Different Seas, Different Phishes – Large-Scale Analysis of Phishing Simulations Across Different Industries

Oskar Braun
AWARE7 GmbH &
Rhine-Waal University of Applied Sciences
Gelsenkirchen, Germany
oskar.braun@hsrw.org

Jan Hörnemann
AWARE7 GmbH &
Westphalian University of Applied Sciences
Gelsenkirchen, Germany
jan@aware7.de

Norbert Pohlmann
Westphalian University of Applied Sciences
Institute for Internet Security
Gelsenkirchen, Germany
pohlmann@internet-sicherheit.de

Tobias Urban
Westphalian University of Applied Sciences
Institute for Internet Security
Gelsenkirchen, Germany
urban@internet-sicherheit.de

Matteo Große-Kampmann
Rhine-Waal University of Applied Sciences
Kamp-Lintfort, Germany
matteo.grosse-kampmann@hochschule-rhein-waal.de

Abstract

Phishing is an increasing threat to the security of end-users, networks, and organizations. Phishing simulations via email are a widespread tool used to measure user awareness, especially in workplace settings. However, current studies focusing on large-scale analysis of phishing simulations often have issues: The phishing simulations were conducted using a small sample size (mostly one or two organizations), or while many emails are sent, the analysis focuses only on specific companies. This study analyzes phishing simulations conducted over three years at 36 organizations with over 68 000 delivered emails. We compare different dimensions of the organizations where these simulations were conducted, such as the economic sector and departments. Furthermore, we evaluate various dimensions of phishing simulation campaigns, such as detection difficulty and the scenario under which the simulation occurs. Our findings indicate significant disparities in the results, such as the industry sector in which the company operates. Moreover, we find substantial differences between the success rates of varying scenarios used for phishing emails.

CCS Concepts

• **Security and privacy → Human and societal aspects of security and privacy**.

Keywords

phishing, awareness, end-users, phishing simulations

ACM Reference Format:
Oskar Braun, Jan Hörnemann, Norbert Pohlmann, Tobias Urban, and Matteo Große-Kampmann. 2025. Different Seas, Different Phishes – Large-Scale Analysis of Phishing Simulations Across Different Industries. In *ACM Asia Conference on Computer and Communications Security (ASIA CCS '25), August 25–29, 2025, Hanoi, Vietnam.* ACM, New York, NY, USA, 15 pages. https://doi.org/10.1145/3708821.3733905

1 Introduction

Phishing emails are still a prevalent threat to organizational security. Several countermeasures like detection mechanisms [53] or machine learning approaches [44] were used in the past. In recent years, the trend has also introduced so-called *security awareness campaigns* (SAC) and *phishing simulation campaigns* (PSC) alongside technical countermeasures to raise awareness about phishing emails. This leads to an added security layer, and nowadays employee training is vital to an organization's cybersecurity and mandatory according to different cybersecurity regulations and standards, e.g., NIS2 [51], ISO 27001 [46], or ISO 27004, which recommend PSC as a measure to verify security trainings [27].

It is currently more complex to quantitatively evaluate human aspects of cybersecurity than to evaluate technical measures, e.g., the blocking rate of a firewall [7, 48]. One approach to obtain quantitative data is to perform PSCs in an organization that generate data that can be evaluated. By analyzing this data, organizations seek to conclude the effectiveness of their security program [4–6].

Other studies researching phishing simulations and their click rates focus primarily on one organization [28, 34, 47]. Some academic studies are also biased toward academic institutions as the authors solely conduct experiments within a university setting [54, 61]. This work addresses these gaps by performing large-scale phishing email campaigns in 36 organizations, sending and analyzing 68 743 phishing emails over 45 months using 96 distinct phishing campaigns. Our approach allows us to assess whether different sectors or departments within a company are more likely to identify potential phishing emails. Our findings help make specific recommendations for organizations to achieve a comprehensive level of awareness and realistically evaluate the phishing threat.
In summary, we make the following contributions:

- By performing a large-scale quantitative analysis of 96 proprietary phishing campaigns across 36 organizations in various

industries, we show the non-generalizability of previous work, which analyzes phishing within one or two organizations. Our large-scale horizontal study fills this research gap.
- We analyze different economic sectors and departments within a company and show significant disparities in the susceptibility to phishing. By comparing different levels of detection difficulty according to the Phish Scale [13, 52], we show that the difficulty is not significant to predict dangerous interactions, but rather other factors are more influential.
- Through the analysis of different phishing scenarios, we show that some (e.g., "Don't Miss Out", where the email contains a limited offer that the user should not miss) lead to considerably more dangerous interactions with the emails than other scenarios.

Research Questions. Our work was guided by the following research questions we answered in our large-scale study.

- *RQ1: Are results from previous studies reproducible if they are evaluated across industries?* The first goal of this study is to understand whether findings from previous studies focusing only on one (or two) organizations can be generalized across multiple companies.
- *RQ2: What are the differences between different dimensions of the phishing simulations such as generality (i.e., generic phishing compared to spear-phishing), detection difficulty, and scenario?* We want to determine whether different dimensions of a phishing email, such as an individualized email, a high detection difficulty, or specific scenarios, lead to significantly higher click rates.
- *RQ3: How do different economic sectors (resp. departments) compare regarding the click rate of the users employed in the respective organizations (resp. departments)?* Lastly, we want to ascertain whether click rates across different economic sectors and departments differ and whether these differences are significant.

Main Findings. Regarding our research questions, we summarize the following key takeaways from our experiments.

- *Findings related to RQ1:* The rates of dangerous interactions with phishing emails vary highly across different organizations (see Section 4.1). Thus, results from studies focusing on single organizations are not reproducible if evaluated across industries.
- *Findings related to RQ2:* First, phishing simulation campaigns that did not request sensitive data from the users led to higher click rates than campaigns with possible data submission (see Section 4.1). Second, spear-phishing emails yield higher interaction rates than generic emails (see Section 4.2). Third, when raising the detection difficulty, there is a slight increase in click rates (see section 4.5). Our findings indicate, however, that other factors are more significant than detection difficulty. Fourth, there are significant differences in the scenarios used. For example, while "Account compromised" was handled rather well by users, "Don't Miss Out" led to many dangerous interactions (see Section 4.6).
- *Findings related to RQ3:* There is a significant difference in the number of dangerous interactions between economic sectors. Mainly, the second economic sector (i.e., manufacturing and industry) performed more securely (in terms of our study) than the others (see Section 4.3). Furthermore, there is a disparity in dangerous interactions between different departments. Significantly, Finance & Administration performed safer, while Sales & Marketing had many more dangerous interactions (see Section 4.4).

2 Background & Terminology

This section provides the necessary background to facilitate the understanding of phishing simulation campaigns 2.1.

2.1 Phishing Simulation Campaigns

Security awareness campaigns (SAC) try to influence employees' security behavior in various organizations. Phishing simulation campaigns (PSC) should protect companies from the prevalent threat of phishing, which is, for example, used by advanced, persistent, and harmful actors to access company infrastructure [57]. The market for SACs is generally growing, and PSCs play a significant role in that growth, particularly since cybersecurity awareness has become a success factor for organizations that were victims of a cyberattack in the past and a firm's market valuation in general [2]. CISOs also see phishing simulation campaigns as a leading driver for human-centered security [23]. Demand for location-independent awareness solutions like PSCs is growing due to the rising numbers of remote workers with remote connections to companies, posing a threat outside of the classical perimeter [39]. Another driver for demand in phishing SACs and PSCs are regulations, for example, the European NIS2 directive, which requires organizations to apply cybersecurity measures across their technical infrastructure and workforce [55]. PSCs can consist of multiple campaigns that differ in the email used and the detection difficulty of that email. At least one or more of these campaigns are combined to create one phishing simulation. Technically speaking, PSCs work similarly regardless of the specific vendors. The PSC provider must register domains and maintain an infrastructure for the phishing simulations. Domains are registered, and an email server is configured from which the emails originate. The simulation provider uses this infrastructure because it wants emails to appear as if they were sent from a specific domain, but in fact they are originating from a spoofed domain: executive@example.com is spoofed by the campaign provider as executive@excample.com, for example [18]. Another method of deceiving a user is using convincing subdomains from otherwise legitimate-seeming domains (e.g., microsoft.security-com.org). As the providers themselves are in control of the domain, they can employ anti-spoofing techniques and protocols like *Sender Policy Framework (SPF)* or DomainKeys Identified Mail (DKIM), among others, to make their emails appear even more legitimate and thus bypass certain technical countermeasures like spam detection [26, 37]. Attackers and providers of commercial phishing campaigns need to properly configure these anti-spoofing techniques for their domains so that the emails are delivered to the respective victims or clients. The phishing or spear-phishing email is crafted either with a proprietary tool or open-source alternatives like GoPhish [60], the Social Engineer Toolkit [31], or KingPhisher [12]. For all toolkits, a PSC provider first needs to create a template for the phishing email that should be sent. Typically, not just one email is sent, but rather multiple emails over a predefined period to deduce a trend from the multiple phishing attempts. The email usually has varying difficulties, defined mainly by the number of cues that the email is malicious and by the alignment of the email with the users. A framework often used to rank difficulty is the *Phish Scale* by Steves et al. [52]. Afterward, a tracking pixel and trackable link are inserted into the email. The tracking pixel helps to measure

email opening rates and the trackable link measures *clicks on the link*. The clickable link guides the victim of the PSC campaign to a prepared landing page, where user behavior is tracked. These landing pages are typically login pages akin to the ones real attackers would use [40, 49]. Measuring whether a user submits login data to a login field on these landing pages is especially interesting. After the user submits login data, there is a redirection towards a *post-mortem* page. This page reveals to the user that the received email was potentially malicious [21].

3 Method

This section describes the surveyed organizations (see Section 3.1), the analyzed phishing campaigns (see Section 3.2), and our approach to analyzing the results (see Section 3.3).

3.1 Surveyed Organizations and Dataset

As part of the study, the researchers collaborated with one phishing service provider (partner organization), which gave them access to the aggregated PSC results of 36 organizations. Organizations that purchased phishing simulations conduct business in various sectors, from insurance providers to IT service providers and manufacturers of special tools. The authors did not influence the selection of the organizations, as the organizations themselves paid the partner company to use the respective phishing simulation service. Initially, we received a list from the organization with the following information: (i) pseudonymous identifier, (ii) economic information (sector, industry, ISIC Classifier and code), (iii) organization size, (iv) balance sheet total (BST), and (v) country (see Table 6 in Appendix B). Due to data protection laws, confidentiality agreements, and individual agreements with, for example, work councils, the data was analyzed pseudonymously. The research team thus had no access to the actual names of the organizations, but only to the results of a specific campaign (see Section 6). In addition to an individual ID, the data records contain the action and the associated organization and department. Any ID can be assigned to one of the following categories of actions: (i) delivered, (ii) opened, (iii) clicked, (iv) data submitted. The category selection is sorted accordingly in ascending order, i.e., a person who has clicked on an email has opened it before and has not been evaluated twice in this study. Before the start of each individual phishing campaign, tests were conducted with a technical contact person of the organization to ensure the following: (i) the sent-out emails are appropriately delivered; (ii) they are not filtered as spam or phishing; (iii) their visual layout appears as intended, and all images are loaded; (iv) whether the human opening of the email is registered; (v) whether the tracked actions were human actions and not triggered by any software.

Organization Sizes. Of the 36 organizations, seven are considered small-sized organizations, 13 are medium-sized organizations, and 16 are large organizations based on the number of employees and using the SME classification scheme of the EU [15]. We analyzed each organization's balance sheet total to strengthen our claim that this dataset is diverse. The balance sheet total is the sum of all items on the assets or liabilities side of the balance sheet at the end of a financial year. Using this approach, we could not obtain results for eight organizations, as they do not need to publish their financial statements at the end of a financial year, e.g., due to their legal entity being a registered association. The average balance sheet total for the remaining organizations is 269 603 448.28 EUR (min.: 1 100 000 EUR, max.: 3 400 000 000 EUR, SD: 696 266 261.4). For small-sized organizations, the balance sheet total average is 4 885 714.29 EUR; for medium-sized organizations, the average is 12 946 153.85 EUR; for large organizations, the average is 421 500 000 EUR. Our dataset thus presents a diverse representation of different organizations with regard to balance sheet total as well.

Economic Diversity of the Organizations. Furthermore, the observed organizations are distributed across different economic sectors. The economy and the organizations that form it are often classified using a three-sector model [17]. As prevailing economic and societal influences have changed over the years, several economists suggested an extension of this model by a fourth quaternary sector as early as the 1980s [19, 30]. We follow the proposed definition of the quaternary sector from [56]. Those organizations contributing to generating, applying, and sharing knowledge are categorized in the quaternary sector. Services unrelated to developing and sharing knowledge and information remain in the tertiary sector [56]. None of the organizations we analyzed belong to the *primary sector*, which describes the extraction of raw materials (e.g., agriculture, fishing, mining). Nine organizations belong to the *secondary sector*, which contains manufacturing of any kind (e.g., construction, manufacturing, chemical production). In the *tertiary sector*, which contains service industries such as hospitality or real estate, we identified 12 organizations. In the *quaternary sector*, which includes knowledge-based services, we surveyed 15 organizations.

Departments in the Organization. Our partner company gave us specific interaction rates for different departments to understand whether there are generalizable differences in phishing susceptibility within each organization over the whole dataset. Since the departments varied greatly across the organizations and some organizations have different names for similar departmental functions, e.g., "Sales" or "Business Development" [41], or regarding the detailed allocation level, we decided on a simple categorization into three fundamental groups: (i) Finance & Administration (F&A), (ii) Production & Operations (P&O), (iii) Sales & Marketing (S&M). Combining, for example, Finance (accounting and paying bills) with different forms of Administration (which includes back office and HR) is reasonable, as their tasks overlap and sometimes even produce increases in efficiency [42]. Similarly, marketing and sales are related and often in many organizations entangled [3]. In the group Production & Operations, besides the apparent departments such as manufacturing, research, development, IT, etc., we also included all miscellaneous or uncommon departments that did not have counterparts in other organizations [1]. This categorization of different departments helps us to answer RQ3.

Dataset Heterogenity. Our dataset is heterogeneous with regard to the number of employees, financial background, economic sectors, and examined departments. The balance sheet total is skewed towards large organizations that comprise a large portion of this dataset. Accordingly, it offers overarching insight into different industries and economic sectors and helps us answer the defined research questions. While all survey organizations are chosen from a single client (i.e., our partner company), they represent a diverse

set of companies to analyze (see Table 6 in Appendix B). Thus, the dataset is valid to analyze how similar phishing simulations affect different companies. Further, the focus on German-speaking countries decreases the likelihood that the measured results might be a result of different cultural perspectives.

Open Science. To foster future research, we make our pseudonymous dataset and analysis scripts available at https://github.com/awareseven/Different-Seas-Different-Phishes.

3.2 Phishing Campaigns Used

The partner company sent phishing emails on behalf of the organizations in different campaigns. One phishing campaign is defined as follows for our measurements: The distinct feature distinguishing campaigns is the email that changes with each campaign. The email contains a link to a website, the "landing page". The landing page and set of users can remain unchanged for multiple campaigns. Interactions with the email are logged until the end of a waiting period of seven days from the last phishing email sent. This waiting period was implemented to account for anyone absent from the workplace (e.g., due to illness or holiday). 68 743 phishing emails were sent out and delivered in 96 campaigns, and employees' interactions with these emails were monitored and analyzed. The emails ranged from least to very difficult according to the Phish Scale [13, 52]. The contents of the phishing emails were discussed and defined in a direct exchange with the participating organization to ensure that their content is relevant to the users. The emails were designed once and then sent out to the organizations. Our partner company sends phishing emails ranging from brand impersonations (e.g., Google-, Microsoft-, or Amazon-branded emails) to individualized "spear phishing" emails. The spear phishing emails were tailor-made from information found on various websites using open-source intelligence [57], or designed to appear as if they originated internally within the organization (e.g., from IT or HR departments). The landing pages mimicked login forms for different popular web services (e.g., Microsoft 365) for generic or organization-branded logins, depending on the specific campaign. The number of phishing simulations per campaign and organization varied between 1 and 13 (mean: 2.72, modal value: 1). The participating organizations are primarily from German-speaking countries (i.e., Germany, Austria, and Switzerland), as sales and customer acquisition of our partner company were mainly conducted in this region. In three campaigns with 17 133 emails, a bilingual email containing the same German and English text was used. This only occurred within organizations where such communication is customary. In eight campaigns, separate German (10 799) and English (9 031) emails were used. Two campaigns with 47 delivered emails were sent in English. The employees were typically not informed that a phishing simulation was planned.

Rating Difficulties of Phishing Emails. The Phish Scale [52] uses a three-step approach. The first step counts the number of cues that an email might be phishing. This results in a cue category: "Few", "Some", and "Many". In the second step, the premise alignment is determined, which results in a premise alignment category: "Weak", "Medium", or "Strong". In the last step, these two categories are combined to yield a detection difficulty: "Least difficult", "Moderately to least difficult", "Moderately difficult", or "Very difficult". For example, an email with the cue category "Some", but a "Strong" premise alignment produces a detection difficulty of "Very difficult". Depending on the difficulty of the phishing campaign that the organization decided on, the partner company manually prepared phishing emails with different cues. For our experiment, one of the authors (Rater 1) categorized all phishing campaigns by their detection difficulty level, closely following the NIST Phish Scale User Guide [13]. To make the classification more reliable, another rater (Rater 2) classified them independently without having access to Rater 1's classifications. Afterward, we calculated the inter-rater reliability, Cohen's Kappa, and weighted it quadratically so that opposite classifications have more weight than similar ones. We obtained a Cohen's Kappa of 0.65, showing substantial agreement between the two raters. Therefore, we use this classification as the categorization for the paper and to answer RQ2. In the following section, the detection difficulties of generic phishing emails range from "Least difficult" to "Moderately difficult". The detection difficulties of spear-phishing emails range from "Moderately difficult" to "Very difficult". One spear-phishing email, which we categorize as *moderately complex* using the Phish Scale and the TCoP scenario, is depicted in Figure 9 in Appendix A.

Table 1: The different scenarios used in the phishing emails by the service provider with an illustrative example

Scenario	Description	Example of content
Don't Miss Out (DMO)	Email contains a limited offer or event that the user should not miss	Information about a discount or winning a lottery
Take Care of Process (TCoP)	Email requests or demands the user to take care of a process	Email about a system update where the user needs to take a specific action
Account Compromised (AC)	Email pretends that a sensitive account was compromised	There was suspicious activity on an account, and now action must be taken
Take Care of Email (TCoE)	Email pretends that the user needs to take care of a (work-related) email	There was an issue with sending the last email. The user must take care of it
Click File/Ticket (CF/T)	Email contains a file or a (work) ticket that is general and lacks details	An email is sent with a download link for the file "salaries_jan2024.xlsx"
Unexpected Order (UO)	Email pretends to be a dispatch, shipping, or delivery without details	User receives an email regarding a product they never ordered
Information (I)	Email offers some (obligatory or optional) information to the user	Click this link to see recent changes in the organizational chart

Overview of the Infrastructure Used. An overview of the phishing infrastructure used and the data collection process can be found in Figure 1. The figure illustrates the aspects of the phishing infrastructure: 1) Each user is assigned a different ID, and emails with the IDs are sent to users; 2) openings of the emails are logged via tracking pixels, with the IDs included in the URLs; 3) clicks in the email are

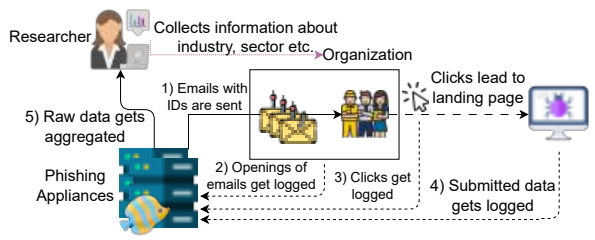

Figure 1: Overview of the Research Process

logged, again with the IDs included in the URLs of the links, and each click directs the user to a landing page; 4) if the landing page allows for it, data submissions by users are logged; 5) these events yield raw data, which is then aggregated for analysis. Additionally, the partner company sent the emails from dedicated servers (i.e., nobody else is using the servers) and used custom-made phishing software that works similarly to open-source implementations like GoPhish [60]. The partner company set up all necessary security options, e.g., DKIM and SPF, for server and domain infrastructure for the emails to reach the users' inboxes.

3.3 Statistical Methodology

For each email delivered to a user, we considered different categories of events: "Opened", "Clicked" and "Data submitted" (the latter is considered for campaigns with possible data submission on the landing page). We compared different average results, e.g., generic vs. spear phishing, economic sectors, departments, etc., for those categories. For example: "How many users working in Marketing & Sales clicked on a fraudulent link, compared with users working in Finance & Administration?" Because of the dichotomous results, each event can be statistically interpreted as a Bernoulli trial. For each separate phishing campaign and each of the three (resp. two) categories of events, we have a sample number n of independent Bernoulli trials, and thus we obtain a Bernoulli process. Since users behave both differently and independently, for each phishing campaign we get separate Bernoulli processes of sample n for the categories "Opened", "Clicked", and "Data submitted" (if the landing page allows for it). We statistically estimate each of these Bernoulli processes to have a binomial distribution, though with different probabilities p_i.

We used the one-tailed binomial test, which uses the binomial cumulative distribution function (CDF), to compare whether the difference between two binomial distributions is statistically significant [14]. Throughout Section 4, we used the strict significance level of $\alpha = 0.001$, recognizing that large-scale datasets are prone to yield statistically significant results even when the effect size is low. In addition, we conduct a Bayesian logistic regression [14, 29] to investigate which categories (e.g., generic or spear-phishing, scenario of email, economic sector of organization) have the most significant impact on the click rate. We used the average click rate ($p \approx 0.09$) of all emails (see Section 4.1) to find a proper prior intercept via the logit-function: $\text{logit}(p) = \ln(\frac{p}{1-p}) \approx -2.31$. For the nominal variables (i.e., sectors and scenarios), we chose certain "baseline categories", with which we compare the other categories.

4 Results

In this section, we analyze the collected data, specifically the emails that were delivered and opened, and in which a link was clicked or data submission logged. Moreover, we provide basic statistics about the results, such as the average email interaction rate. In the respective subsections, we analyze the data across different categories. First, we present the overall results (see Section 4.1). Next, we compare the success rates of generic and specific phishing simulations (Section 4.2). Furthermore, we compare how the clicking behavior changes per sector for different economic sectors (see Section 4.3). We compare (1) different kinds of departments and their clicking behavior across organizations (see Section 4.4), (2) different detection difficulty categories according to the Phish Scale (see Section 4.5), and (3) interaction rates for different scenarios of content in the phishing emails (see Section 4.6).

4.1 Overall Results

We analyzed 96 phishing campaigns at 36 organizations with 68 742 delivered emails. Of these, 28 campaigns allowed for data submission (29.17%), comprising 17 737 delivered emails (25.8%), a subset of the total delivered emails. For the rest of this section, the percentage values of the presented submission rates are solely based on this subset of emails that allowed data submission.

The overall results are presented in Tables 2 (all phishing simulations) and 3 (phishing simulations with data submission) and indicate that overall 20 085 emails (29.22%) were opened by people in the 36 organizations. In total, 6 127 (8.91%) users clicked on a link in the emails. Of these users, 422 (0.61%) submitted data on a landing page. In Figure 2, we present the interaction rates for each organization. The organizations are pseudonymously listed in descending order, and the respective rates are averaged over all campaigns conducted at each organization. The missing cells in the third column indicate that no campaign with data submission was conducted. The results show that the rates vary across different organizations. The differences between the rates of campaigns with and without data submission are significant ($p < 0.001$).

Figure 3 shows the distributions of the log-odds coefficients for the predictor variables, resulting from the Bayesian logistic regression (see Section 3.3). The higher the absolute value of the mean of a distribution, the greater the impact this category has on the click rate. Positive values imply an increase in the click rate, while negative values imply a decrease. They are sorted in descending order by impact. The results indicate that the scenario "Account Compromised" would significantly decrease the click rate, while the scenario "Unexpected Order" would increase it notably. Further, phishing in the quaternary sector and spear-phishing both lead to an increase, while phishing in the secondary sector leads to a decrease in the click rate. The other categories have less impact.

No email template was sent to all organizations, but two cases were sent to several organizations: (i) An email sent to 15 organizations with the difficulty "Moderately to least difficult" and scenario "Take Care of Email". It did not allow data submission and had an average opening rate of 18% with a click rate of 9%. (ii) An email sent to 5 organizations with the difficulty "Moderately difficult" and scenario "Account Compromised". It had an average opening rate

Table 2: Results of the phishing campaigns for different categories. Results are rounded for better readability.

Type	Delivered	Opened	Clicked
Overall	68 742	20 085 (29%)	6 127 (9%)
Generic	33 769	9 338 (28%)	2 462 (7%)
Spear Phishing	34 973	10 747 (31%)	3665 (10%)
Economic Sectors			
Secondary	30 488	9 259 (30%)	2 114 (7%)
Tertiary	17 893	4 066 (23%)	1 384 (8%)
Quaternary	20 361	6 760 (33%)	2 629 (13%)
Departments			
Finance & Administration	714	159 (22%)	47 (7%)
Production & Operations	6 479	1 931 (30%)	596 (9%)
Sales & Marketing	562	299 (53%)	77 (14%)
Detection Difficulties			
Least difficult	4 923	1 576 (32%)	414 (8%)
Moderately to least difficult	9 879	1 904 (19%)	904 (9%)
Moderately difficult	31 014	10 154 (33%)	2 136 (7%)
Very difficult	20 134	5 517 (27%)	2 304 (11%)
Scenarios			
Don't Miss Out	18 105	5 697 (31%)	2 262 (12%)
Take Care of Process	13 934	4 694 (34%)	1 091 (8%)
Account Compromised	10 893	3 626 (33%)	355 (3%)
Take Care of Email	9 631	1 825 (19%)	853 (9%)
Click File/Ticket	5 857	852 (15%)	467 (8%)
Unexpected Order	3 939	1 385 (35%)	407 (10%)
Information	3 591	1 072 (30%)	323 (9%)

of 37%, an average click rate of 4% with an average data submission rate of 2%. These results align with our overall findings.

Summary. The data reveals that the average email campaign performance metrics show an opening rate of 29%, a click rate of 9%, and a submission rate of 2%. Campaigns that do not involve data submission exhibit a click rate nearly twice as high as those requiring data submission. However, within the subset of campaigns involving data submission, 43% of users who clicked on a link proceeded to submit their data. These figures highlight the differential impact that a data submission requirement has on user engagement and action. Also, interaction rates vary broadly across organizations.

4.2 Generic vs. Spear-Phishing

We analyzed the different rates of interactions with generic emails compared to spear-phishing emails, i.e., when specific information connected to the respective organization was used. Of the 96 campaigns, 32 (33.33%) fall into the spear-phishing category, and 18 organizations (50%) were targeted. According to the Phish Scale, these emails had two levels of detection difficulty: either "Moderately difficult" or "Very difficult." Seven of these campaigns also used specific spear-phishing landing pages, while the others did not.

Table 3: Results of different categories, only for campaigns with data submission

Type	Delivered	Opened	Clicked	Submitted
Overall	17 737	4 860 (27%)	984 (6%)	422 (2%)
Generic	10 962	3451 (31%)	405 (4%)	135 (1%)
Spear Phishing	6 775	1 409 (21%)	579 (9%)	287 (4%)
Economic Sectors				
Secondary	10 477	2 555 (24%)	384 (4%)	123 (1%)
Tertiary	4 778	1 433 (30%)	262 (5%)	95 (2%)
Quaternary	2 482	872 (35%)	338 (14%)	204 (8%)
Departments				
Finance & Administration	252	54 (21%)	24 (10%)	9 (4%)
Production & Operations	3 656	1 112 (28%)	327 (8%)	160 (4%)
Sales & Marketing	228	102 (45%)	18 (8%)	7 (3%)
Detection Difficulties				
Least difficult	241	4 (2%)	1 (0%)	0 (0%)
Moderately difficult	14 055	4 596 (33%)	761 (5%)	351 (3%)
Very difficult	3 441	260 (8%)	222 (6%)	71 (2%)
Scenarios				
Don't Miss Out	382	151 (40%)	92 (24%)	42 (11%)
Take Care of Process	3 279	1 017 (31%)	239 (7%)	148 (5%)
Account Compromised	9 011	3 103 (34%)	272 (3%)	109 (1%)
Take Care of Email	1 223	243 (20%)	113 (9%)	26 (2%)
Click File/Ticket	3 410	249 (7%)	207 (6%)	61 (2%)
Unexpected Order	241	4 (2%)	1 (0%)	0 (0%)
Information	191	93 (49%)	60 (31%)	36 (19%)

In total, 34 973 emails were delivered in these spear-phishing campaigns (50.88% of the total number of delivered emails). In Figure 4, we compare the interaction rates in generic and spear-phishing campaigns. We compare the averages of the opening rates for generic (mean: 29.56%; SD: 13.64%) and spear-phishing campaigns (mean: 37.77%; SD: 14.68%), the averages of the click rates for generic (mean: 7.24%; SD: 6.2%) and spear-phishing campaigns (mean: 18.09%; SD: 11.03%), and the averages of the submission rates for generic (mean: 1.55%; SD: 2.23%) and spear-phishing campaigns (mean: 7.11%; SD: 7.12%). Note that here the average value of the opening (or click, submission) rates is calculated as the average of each campaign's opening (or click, submission) rates. With this approach, small campaigns are weighted the same as large ones. This leads to values that are different from those generated by calculating the average rates across all delivered emails. The differences between generic and spear-phishing campaign rates are significant with $p < 0.001$.

Summary. Spear-phishing campaigns demonstrate significantly higher interaction rates compared to generic campaigns. Notably, in spear-phishing campaigns, 50% of the users who click on a link also submit their data, indicating a high level of engagement. In contrast, generic campaigns have a relatively low submission rate, underscoring the effectiveness of spear-phishing tactics.

4.3 Comparison of Economic Sectors

In this section, we compare the rates of interactions with phishing emails from organizations across different economic sectors. In 29 (30.21%) campaigns in the secondary sector, 30 488 (44.35%) emails were delivered. In 46 (47.92%) campaigns in the tertiary sector, 17 893 (26.03%) emails were delivered. In 21 (21.88%) campaigns

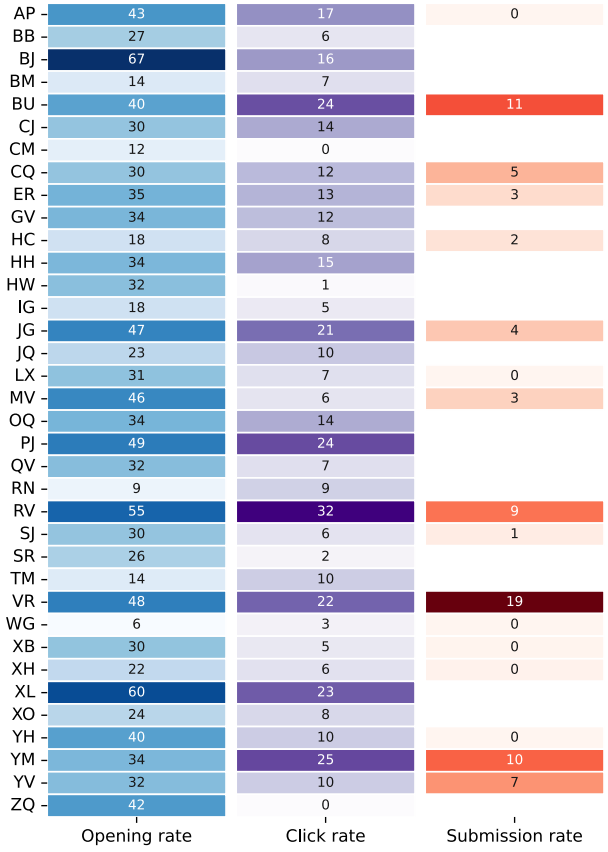

Figure 2: Heatmap of overall interaction rates (in percent) for each organization's event category. Organizations are pseudonymously listed in descending order

in the quaternary sector, 20 361 (29.62%) emails were delivered. The interaction rates are visualized in Figure 5. We compare: (i) the averages of the opening rates of the secondary sector (mean: 31.05%; SD: 11.21%), the tertiary sector (mean: 29.18%; SD: 15.13%), and the quaternary sector (mean: 40.86%; SD: 14.1%); (ii) the averages of the click rates of the secondary sector (mean: 11.78%; SD: 10.87%), the tertiary sector (mean: 8.56%; SD: 7.92%), and the quaternary sector (mean: 14.67%; SD: 9.98%); and (iii) the averages of the submission rates of the secondary sector (mean: 4.22%; SD: 6.05%), the tertiary sector (mean: 0.72%; SD: 1.48%), and the quaternary sector (mean: 7.98%; SD: 5.93%). Again, note that here the average values are calculated as the averages of the rates of each campaign, contrary to averaging over all delivered emails. The differences between the rates of economic sectors are all significant with $p < 0.001$.

Summary. Phishing susceptibility varies significantly across different economic sectors, with the secondary sector exhibiting the lowest levels of vulnerability. Specifically, the secondary sector has a click rate of 7.02% and a submission rate of 0.86%, indicating a comparatively lower likelihood of users in this sector engaging with phishing attempts.

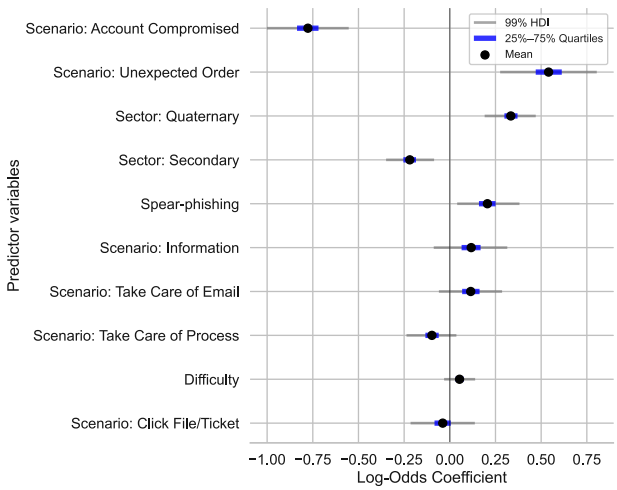

Figure 3: Distributions of log-odds coefficients for predictors. Baseline economic sector is the tertiary sector, and the baseline scenario is "Don't Miss Out".

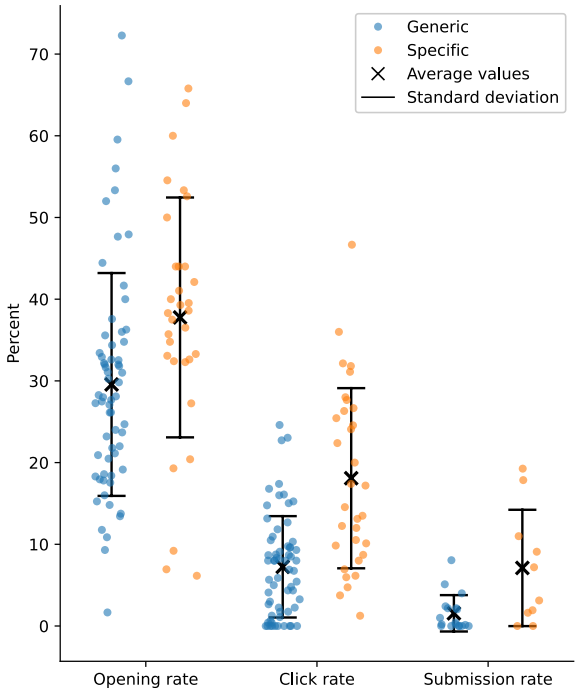

Figure 4: Comparison of interaction rates (in percent) in generic and spear-phishing campaigns

4.4 Comparison of Departments

This subsection compares the different rates of interactions with phishing emails in various departments across organizations. Of all phishing campaigns, 35 campaigns (36.46%) were conducted in a

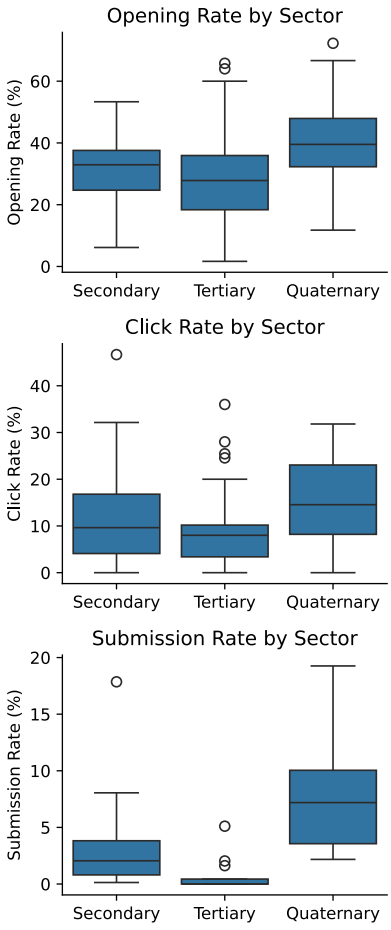

Figure 5: Interaction rates across the economic sectors.

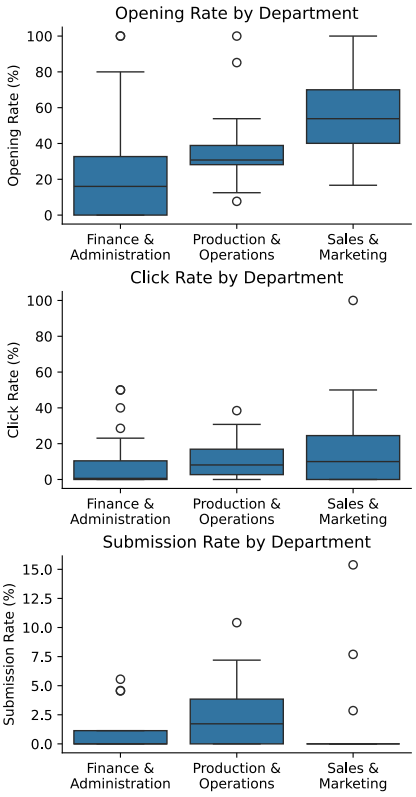

Figure 6: Interaction rates of the analyzed departments.

way that included information to compare different departments inside each organization. These 35 campaigns were distributed across 14 organizations (38.89%). In these 35 campaigns, 7 755 emails were delivered, of which 2 389 (30.83%) were opened. 720 (9.28%) users clicked on a link and 176 (2.27%) of these users submitted data. The results are visualized in Figure 6. We compare: (i) the averages of the opening rates of F&A (mean: 23.55%; SD: 28.54%), P&O (mean: 35.62%; SD: 18.05%), and S&M (mean: 55.93%; SD: 20.4%); (ii) the averages of the click rates of F&A (mean: 8.25%; SD: 14.02%), P&O (mean: 10.88%; SD: 9.89%), and S&M (mean: 16.49%; SD: 20.44%); and (iii) the averages of the submission rates of F&A (mean: 1.22%; SD: 2.23%), P&O (mean: 2.64%; SD: 3.29%), and S&M (mean: 1.99%; SD: 4.59%). Again, note that the average values are calculated as the averages of the rates of each campaign, contrary to averaging over all delivered emails. The differences between the rates of departments are all significant with $p < 0.001$, except for the difference between the click rates of F&A and P&O. If we focus on campaigns with data submission, the differences in the opening rates between F&A and S&M and between P&O and S&M are significant with $p < 0.001$, but the other differences are *not* significant.

Summary. The response to phishing emails varies notably across different departments, with Sales and Marketing being particularly susceptible. This department exhibits a high opening rate of 53.2% and a click rate of 13.7%, indicating a heightened vulnerability to phishing attacks compared to other areas within an organization.

4.5 Comparison of Phish Scale Difficulties

This section considers different difficulty levels of recognizing phishing emails according to the Phish Scale [13, 52]. The results are visualized in Figure 7. We compare: (i) the averages of the opening rates of "Least difficult" (mean: 30.06%; SD: 14.96%), "Moderately to least difficult" (mean: 25.9%; SD: 10.81%), "Moderately difficult" (mean: 35.59%; SD: 15.05%) and "Very difficult" (mean: 31.84%; SD: 15.26%); (ii) the averages of the click rates of "Least difficult" (mean: 6.51%; SD: 7.66%), "Moderately to least difficult" (mean: 6.9%; SD: 5.35%), "Moderately difficult" (mean: 11.83%; SD: 10.59%) and "Very difficult" (mean: 14.76%; SD: 9.81%). Again, note that here the average values are calculated as the averages of the rates of each campaign, contrary to averaging over all delivered emails.

The differences between the detection difficulties are all significant with $p < 0.001$, except for the difference between the click rate of "Least difficult" and "Moderately to least difficult", which does not achieve the significance level. If we focus on campaigns with data submission, our dataset lacks data for "Moderately to least difficult", so we don't have any statistical significance. The other

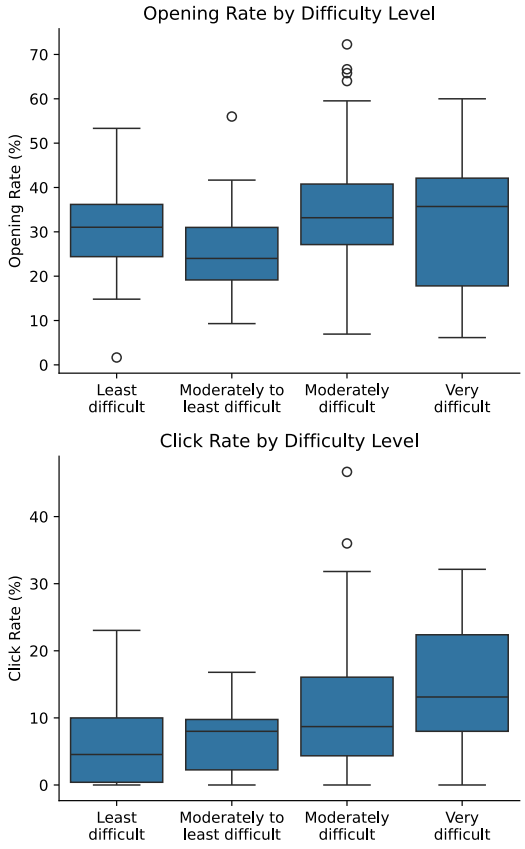

Figure 7: Opening and click rates of different difficulties.

differences in the rates are significant with $p < 0.001$, except the data submission rates between "Least difficult" and "Moderately difficult", and between "Least difficult" and "Very difficult", which fail the significance level.

Summary. Click rates tend to increase along the Phish scale, except for "moderately difficult" emails, which exhibit the lowest click rate. If an attacker gets a user to open an email classified as "moderately to least difficult", nearly half of those users click on the link, indicating a notable engagement level.

4.6 Comparison of Scenarios

In this section, we consider different categories of scenarios that were presented in the phishing emails and compare their rates of interaction. The total number of campaigns containing information about the scenario is 94, with 65 951 delivered emails. A distribution of these categories of scenarios throughout all campaigns and emails is presented in Table 4. Lastly, a comparison of click rates is visualized in Figure 4.6. Again, note that here the average values are calculated as the averages of the rates of each campaign and not averaged over all delivered emails.

Focusing solely on campaigns with data submission, nearly all rate differences are significant with $p < 0.001$, with the exception of

Table 4: Distribution of the different interaction rates for scenarios over all campaigns; the percentages are rounded.

Scenario	Campaigns	Emails
DMO	9 (10%)	18 105 (27%)
TCoP	17 (18%)	13 934 (21%)
AC	17 (18%)	10 893 (17%)
TCoE	20 (21%)	9631 (15%)
CF/T	15 (16%)	5857 (9%)
UO	7 (7%)	3939 (6%)
Information	9 (10%)	3591 (5%)

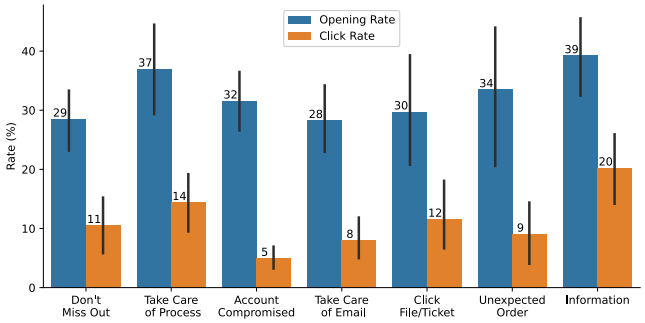

Figure 8: Average interaction rates of the campaigns.

the opening rates between "DMO" and "AC", the click rates between "TCoP" and "CF/T", and the data submission rates between "TCoE" and "CF/T". If we consider all campaigns, again, the majority of the differences are significant, except for the difference in the opening rates of "TCoP" and "AC", the click rates between (i) "TCoP" and "CF/T"; (ii) "TCoE" and "Information; (iii) "CF/T" and "Information"; and (iv) "UO" and "Information".

Summary. There are notable variations in click rates across different scenarios, ranging from 3.26% for AC to 12.49% for DMO. Once a user opens an email, they are particularly likely to click on a link if the scenario involves TCoE, with a click rate of 46.74%, or CF/T, with an even higher rate of 54.81%. These findings suggest that specific scenarios are more effective at compelling users to engage further with phishing emails after the initial interaction.

5 Discussion

As mentioned in Section 4.1, some aspects of the overall results stand out in particular. Click rates being higher in campaigns without data submission is noteworthy, for example, as these emails generally did not mention that they would request sensitive data on the landing page. This reflects the dangerous and fear-eliciting tone which emails in campaigns involving data submission often have [9], e.g., "change your password, your account has been breached". In other words, it indicates that users would rather click on a link in a harmless-looking email than a link in an email warning of a severe incident. In campaigns with data submission, the high number of users who clicked on a link and submitted data is noticeable.

This suggests that a small number of users have low cybersecurity awareness and are very susceptible to phishing emails. This minority constitutes a higher risk to an organization's cybersecurity than the majority of employees with at least average awareness of cybersecurity. However, this is not the case in all observed organizations. Indeed, the vastly varying interaction rates across different organizations reflect the considerably diverse backgrounds and experiences in cybersecurity awareness that employees of different organizations have [33, 38, 59].

Comparing just the difficulty levels "Least difficult", "Moderately to least difficult", and "Very difficult", the click rate increases with difficulty, as expected. But the click rate of "Moderately difficult" is actually lower than that of each of the other difficulties. This is likely due to the large number of campaigns with data submission in this difficulty: out of 31 014 delivered emails, 14 055 (45.32%) belonged to campaigns with data submission. The overall results indicate that campaigns with data submission have a much lower click rate than those without. Detection difficulty does not seem to have as much influence as other dimensions of phishing simulations (see Figure 3). The results for "Moderately difficult" are strongly influenced by six campaigns within the same organization and thus with a very similar user set. These campaigns comprise 17 939 emails, or around 57.8% of the emails of this difficulty. When investigating the phishing campaigns for this organization, we noticed that after the first four phishing campaigns, the click rates in the subsequent four campaigns decreased considerably. This might be due to increased awareness from earlier campaigns or even the organization's training and education of the users. Analogously, the results of the level "Very difficult" are heavily impacted by an extensive campaign in an organization with 14 182 emails, or around 70.4% of the emails of this difficulty.

It is remarkable that the click rate for "Account Compromised" is very low, even though the opening rate is high. This email scenario is strongly linked to data submission on the landing page. Hence, the low click rate likely stems from the large number of emails involved with such a campaign: out of 10 893 delivered emails for this scenario, 9 011 (82.72%) allowed for data submission. The same argument as above also holds for this scenario: The fear-eliciting tone of the emails might lead to more cautious behavior by the users. In contrast, "Don't Miss Out" has the highest click rate over all emails (see Table 2). This is probably because many campaigns were not conducted with data submission. Of the 18 105 emails delivered in this scenario, 17 723 (97.89%) emails did not allow data submission. Again, this follows the trend of users clicking more often when the email appears harmless. The same argument applies to another notable result: For the scenarios "Take Care of Email" and "Click File/Ticket", we note that around half of the users who open an email also click on a link.

With regard to different departments, the overall low click rate for Finance & Administration stands out. Contrarily, Sales & Marketing has a high opening rate and by far the highest click rate among departments, especially for campaigns without data submission. This seems natural, as these departments are usually in contact with external parties for work and must be receptive to incoming opportunities. Another explanation could be that employees from Sales & Marketing may tend to have higher levels of neuroticism or agreeableness [22, 35], which might be a correlating factor in determining whether a user interacts with a phishing email [8].

Lastly, as expected, spear-phishing interaction rates are higher than those for generic campaigns. Some spear-phishing emails used specific information about the organization to appear authentic. However, another approach that might be more effective (and which was indeed employed in some campaigns) is to make the emails more visually authentic, e.g., by using organizational themes, email footers, logos, or fonts, in combination with organizational-specific information. Finally, generic campaigns' click and submission rates with data submission are extremely low. Again, arguing as above, it would appear that fear-provoking emails lead to more cautious behavior by the users. In that case, general emails are far less convincing than spear-phishing emails and point users to a somewhat correct assessment of the situation.

5.1 Comparison with Previous Works

This section discusses the differences between our results and studies published by companies (e.g., white papers) and academia.

Comparison to Industry Reports. Vendors of phishing platforms publish their own reports on click rates in the industry. For example, KnowBe4 mentions that 33% of untrained users clicked on a phishing link [32]. StationX reports that in 2021 the average click rate for a phishing campaign was 18% for generic emails, while spear-phishing emails had a click rate of 53% [36]. This report indicates the following three industry sectors as having the highest click rates: Education (28%), Finance & Insurance (27%), and Information Technology (26%). In 2022, SoSafe reported that they analyzed 1 350 Internet users and had a click rate of 31% for this cohort [10]. SoSafe publishes these numbers annually, and the last report from 2023 also indicates an identical click rate of 31.0%. They state that they analyzed over 8 million emails from 3 000 clients [20].

Comparison to Academic Studies. Compared to other academic studies, our results depict a broader evaluation of phishing campaigns across different sectors. Williams et al. [58] report an average click rate of 19.44% for approximately 62 000 users, much higher than our average click rate over all sectors, which was 9.19%. Daengsi et al. [11] conducted a phishing simulation at a bank in Thailand and compared different departments. They found that the technology-related departments performed more securely than the social-related departments and that the cybersecurity awareness of the latter improved afterward. Rizzoni et al. [47] performed a study involving 6 000 participants in one organization over three campaigns. Their results vary from 3% for one generic campaign to 55% for one customized email (their average click rate is 20.14%). In comparison, our result for customized emails sent out to the different organizations in our research is much lower. Lain et al. [34] performed a large-scale, long-term study with more than 1 400 participants over 15 months in one organization. They sent out 117 864 phishing emails with an average click rate of 5.67%. Our analysis also supports this low number of click rates. In a passive measurement of phishing websites Oest et al. [40] found that the submission rate on phishing websites is at least 7.42%, supporting our findings from the phishing simulation campaigns conducted. Sutter et al. [54] performed another large-scale study with 31 940

participants and 288 000 emails at one university of applied sciences. They report that 27% of their observed users clicked at least once during their measurement, which spanned several campaigns. Again, their number is much higher than our overall click rate of 8.91% for all delivered emails.

In conclusion, industry reports generally indicate higher click rates than phishing reports from academic publications; for a summary, see Table 5. One reason might be that academic publications rely on only one organization as the focus of their study. On the other hand, industry reports might be biased towards higher click rates for sales and marketing purposes of the respective platforms. However, our findings indicate that, in general, lower opening, click, and submission rates across different organizations are applicable in comparison with prior work.

Table 5: Click rates and metadata about simulations performed in industry and academia. The arrows show whether the click rate is higher or lower than ours.

Reference	No.	Campaign Type	Industry	Click Rate
[36]	Many	Spear-Phishing	Aggregated	53% ↑
		Generic	Aggregated	18% ↑
[32]	Many	Aggregated	Aggregated	33% ↑
[10]	Many	Aggregated	Aggregated	31% ↑
[54]	One	Generic	Academia	27% ↑
[47]	One	Generic	Healthcare	20% ↑
[58]	One	Spear-Phishing	Public Sector	19% ↑
[34]	One	Generic	Public Sector	6% ↓

5.2 Recommendations

The results and our discussion show that phishing simulations are less effective and sometimes even counter-effective. In general, organizations and researchers should enhance technical detection measures to reduce the number of phishing emails that reach users inboxes. Moreover, a reassessment of detection difficulty metrics is recommended, as detection difficulty seems to have a limited impact on click rates. Contextual factors, like culture or employee role, might also influence rates. However, since current research is often limited to only one organization, this should be investigated on a larger scale. Furthermore, organizations must compare and evaluate factors such as the difficulty level of the phishing email, the economic sector and departmental structure of their organization, and the phishing scenarios provided by the phishing provider, as variances in these can lead to significantly different interaction rates. Therefore, we recommend that organizations evaluate the rigor of the phishing simulation provider and compare the dimensions of the phishing simulation emails and reports generated by the provider. For the research community, given the variability of the interaction rates for different organizations, general findings from single-organization studies should be avoided for other industries or companies. For example, the scenario "Account compromised" should be used with caution, as our results show that this scenario leads to low interaction rates and might produce false conclusions. In comparison, a harmless-seeming email that shares information might lead to very different results. Furthermore, different departments might exhibit varying levels of awareness and require different training approaches. Over-reliance on simplified phishing simulations should be avoided by using varying detection difficulty levels and data submission requests to prepare and test for diverse attack vectors.

6 Limitations

In this section, we discuss the limitations of the approach we used. We discuss threats to validity, how we mitigated them, and ethical considerations when conducting such a study. Further, we also clarify which protective measures we took to safeguard our data.

Threats to Validity. Our work generally does not focus on "why" users click on a phishing mail. Instead, we performed a broad, quantifiable study that does not have the specific shortcomings that comparable literature has, in particular (1) focusing only on one organization for (2) a short period of time, and (3) employing only a small number of campaigns. We wanted to ensure that our study's results apply to a variety of companies. While performing our experiments, our approach also had some shortcomings. The emails sent are categorized as "Open" and "Clicked". However, while conducting our research, we found that some campaigns might have inflated "Opened" rates, as some protection tools might load the tracking pixel automatically. We can only be sure that a human clicked the link if data was submitted at a later stage. On the other hand, a protection mechanism might also be that external content (e.g., images) is blocked by the security configurations of the organization, which possibly leads to a decrease in openings. In the category "Clicked" there are also cases where a software "clicked" on a link. However, the data was cleared for this, as software that clicks on the links in most delivered emails is distinguishable from users clicking. This was done by analyzing the combinations of IP address and *user-agent* string. Additionally, in our case, the software's user-agent was more outdated than the user-agents of the human users. Thus, we removed these automatic interactions prior to analysis. Admittedly, we cannot evaluate whether users actually read the phishing email, so we can only make assumptions about how an email is treated in a user's inbox. Merely opening an email generally results in little risk of a successful phishing attack (i.e., credentials being stolen), provided that the software in use is not significantly outdated. Thus, in this study, we focused on click and data submission rates. Nevertheless, the potential for opening an email to introduce an additional attack vector should not be disregarded. Ensuring that users' software is up-to-date may potentially mitigate the risks associated with clicking on malicious links. However, in practice there are three critical reasons to avoid clicks altogether: (i) the existence of zero-day vulnerabilities, (ii) the frequent lack of timely software updates, and (iii) the increased value of email addresses from individuals who click on links on the black market, which may render them specific targets for future attacks. One issue with using the Phish Scale for determining detection difficulties is its process of categorizing emails. Counting the cues to get a cues category for the email can be done in different ways and may lead to ambiguous results; e.g., when counting distracting details in the email, does the rater count every sentence or count the individual segments of sentences? Due to this ambiguity, checking the inter-rater reliability to obtain scientifically sound detection difficulties is necessary. We ensured this by achieving a reliable

Cohen's Kappa coefficient of > 0.6. Future work should explore the distinction between generic and spear-phishing attacks in more detail, especially considering how users interact with the different categories of fraudulent emails. Another future direction is to identify structured training that reduces phishing susceptibility and protects organizations in this complex landscape. Analyzing the effect different trainings have in various sectors could also generate insight into the efficacy of phishing simulations as a training tool.

Ethics. The clients acknowledged the phishing campaigns and pseudonymous data analysis with the partner company. During the study, we never had access to any PII and were only given access to anonymized and aggregated data after collection. According to our institution's guidelines, the analysis of anonymized data does not require IRB approval, so we did not submit a formal request. Organizations sometimes notified users of the simulated phishing campaigns, i.e., the participants knew that there was a simulated phishing campaign coming, but did not know the exact timing or content of the messages. As part of this study, participants were subject to minimal risk. In fact, they are exposed to greater risk as part of their everyday work, as they receive real phishing emails and other malicious emails on a regular basis [16]. This was corroborated by Lain et al. [34] who stated in their study that phishing simulations do not expose participants to a greater risk than what they would encounter during their typical days in the office because the participants are regularly exposed to real phishing emails and spam. However, phishing simulations themselves can have negative effects. For example, they waste employees' limited time or sow distrust towards the company. To mitigate this, our research was conducted as part of the respective organizations' broader cybersecurity awareness programs. Yet, we acknowledge that conducting these simulations exposed the users to minimal risks, but similar to other researchers, we believe that the participants' positive experiences merited these risks. Combining these arguments, we argue that conducting the simulations in a considerate way is valid and these arguments do not pose a threat to validity [45].

Data Collection and Protection. We decided to report phishing campaigns with data submission separately from those without data submission to avoid the potential for wrong conclusions to arise. We did so primarily for two reasons. First, the ratio of users submitting data compared to the number of emails delivered – including campaigns without data submission – gives the wrong impression that very few users submit their credentials. Second, comparing click rate with the associated submission rate is also insightful. This produces a "conversion rate" quantifying how many users who clicked a link also submit credentials. This conversion rate would be skewed when comparing a submission rate with the click rate of all delivered emails, which included campaigns without any data submission at all. During the phishing simulations, data on clicks, email openings, and data submissions on a phishing site were collected. If a study participant entered PII, e.g., his email or password, a record was created that the data was submitted, but not which data exactly. In theory, a user could enter "wrong" credentials at this point. However, we assume that most users will enter their correct credentials instead. The dataset for each campaign is stored in the EU at an ISO 27001-compliant company.

7 Related Work

Much work on phishing simulations and their effectiveness has already been done. However, these previous studies all are smaller in terms of the total number of study participants and included sectors, or they had different analysis goals. Burda et al. [5] analyzed susceptibility to phishing for two different organizational types (a university and a consultancy). They conclude that an overarching study across various domains is needed. Burns et al. [6] choose a residential MBA program for their phishing simulation. They also conclude that future research should focus on generalizing results from phishing campaigns to draw overarching conclusions. Rizzoni et al. analyzed a phishing simulation exercise at a large hospital. They found that customization of phishing emails leads to more people opening a potential phishing email [47]. This study verifies the work of Jalali et al. [28], who also found that hospital employees were particularly susceptible to phishing. They add that the high workload generally leads to a greater susceptibility. Ho et al. [24] had a dataset that contained 92 organizations and 113 083 695 unique, employee-sent emails. However, they focused on lateral phishing between different organizations. There are also other approaches to understanding and interpreting phishing simulations, for example, role-playing. Sheng et al. [50] used one such role-playing approach to analyze demographic effects on users' susceptibility to clicking on phishing emails. Our findings align with the recent work of Ho et al. [25], who found that annual security training does not affect whether a user clicks on links in simulated phishing emails. Having examined the related works, we conclude that our study presents the first longitudinal and intersectional large-scale study of the phishing phenomenon, which several previous and related works identified as a current research gap.

8 Conclusion

We conducted a quantitative, large-scale, and intersectional study on phishing simulations in the workplace. Our findings suggest that the prevalent body of research on phishing metrics is skewed with quantitative results on click and submission rates that are not generalizable. We found that the industry reports even higher click rates than researchers, which makes existing figures even more unreliable. Our findings indicate that phishing is still a prevalent problem, though it seems to be inflated to a certain extent.

Acknowledgments

The authors gratefully acknowledge funding from the *Federal Ministry of Education and Research* (grant 16KIS1648 "DigiFit"). The authors extend their appreciation to the partner company AWARE7 GmbH for making the analyzed data available. Further, the authors thank Brett Ellis for his support in creating the final version of this work.

References

[1] Atalay Atasu, Charles J Corbett, Ximin Huang, and L Beril Toktay. 2020. Sustainable operations management through the perspective of manufacturing & service operations management. *Manufacturing & service operations management* 22, 1 (2020), 146–157.

[2] Henk Berkman, Jonathan Jona, Gladys Lee, and Naomi Soderstrom. 2018. Cybersecurity awareness and market valuations. *Journal of Accounting and Public Policy* 37, 6 (2018), 508–526.

[3] Wim Biemans, Avinash Malshe, and Jeff S Johnson. 2022. The sales-marketing interface: A systematic literature review and directions for future research. *Industrial Marketing Management* 102 (2022), 324–337.

[4] Lina Brunken, Annalina Buckmann, Jonas Hielscher, and M Angela Sasse. 2023. "To" Do This Properly, You Need More Resources: The Hidden Costs of Introducing Simulated Phishing Campaigns. In *32nd USENIX Security Symposium (USENIX Security 23)*. USENIX Association, Anaheim, CA, 4105–4122. https://www.usenix.org/conference/usenixsecurity23/presentation/brunken

[5] Pavlo Burda, Tzouliano Chotza, Luca Allodi, and Nicola Zannone. 2020. Testing the effectiveness of tailored phishing techniques in industry and academia: a field experiment. In *Proceedings of the 15th International Conference on Availability, Reliability and Security* (Virtual Event, Ireland) *(ARES '20)*. Association for Computing Machinery, New York, NY, USA, Article 3, 10 pages. doi:10.1145/3407023.3409178

[6] AJ Burns, M Eric Johnson, and Deanna D Caputo. 2019. Spear phishing in a barrel: Insights from a targeted phishing campaign. *Journal of Organizational Computing and Electronic Commerce* 29, 1 (2019), 24–39.

[7] Sunil Chaudhary, Vasileios Gkioulos, and Sokratis Katsikas. 2022. Developing metrics to assess the effectiveness of cybersecurity awareness program. *Journal of Cybersecurity* 8, 1 (2022), tyac006.

[8] Jin-Hee Cho, Hasan Cam, and Alessandro Oltramari. 2016. Effect of personality traits on trust and risk to phishing vulnerability: Modeling and analysis. In *2016 IEEE International Multi-Disciplinary Conference on Cognitive Methods in Situation Awareness and Decision Support (CogSIMA)*. IEEE Operations Center, 445 Hoes Lane, Piscataway, NJ 08854, 7–13. doi:10.1109/COGSIMA.2016.7497779

[9] Anargyros Chrysanthou, Yorgos Pantis, and Constantinos Patsakis. 2024. The anatomy of deception: Measuring technical and human factors of a large-scale phishing campaign. *Computers & Security* 140 (2024), 103780.

[10] Julia Mutzbauer CSO Online. 2022. Digital Natives sind anfälliger für Phishing. https://www.csoonline.com/de/a/digital-natives-sind-anfaelliger-fuer-phishing,3674227#:~:text=31%20Prozent%20der%20Teilnehmer%20haben,drei%20Phishing%2DAttacken%20erfolgreich%20gewesen. Accessed: 2024-07-02.

[11] Therdpong Daengsi, Pongpisit Wuttidittachotti, Phisit Pornpongtechavanich, and Nathaporn Utakrit. 2021. A Comparative Study of Cybersecurity Awareness on Phishing Among Employees from Different Departments in an Organization. In *2021 2nd International Conference on Smart Computing and Electronic Enterprise (ICSCEE)*. IEEE Operations Center, 445 Hoes Lane, Piscataway, NJ 08854, 102–106. doi:10.1109/ICSCEE50312.2021.9498208

[12] Erik Daguerre, Brandan Geise, Jeff McCutchan, and Spencer McIntyre. 2024. King Phisher. https://github.com/rsmusllp/king-phisher. Accessed: 2024-04-19.

[13] Shanee Dawkins and Jody Jacobs. 2023. NIST Phish Scale User Guide. doi:10.6028/NIST.TN.2276

[14] Yadolah Dodge. 2008. *The Concise Encyclopedia of Statistics*. Springer New York, New York, NY. 165–169 pages. doi:10.1007/978-0-387-32833-1_36

[15] European Commission, Directorate-General for Internal Market, Industry, Entrepreneurship and SMEs. 2024. SME Definition. https://single-market-economy.ec.europa.eu/smes/sme-fundamentals/sme-definition_en.

[16] Peter Finn and Markus Jakobsson. 2007. Designing ethical phishing experiments. *IEEE Technology and Society Magazine* 26, 1 (2007), 46–58.

[17] Allan GB Fisher. 1939. Production, Primary, Secondary and Tertiary. *Economic record* 15, 1 (1939), 24–38.

[18] Brett Fouss, Dennis M Ross, Allan B Wollaber, and Steven R Gomez. 2019. PunyVis: A Visual Analytics Approach for Identifying Homograph Phishing Attacks. In *2019 IEEE Symposium on Visualization for Cyber Security (VizSec)*. IEEE, IEEE Operations Center, 445 Hoes Lane, Piscataway, NJ 08854, 1–10.

[19] Yona Friedman. 1981. The quaternary sector. *Human Systems Management* 2, 1 (1981), 44–52.

[20] SoSafe GmbH. 2023. Human Risk Review 2023. https://lp.sosafe.de/hubfs/SoSafe%20-%20Human%20Risk%20Review%202023%20-%20DE.pdf Accessed: 2024-07-02.

[21] Srishti Gupta and Ponnurangam Kumaraguru. 2014. Emerging phishing trends and effectiveness of the anti-phishing landing page. In *2014 APWG Symposium on Electronic Crime Research (eCrime)*. IEEE, Piscataway, NJ, 36–47. doi:10.1109/ECRIME.2014.6963163

[22] Johannes Habel, Selma Kadić-Maglajlić, Nathaniel N Hartmann, Ad de Jong, Nicolas A Zacharias, and Fabian Kosse. 2024. Neuroticism and the sales profession. *Organizational Behavior and Human Decision Processes* 184 (2024), 104353.

[23] Jonas Hielscher, Uta Menges, Simon Parkin, Annette Kluge, and M. Angela Sasse. 2023. "Employees Who Don't Accept the Time Security Takes Are Not Aware Enough": The CISO View of Human-Centred Security. In *32nd USENIX Security Symposium (USENIX Security 23)*. USENIX Association, Anaheim, CA, 2311–2328. https://www.usenix.org/conference/usenixsecurity23/presentation/hielscher

[24] Grant Ho, Asaf Cidon, Lior Gavish, Marco Schweighauser, Vern Paxson, Stefan Savage, Geoffrey M. Voelker, and David Wagner. 2019. Detecting and Characterizing Lateral Phishing at Scale. In *28th USENIX Security Symposium (USENIX Security 19)*. USENIX Association, Santa Clara, CA, 1273–1290. https://www.usenix.org/conference/usenixsecurity19/presentation/ho

[25] Grant Ho, Ariana Mirian, Elisa Luo, Khang Tong, Euyhyun Lee, Lin Liu, Christopher A. Longhurst, Christian Dameff, Stefan Savage, and Geoffrey M. Voelker. 2025. Understanding the Efficacy of Phishing Training in Practice . In *2025 IEEE Symposium on Security and Privacy (SP)*. IEEE Computer Society, Los Alamitos, CA, USA, 76–76. doi:10.1109/SP61157.2025.00076

[26] Hang Hu, Peng Peng, and Gang Wang. 2018. Towards understanding the adoption of anti-spoofing protocols in email systems. In *2018 IEEE Cybersecurity Development (SecDev)*. IEEE, Piscataway, NJ, 94–101.

[27] ISO Central Secretary. 2016. *Information technology - Security techniques - Information security management - Monitoring, measurement, analysis and evaluation*. Standard. International Organization for Standardization, Geneva, CH.

[28] Mohammad S Jalali, Maike Bruckes, Daniel Westmattelmann, and Gerhard Schewe. 2020. Why employees (still) click on phishing links: investigation in hospitals. *Journal of medical Internet research* 22, 1 (2020), e16775.

[29] Alicia A Johnson, Miles Q Ott, and Mine Dogucu. 2022. *Bayes rules!* Chapman & Hall/CRC, Philadelphia, PA.

[30] Zoltan Kenessey. 1987. THE PRIMARY, SECONDARY, TERTIARY AND QUATERNARY SECTORS OF THE ECONOMY. *Review of Income and Wealth* 33, 4 (Dec. 1987), 359–385. doi:10.1111/j.1475-4991.1987.tb00680.x

[31] David Kennedy. 2024. Social Engineer Toolkit. https://github.com/trustedsec/social-engineer-toolkit. Accessed: 2024-04-19.

[32] KnowBe4, Inc. 2023. Phishing by Industry Benchmark Report. https://blog.knowbe4.com/knowbe4-2023-phishing-by-industry-benchmarking-report. Accessed: 2024-07-02.

[33] Anastassija Kostan, Sara Olschar, Lucy Simko, and Yasemin Acar. 2024. Exploring digital security and privacy in relative poverty in Germany through qualitative interviews. In *33rd USENIX Security Symposium (USENIX Security 24)*. USENIX Association, Philadelphia, PA, 2029–2046. https://www.usenix.org/conference/usenixsecurity24/presentation/kostan

[34] Daniele Lain, Kari Kostiainen, and Srdjan Čapkun. 2022. Phishing in Organizations: Findings from a Large-Scale and Long-Term Study. In *2022 IEEE Symposium on Security and Privacy (SP)*. IEEE, Piscataway, NJ, 842–859. doi:10.1109/sp46214.2022.9833766

[35] John W Lounsbury, Nancy A Foster, Jacob J Levy, and Lucy W Gibson. 2014. Key personality traits of sales managers. *Work* 48, 2 (2014), 239–253.

[36] STATIONX LTD. 2024. Top Phishing Statistics for 2024: Latest Figures and Trends. https://www.stationx.net/phishing-statistics/#:~:text=Global%20Average%20Phishing%20Email%20Click,average%20click%20rate%20of%2053.2%25. Accessed: 2024-07-02.

[37] Sourena Maroofi, Maciej Korczynski, and Andrzej Duda. 2020. From Defensive Registration to Subdomain Protection: Evaluation of Email Anti-Spoofing Schemes for High-Profile Domains. In *Proc. Network Traffic Measurement and Analysis Conference (TMA)*. IFIP, Modling, Austria, 9 pages.

[38] Xenia Mountrouidou, David Vosen, Chadi Kari, Mohammad Q. Azhar, Sajal Bhatia, Greg Gagne, Joseph Maguire, Liviana Tudor, and Timothy T. Yuen. 2019. Securing the Human: A Review of Literature on Broadening Diversity in Cybersecurity Education. In *Proceedings of the Working Group Reports on Innovation and Technology in Computer Science Education* (Aberdeen, Scotland Uk) *(ITiCSE-WGR '19)*. Association for Computing Machinery, New York, NY, USA, 157–176. doi:10.1145/3344429.3372507

[39] Joseph K Nwankpa and Pratim Milton Datta. 2023. Remote vigilance: The roles of cyber awareness and cybersecurity policies among remote workers. *Computers & Security* 130 (2023), 103266.

[40] Adam Oest, Penghui Zhang, Brad Wardman, Eric Nunes, Jakub Burgis, Ali Zand, Kurt Thomas, Adam Doupé, and Gail-Joon Ahn. 2020. Sunrise to Sunset: Analyzing the End-to-end Life Cycle and Effectiveness of Phishing Attacks at Scale. In *29th USENIX Security Symposium (USENIX Security 20)*. USENIX Association, Anaheim, CA, 361–377. https://www.usenix.org/conference/usenixsecurity20/presentation/oest-sunrise

[41] Alhassan Ohiomah, Morad Benyoucef, and Pavel Andreev. 2020. A multidimensional perspective of business-to-business sales success: A meta-analytic review. *Industrial Marketing Management* 90 (2020), 435–452.

[42] Andrew Owens. 2013. Improving the Performance of Finance and Accounting Shared Service Centres. *Journal of payments strategy & systems* 7, 3 (2013), 250–261.

[43] Ryan L Phillips and Rita Ormsby. 2016. Industry classification schemes: An analysis and review. *Journal of Business & Finance Librarianship* 21, 1 (2016), 1–25.

[44] Paul Prasse, Christoph Sawade, Niels Landwehr, and Tobias Scheffer. 2015. Learning to identify concise regular expressions that describe email campaigns. *The Journal of Machine Learning Research* 16, 1 (2015), 3687–3720.

[45] David B. Resnik. 2024. *The Ethics of Research with Human Subjects: Protecting People, Advancing Science, Promoting Trust*. Springer Nature Switzerland, Cham. doi:10.1007/978-3-031-82757-0

[46] Demas Muhammad Rijal, Mukhamad Fahmi Assydiqi, Yoel Rensisko Prasetya, Lidya Nurhapsari Prasetya Ningsih, Nisya Kayla Putri Anindra, and Pandu Dwi Luhur Pambudi. 2024. Information Security Awareness Analysis of the Threat of Data Leakage in Educational Institutions with the ISO 27001 Framework. *Journal of Digital Business and Innovation Management* 3, 1 (2024), 36–52.

[47] Fabio Rizzoni, Sabina Magalini, Alessandra Casaroli, Pasquale Mari, Matt Dixon, and Lynne Coventry. 2022. Phishing simulation exercise in a large hospital: A

case study. *Digital Health* 8 (2022), 20552076221081716.
[48] Rohani Rohan, Debajyoti Pal, Jari Hautamäki, Suree Funilkul, Wichian Chutimaskul, and Himanshu Thapliyal. 2023. A systematic literature review of cybersecurity scales assessing information security awareness. *Heliyon* 9, 3 (2023), e14234. doi:10.1016/j.heliyon.2023.e14234
[49] Manuel Sánchez-Paniagua, Eduardo Fidalgo Fernández, Enrique Alegre, Wesam Al-Nabki, and Victor Gonzalez-Castro. 2022. Phishing URL detection: A real-case scenario through login URLs. *IEEE Access* 10 (2022), 42949–42960.
[50] Steve Sheng, Mandy Holbrook, Ponnurangam Kumaraguru, Lorrie Faith Cranor, and Julie Downs. 2010. Who falls for phish? a demographic analysis of phishing susceptibility and effectiveness of interventions. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Atlanta, Georgia, USA) *(CHI '10)*. Association for Computing Machinery, New York, NY, USA, 373–382. doi:10.1145/1753326.1753383
[51] C Singh. 2023. The European Approach to Cybersecurity in 2023: A Review of the Changes Brought in By the Network and Information Security 2 (NIS2) Directive 2022/2555. *International Company and Commercial Law Review* 5 (2023), 251–261.
[52] Michelle Steves, Kristen Greene, and Mary Theofanos. 2020. Categorizing human phishing difficulty: a Phish Scale. *Journal of Cybersecurity* 6, 1 (09 2020), tyaa009. doi:10.1093/cybsec/tyaa009 arXiv:https://academic.oup.com/cybersecurity/article-pdf/6/1/tyaa009/33746006/tyaa009.pdf
[53] Salvatore J Stolfo, Eleazar Eskin, Shlomo Herskop, and Manasi Bhattacharyya. 2010. System and methods for detecting malicious email transmission. US Patent 7,657,935.
[54] Thomas Sutter, Ahmet Selman Bozkir, Benjamin Gehring, and Peter Berlich. 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 (2022), 100540–100565.
[55] The European Parliament and the Council of the European Union. 2022. Directive (EU) 2022/2555 of the European Parliament and of the Council of 14 December 2022 on measures for a high common level of cybersecurity across the Union, amending Regulation (EU) No 910/2014 and Directive (EU) 2018/1972, and repealing Directive (EU) 2016/1148 (NIS 2 Directive) (Text with EEA relevance).
[56] Kamila Turečková and Stanislav Martinát. 2015. *Quaternary sector and extended sectoral structure of the economy in the selected European countries.* Working Papers 0010. Silesian University, School of Business Administration. https://ideas.repec.org/p/opa/wpaper/0010.html
[57] Tobias Urban, Matteo Große-Kampmann, Dennis Tatang, Thorsten Holz, and Norbert Pohlmann. 2020. Plenty of Phish in the Sea: Analyzing Potential Pre-attack Surfaces. In *Computer Security – ESORICS 2020*, Liqun Chen, Ninghui Li, Kaitai Liang, and Steve Schneider (Eds.). Springer International Publishing, Cham, 272–291.
[58] Emma J Williams, Joanne Hinds, and Adam N Joinson. 2018. Exploring susceptibility to phishing in the workplace. *International Journal of Human-Computer Studies* 120 (2018), 1–13.
[59] Sarunyoo Wongkrachang. 2023. Cybersecurity awareness and training programs for racial and sexual minority populations: An examination of effectiveness and best practices. *Contemporary Issues in Behavioral and Social Sciences* 7, 1 (2023), 35–53.
[60] Jordan Wright. 2024. GoPhish. https://getgophish.com/. Accessed: 2024-04-19.
[61] Aiping Xiong, Sian Lee, Zekun Cai, Ephraim Govere, and Harish Kolla. 2023. You Received an Email from Your Advisor? A Case Study of Phishing Scams in a University Setting.

A Example of a Spear Phishing Email Used

Figure 9 provides an example of a spear phishing email used for the scenario "Take Care of Process." Personal information has been redacted, and the email features a fictitious company and logo.

Figure 9: A spear phishing email we used.

B Overview of Surveyed Organizations

Table 6 provides an overview of the pseudonymous organizations we surveyed in our research. We use the United Nations International Classification of All Economic Activities system (ISIC) to classify the industry. Using this classifier allows a majority of organizations to start benchmarking, as many countries classify their economic activities using a comparable system to ISIC or have even adopted ISIC as their national classification system [43].

Table 6: This table provides an overview of the observed organizations. Sector describes primary (1), secondary (2), tertiary (3), and quaternary (4). "Consulting, IT services" in our table corresponds to "Consulting & telecommunications, computer programming, consultancy, computing infrastructure, and other information service activities" in the ISIC classifier scheme. BST is the balance sheet total. "n.a." in the BST column means that the organization is not legally required to publish a yearly financial report. Thus, in these cases, we were not able to determine the BST.

ID	Sector	Industry	ISIC Classifier	ISIC Code	Organization Size	BST (Million €)	Country
HW	4	Finance	Financial and insurance activities	L6433	Medium	31	DE
YM	2	Industry	Manufacturing	C26	Small	11	DE
JG	4	Consulting	Consulting, IT services	K6219	Small	5	DE
CM	4	Consulting	Other service activities	T9411	Small	n.a.	DE
AP	3	Security	Financial and insurance activities	L6419	Medium	13	DE
QV	4	IT	Publishing and content distribution	J582	Medium	1.6	DE
OQ	4	Education	Education	Q8540	Large	n.a.	DE
YH	3	Industry	Manufacturing	C26	Medium	1.8	DE
SJ	2	Industry	Manufacturing	C2511	Large	332	DE
SR	4	Education	Manufacturing	Q8540	Large	n.a.	DE
MV	3	Real Estate	Real estate activities	M68	Large	3 400	DE
RN	3	Social	Human health and social work activities	R8890	Small	n.a.	DE
BU	4	IT	Publishing and content distribution	J582	Large	58.3	DE
XH	3	Gambling	Arts, sports and recreation	S9200	Large	n.a.	DE
BJ	4	Consulting	Education	Q8559	Medium	4	DE
TM	3	Healthcare	Human health and social work activities	R86	Large	429	DE
HH	2	Chemical Industry	Electricity, gas, steam, air conditioning	D3520	Large	1 700	EU
WG	3	Automobile	Wholesale and retail trade	G4781	Large	109	DE
JQ	2	Industry	Manufacturing	C2395	Medium	24.7	DE
GV	3	Culture	Arts, sports and recreation	S9020	Medium	7.2	DE
VR	4	Consulting	Consulting, IT services	K6219	Medium	9	DE
ER	4	Consulting	Consulting, IT services	K6219	Medium	12.9	DE
PJ	2	Chemical Industry	Manufacturing	C201	Medium	17.1	DE
IG	2	Food	Manufacturing	C1050	Large	69	DE
CQ	2	Healthcare	Manufacturing	C2100	Medium	37	DE
XB	2	Industry	Wholesale and retail trade	G4659	Large	300	DE
ZQ	4	IT	Wholesale and retail trade	K6219	Small	1.6	DE
BB	3	Municipality	Public administration	P8411	Large	n.a.	DE
BM	3	Municipality	Public administration	P8411	Medium	n.a.	DE
LX	3	Water Supply	Water supply and waste management	E3600	Small	15.5	DE
XO	4	Consulting	Other service activities	T9510	Large	237	JP
RV	4	Consulting	Consulting, IT services	K6219	Small	1.1	DE
CJ	2	Environment	Manufacturing	C2651	Large	34.2	DE
YV	4	Research	Scientific and technical activities	N7210	Large	n.a.	DE
XL	4	Consulting	Consulting, IT services	K6219	Medium	9	DE
HC	3	Security	Financial and insurance activities	L6419	Large	75.5	EU