The Impact of Emerging Phishing Threats: Assessing Quishing and LLM-generated Phishing Emails against Organizations

Marie Weinz
marie.weinz@uni.li
University of Liechtenstein
Vaduz, Liechtenstein

Nicola Zannone
n.zannone@tue.nl
Eindhoven University of Technology
Eindhoven, Netherlands

Luca Allodi
l.allodi@tue.nl
Eindhoven University of Technology
Eindhoven, Netherlands

Giovanni Apruzzese
giovanni.apruzzese@uni.li
University of Liechtenstein
Vaduz, Liechtenstein

Abstract

Modern organizations are persistently targeted by phishing emails. Despite advances in detection systems and widespread employee training, attackers continue to innovate, posing ongoing threats. Two emerging vectors stand out in the current landscape: QR-code baits and LLM-enabled pretexting. Yet, little is known about the effectiveness of current defenses against these attacks, particularly when it comes to real-world impact on employees. This gap leaves uncertainty around to what extent related countermeasures are justified or needed. Our work addresses this issue.

We conduct three phishing simulations across organizations of varying sizes—from small-medium businesses to a multinational enterprise. In total, we send over 71k emails targeting employees, including: a "traditional" phishing email with a click-through button; a nearly-identical "quishing" email with a QR code instead; and a phishing email written with the assistance of an LLM and open-source intelligence. Our results show that quishing emails have the same effectiveness as traditional phishing emails at luring users to the landing webpage—which is worrying, given that quishing emails are much harder to identify even by operational detectors. We also find that LLMs can be very good "social engineers": in one company, over 30% of the emails opened led to visiting the landing webpage—a rate exceeding some prior benchmarks. Finally, we complement our study by conducting a survey across the organizations' employees, measuring their "perceived" phishing awareness. Our findings suggest a correlation between higher self-reported awareness and organizational resilience to phishing attempts.

CCS Concepts

• **Security and privacy** → **Phishing**.

Keywords

phishing, quishing, chatgpt, education, user study, awareness, email

ACM Reference Format:

Marie Weinz, Nicola Zannone, Luca Allodi, and Giovanni Apruzzese. 2025. The Impact of Emerging Phishing Threats: Assessing Quishing and LLM-generated Phishing Emails against Organizations. In *ACM Asia Conference on Computer and Communications Security (ASIA CCS '25), August 25–29, 2025, Hanoi, Vietnam.* ACM, New York, NY, USA, 20 pages. https://doi.org/10.1145/3708821.3736195

1 Introduction

Phishing is endemic in the threat landscape of modern organizations. According to Proofpoint's 2024 State of the Phish report [10], among the top-5 most prevalent attacks suffered by organizations, four revolved around phishing. Worryingly, over 66 million business-email-compromise attacks have been detected *every month* in 2023, a finding that is attributed to recent advances in generative artificial intelligence (AI). The report also stated that even though nearly every company implements some form of phishing education, only 23% educate on generative AI safety—and only half (53%) provide training for everyone, leading to "gaps" exploitable by phishers.

Phishing remains a persistent and evolving threat, despite having been studied for decades [22, 32]. In response, the research community and industry have developed a variety of countermeasures, including phishing email detectors [62, 80, 98], as well as simulation, training, and educational campaigns [48, 59, 60, 84]. Yet, attackers continue to bypass these defenses and achieve their objectives. This ongoing challenge stems from the absence of a universal solution: phishing exploits both the technical limitations of automated systems and the cognitive biases of human users. Recent developments further complicate the landscape. Adversaries are increasingly using large language models (LLMs) to generate convincing phishing content [10, 25, 34, 92], and are leveraging QR codes as a novel delivery mechanism for phishing attacks [4, 11, 15]. Addressing this threat requires continuous monitoring of emerging techniques and the development of adaptive, targeted defenses.

Unfortunately, development and implementation of such countermeasures are progressing at a slow pace. Take "quishing" emails, for instance: despite some works proposing ways to detect malicious QR-codes in emails (e.g., [97, 106]), quishing emails still evade most "automated" filters—and attackers are aware of this. Recent reports by Cisco Talos [5] state that over 60% of the emails containing a QR code are not benign—a trend confirmed also by other recent reports [11]. At the same time, despite the increasing reliance on

LLMs by phishers [25, 34, 92], there is still a lack of education on how to spot (deceitful) LLM-written content [10].

We argue that such deficiencies are due to an overall poor understanding of such emerging phishing threats (discussed in §2). However, such technologies are now becoming an important asset for professional phishers [10]. Therefore, it is necessary to analyze the effectiveness of these emerging phishing threats against employees, and how this might be expected to vary (or not) across companies. This paper seeks to fulfill this gap.

To achieve our goal, we first find an agreement with three distinct companies: doing so enables us to gauge the extent to which our findings hold across different companies. Then, we carry out phishing simulations focusing on investigating: *(i)* the effectiveness of QR-code phishing emails compared to traditional click-through phishing emails; and *(ii)* the effectiveness of LLM-based phishing emails that have been fed with open-source intelligence (OSINT) information. Finally, we carry out a complementary investigation on the relationship between *perceived phishing awareness* and *actual phishing susceptibility* across our considered companies' employees. Connecting all three aspects (perceived phishing awareness, actual phishing susceptibility, diverse companies) is of great value for research, since it would enrich the findings of all prior work that investigated only one or two of these aspects in an organizational context (e.g., [40, 60, 78, 103]), as well as provide practical suggestions for real organizations (e.g., if there is a correlation between perceived phishing awareness and actual phishing susceptibility, then one can be a predictor of the other) or providers of security services (e.g., prioritizing the development of automated countermeasures).

Contributions and Findings. We provide factual data on how a diverse set of real-world companies deal with, and are affected by, the multi-faceted threat of phishing emails. We carry out a *large-scale and fine-grained assessment* of employees' susceptibility to three types of targeted-phishing emails reflecting emerging trends. Our sample (71 309 total emails sent) refers to companies of *different business size* (respectively: <250, ≈1 500, >30 000 employees). Our setup enables comparison of the results across companies.

- We scrutinize whether **QR-based phishing emails** are more (or less) effective than traditional button-based phishing emails at luring users to a malicious webpage. We find *no statistically significant* (p=.552) difference, and a TOST test of equivalence confirms that the two groups are indistinguishable at a tolerance level of less than 1%. This apparently counterintuitive result (scanning a QR code in an email is not as straightforward as clicking on a button) has unfortunately several concerning security implications (described in §2.1.1, and verified by an experiment).
- We study the effectiveness of **combining a LLM with OSINT** to craft targeted phishing emails. Our assessment reveals that even employees with prior phishing training can be deceived using freely available AI tools and public information. For instance, for the second company, ≈10% of the recipients (n=589) submitted their credentials, and ≈21% visited the webpage.
- Through an informed survey with a subset of our companies' employees (n=131), we measure their degree of **perceived phishing awareness** (PPA). We then cross-analyse the PPA of each company with the overall effectiveness of our phishing simulations. We find that the PPA can be a statistically significant (p=.044) predictor of the effectiveness of a phishing campaign.

Finally, we provide all our fine-grained results in Appendix E E. Such details are not only important for transparency, but are also useful for benchmarking and comparative purposes.

2 Background and Motivation

Numerous reports from cybersecurity agencies underscore the impact of phishing, and particularly phishing emails,[1] on modern organizations [10, 11, 20, 34]. Below, we summarize the risks posed by two emerging phishing-email threats in corporate contexts (§2.1), and then highlight the research gap that we aim to fill (§2.2).

2.1 Emerging Trends in Phishing Emails

We motivate the problem tackled in our paper by outlining the properties of QR-code phishing (§2.1.1) and LLM-based phishing (§2.1.2), explaining why they are particularly problematic today.

2.1.1 *Quishing: Social Engineering via QR codes*

QR-code phishing, also referred to as "Quishing", is a form of social engineering attack which attempts to deceive individuals into scanning QR codes that point to a malicious website [18, 90]. In some cases, such malicious QR codes are spread in the real world (e.g., physically glued to objects [5]); however, the majority of quishing incidents originate from emails [4], which can have a malicious QR code included either as an image attachment, or embedded in the email's body. To deceive their targets, quishing emails leverage the same methods as regular phishing emails, e.g., mimicking reputable sources and emphasizing urgency [8, 58]. However, two peculiarities make quishing emails particularly subtle.

- *Quishing emails are not detected by spam/phishing filters.* Many providers of email services now integrate automated blocking mechanisms that prevent delivery of emails containing malicious URLs, thereby defusing most phishing attacks. However, if the malicious URL is concealed by a QR code, then such filters would not work [4]. We have verified this property with an original experiment (discussed in Appendix D). Further, a user cannot evaluate the safety or trustworthiness of a QR-code in the same way they may have been trained for phishing links.
- *Quishing emails bypass organization-wide security barriers.* Most workplaces adopt security mechanisms such as firewalls, VPN or private DNS. Therefore, even under the assumption that an employee receives an email with a phishing URL (potentially concealed in a click-through button [63]), clicking the URL may lead to a response of the security system—preventing the user from reaching the malicious website. However, QR codes are typically scanned with a different device, such as smartphones; such devices may not be connected to organization's network, but to that of, e.g., the telecommunication provider of the user—which operates outside the organization's security control. Hence, a user receiving a quishing email would scan the QR code and visit the linked website with an "unprotected" device, increasing the likelihood of falling victim to the phishing attempt [4].

In short, quishing emails sidestep most filters and security controls. We highlight two recent works that have addressed the problem of quishing (but differently from us). Sharevski et al. [89] conducted

[1] We focus on phishing (and not spam [50]) *emails*: other forms of phishing (e.g., websites [70], SMS [49, 68], or via vocal telephony [104]) are outside our scope.

a user study with 173 Amazon Mechanical Turk workers (i.e., humans) exposed to a fictitious malicious QR code. The goal was to determine whether users would scan the code and visit the associated URL or refrain. Only 14.5% chose not to scan, indicating high susceptibility to quishing in this population. However, the study did not consider organizational settings or compare the effectiveness of QR-based phishing emails to traditional link-based ones, both central to our investigation. Roy et al. [79] examine the use of large language models (LLMs) to generate phishing emails, some featuring malicious QR codes. While they show that commercial LLMs (e.g., ChatGPT, Claude, Bard) can easily produce convincing quishing content, they do not assess the real-world effectiveness of these emails on human targets, especially in organizational contexts.

2.1.2 *LLM-generated Phishing Emails*

Advancements in Artificial Intelligence (AI), particularly the emergence of publicly accessible large language models (LLMs), represent a double-edged sword [87]. On one hand, LLMs enhance productivity by supporting tasks such as text generation and analysis [28, 36, 100]. On the other hand, these same tools can be exploited by malicious actors to streamline and scale cyberattacks, lowering the barrier to crafting sophisticated offensive content.

Importantly, LLM-based tools not only are effective at *(i)* imitating human writing to create persuasive texts [38], but they can also *(ii)* facilitate the summarization of unstructured information [108], such as that acquired via open-source intelligence (OSINT). Moreover, *(iii)* LLM can now be used by anyone (essentially) for free [87]. The combination of these three factors makes LLM very attractive for phishers. It is hence not surprising that numerous company executives and technical reports from renowned cyber-security companies affirm that there is an increasing usage of LLM (or "generative AI") to convey phishing attacks [10, 25, 34, 92, 93].

Prior works have explored the potential of LLM in the phishing-email context. Most research papers (e.g., [61, 79], as well as various preprints [43, 44, 53]) depict ways in which LLM can be used to craft phishing-related content. However, and to our knowledge, there are only three (unpublished, at the time of writing) works that analysed the effectiveness of LLM-based phishing emails against humans: a case report by IBM's X-Force [29] across 800 employees of a healthcare company evaluated the effectiveness of emails generated by feeding OSINT-acquired information (from online social networks) to an LLM (ChatGPT); Bethany et al. [24] carried out a simulation on 9000 members of a university by asking ChatGPT to write phishing emails in a style that imitated that of some official webpages of the university; Heiding et al. [45] consider various LLMs to craft spear-phishing emails targeting one among 101 volunteers in a user study. Hence, despite the great interest in LLM in the phishing-email context, there is still a lack of understanding of how effective these tools can be at deceiving humans. (Note: using LLMs as phishing-email *detectors* [46] is orthogonal to our work.)

2.2 Research and Knowledge Gap

As acknowledged in Proofpoint's latest report, QR-code and LLM-based phishing emails are becoming trendy vectors to convey phishing email attacks in the real world [10]. In this paper, we aim to understand these threats by focusing on their effectiveness against humans—the true target of phishing [107].

Specifically, we focus on addressing two gaps in our knowledge—which, if filled, would enable current organizations to better cope with the never-ending struggle against phishers. Namely:

(1) Quishing emails are gaining traction as a phishing vector [5, 10]. This raises a critical question: how effective are QR-codes at luring potential victims to a phishing website compared to "traditional" phishing vectors? Although quishing emails can evade automated detection mechanisms (§2.1.1), *it remains unclear whether QR codes are equally effective at deceiving human users*. One might expect traditional phishing emails to perform better in this regard,[2] as the additional burden introduced by QR codes could reduce user engagement. If this expectation does not hold, however, it would suggest an urgent need to adapt both *(i)* automated detection systems and *(ii)* phishing education programs to better address the rising threat of quishing.
(2) There is increasing evidence of LLMs being used by phishers in the wild [10, 34]. This raises the question: how susceptible are a given company's employees against phishing emails written by an LLMs fed with OSINT information pertaining to their targeted company? Indeed, attackers are increasingly refining their tactics and can develop automated OSINT pipelines capable of targeting the entire workforce of a given organization.

Moreover, to provide an additional human-centered perspective, we attempt to establish whether employees "perceived" *phishing awareness* has any relationship with phishing susceptibility to our considered phishing threats. Such an investigation is motivated by the many works that address the topic of phishing education [48, 59, 84], which often show contrasting results (c.f. [60] with [48]).

We were unable to identify any prior work that specifically addresses the first gap.[3] For the second gap, the only real-world evidence we found comes from two studies [24, 29], each focused on a single organization—reflecting a broader trend in phishing research [27, 59, 60, 84]. While these studies—like ours—do not claim universal generalizability, it remains unclear whether particular attack methodologies can be broadly applied across different organizational contexts. It is also of interest to explore how the same phishing approach might yield varying outcomes depending on the organizational environment. We therefore aim to investigate these knowledge gaps through a multi-organization study.

3 Research Questions and Problem Definition

In this work, we tackle a broad research question (RQ): "*How resilient employees across organizations of different size are to emerging social engineering techniques used to deliver phishing email attacks?*" Specifically, to align such an RQ with the previously identified research gaps, we disentangle this RQ into three sub-RQs:

RQ1 Are Quishing emails more (or less) effective at deceiving end users than traditional button-based "click-through" emails?
RQ2 What are the effects of LLM-generated and OSINT-based phishing emails against modern organizations' employees?
RQ3 Is there a correlation between employees' *(a)* perceived phishing awareness and their *(b)* actual susceptibility to phishing?

[2] Clicking a link is nearly effortless, whereas scanning a QR code involves multiple steps: *(i)* retrieving a secondary device such as a smartphone, *(ii)* opening a QR-code reader, *(iii)* scanning the code, and *(iv)* following the link on the separate device.
[3] We systematically review the (lack of) coverage of "quishing" in Appendix C.

To better understand the framing of our research questions, we now present the fundamental assumptions that drive our experiments.

3.1 Threat Model

It is evident that, to address RQ1 and RQ2, we need to craft phishing emails and measure their effectiveness. Let us elucidate the quintessential security elements of the overarching scenario.

We assume an attacker that wants to steal sensitive credentials (i.e., userid and password) of employees of a given target company. The attacker seeks to do so via targeted phishing emails, which include elements mimicking those of the target company, which are sent to an unspecified set of employees of such a company. Therefore, we assume the attacker knows the email addresses of some employees as well as their name and surname (inferring the email given the name/surname, or vice versa, is easy [72, 99]). The attacker also has some knowledge on the target company, such as what provider is used to handle company-related emails (e.g., Microsoft or Google; inferring such information is doable, e.g., via MX lookups [31]). To harvest credentials, the attacker sets up a malicious webpage that mimics the organization's branding (e.g., logos, banners) to foster a sense of authenticity [54, 102]. In our experiments, we assume that the URL of the malicious site is not (yet) listed in any blocklist.[4] Practically, this can be achieved by deploying cloned versions of the webpage to different URLs [57]. The attacker leverages their (limited) knowledge of the target company to craft a phishing email designed to lure recipients to a malicious webpage. To conceal the suspicious nature of the URL, the attacker may either embed it in a click-through button or encode it into a QR code included in the email body. RQ1 investigates the relative effectiveness of these two phishing tactics.

Moreover, we consider a scenario where the attacker leverages openly accessible LLMs to (cheaply) generate the phishing email. To this end, the attacker gathers publicly available information about the target company (e.g., from social media or the company's website) and provides it as input to the LLM. The model is then tasked with extracting relevant details and generating a persuasive email that could deceive recipients. RQ2 explores how OSINT can be combined with LLMs to craft phishing emails and evaluates the practical implications of such strategies.

3.2 Experimental Approach and Choices

We describe our approach, justifying two crucial design choices.

Generic approach (and challenges). First, we must find some organizations that enable us to collect data for our RQ. Specifically, these organizations must grant us the following permissions:

- Carry out phishing simulations (or give us data about phishing simulations) which entail a large share of their employees.
- Give us some freedom on such simulations, so that we can craft the "phishing" emails in such a way that we can answer RQ1–2.
- Enable some form of interactivity with some of their employees to measure the "perceived" phishing awareness for RQ3.

More specifically, for RQ1, we need to test the effectiveness of two emails—which should be identical, aside from: one leading to the credential-harvesting webpage via a click-through button; and another one via a QR-code. For RQ2, we need a third email created by collecting OSINT on the company (hence, even though OSINT entails publicly available information, we must still obtain the company's permission to carry out OSINT activities). Nevertheless, to ensure consistency in the data we collect (and hence provide a meaningful answer to our RQs), we must design our experiments by minimizing the underlying differences that exist between the companies that accept to collaborate in this research. In what follows, we provide more details on how we seek to answer our RQs.

Phishing email effectiveness (and susceptibility). Our research is centered on a core objective: measuring the impact of phishing *emails' content* to deceive users. In simple terms, we want to measure the following: "given a (phishing) email that is *read* by a user, will such a user be fooled and hence *visit* the (malicious) webpage pointed by the email?" We do not consider emails that have not been opened, since such an outcome is not related to the email's content—but rather to its metadata or external factors (e.g., subject or time of delivery); at the same time, what happens after the user visits the landing webpage can be influenced by other factors, such as browser, device, or even the landing page itself—all of which have little relevance to the email's content. Therefore, we measure the effectiveness of a set of phishing emails (or the susceptibility of a company's employees to a set of phishing emails) by computing the ratio of those phishing emails that successfully bring a user to the corresponding landing webpage with respect to the number of phishing emails that have been read: a higher ratio denotes more effective emails (or more susceptible employees). Such a metric is typical in related studies [45, 47, 71, 83, 91].

Measuring the PPA. There are various ways (e.g., [16, 30, 65, 77]) to collect data that can be used to measure the awareness of a given set of subjects with respect to phishing threats—which is a core theme in cybersecurity awareness programs (CSA). For our research, we focus on the perception of the end users, i.e., the "perceived" phishing awareness (PPA). We do so by following the guidelines by Chaudhary et al. [30], who indicate *surveys* as the most appropriate (and least intrusive) mechanism to collect data useful for our goal.[5] Building on the foundations of prior work, our survey is rooted on the "knowledge-attitude-behavior" principles [86]: *knowledge* denotes "familiarity, awareness, or understanding" of security policies/procedures/standards/directives/regulations/laws/guidelines/strategies/technologies. *attitude* denotes "beliefs/opinions/thinking/feelings" toward security; *behavior* denotes the way a person "acts" when faced with security issues [30]. In this context, RQ3 seeks to correlate information related to the PPA of a given company's (subset of) employees with the overall susceptibility of such a company's employees to phishing emails—and then see if the same result holds across different companies.

4 Methodology

To systematically answer our RQs, we found agreements with three companies (described in §4.1). We first carried out three phishing

[4] If the URL is included in some blocklists, then the phishing campaign would likely fail because any victim may not reach the credential-harvesting webpage—either because the webpage is blocked by the browser, or by a firewall; or even because the email may be blocked by the automatic filters and hence not read in the first place.

[5] Chaudhary et al. [30] synthesized 32 papers on cybersecurity awareness programs, and proposed four indicators: *impact* (which is our focus, given that its relatedness to phishing susceptibility), *sustainability*, *accessibility*, *monitoring*. Each indicator comprises nine factors: *attitude towards cybersecurity*, *interest*, *usability*, *self-reported behavior*, *knowledge and competence gain*, *value added*, *reachability*, *touchability*, *overall feedback*.

simulations in these companies (described in §4.2), focused on answering RQ1 and RQ2. Afterwards, we focused on RQ3 and carried out a user study (described in §4.3) with our companies' employees. We discuss ethical concerns pertaining to our methodology in §4.4.

4.1 Description of Companies

We contacted various companies located in Central Europe, asking for their collaboration in a research project on the topic of phishing email susceptibility and assessment. After numerous exchanges, we eventually found an agreement with three companies, which we summarize below (an overview is provided in Table 1).

- *Small-sized Company* ($\mathbb{C}_s$). A small-medium enterprise (SME) operating in the hospitality sector [52]. $\mathbb{C}_s$ has between 50 and 250 employees. The customers of $\mathbb{C}_s$ are located only in two countries in Europe. This company has no self-administered IT department, and their IT is outsourced to a third-party vendor. In terms of CSA training, only basic dissemination methods (e.g., educational texts and slides covering various attack vectors seen in the past) are adopted by $\mathbb{C}_s$, which are provided to its employees on a yearly basis. Each employee receives the exact same CSA training, and there is no mechanism meant to assess the effectiveness of CSA training in $\mathbb{C}_s$. Notably, $\mathbb{C}_s$ had never carried out in-house phishing simulations before.
- *Medium-sized Company* ($\mathbb{C}_m$). An enterprise with less than 2 000 employees, operating in the financial sector. This company operates in multiple German-speaking countries as well as in some middle-eastern countries. The employees of $\mathbb{C}_m$ span across the typical roles of financial organizations, including administration, marketing, and IT. Indeed, $\mathbb{C}_m$ has its own IT infrastructure and a dedicated cyber security team, which is also tasked to carry out phishing simulations at least twice per year. The CSA training provided by $\mathbb{C}_m$ to its employees includes slides, videos, texts, as well as physical classes covering even recent/emerging phishing trends. Each employee receives the same type of CSA training (on a yearly basis) and employees are required to pass a dedicated exam to demonstrate their understanding of such CSA training.
- *Huge-sized Company* ($\mathbb{C}_h$). A multi-national enterprise (having over 30 000 employees) conducting business on a global scale in the manufacturing sector. They have a sophisticated IT infrastructure and a large cybersecurity team. $\mathbb{C}_h$ regularly carries out phishing simulations across all its employees. CSA training is provided every two years, and is tailored for the specific role of each employee (e.g., managers receive different training than members of HR). Such training includes slides, videos, texts, in-person training, and even leverages eLearning platforms with additional content (covering also recent trends); at the end of each CSA training campaign, employees must pass an exam.

Information on the CSA of each company has been derived via a structured questionnaire (Table 6 in Appendix E) with knowledgeable representatives of each company. Due to non-disclosure agreements (NDA), we cannot provide more details.

4.2 Experimental Setup of the Simulations

We create three "phishing" emails: two (described in §4.2.2) for RQ1, whereas a third (described in §4.2.3) is for RQ2. Let us first introduce the common elements of our experimental testbed (§4.2.1).

Table 1: Overview of Companies. For our research, we considered three companies whose businesses is predominantly located in Central Europe.

	Small Company ($\mathbb{C}_s$)	**Medium Company** ($\mathbb{C}_m$)	**Huge Company** ($\mathbb{C}_h$)
# Employees	between 50 and 250	≈1 500	>30 000
Industry	Hospitality	Finance	Manufacturing
CSA Training Frequency	Yearly	Yearly	Biyearly
CSA Training Approaches	Slides, Texts	Slides, Videos, Texts, Classes	Slides, Videos, Text, Classes, eLearning
In-house Simulations?	✗	✓	✓
CSA Training Specificity	Generic	Generic	Group-specific
Emerging Trends in CSA?	✗	✗	✓
Simulation Framework	(GoPhish [3])	MS Defender [6]	MS Defender [6]

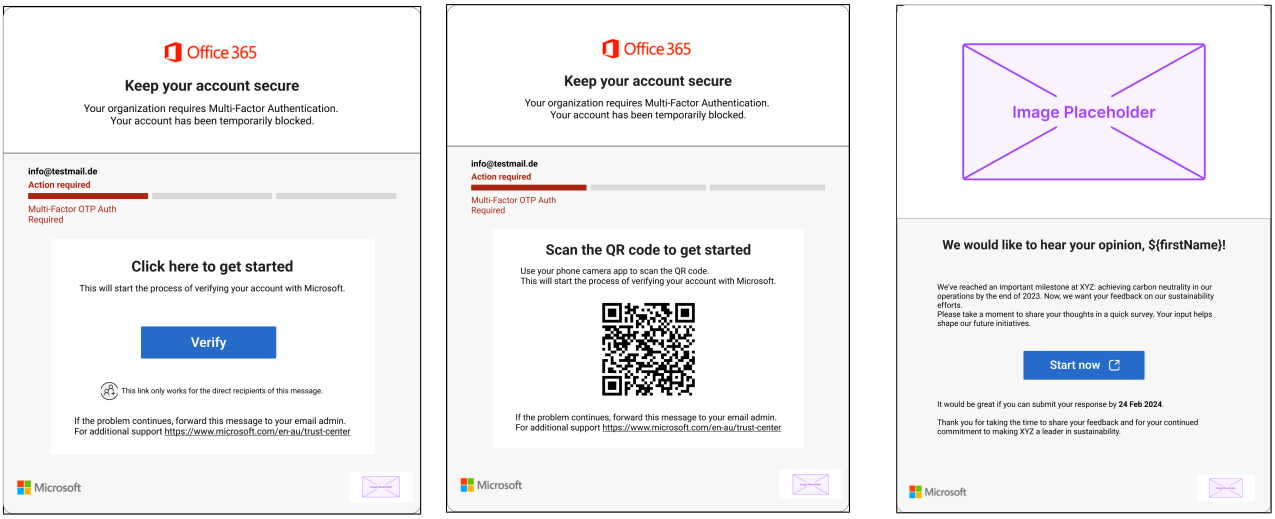
4.2.1 *Common elements*

All phishing simulations carried out in our study have been designed in accordance with the chosen companies' policies. Specifically, our simulations were meant to serve as a periodical assessment of $\mathbb{C}_m$ and $\mathbb{C}_h$; whereas, for $\mathbb{C}_s$, our simulations were the very first phishing assessment done in $\mathbb{C}_s$. Such a context allowed us to develop a customized simulation framework for $\mathbb{C}_s$, for which we aligned with $\mathbb{C}_m$ and $\mathbb{C}_h$ to minimize inter-company differences. In what follows, we provide the most relevant technical details.

- *Infrastructure.* At a high-level, our experiments entail sending "phishing" emails to employees and see how they react (e.g., whether they open the email, or click on the phishing link). We used the existing infrastructure (i.e., Microsoft Defender [6]) for $\mathbb{C}_m$ and $\mathbb{C}_h$. For $\mathbb{C}_s$, we deployed our own infrastructure by leveraging the open-source GoPhish framework [3] (used also, e.g., in [27, 64]). Altogether, these frameworks enable collection of the data required for our RQs (see §3.2). We provide low-level details on the experimental infrastructure, as well as on the challenges we had to overcome to set it up, in Appendix A.
- *Landing webpage.* A common practice [60] in phishing simulations is to embed a link in the email that points to a "credential-harvesting" webpage that invites the user to submit some sensitive information—which aligns with our threat model (§3.1). Such a webpage is typically designed to enable tracking of those users that land on it, or who submit their data. Both $\mathbb{C}_m$ and $\mathbb{C}_h$ used the typical login page of Microsoft for their simulations, so we designed the landing page for $\mathbb{C}_s$ (shown in Fig. 3 in the Appendix) accordingly, given that $\mathbb{C}_s$ also relies on Microsoft.[6]
- *Data collection.* We coordinated with the companies so that each (randomly chosen) employee could only receive at most one email among those we crafted. We timed the delivery so that each recipient would receive the email in the morning (around 8am, accounting for timezones) of a work day. All emails include a "Microsoft" component (e.g., a logo) since all companies use the Microsoft Office suite. The sender of these emails was always related to an identity resolving to "@mircosoft.com". When crafting our emails, since they entailed HTML objects (e.g., images), we ensured that they rendered correctly on the email clients most commonly used by the respective company's employees.[7] Finally, for each email sent, we obtained: the number of recipients that opened/read it; the number of recipients that visited the landing

[6] The landing webpage for $\mathbb{C}_s$ was hosted on a domain we purchased ourselves, and we made the URL very similar to that of the official webpage of $\mathbb{C}_s$. Specifically, its URL was "$Company_sName.email". The landing webpages for $\mathbb{C}_h$ and $\mathbb{C}_m$ were hosted on their premises; we cannot provide details on these URL due to NDA.

[7] While setting up our testbed for $\mathbb{C}_s$ we noticed that their default configuration of Microsoft Outlook forced the user to explicitly allow displaying images (including QR codes) in an email before showing them—if the email comes from an "unknown" sender. To overcome this problem, the sender of our emails was added to the "trusted sender" list. This ensured that any images (including the QR code) would be displayed—thereby also guaranteeing a correct counting of the opened emails by GoPhish (see [26]).



(a) Example of button "click-through" email ($\mathbb{E}_B$). The "info@testmail.de" was replaced with a company-related email address.

(b) Example of QR-code phishing email ($\mathbb{E}_Q$). Note that the design is identical to $\mathbb{E}_B$ aside from the button being replaced with a QR-code.

(c) Example of OSINT+LLM phishing email ($\mathbb{E}_L$). The large "image placeholder" was replaced with an image taken from a press release of the specific company.

Fig. 1: Emails used in our experiments. Our emails shared a similar design, but each email presented some company-specific traits to increase authenticity (e.g., we put the company logo at the bottom right). All emails bring the user to the same landing webpage (which was also specific to each company).

webpage; the number of recipients that reported the message, and the number of submitted credentials.

The simulations occurred between April 24th and May 10th, 2024 (depending on the company), and lasted around 3 days each.

4.2.2 *Quishing & Traditional-phishing Emails*

To answer RQ1, we carried out two phishing simulations, each revolving around a specific email created ad-hoc for our experiments.

- *Button email* ($\mathbb{E}_B$). This email contains an URL integrated in a "click-through" button that brings the user to the landing page.
- *QR-code email* ($\mathbb{E}_Q$). This email is identical to $\mathbb{E}_B$, and the only difference is that $\mathbb{E}_Q$ includes a QR code (instead of the button).

Fig. 1a shows an example of $\mathbb{E}_B$, while Fig. 1b presents its quishing counterpart, $\mathbb{E}_Q$. The two emails are visually identical, except for the interaction mechanism: $\mathbb{E}_B$ prompts users to click a button, whereas $\mathbb{E}_Q$ requires scanning a QR code (e.g., via smartphone) to access the concealed URL. This controlled design isolates the variable of interest, enabling us to address our first RQ.

Email design. To create the "phishing hook" of these emails, we took inspiration from the common tactics adopted by phishers. Specifically, the email urges the recipient to set up multi-factor authentication to reactivate their account—which should be done by following the link included in the email. We chose such a hook because it is well-known (e.g., [40]) that emails containing IT-related topics are very successful at deceiving users. Moreover, we created the emails so that they had an "authentic" design [17], resembling that of communications sent by the respective company (i.e., we used the company's logo and also elements of the Microsoft Office suite). We also included urgency cues (e.g., "action required") and loss ("your account has been temporarily blocked"), since they have all been found to be very effective [101]. Note that all of our design choices comply with the overarching threat model (§3.1).

Company-specific differences. Full alignment across companies was not always feasible, so minor differences affect our setup. First, while the emails were in English for $\mathbb{C}_h$, they were translated into German for $\mathbb{C}_s$ and $\mathbb{C}_m$, where German is the primary language. Second, a security banner ("This is an external email...") was included for $\mathbb{C}_m$ and $\mathbb{C}_h$, but not for $\mathbb{C}_s$, which does not use such warnings by default. Third, $\mathbb{C}_h$ did not require $\mathbb{E}_Q$, as it had recently conducted a QR-based simulation (in Jan. 2024), for which it shared results with us. Accordingly, we designed $\mathbb{E}_B$ to match that version of $\mathbb{E}_Q$ for $\mathbb{C}_h$, differing only in the use of a button instead of a QR code; we cannot disclose the exact emails due to NDA. Importantly, these variations do not affect our investigation of RQ1, as $\mathbb{E}_Q$ and $\mathbb{E}_B$ retain consistent properties within each organization.

Data analysis. Overall, 18 339 $\mathbb{E}_B$ (21 for $\mathbb{C}_s$, 567 for $\mathbb{C}_m$, 17 751 for $\mathbb{C}_h$) and 34 610 $\mathbb{E}_Q$ (21 for $\mathbb{C}_s$, 558 for $\mathbb{C}_m$, 34 031 for $\mathbb{C}_h$) were sent.[8] As we explained (§3.2), the answer to our first RQ is determined by calculating, for both $\mathbb{E}_Q$ and $\mathbb{E}_B$, the ratio of those users that visited the landing webpage with respect to those that read the email; and then compare these two numbers via statistical tests and analytics.

4.2.3 *Phishing Email by using OSINT and LLM*

To address RQ2, we carry out a simulation revolving around a single email, $\mathbb{E}_L$, crafted by providing OSINT-acquired information as input to a publicly accessible (and free to use) LLM. At a high level, our methodology resembles that used in the test by X-Force in 2023 [29]. Specifically, our email aimed to imitate the invitation to participate in survey, organized by the targeted company, on topics aligning with the company's agenda. To create $\mathbb{E}_L$, we followed a systematic approach (visualized in Fig. 2), rooted on the assumptions of our threat model (see §3.1). We describe it below.

Gathering OSINT. We leveraged three publicly available sources: the most prominent German-speaking employer rating website (Kununu [13]), an online professional social network (LinkedIn [14]),

[8] For $\mathbb{C}_h$, the numbers for $\mathbb{E}_Q$ are higher than $\mathbb{E}_B$ because the quishing simulation in $\mathbb{C}_h$ had been carried out by $\mathbb{C}_h$, and it hence targeted all employees that take part in these phishing simulations; in contrast, the simulation for $\mathbb{E}_B$ has been carried out by us, and the emails were sent by randomly choosing half of the employees of $\mathbb{C}_h$ (the remaining half received the email used for RQ2, discussed in §4.2.3). Across all companies, our emails were addressed to approximately 30–50% of employees. Due to NDA restrictions, we cannot disclose the exact percentages.

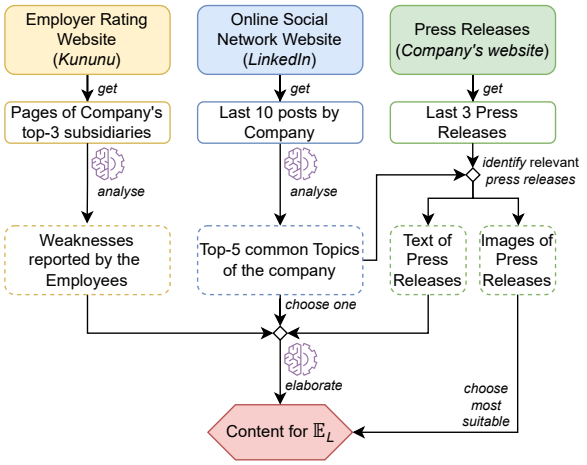


Fig. 2: Extraction and exploitation of OSINT for $\mathbb{E}_L$. Operations denoted with a "brain-cog" image have been carried out with an LLM.

and press releases on the company. We chose these sources because of their popularity and relevance—given the considered companies' public visibility. Let us explain how we used each of these sources.

- *Kununu.* We began by surveying the Kununu page for each company, seeking to find "issues/weaknesses" reported by the employees. We found that $\mathbb{C}_s$ did not have a dedicated entry on this website. For $\mathbb{C}_m$ and $\mathbb{C}_h$, we first identified the three largest subsidiaries of each company; then, we looked up the Kununu page of each subsidiary; finally, we saved the Kununu's page (if available) of each subsidiary in a single HTML file—which would be later used as input to the LLM to find "potential weaknesses" that could be exploited for phishing attacks.
- *LinkedIn.* Afterwards, we looked at the LinkedIn page of each company. The goal was finding a topic that could be of interest to the employees of the company—the recipients of our email. All companies have an official page on LinkedIn. We selected the 10 last posts made by the company (excluding reposts), and saved the contents of such posts into a single HTML file. Such a file would be later used as input to the LLM to (automatically) find "the 5 most common topics covered in these posts".
- *Press releases.* Finally, we considered the public press releases of each company, for which we relied on the official communication channels in each company's website. The idea was to further enrich our previous analyses by also integrating visual assets that could improve the quality of our phishing email. Hence, for each company, we considered the three most recent press releases: after identifying those that deal with the topics found via LinkedIn, we downloaded all usable assets (e.g., images or banners) and we saved the text of the press release in a file—which we later used as input to the LLM to write "an introductory text to a survey about a relevant topic for this company".

The above-mentioned operations were done in February 2024.

Feeding OSINT to the LLM. We considered ChatGPT 3.5 Turbo for our LLM. Our choice is because it was free: ChatGPT 4.0 required a paid subscription (at the time) which may have discouraged real attackers from using it (phishing campaigns tend to be cheap [19]). Nonetheless, to generate the content of each email, we assembled a sequence of five prompts (see Table 4 in Appendix A) that integrate the information extracted via our OSINT operations. The resulting text was used to compose the main body of our email $\mathbb{E}_L$. Then, we used the most appropriate visual assets we collected from the press releases and used them to improve the aesthetics of the email. Next, to provide a sense of authenticity, we added the company's and Microsoft's logos at the bottom of the email. Finally, we added a click-through button that embedded a link to our landing page (ideally, the survey was meant to be organized by the company so that only its employees could participate—which is why a login was required) and added text soliciting the recipient to submit their responses within a few days. To sum up, we used the LLM to both *(i)* summarize content and *(ii)* write the email's text.

Data analysis. Overall, 18 360 $\mathbb{E}_L$ (18 for $\mathbb{C}_s$, 589 for $\mathbb{C}_m$, 17 753 for $\mathbb{C}_h$) have been sent. To investigate RQ2, we qualitatively analyse all numerical data we can collect related to $\mathbb{E}_L$. We predominantly focus on the ratio of users that visited the landing webpage w.r.t. those that read the email; but we also gauge the ratio of users that submitted their credentials w.r.t. those that visited the landing webpage. This is because, for this email, we do not leverage the sense of loss as we did for $\mathbb{E}_B$ and $\mathbb{E}_Q$ (i.e., "your account has been temporarily blocked"). Thus, users may not expect to land on a webpage that asks them to input their credentials. Therefore—contrarily to $\mathbb{E}_B$ and $\mathbb{E}_Q$—for $\mathbb{E}_L$ the submission of credentials is strongly dependent on how persuasive $\mathbb{E}_L$'s content (which depends on OSINT and LLM, i.e., the crux of RQ2) is in deceiving the end user.

4.3 Perceived Phishing Awareness

To answer RQ3, we conducted user surveys among companies' employees to estimate their (perceived) phishing awareness. The surveys (implemented via MS Forms) consist of anonymous closed-answer questionnaires, developed in agreement with each company.

Questionnaire. The questionnaire follows scientific guidelines on empirical social research [23], and is rooted on the knowledge-attitude-behavior principles [86]. As we explained (§3.2), our questionnaire is built on the work of Chaudhary et al. [30]. Specifically, to allow complete coverage of the "impact" indicator (which is related to phishing susceptibility), we consider the following factors proposed by Chaudhary et al: *attitude towards cybersecurity*, *interest*, *usability*, *behavior*, *knowledge and competence gain*.[9] Overall, the questionnaire spans across 40 questions, distributed in five sections: *(i)* attitude towards cybersecurity—for which we provide a snippet in Fig. 9; *(ii)* cybersecurity routines—which focuses on the *(self-reported) behavior*; *(iii)* cybersecurity awareness training—which focuses on *CSA training experience* (related to *interest*) and *training usability*; *(iv)* quick assessment—for which we provide a snippet in Fig. 10, and which focuses on *knowledge and competence gain*; *(v)* socio demographics. The answers to most of the questions were based on a 5-point Likert Scale [69] (1=strongly disagree; 5=strongly agree). Given the multi-lingual nature of our companies, the questionnaire was created both in German and in English. One of the authors has native German fluency. The complete list of questions in our questionnaire is provided in Table 5 (in Appendix E).

Distribution. For $\mathbb{C}_s$ and $\mathbb{C}_m$, our questionnaire was distributed among employees via the official internal communication platform;

[9] N.b.: our preliminary survey with the companies' representatives (shown in Table 6), used to derive the CSA profile of each company (summarized in §4.1), covered the remaining four factors (i.e., *value added*, *reachability*, *touchability*, *overall feedback*). Hence, our study provides a complete coverage of the "impact" indicator.

for $\mathbb{C}_h$, it was distributed via dedicated "team" channels as well as by convenience sampling [35] (e.g., by sending emails to employees). For each company, we disseminated the questionnaire a few days after concluding our phishing simulations and collected responses for approximately one week. While we could not control exactly who chose to participate, it is reasonable to expect that respondents had also taken part in the phishing simulation.

Data analysis. To answer RQ3, we first scrutinize the data we collected with our questionnaire, and then cross-analyse our results with those of our phishing simulations. Specifically, for each company, we first compute the mean and variance of the five-point Likert scale for each item of our questionnaire; then, we derive a general "Perceived Phishing Awareness" score (PPA-score) by aggregating all responses. Next, we define a "Phishing Susceptibility" score (PS-score) by computing the ratio of those emails (accounting for $\mathbb{E}_B$, $\mathbb{E}_Q$, and also $\mathbb{E}_L$) that brought a user to a landing webpage w.r.t. those that have been read. Considering a single pool which aggregates the results of all our emails (i.e., 71 309 in total) is valid because, despite some differences: the emails are ultimately all "phishing", leveraged the same structural properties, and there was no targeted or arbitrarily-chosen selection of recipients (any employee could be eligible).[10] Finally, we statistically compare the PPA-score with the PS-score, and draw our conclusions.

4.4 Ethical Considerations

To ensure we perform our simulations ethically, we followed well-known and established scientific practices [21, 23, 56].

Our experiments (§4.2) have been designed in accordance with company $\mathbb{C}_h$ and $\mathbb{C}_m$'s ethical standards, whereas $\mathbb{C}_s$ subcontracted us to carry out the assessment within their premises. Accordingly, all activities were approved by the relevant ethical bodies within the involved organizations. Employees that received one of our phishing emails were aware that their companies carried out phishing assessments. No harm was caused during the study: both the emails and landing pages were under our control, and no credentials were persistently stored. Participants of our (anonymous) user study (§4.3) did so willingly and had been made aware that their responses would be collected and used only for scientific purposes.

Finally, we mention that ChatGPT refused to answer our request to provide "weaknesses that could be leveraged for a phishing attack" (§4.2.3), so we had to find a workaround that would bypass its automatic censorship mechanisms (we do not provide our exact prompts to avoid helping attackers crafting phishing emails).

5 Phishing Simulations [RQ1, RQ2]

We first present the results (§5.1) of our three simulations entailing the QR-code phishing email ($\mathbb{E}_Q$), the traditional phishing email leveraging a click-through button ($\mathbb{E}_B$), and the OSINT-fed LLM-written email ($\mathbb{E}_L$). Then, we address our first research question with a statistical test (§5.2). Finally, we address our second research question (§5.3) with a comprehensive qualitative analysis.

5.1 Overall Results of our Phishing Simulations

We report in Table 2 the results of our simulations. Specifically, for each email ($\mathbb{E}_B$, $\mathbb{E}_Q$, $\mathbb{E}_L$) we provide: the overall number of emails sent; the number of emails that have been read (we consider an email as "read" if it has been opened); the number of emails for which the landing page has been visited at least once; the number of emails for which the credentials of the recipient have been submitted; as well as the "page visited / email read" ratio and the "credential submitted / email read" ratio. The results are provided for each company, and the rightmost columns report the aggregated results across all companies. Let us analyse our results at a high level.

For $\mathbb{C}_s$, we can observe that $\mathbb{E}_B$ and $\mathbb{E}_Q$ led to a similar outcome in terms of "page visited / email read". We also find it worrying that 11.1% (and 7.7%) of those who read $\mathbb{E}_B$ (and $\mathbb{E}_Q$) eventually submitted their credentials; even more worrying is the effectiveness of $\mathbb{E}_L$. However, the relatively-small sample size for $\mathbb{C}_s$ prevents one from drawing sound conclusions from these numbers. Nevertheless, after we launched our simulations, we were contacted by some representatives of $\mathbb{C}_s$ in charge of IT matters: they reported that they had been messaged/called by 11 employees, asking about the legitimacy of the emails they had just received.

For $\mathbb{C}_m$, we also see (as for $\mathbb{C}_s$) that the trends for $\mathbb{E}_B$ and $\mathbb{E}_Q$ are similar, with 3.9% (resp. 5.4%) of those that opened $\mathbb{E}_B$ (resp. $\mathbb{E}_Q$) visiting the landing page. Moreover (and also in line with $\mathbb{C}_s$), $\mathbb{E}_L$ seems to have been more effective than both $\mathbb{E}_B$ and $\mathbb{E}_Q$. Finally, we mention that, for $\mathbb{E}_B/\mathbb{E}_Q/\mathbb{E}_L$, 241/155/182 employees reported the email, whereas 352/377/312 deleted it.[11]

For $\mathbb{C}_h$, the outcome of $\mathbb{E}_B$ and $\mathbb{E}_Q$ are also somewhat similar, with 8.1% (resp. 7.9%) of those who opened $\mathbb{E}_B$ (resp. $\mathbb{E}_Q$) visiting the landing page. However (and differently from $\mathbb{C}_s$ and $\mathbb{C}_m$) the impact of $\mathbb{E}_L$ had a lower impact than both $\mathbb{E}_B$ and $\mathbb{E}_Q$ (despite $\mathbb{E}_L$ having been read by comparatively the same amount of recipients as both $\mathbb{E}_B$ and $\mathbb{E}_Q$). Finally, we mention that, for $\mathbb{E}_B/\mathbb{E}_Q/\mathbb{E}_L$, 3 039/10 268/1 637 employees reported the email, whereas 6 567 deleted $\mathbb{E}_B$ and 6 375 deleted $\mathbb{E}_L$ (we do not have such data for $\mathbb{E}_Q$).

5.2 Statistical Assessment of $\mathbb{E}_B$ and $\mathbb{E}_Q$ [RQ1]

To objectively answer RQ1, we rely on one-tailed chi-square tests [76].

We define our core hypothesis as follows: "An employee opening $\mathbb{E}_B$ is *more likely* to visit the landing webpage than an employee opening $\mathbb{E}_Q$." Indeed, our expectation is that QR-codes are less effective than click-through buttons (see §2.1.1), since a button simply needs to be clicked/tapped, whereas a QR-code must be scanned first. Such a procedure can be cumbersome, and some employees may think twice before doing so, potentially leading to postponement or forgetfulness of the task; an employee may even realize that the email is suspicious and not proceed at all. We stress, as we stated (§4.2.2), that we seek to measure whether there is any statistically significant difference in the ability of a QR code to bring a potential victim to a phishing webpage w.r.t. a traditional click-through button (note that both cases conceal the URL). What happens "after" the user lands on such a webpage is outside the scope of RQ1 (and, hence, of our null hypothesis for this test).

We perform the chi-square test four times: first on the aggregated data from all companies (yielding a larger sample size), and then separately for each company. This procedure is statistically valid, as the measured phenomenon is consistent across companies.

[10] Moreover, given that we do not know which email (among $\mathbb{E}_B$, $\mathbb{E}_Q$, $\mathbb{E}_L$) was received by our participants, it would be unfair to consider the results of a single simulation.

[11] These results are not available for $\mathbb{C}_s$ but are available for $\mathbb{C}_h$, because GoPhish does not provide such a functionality—which is, however, integrated in Microsoft Defender.

Table 2: Results of $\mathbb{E}_B$, $\mathbb{E}_Q$, and $\mathbb{E}_L$. We recall (§4.2.2) that, for $\mathbb{C}_h$, the simulation of $\mathbb{E}_Q$ was not managed by us: the email was sent to more employees and no data was logged about the credentials submitted. Therefore, numbers with an asterisk (*) have been derived by removing the $\mathbb{E}_Q$ of $\mathbb{C}_h$ from the pool.

Company	$\mathbb{C}_s$ (Small Company)			$\mathbb{C}_m$ (Medium Company)			$\mathbb{C}_h$ (Huge Company)			AGGREGATE		
Email	$\mathbb{E}_B$	$\mathbb{E}_Q$	$\mathbb{E}_L$	$\mathbb{E}_B$	$\mathbb{E}_Q$	$\mathbb{E}_L$	$\mathbb{E}_B$	$\mathbb{E}_Q$	$\mathbb{E}_L$	$\mathbb{E}_B$	$\mathbb{E}_Q$	$\mathbb{E}_L$
Emails sent	21	21	18	567	558	589	17 751	34 031	17 753	18 339	34 610	18 360
Emails read	9	13	12	312	317	397	11 538	24 842	11 025	11 859	25 172	11 434
Page visited	2	3	8	12	17	125	936	1 950	499	950	1 970	632
Credentials submitted	1	1	3	9	6	59	531	n/a	243	541	7*	305
Page visited / Email read	22.2%	23.1%	66.6%	3.9%	5.4%	31.5%	8.1%	7.9%	4.5%	8.0%	7.8%	5.5%
Cred. sub. / Email read	11.1%	7.7%	25.0%	2.9%	1.9%	14.9%	4.6%	n/a	2.2%	4.6%	2.1%*	2.7%

- *Aggregate.* Therefore, we aggregate the results for $\mathbb{E}_B$ and $\mathbb{E}_Q$ across our three companies. For $\mathbb{E}_B$, 11 859 employees opened it, and 950 (8.01%) visited the landing page; whereas 10 909 did not visit the landing page despite opening $\mathbb{E}_B$. For $\mathbb{E}_Q$, 25 172 employees opened it, and 1 970 (8.49%) visited the landing page, whereas 23 202 did not visit the landing webpage despite opening $\mathbb{E}_Q$. The result of the test is a chi-square statistic of 0.353. The corresponding (one-tailed) p-value is .276, indicating no statistically significant difference (assuming a significance level of .05). Therefore, this test indicates that our hypothesis cannot be accepted. Moreover, the effect size is small (0.0031), further confirming that any difference between $\mathbb{E}_B$ and $\mathbb{E}_Q$ is negligible.[12]
- *Company-specific.* We repeat the chi-square test for each company to verify if our findings hold even in specific contexts; for simplicity, we only report the results. For $\mathbb{C}_s$, chi-square=0.0, one-tailed p-value=1.0, effect size=0.0. For $\mathbb{C}_m$, chi-square=0.514, one-tailed p-value=1.0, effect size=0.029. For $\mathbb{C}_h$, chi-square=0.709, one-tailed p-value=.2, effect size=0.004. Hence, our hypothesis cannot be accepted for each of these tests. Note that, for $\mathbb{C}_s$, the sample size is small (so it is possible that the test is inconclusive here). However, the results of $\mathbb{C}_m$ and $\mathbb{C}_h$ are more informative, and the effect sizes (which are almost negligible) confirm that differences between $\mathbb{E}_B$ and $\mathbb{E}_Q$ are not statistically significant.[13]

Finally, we carry out a "two one-sided test" [88] (or "TOST") to establish numerical boundaries that allow one to consider $\mathbb{E}_B$ and $\mathbb{E}_Q$ to statistically have the same effectiveness. For simplicity, we carry out this test only for the "aggregated" results. We assume an equivalence margin of ±1%. We find that the lower bound p-value is .000042, and the upper bound p-value is .0034. Given that both of these values are below .05, we can conclude that there is no practically meaningful difference between $\mathbb{E}_B$ and $\mathbb{E}_Q$ in leading recipients to the landing webpage.

Answer to RQ1. $\mathbb{E}_B$ and $\mathbb{E}_Q$ have practically the same effectiveness at bringing a potential victim to a phishing website. Such a finding is alarming: quishing emails are harder to detect (§2.1.1) but our expectation was that they were less effective at luring users w.r.t. traditional click-through phishing emails. Our findings suggest that such an hypothesis is not true.

5.3 Qualitative Analysis of $\mathbb{E}_L$ [RQ2]

There are numerous insights that can be drawn by qualitatively analysing the results pertaining to $\mathbb{E}_L$.

First, it is evident that $\mathbb{C}_s$ was the company with the highest ratio of employees that visited the landing page or submitted their credentials after reading $\mathbb{E}_L$. While the sample for $\mathbb{C}_s$ was relatively small, the effectiveness of $\mathbb{E}_L$ on $\mathbb{C}_m$ seems to confirm the effectiveness of such an email to deceive employees—and are based on a much larger sample (hundreds of emails). Importantly, the impression is that $\mathbb{E}_L$ tends to be much more effective than both $\mathbb{E}_B$ and $\mathbb{E}_Q$ against the employees of $\mathbb{C}_s$ and $\mathbb{C}_m$ (see Table 2).

It is intriguing to observe that the percentage of "fooled" users for $\mathbb{C}_h$ is comparatively much lower than that of $\mathbb{C}_s$ and $\mathbb{C}_m$ (and much lower also w.r.t. $\mathbb{E}_B$ and $\mathbb{E}_Q$).[14] However, such a result can be due to the multi-national nature of $\mathbb{C}_h$. Indeed, $\mathbb{C}_s$ and $\mathbb{C}_m$ are mostly based in a single country, and it is reasonable to assume that their employees may share similar views that align with the respective company's agenda. Therefore, the $\mathbb{E}_L$ we crafted for $\mathbb{C}_s$ and $\mathbb{C}_m$ could have been very effective at deceiving their employees. In contrast, $\mathbb{C}_h$ is not very localized and its employees may not have a strong sense of attachment to such a company (evidence of this can be found in Tables 11 and 12, given that the percentage of respondents that worked for 6+ years for the same company was much higher for $\mathbb{C}_s$ and $\mathbb{C}_m$ compared to $\mathbb{C}_h$). It is also possible that the email we crafted leveraged cues that $\mathbb{C}_h$'s employees did not find captivating (potentially because the press releases of $\mathbb{C}_h$ may not be interesting for its employees). In contrast, the underlying theme of $\mathbb{E}_B$ and $\mathbb{E}_Q$ (i.e., "your account has been locked") may have been more effective at capturing the attention of $\mathbb{C}_h$'s employees. Another explanation is that some employees believed the emails were legitimate but were not sufficiently motivated by the content to click, resulting in a lower click-through rate (w.r.t. $\mathbb{E}_B$ and $\mathbb{E}_Q$) despite the deception being successful at a cognitive level.

Another contributing factor may be the intrinsic difficulty of crafting a single "generic" phishing email (whether human- or LLM-written) that resonates across the diverse workforce of a large multinational company such as $\mathbb{C}_h$. As prior work has shown [41, 82, 94], the *contextual relevance* of a phishing email (e.g., in terms of timing, location, pretext, or personal relevance) plays a key role in its effectiveness. Smaller organizations are more likely to have employees who are contextually aligned [27], while large companies typically exhibit greater diversity in roles, experiences, and expectations [67].

Nonetheless, it would be misleading to conclude that $\mathbb{E}_L$ is ineffective for $\mathbb{C}_h$ in absolute terms. Our results demonstrate that it is possible to obtain credentials from 243 employees of a multinational

[12] The difference in click-through rates of $\mathbb{E}_B$ and $\mathbb{E}_Q$ is 0.18%, with a 95% confidence interval of $(-0.41\%, 0.78\%) \in \pm 1\%$, confirming no statistically-significant difference.

[13] The 95% confidence intervals for the differences in click-through rates ($\mathbb{E}_B - \mathbb{E}_Q$) for each company are: $\mathbb{C}_s = (-0.363, 0.346)$; $\mathbb{C}_m = (-0.047, 0.017)$; $\mathbb{C}_h = (-0.003, 0.008)$.

[14] While it is possible to carry out statistical comparisons of $\mathbb{E}_L$ w.r.t. $\mathbb{E}_B$ (or $\mathbb{E}_Q$), we refrain from doing so because there are too many differences between these emails and any test would prevent any sound conclusion. For instance, $\mathbb{E}_B$ and $\mathbb{E}_Q$ were designed to impersonate the IT team urging the recipient to setup multi-factor authentication to unblock their account—which is a very important task (if true); whereas $\mathbb{E}_L$ merely requires the employee to participate in an optional survey related to their company.

company with minimal effort. The phishing email was generated using the free version of ChatGPT (in Q2 2024) and relied solely on OSINT from publicly available sources. This low-cost setup highlights the appeal of such tactics to real-world attackers.

Answer to RQ2. Using OSINT data as input to an LLM can result in phishing emails that are cheap to craft while being highly effective—especially against smaller companies.

6 PPA & Phishing Susceptibility [RQ3]

We first present the results of our phishing awareness questionnaire (§6.1), and then answer RQ3 via a statistical assessment (§6.2).

Sample description. Overall, we obtained 131 responses to our questionnaires (13 for $\mathbb{C}_s$, 82 for $\mathbb{C}_m$, and 36 for $\mathbb{C}_h$)[15]. Respondents varied in age: 57 were younger than 34 years, 56 were 34–54 years old, 17 older than 55 (one preferred not to say). Our sample is also relatively well-educated, with 90 participants having a degree (BSc., MSc., or PhD). Respondents also belonged to various departments: the three most prevalent ones being IT (36), operations (24), administration (17). Digital devices were used at work more than 75% of the time for 117 participants. Most of our sample (87) has more than ten years of work experience, with only a minority (10) having worked for less than two years. The complete demographic details (including the repartition across companies) are in Appendix E.2.

6.1 Phishing Awareness Questionnaire: Results

Table 3 reports the results, for each company, of each factor considered in our questionnaire. These numbers have been obtained by averaging the responses of each question across a specific company—which we report in full in Tables 13, 14, 15, 16, in the Appendix E.

For $\mathbb{C}_s$, the *attitude towards cybersecurity* is quite high (avg=4.223 out of 5) indicating that its employees do care about cybersecurity. However, for some specific items (see ACS3, ACS4 and ACS7 in Table 13), the scores are comparatively lower (≤4). This could be due to $\mathbb{C}_s$ employees not being strongly confident about how to act upon arising threats. The *self-reported behavior* is also somewhat high (avg=4.0); the lowest scored items (i.e., BHV2 and BHV3 in Table 14) likely stem from some uncertainty in how to integrate cybersecurity practices into their daily routines. The score for *CSA training experience* is low (avg=1.667): this is expected because $\mathbb{C}_s$ does not carry out regular training with its employees. Moreover, for *training usability*, the mediocre score (avg=2.792) denotes that the employees of $\mathbb{C}_s$ may not perceive training as useful. In terms of *knowledge and competence gain*, there is mediocre (avg=2.862) with high fluctuations (variance=1.258): a detailed look at Table 16 shows that the scores of some individual items (KCG6, KCG8, KGG9) are very low (≤2), due to the employees of $\mathbb{C}_s$ mistakenly deeming that some benign links were actually malicious.

For $\mathbb{C}_m$, the *attitude towards cybersecurity* is very high (avg=4.398). A detailed look at the individual items (Table 13) reveals that $\mathbb{C}_m$'s employees have similar difficulties as those of $\mathbb{C}_s$ (i.e., ACS3 and ACS4, both having averages ≤3.8). Such a tendency also pertains to *self-reported behavior*: despite a higher overall score (avg=4.168) than $\mathbb{C}_s$, the lowest scores pertain to the same items (i.e., BHV2 and BHV3 in Table 14). In contrast, for *CSA training experience*, the score of $\mathbb{C}_m$ (avg=4.091) and of *training usability* (avg=3.982) are substantially higher than those of $\mathbb{C}_s$: this denotes that $\mathbb{C}_m$'s employees believe that the training provided by their company to be useful in protecting them against emerging threats. Finally, for *knowledge and competence gain*, the scores (avg=3.884) are generally much higher than for $\mathbb{C}_s$ indicating that the employees of $\mathbb{C}_m$ may have improved their competences after training.

For $\mathbb{C}_h$, it stands out that the *attitude towards cybersecurity* has the highest score (avg=4.631) among all companies; however, $\mathbb{C}_h$ also had the lowest individual scores for the same two items for which the employees of both $\mathbb{C}_m$ and $\mathbb{C}_s$ struggled (i.e., ACS3 and ACS4 in Table 13). The situation is similar for *self-reported behavior*: $\mathbb{C}_h$ has the highest scores (avg=4.507) and the item with the lowest score was also the lowest for $\mathbb{C}_m$ and $\mathbb{C}_s$ (i.e., BHV3 in Table 14). In terms of *CSA training experience* and *training usability*, the scores (avg=4.667 and 4.351, respectively) were always above 4, denoting that the employees of $\mathbb{C}_h$ appreciate the training they receive—and do so much more than either $\mathbb{C}_s$ or $\mathbb{C}_m$. (These results suggest the employees of $\mathbb{C}_h$ have a similar profile to those of the organization considered by Schiller et al. [84]). Finally, $\mathbb{C}_h$ also had the highest scores for *knowledge and competence gain* (avg=4.103), and the two items for which the scores were the lowest (i.e., KCG3 and KCG4 in Table 16) were the same ones as for $\mathbb{C}_m$.

Summary. The employees of all companies have a strong attitude towards cybersecurity, and their self-reported behavior shows that, in general, they are confident in their own ability to deal with cyber threats. However, while the employees of $\mathbb{C}_m$ and $\mathbb{C}_h$ appreciate the training they receive, and believe that it increases their security awareness, this is not the case for $\mathbb{C}_s$'s employees—who believe that their training is poor and not very useful.

6.2 Statistical Assessment of PPA and PS [RQ3]

To objectively answer RQ3, we use a linear regression [66]. We set our hypothesis as "The perceived phishing awareness (i.e., PPA-score) is not a statistically significant predictor of phishing susceptibility (i.e., PS-score)." The PPA-score of each company, taken from the last row of Table 3, is: 3.108 for $\mathbb{C}_s$, 4.068 for $\mathbb{C}_m$, and 4.393 for $\mathbb{C}_h$. The PS-score of each company (computed by aggregating the "page visited / email read" ratios across $\mathbb{E}_B$, $\mathbb{E}_Q$, $\mathbb{E}_L$—see Table 2) is: 38.24 for $\mathbb{C}_s$, 15.01 for $\mathbb{C}_m$, 7.14 for $\mathbb{C}_h$.

After fitting a linear regression model (shown in Fig. 11 in Appendix E.3), we obtain the following results. First, the coefficient of determination is 1.0, meaning that the model can explain possible variance in the variables (i.e., the model is a perfect fit). Second, the slope is −24.2, with an intercept of 113.45. Third, and more importantly, the p-value is $< .001$. Therefore, we must reject our hypothesis. We further validated this finding via a Spearman's Rank Correlation test [39], obtaining ρ=−1.0 with p-value=0.

Given that the PPA score ranges between 1–5, we use our linear regression model to estimate the corresponding PS-score. For instance, a hypothetical company with a PPA-score=1, an extremely low level of perceived phishing awareness, would be expected to

[15] The significantly lower participation rate of $\mathbb{C}_h$ (w.r.t. $\mathbb{C}_m$ and $\mathbb{C}_s$) is, we believe, due to the intrinsic nature of $\mathbb{C}_h$'s employees. While participation in the phishing simulation was mandatory at $\mathbb{C}_h$, participation in the survey was optional, and employees who were less interested may have chosen not to participate—especially given the lack of compensation. In contrast, $\mathbb{C}_m$ and $\mathbb{C}_s$ are smaller companies with close-knit teams, where the corporate culture may naturally encourage participation in voluntary professional activities such as this survey.

Table 3: Summary of the PPA (mean and variance) for each company. We provide the data used to derive these results in Appendix E.

	$\mathbb{C}_s$		$\mathbb{C}_m$		$\mathbb{C}_h$	
	Mean	Var.	Mean	Var.	Mean	Var.
Attitude towards Cybersecurity	4.223	0.596	4.398	0.416	4.631	0.316
Self-reported Behavior	4.000	0.728	4.168	0.569	4.507	0.387
CSA Training experience	1.667	0.222	4.091	0.550	4.667	0.333
Training Usability	2.792	1.972	3.982	0.797	4.351	0.638
Knowledge & Competence Gain	2.862	1.258	3.884	1.197	4.103	1.176
OVERALL (PPA-score)	3.108	0.955	4.068	0.788	4.393	0.671

have a PS-score=89.26, indicating that nearly 90% of its employees would click through to the phishing webpage upon receiving an email such as $\mathbb{E}_B$, $\mathbb{E}_Q$, or $\mathbb{E}_L$. In contrast, any company with a PPA-score≥4.7 would have a PS-score≤0 (i.e., nobody would land on the phishing webpage if they received an email similar to our $\mathbb{E}_B$, $\mathbb{E}_Q$, or $\mathbb{E}_L$). Of course, these are just extreme scenarios.

Answer to RQ3. We found a strong correlation between the perceived phishing awareness (PPA) and the phishing susceptibility of a company's employees. By estimating the PPA of (a subset of) the employees of a given company, it is possible to predict their phishing susceptibility—potentially of the entire company.

7 Discussion and Critical Analysis

We draw lessons learned from our research (§7.1), discuss limitations of our study (§7.2), and suggest avenues for future work (§7.3). In doing so, we also position our findings within related work.

Moreover, we further examine our findings (discussing, e.g., the report-rate and credential-submitted) in Appendix D.

7.1 Major Findings and Lessons Learned

We distill three lessons learned that can be used as a foundation for future research on the threat of phishing in organizations.

The first lesson learned is that *quishing emails are dangerous*. The findings of RQ1 show that quishing emails have the same effectiveness as traditional click-through emails in "hooking" users to a phishing website. Such a result, combined with the intrinsic and subtle characteristics of quishing emails (§2.1.1), makes this form of phishing attack particularly problematic. In a sense, quishing emails are more threatening than other types of phishing emails—such as those leveraging click-through buttons or plaintext URLs, both of which can be detected more easily via automated mechanisms [71, 95]. Some works (e.g., [97, 106]) propose ways to detect quishing emails; yet, as we argued (in §2.1.1), and as also confirmed by the increasing popularity of QR-codes as a phishing-email vector [5], currently deployed defenses do not seem to be effective. We endorse future studies to put more attention to quishing.

The second lesson learned is that *LLMs, if fed with OSINT, can be very effective at crafting phishing emails*. For instance, for $\mathbb{C}_m$, 31.5% (resp. 14.9%) of the employees who read the email visited the landing webpage (resp. submitted their credentials). This result is worrying, given that $\mathbb{E}_L$ was created without exploiting tactics such as urgency or sense of loss. Notably, for $\mathbb{C}_m$, this email was up to 5 times more effective than either $\mathbb{E}_B$ and $\mathbb{E}_Q$ at "persuading" users (likely because $\mathbb{C}_m$'s employees already use two-factor authentication and do not need to set it up). Our results can be compared to those of some (unpublished) works that measured the effectiveness of OSINT-fed LLM on employees of a single organization: in both [29, 45], ≈11% of the employees who opened the email visited the landing webpage; these numbers are higher than those we obtained for $\mathbb{C}_h$, but much lower than those of $\mathbb{C}_m$ and $\mathbb{C}_h$. Such differences call for more work to explore the effectiveness of OSINT-fed LLMs to write phishing emails against employees of a wide range of companies. Nonetheless, given that *(i)* our approach to craft $\mathbb{E}_L$ can be applied to most large companies, and that *(ii)* LLMs are getting increasingly better at elaborating and generating data [85], we can expect that feeding OSINT to LLMs will become commonplace in crafting phishing emails, which can also be targeting the specific employee rather than their company (as done in [45]). We thus endorse future work to develop ways to counter LLM-written phishing emails.

The third lesson learned is that the *employees' perceived phishing awareness can be a predictor of a company phishing susceptibility*. Importantly, our findings seem to hold across different companies. The outcome of RQ3 complements that of recent works [59, 84]: phishing training may provide a false sense of security because employees who did well during training still fell for some phishing traps. Our study, however, had a different purpose, given that our PPA-score is computed by accounting for five different factors—which are a superset of the "knowledge and competence gain" from CSA training [30]. Nonetheless, our fine-grained results (in Appendix E.3) enable future work to consider different elements to derive a different PPA-score. We stress that the findings of RQ3 were possible thanks to our collaboration with companies which allowed us to *(i)* conduct phishing simulations to gauge the phishing susceptibility, and *(ii)* disseminate a survey to measure the perceived phishing awareness of their employees. Accomplishing such an effort in three different contexts required us to overcome many challenges (see, e.g., Appendix A), which may explain why we could not find papers that shared a similar goal. For instance, some very recent works (e.g., [48, 59, 84]) only consider a single company.

Remark. Our findings depend *(a)* on our chosen companies—predominantly based in Europe, and pertaining to three specific businesses (see Table 1); and *(b)* on our emails—which could have been created in a plethora of other ways. For instance, the generic idea for $\mathbb{E}_B$ and $\mathbb{E}_Q$ was to imitate an email related to institutional credentials (similarly to, e.g., [60, 105]), but there exist other hooks that could have been used to compare a click-through button with a QR code (e.g., [83, 84]). At the same time, there are many ways (e.g., [45, 61, 79]) in which LLMs can be used to craft the email valid for RQ (ours is inspired by [29]). Therefore, *we do not seek to generalize our conclusions*, nor claim generalizability.

7.2 Threat to Validity and Limitations

Our study is complex and entailed carrying out a number of experiments in different companies. Let us discuss the most evident issues that could have threatened the validity of our conclusions.

First, *the fact that, for $\mathbb{E}_Q$ in $\mathbb{C}_h$, we did not have complete control of the experiment*. Since $\mathbb{C}_h$ had recently carried out a quishing simulation on their own, we could not carry out another experiment so soon. Moreover, and unfortunately, $\mathbb{C}_h$ did not collect data pertaining to the "submitted credentials" for their own quishing simulation. However, neither of these facts threaten our conclusions:

- For RQ1, we compare the effectiveness of a quishing vs. click-through phishing email in leading a user to a malicious webpage. Therefore, what the user does *after* landing on such a webpage is outside the scope of RQ1 (see also §3.2). Hence, lack of data on the submitted user credentials is irrelevant for RQ1.
- For RQ3, we consider the PS-score, i.e., the ratio between the "visited webpage / email read", which does not depend on the number of users that submitted their credentials.

Finally, for RQ2, we carry out a qualitative analysis which is based on a different email (i.e., $\mathbb{E}_L$), for which we have all data.

Second, *the fact that the timeline of our analysis may have affected our results.* Indeed, we first carried out the phishing simulations, and a couple of days later we carried out the survey for assessing the PPA (see §4). Hence, it is possible that some answers to the PPA-related questions had been influenced by the recent phishing simulation. However, such a possibility only pertains $\mathbb{C}_s$: the two other companies are used to carry out phishing assessments quite often. Nevertheless, such a possibility does not impact our conclusions whatsoever for RQ1 and RQ2, and it may have had only a minor effect on RQ3 (because $\mathbb{C}_s$ is the smallest company).

Third, *the fact that we have no information on "who" participated in the PPA survey.* Therefore, we do not know how such participants performed in the previous phishing simulations. If we could know what the participants of our PPA survey did with the respective emails, we could carry out a more fine-grained assessment. Unfortunately, we do not have access to such (confidential) data. Nonetheless, we are unsure of how such data could be fairly obtained in the first place: having an employee perform the PPA right after the phishing simulation may bias the results; moreover, we could not force an employee to participate in the PPA survey (which is voluntary), hence complete coverage may be impossible to attain.

Fourth, and extending the previous point, *the small sample size for RQ3.* Only 131 employees participated in our PPA survey, whereas we sent 71 309 emails for our phishing simulations. When answering RQ3, we are using the relatively small sample size (i.e., a few dozen employees per company) as a representative indicator of the entire company—which is a gross generalization. Moreover, only three datapoints have been used to answer RQ3 (i.e., we have the PS-score and PPA-score for three companies) for our linear regression model. However, to validate our results, we also attempted to consider all companies as a single entity (with a PS-score of 7.3% and a PPA-score of 4.06): this yields a fourth datapoint that can be used to compute the linear regression anew—which results in a p-value=.036<.05 which still supports our conclusion. Nonetheless, it is known (see, e.g., [59, 84]) that finding volunteers for similar studies is tough.

7.3 Future Work

Our study opens new grounds for future research addressing the problem of phishing in organizations. In what follows, we emphasize three areas that deserve particular attention.

Countermeasures. Our findings (§7.1) highlight the need for dedicated mitigations. First, defenses against quishing emails are essential although difficult to implement [37]. Server-side approaches are resource-intensive, while client-side solutions offer a promising alternative (e.g., QR-code scanners that automatically flag suspicious URLs [75]). Second, addressing malicious content generated by OSINT-fed LLMs remains an open challenge. Humans struggle to distinguish between human- and LLM-written text [38]. In the phishing context, one option is to apply detectors of machine-generated content (e.g., [42]), though they have known limitations. We expand on these and other potential defenses in Appendix D. We encourage future research to build on our findings, both as motivation and as a foundation for developing effective countermeasures.

More LLMs. Our study relied on GPT-3.5 Turbo, which was the free version of ChatGPT available at the time (early 2024). However, the LLM landscape is rapidly evolving, and newer models often surpass their predecessors. This ongoing progress also benefits attackers, who gain access to increasingly capable (and often free) tools against which countermeasures remain limited. We anticipate that LLMs' ability to craft persuasive, and thus more deceptive, phishing emails will only improve over time [81], thus making findings reported in this paper a 'lower bound' of what is to come. Future research should evaluate the phishing potential of emerging models, including newer versions of ChatGPT and offerings from other vendors (e.g., Claude, Llama, Gemini). Understanding the capabilities of these models, how they fit into offensive practices, and how they affect attackers' *modus operandi* and capabilities is essential for designing effective defenses. LLM providers have begun to acknowledge these risks (see [79]), and—ideally—will support efforts to study and mitigate such threats.

More Organizations. Our study focuses on three companies operating in the financial, hospitality, and manufactoring sectors. However, *our considered phishing threats can affect virtually any organization*—as evidenced by ProofPoint's recent report [10]. For instance, prior work has shown the simplicity of carrying out OSINT operations against employees of critical infrastructures [33], which can be leveraged to craft specific phishing emails against similar organizations. More generally, the public sector (including, e.g., higher education [51], healthcare [40, 74], or government [55]) is at constant risk of phishing attacks. Whereas our findings do not directly map to these organizations (since they have a different workforce), our research methods can be applied to study these complementary contexts—an intriguing avenue for future work.

8 Conclusions and Recommendations

Our study is a stepping stone towards understanding the impact of quishing and LLM-based phishing emails across organizations.

We found that embedding malicious QR-codes in phishing emails has the same effectiveness at luring users to a landing webpage as a traditional click-through button. This result is alarming, given that quishing emails can bypass most filters (as also demonstrated by our experiment). We recommend security developers and researchers alike to prioritize the implementation of automated defenses that can mitigate the widespread usage of malicious QR codes. We also promote the inclusion of quishing in CSA training exercises.

The results of our assessment of LLM-based phishing emails, as well as those of our PPA survey, should also serve as an inspiration for future work. Ultimately, and unfortunately, there are many ways to use LLMs to craft phishing emails. Moreover, every company has different employees. By providing all our results and methods, we hence enable downstream research to carry out similar assessments and compare their results with ours—thereby further expanding our understanding of emerging phishing threats.

Acknowledgements

We thank the anonymous reviewers and our considered companies. This work has been partially supported by the Hilti Foundation, and the INTERSECT project, Grant No. NWA.1162.18.301, and the SeReNity project, grant no. CS.010, funded by the Netherlands Organization for Scientific Research (NWO). Any opinions, findings, conclusions, or recommendations expressed in this work are those of the author(s) and do not necessarily reflect the views of NWO.

References

[1] 2023. A Comprehensive Guide to SMTP Relay: Definitions, Examples, and Best Practices. https://mailtrap.io/blog/smtp-relay/.
[2] 2024. Anti-phishing protection in Microsoft 365. https://learn.microsoft.com/en-us/defender-office-365/anti-phishing-protection-about.
[3] 2024. GoPhish – Open Source Phishing Toolkit. https://getgophish.com/.
[4] 2024. *How are attackers using QR codes in phishing emails and lure documents?* Technical Report. Cisco Talos. https://blog.talosintelligence.com/how-are-attackers-using-qr-codes-in-phishing-emails-and-lure-documents/.
[5] 2024. *Malicious QR Codes: How big of a problem is it, really?* Technical Report. Cisco Talos. https://blog.talosintelligence.com/malicious_qr_codes/.
[6] 2024. Phishing Attack Simulation Training–Microsoft Security. https://microsoft.com/en-us/security/business/threat-protection/attack-simulation-training.
[7] 2024. PhishTank. www.phishtank.org.
[8] 2024. QR Code Phishing Scams Are on the Rise—Don't Get Caught - Hoxhunt. https://www.hoxhunt.com/blog/qr-code-phishing-scams
[9] 2024. Simulate a phishing attack with Attack simulation training. https://learn.microsoft.com/en-us/defender-office-365/attack-simulation-training-simulations.
[10] 2024. *State of the Phish 2024.* Technical Report. ProofPoint. https://www.proofpoint.com/it/resources/threat-reports/state-of-phish.
[11] 2024. *Top Email Threats and Trends.* Technical Report. Barracuda. https://assets.barracuda.com/assets/docs/dms/top-email-threats-and-trends-vol1.pdf.
[12] 2024. Train Your Office 365 Users Against Phishing Attacks using Attack Simulation Training. https://web.archive.org/web/20250116155243/https://o365reports.com/2022/02/16/train-your-office-365-users-against-phishing-attacks-using-attack-simulation-training/.
[13] 2025. Kununu. https://www.kununu.com/.
[14] 2025. LinkedIn. https://www.linkedin.com/.
[15] 2025. *WhatsApp spear phishing campaign uses QR codes to add device.* Technical Report. Malwarebytes Labs. https://www.malwarebytes.com/blog/news/2025/01/whatsapp-spear-phishing-campaign-uses-qr-codes-to-add-device.
[16] Abigail N. W Prah Angela Aba and Otchere Kojo Ennin Opan. 2016. The perceived effectiveness of information security awareness. *Information and Knowledge Management* (2016).
[17] Dania Aljeaid, Amal Alzhrani, Mona Alrougi, and Oroob Almalki. 2020. Assessment of end-user susceptibility to cybersecurity threats in Saudi Arabia by simulating phishing attacks. *Information* (2020).
[18] G Amoah and J Hayfron-Acquah. 2022. QR Code security: mitigating the issue of quishing (QR Code Phishing). *Int. J. Comput. Appl* 975 (2022), 8887.
[19] Giovanni Apruzzese, Mauro Conti, and Ying Yuan. 2022. SpacePhish: The Evasion-Space of Adversarial Attacks against Phishing Website Detectors Using Machine Learning. In *Proc. ACSAC*.
[20] APWG. 2024. *Phishing Activity Trends Report, Q3.* Technical Report. APWG.
[21] Michael Bailey, David Dittrich, Erin Kenneally, and Doug Maughan. 2012. The Menlo report. *IEEE Security & Privacy* (2012).
[22] Shahryar Baki and Rakesh M Verma. 2023. Sixteen years of phishing user studies: What have we learned? *IEEE TDSC* (2023).
[23] Nina Baur and Jörg Blasius. 2014. *Handbuch methoden der empirischen sozialforschung*. Vol. 13. Springer.
[24] Mazal Bethany, Athanasios Galiopoulos, Emet Bethany, Mohammad Bahrami Karkevandi, Nishant Vishwamitra, and Peyman Najafirad. 2024. Large language model lateral spear phishing: A comparative study in large-scale organizational settings. *arXiv:2401.09727* (2024).
[25] Violino Bob. 2023. AI tools such as ChatGPT are generating a mammoth increase in malicious phishing emails. *CNBC* (2023). https://www.cnbc.com/2023/11/28/ai-like-chatgpt-is-creating-huge-increase-in-malicious-phishing-email.html
[26] Pavlo Burda, Luca Allodi, and Nicola Zannone. 2020. Don't forget the human: a crowdsourced approach to automate response and containment against spear phishing attacks. In *WACCO (co-located with IEEE EuroS&P)*.
[27] Pavlo Burda, Abdul Malek Altawekji, Luca Allodi, and Nicola Zannone. 2023. The Peculiar Case of Tailored Phishing against SMEs: Detection and Collective DefenseMechanisms at a Small IT Company. In *WACCO (IEEE EuroS&PW)*.
[28] Alexia Cambon, Brent Hecht, Ben Edelman, Donald Ngwe, Sonia Jaffe, et al. 2023. Early LLM-based Tools for Enterprise Information Workers Likely Provide Meaningful Boosts to Productivity. *Microsoft Research* (2023).
[29] Stephanie Carruthers. 2023. *AI vs. human deceit: Unravelling the new age of phishing tactics.* Technical Report. IBM. https://web.archive.org/web/20241009154831/https://securityintelligence.com/x-force/ai-vs-human-deceit-unravelling-new-age-phishing-tactics/.
[30] Sunil Chaudhary, Vasileios Gkioulos, and Sokratis Katsikas. 2022. Developing metrics to assess the effectiveness of cybersecurity awareness program. *Journal of Cybersecurity* (2022).
[31] Casey Deccio, Tarun Yadav, Nathaniel Bennett, Alden Hilton, Michael Howe, Tanner Norton, Jacob Rohde, Eunice Tan, and Bradley Taylor. 2021. Measuring email sender validation in the wild. In *CoNEXT*.
[32] Rachna Dhamija and J Doug Tygar. 2005. The battle against phishing: Dynamic security skins. In *SOUPS*.
[33] Matthew Edwards, Robert Larson, Benjamin Green, Awais Rashid, and Alistair Baron. 2017. Panning for gold: Automatically analysing online social engineering attack surfaces. *Comp. Secur.* (2017).
[34] Egress. 2024. *Email Security Risk Report.* Technical Report. Egress.
[35] Robert Wall Emerson. 2015. Convenience sampling, random sampling, and snowball sampling: How does sampling affect the validity of research? *Journal of Visual Impairment & Blindness* (2015).
[36] Chiarello Filippo, Giordano Vito, Spada Irene, Barandoni Simone, and Fantoni Gualtiero. 2024. Future applications of generative large language models: A data-driven case study on ChatGPT. *Technovation* (2024).
[37] Jason Ford and Hala Strohmier Berry. 2024. Feasibility of Machine Learning-Enhanced Detection for QR Code Images in Email-based Threats. In *IEEE Cyber Awareness and Research Symposium*.
[38] Joel Frank, Franziska Herbert, Jonas Ricker, Lea Schönherr, Thorsten Eisenhofer, Asja Fischer, Markus Dürmuth, and Thorsten Holz. 2024. A representative study on human detection of artificially generated media across countries. In *S&P*.
[39] Thomas D Gauthier. 2001. Detecting trends using Spearman's rank correlation coefficient. *Environmental forensics* (2001).
[40] William J Gordon, Adam Wright, Ranjit Aiyagari, Leslie Corbo, Robert J Glynn, et al. 2019. Assessment of employee susceptibility to phishing attacks at US health care institutions. *JAMA network open* (2019).
[41] Kristen K Greene, Michelle Steves, Mary Theofanos, Jennifer Kostick, et al. 2018. User context: an explanatory variable in phishing susceptibility. In *USEC*.
[42] Abhimanyu Hans, Avi Schwarzschild, Valeriia Cherepanova, Hamid Kazemi, Aniruddha Saha, Micah Goldblum, Jonas Geiping, and Tom Goldstein. 2024. Spotting LLMs With Binoculars: Zero-Shot Detection of Machine-Generated Text. In *Forty-first International Conference on Machine Learning (ICLR)*.
[43] Julian Hazell. 2023. Large language models can be used to effectively scale spear phishing campaigns. *arXiv:2305.06972* (2023).
[44] Julian Hazell. 2023. Spear phishing with large language models. *arXiv preprint arXiv:2305.06972* (2023).
[45] Fred Heiding, Simon Lermen, Andrew Kao, Bruce Schneier, and Arun Vishwanath. 2024. Evaluating Large Language Models' Capability to Launch Fully Automated Spear Phishing Campaigns: Validated on Human Subjects. *arXiv:2412.00586* (2024).
[46] Fredrik Heiding, Bruce Schneier, Arun Vishwanath, Jeremy Bernstein, and Peter S Park. 2024. Devising and detecting phishing emails using large language models. *IEEE Access* (2024).
[47] Doron Hillman, Yaniv Harel, and Eran Toch. 2023. Evaluating organizational phishing awareness training on an enterprise scale. *Comp. Secur.* (2023).
[48] Grant Ho, Ariana Mirian, Elisa Luo, Khang Tong, Euyhyun Lee, et al. 2025. Understanding the Efficacy of Phishing Training in Practice. In *IEEE S&P*.
[49] Markus Jakobsson. 2018. Two-factor inauthentication–the rise in SMS phishing attacks. *Computer Fraud & Security* (2018).
[50] Francisco Jáñez-Martino, Rocío Alaiz-Rodríguez, Víctor González-Castro, Eduardo Fidalgo, and Enrique Alegre. 2023. A review of spam email detection: analysis of spammer strategies and the dataset shift problem. *Artificial Intelligence Review* (2023).
[51] Matthew L Jensen, Michael Dinger, Ryan T Wright, and Jason Bennett Thatcher. 2017. Training to mitigate phishing attacks using mindfulness techniques. *Journal of Management Information Systems* (2017).
[52] Prasanna Kansakar, Arslan Munir, and Neda Shabani. 2019. Technology in the hospitality industry: Prospects and challenges. *IEEE Consumer Electronics Magazine* (2019).
[53] Rabimba Karanjai. 2022. Targeted phishing campaigns using large scale language models. *arXiv:2301.00665* (2022).
[54] Doowon Kim, Haehyun Cho, Yonghwi Kwon, Adam Doupé, Sooel Son, Gail-Joon Ahn, and Tudor Dumitras. 2021. Security analysis on practices of certificate authorities in the HTTPS phishing ecosystem. In *ACM AsiaCCS*.
[55] Michael Koddebusch. 2022. Exposing the phish: the effect of persuasion techniques in phishing e-mails. In *Annual International Conference on Digital Government Research*.
[56] Tadayoshi Kohno, Yasemin Acar, and Wulf Loh. 2023. Ethical frameworks and computer security trolley problems: Foundations for conversations. In *USENIX Sec.*

[57] Takashi Koide, Naoki Fukushi, Hiroki Nakano, and Daiki Chiba. 2023. PhishReplicant: A Language Model-based Approach to Detect Generated Squatting Domain Names. In *ACSAC*.

[58] Katharina Krombholz, Peter Frühwirt, Peter Kieseberg, Ioannis Kapsalis, Markus Huber, and Edgar Weippl. 2014. QR code security: A survey of attacks and challenges for usable security. In *HAS (part of HCI International)*. Springer.

[59] Daniele Lain, Tarek Jost, Sinisa Matetic, Kari Kostiainen, and Srdjan Capkun. 2024. Content, Nudges and Incentives: A Study on the Effectiveness and Perception of Embedded Phishing Training. In *ACM CCS*.

[60] Daniele Lain, Kari Kostiainen, and Srdjan Čapkun. 2022. Phishing in organizations: Findings from a large-scale and long-term study. In *S&P*.

[61] Tyson Langford and Bryson Payne. 2023. Phishing faster: Implementing chatgpt into phishing campaigns. In *Future Technologies Conference*.

[62] Jehyun Lee, Farren Tang, Pingxiao Ye, Fahim Abbasi, Phil Hay, and Dinil Mon Divakaran. 2021. D-fence: A flexible, efficient, and comprehensive phishing email detection system. In *IEEE EuroS&P*.

[63] Tian Lin, Daniel E Capecci, Donovan M Ellis, Harold A Rocha, Sandeep Dommaraju, Daniela S Oliveira, and Natalie C Ebner. 2019. Susceptibility to spear-phishing emails: Effects of internet user demographics and email content. *ACM Transactions on Computer-Human Interaction (TOCHI)* (2019).

[64] M Lokesh, A Keerthi Devi, U Dinesh Chowdary, PVNS Divya Lakshmi, and G Rama Koteswara Rao. 2023. Data Redundancy, Data Phishing, and Data Cloud Backup. In *IEEE ICECCT*.

[65] Charalampos Manifavas, Konstantinos Fysarakis, Konstantinos Rantos, and George Hatzivasilis. 2014. DSAPE–dynamic security awareness program evaluation. In *HAS (part of HCI International)*. Springer.

[66] Douglas C Montgomery, Elizabeth A Peck, and G Geoffrey Vining. 2021. *Introduction to linear regression analysis*. John Wiley & Sons.

[67] Santana N Morris. 2023. Cultural diversity in workplace and the role of management. *American Journal of Industrial and Business Management* (2023).

[68] Aleksandr Nahapetyan, Sathvik Prasad, Kevin Childs, Adam Oest, Yeganeh Ladwig, Alexandros Kapravelos, and Bradley Reaves. 2024. On sms phishing tactics and infrastructure. In *IEEE Symposium on Security and Privacy*.

[69] Geoff Norman. 2010. Likert scales, levels of measurement and the "laws" of statistics. *Advances in health sciences education* (2010).

[70] Adam Oest, Yeganeh Safaei, Penghui Zhang, Brad Wardman, Kevin Tyers, Yan Shoshitaishvili, and Adam Doupé. 2020. {PhishTime}: Continuous longitudinal measurement of the effectiveness of anti-phishing blacklists. In *USENIX Security*.

[71] Justin Petelka, Yixin Zou, and Florian Schaub. 2019. Put your warning where your link is: Improving and evaluating email phishing warnings. In *ACM CHI*.

[72] Iasonas Polakis, Georgios Kontaxis, Spiros Antonatos, Eleni Gessiou, Thanasis Petsas, and Evangelos P Markatos. 2010. Using social networks to harvest email addresses. In *ACM WPES*.

[73] Rana Pourmohamad, Steven Wirsz, Adam Oest, Tiffany Bao, Yan Shoshitaishvili, et al. 2024. Deep Dive into Client-Side Anti-Phishing: A Longitudinal Study Bridging Academia and Industry. In *AsiaCCS*.

[74] Ward Priestman, Tony Anstis, Isabel G Sebire, Shankar Sridharan, and Neil J Sebire. 2019. Phishing in healthcare organisations: Threats, mitigation and approaches. *BMJ Health & Care Informatics* (2019).

[75] Ahmad Sahban Rafsanjani, Norshaliza Binti Kamaruddin, Hazlifah Mohd Rusli, and Mohammad Dabbagh. 2023. Qsecr: Secure qr code scanner according to a novel malicious url detection framework. *IEEE Access* (2023).

[76] Rakesh Rana and Richa Singhal. 2015. Chi-square test and its application in hypothesis testing. *Journal of Primary Care Specialties* (2015).

[77] Konstantinos Rantos, Konstantinos Fysarakis, and Charalampos Manifavas. 2012. How effective is your security awareness program? An evaluation methodology. *Information Security Journal: A Global Perspective* (2012).

[78] Benjamin Reinheimer, Lukas Aldag, Peter Mayer, Mattia Mossano, Reyhan Duezguen, Bettina Lofthouse, Tatiana Von Landesberger, and Melanie Volkamer. 2020. An investigation of phishing awareness and education over time: When and how to best remind users. In *SOUPS*.

[79] Sayak Saha Roy, Poojitha Thota, Krishna Vamsi Naragam, and Shirin Nilizadeh. 2024. From Chatbots to Phishbots?: Phishing Scam Generation in Commercial Large Language Models. In *IEEE S&P*.

[80] Said Salloum, Tarek Gaber, Sunil Vadera, and Khaled Shaalan. 2021. Phishing email detection using natural language processing techniques: a literature survey. *Procedia Computer Science* 189 (2021).

[81] Francesco Salvi, Manoel Horta Ribeiro, Riccardo Gallotti, and Robert West. 2024. On the conversational persuasiveness of large language models: A randomized controlled trial. *Nature Human Behaviour* (2024).

[82] Orvila Sarker, Asangi Jayatilaka, Sherif Haggag, Chelsea Liu, and M Ali Babar. 2024. A Multi-vocal Literature Review on challenges and critical success factors of phishing education, training and awareness. *J. Systems and Software* (2024).

[83] Dawn M Sarno and Mark B Neider. 2022. So many phish, so little time: Exploring email task factors and phishing susceptibility. *Human Factors* (2022).

[84] Katharina Schiller, Florian Adamsky, Christian Eichenmüller, Matthias Reimert, and Zinaida Benenson. 2024. Employees' Attitudes towards Phishing Simulations:"It's like when a child reaches onto the hot hob". In *ACM CCS*.

[85] Philipp Schoenegger, Indre Tuminauskaite, Peter S Park, Rafael Valdece Sousa Bastos, and Philip E Tetlock. 2024. Wisdom of the silicon crowd: LLM ensemble prediction capabilities rival human crowd accuracy. *Science Advances* (2024).

[86] Paul G Schrader and Kimberly A Lawless. 2004. The knowledge, attitudes, & behaviors approach how to evaluate performance and learning in complex environments. *Performance Improvement* (2004).

[87] Saskia Laura Schröer, Giovanni Apruzzese, Human Soheil, Pavel Laskov, Hyrum S. Anderson, et al. 2025. SoK: On the Offensive Potential of AI. In *IEEE SaTML*.

[88] Donald J Schuirmann. 1987. A comparison of the two one-sided tests procedure and the power approach for assessing the equivalence of average bioavailability. *Journal of pharmacokinetics and biopharmaceutics* (1987).

[89] Filipo Sharevski, Amy Devine, Emma Pieroni, and Peter Jachim. 2022. Phishing with malicious QR codes. In *EuroUSEC*.

[90] Filipo Sharevski, Mattia Mossano, Maxime Fabian Veit, Gunther Schiefer, and Melanie Volkamer. 2024. Exploring Phishing Threats through QR Codes in Naturalistic Settings. In *USEC*.

[91] Hossein Siadati, Sean Palka, Avi Siegel, and Damon McCoy. 2017. Measuring the effectiveness of embedded phishing exercises. In *USENIX CSET 17*.

[92] Slashnext. 2023. *The State of Phishing*. Technical Report. Slashnext. https://slashnext.com/wp-content/uploads/2023/10/SlashNext-The-State-of-Phishing-Report-2023.pdf.

[93] Stephanie Stacey. 2025. AI-generated phishing scams target corporate executives. *Financial Times* (2025). https://www.ft.com/content/d60fb4fb-cb85-4df7-b246-ec3d08260e6f

[94] Michelle Steves, Kristen Greene, and Mary Theofanos. 2020. Categorizing human phishing difficulty: a Phish Scale. *Journal of Cybersecurity* (2020).

[95] Karthika Subramani, William Melicher, Oleksii Starov, Phani Vadrevu, and Roberto Perdisci. 2022. PhishInPatterns: measuring elicited user interactions at scale on phishing websites. In *ACM Internet Measurement Conference*.

[96] Zhibo Sun, Faris Bugra Kokulu, Penghui Zhang, Adam Oest, Gianluca Stringhini, et al. 2024. From Victims to Defenders: An Exploration of the Phishing Attack Reporting Ecosystem. In *RAID*.

[97] Saranya Vaithilingam and Santhosh Aradhya Mohan Shankar. 2024. Enhancing Security in QR Code Technology Using AI: Exploration and Mitigation Strategies. *International Journal of Intelligence Science* (2024).

[98] Rohit Valecha, Pranali Mandaokar, and H Raghav Rao. 2021. Phishing email detection using persuasion cues. *IEEE TDSC* (2021).

[99] Suresh Veluru, Yogachandran Rahulamathavan, P Viswanath, Paul Longley, and Muttukrishnan Rajarajan. 2013. E-mail Address Categorization based on Semantics of Surnames. In *IEEE CIDM*.

[100] Thomas Weber, Maximilian Brandmaier, Albrecht Schmidt, and Sven Mayer. 2024. Significant Productivity Gains through Programming with Large Language Models. *Proceedings of the ACM on Human-Computer Interaction* (2024).

[101] Emma J Williams, Joanne Hinds, and Adam N Joinson. 2018. Exploring susceptibility to phishing in the workplace. *Int. J. Human-Computer Studies* (2018).

[102] Emma J Williams and Danielle Polage. 2019. How persuasive is phishing email? The role of authentic design, influence and current events in email judgements. *Behaviour & Information Technology* (2019).

[103] Rundong Yang, Kangfeng Zheng, Bin Wu, Di Li, Zhe Wang, and Xiujuan Wang. 2022. Predicting User Susceptibility to Phishing Based on Multidimensional Features. *Computational Intelligence and Neuroscience* (2022).

[104] Ezer Osei Yeboah-Boateng and Priscilla Mateko Amanor. 2014. Phishing, SMiShing & Vishing: an assessment of threats against mobile devices. *Journal of Emerging Trends in Computing and Information Sciences* (2014).

[105] William Yeoh, He Huang, Wang-Sheng Lee, Fadi Al Jafari, and Rachel Mansson. 2022. Simulated phishing attack and embedded training campaign. *Journal of Computer Information Systems* (2022).

[106] Kelvin SC Yong, Kang Leng Chiew, and Choon Lin Tan. 2019. A survey of the QR code phishing: the current attacks and countermeasures. In *IEEE ICSCC*.

[107] Ying Yuan, Qingying Hao, Giovanni Apruzzese, Mauro Conti, and Gang Wang. 2024. "Are Adversarial Phishing Webpages a Threat in Reality?" Understanding the Users' Perception of Adversarial Webpages. In *TheWebConf*.

[108] Tianyi Zhang, Faisal Ladhak, Esin Durmus, Percy Liang, Kathleen McKeown, and Tatsunori B Hashimoto. 2024. Benchmarking large language models for news summarization. *T. of the Association for Computational Linguistics* (2024).

Appendix A Technical details (and challenges)

We expand the information provided in §4.2, which covers our setup for $\mathbb{C}_s$ (for which we show our custom landing webpage in Fig. 3). Additionally, we provide the prompts used to craft $\mathbb{E}_L$ in Table 4.

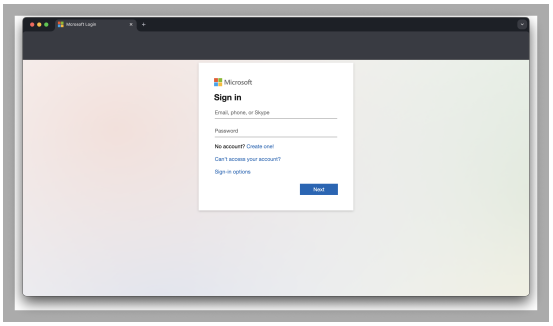


Fig. 3: Landing page. All of our emails would point to a webpage with a similar design as this one, showing the typical "Microsoft login".

Table 4: Sequence of Prompts used to craft $\mathbb{E}_L$. Text in regular font are not part of the prompt; the last prompt is optional. We do not show the prompts used to "jailbreak" the model (to avoid helping attackers).

#	Prompt
1	`Please help me summarize the weaknesses this company has according to this employer rating website.` [Extra input: data extracted from Kununu]
2	`If I were an attacker, which weakness would be the best to leverage in a phishing attack?`
3	`Please give me one concrete example of a potential phishing mail leveraging this weakness.`
4	`Please analyse these postings for me and give me the 5 most common topics that this company cares about.` [Extra input: data extracted from LinkedIn]
5	`Please write me a brief introduction to a company survey directed at employees regarding the latest company efforts in relation to [topic from prompt #4] at [company]. The introduction is meant to accompany the link to the survey. Here is some additional information the employees are already aware of.` [Extra input: text from press releases]
	`Shorter please` [Note: only added if the output was longer than 100 words so that it would still be readable]

A.1 Microsoft Defender

The simulations for $\mathbb{C}_m$ and $\mathbb{C}_h$ leveraged the "Microsoft Defender Attack Simulation Training" (MADST) module, which is part of Microsoft's Office 365 licensing for large enterprises [6].

This module helps organizations to run realistic simulations in their workplace using Microsoft's own ecosystem. We show in Fig. 4 a sample visualization of its interface. For $\mathbb{C}_m$ and $\mathbb{C}_h$, we used the same module deployed by the respective security team. Given the highly-confidential nature of our research, we spent a lot of time discussing with $\mathbb{C}_m$ and $\mathbb{C}_h$ so that we could find an agreement on how to use their framework for our experiments.

Among the greatest challenges we encountered was integrating the QR-code email simulation in MADST. Indeed, MADST does not support QR-code emails natively. Hence, we had to deploy a dedicated webapp that would create a custom QR code from the specific link of an email; doing this was not simple from a bureaucratic viewpoint, given the "external" nature of such a webapp.

Nonetheless, the MADST module supports various "attack scenarios". For our experiments, we opted for the "credential harvesting scenario" (see [9]), since it aligned with our goals and was also supported by GoPhish.

Finally, for $\mathbb{C}_h$ and $\mathbb{C}_m$, we also relied on their "feedback" page (similar to the one shown in [12]: we cannot show the actual one due to NDA) that informs users who submitted their credentials that they have been "phished". We did not do this for for $\mathbb{C}_s$ since it was not deemed necessary by their representatives. Indeed, due to the small size of $\mathbb{C}_s$, it was possible to directly reach out to each user who submitted their credentials and let them know that they "fell" in a phishing trap. Regardless, such a discrepancy has no impact on the goal of our study.

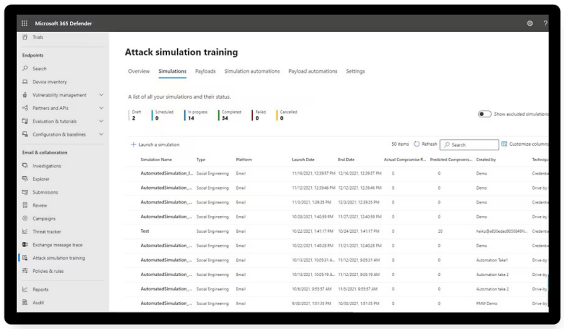


Fig. 4: Interface of Microsoft Defender Attack Simulation module. This is just an example, no confidential information is shown.

A.2 GoPhish

GoPhish is an open-source phishing framework written in the programming language "Go" [3]. GoPhish seeks to make phishing assessments and training available and accessible for everyone. We provide in Fig. 5 a sample of GoPhish dashboard. We used GoPhish for $\mathbb{C}_s$: unfortunately, setting up GoPhish for our experiment revealed to be much more complex than what we had foreseen.

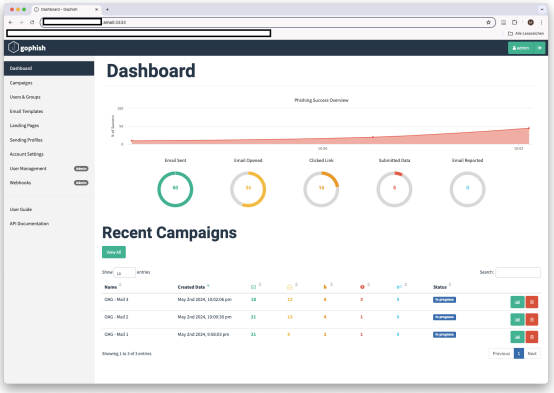


Fig. 5: Interface of GoPhish. This is just an example, no confidential information is shown.

Internet deployment. We deployed our instance of GoPhish on a virtual private server (VPS) "exposed" to the internet, since we needed it to be operational and accessible during the entire time of our simulations. To this end, we licensed a small (specs: OS=Ubuntu 23.10; CPU Type=Regular Intel (1 CPU); RAM=512MB; SSD=10GB; Location=Germany) VPS with a well-known cloud-service provider, which cost us 4$ per month. We also purchased a domain (which would resemble $\mathbb{C}_s$'s name, to which we added ".email" to it) with a well-known domain registrar, which cost us 9$. Such a cost was necessary to increase the realistic fidelity of our campaign: otherwise, users could be suspicious if, e.g., they saw IP

addresses or weird domains in the emails they received. To further increase the credibility of our infrastructure, we also set up an SSL certificate for our domain (we used ZeroSSL, which provided a free service for the first 90 days).

SMTP relay. The provider of our VPS automatically blocks the SMTP protocol ("to avoid misuse by criminals"). Unfortunately, such a protocol was, of course, required for our simulation. To overcome this challenge, we set up an SMTP relay [1]. This required us to authenticate our domain by creating four CNAME records, and add them to the VPS. This way, we verified our domain and were able to set up the sender identity (which we chose as described in §4.2.1). The SMTP relay we chose was free for up to 100 emails per day, so we did not incur in any costs—given the small size of $\mathbb{C}_s$.

Blocked emails. Once we set up the abovementioned infrastructure, we began doing some tests by sending some emails. Unfortunately, we found that such emails, when sent to Gmail addresses, were blocked by Google's spam filters and put in the "Junk" folder; other email providers blocked the emails entirely (no email was received even after 48 hours). To overcome this problem, we agreed to have $\mathbb{C}_s$ enter the sender of our emails among the whitelisted senders for $\mathbb{C}_s$ mail client. However, the issues did not stop here: when we tested the landing page we created for $\mathbb{C}_s$ (hosted on the VPS), we found out that it had been blocked by Google's SafeBrowsing. We reached out to Google, reporting the page as benign: thankfully, the block was lifted after 24hours.

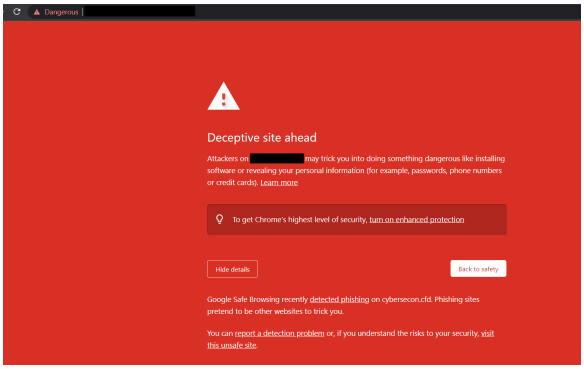


Fig. 6: Our landing page was initially blocked by Google SafeBrowsing. We reached out to Google who lifted the block after 24 hours.

Appendix B Do quishing emails evade operational detectors?

To provide real-world evidence that QR-code emails are "more stealthy" than traditional URL-based phishing emails, we carried out an original experiment on an *operational detector*.[16] In November 2024, we retrieved a malicious URL (i.e., https://arub330011.page.link/jdF1) from phishtank [7] (see Fig. 7a). We first verified that it was included among common blocklist: we tried visiting the URL, and we were shown a warning webpage (see Fig. 7c). Then, we generated a QR-code for such a URL (see Fig. 7b). At this point, we sent four emails—all from the same email account (i.e., the personal Gmail account of one of the authors) to the same email account (i.e., the institutional email account of the same author). Specifically:

- The first email was just a sanity check, and it simply included a link to a well-known (benign) website, asking the recipient to "click on the link". This email, as expected, was put in the *inbox folder* of the recipient account.
- The second email was the URL-based phishing email: it was the same as the first email, but instead of the benign link we put the malicious link mentioned above. This email was put in the *junk folder* of the recipient account.
- The third email was an exemplary quishing email in which the QR code was provided as an image attachment; the text invites the reader to "check out the link in the qr code". This email was put in the *inbox folder* of the recipient account.
- The fourth email was also a quishing email in which we put the QR code in the body of the email (i.e., as an HTML object). This email was put in the *inbox folder* of the recipient account.

The four emails above all had the same subject ("2FA") and had been sent within a timespan of 7 minutes. It is possible to visualize the results of this experiment in Fig. 8. These results demonstrate that both quishing emails have "evaded" a commercial phishing/spam filter—despite the corresponding URL-based phishing email being (correctly) deemed as junk.

Appendix C Systematic Literature Review

Quishing has been somehow overlooked by prior research—at least from the viewpoint of papers focused on user studies.

Indeed, we have systematically analysed the 2014–2024 proceedings of 11 top-venues related to Security, Human Factors and the Web: WWW, S&P, EuroS&P, CCS, USENIX SEC, NDSS, AsiaCCS, ACSAC, IMC, WSDM, CHI. We searched for full papers (excluding, e.g., workshops) having "phish" in the title and found 66 papers. Then, we inspected their text, searching for occurrences of the term "qr". We found only two papers with such a string: [73] (where "qr" was mentioned only once) and [79] (here, it occurs 13 times). Neither of these, however, carried out user studies.

More generally, however, we can state that QR-code phishing is not a commonly researched theme among these 11 top-tier venues.

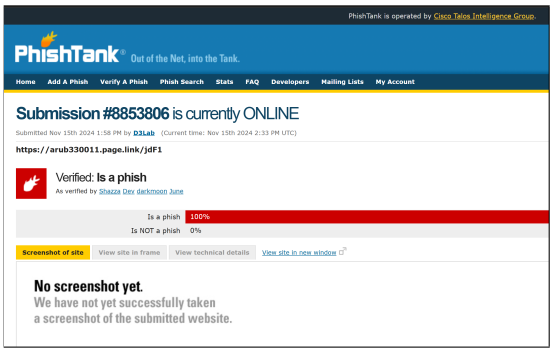
Appendix D Additional Considerations

We provide some additional critical remarks on our research, and further support our arguments with original analyses.

Credentials submitted. For RQ1, we did not consider what happens after the user lands on the webpage. A look at these results in Table 2 shows that, for $\mathbb{E}_Q$, a comparatively lower number of users submitted their credentials with respect to $\mathbb{E}_B$ (around 33% less for $\mathbb{C}_m$ and $\mathbb{C}_s$). This result may suggest that even though quishing emails have the same effectiveness in terms of bringing a potential victim to a phishing webpage, such a victim may be somewhat more reluctant to submit their credentials. A possible explanation of this result, however, lies in the experimental setup of our experiments. Users were ultimately required to type their userid and passwords, which could be stored in a password manager accessible only from, e.g., the laptop or desktop used for work. If this is true, then less users submitted their credentials for $\mathbb{E}_Q$ (w.r.t. $\mathbb{E}_B$) because such users simply *could not do so*—given that the landing webpage was visited on a smartphone, i.e., the device used to scan the QR code.

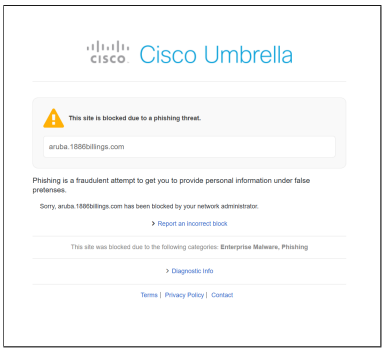
Reported emails. Let us provide some remarks on the number of "reported" emails of our simulation—especially with regard to

[16] The infrastructure that supports the institutional email services of (some of) the authors of this paper is provided by Microsoft, which integrates anti-phishing tools [2].



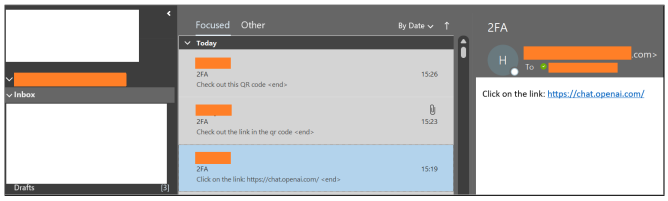
(a) Details of the malicious URL (https://arub330011.page.link/jdF1) according to Phishtank [7] (in November 2024).

(b) QR-code of the malicious URL used as a basis for this experiment.



(c) Verification that the URL was known to be malicious by well-known providers (e.g., CISCO).

Fig. 7: Original QR-code test: preliminaries. We took a URL pointing to a phishing webpage from Phishtank (Fig. 7a), we generated the corresponding QR code (Fig. 7b) and also checked that the webpage had been included in operational blocklists (Fig. 7c) used by popular browsers.



(a) Aside from the "benign" email (sent at 15:19), the two "quishing" emails (one having the QR code as attachment, the other embedded in the email's content) were not put in the junk folder—thereby evading the phishing/spam detection filter.

(b) The only mail put in the "junk" folder was the one with the URL in plaintext.

Fig. 8: Original QR-code test: results. We sent four emails to the institutional email address of one of the authors, managed by Microsoft (i.e., the same provider of the companies considered in our paper—see §4.1).

$\mathbb{C}_s$. Recall that, across $\mathbb{E}_B/\mathbb{E}_Q/\mathbb{E}_L$, 11 employees of $\mathbb{C}_s$ reached out to the IT managers about the "phishing" emails they had just received (note: 60 emails were sent in $\mathbb{C}_s$). What is intriguing, however, is that $\mathbb{C}_s$ does not have a dedicated security team (see §4.1). Likely, such employees were somewhat suspicious of the email and opted for the easiest way of support they were aware of: contacting the most IT-savvy person in the company to ask for advice. Such an occurrence highlights the benefits of a low hierarchy organization (typical of small companies) and their close-knit structures of communication on phishing susceptibility. This aligns with the findings of Burda et al. [27], that phishing attacks towards SMEs can be stopped by the high level of direct communication, thus leading to users being alerted by coworkers quickly after an attack has been discovered. Comparatively speaking, 21.5% (resp. 40.2%) of the emails sent in $\mathbb{C}_h$ (resp. $\mathbb{C}_m$) had been reported. Intriguingly, some prior works carried out in different contexts found that the "report ratio" of phishing emails (in a simulation) tends to be much lower—typically below 10% (e.g., [26, 59]). A potential explanation lies in the heterogeneity of the reporting ecosystem across companies [96]: for instance, a company that makes it easy to report emails (e.g., via a dedicated button) and that encourages their employees to be suspicious is likely to have a higher report rate [84]. Evidence that this is the case can be found in the results of our PPA questionnaire: many of our participants (across all companies) thought that a benign link was actually suspicious—which reflects an overly skeptical attitude. Hence, our numbers suggest that our considered companies promote reporting of suspicious emails.

Some potential countermeasures. Let us expand our suggestions in Section §7.3 with additional insight and justifications.

- *Defenses against quishing emails* are tough to realize. The issue is that implementation of automated mechanisms that can reliably detect the presence of (malicious) QR-codes in emails is a hard problem [37]. This is because it is not known, a priori, if an email contains a QR-code—and, if so, where (e.g., it can be in an attachment, or embedded as an HTML object). Therefore, such mechanisms would necessitate a thorough scan of every email received by a given user, which would pose a lot of stress to the respective servers. Client-side solutions (e.g., implementing QR-code scanners with automated mechanisms that warn a user of a suspicious URL [75]; or dedicated browser-extensions that perform their analyses at the client level) may not present substantial computational overhead, but may not be universally applicable, and/or may induce software lock-in. Nevertheless, as also recommended by [89], we advocate future work to emphasize the importance of "quishing education/training": users should be aware that QR-codes may conceal cyber threats, and hence users should not blindly trust (and visit) the URLs interpreted by any given QR-code scanner.
- *Dealing with malicious content generated by (OSINT-fed) LLMs* is an open issue. Humans can hardly distinguish human- from LLM-generated content [38]. In the context of (OSINT-fed) LLM-written phishing emails, a plausible mitigation entails using detectors of machine-generated text (e.g., [42]). For instance, if *(i)* a company tells its employees that company-related emails are not written by LLMs, then *(ii)* an automated mechanism that warns an employee that a given email contains LLM-written text which allegedly is company-related would induce the user to be more suspicious of the legitimacy of such an email.

We stress that the aforementioned mitigations (which are based on our educated guesses) *should not be taken as universal solutions to these problems*. Firstly, because they present tradeoffs; secondly, because they can be exploited by attackers (e.g., detectors of machine-generated text are not perfect, and require constant updates to be able to detect content generated by state-of-the-art LLMs).

Table 5: Questionnaire for measuring the PPA. Some questions (e.g., KCG1–5) have been provided with links or emails that were specific to the corresponding company (we cannot provide more details due to NDA). We did not use CSA1 in our main paper because, ultimately, nobody filled it for $\mathbb{C}_s$.

ID	Category	Item	Question (English version)	Question (German version)
1	Attitude towards Cybersecurity	ACS1	I believe cybersecurity is important for protecting my personal information and online accounts.	Ich glaube, dass Cybersicherheit wichtig ist, um meine persönlichen Daten und Online-Konten zu schützen.
2	Attitude towards Cybersecurity	ACS2	I feel confident in my ability to identify cyber threats.	Ich fühle mich sicher in meiner Fähigkeit, Cyber-Bedrohungen zu erkennen.
3	Attitude towards Cybersecurity	ACS3	I feel confident in my ability to protect myself from cyber threats.	Ich habe Vertrauen in meine Fähigkeit, mich vor Cyber-Bedrohungen zu schützen.
4	Attitude towards Cybersecurity	ACS4	I believe that everyone has a role to play in protecting against cyber threats.	Ich glaube, dass jeder eine Rolle beim Schutz vor Cyber-Bedrohungen spielen muss.
5	Attitude towards Cybersecurity	ACS5	I feel a sense of responsibility to protect myself and others from cyber threats.	Ich fühle mich dafür verantwortlich, mich und andere vor Cyber-Bedrohungen zu schützen.
6	Attitude towards Cybersecurity	ACS6	I believe that staying informed about cybersecurity helps me to react effectively to unexpected situations.	Ich glaube, dass es mir hilft, auf unerwartete Situationen effektiv zu reagieren, wenn ich über Cybersicherheit informiert bin.
7	Attitude towards Cybersecurity	ACS7	I believe that cyber threats are becoming more common and sophisticated.	Ich glaube, dass Cyber-Bedrohungen immer häufiger und raffinierter werden.
8	Attitude towards Cybersecurity	ACS8	I am willing to take steps to improve my cybersecurity practices.	Ich bin bereit, Maßnahmen zu ergreifen, um meine Cybersicherheitspraktiken zu verbessern.
9	Attitude towards Cybersecurity	ACS9	I am willing to adapt my cybersecurity practices to new threats and challenges.	Ich bin bereit, meine Cybersicherheitspraktiken an neue Bedrohungen und Herausforderungen anzupassen.
10	Self-reported Behavior	BHV1	I believe that increased awareness of phishing scams would help to reduce the overall level of cybersecurity risky behavior.	Ich glaube, dass eine stärkere Sensibilisierung für Phishing-Betrügereien dazu beitragen würde, das allgemeine Risikoverhalten im Bereich der Cybersicherheit zu verringern.
11	Self-reported Behavior	BHV2	I believe that I am less likely to click on suspicious links or open attachments in emails because I am afraid of being phished.	Ich glaube, dass ich weniger wahrscheinlich auf verdächtige Links klicke oder Anhänge in E-Mails öffne, weil ich Angst habe, Opfer eines Phishings zu werden.
12	Self-reported Behavior	BHV3	My understanding of phishing scams affects my overall behavior when working with a digital device.	Mein Verständnis von Phishing-Betrug beeinflusst mein allgemeines Verhalten bei der Arbeit mit einem digitalen Gerät.
13	Self-reported Behavior	BHV4	I am confident in my ability to identify phishing emails.	Ich habe Vertrauen in meine Fähigkeit, Phishing-E-Mails zu erkennen.
14	Self-reported Behavior	BHV5	I recognized and avoided a phishing scams at least once in the past thanks to my phishing education.	Ich habe in der Vergangenheit mindestens einmal einen Phishing-Betrug dank meiner Phishing-Aufklärung erkannt und vermieden.
15	CSA Training experience	CSA1	Have you ever participated in an organizational cybersecurity training?	Haben Sie jemals an einer organisatorischen Cybersicherheitsschulung teilgenommen?
16	CSA Training experience	CSA2	I regularly complete my organization's cybersecurity awareness training.	Ich nehme regelmäßig an den Cybersicherheitsschulungen meiner Organisation teil.
17	Training Usability	TUB1	I believe that the information presented in my organization's cybersecurity awareness training was relevant and applicable to my work or personal life.	Ich bin der Meinung, dass die Informationen, die in den Schulungen meines Unternehmens zum Thema Cybersicherheit vermittelt wurden, für meine Arbeit relevant und anwendbar waren.
18	Training Usability	TUB2	I feel more knowledgeable about cybersecurity threats and prevention methods since participating in my organization's cybersecurity awareness training.	Ich fühle mich besser informiert über Bedrohungen der Cybersicherheit und Präventionsmethoden, seit ich an der Schulung zum Thema Cybersicherheit in meiner Organisation teilgenommen habe.
19	Training Usability	TUB3	I feel more confident in my ability to protect myself and my organization from cyber threats since participating in my organization's cybersecurity awareness training.	Ich bin zuversichtlicher, dass ich mich und mein Unternehmen vor Cyber-Bedrohungen schützen kann, seit ich an der Schulung teilgenommen habe.
20	Training Usability	TUB4	I am more likely to apply the knowledge and skills I learned in my organization's cybersecurity awareness training to my future cybersecurity practices.	Ich werde das Wissen und die Fähigkeiten, die ich in der Schulung zum Bewusstsein für Cybersicherheit in meiner Organisation gelernt habe, in meinen zukünftigen Cybersecurity-Praktiken eher anwenden.
21	Training Usability	TUB5	I feel more prepared to protect myself from online threats after learning about phishing in my company's cybersecurity awareness training.	Ich fühle mich besser vorbereitet, mich vor Online-Bedrohungen zu schützen, nachdem ich in der Schulung meines Unternehmens über Phishing gelernt habe.
22	Training Usability	TUB6	I believe that the skills I learned in my organization's cybersecurity awareness training will help me to better identify and respond to cybersecurity threats.	Ich glaube, dass die Fähigkeiten, die ich in der Cybersecurity-Schulung meines Unternehmens gelernt habe, mir helfen werden, Cybersecurity-Bedrohungen besser zu erkennen und auf sie zu reagieren.
23	Training Usability	TUB7	The topics covered in my organization's cybersecurity awareness training were relevant to my work and private life.	Die Themen, die in der Schulung zum Thema Cybersicherheit in meinem Unternehmen behandelt wurden, waren für meine Arbeit und mein Privatleben relevant.
24	Training Usability	TUB8	I found my organization's cybersecurity awareness training to be informative and engaging.	Ich fand die Schulung zum Thema Cybersicherheit in meinem Unternehmen informativ und ansprechend.
25	Knowledge and competence gain	KCG1	How suspicious are you of this link? (malicious link)	Wie misstrauisch sind Sie gegenüber diesem Link?
26	Knowledge and competence gain	KCG2	How suspicious are you of this link? (malicious link 2)	Wie misstrauisch sind Sie gegenüber diesem Link?
27	Knowledge and competence gain	KCG3	How suspicious are you of this link? (benign link)	Wie misstrauisch sind Sie gegenüber diesem Link?
28	Knowledge and competence gain	KCG4	How suspicious are you of this email? (benign email)	Wie misstrauisch sind Sie gegenüber dieser E-Mail?
29	Knowledge and competence gain	KCG5	How suspicious are you of this email? (malicious email)	Wie misstrauisch sind Sie gegenüber dieser E-Mail?
30	Knowledge and competence gain	KCG6	An email from my colleague cannot be a phishing email.	Eine E-Mail von meinem Kollegen kann keine Phishing-E-Mail sein.
31	Knowledge and competence gain	KCG7	Phishing emails always contain grammatical errors or poor spelling.	Phishing-E-Mails enthalten immer grammatikalische Fehler oder schlechte Rechtschreibung.
32	Knowledge and competence gain	KCG8	Phishing scams are not only a threat to people who use personal computers; mobile devices are susceptible to phishing attacks, too.	Phishing-Betrügereien sind nicht nur eine Bedrohung für Menschen, die einen Computer benutzen; auch mobile Geräte sind anfällig für Phishing-Angriffe.
33	Knowledge and competence gain	KCG9	Phishing scams are not only used to steal financial information; they are used to steal other types of data, such as personal information or login credentials.	Phishing-Betrügereien dienen nicht nur dazu, finanzielle Informationen zu stehlen, sondern auch andere Arten von Daten, z. B. persönliche Informationen oder Anmeldedaten.
34	Knowledge and competence gain	KCG10	Phishing can only occure if I am clicking on a link.	Phishing kann nur stattfinden, wenn ich auf einen Link klicke.
35	Socio-Demographics	SDG1	What is your age?	Wie alt sind Sie?
36	Socio-Demographics	SDG2	What is the highest degree or level of education you have completed?	Welchen höchsten Abschluss haben Sie erreicht?
37	Socio-Demographics	SDG3	How many years have you been with your current company?	Wie viele Jahre arbeiten Sie bereits bei Ihrem derzeitigen Unternehmen?
38	Socio-Demographics	SDG4	How many years of work experience do you have in total (not just at your current company)?	Wie viele Jahre Berufserfahrung haben Sie insgesamt (nicht nur in Ihrem jetzigen Unternehmen)?
39	Socio-Demographics	SDG5	How often do you use a digital device (e.g. Laptop, Desktop PC, Smartphone) for doing your work?	Wie oft benutzen Sie ein digitales Gerät (z. B. Laptop, Desktop-PC, Smartphone) für Ihre Arbeit?
40	Socio-Demographics	SDG6	Which department do you work in?	In welcher Abteilung arbeiten Sie?

Appendix E User Surveys and Questionnaires

Here, we provide additional information on our questionnaires.

E.1 Implementation

We provide in Table 5 (done with the employees) and Table 6 (done with executives/managers of the companies) the complete questionnaires used in the user surveys with our companies. We also show snippets of questionnaire with employees in Fig. 9 and Fig. 10.

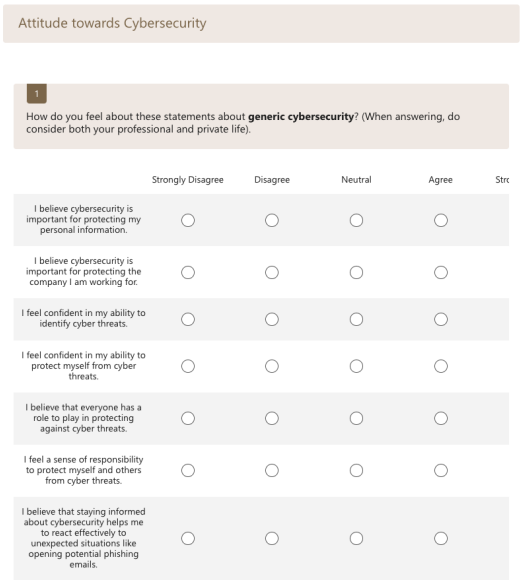


Fig. 9: Snippet of the "Attitude towards Cybersecurity" section of questionnaire. Every question could be answered in a 5-point Likert scale.

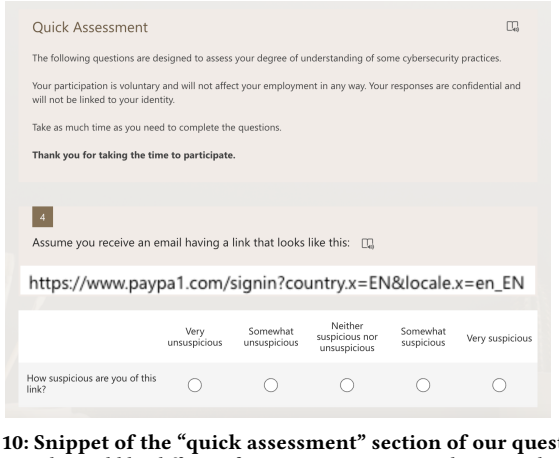


Fig. 10: Snippet of the "quick assessment" section of our questionnaire. Links could be different for every company. We also warn the user that an incorrect answer would not affect their employment.

We stress that, before analysing our data to answer our third RQ, we performed *preprocessing operations*. For instance, we manually checked if the employees could correctly identify phishing links, and hence had the required knowledge to answer the questions in the sections of "competence and knowledge gain" (KCG). Moreover, for items KCG3, KCG4, KCG8, and KCG9, we reversed the five-point Likert scale because these were benign examples, and respondents were expected to not be suspicious of these emails or statements.

We did not include "attention-check" questions, as the questionnaire was intentionally short and participation was voluntary, reducing the likelihood of disengaged responses. However, we manually reviewed all submissions for inconsistencies and found none. On average, participants took approximately 10 minutes to complete the questionnaire, with no anomalous deviations. As a result, no responses were discarded.

Table 6: Questionnaire with the companies' representatives. We used these answers to derive a profile of these companies (§4.1). Altogether, these questions also enable one to derive the remaining four indicators proposed by Chaudhary et al. [30] to investigate the "impact" factor (i.e., *value added*, *reachability*, *touchability*, *overall feedback*—see §4.3).

ID	Question	Type
1	Do you carry out cybersecurity awareness trainings?	Single Choice
2	How often do you train your emloyees?	Short Answer
3	When did you start training your employees?	Short Answer
4	What are the topics which are being taught in the cybersecurity awareness trainings?	Long Answer
5	How often do you update the training?	Short Answer
6	Do you implement recent threats / attack trends into the training?	Short Answer
7	Which methods of delivery are being used for the training?	Multiple Choice
8	Do you differentiate the training's content among different target groups?	Short Answer
9	What kind of feedback do you receive for the cybersecurity awareness training?	Single Choice
10	How many phishing simulations do you carry out each year?	Short Answer
11	Do you implement recent threats / attack trends into the simulations?	Short Answer

E.2 Demographics (PPA questionnaire)

We report in Tables 7–12 the complete demographic details of the participants of our PPA-related questionnaire (§4.3).

Table 7: Demographics: Age.

Age range	$\mathbb{C}_s$	$\mathbb{C}_m$	$\mathbb{C}_h$
18–24 years	0	7	3
25–34 years	4	22	21
35–44 years	4	19	7
45–54 years	4	17	5
55+ years	1	16	0
not provided	0	1	0

E.3 Detailed Results (PPA questionnaire)

We provide in Tables 13–16 the aggregated results of every question asked in our PPA-related questionnaire (refer to Table 5 for the mapping between "ItemCode" and actual question). We also report in Fig. 11 the regression line of our statistical test (see §6.2).

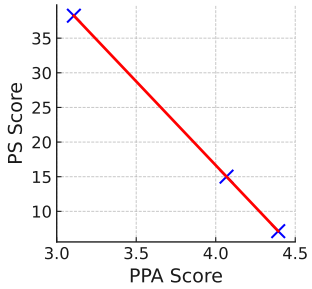


Fig. 11: Regression model of the perceived phishing awareness (PPA) w.r.t. phishing susceptibility (PS) for each company. The blue markers denote our datapoints, whereas the red line is the fitted regression model.

Table 14: PPA assessment: Self-reported Behavior.

	$\mathbb{C}_s$		$\mathbb{C}_m$		$\mathbb{C}_h$	
ItemCode	Mean	Var.	Mean	Var.	Mean	Var.
BHV1	4.462	0.556	4.305	0.456	4.667	0.278
BHV2	3.846	1.207	3.963	0.889	4.528	0.305
BHV3	3.538	0.710	3.732	0.635	4.111	0.710
BHV4	4.154	0.438	4.671	0.294	4.722	0.256
overall	4.000	0.728	4.168	0.569	4.507	0.387

Table 15: PPA assessment: CSA Training Experience and Training Usability. Only one valid question refers to CSA training experience in our questionnaire, i.e., CSA2; the other codes here refer to Training Usability.

	$\mathbb{C}_s$		$\mathbb{C}_m$		$\mathbb{C}_h$	
ItemCode	Mean	Var.	Mean	Var.	Mean	Var.
CSA2	1.667	0.222	4.091	0.550	4.667	0.333
TUB1	3.333	2.889	4.260	0.608	4.583	0.354
TUB2	2.333	0.889	3.779	1.029	4.306	0.768
TUB3	2.667	1.556	3.792	0.892	4.194	0.768
TUB4	3.000	2.667	4.273	0.536	4.306	0.712
TUB5	3.000	2.667	3.816	0.940	4.278	0.756
TUB6	3.000	2.667	3.948	0.777	4.278	0.766
TUB7	2.667	1.556	4.117	0.675	4.444	0.469
TUB8	2.333	0.889	3.870	0.918	4.417	0.521
overall	2.792	1.972	3.982	0.797	4.351	0.638

Table 16: PPA assessment: Knowledge and Competence Gain. The codes KCG1–KCG5 refer to "email and link identification", whereas KCG6–KCG10 refer to "knowledge assessment".

	$\mathbb{C}_s$		$\mathbb{C}_m$		$\mathbb{C}_h$	
ItemCode	Mean	Var.	Mean	Var.	Mean	Var.
KCG1	4.769	0.178	4.768	0.398	4.778	0.340
KCG2	3.692	0.828	3.646	1.351	4.361	0.564
KCG3	2.154	1.207	2.580	1.651	2.944	1.886
KCG4	2.462	2.710	2.695	1.700	3.000	2.111
KCG5	4.769	0.178	4.573	0.757	4.444	1.247
KCG6	1.923	1.302	3.976	1.268	4.417	0.465
KCG7	2.462	1.325	3.988	0.866	4.056	1.386
KCG8	1.462	1.172	4.354	1.326	4.500	1.250
KCG9	1.417	1.243	4.463	1.468	4.417	1.576
KCG10	3.154	2.438	3.793	1.189	4.111	0.932
overall	2.862	1.258	3.884	1.197	4.103	1.176

Table 8: Demographics: Highest level of education.

Education	$\mathbb{C}_s$	$\mathbb{C}_m$	$\mathbb{C}_h$
High School	1	11	1
Bachelor's Degree	5	14	6
Master's Degree	2	31	29
PhD or higher	0	3	0
Other	4	21	0
not provided	1	2	0

Table 9: Demographics: Area of Work.

Area	$\mathbb{C}_s$	$\mathbb{C}_m$	$\mathbb{C}_h$
Administration & Support	6	8	3
Finance, Risk & Audit	0	14	0
Human Resources	0	3	0
IT & Digital Banking	0	18	18
Legal & Compliance	0	8	0
Logistics	0	0	2
Management	1	1	3
Marketing & Communications	0	2	6
Operations	4	19	1
not provided	2	9	3

Table 10: Demographics: Work-related usage of digital devices.

percentage	$\mathbb{C}_s$	$\mathbb{C}_m$	$\mathbb{C}_h$
0–25% of the time	2	0	0
26–50% of the time	1	2	0
51–75% of the time	2	5	2
75+% of the time	8	75	34
not provided	0	0	0

Table 11: Demographics: Years of affiliation to the same company.

Affiliation	$\mathbb{C}_s$	$\mathbb{C}_m$	$\mathbb{C}_h$
0–2 years	2	32	17
3–5 years	4	16	10
6–10 years	4	10	4
10+ years	2	23	5
not provided	1	1	0

Table 12: Demographics: Work Experience in Years.

Experience	$\mathbb{C}_s$	$\mathbb{C}_m$	$\mathbb{C}_h$
0–2 years	0	2	8
3–5 years	0	8	4
6–10 years	0	8	4
10+ years	12	64	11
not provided	1	0	0

Table 13: PPA assessment: Attitude towards Cybersecurity.

	$\mathbb{C}_s$		$\mathbb{C}_m$		$\mathbb{C}_h$	
ItemCode	Mean	Var.	Mean	Var.	Mean	Var.
ACS1	4.692	0.213	4.768	0.178	4.917	0.076
ACS2	4.769	0.178	4.890	0.098	5.000	0.000
ACS3	3.615	1.006	3.716	0.598	4.111	0.710
ACS4	3.154	1.207	3.457	0.816	3.972	0.860
ACS5	4.385	0.544	4.695	0.285	4.694	0.268
ACS6	4.385	0.544	4.333	0.543	4.611	0.293
ACS7	4.000	0.769	4.476	0.493	4.750	0.188
ACS8	4.692	0.367	4.805	0.181	4.861	0.231
ACS9	4.154	0.746	4.366	0.598	4.667	0.333
ACS10	4.385	0.391	4.476	0.371	4.722	0.201
overall	4.223	0.596	4.398	0.416	4.631	0.316