysis, these reports were not used for that particular analysis.

Third, some employees have multiple roles. This is possible if an employee changed their job in the analysed period but stayed within the company. In this case, we attributed the most recent job description to the employee. Because the analysis will focus on the aggregated departments, and employees mostly change jobs vertically within a business line instead of horizontally between business lines, this should not influence the outcomes significantly.

Fourth, some emails were classified incorrectly while they were part of the simulation wave. These emails have been found based on their title and given the correct classification. Additionally, it was checked the sender was as expected, which was the case. As the number of emails was very select (less than 1% of the simulation emails) and the title of the email was known, using only the title to locate these emails was sufficient in this specific case.

3.2.2. Data Characteristics

The data contains reported emails with several attributes. The most important attributes which are used for the analysis are outlined below. The dataset contains all emails employees have reported between January 1st 2022, and April 1st 2023. The emails have been categorised into five distinct classes, which are explained below.

No Threat Detected: The email does not contain a threat, and normal interaction with the email by the employee is safe. The employee incorrectly identified the email as potential phishing. In the study, these emails are further referred to as benign emails.
Malicious: The email contains a threat and is correctly identified as potential phishing by the employee.
Simulation: The email is part of a simulation sent from within the company. The employee correctly identified the email as potential phishing.
Do Not Engage: Although the email does not contain a direct threat, the motivation behind the email does not seem legitimate. Therefore the advice is to not interact with the email further. The employee

correctly identified the email as potential phishing.
Further Investigation Required: The report does not include enough information to classify the email. The main reason for this is an incorrect method for reporting the email, for example, without the suspicious email body or title attached, making it impossible to classify the email and determine whether it is a threat.

If an employee reports an email, the email is analysed, and the classifications, as stated above, are provided as feedback to the employee. The percentages of the classifications in the dataset can be found in Table 3.1.

Table 3.1: All classifications and the percentage of emails classified as such.

Classification	Percentage
No Threat Detected	34.8 %
Malicious	26.1 %
Simulation	23.4 %
Do Not Engage	13.3 %
Further Investigation Required	2.5 %

Besides the classification, each reported email can be assigned to a specific business line. The company has several business lines, and each employee belongs to one. For each business line, it is shown what percentage of emails are reported by someone from that line and how many of the employees are part of it. Several reported emails are reported by an employee not part of any business line, as it was impossible to link the employee to one. As described in Section 3.2.1, this can be if the employee has left the company since reporting the email or the job description is no longer active.

Table 3.2: All business lines and the percentage of emails classified as such.

Business line	% reports	Total % employees
A	29.9 %	38.7 %
B	28.9 %	28.6 %
C	16.2 %	10.4 %
D	10.4 %	14.3 %
E	5.2 %	2.8 %
F	3.2 %	4.8 %
No assigned business line	6.2 %	-

3.3. Research Design

The design of this research consists of two parts. First, quantitative research is conducted by analysing datasets provided by the company. This quantitative research includes the analysis of the reported emails over time and an in-depth analysis of the email content in the dataset. Second, we conduct smaller qualitative research through interviews with several employees.

3.3.1. Patterns over Time

Locating the timing of the E-learning and simulation wave in the plots over time provides the opportunity to see the changes in the factors in relation to the training.

Each email has a timestamp of the moment of reporting. To study the effect of the training on the reported emails, we will use the timestamp to analyse the patterns. To this end, the data is grouped by the value of the characteristic and the moment of reporting for each characteristic of interest. Next, the email counts are aggregated per week or day, depending on the intended goal of the analysis. Then, the counts of the items in the group are put into a data frame and plotted over time with the counts of emails plotted by the characteristic of interest.

After thorough experimentation, we opted to examine these specific characteristics.

- The classification
- The unique reporters
- The unique reports

Classification
The classification of an email shows the accuracy of the employees in reporting emails. Comparing true and false positives can provide insight into high benign report rates in the company established in Table 3.1, making malicious and benign classifications of particular interest. Therefore, we analyse the difference in reporting rates between emails classified as benign and malicious. This will be done in relation to the two types of training the employees have received, the E-learning and simulation wave. Furthermore, we adapt two additional measures to determine the significance of the results.

Correlation
Besides the visual cues of seeing counts change over time, it is possible to determine the correlation between the two variables. Pearson's Correlation Coefficient provides a simple and commonly used metric to measure linear correlation using Equation 3.1. Here r is the correlation coefficient, x is the first variable, and y is the second variable. For each element i, the difference between the element and the mean ($\overline{x}$ and $\overline{y}$) is calculated and summed. This is divided by the multiplication of the standard deviation of both variables. In this research, x is the measurement of benign emails, while y denotes the measurement of malicious emails.

$$r = \frac{\sum(x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum(x_i - \overline{x})^2 \sum(y_i - \overline{y})^2}} \tag{3.1}$$

The value lies between -1, meaning complete negative correlation, to 1, meaning complete positive correlation. 0 indicates there is no correlation. Generally, a coefficient between 0 and 0.4 indicates a weak or no correlation, between 0.4 and 0.7 a moderate correlation and higher than 0.7 a strong correlation (Thomas, 2023).

P-Value
To determine the significance of changes over time, we use the independent two-sample t-test, or Welch's t-test, to compare the data before and after the simulation. The p-value is the probability of obtaining your, or more extreme, results under the assumption that a chosen event does not affect the measurement. In this research, we test the (null) hypothesis that the counts in reports before the simulation have the same mean as those after the simulation. The value is calculated using Equation 3.2. Hereby $\overline{X}_i$ is the sample mean of the i^{th} sample and $s^2_{\overline{X}_i}$ its standard error.

$$t = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{s^2_{\overline{X}_1} + s^2_{\overline{X}_2}}} \tag{3.2}$$

Rejecting the null hypothesis based on the t-test does not prove the relation between factors; it can only support or give ground to dismiss a hypothesis.

Unique reporters
Not only the classification is an interesting factor to observe over time. Instead of the total email count, the unique reported emails are a measure to increase the understanding of the data. Additionally, the number of unique people who report an email can be a metric of the security behaviour in the company. With this metric, it is possible to see whether new people report emails or if the same people keep reporting emails with few additional employees using the report function. Additionally, the number of unique reporters can increase over time, indicating a security culture where various employees participate in reporting emails.

Unique reports
We can gain an understanding of the underlying behaviour of employees if we compare unique reports with the total reports. However, unique reported emails are ambiguous to group as unique because emails with duplicate titles can have different content. Additionally, attackers can change email addresses and headers slightly to circumvent filters (Aneke et al., 2019). Therefore, we chose to group the emails based on their title, but only if the emails were sent within two days of each other. This period is determined because most phishing campaigns only last for a short period, less than a day (Cimpanu, 2020). Additionally, 95 % of the reported emails in the data are reported within a day, with a few outliers.

External Factors
There are external factors that can influence the reporting behaviour of employees, and we made several assumptions to perform the analysis. The results must be examined in relation to several of these assumptions, as stated in Chapter 2. To account for these factors, they have been mapped over time. It has been determined whether the expectation is that the reports would in- or decrease due to the external factor.

3.3.2. Text Similarities

To increase the understanding of the reported emails and their relation to the simulation emails, we compared their contents based on the text in the body to answer Sub-Question 1.2. For the text analysis, emails must first be preprocessed, after which we use several techniques to compare the text in the emails.

Email text preprocessing
Before comparing emails, it is necessary to preprocess the text in the body. As the method depends on the text's language, the decision is made only to include English emails in this part of the analysis. The majority of the emails are in English (76.7 %) and evenly distributed over the data in terms of classification. The following steps are taken during the preprocessing, using the NLTK library in Python (Bird et al., 2009).

- All non-alphanumeric characters are removed.
- Stopwords were removed.
- The words have been lemmatised.

The non-alphanumeric characters and stopwords are removed to retrieve results based on the core words of a text instead of also including nonsense and filler words which would dominate the results (ProjectPro, 2023).

Lemmatisation of a word brings it back to its root meaning (Yash, 2023). For example, running and ran are both brought back to run. This step cleans the text and creates more human-readable results than alternatives such as stemming.

The chosen NLTK library is user-friendly, can be combined with other techniques used during the analysis and has extensive options for preprocessing the data, which can be customised where necessary (Bak, 2019). With the preprocessed text present, there are several possible approaches. Different NLP techniques are combined to get more insightful results.

Topic Modelling
Topic models aim to uncover hidden structures in a collection of texts. The chosen topic modelling model operates on numeric data instead of raw text. Therefore, every word gets a unique ID to make comparison possible. Next, a Latent Dirichlet Allocation (LDA) model (Blei, 2012) is used to discover the latent topics in the emails. This type of model is the most used for this type of analysis, making the use simple and well-documented (Asmussen & Møller, 2019).

The LDA model sees the text as a mixture of all the topics. The model's outcome represents the 'topics' that best represent the information in them (Kapadia, 2019). The LDA model requires a predefined number of words per topic. There is no guideline for the number of topics, and the recommended approach is to try a few different numbers and compare the analysis outcomes (CR, 2020). If words are repeated among multiple topics, the number is too large.

A downside of the LDA model is that some topics may be difficult to interpret, influencing the interpretability of the results. While an important limitation, the study's exploratory nature calls for flexible methods. Additionally, as the method is new for this type of analysis, the results can determine whether this method is useful for future metrics. Starting with this approach can show whether future research should adopt more advanced methods, repeat the current method or exclude this type of analysis completely.

Alternative approaches include Bidirectional Encoder Representations from Transformers (BERT), a NLP method, which can use the context of a word to determine the meaning of ambiguous language (Devlin et al., 2019). This research chooses not to apply BERT due to its computationally extensive nature (Simha, 2021). With the large amounts of emails in the dataset, running BERT significantly increases the computation time. As the research is mainly exploratory and subject to changes, the choice has been made to use a more flexible method.

Analysing the topics can be done for subgroups of the data to answer part of Sub-Question 1.2 and explain potential differences. The benign and malicious emails will be analysed separately. The topics can then be compared to see if there are striking similarities or differences in the words that indicate the most popular topics in the emails. Additionally, the analysis will be done per aggregated department. For this, the departments have to be aggregated into sensible larger groups. There are too many small departments to analyse every small group of people. The challenge lies in classifying the departments into useful groups. The choice has been made to use the business lines the company already uses.

Text comparison
Besides extracting topics from the text, the emails have been compared based on the similarity in their email body to add to the interpretation of Sub-Question 1.2. The benign and malicious reported emails are compared to the content of the emails used in the simulation wave. The most frequently used method to compare emails on their content is TF-IDF (Salloum et al., 2022). With this method, text files are converted to a vector representation. The resulting vectors can then be compared with the cosine similarity score.

Cosine similarity
The cosine similarity score is often used to measure document similarity in text analysis (Agarwal, 2013). First, a text must be converted to a vector representation to apply cosine similarity. The score ($sim(A,B)$) is then calculated with Equation 3.3, where A and B are the two texts we compare, but in their vector representation.

$$sim(A,B) = \frac{\sum_{i=1}^{n} A_i B_i}{\sqrt{\sum_{i=1}^{n} A_i^2 \cdot \sum_{i=1}^{n} B_i^2}} \tag{3.3}$$

The similarity score ranges from 0, indicating no similarity between the two documents, to 1, meaning exactly the same.

Component comparison
Sub-Question 1.2 can be answered further and in-depth by analysing the component in an email. These components relate to the E-learning, and an employee is taught to look at them to identify the nature of an email are numerous, as described in chapter 2. Most components have to be extracted from the body of the email, as they are not provided separately for each email. The components in the analysis are based on previous research and preliminary documents from the company to minimise confirmation and success bias. The preliminary documents include a difficulty rating the company uses to determine

the difficulty of their simulation emails.

These components allow us to identify how many cues an employee could use to determine the email's classification. We can also compare the components in the simulation emails with those in the reported emails.

3.3.3. Interviews

Three interviews are conducted to increase the understanding of the phishing culture at the company. We can use the interviews to validate and enhance the credibility of the research findings. It is important to note that the conclusions drawn from these interviews may not be broadly generalisable due to the limited sample size. However, the insights obtained through these interviews offer valuable perspectives and supplement the larger body of research. While one should not deem these conclusions representative of the entire security culture, they provide useful qualitative information and contribute to a deeper understanding of the system.

As the company is large and provides several services, it is impossible to create a complete overview of all perceived threats and opinions among employees. The three interviews aim to get a sense of the possible opinions in the company about phishing practices and the training employees receive. Additionally, a sense of the work environment can be gained. Managers within three different departments of the company have been contacted. This is because department managers have an overview of their department and can represent more people than a single employee. An assumption is that a manager has a general sense of their employees' average tasks and can estimate their way of working. As this is not scientifically validated, this will be checked explicitly, and the participant will be asked if they know what actions their team members take upon receiving a phishing email.

The three participants have different tasks and have different relations with phishing emails. One employee provides phishing-related services to clients and employees, and the second deals with security but not specifically phishing. At the same time, the third does not have phishing and (cyber)security as daily topics in their tasks. Additionally, the three participants have opposed interaction with external entities. The contrary job descriptions result in a broader understanding of the possible opinions and differences between departments.

Questions about the behaviour of the participant and their team have been asked. The design of the interview questions plays a critical role in preventing personal backlash for the participants and ensuring unbiased and reliable information. To minimise the potential for leading questions, these questions are constructed with the company's aid and have been discussed. Appendix B contains a list of the questions serving as the guideline for the interview.

The interviews will provide a backdrop of the company culture for the results and ensure the recommendations do not suggest existing measures the company already deploys. Furthermore, the results of the interviews are not generalisable to other companies.

3.3.4. Recommendations

This research ends with recommendations to the company to improve its phishing training and simulation emails. The recommendations are based on the results from the study in combination with existing literature to validate the findings. Additionally, the interviews inspire and help determine which tactics are already deployed in the company. Unfortunately, these recommendations cannot be tested, and their effectiveness cannot be measured before implementation.

3.4. Ethical considerations

As attackers can misuse the information in this research, precautionary measures had to be taken to ensure the safety of the company and individual employees. To safeguard the participants, an ethics application has been filled out, which is approved by The Human Research Ethics Committee (HREC) of the TU Delft. As part of this application, several risks have been explored and mitigated where possible.

The research has been limited if it was impossible to guarantee safety and anonymity. Specifically, the E-learning data is not linked to individuals, and no analysis is done on individual behaviour by linking reports to personal information.

3.4.1. Data safety

Due to the confidential nature of the data, not all results will be publicly accessible. Measures will be taken to guarantee confidentiality along with proper results. To safeguard the identity of the employees who reported the emails, the results are aggregated and cannot be linked back to individuals.

Strict data protection protocols were followed throughout the research process to ensure the secure handling of sensitive data. Discussions and collaboration regarding the data and code were limited to the internal research team and authorised individuals. Additionally, an expert from the company has considered the choices made in the data cleaning process. Furthermore, any external discussions involving the data were conducted under the appropriate confidentiality agreements and in compliance with the company's data protection policies. These measures were taken to safeguard the organisation's and its employees' privacy and security.

The data availability partly drives the direction of this research. Depending on the quality and quantity of the data, the findings of this report can vary in their reproducibility and extent. Furthermore, only factors that can be observed can be analysed, and findings can only be externally validated if similar data is available. However, by determining the factors or components to look into beforehand based on the literature, the availability does not entirely determine the entire research. Additionally, the study's results are linked to the company because the project will be within a bank. The results may be exclusively applicable in this context, making them unsuitable for general conclusions to apply within the field of cybersecurity. However, as stated by Flyvbjerg (2006), *"formal generalization ... is considerably overrated as the main source of scientific progress"*. Non-generalisation is a limitation that must be considered, and aiming to find general conclusions is a sub-goal. However, it does not make the research a failure if the results are not widely applicable.

3.4.2. Interview limitations

The interviewees' identities have to be protected. There are a limited number of participants, and not all provided information can be used, as this would make it possible to identify individuals based on their job descriptions or other attributes. Therefore, statements have been generalised to protect the participants' identity in the interviews. Additionally, the questions in the interview do not steer a participant towards admitting unwanted behaviour. This information could be harmful to the prospect of their career.

The interviews do not necessarily reflect the opinion of the company or researcher.

4
Results

4.1. E-Learning and Simulation

The timing of the E-learning and simulation wave has been laid out to investigate their effect. To determine the spike in E-learning, the employees have received, we plotted the number of E-learning courses completed per week in Figure 4.1. The first E-learning was completed on February 3, 2022. Employees were given four weeks to complete the training, and most employees completed the training in the mandatory period. However, some employees complete the training later in the year due to external factors such as a sabbatical or a pregnancy resulting in the employee's absence.

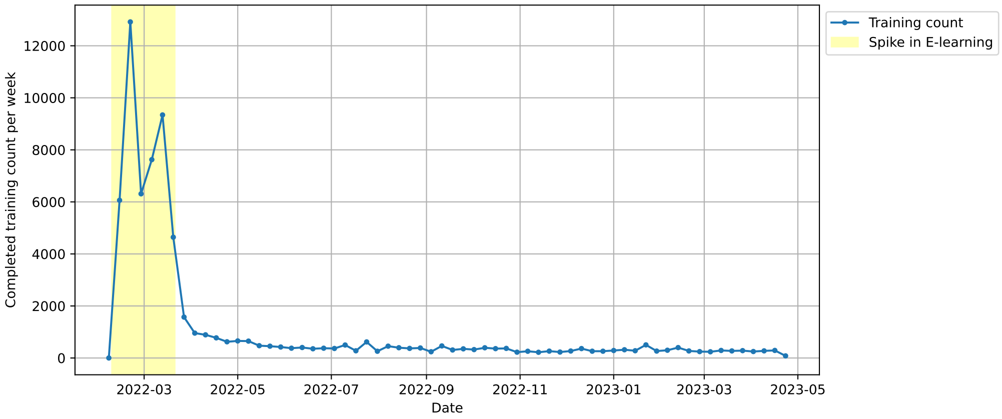


Figure 4.1: The development of the number of completed E-learnings per week. The counts are accumulated per week, where each dot presents the Sunday of the week the training was completed.

Besides the mandatory E-Learning, employees receive simulated phishing emails containing elements discussed in the training. The company-wide simulation campaign started on November 14 2022 and ended on November 25 2022. A smaller campaign was performed at the end of February 2022, limited to one division. In April 2023, there was a simulation in one of the Tech hubs of the company. Additionally, throughout the year, there are several simulations sent to employees. These simulations are limited to specific countries. An overview of all reported simulation emails and the duration of the simulation wave can be found in Figure 4.2. With this information, we can examine the development of the reported emails over time in relation to the training the employees have received.

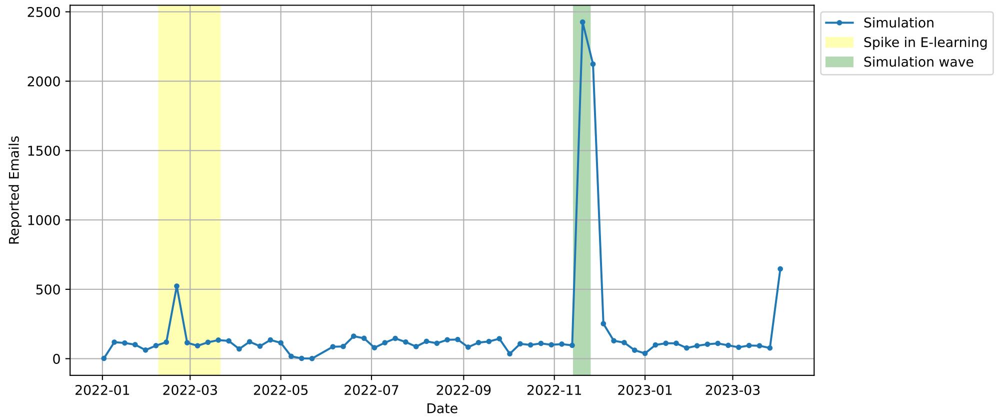


Figure 4.2: The development of the reported simulation emails.

4.2. Development over time

The analysis of the data has been split into two parts. First, the effect of the simulation wave on the reports over time is examined to answer Sub-Question 1.3.

4.2.1. Classification

The peak in the simulation emails, as shown in Figure 4.2, covers the results of the other classifications. Therefore, in Figure 4.3, the simulation emails have been removed from the data. The reported emails are mainly classified as No Threat Detected or Malicious, as already seen in Table 3.1. Several peaks can be identified. For the benign emails, this includes a period after the simulation wave. A clear drop in reports is present at the end of December, explained by the many employees on holiday during this period.

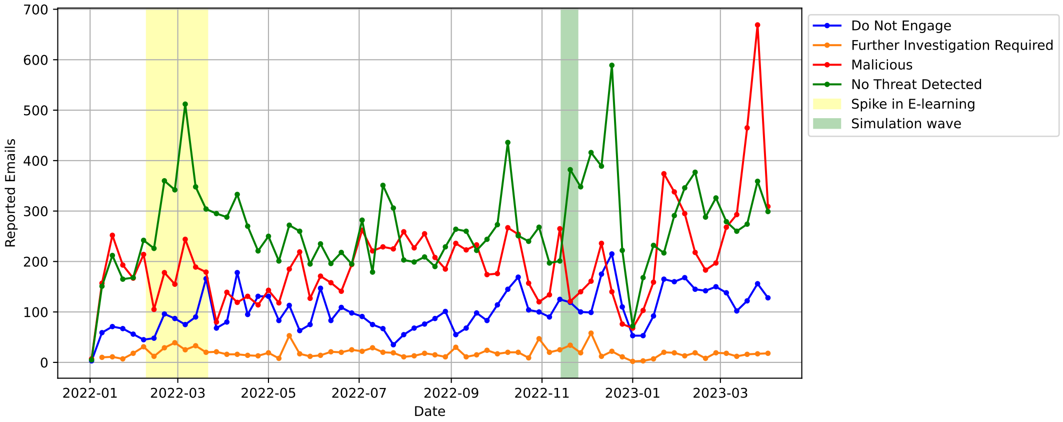


Figure 4.3: The development of the reported emails where the simulation emails are taken from the dataset. The counts are accumulated per week, where each dot presents the Monday of the week. The legend shows the classification of the emails and the two time periods for the training and simulation wave.

We can zoom in on the area around the simulation weeks in Figure 4.4. The days of the week can be seen clearly, as the reports draw near zero during the weekend. There is no increase in Malicious reported emails after the simulation compared to before. The benign emails do seem to increase during and after the simulation.

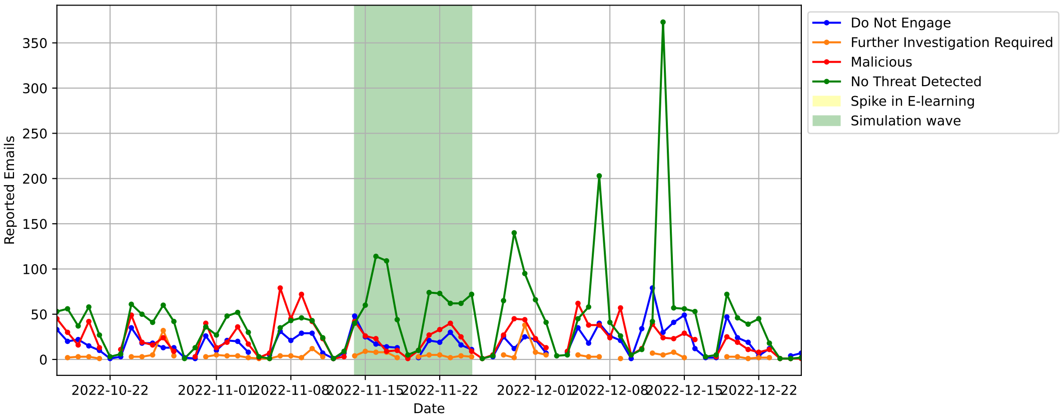


Figure 4.4: The development of the reported emails where the simulation emails are taken from the data. The counts are portrayed per day. There has been zoomed in to the four weeks before and after the simulation.

Pearson's correlation coefficient

Pearson's correlation coefficient between benign and reported emails with the count per week is 0.247. This indicates no or a weak correlation. The same factor with the data aggregated per day is 0.465, which indicates a moderate correlation between the reports of benign and malicious emails.

P-Value

The significance of the changes in the reporting counts after the simulation wave is calculated, with the p-values shown in Table 4.1. The p-values are calculated for the aggregation per week. Besides the entire period after the simulation, the p-values are also calculated for a shorter period until December 31.

Table 4.1: The p-value of the evaluation of the count in reported emails per classification before and after the simulation.

Classification	P-value week	P-value week short
No Threat Detected	0.039	0.032
Malicious	0.097	0.159
Simulation	0.13	0.16
Do Not Engage	0.00054	0.06
Further Investigation Required	0.47	0.42

The p-values for the period before December 31, the short period, only the value for *No Threat Detected* is below 0.05, indicating a significant value to reject the hypothesis that the simulation does not impact the number of reported emails.

4.2.2. Unique reports

To understand the development of the reported emails, other metrics besides the total number of reported emails can provide insight. We use the number of unique reported emails over time to explain outliers present in the data. The unique reported emails per week show that the number of unique emails fluctuates less than the total number of reported emails in Figure 4.5. With this, an email is unique if it has not been reported in two days of another email with the same title. Alternatively, the global unique emails show the unique titles throughout the observed period. These globally unique emails show the same pattern. The spikes in the reported emails can be explained by one or some specific emails that more employees have reported. This situation can occur only if the emails are sent to numerous employees, which could happen in the case of a harmless mass email or a phishing attack aimed at multiple employees.

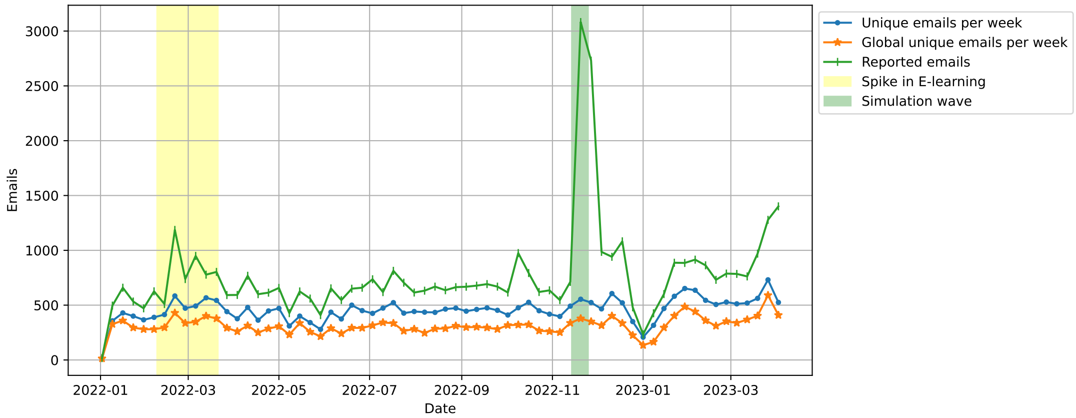


Figure 4.5: The unique reported emails over time. Additionally, the total email count and the global unique emails are shown. An email is globally unique if the title has not been reported in the data before. An email is unique if it has not been reported within two days of another email with the same title.

From these unique emails, the classification is known, as the emails with the same title have been given the same classification throughout the dataset. Looking at the unique emails per classification in Figure 4.6, the changes in reported emails are distinctly different than those seen in Figure 4.3. There is a peak in unique benign reported emails during the E-learning. The training seems to correlate with an increase in benign reported emails. The unique reported malicious emails peak in July 2022 and February and March 2023.

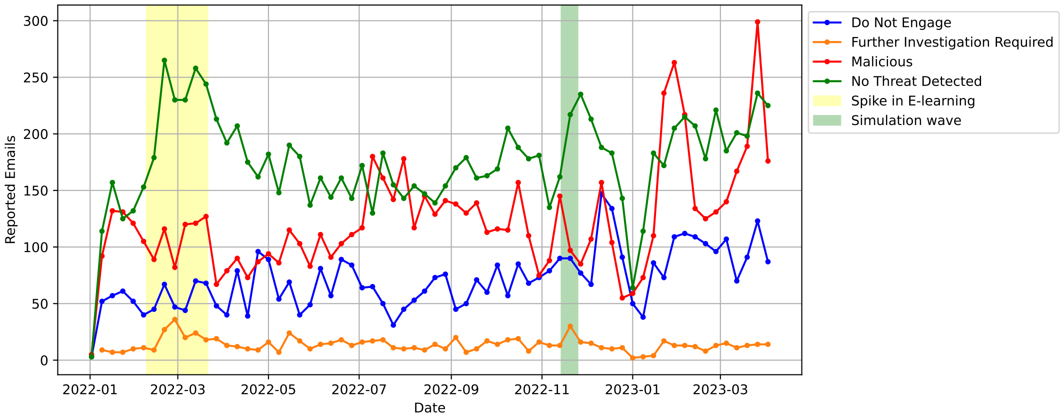


Figure 4.6: The unique reported emails per classification.

Suspicious email

It is possible to analyse falsely reported benign emails by looking at the false positives in the data. We can observe a peak in the total reported emails around mid-December by examining both Figure 4.3 and Figure 4.6. This peak can be attributed to a specific benign email reported by multiple employees. Upon closer examination of this period, it becomes clear why this happened. One benign email has been reported 330 times, accounting for over half of the benign emails reported this week. This specific email was very suspicious, asking employees to change their expired passwords and giving them 60 minutes to do so. Although benign reports are not the intended response from employees, the report of this benign email is primarily due to the poor insight of the sender and the resulting suspicious content.

4.2.3. Unique reporters

Understanding the reports over time continues with understanding the employees who report emails. Every week, around 500 unique employees report an email, as seen in Figure 4.7. During the simulation wave, there is an apparent increase in unique reports. The changes in the number of unique employees who reported an email follow a similar pattern as the changes in reports, more closely than unique emails. Reports increase are explained by multiple employees reporting the same email(s). During E-learning, there is also a peak in unique reporters.

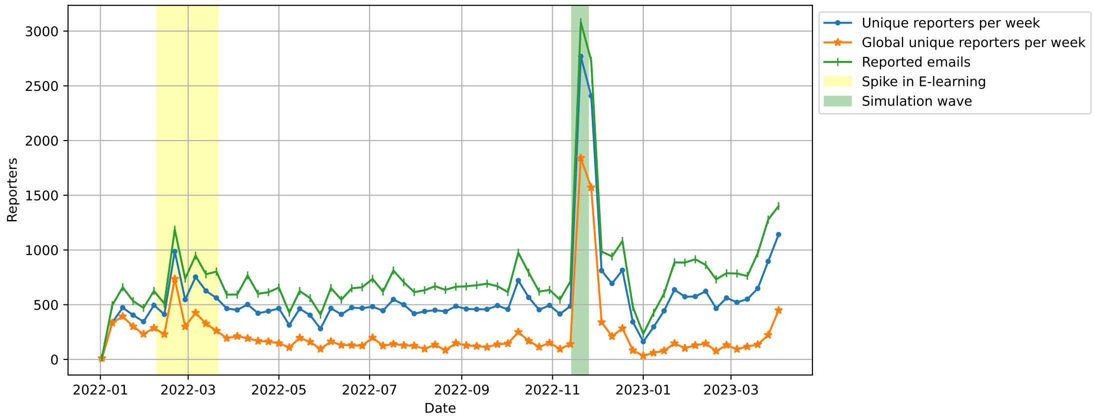


Figure 4.7: The unique reporters aggregated per week. Blue shows the unique reporters per week, whereas orange shows the addition of new reporters from the beginning per week.

4.2.4. External factors

As explained in Section 2.9.1, external factors can influence the employees' actions to report. The events over time that can be analysed can be found in Figure 4.8. Throughout the year, newsletters are sent to employees. Often they contain topics related to phishing emails to keep employees engaged.

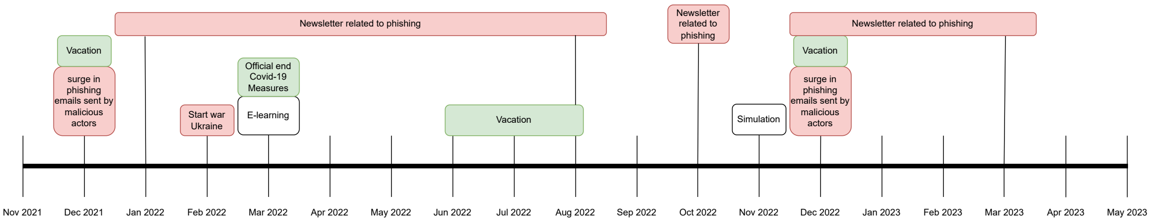

Newsletter related to phishing
Newsletter related to phishing
Newsletter related to phishing
Vacation
surge in phishing emails sent by malicious actors
Start war Ukraine
Official end Covid-19 Measures
E-learning
Vacation
Simulation
Vacation
surge in phishing emails sent by malicious actors
Nov 2021
Dec 2021
Jan 2022
Feb 2022
Mar 2022
Apr 2022
May 2022
Jun 2022
Jul 2022
Aug 2022
Sep 2022
Oct 2022
Nov 2022
Dec 2022
Jan 2023
Feb 2023
Mar 2023
Apr 2023
May 2023

Figure 4.8: A timeline of events or circumstances that can influence the reporting behaviour of employees. A red box indicates an expected higher reporting rate. A green box indicates an expected lower reporting rate. The E-learning and the simulation wave are represented in white boxes as these are under consideration during the research.

4.3. Text Comparisons

The analysis of the text similarities between several subgroups answers Sub-Question 1.2. Comparing the email content gives us an understanding of the emails employees find suspicious.

4.3.1. Topic Modelling

From the topic analysis between 'No Threat Detected' emails and 'Malicious' emails, the topics as shown in Table 4.2 can be identified. If the company's name was present in the email, this name is replaced with "companyname". Only English emails are used in generating these results. However, either the language detection does not work optimally, or emails are written in multiple languages, as there are several topics with non-English words. Not all discovered topics allow for a straightforward interpretation.

Table 4.2: Topics of malicious and no threat detected emails for the English emails in the dataset.

Classification	Topic Words
No Threat Detected	e, mail, sie, und, diese, die, der, oder, von, egencia
	de, je, en, van, companyname, password, le, e, het, vous
	0, style, font, table, td, color, size, tr, align, width
	companyname, 1, 2022, 2, 2023, message, e, 10, 0, 15
	information, data, customer, may, personal, business, privacy, lusha, contact, email
	unsubscribe, click, risk, financial, security, 00, 2022, u, 2023, management
	event, digital, 2023, business, amp, industry, innovation, register, head, global
	intended, email, message, please, communication, companyname, sender, recipient, use, information
	role, action, need, participant, true, rsvp, cn, attendee, partstat, req
	please, email, team, companyname, time, service, like, new, u, would
Malicious	management, 19, issue, unit, abidah, botannia, parc, take, time, 35
	com, message, 2022, outlook, 1, companyname, 2, email, 0, rule
	companyname, de, information, message, mail, e, attachment, belgium, intended, click
	password, email, keep, 2023, today, event, help, video, mar, right
	e, n, r, p, de, l, c, u, 2, w
	font, 0, style, td, width, tr, text, table, border, p
	please, sent, email, 2022, subject, thank, attached, document, hi, pm
	state, ukraine, united, court, 1, shall, president, london, bank, office
	fund, bank, email, u, mr, contact, name, ox, money, address
	intended, companyname, message, please, communication, email, mail, recipient, use, sender

For both classifications, several words are frequently used in the topic. Notable to see is the presence of Ukraine in one of the topics for malicious emails. Attackers use current affairs to send phishing emails, but this event, relatively unrelated to the employees' tasks, does get picked up as potential phishing.

There are some similarities between the topics. The presence of **please** in both a topic for malicious and two for benign indicates a raised suspicion for emails that asks an employee to act. The same holds for the word **password**. Furthermore, there one of the topics for the benign emails is very similar to one of the malicious topics. They both mention **intended, recipient**. This topic can relate to the email footer the company uses to warn employees the email is intended for the original recipient. The presence of this topic in malicious emails can be because attackers use this banner in their emails or because while reporting an email, the banner is added to the report.

The same analysis is done between the business roles in Table 4.3. All classified emails are considered here; no distinction is made between benign or malicious reports. The word **please** is present in at least one topic for each business line, as well as the company name. Again, several non-English words are present in the topics.

Table 4.3: The topics of emails per business role for the English emails in the dataset.

Business line	Topic Words
A	e, mail, sie, please, diese, und, die, parcel, information, der
	companyname, intended, communication, sender, please, message, use, email, recipient, error
	information, data, may, business, email, customer, u, microsoft, personal, privacy
	please, account, email, click, password, service, new, team, companyname, update
	20, utm, please, loan, email, intended, office, colleague, expense, line
	need, role, action, card, team, true, gift, rsvp, attendee, participant
	de, r, n, e, en, la, le, l, companyname, je

Business line	Topic Words
	e, 0, 1, ox, 6150a913c6, id, 2022, meeting, c, com
	email, please, sent, 2023, message, companyname, day, received, 2022, regard
	digital, event, risk, business, best, unsubscribe, 2, management, manager, financial
B	attendee, need, role, participant, cn, action, rsvp, true, partstat, req
	microsoft, account, join, team, device, meeting, file, security, recent, see
	email, u, risk, please, company, business, new, unsubscribe, event, financial
	companyname, de, information, mail, message, e, attachment, belgium, click, intended
	password, de, reset, companyname, e, mycompliance, button, r, n, l
	com, 7c, n, 3d, 0, x, 1, 2, 7c0, microsoft
	van, de, en, te, een, op, het, u, voor, met
	email, please, information, account, message, companyname, 2022, u, sent, contact
	order, email, please, mail, message, intended, e, sent, recipient, way
	font, style, 0, td, tr, p, width, size, color, text
C	need, action, role, participant, cn, attendee, true, rsvp, partstat, req
	letter, take, ukraine, bmg, war, russian, russia, companyname, legal, merger
	message, 2022, email, companyname, 1, sent, com, subject, 2, pm
	account, email, please, file, password, microsoft, bank, sender, standard, chartered
	update, click, device, install, please, mail, keep, information, companyname, blocked
	energy, egypt, summit, storage, hydrogen, r, clear, cache, e, renewable
	email, please, u, company, would, new, business, event, time, unsubscribe
	0, nbsp, style, td, border, table, tr, font, width, join
	external, state, 1, ukraine, united, employee, court, yunlin, president, office
	de, r, e, en, je, sie, die, f, und, van
D	account, microsoft, security, verify, join, team, office, recent, device, meeting
	executive, gen, global, event, 2023, e, network, venue, n, g
	data, email, security, time, business, 20, risk, management, company, team
	companyname, de, please, e, password, link, w, je, regard, mail
	onedrive, icon, microsoft, sharepoint, corporation, action, 2019, help, support, code
	0, style, font, table, td, tr, width, align, padding, border
	de, n, com, romero, e, v, microsoft, bank, en, mr
	information, survey, file, data, day, questionnaire, please, companyname, lusha, 3
	restaurant, group, information, email, mbl, cid, click, please, worsley, investigation
	email, please, message, companyname, 2022, information, mail, financial, e, view
E	client, industry, financial, risk, onboarding, leader, expert, report, kyc, cefpro
	de, n, record, companyname, le, r, 7c, vous, mail, com
	geico, company, information, insurance, message, mail, e, nhsmail, intended, secure
	kohl, sender, image, removed, specified, error, filename, u, shop, e
	com, message, 2022, email, u, 1, outlook, 2, change, rule
	email, information, data, business, paulo, utm, 20, personal, service, risk
	email, please, password, account, u, link, 2022, companyname, contact, using
	email, risk, business, need, right, time, employee, webinar, 2023, event
	de, van, je, support, request, 1, c, document, u, companyname
	bank, 1, 3, investment, fund, please, financial, payment, 4, 2
F	information, account, personal, lusha, data, verify, customer, may, email, business
	file, deleted, microsoft, onedrive, account, number, restore, bank, day, 0
	je, mail, companyname, e, de, data, please, email, information, op
	please, email, data, business, team, successful, new, would, like, regard
	message, com, email, 2022, 2, companyname, outlook, 1, 8
	account, microsoft, join, recent, security, team, device, password, activity, meeting
	de, e, je, n, en, te, mail, companyname, la, le
	e, new, please, intended, cisco, recipient, email, message, information, mail
	click, unsubscribe, email, companyname, 2021, 15, please, request, 11, sent
	email, u, security, please, event, view, 00, contact, register, 2022

4.3.2. Similarity Scores

With the preprocessed emails, we determine the similarity score between the emails and the simulation emails. In Figure 4.9, the similarity scores between the ten simulation emails can be found. As required, all emails have a score of 1 with themselves. In the figure, emails 2 and 7 have the highest similarity score. A closer examination of these two emails shows they discuss partially the same topic and mention a specific service. A score of 0.24 can thus already cover a similar topic between emails. However, both emails differ in length, likely attributing to a lower similarity score.

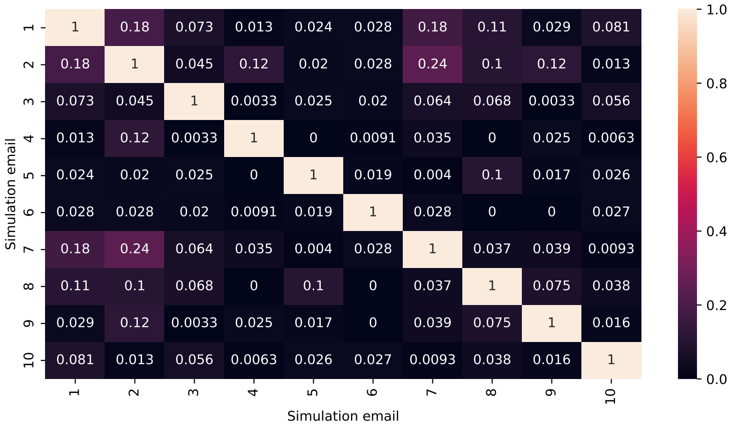


Figure 4.9: Similarity matrix of the scores between the ten simulation emails.

We determine the same similarity score for each email in the data, which is not part of the simulation. Each email is compared to the ten simulation emails in the wave in November 2022. The results are shown in a boxplot in Figure 4.10. Several outliers can be seen. The outlier in the 1st similarity score is due to a test with the simulation emails. These are not classified as simulations while they should be, which is also the reason for the outlier in score 10. Thus, these outliers are similar to the simulation emails but still get a score below 0.5.

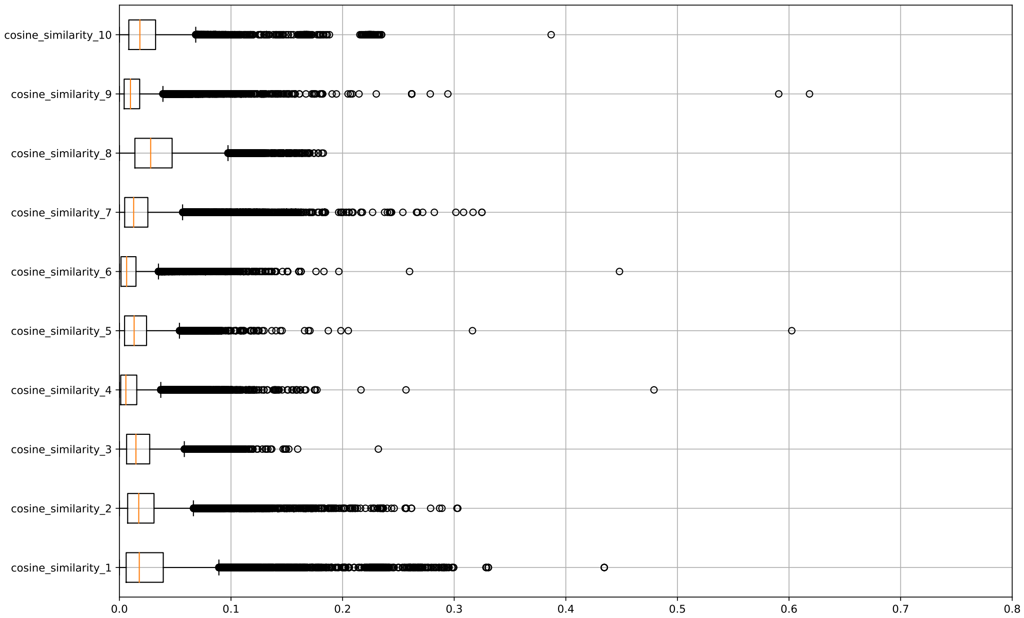
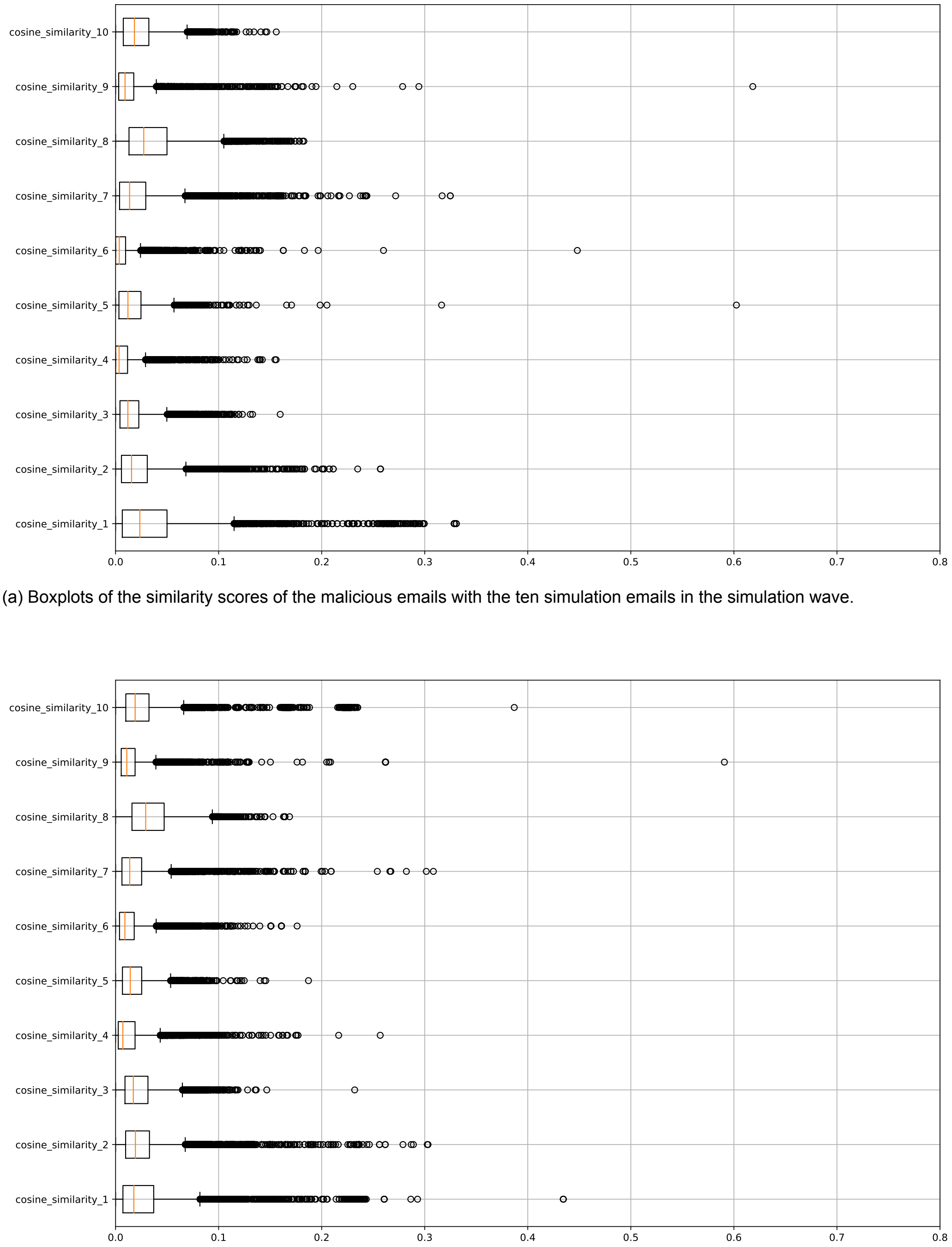


Figure 4.10: Boxplots of the similarity scores of all emails not classified as simulation with the ten simulation emails in the simulation wave of November 2022.

The outlier for score 9 in Figure 4.10 can be explained by the fact that simulation email nine was based on a previously reported email. For similarity score 4, there is no obvious external explanation for the higher similarity score other than that the content of the emails is just very similar.

The similarity scores are also analysed separately for the benign and malicious emails in Figure 4.11. The scores do not differ significantly between the two classifications. Additionally, both classifications do not contain emails with a score above 0.2, except for some outliers.



(a) Boxplots of the similarity scores of the malicious emails with the ten simulation emails in the simulation wave.

(b) Boxplots of the similarity scores of the benign emails with the ten simulation emails in the simulation wave.

Figure 4.11: Boxplots of the similarity scores of the malicious and benign emails with the ten simulation emails in the simulation wave.

The similarity scores can be further investigated by identifying the behaviour of the scores over time. In Figure 4.12, the similarity score of the first email with all emails can be found. The increase in the value of the score in November is because the simulation wave was held during this time. The high scores are, therefore, of the simulation email itself. The boxplots for each similarity score over time with the benign and malicious emails can be found in Appendix C, Figure C.2.

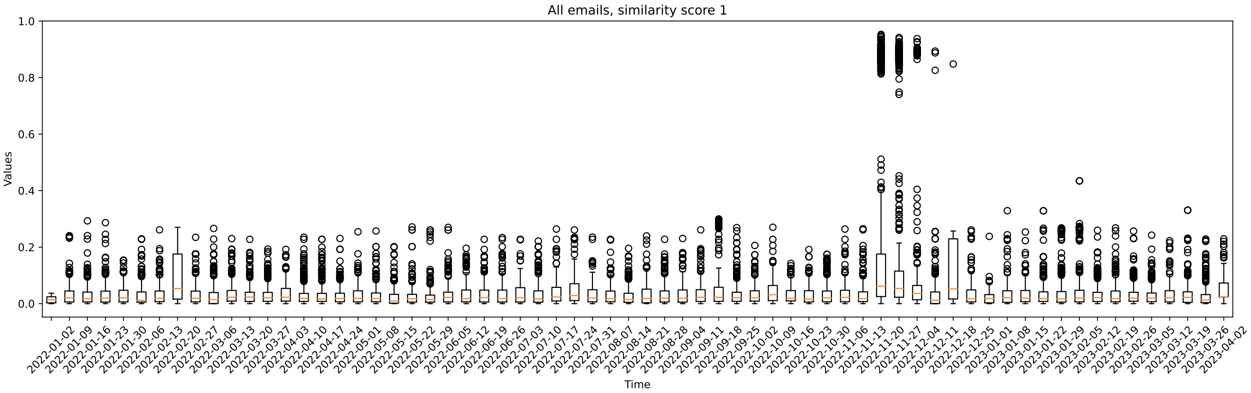


Figure 4.12: Boxplots over time of the similarity scores of all emails with the 1st simulation emails in the simulation wave.

4.3.3. Component comparison

Another comparison can be made on the components present in the emails. The comparison between the simulation, malicious and benign emails is provided in Table 4.4. For security reasons, the components are not specified. All components are based on previous research described in Section 2.5. Components 10 to 12 look into the presence of a topic in the emails. As there are only ten emails in the simulation wave, the percentages of this column are multiples of 10.

Table 4.4: Percentage of emails where the component is present in the body of the text. The last column specifies whether specific attention is spent on this component during the E-learning.

Component	Benign	Malicious	Simulation all	Simulation wave	Focus in E-learning
1.	81 %	59 %	100 %	100 %	No
2.	93 %	65 %	100 %	100 %	Yes
3.	59 %	69 %	72 %	100 %	Yes
4.	39 %	11 %	51 %	30 %	No
5.	75 %	86 %	76 %	100 %	Yes
6.	30 %	18 %	54 %	60 %	Yes
7.	71 %	49 %	68 %	70 %	Yes
8.	46 %	19 %	37 %	30 %	No
9.	60 %	40 %	32 %	30 %	No
10.	9 %	6 %	13 %	20 %	Yes
11.	14 %	10 %	9 %	20 %	No
12.	53 %	21 %	55 %	60 %	No

Some remarkable occasions can be observed in Table 4.4. Several components are present in all simulation emails during the wave in November, while the reported emails contain lower percentages of these components.

Component 12 refers to words associated with a topic. Benign emails frequently mention this topic, while malicious emails have a lower frequency of these words.

4.4. Interviews

The interviews are analysed to incorporate the company's context and provide additional explanations for the results. In Appendix B, the questions for the interview can be found along with the summary of the three conducted interviews. The participants have approved the content of the summaries. From the interviews, several points stand out. There is overlap and disagreement between the participants and between the participants and the literature. Some insights are uncovered which could aid the company specifically but are not generalisable.

4.4.1. Background

The employees' background varies, where one of the participants interacts with a lot of external email traffic, while this is not the case for the other two participants. The relation between their job and security practices also varies between the participants.

4.4.2. Personal behaviour

None of the three participants personally dealt with phishing attacks aimed at them. This limits our conclusions from their personal behaviour when interacting with phishing emails. They do have experience with reporting emails and have all reported suspicious emails in the past.

4.4.3. Company situation

All participants agree the progress made by the company regarding the phishing culture has improved in the last few years. The combination of different types of training stand out and have their benefits. They are familiar with the reporting button and have used it in the past year.

Additionally, all three participants mention the perceived increased risk for new employees to be targeted. They experience that these people need more extensive training to get their security knowledge up to date.

The first participant mentioned they would personally benefit from putting the topic of cybersecurity on the agenda of general department meetings. From the third participant, we learn this is already the case in the department this employee works. This shows not all security measures are implemented throughout the company, and there are differences in the strategies and procedures. The company could benefit from implementing the tested and proven strategies of specific departments company-wide.

4.4.4. Training

The second and third participants mentioned their team members approach them if they have questions about specific emails or actions related to reporting phishing emails. For the participants of the interviews, this indicates their team members feel their managers are approachable in cybersecurity-related topics.

4.4.5. Improvements

One last noticeable point during the interviews was the perceived threat of macros in email attachments. One of the participants specifically mentioned this could be an additional aspect explicitly addressed in the simulation emails. However, overall, the interviewees are content with the current practices of the company.

5 Discussion and Recommendations

5.1. Discussion

This study aims to explore which metrics can be used to understand the behaviour of employees surrounding phishing emails. With the exploratory approach, the numerous results show the possibilities of the chosen methods. To delve into the results, we will examine the statistics of the dataset and, combined with the outcomes, address the sub-questions. Consequently, the Figures are analysed in more depth, after which several conclusions are presented to answer the remaining questions. The main research question is answered, after which several recommendations are made to the company for implementing the findings.

5.1.1. Measurable components

From the descriptive statistics in Section 3.2.2, the percentage of 34.8% of emails classified as *No Threat Detected* compared to 39.4% of *Malicious* and *Do Not Engage* combined is striking. Employees report almost as many benign emails as correctly reported emails. Assuming employees only report emails if they suspect they are malicious suggests that they have difficulty distinguishing benign from malicious emails and are ineffective in classifying emails or that many benign emails contain suspicious attributes similar to malicious emails, making the classification difficult.

Further examination of several individual emails and peaks in benign reports suggest several benign emails have many components pointing to phishing. In Table 4.4, it can be seen that the reported benign emails contain components addressed in E-learning, and thereby the answer to Sub-Question 1.2 is that several components employees are aware could indicate phishing are also present in benign emails. Employees receive feedback on the emails they reported and their true classification, indicating if an email was reported correctly. If reported emails keep returning as benign, this can confuse the employees and signal their actions are incorrect while they apply the training correctly. Adapting the benign emails to include fewer components can aid employees in evaluating the emails and keep the learned actions from the training intact.

This finding adds to the research of Steves et al. (2020), which argues that the difficulty of an email determines the click rate but lacks real reported emails to complete the analysis. The difficulty of an email, not the classification, also seems to determine the actual reporting rate. While the scale by Steves et al. (2020) looks at both the cues of an email and the premise alignment, this research solely analyses the clues present in the email, which is already an indication of the difficulty.

5.1.2. Trends over time

The answer to Sub-Question 1.3 consists of several parts based on the development of the reported emails.

Benign reports

The results of the total reported emails per classification in Figure 4.3 show no noticeable change in the reporting behaviour of employees before and after the phishing simulation. However, looking at the p-values for the benign reports in Table 4.1, the values are below 0.05 and thus suggest the simulation has a significant effect on the reported benign emails. However, there is an apparent decrease in reports at the end of 2022, likely caused by the absence of employees during the holidays in this period, as shown in Figure 4.8. Furthermore, four weeks after the simulation, there was an apparent increase in the benign reports, with a count above 350 in a day in Figure 4.4 while the counts were generally around 50 before the simulation. As outlined in Figure 4.2.2, this high count was caused by the report of one distinct email, and the reporting numbers are thus likely not caused by the simulation.

To conclude, the in- and decrease of the reported benign emails after the simulation can be explained by factors other than the simulation. With this finding, the benefit of using actual reported emails is clear. This agrees with Hillman et al. (2023), who state that by researching interactions with real phishing emails in real-world situations, the organisational aspects of susceptibility can be investigated instead of solely the individual's susceptibility.

Correlation between benign and malicious

The Pearson correlation coefficient of 0.247 for aggregation per week and 0.465 for aggregation per day indicates a weak to moderate correlation between the number of malicious and benign reported emails over time. The increase in correlation from aggregation per week to aggregation per day can be explained by the decrease in reports during the weekend for all classifications. The absence of a correlation between the malicious and benign reports over time suggests different reasons cause the fluctuations in reporting numbers.

Unique reporters

The relation between the number of unique reporters and the total reported emails in Figure 4.7 provides insight into the company's security behaviour (K. Greene et al., 2018). During the three spikes in simulation waves, in February 2022, November 2022 and March 2023, as seen in Figure 4.2, there are noticeable peaks in the number of unique reporters. This shows that the simulation emails manage to reach employees who do not report other emails during the year, as seen by the peaks in the orange line in Figure 4.7. As these peaks appear during all three simulation waves, they are likely an effect of the simulations.

The unique reporters provide a noticeable trend, which partially answers Sub-Question 1.3. From the progress of unique reporters, we see the simulation wave reaches people who have not reported an email in the period before the simulation. Assuming these employees get real phishing emails that they do not report but report simulation emails, this could mean the simulation emails are not similar to the phishing emails. Alternatively, employees discuss the simulation emails more because they all get the same emails, resulting in different behaviour from an employee than they would have if there were malicious emails.

The development over time of the unique reporters also shows how well simulation emails reach employees and whether these employees are familiar with reporting emails before the simulation. This is a new perspective towards the way simulation waves can be used. Where research such as (Lain et al., 2022) look at the reporting rates separately, combining these rates with the actual reports shows a broader picture.

Suspicious email

From the single suspicious email, as described in Section 4.2.2 and the resulting peak in reports, there is an indication that peaks in the reported emails can indicate the presence of an email campaign, primarily benign. From the simulation wave in November 2022, the reporting rate as the percentage of the total sent emails is known. This is around 10%, similar to previous simulation waves at the company before 2022. If this is similar to the reporting rate of real malicious emails and email campaigns, many missed phishing emails are not being reported. A spike in # reporters, which could indicate the presence of a campaign, could be used as a trigger to notify all employees of the ongoing campaign, depending on the threat in the email. This finding aligns with Lain et al. (2022), who found that employees can function as a collective phishing detection mechanism and detect new phishing campaigns quickly.

Unique reports
In Figure 4.5, it can be seen that the unique reported emails over time are relatively consistent and do not have apparent peaks or drops correlated with the total number of reported emails. Only in May and June 2022, a noticeable decrease goes below 400 unique emails. This could indicate the holidays during this period, but other explanations are possible such as a decrease in the sent phishing emails by attackers in June according to CISCO (2022).

Security Culture
The relation between unique reporters and total reported emails can expose the influence of email campaigns, leading to better detection of said campaigns. Furthermore, the number of unique reporters can indicate the security culture of a bank better than the total number of reports. Numerous unique reporters suggest a better security culture than a few employees who report a lot. This difference cannot be uncovered by solely looking at the total number of reported emails.

External Factors
In Figure 4.8, several identified factors can be found. The newsletter sent out by the company every month often contains topics related to phishing. As there are few fluctuations in these topics, fluctuations in effects are likely small. On the contrary, the vacation period does explain decreases in reports as shown in Figure 4.3, especially during the vacation period in December.

From the research into external factors that can influence reporting behaviour, previous research shows an increase in phishing emails during some months of the year, mainly in December. The expectation would be this would be visible in the reports of malicious emails, but remarkably this does not seem to be the case. This could mean that employees miss more emails during this period resulting in a higher chance of success for attackers, or employees spend less time reporting. Alternatively, there is a delay in reports for this period, which would explain the increase in reported emails in February.

5.1.3. Text similarities

The trends over time give an overview of the reporting behaviour of the employees. A more in-depth analysis of the content of the emails can provide an understanding of why the employees behave this way and how the content relates across different subsets. The answer to Sub-Question 1.4 can be found in the topic modelling, text analysis and component comparison.

Topic Modelling
In Table 4.2, the results from the topic analysis show the topics for the Benign and Malicious emails. As noted, there are several topics in the benign emails containing words not in English, which is not true for malicious emails. This is because the company writes company-wide emails in multiple languages in one email, confusing the language detection method. These emails are classified as English, as the email is partly English, and the method can classify ambiguous text incorrectly (Danilk, 2014). It is worth noting that the topic modelling technique emphasises uncommon words. Additionally, since all the other emails are written in English, the non-English words may appear more frequently than expected in the generated results.

The email bodies were cleaned before the text analysis was performed, and all links and email addresses were taken out. However, besides non-English words, single letters and numbers are also present in the topics. For example, one of the topics for malicious emails is *e, n, r, p, de, l, c, u, 2, w*. This indicates there are still cluttering parts of the emails. Consequently, improving the cleaning process of the texts can lead to better results.

The results for the differences between the business lines in Table 4.3 do not provide obvious contrasts or similarities in the email topics across the business lines, answering part of Sub-Question 1.4. The company name is present for all subgroups in at least one of the topics.

When comparing the different business lines, the topics are challenging to interpret and match between the six groups, and there is no apparent reason for specific topics being present in a different subgroup. This can be explained by the chosen sizes of the comparing groups, which are too large and do not have distinctly different job descriptions. As the company is large, the selected departments do not necessarily reflect various tasks within the company well. The topic modelling does show what results this kind of analysis can provide. The method used can be a possible exploration strategy for future research, after which the found topics can be examined further.

The topics of the simulation emails cannot directly be linked to the topic modelling as presented in Table 4.2. Some of the words in the topics, such as password, coincide with some simulation emails, making a similarity likely. We understand why some employees can find specific topics suspicious from the topic comparison results. In the results of the malicious emails, the presence of current news items shows attackers use this approach. This can be used to set up a future counterfeit phishing campaign scenario. As the campaigns will be spread more throughout the year, the relevant topics to include can be varied based on the time of year to appeal to the current situation.

Similarity scores

Furthermore, Figure 4.10 show a slight variation in the similarity scores, but none have a high overall similarity. Some differences are visible in Figure 4.11 between the scores of benign and malicious emails. For example, simulation email 4 is more similar to benign than malicious emails.

A closer examination of the scores shows that a score of 0.24 between simulation emails 2 and 7 can already indicate the emails cover a topic similar to the human eye.

Component comparison

For Sub-Question 1.2, the components of the simulation emails are compared with the benign and malicious emails and the components discussed in the E-learning. As seen in Table 4.4, all simulations and most benign emails contain components 1 and 2, while these are less present in the reported malicious emails. It is not the case that employees do not take components 1 and 2 into account when determining if an email is suspicious, as this is refuted by research as presented in Section 2.5. The components were even specifically selected because people use them to determine the nature of an email. This indicates that employees can apply the training to emails containing components other than those discussed in the training and used in the simulation. Additionally, Table 4.4 shows the types of malicious attacks not covered by the simulations. As supported by Hillman et al. (2023), ideally, these cases would also be included in the simulation emails.

The connection of the content comparison with the E-learning also shows that components not discussed in the E-learning are generally less represented in the reported malicious emails, and the focus of the simulation emails seems to lie with the factors discussed in the E-learning. To answer Sub-Question 1.2, several components in the reported emails are underrepresented in the simulation emails. Relating these components to the metrics discussed in the E-learning links the two together.

As stated, some of the reported malicious emails differ from the E-learning, as they contain additional or miss components that are present in all simulation emails but are also not discussed during the E-learning. This could indicate that employees can extend their knowledge to other types of emails or, previous to the learning, already knew how to report.

5.1.4. Interviews

From the interviews, directions for improvements in anti-phishing tactics are found. There are tactics to combat phishing which are only applied in specific parts of the company. As others suggest these approaches would benefit their department, mapping the different options not used company-wide can be an easy approach to extend the measures taken. Departments or countries performing better on the identified metrics can be used as an example.

During the interviews, all three participants expressed their concerns regarding new employees. They considered this group particularly interesting regarding interactions with phishing emails. This statement suggests that the company should investigate the idea that new employees might be more vulnerable and base their approach on their findings. If they are indeed more susceptible, which would be in accordance with Kumaraguru et al. (2009) and Lain et al. (2022), there can be more attention to this group, providing a possible improvement for the company. If they are not, which would agree with Baillon et al. (2019), the mindset of employees who are with the company longer can be false, and they can feel too secure. In this case, this notion can be incorporated into the training. Based on the research conducted in this study and the relation between unique emails and the total number of reports, emails' content seems more important than an employee's experience. However, experience level could provide an alternative explanation to the conflicting results for the relation between age and susceptibility as uncovered by (Zhuo et al., 2022), and be a legitimate influencing factor.

5.1.5. New Metrics

According to Sanders (2014), finding accurate measures for cybersecurity is an ongoing challenge and research priority. Since this research came out, new metrics have been found, but there are still uncertainties when evaluating companies' cybersecurity. This research proposes new, additional metrics to assess the security culture and the effectiveness of phishing training.

- Unique reporters and global unique reporters over time.
    - Can be a proxy for the security culture of a company. Changes in unique reporters can explain in- or decreased reports and the impact of training throughout the company.
- Unique reported emails and the relation to the total reported emails.
    - A mismatch in peaks and dips between these two metrics can indicate the presence of email campaigns. A match can indicate there are more targeted individual (spear-) phishing emails reported.
- Presence of components in the emails
    - This metric can link training and reported emails and provide a rating of an email different from the classification of the difficulty of an email.
- Topic comparison
    - Topics can provide an understanding of phishing tactics employed by attackers. Additionally, if this metric is developed further, emails can be compared based on content.

As stated by Canfield et al. (2017), *"Any metric faces three challenges: (a) it must differentiate between users' ability and the technology in place to protect them; (b) it must account for the low base rate of phishing attacks; and (c) it must be able to extrapolate from the observed circumstances to those where users are faced with actual attacks."*. Using unique reporters as a metric for the security culture overcomes these three challenges. (a) Every employee has the same technology measures in place; differences in reports originate from the employee's ability. Over time and between organisations, not all metrics overcome this challenge. Therefore, caution is advised. (b) The unreported emails do not influence the metrics, and the low base rate is accounted for. (c) As the reported emails contain actual attacks, the reported emails provide insight into the behaviour when employees face actual attacks. Additionally, since the simulation emails are also part of the data, the numbers of the observed behaviour surrounding these simulation emails can be extrapolated to actual malicious emails when the simulation emails cover the possible scenarios of real attacks.

If the proposed metrics are widely used in conjunction with existing ones, they can effectively address the challenges presented in the paper. To use these metrics across organisations, the individual reports of employees have to be measured by the company, including their names or other ways to identify individual reporters. The same holds for the unique emails, as there should be a way to determine if two emails are identical. Besides comparison between organisations, the metrics can be evaluated over time within a company as well.

The challenges of these metrics include access to the necessary datasets, as previous research shows this is a limitation of several studies.

Several of the analysed metrics in this report do not provide clear insights in their current form. This mainly relates to the topic modelling and the similarity scores. However, they can be improved by spending more time preparing the data. Therefore, it is imperative to conduct additional research to thoroughly examine the potential of these techniques and enhance their results.

5.2. Recommendations

From the results and ensuing discussion, several recommendations can be constructed for the company to proceed with their anti-phishing approaches.

The benign reported emails contain several company- or department-wide emails. This indicates that the company should focus on improving its mass emails to reduce the number of benign reports. One of the identified metrics answering the Main Research Question is the components present in the emails. The senders need to remove the components raising suspicion from the benign emails that cause employees to feel suspicious about them. To accomplish this, composers of these emails must be aware of their email's impact on the employees and improve the texts with this effect in mind. This can decrease the ratio of benign reported emails and reduce the costs of reporting false positives. This recommendation is mainly addressed to the HR department within the company, as they are responsible for internal communication. The suspicious components can be identified by analysing the benign campaigns in more depth. This has to be done by the cybersecurity department, as they have the skills and access to perform these actions. Implementing this approach can reduce the number of false positives in the reports, leading to greater clarity for the employees.

From the answer to Sub-Question 1.1 in Table 4.4, it can be seen that component two is present in a high percentage of the benign reported emails, which may be the cause for suspicion among the employees. Minimising the use of this specific component in benign emails can be a suitable practice to adopt by the HR department and other entities that send numerous emails to many employees.

Currently, the emails used in the simulation wave do not represent all possible malicious emails, as found in answering Sub-Question 1.2. From the component comparison, the recommendation follows to extend the types of emails used in the simulation to include scenarios where components 1 and 2 are absent. If the simulation emails represent the possible attacks accurately, then the statistics of the simulation emails are more representative of the overall security behaviour in the company. This practice can help the CISO department evaluate the results of the simulations, as they are in charge of conducting these simulations. Additionally, one of the simulation emails was similar to a benign email, giving the wrong signal to employees. Only if there is an indication that attackers copy these specific benign emails, these simulation emails are in addition to the measurement.

From the results to answer Sub-Question 1.3, the decrease in reported emails in December 2022 and the lack of an increase in reports at the beginning of January suggest that more support is necessary in these months. Because many employees are absent during this period and have to deal with their full inbox when they get back, their workload is higher, and the low report rates indicate they spend less time on reporting emails, especially with the increase in phishing emails in December as found by CISCO (2022).

From the interviews, the need for closer contact for employees with an expert within their tribe or department arose. The company has already developed such roles in certain areas, but not company-wide. This need coincides with findings from the literature, as Tally et al. (2023) recommends that employees benefit from a one-to-one connection with a cybersecurity-knowledgeable person. The size of the company limits the opportunities for one-on-one contact, but a step in the right direction could be to extend the existing concept, as discussed in the second interview, in a broader approach.

The current E-learning training and simulation emails do not seem to have the desired effect of increasing the reported emails. Therefore, with these recommendations, we urge the shareholders to invest in additional anti-phishing approaches and take the next step in protecting the company against attacks.

6 Limitations and Future research

6.1. Limitations

The conclusions that cannot be drawn from the analysis are also important. Several limitations can be identified in all aspects of the research. They indicate the boundaries of the possibilities of analysing reported emails with the proposed methods. As the research focuses on characterisation, several limitations can be addressed in future research with a more dedicated study.

6.1.1. Data collecting

The data's quality and quantity influence this study's specific findings. The data availability partly determines the possibilities in this research and the options at several stages in the research. For example, only factors that can be observed can be analysed for the component comparison of the emails. Furthermore, it would have been advantageous to include certain identified factors which were not observable and thus not included. As a consequence, several factors which influence the behaviour of employees are not analysed. To reduce the impact of this constraint, we identify the relevant factors in advance with literature research. As a result, some available elements were excluded as they were not deemed significant based on the literature. Using this method, the factors available in the data do not necessarily dictate the components for the comparison.

Another limitation of the data is the absence of False Negatives, the unreported malicious emails. The data only contains the True and False Positives, as identified in Table 1.1. Assumptions about the False Negatives based on this data are hard to validate. It is necessary to include a disclaimer with the recommendation to base simulation emails on the components in the reported emails, as anyone can fall for a phishing email, and there are likely emails that no employee reports (de Meulebroucke, 2021).

6.1.2. Trends

To determine unique emails over time, the titles of the emails have been cleaned and compared to group the identical titles. While this tactic is quite effective for benign emails, as campaigns are the same for each employee, the approach is likely less suitable for detecting malicious campaigns which actively try to avoid easy grouping (Ke et al., 2016). A more sophisticated approach to identifying unique emails can improve the analysis of these trends. This can also reduce the number of emails that were incorrectly grouped together.

6.1.3. External Factors

The identified external factors that can influence employees' reporting behaviour threaten the interpretability and significance of the results from the text comparisons. For example, a news topic can be excessively present in the topic analysis, while the topic is only present for a short time and does not represent a topic used throughout the year. To cope with external factors and minimise their impact, the analysis can be done over a more extended period to increase the understanding of the influence of external factors. Furthermore, although based on literature, the identified external factors are limited to the author's experience and interpretation.

6.1.4. Text analysis

Where humans can quickly see whether emails contain similar topics, this is more difficult with ML. Although language models can detect differences and similarities, interpreting the outcomes is difficult as semantic meaning is lost (Simha, 2021). The topic modelling and text comparison provide little insight into the differences between malicious and benign emails. Unfortunately, eXplainable Artificial Intelligence (XAI) techniques such as LIME or SHAP do not provide a solution to this problem. SHAP cannot easily be used as a global explainer for text analysis, and LIME only provides a local explanation which cannot be shared due to the confidentiality of the emails. More advanced NLP methods are required to increase the interpretability of the topic analysis. There is promise in the new techniques that are quickly developing. Conducting a dedicated study solely focused on topic analysis as an extension of the work from this research can take the analysis to a higher level.

An additional limitation of TF-IDF for topic modelling is that rare words in the combined text can be over-emphasised in the results for the topic modelling. This limitation explains the presence of non-English words in the topics. Because several emails are wrongly identified as English, these words are only present in incorrectly evaluated emails. Fortunately, as some topics contain the same words, there are enough topics to extract the most important general topics. Therefore, it is not the case that important topics are overlooked. However, if this method is to be extended, the topic modelling could benefit from a threshold for the word count in the entire dataset before the topics are considered in the final results.

6.1.5. Interviews

The participants of the interviews have been asked whether they would be interested in participating in this research with information about the topic of the interview. This can create selection bias, where employees with a preexisting interest in and attention to phishing are likelier to participate. The resulting conclusions from the interview could be limited to a specific subgroup of employees and departments with similar security cultures. This can result in thoughts and suggestions that may not represent the entire company accurately. To overcome this limitation, more interviews can be conducted, or the conclusions from the interviews can be shared to ask whether other employees agree.

This study involved only three interviews, which is insufficient to obtain a comprehensive overview of the company. Therefore, the conclusions based on the interviews should be taken considering this limitation. More interviews or questionnaires are necessary to validate the shared opinions.

6.1.6. Recommendations

One of the recommendations is to base new simulation emails on reported malicious emails with components currently not represented in the simulation. Basing the simulation emails on the reported malicious emails ignores the unreported phishing emails, which can cause survivor bias. However, this strategy does offer assistance to workers who may not have reported the harmful email themselves, which is preferable to disregarding the reported malicious emails during the creation of the simulation emails. Additionally, the reported emails can show insight into the emails not caught by filters (K. Greene et al., 2018) and allow email providers to detect attacks early to inform other potential victims (Lain et al., 2022; Wang & Song, 2021).

As the phishing challenge is part of the larger wicked problem of cybersecurity in general, proposed solutions can easily lead to unexpected consequences (Sharkov, 2016). The proposed solutions have not been implemented and therefore have not been empirically validated. There are scientific and evidence-based arguments for the components, but an effective study of the solutions is still lacking and could provide an opportunity for future research. Therefore, a follow-up analysis of the behaviour is necessary to validate their effectiveness and mitigate possible unforeseen negative side effects.

Phishing developments are pointing towards an increase in spear-phishing and whaling (Bhardwaj et al., 2020). These personalised attacks are not given explicit attention during the research, and the analysis of the proposed metrics is currently more focused on campaigns. To broaden the metrics, it would be helpful to delve deeper into the differences between spear-phishing strategies and more extensive malicious campaigns. This would decrease the limitations of the proposed metrics.

6.2. Future Research

During this research, we explored various directions, which can be a start for future research. Within this research, there has been zoomed in on a small part of the phishing system. External influential factors have been tried to incorporate or account for, but a more extensive study could provide a more detailed interaction between external factors and reported emails. Previous research agrees, as there seems to be more promise in understanding situational effects than personal characteristics (Canfield & Fischhoff, 2018). Such a study could include more specific data from the company to account for changes in what we deemed external factors in this research. Thereby, the recommended focus is to create an insight into the regular email traffic of employees and incorporate their daily email behaviour.

Because the research only analyses the reported emails, which are known in many industries, there is the possibility to use the proposed metrics in other companies or sectors, adding to existing measures (Sas et al., 2021). Comparing the results of the trends over time can provide a better understanding of different security cultures and could separate the effect of external factors versus training exercises. If a company records the reported emails and their attributes, such as the email body and sender, the implementation of the analysis in this research is straightforward.

Diving into the differences between departments and creating subgroups based on tasks and daily workload can aid in understanding the struggles of individuals or groups of employees. This would provide input for more personalised recommendations for training, which can increase the efficiency of the training (Jampen et al., 2020). Compared to the focus on departments in this research, a step could be made in choosing suitable subgroups. A recommendation would be to organise sub-groups based on their primary task, considering factors such as interaction with external entities, on- or offline tasks, and responsibilities, as user context is found to be extremely important in phishing susceptibility (K. K. Greene et al., 2018). Dividing the departments on such characteristics creates the opportunity to add a factor of premise alignment to each email, with which the phish scale as determined by Steves et al. (2020) can be determined for each email. Additionally, the analysis of topics and text similarities between groups demonstrated in this research could be repeated and create an understanding of the strengths and challenges of roles in the company. Combining these results with a comprehensive understanding of a company's operations can expose directions for improvement in the security approach. Furthermore, if large differences are found, the anti-phishing training can be customised based on an employee's tasks.

To get a more accurate sense of the effect of the training, the entries of the training could be linked to individuals, making it possible to perform analysis on individual behaviour. This can create a more direct link between employees who finished the E-learning and whether changes occurred in their reporting behaviour compared to the period before and after the training and additional effects of the simulation.

With developments in the field of NLP, where ChatGPT can write an email in seconds, the phishing problem is bound to grow (Grbic & Dujlovic, 2023). Future research could look at the development of reported phishing emails over time and analyse whether the content of the emails changes. The analysed components from Table 4.4 can provide a structure for the interesting components and the development of their presence in phishing emails.

In the comparison between different types of emails, the focus lies on the text comparison and exact words, which can create results that are not easy to interpret for people. Future research can use the same approach but take the emotion of a text into account instead of the used words to compare the simulation, benign and malicious emails as emotion and persuasion principles are used to convince people to interact with an email (Lain et al., 2022). Alternatively, instead of emotion, the division of the emails into the persuasion principles proposed by Cialdini (Cialdini, 2009) can show which types of persuasion techniques evoke suspicion among employees (Wright et al., 2014).

For the company specifically, as its approach in the embedded phishing campaign is changing, the effects of these changes compared to the outcomes of this report are helpful to assess the validity of the choice to change its approach. The proposed metrics, as determined in this report, can show whether the changes contribute to a better security culture and improved reporting behaviour.

7 Conclusions

Phishing is the major cause of data breaches, and human interference is necessary to decrease the success rates of attackers. This research aimed to see how reported emails can be compared to simulation emails to answer the research question *How can email reporting patterns in a large organisation measure the relationship between phishing training, reported emails and employee behaviour?* This research offers valuable insights that can help enhance anti-phishing measures.

The answer to the main research question and the main result of this research is a set of metrics for evaluating a company's security culture through the analysis of email reports. Newfound metrics to quantify the security culture of a company include the unique reporters and the unique reported emails in combination with the total reported emails. The development of these metrics over time provides a broader indication of the reporting behaviour of employees. Spikes in reports can be explained by combining these two metrics. This is exemplified by the rates for the company, where we found that spikes in benign email reports, from 50 reported emails per day on average to 350 at the end of December, are the result of single emails sent to a multitude of employees. The specific findings assist in identifying areas of potential improvement in how emails are sent to employees within the company. The proposed metrics can be used in other companies and sectors as well. Additionally, the metrics can provide an improved monitoring environment as requested by Hillman et al. (2023) to measure the security culture of a company.

This research adds to the findings of Steves et al. (2020), who argue that phishing analysis should focus on the components of an email pointing to phishing instead of the total percentages. The unique email report(er)s combined with the total reports is an example of applying this practice.

The benign and malicious reports have different patterns, quantified with a low Pearson correlation coefficient of 0.247 if the counts are aggregated per week. This suggests that different factors influence their variances over time.

The methods to conduct this research include NLP to analyse the text of emails. The results show similarities between benign, malicious and simulation emails, and keywords such as **please** and **password** are present among all three groups of emails. However, the approach needs improvement and more advanced ML techniques to be interpretable. Numerous possibilities for improvement in the anti-phishing measures of a bank are proposed. One of the proposals is to improve its simulation emails by adding or removing components to address a wider range of potential phishing emails. This recommendation is found by analysing the components of emails rather than raw text.

The company is only better protected if the employees are part of the solution. Several interviews show how employees experience phishing training and the security culture. Providing anti-phishing training is seen as valuable by employees. Additionally, employees believe that new employees are more vulnerable to phishing emails. Experience level could provide an alternative explanation to the conflicting results for the relation between age and susceptibility as uncovered by (Zhuo et al., 2022).

Alternatively, this belief is unfounded and based solely on personal experience.

Future research should focus on applying the proposed and tested method to other companies and sectors to make comparisons and further validation of the proposed metrics possible. In this process, advancement can be made in analysing different business lines or departments to consider more components, such as premise alignment.

To keep employees in touch with the rapid developments in cybersecurity, companies provide the tools to navigate the complexity of dealing with phishing emails. This research provides the metrics to evaluate the security culture and make vital steps in the fight against phishing attacks.

Bibliography

Agarwal, S. (2013). Data mining: Data mining concepts and techniques. *2013 international conference on machine intelligence and research advancement*, 203–207.

Al-Alawi, A. I., & Al-Bassam, S. A. (2019). Assessing the factors of cybersecurity awareness in the banking sector. *Arab Gulf Journal of Scientific Research*, *37*(4), 17–32.

Al-Shanfari, I., Yassin, W., Abdullah, R. S., Al-Fahim, N. H., & Ismail, R. (2021). Introducing a novel integrated model for the adoption of information security awareness through control, prediction, motivation, and deterrence factors: A pilot study. *Journal of Theoretical & Applied Information Technology (JATIT)*.

Aneke, J., Ardito, C., & Desolda, G. (2019). Designing an intelligent user interface for preventing phishing attacks. *IFIP Conference on Human-Computer Interaction*, 97–106.

Arshad, A., Ur Rehman, A., Javaid, S., Ali, T., Sheikh, J., & Azeem, M. (2021). A systematic literature review on phishing and anti-phishing techniques.

Asghari, H., van Eeten, M., & Bauer, J. M. (2016). Economics of cybersecurity. *Handbook on the Economics of the Internet*.

Asmussen, C. B., & Møller, C. (2019). Smart literature review: A practical topic modelling approach to exploratory literature review. *Journal of Big Data*, *6*(1), 1–18.

Baillon, A., De Bruin, J., Emirmahmutoglu, A., Van De Veer, E., & Van Dijk, B. (2019). Informing, simulating experience, or both: A field experiment on phishing risks. *PloS one*, *14*(12), e0224216.

Bak, T. (2019). Python nlp libraries: Features, use cases, pros and cons. https://medium.com/@tomaszbak/python-nlp-libraries-features-use-cases-pros-and-cons-da36a0cc6adb

Bhardwaj, A., Sapra, V., Kumar, A., Kumar, N., & Arthi, S. (2020). Why is phishing still successful? *Computer Fraud & Security*, *9*, 15–19. https://doi.org/10.1016/S1361-3723(20)30098-1

Bird, S., Klein, E., & Loper, E. (2009). *Natural language processing with python.* O'Reilly Media Inc. https://www.nltk.org/book/

Bispham, M., Creese, S., Dutton, W. H., Esteve-González, P., & Goldsmith, M. (2022). An Exploratory Study of Cybersecurity in Working from Home: Problem or Enabler? *Journal of Information Policy*, *12*, 353–386.

Blei, D. M. (2012). Probabilistic topic models. *Communications of the ACM*, *55*(4), 77–84.

Blythe, M., Petrie, H., & Clark, J. A. (2011). F for fake: Four studies on how we fall for phish. *Proceedings of the SIGCHI conference on human factors in computing systems*, 3469–3478.

Canfield, C. I., Davis, A., Fischhoff, B., Forget, A., Pearman, S., & Thomas, J. (2017). Replication: Challenges in using data logs to validate phishing detection ability metrics. https://www.usenix.org/conference/soups2017/technical-sessions/presentation/canfield

Canfield, C. I., & Fischhoff, B. (2018). Setting priorities in behavioral interventions: An application to reducing phishing risk. *Risk Analysis*, *38*. https://doi.org/10.1111/risa.12917

Caputo, D. D., Pfleeger, S. L., Freeman, J. D., & Johnson, M. E. (2014). Going spear phishing: Exploring embedded training and awareness. *IEEE Security & Privacy*, *12*(1), 28–38. https://doi.org/10.1109/MSP.2013.106

Cavusoglu, H., Mishra, B., & Raghunathan, S. (2004). A model for evaluating it security investments. *Communications of the ACM*, *47*(7), 87–92.

Chakraborty, S., & Nisha, T. (2022). Next generation proactive cyber threat hunting-a complete framework. *AIP Conference Proceedings*, *2519*(1), 030093.

Chatfield, C. (1995). *Problem solving: A statistician's guide*. CRC Press.

Chowdhury, N., Katsikas, S., & Gkioulos, V. (2022). Modeling effective cybersecurity training frameworks: A delphi method-based study. *Computers & Security*, *113*, 102551. https://doi.org/10.1016/j.cose.2021.102551

Cialdini, R. B. (2009). *Influence: Science and practice* (Vol. 4). Pearson education Boston, MA.

Cimpanu, C. (2020). Phishing campaigns, from first to last victim, take 21h on average. https://www.zdnet.com/article/phishing-campaigns-from-first-to-last-victim-take-21h-on-average/

CISCO. (2022). *Cybersecurity threat trends: Phishing, crypto top the list* (tech. rep.). CISCO.

CR, A. (2020). Topic modeling using gensim-lda in python. https://medium.com/analytics-vidhya/topic-modeling-using-gensim-lda-in-python-48eaa2344920#:~:text=One%20approach%20to%20find%20optimum,'k'%20is%20too%20large.

Danilk, M. (2014). Langdetect 1.0.9. https://pypi.org/project/langdetect/

De Nederlandsche Bank N.V. (2022). Jaarverslag 2022, koers houden. https://www.dnb.nl/publicaties/publicaties-dnb/jaarverslag/jaarverslag-2022/

de Meulebroucke, A. V. (2021). https://phished.io/es/blog/an-end-to-pride-and-prejudice-everyone-is-susceptible-to-phishing?__geom=%E2%9C%AA

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). Bert: Pre-training of deep bidirectional transformers for language understanding. https://doi.org/10.48550/arXiv.1810.04805

Din, A., & Soare, B. (2023). Most common remote work security risks & best practices. *Heimdal*.

Dutch Government. (2023). Wet op het financieel toezicht bwbr0020368. https://wetten.overheid.nl/BWBR0020368/2023-01-01

European Banking Authority. (2018). Guidelines on security measures for operational and security risks under the psd2. https://www.eba.europa.eu/regulation-and-policy/payment-services-and-electronic-money/guidelines-on-security-measures-for-operational-and-security-risks-under-the-psd2

Feeley, A., Lee, M., Crowley, M., Feeley, I., Roopnarinesingh, R., Geraghty, S., Cosgrave, B., Sheehan, E., & Merghani, K. (2022). Under viral attack: An orthopaedic response to challenges faced by regional referral centres during a national cyber-attack. *The Surgeon*, *20*(5), 334–338.

Flyvbjerg, B. (2006). Five misunderstandings about case-study research. *Qualitative inquiry*, *12*(2), 219–245.

Foroughi, F., & Luksch, P. (2018). Observation measures to profile user security behaviour. *2018 International Conference on Cyber Security and Protection of Digital Services (Cyber Security)*, 1–6. https://doi.org/10.1109/CyberSecPODS.2018.8560686

Georgiadou, A., Michalitsi-Psarrou, A., & Askounis, D. (2022). Cyber-security culture assessment in academia: A covid-19 study: Applying a cyber-security culture framework to assess the academia's resilience and readiness. *ACM International Conference Proceeding Series*. https://doi.org/10.1145/3538969.3544467

Goel, S., Williams, K., & Dincelli, E. (2017). Got phished? internet security and human vulnerability. *Journal of the Association for Information Systems*, *18*, 2. https://doi.org/10.17705/1jais.00447

Gordon, L. A., & Loeb, M. P. (2002). The economics of information security investment. *ACM Transactions on Information and System Security (TISSEC)*, *5*(4), 438–457.

Grbic, D. V., & Dujlovic, I. Social engineering with chatgpt. In: 2023. https://doi.org/10.1109/INFOTEH57020.2023.10094141.

Greene, K., Steves, M., & Theofanos, M. (2018). No phishing beyond this point. *Computer*, *51*, 86–89. https://doi.org/10.1109/MC.2018.2701632

Greene, K. K., Steves, M., Theofanos, M. F., Kostick, J., et al. (2018). User context: An explanatory variable in phishing susceptibility. *in Proc. 2018 Workshop Usable Security*.

Griffiths, C. (2023). The latest 2023 phishing statistics. https://aag-it.com/the-latest-phishing-statistics/#:~:text=In%202021%2C%20the%20average%20click,12%25%20delivered%20malware

Halevi, T., Memon, N., & Nov, O. (2015). Spear-phishing in the wild: A real-world study of personality, phishing self-efficacy and vulnerability to spear-phishing attacks. *Phishing Self-Efficacy and Vulnerability to Spear-Phishing Attacks (January 2, 2015)*.

Hillman, D., Harel, Y., & Toch, E. (2023). Evaluating organiza-tional phishing awareness training on an enterprise scale. *Computers & Security*. https://doi.org/10.1016/j.cose.2023.103364

IBM. (2023). What is phishing? https://www.ibm.com/topics/phishing

ING. (2021). Ing becomes first bank in europe to launch training platform to combat cybercrime. https://newsroom.ing.be/ing-becomes-first-bank-in-europe-to-launch-training-platform-to-combat-cybercrime

ING group N.V. (2022). Annual report 2022. https://www.ing.com/Investor-relations/Financial-performance/Annual-reports.htm

Iuga, C., Nurse, J. R., & Erola, A. (2016). Baiting the hook: Factors impacting susceptibility to phishing attacks. *Human-centric Computing and Information Sciences*, *6*, 1–20.

Jagatic, T. N., Johnson, N. A., Jakobsson, M., & Menczer, F. (2007). Social phishing. *Communications of the ACM*, *50*(10), 94–100.

Jampen, D., Gür, G., Sutter, T., & Tellenbach, B. (2020). Don’t click: Towards an effective anti-phishing training. a comparative literature review. *Human-centric Computing and Information Sciences*, *10*. https://doi.org/10.1186/S13673-020-00237-7

Kaddoura, S., Chandrasekaran, G., Elena Popescu, D., & Duraisamy, J. H. (2022). A systematic literature review on spam content detection and classification. *PeerJ. Computer science*, *8*, e830. https://doi.org/10.7717/peerj-cs.830

Kam, H.-J., Mattson, T., & Goel, S. (2020). A cross industry study of institutional pressures on organizational effort to raise information security awareness. *Information Systems Frontiers*, *22*(5), 1241–1264.

Kamruzzaman, A., Thakur, K., Ismat, S., Ali, M. L., Huang, K., & Thakur, H. N. (2023). Social engineering incidents and preventions. https://doi.org/10.1109/CCWC57344.2023.10099202

Kanwal, K., Shi, W., Kontovas, C., Yang, Z., & Chang, C.-H. (2022). Maritime cybersecurity: Are onboard systems ready? *Maritime Policy & Management*, 1–19.

Kapadia, S. (2019). Topic modeling in python: Latent dirichlet allocation (lda). https://towardsdatascience.com/end-to-end-topic-modeling-in-python-latent-dirichlet-allocation-lda-35ce4ed6b3e0

Ke, L., Li, B., & Vorobeychik, Y. (2016). Behavioral experiments in email filter evasion. *Proceedings of the AAAI Conference on Artificial Intelligence*, *30*(1). https://doi.org/10.1609/aaai.v30i1.10061

Kelvas, D. (2023). Slam method: How to prevent hipaa email phishing attacks. https://www.hipaaexams.com/blog/slam-method#:~:text=The%20SLAM%20method%20is%20an,Link%2C%20Attachment%2C%20and%20Message.

Khusainov, P., Toliupa, S., Bakanov, V., & Shtanenko, S. (2022). Substantial formulation of the task of improving the information model of decision-making in the prompt (crisis) response to cyber incidents. *2022 IEEE 16th International Conference on Advanced Trends in Radioelectronics, Telecommunications and Computer Engineering (TCSET)*, 287–290.

Kuchumov, A., Pecheritsa, E., Chaikovskaya, A., & Maslova, E. Digitalization of economics: Modern financial technologies and their influence on economic security. In: 2021. https://doi.org/10.1145/3487757.3490866.

Kumaraguru, P., Cranshaw, J., Acquisti, A., Cranor, L., Hong, J., Blair, M. A., & Pham, T. (2009). School of phish: A real-world evaluation of anti-phishing training. *Proceedings of the 5th Symposium on Usable Privacy and Security*, 1–12.

Lain, D., Kostiainen, K., & Capkun, S. (2022). Phishing in organizations: Findings from a large-scale and long-term study. *Proceedings - IEEE Symposium on Security and Privacy*, *2022-May*, 842–859. https://doi.org/10.1109/SP46214.2022.9833766

Lev, A. (2010). Spammer tricks: Link shenanigans. https://www.computerworld.com/article/2468969/spammer-tricks--link-shenanigans.html

Marin, I., Allodi, L., Burda, P., & Zannone, N. (2023). The infuence of human factors on the intention to report phishing emails. *18*. https://doi.org/10.1145/3544548.3580985

Moore, T., Dynes, S., & Chang, F. R. (2016). Identifying how firms manage cybersecurity investment. *Workshop on the Economics of Information Security (WEIS)*, 1–27.

Morgan, P. L., Asquith, P. M., Bishop, L. M., Raywood-Burke, G., Wedgbury, A., & Jones, K. (2020). A new hope: Human-centric cybersecurity research embedded within organizations. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *12210 LNCS*, 206–216. https://doi.org/10.1007/978-3-030-50309-3_14

Nederlandse Vereniging van Banken. (2023). Phishing. https://veiligbankieren.nl/fraude/phishing/

Nikolopoulou, K. (2022). What is survivorship bias? | definition & examples. https://www.scribbr.com/research-bias/survivorship-bias/#:~:text=Survivorship%20bias%20occurs%20when%20researchers,a%20subset%20of%20the%20population.

Ortiz, E., Reinerman-Jones, L., & Matthews, G. (2016). Developing an insider threat training environment. In *Advances in human factors in cybersecurity* (pp. 267–277). Springer.

Osterman Research. (2019). The roi of security awareness training. https://www.infosecinstitute.com/wp-content/uploads/2021/03/IQ-Whitepaper-The-ROI-of-Security-Awareness-Training.pdf

Parsons, K., Mccormac, A., Pattinson, M., Butavicius, M., & Jerram, C. (2014). Using actions and intentions to evaluate categorical responses to phishing and genuine emails. *Proceedings of the 8th International Symposium on Human Aspects of Information Security and Assurance, HAISA 2014*, 30–41.

ProjectPro. (2023). 10 nlp techniques every data scientist should know. https://www.projectpro.io/article/10-nlp-techniques-every-data-scientist-should-know/415

Reason, J. (1990). *Human error*. Cambridge university press.

Reeves, A., Parsons, K., & Calic, D. (2020). Whose risk is it anyway: How do risk perception and organisational commitment affect employee information security awareness? *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *12210 LNCS*, 232–249. https://doi.org/10.1007/978-3-030-50309-3_16

Reinheimer, B., Aldag, L., Mayer, P., Mossano, M., Duezguen, R., Lofthouse, B., Von Landesberger, T., & Volkamer, M. (2020). An investigation of phishing awareness and education over time: When and how to best remind users. *Sixteenth Symposium on Usable Privacy and Security (SOUPS 2020)*, 259–284.

Rocha Flores, W., Holm, H., Svensson, G., & Ericsson, G. (2014). Using phishing experiments and scenario-based surveys to understand security behaviours in practice. *Information Management & Computer Security*, *22*(4), 393–406.

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*. https://doi.org/10.1109/ACCESS.2022.3183083

Sanders, W. H. (2014). Quantitative security metrics: Unattainable holy grail or a vital breakthrough within our reach? *IEEE Security & Privacy*, *12*(2), 67–69.

Sas, M., Hardyns, W., Van Nunen, K., Reniers, G., & Ponnet, K. (2021). Measuring the security culture in organizations: A systematic overview of existing tools. *Security Journal*, *34*, 340–357.

Schulze, H. (2020). *2020 phishing attack landscape.* (tech. rep.). Greathorn. https://info.greathorn.com/report-2020-phishing-attack-landscape/

Security, I. (2022). Cost of a data breach report. *Computer Fraud & Security*. https://doi.org/10.1016/S1361-3723(21)00082-8

Sharkov, G. (2016). From cybersecurity to collaborative resiliency. *Proceedings of the 2016 ACM workshop on automated decision making for active cyber defense*, 3–9.

Sheng, S., Magnien, B., Kumaraguru, P., Acquisti, A., Cranor, L. F., Hong, J., & Nunge, E. (2007). Anti-phishing phil: The design and evaluation of a game that teaches people not to fall for phish. *Proceedings of the 3rd symposium on Usable privacy and security*, 88–99.

Simha, A. (2021). https://www.capitalone.com/tech/machine-learning/understanding-tf-idf/

Slachtoffer Hulp Nederland. (2023). Wat is phishing? https://www.slachtofferhulp.nl/gebeurtenissen/fraude/phishing/

Steves, M., Greene, K., & Theofanos, M. (2020). Categorizing human phishing difficulty: A phish scale. *Journal of Cybersecurity*, *6*. https://doi.org/10.1093/CYBSEC/TYAA009

Sutter, T., Bozkir, A. S., Gehring, B., & Berlich, P. (2022). Avoiding the hook: Influential factors of phishing awareness training on click-rates and a data-driven approach to predict email difficulty perception. *IEEE Access*, *10*, 100540–100565. https://doi.org/10.1109/ACCESS.2022.3207272

Sveikauskas, D. (2021). Security by design principles according to owasp. *Patchstack, https://patchstack.com/articles/security-design-principles-owasp*.

Tally, A. C., Abbott, J., & Bochner, A. (2023). What mid-career professionals think, know, and feel about phishing: Opportunities for university it departments to beter empower employees in their anti-phishing decisions. *Proc. ACM Hum.-Comput. Interact*, *7*, 27. https://doi.org/10.1145/3579547

Thomas, S. (2023). Understanding the pearson correlation coefficient. https://articles.outlier.org/pearson-correlation-coefficient

Tulane University. (n.d.). Four reasons the cybersecurity field is rapidly growing. https://sopa.tulane.edu/blog/four-reasons-cybersecurity-field-rapidly-growing

Tyler, D., & Davey, J. (2020). The cyber coalface: Cognitive behaviour & phishing [Interview with Lance Wantenaar]. https://thecybercoalface.buzzsprout.com/

Vector, A. (n.d.). https://www.shutterstock.com/image-vector/login-into-account-email-envelope-fishing-702210526

Violino, B. (2023). Phishing attacks are increasing and getting more sophisticated. here's how to avoid them. https://www.cnbc.com/2023/01/07/phishing-attacks-are-increasing-and-getting-more-sophisticated.html

Vishwanath, A., Herath, T., Chen, R., Wang, J., & Rao, H. R. (2011). Why do people get phished? testing individual differences in phishing vulnerability within an integrated, information processing model [Cited By :244]. *Decision Support Systems*, *51*(3), 576–586. https://doi.org/10.1016/j.dss.2011.03.002

Wang, M., & Song, L. (2021). An incentive mechanism for reporting phishing e-mails based on the tripartite evolutionary game model. *Security and Communication Networks*, *2021*, 1–8.

Williams, E. J., Hinds, J., & Joinson, A. N. (2018). Exploring susceptibility to phishing in the workplace. *International Journal of Human-Computer Studies*, *120*, 1–13. https://doi.org/10.1016/j.ijhcs.2018.06.004

Wright, R. T., Jensen, M. L., Thatcher, J. B., Dinger, M., & Marett, K. (2014). Research note—influence techniques in phishing attacks: An examination of vulnerability and resistance. *Information systems research*, *25*(2), 385–400.

Xu, T., Singh, K., & Rajivan, P. (2023). Personalized persuasion: Quantifying susceptibility to information exploitation in spear-phishing attacks. *Applied Ergonomics*, *108*, 103908.

Yash, R. (2023). Python | lemmatization with nltk. https://www.geeksforgeeks.org/python-lemmatization-with-nltk/

Zhuo, S., Biddle, R., Koh, Y. S., Lottridge, D., & Russello, G. (2022). Sok: Human-centered phishing susceptibility. *ACM Transactions on Privacy and Security*. https://doi.org/10.1145/3575797

A

Data contents

Reported phishing emails

- IncidentType: Is the same for every datapoint, namely SEA (Suspicious Email Analysis)
- IncidentStatus: The status of the incident, for all emails this is ’Closed’
- IncidentID: ID of the incident, if an email has multiple indicators these IDs are the same for the different rows. So one reported email has one IncidentID, but can have multiple data points.
- Classification: The type of email, and thus the result of the analysis. Can be: Malicious, Simulation, No Threat Detected, Do Not Engage.
- ThreatType: The type of threat that is identified in the email. Can be: Link, Response, or Payload.
- subClassification: Specification of the classification
- EmailReportedBy: The email address of the person who reported an email as phishing
- Subject: The subject header of the email
- Sender: The email address of the person who send the supposed phishing email
- Reported: Date and time the email was reported
- Modified: Not sure what this entails. Format: Date and time.
- Resolved: Time at which the email was classified and closed by the company
- Age: Time between the reported and resolved moment. Format: time in seconds
- Message ID: ID given to the message. If an email has multiple indicators these IDs are the same for the different rows.
- Indicator Type: Type of indicator that helps to indicate the email is phishing [nan, URL, email address, payload]
- Indicator: In case an indicator is present, the indicator is given here.
- Campaign ID: If the email is marked as part of a campaign the id of the campaign is given here

Department specification

- Email
- Given name
- Country code

- Organisation
- Company
- Department
- pwdLastSet: date the password has last been reset
- Manager: Specifications of the manager of the employee
- DistinguishedName

Employee training: Mandatory e-learning

- Enroll time: time the employee was enrolled for the training and thus when someone could get started
- Start time: the time an employee started with the training
- End time: The time an employee finished the training
- Status: status of the training, whether someone finished the training
- Name/email

Counterfeit phishing emails
This analysis has already been made by the company

- Email body
- Title
- Click rates
- Compromise rates
- Reporting rates

B

Interviews

B.1. Questions

B.1.1. Background information

- What is your role within the company?
- What are your specific responsibilities in your current role?
- Could you describe the primary tasks your department is responsible for?
- How would you define a phishing email?
- In what ways does your job intersect with phishing emails, if any?
- In your normal email traffic, do you receive many emails with attachments?

B.1.2. Personal behaviour

- How frequently do you receive suspicious emails?
- What is your immediate response upon receiving a suspicious email?
- Are you familiar with the process of reporting a phishing email? Have you ever reported one? If not, why not?
- What additional tools or information would aid you in identifying phishing emails?
- What actions do you take when you receive a potential phishing email, and why?
- Are you aware of the correct point of contact for phishing-related queries?
- Are you familiar with the SDC helpdesk? Have you found the helpdesk to be responsive?
- Have you ever had a discussion about phishing within your department? If so, when?
- Do you find it challenging to distinguish between a phishing email and a legitimate work-related email? If so, could you describe the types of emails you find tricky?
- If you realised you had interacted with a phishing email and potentially compromised security, what do you believe the correct response would be? To whom would you report it?
- How confident do you feel in your ability to recognise a phishing attempt?

B.1.3. Situation within the department

- What kind of tasks do your employees mainly have within your department?
- To the best of your knowledge, is security a topic of conversation within your department?
- Have you had a discussion about phishing within your department?
- Do you know what actions your team members take when they receive a suspicious email?
- Could you describe the typical interactions among team members in your department?
- Do you know what your employees within your department do if they receive a suspicious email?
- Do you know if your department ever had a security breach due to a phishing email?

B.1.4. Training

- How would you evaluate the phishing training provided by the company?
- In your opinion, do employees make time to complete the phishing training?
- How would you rate the existing training program? What aspects do you appreciate, and what areas do you believe require enhancement? (This applies to both mandatory e-learning and simulated phishing emails.)
- In your opinion, how has the company's approach to phishing prevention evolved?
- What measures would you suggest to enhance the security culture and mindset within your department?
- Have there been any incidents of successful phishing attacks within your department? If yes, how was it handled?

B.1.5. Improvements

- From your perspective, what aspects could the company improve upon? Have you noticed any problematic behaviours?
- Have you seen/experienced weird behaviour?
- What are some common questions you receive from team members regarding phishing?
- What additional training or resources would you find beneficial?
- Which training components were instrumental in helping you identify phishing attempts?

Interview 1

The participant confirms that they deal with a lot of email traffic and occasionally receive emails that they are unsure about. In such cases, they forward the emails to a department that checks them for phishing. They mention using a phishing button in Microsoft and have used it once or twice but don't remember the response they received. When asked why they decided to report the emails, the participant explains that they were unsure about the reliability of those emails. They remain vigilant and suspicious, particularly when receiving emails with links.

The interviewer discusses the prevalence of email traffic from outside the company and how the participant's vigilance may be higher compared to others. The participant agrees and believes that people rely on the company's firewall and may not be as vigilant as they should be. They mention time pressure as a reason for being less careful with emails and how training on phishing awareness helps but needs to be reinforced regularly. The interview touches on a security breach in the participant's department, where a colleague's identity was compromised, leading to phone calls to clients and monetary loss. This incident made the participant feel unsafe and highlights the need for increased caution. They mention removing phone numbers from the website and LinkedIn to prevent such incidents.

Regarding training, the participant acknowledges the importance of reminders and variation in learning methods. In the interview, the participant shared their evaluation of the training and online learning about phishing within the company. They suggest combining e-learning with face-to-face meetings and discussions within departments. They mentioned that while the e-learning modules are good to have as reminders, there are so many of them that it becomes challenging to effectively inform people. The participant suggested that it would be better to have a variety of training methods, such as addressing the topic in department meetings or through more traditional/classical approaches where managers discuss it with their teams. They emphasized that e-learning alone may not be the ideal way to engage employees, as some people may not be fully attentive or engaged while going through multiple e-learning modules.

The participant also mentions that new developments and trends in phishing should be shared to keep employees informed and engaged. They suggested that newsletters could be an effective way to provide this information, and they personally read the quarterly newsletter carefully. They highlighted that when there is something new or trending in the realm of phishing, it tends to capture people's attention and generate a sense of urgency to stay informed. However, the participant expressed their observation that as media attention on cybersecurity decreases, the topic of phishing is becoming less prominent in people's minds.

In the interview, the participant also mentioned the vulnerability of new employees and the importance of providing specialized training regarding phishing and cybersecurity awareness. The participant highlighted that it is crucial for new employees to be aware of the risks and have the knowledge and tools to protect themselves and the company from phishing attacks right from the start of their employment. Overall, the participant believes that while there is some awareness about phishing, it may diminish over time if not reinforced. They emphasize the need for continuous training and sharing of information to stay vigilant against evolving phishing techniques.

Interview 2

The interviewee is extensively involved in cyber operations. They are responsible for various tasks related to preventing fraud, including taking down fake sites and collaborating with the marketing team on awareness campaigns for customers. From an employee perspective, they rely on global services provided by the company, but they also prioritize local awareness initiatives since they are the face of security for their colleagues. They deliver specific awareness content to employees and encourage them to complete global learning activities. The interviewee mentions that there are differences in the tactics used by attackers to target customers and employees. For customers, phishing attacks have evolved from email-based to SMS-based attacks. On the other hand, employees are primarily targeted through email, and attackers often aim to obtain credentials or deliver malware or malicious documents.

The interviewee personally doesn't receive many phishing emails, but they are familiar with the phishing emails that employees encounter. They often advise employees to report phishing emails through the appropriate channels, such as using the phishing reporting button. They also provide guidance on how to identify suspicious emails and frequently refer employees to articles, local or global awareness materials, and specific training campaigns. The company employs various strategies to raise awareness, including gamification platforms, webisodes, quizzes, and talks given by the CISO or other security professionals in different departments. The goal is to reach all employees, but it can be challenging due to the company's size. The company faces a significant task in reaching everyone. Mandatory learning programs are in place, but additional efforts are made to tailor awareness talks for specific groups within the company, such as marketing and social media teams who are exposed to customer-related risks, and financial teams who may face business email compromise attempts. They also participate in company-wide meetings and events, leveraging different platforms to deliver the security message. For new employees, the company organizes separate meetings to introduce them to various departments, including the security team, to establish a human connection and provide a point of contact for security-related inquiries.

While there is no concrete evidence to suggest that new employees are targeted more frequently, the interviewee acknowledges that attackers may search platforms like LinkedIn to identify new employees and potentially target them. However, they do not have specific metrics to support this assumption.

The interviewee believes that discussing phishing and cybersecurity is important within the company. They encourage dialogue and aim to create a culture where employees actively engage in discussions and share information related to phishing attacks or cybersecurity in general. However, they acknowledge that discussions may vary, and some employees may only focus on completing mandatory training without extensive discussion on the topic.

During the interview, the participant mentioned the concept of having a "champion in every tribe" as a crucial element in their company's security program. The participant explained that each department or team within the organization has an individual who takes the lead in promoting and advocating for security measures within their respective areas. These champions serve as the liaison between the security team and their colleagues, helping to raise awareness, address concerns, and ensure that security practices are effectively implemented.

The participant discussed the training programs and the room for improvement in the mandatory e-learning courses. While the content of the courses is accurate, the participant mentioned employees find them difficult to read and not engaging. To enhance the training experience, the company has incorporated expert videos as complementary resources, where employees record short videos discussing the learning content. This approach aims to make the material more relatable and engaging for the employees. The participant also expressed their preference for practical learning methods, providing an example of a program that includes a quiz where employees had to identify phishing emails. They found such interactive exercises more interesting and beneficial for the employees. Additionally, the participant mentioned the effectiveness of phishing employee campaigns conducted annually, which help employees learn by example rather than solely relying on reading materials.

The conversation then shifted towards the frequency of employees working from home and the flexibility in office attendance. Depending on the department and project, employees have different agreements regarding office presence. The participant mentioned going to the office approximately once a week or more based on project requirements. This flexibility allows for a balance between remote work and in-person collaboration.

The interviewer inquired about changes in employee behavior over time. The participant noted significant changes, especially in terms of the company's security culture. In the past, the security team was not actively involved in projects, but now they actively participate, propose security measures, and provide guidance. The participant also highlighted technological advancements, such as a global build and two-factor authentication (2FA), which have significantly improved security.

When discussing the effectiveness of training and resources, the participant expressed their opinion that instead of adding new programs, the focus should be on improving the existing ones. They believe that productivity and usefulness can be enhanced by transforming and refining the current methods rather than introducing more content. The participant emphasized the importance of utilizing the available time and resources effectively.

Regarding feedback, the participant mentioned that although they do not receive formal metrics, they receive positive insights and feedback from employees who approach the security team with specific doubts or questions, indicating that they see the security team as a valuable resource in the company.

The participant's concern about attachments, specifically office documents with macros, are discussed. Despite awareness efforts, macros are commonly used in daily work, posing a potential security risk. The participant mentioned that the integration of technical applications has helped mitigate some malware threats, but the reliance on macros remains a concern.

Regarding successful phishing attempts and data breaches, the participant stated that they have not experienced any major breaches resulting from real phishing attempts. However, they mentioned the occurrence of compromised credentials, shared on the dark web, usually detected through local services for monitoring compromised accounts. Recovered credentials are shared internally to reset passwords and raise awareness among employees.

The participant concluded by noting that as long as cybersecurity threats continue to evolve, there will always be a need for security professionals. They highlighted the ongoing importance of staying vigilant and adapting to changing circumstances.

Interview 3

During the interview, the interviewee was asked about their role within the company and their experience with phishing emails. The interviewee stated their work mainly involves risk control, whereby their tasks relate to conversations, meetings, and email communication. The interviewee spends about half of their time in meetings and the other half on other tasks. The interviewee also receives many reports through a central mailbox. Therefore, there are many unexpected issues that arise. Most of the contact is internal, with people from within the company.

Regarding phishing emails, the interviewee described them as emails in which the sender poses as a specific organization but tries to entice the recipient to click on a link in order to gain access to their phone, laptop, or computer. The interviewee mentioned receiving phishing emails regularly and usually marks them as spam. The interviewee has also reported a few phishing emails using the phishing report button, one of which was sent as part of a phishing simulation. Particularly, attractive titles such as a Christmas gift seem to be successful.

The interviewee stated that they have always been able to report and avoid falling for phishing emails so far. They noted that their role, which focuses on risk management, may make them more vigilant than other employees. If they have doubts about an email, they report it. The sender is an important trigger for them to determine whether an email should be reported as phishing.

The interviewee also discussed situations where they received unexpected emails, such as receiving an email about a task they didn't expect to receive. In these cases, they reported the email as phishing because they didn't recognize it. The email turned out not to be intended for them. They noted that phishing attempts are becoming increasingly sophisticated, such as receiving an email from a familiar delivery service after placing an order, which turned out to be phishing. The interviewee feels particularly vulnerable when the context of the email seems to align with their actions, such as receiving an email from a delivery service after placing an order.

The interviewee speculates that older people may be more susceptible to phishing attempts and that it is important to be extra cautious when receiving emails with links.

Regarding the Security Defence Centre (SDC) helpdesk, the interviewee mentioned that they only use the phishing report button and do not directly contact the helpdesk.

Within their department, phishing is not often discussed, except during risk awareness presentations for new employees and after a phishing simulation conducted by the company. The interviewee suggested making the "Report this" button more prominent in the awareness training they themselves provides to new employees, so that people have easier access to it.

Overall, the interviewee stated that the simulation training is more memorable than e-learning. The simulation training was also perceived positively and appeared professional. A suggestion would be to also address phishing through attachments rather than just links in an email, especially considering the high volume of email traffic within the company involving attachments. Additionally, they have noticed an increase in the attention given to phishing awareness, which is positive. It is particularly beneficial for new employees who may be less aware of the risks. The interviewee has never experienced a breach within the department. The high volume of email traffic acts as a kind of distraction, making it challenging to make sound judgments.

During the interview, the discussion also touched on remote work and online communication within the company. The interviewee confirmed that there is indeed more remote work taking place. On average, employees spend about half of the week in the office and the other half at home. Communication primarily takes place through Microsoft Teams, where conversations and chats with colleagues occur. Email is mainly used for exchanging larger documents and contacting external parties.

During the conversation, the topic of employees approaching the interviewee with suspicious emails, especially phishing emails, was also raised. The interviewee mentioned that during a testing period, employees did come to them with questions about the reliability of emails. At that time, there was no phishing button available, so reporting required multiple steps. Since the new method with the button, reporting phishing emails has been more effective. Employees also sometimes approach the interviewee to inquire about the trustworthiness of an email.

C

Additional figures

C.1. Barplots

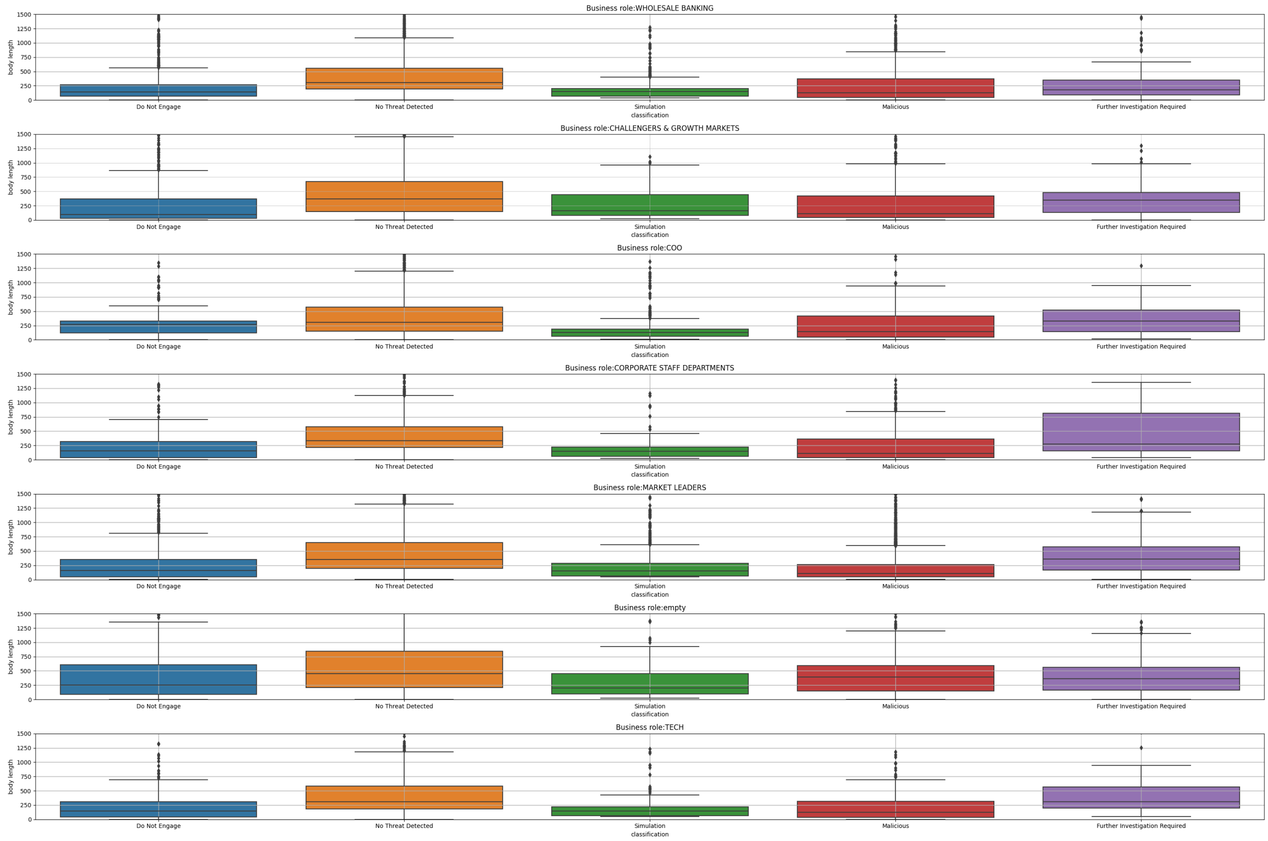
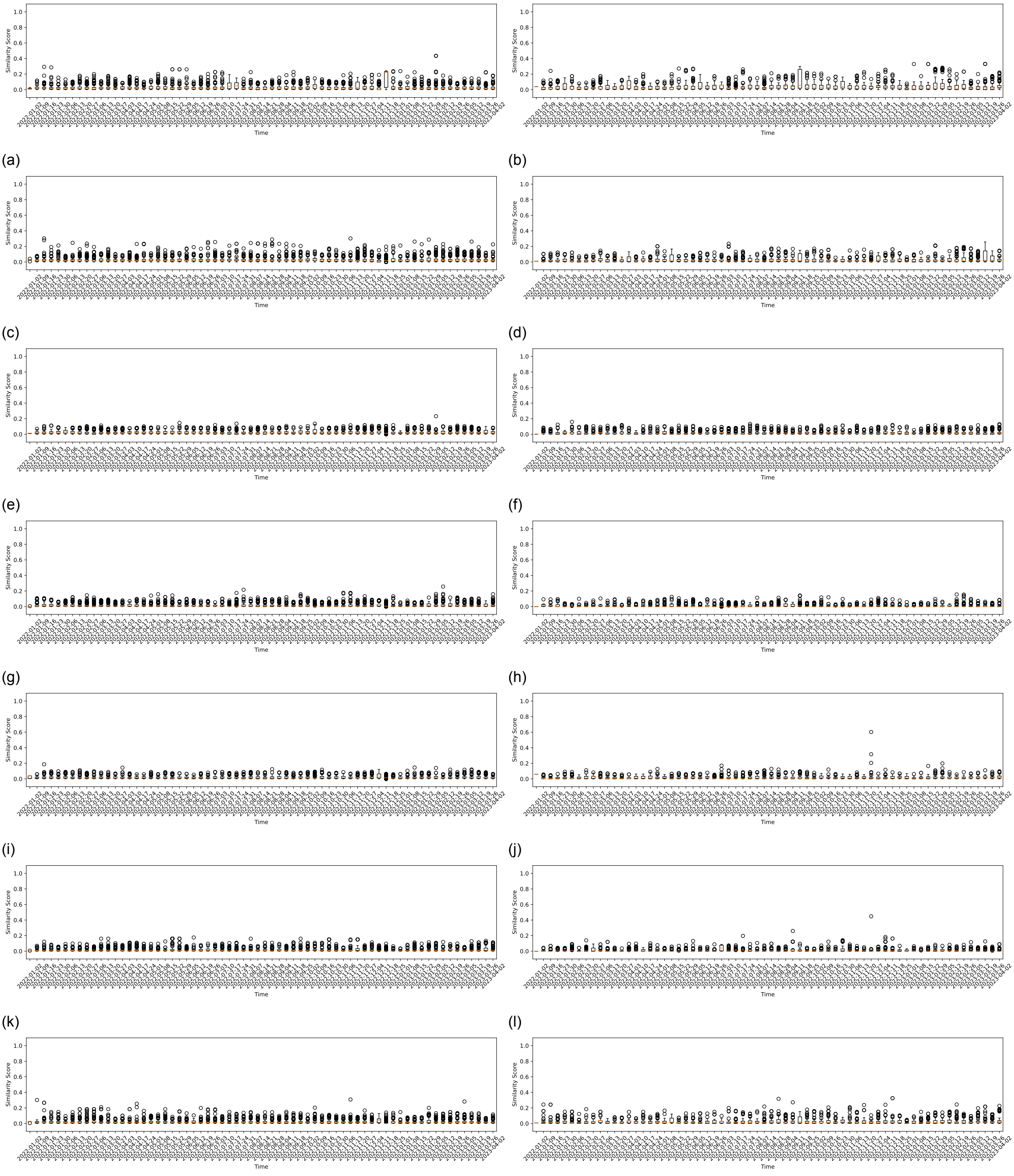


Figure C.1: Boxplots of the tokenised lengths of the email body per business line for the malicious and benign emails.

C.2. Similarity scores

Similarity scores over time.



(a)

(b)

(c)

(d)

(e)

(f)

(g)

(h)

(i)

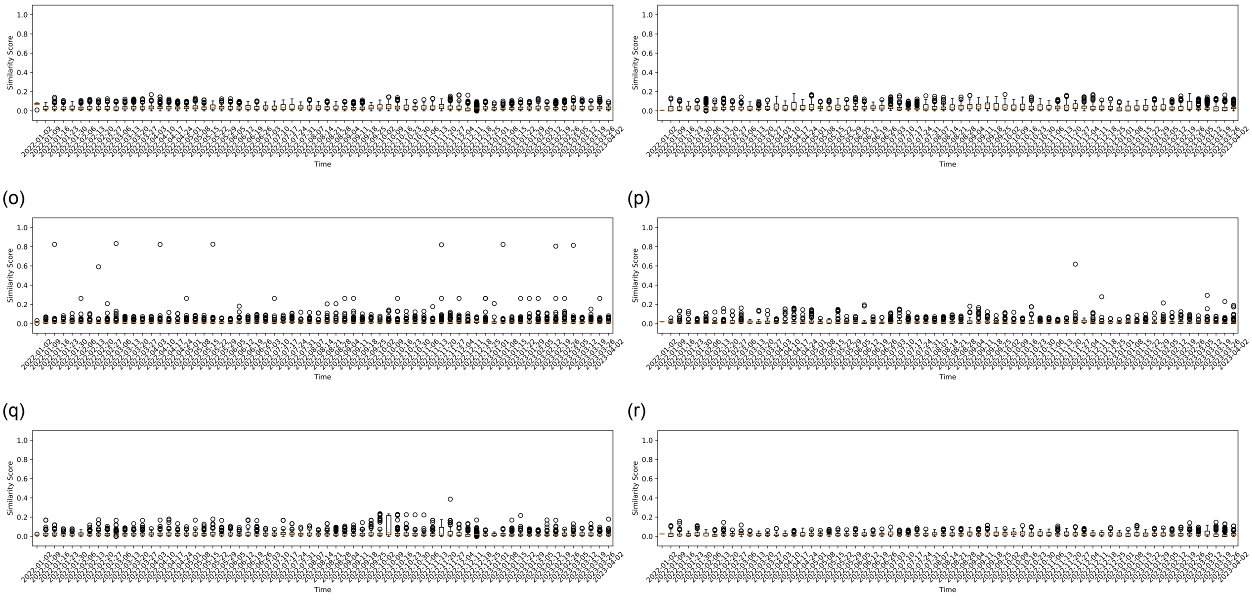
(j)

(k)

(l)

(m)

(n)



(o)

(p)

(q)

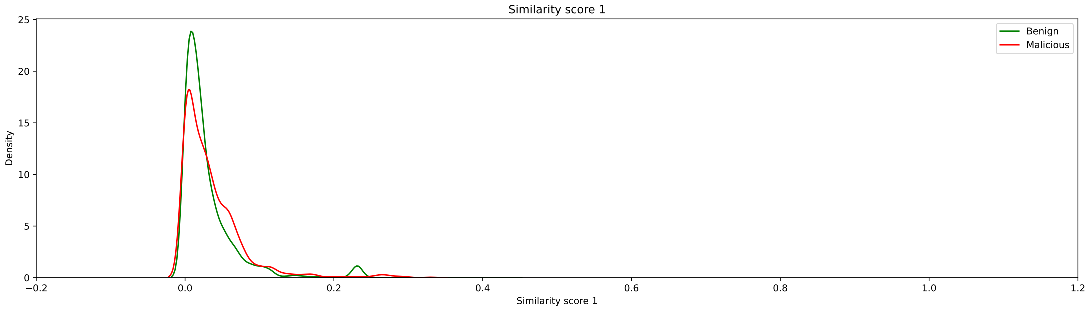
(r)

(s)

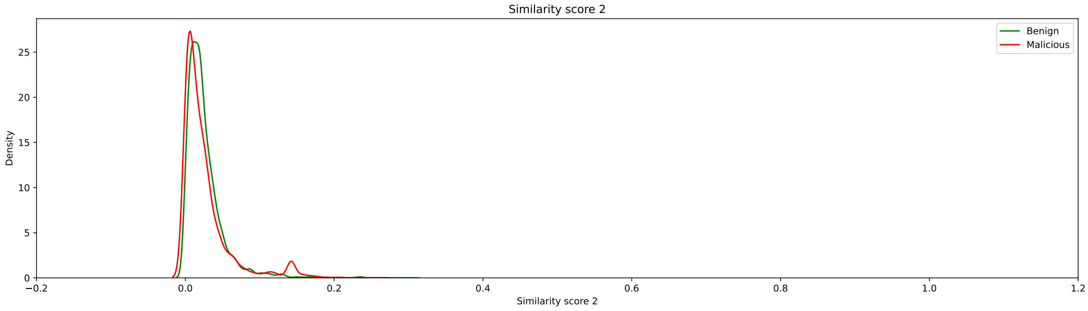
(t)

Figure C.2: Similarity scores presented in boxplots over time.

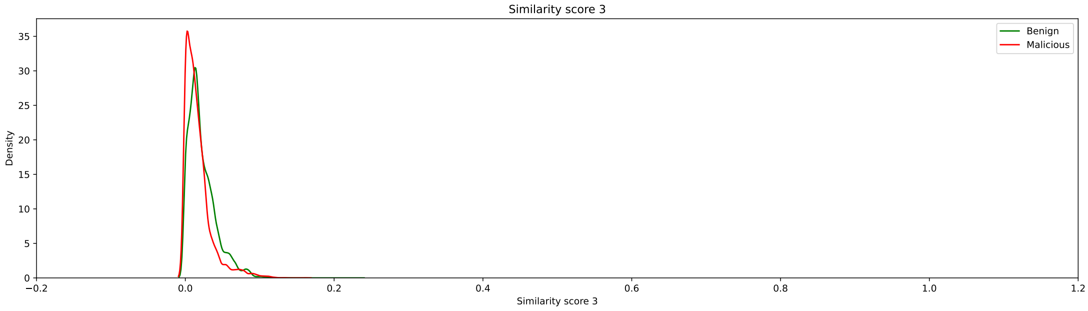
C.3. KDE plots



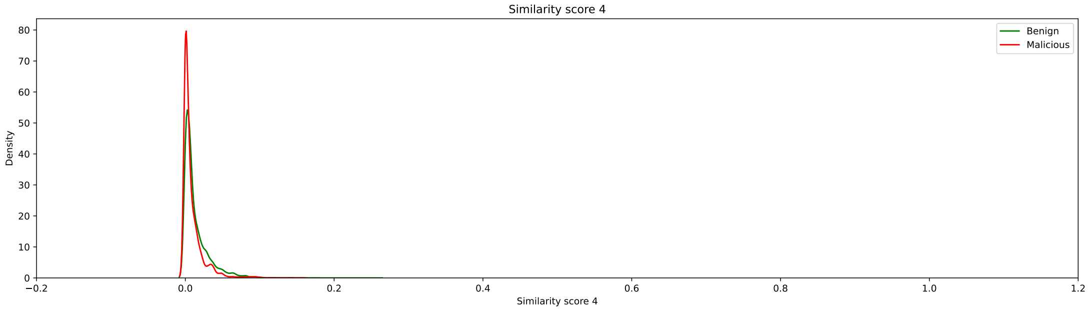
(a)



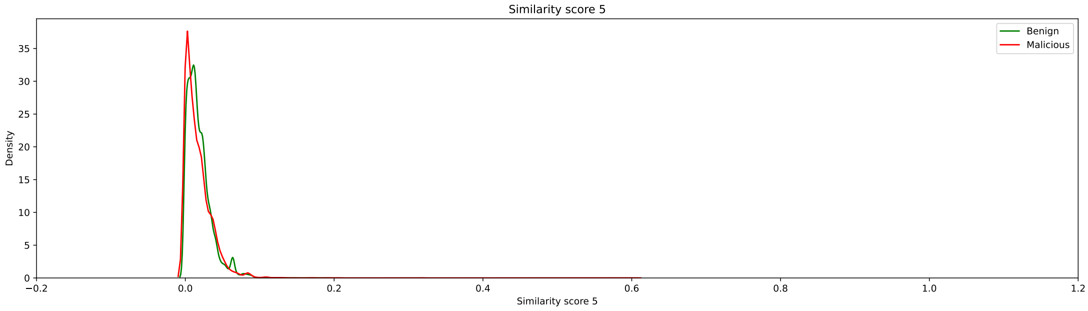
(b)



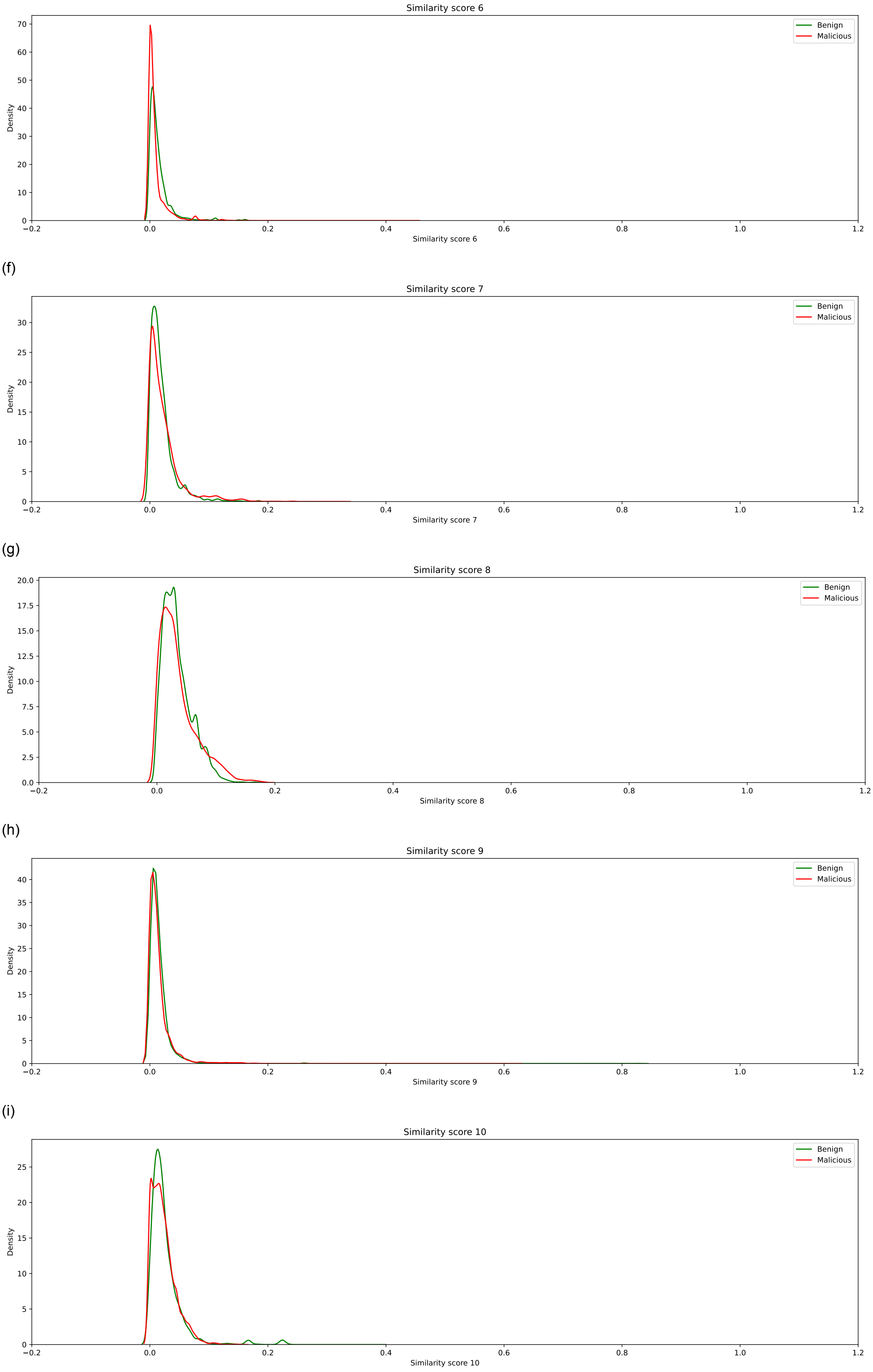
(c)



(d)



(e)



(f)

(g)

(h)

(i)

(j)

Figure C.3: The Kernel Density Estimate of the similarity scores between the malicious and benign emails and the 10 different simulation emails.