Does Phishing Training Work? A Large-Scale Empirical Assessment of Multi-Modal Training Grounded in the NIST Phish Scale

Andrew T. Rozema
Purdue University

James C. Davis
Purdue University

Abstract

Phishing remains a critical cybersecurity threat, often leading to operational incidents and data breaches. Prior research on the effectiveness of cybersecurity awareness training has yielded mixed results, especially concerning the impact of training on responses to phishing lures of varying difficulty. This paper presents a large-scale measurement study ($N \approx 4\,000$) conducted at a US-based international fintech firm, evaluating the effectiveness of different phishing training modalities. We compared a control group (no training), traditional lecture-based training, and the same traditional training augmented with an interactive phishing exercise.

We observed statistically significant differences in reporting rates following training. Although lure efficacy—approximated using the NIST Phish Scale—significantly impacted click-through rates, our analysis indicates that interactive training resulted in a statistically significant improvement in reporting behavior. Specifically, the interactive training group reported phishing attempts 37% more often than the baseline group and 25% more frequently than those receiving traditional training. However, the effect size remains modest. While interactive training does enhance phishing reporting, its impact is limited. This large-scale study contributes by demonstrating the practical utility of the NIST Phish Scale and the limited benefits of interactive training exercises in bolstering organizational defenses against phishing.

1 Introduction

Email social engineering, or "phishing", is a longstanding cybersecurity hazard [1], and it remains one of the most prevalent cybersecurity threats to modern organizations [2]. According to multiple industry studies, it is the primary attack vector used to breach organizations [3]. The Federal Bureau of Investigation reports that phishing is among the most detrimental frauds targeting enterprises [4]. Such attacks critically impact organizations by damaging their reputations and their financial bottom lines.

Although organizations have technological safeguards, such as spam detectors and attachment analysis [5], attackers continue to evade them. Therefore, organizations complement their anti-phishing technology [6] with cybersecurity training programs [7]. However, evaluating the efficacy of awareness training continues to pose a challenge [8]. New research has begun to look at this variation in phishing lures, but it has not yet been applied to large-scale training programs outside of academic environments [9, 10]. These small-scale studies have leveraged the Phish Scale to study phenomena like "repeat clickers", yet the NIST Phish Scale has not yet been thoroughly evaluated in the context of large, multi-modal training programs.

Standard metrics, such as click-through rates from simulated phishing emails, often fail to capture the variations in the effectiveness of phishing lures, treating bad phishing lures the same as highly deceptive emails that look exactly like the messages users receive daily [10].

In this paper, we report the results of a measurement study that answers some of the questions about how best to train users to resist phishing attacks [11]. We situated our study in a large corporate environment, working with approximately 4,000 employees across multiple business units. Working with a third-party cybersecurity training company, we developed and delivered a traditional lecture-style training program to half of the subjects and an additional enhanced interactive training program to the other half. After training, each subject was put through simulated phishing attempts that ranged in difficulty from simple scams to complex, situationally-adjusted lures, as measured by the NIST Phish Scale [12]. This study design lets us distinguish the efficacy of the two training programs and evaluate how training effectiveness varies with difficulty in phishing attempts. Currently, we have phished approximately 4,000 of those users.

We summarize our main contributions below:

1. **Real-World Scale Application of the NIST Phish Scale.**We present a concrete path to using the Phish Scale enterprise-wide in phishing simulation assessments, incor-

porating a more sophisticated perspective on vulnerability than basic click-through methods offer.

2. **Examining Phish Scale Difficulty and the Impact on Clicking Malicious Links.**By evaluating the varying efficacy of different lures - via click-through rates, open rates, and reporting rates- we can see where training impacts the ability to detect phishing lures with varying difficulties.

3. **Evaluating Varied Training Modes.** We examine varying instructional modalities (e.g., video-based tutorials 1alone vs. video-based and interactive modules) to determine which modalities most effectively lower phishing risks in challenging contexts.

Significance We provide information security managers with actionable advice on improving phishing awareness training. This directly links how well training is executed to how resilient people may be in real-world situations, and it allows for potential reworking/input for the supporting strategies that feed the detection rates and reporting mechanisms, helping to prevent exploitation from happening and ideally reporting the attempt before it happens.

2 Background and Related Work

This section provides an overview of current defenses against phishing, categorizing them into technological defenses (2.1) and human-focused measures (2.1.2) [3]. We then discuss the effectiveness of phishing lure emails used by attackers and trainers alike to trick users into clicking on potentially malicious links [13] and introduce an approximation of the NIST Phish Scale which endeavors to provide us with a more objective measure of the efficacy of these messages [12]. Finally, we review related work, identify gaps, and state our objective and situate our study within the current literature [14].

2.1 Phishing Defenses

2.1.1 Automated Defenses

Organizations deploy increasingly sophisticated technological tools to combat email-based social engineering attacks. Traditional methods, such as tools that do spam filtering, attachment analysis, and look for suspicious links [5], remain relevant and are an important component of a layered defense against these attacks. More recent approaches, which make use of the latest technology, such as machine learning based analysis of email content and metadata [15], are becoming increasingly effective. Recent studies show promising results with AI-based approaches. In a recent study, for example, LSTM models have demonstrated high accuracy in detecting phishing emails [16], and optimized random forest algorithms achieve similar accuracy in identifying malicious websites [17].

Despite (and because of) advances in automated detection, adversaries constantly evolve their methods [3]. Artificial intelligence is a double-edged sword [16] and is currently employed in attacks to produce personalized social engineering content as well as to defend against it [18]. When combined with other deceptive techniques like watering hole attacks [19], typosquatting, and domain spoofing, an unsuspecting user may perceive malicious social engineering attack emails as completely legitimate [18]. The latest phishing attacks leverage context-aware content generation to create highly convincing, targeted messages that evade traditional detection methods [20]. This changing threat landscape highlights the importance of holistic defense strategies [17], combining technological defenses with effective human-focused strategies [21] to provide true defense in depth.

2.1.2 Human-Focused Defenses: Phishing Training

Given the limitations of automated defenses, and the costs of successful phishing exploits, organizations have long incorporateed cybersecurity awareness training [22] to harden the human in the loop. We first discuss the external mandates to offer this training, and then describe the various [23] kinds of training [11].

Legal Requirements and Industry Standards Several legal and regulatory frameworks mandate or strongly encourage phishing awareness training. Key examples include:

- *USA: HIPAA (Health Insurance Portability and Accountability Act):* Requires "covered entities" — typically healthcare providers and other organizations that deal with personally identifiable healthcare information — to implement security awareness training, including phishing and frequently with phishing simulation [24].
- *EU: GDPR (General Data Protection Regulation):* While not explicitly mentioning phishing, GDPR Article 39 requires Data Protection Officers to raise awareness and train staff involved in processing operations. Again, it is widely considered a best practice to include phishing and user awareness training to meet this requirement [25].
- *International: PCI DSS (Payment Card Industry Data Security Standard):* This standard requires that all organizations that use, store, or retain credit card information comply with these requirements in order to maintain access to the credit card networks. Requirement 12.6 mandates security awareness programs for all personnel handling cardholder data, and phishing simulations are a common best practice [26].
- *International: ISO 27001:* This international standard for information security management systems (ISMS) includes controls related to human resource security (Annex A.7), which implicitly require security awareness training, including phishing awareness [27].
- *NIST Cybersecurity Framework:* While not a law, the NIST Cybersecurity Framework is a commonly used collection of

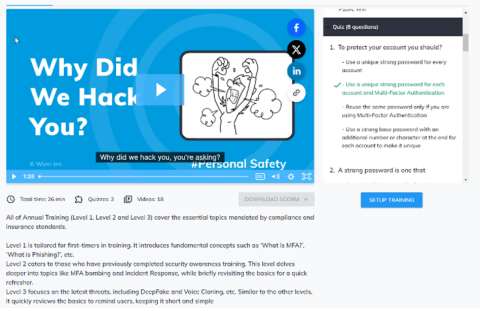

Figure 1: Screenshot of Traditional Mandatory Security Training - Videos and Quiz

rules for managing cybersecurity risk. Although it is not a regulation, it is considered best practice and is frequently the underpinning of regulations. Specifically, security awareness training, including phishing simulations, is advised by the "Identify" function (ID.BE-2) [28].

Training Approaches Phishing training programs typically employ various approaches, broadly categorizable as:

- *Lecture-style training:* This method allows users to learn about phishing threats, common attack vectors, and identifiers through presentations, videos, and written documents. An example is shown in Figure 1. These trainings often conclude with an assessment in the form of a quiz or small test to verify that the user has, in fact, understood the material. The SANS Institute provides training of this nature, particularly focusing on security awareness [29].
- *Feedback from simulations:* One common approach to providing training [30] is to give immediate feedback when users click on phishing links in simulated phishing emails. Research has shown that this approach tends not to be particularly effective as implemented [31]. Vendors such as Infosec IQ Phishing Simulations focus on this aspect of training exclusively [32].
- *Interactive training:* This approach goes beyond passive learning by incorporating some sort of interactive element beyond passively reading or watching material. Oftentimes, a scoring component or other gamification [33] element is included in these sorts of materials [34]. ESET Cybersecurity Awareness Training offers gamified modules to actively engage users in learning [35].
- *Just-in-time training:* A new approach to training is monitoring users as they work on their systems and providing a nudge or alert to those users when they engage in risky behavior. These technologies are relatively new and lack thorough review. Terranova Security's Phishing Simulation Platform provides just-in-time training by delivering immediate feedback to users upon risky actions [36].

Phishing training programs are sometimes developed in-house, but many organizations purchase training from external vendors [14]. Depending on the vendor, these methods may be employed alone or in combination. Phishing awareness programs from commercial training vendors involve a multi-pronged approach. A typical methodology consists of (1) baseline assessments of susceptibility to phishing; (2) tutorial question-and-answer sessions on phishing techniques and detection skills; (3) staged phishing campaigns that assess employee responses in real-world simulations; and (4) follow-up analytics to calculate improvement and optimize training intensity.

2.2 Empirical Studies of Phishing Training

2.2.1 Metrics

Metrics for subject behavior Among the common metrics for assessing the effectiveness of phishing training in terms of the behavior of the subjects, are:

- *Open Rate:* The percentage of recipients who open the simulated phishing email. Opening is usually defined as loading the tracking pixel included with the email message, which generally happens when opened. A lower open rate is generally desirable but can be difficult to interpret without context (*e.g.,* a very relevant, well-crafted email might have a high open rate even among vigilant users).
- *Click-Through Rate (CTR):* The percentage of recipients who click on a link or attachment within the simulated phishing email. Lower CTR is desirable.
- *Reporting Rate:* The percentage of recipients who report the simulated phishing email to the appropriate security personnel or system. Higher reporting rates are desirable, indicating proactive security behavior.

These metrics, however, may not capture the full complexity of user behavior. Some phishing lures are obvious (*e.g.,* "lottery winner" scams), while more sophisticated attacks mimic legitimate communications. After much debate on how to measure the degree of complexity of *phishing lures*, the US National Institute of Standards and Technology (NIST) recently introduced the NIST Phish Scale for this purpose.

Measuring phishing lure complexity The NIST Phish Scale is a standardized framework for assessing the difficulty of phishing lures [10, 37]. The NIST Phish Scale evaluates phishing lures on two dimensions. First, *Phishing Cues* are observable features such as typos, suspicious hyperlinks, or inconsistencies in sender information. Generally, the fewer cues an email exhibits, the easier it is to classify as phishing. Second, *Premise Alignment* describes the degree to which a phishing email resembles the regular correspondence received by the recipient. Higher alignment indicates a more deceptive message [38].

Phishing lures with fewer cues [39] and greater premise alignment are considered to be likelier to deceive a user. Thus, the Phish Scale provides a means of characterizing the effectiveness of phishing training on weaker and harder phishing

lures. This framework can help organizations contextualize phishing simulation results and use that data to enhance security training programs. Where results are surprising — for example, a high click rate in a seemingly obvious phishing attempt — additional training may be necessary. We note, however, that applying this scale is probably meaningful only with a sufficiently large [40] sample. Different people vary in their ability to detect phishing emails, and an individual's abilities may also vary based on contextual factors such as their degree of awareness (*e.g.,* recency of training) and their cognitive load and time sensitivity (*e.g.,* if distracted or in a hurry) [41].

2.2.2 Experiments and Reports

Despite evidence supporting the value of phishing training, many studies fail to account for the efficacy of phishing attacks when evaluating the effectiveness of training modalities, leading to potentially misleading conclusions about real-world preparedness. Research has shown that well-structured training can lower phishing click rates [33, 34], especially when training includes interactive and gamified components. However, a closer examination reveals a more nuanced view of these programs' effectiveness. Quiz-based phishing awareness may yield different detection results compared to interactive, hands-on phishing simulations: while repeated exposure can lead to increased detection rates in some contexts, methodologies designed for hands-on scenarios may also have unwanted effects [42]. Further complicating the matter, Lain et al. [43] argue that frequent phishing tests can lead employees to feel overconfident, thereby increasing the likelihood of falling victim to advanced attacks [10]. Moreover, commercial training content is often focused on meeting compliance metrics, such as click-through rates, which do not necessarily correlate with employees' ability to identify more sophisticated social engineering exploits [44].

Beyond the immediate impact on measurements such as click rates, the implementation of phishing training programs raises important ethical and organizational considerations. Some enterprises have even "rewarded" employees for falling for a phishing simulation by temporarily suspending their accounts or requiring remedial training [45]. Although these measures can promote accountability, they may also foster a blame-oriented atmosphere, where companies react to threats instead of preemptively implementing cybersecurity best practices. *In contrast to this reactive approach, a more effective strategy incorporates security awareness into day-to-day workflows, reinforcing training through continuous education rather than through periodic evaluations alone [33].*

While commercial cybersecurity awareness training aims to address the evolving threat landscape, critical questions remain about the long-term effectiveness of these programs and their ability to foster genuine behavioral change. Many current solutions focus on easily quantifiable metrics like click-through rates, which may not accurately reflect an employee's ability to recognize sophisticated attacks, especially sincetraining tools may not account forthe efficacy of those attacks. Some providers are exploring the use of artificial intelligence (AI) to personalize training and improve behavioral analytics [16], but rigorous, independent evaluations of these approaches are still needed.

To combat the persistent threat of phishing, many organizations have turned to commercial cybersecurity awareness training platforms, such as KnowBe4, Proofpoint, and Cofense, which often incorporate hands-on phishing simulations as a core component [44]. These simulations offer valuable experiential learning opportunities [46]. However, while these platforms provide organized learning materials and behavioral assessments, many initiatives lack rigorous follow-up training or, crucially, a systematic measure of lure difficulty. Notably, only a few practical implementations have integrated standardized frameworks—such as the Phish Scale—to benchmark the complexity of phishing emails, hindering the ability to accurately assess and improve training effectiveness.

Academic research has extensively investigated various cybersecurity awareness training modalities [47], including video-based instruction, email-based tips [48], and interactive gamified exercises. While studies generally report improved knowledge retention and engagement following simulation-based training, these often rely on generic or overly simplistic phishing scenarios, failing to capture the nuanced sophistication of real-world attacks [49]. Furthermore, large-scale investigations that explicitly account for phishing lure complexity using standardized measures remain scarce [50]. Passive video-based instruction, a common modality, often lacks the interactive or adaptive elements necessary to foster sustained behavioral change [51].

2.3 Research Gap and Objectives

In light of the evolving sophistication of phishing attacks and the limitations inherent in both technological and human-focused defenses [4], a substantial research gap persists in systematically assessing phishing awareness training relative to lure complexity. This study aims to address this gap by:

- **Developing a Comprehensive Framework:** Integrating the NIST Phish Scale into the assessment of phishing training outcomes to provide a standardized measure of lure difficulty.
- **Analyzing Phishing Complexity:** Examining the impact of lure sophistication, including the challenges posed by AI-generated phishing emails, on user vulnerability.
- **Optimizing Training Interventions:** Tailoring training programs to address specific weaknesses identified through standardized assessments and user performance data.
- **Enhancing Organizational Resilience:** Improving employee preparedness and response to sophisticated phishing attacks, thereby reducing the organization's overall risk.

By using a more nuanced and standardized evaluation of phishing simulations, this study aims to give organizations a better idea of how effective training is, which will help them defend themselves against phishing attacks that grow more capable each year.

3 Methodology

3.1 Author Positionality and Ethical Concerns

Positionality This study reports the results of a phishing training performed for BLINDED, which we call *ACME Corp.*, in conjunction with a third-party training vendor, BLINDED, which we call *PhishVendor*. The lead author is employed by ACME Corp. as a security professional leading the annual phishing training, and is also studying phishing as part of his doctoral work. The author team embraced the opportunity to evaluate and report on a real-world, large-scale phishing training, which can help address some of the oepn questions in the academic literature. We recognize this needs to be balanced by awareness of potential sources of bias.

The author team's interest in performing this study is to ascertain whether the results of PhishVendor's phishing training merit the costs that ACME Corp. incurs for security awareness training, both in contracts with PhishVendor and in employee time. Neither ACME Corp. nor PhishVendor had a role in the study's design, conduct, nor data analysis. None of the authors has a financial stake in PhishVendor.

Ethical considerations This study involves human subjects data. The author team has completed ethics training from ACME Corp. and the partner academic institution. ACME Corp. has policies alerting users that they will be assigned cyber security training and subjected to regular phishing testing. Users give consent to the process during onboarding and in subsequent policy acknowledgment. ACME Corp. had no concerns about analyzing the results to assess training efficacy, and approved our presentation of this manuscript.

Finally, we acknowledge that poor performance on the phishing training could lead to remedial training or further consequences for the employees in the study. Note, however, that the phishing training would happen whether or not we reported on the results in this paper. Thus, this aspect is out of scope for ethics consideration.

3.2 Hypotheses

Based on prior work on phishing training efficacy and documented differences in risk perception, we formulated the following hypotheses:

1. **H1 (Phish Scale Effect):** Phishing emails with higher difficulty, as measured by our approximation of the NIST Phish Scale, will exhibit higher click-through rates than those with lower difficulty.

2. **H2 (Training Impact - General):** Training will increase the detection and reporting of phishing emails.

3. **H3 (Interactive Training Effect - Reporting):** The inclusion of interactive components (*i.e.,* the Trained+Exercise modality) will lead to significantly higher reporting rates compared to traditional lecture-based training (*i.e.,* Trained Only).

4. **H4 (Interaction Effect):** There will be a significant interaction effect between training modality and phishing lure difficulty, such that the benefit of interactive training on reporting rates is more pronounced for less complex phishing lures.

We justify these hypotheses as follows. **H1** is the assertion of NIST. **H2** and **H3** are the intended effects of training. **H4** is based on our intuition — we believe that at this point, advanced phishing emails are beyond the capability of individuals to detect. We discuss this point further in §6.

3.3 Materials and Instruments

In light of these hypotheses, we examined the resources available to us from PhishVendor. We selected the materials that offered appropriate experimental controls.

Training Materials PhishVendor had 2 relevant offerings.

- *Lecture-style Training:* PhishVendor offers a series of fifteen 1-2 minute instructional videos covering cybersecurity awareness topics, including phishing tactics and appropriate behavior for reporting phishing emails (clicking the "Report Phish") button in the email client. After training, subjects complete a quiz that reviews the content within the videos.
- *Phishing Simulation:* PhishVendor offers a learning module in which users identify simulated phishing attempts in real time, receiving immediate feedback.

Phishing Lures along the Phish Scale For the phishing simulation module, we wanted lures that cover both dimensions of the Phish Scale — cues and premise alignment. We started by selecting emails from PhishVendor's training library. PhishVendor provided an assessment of difficulty (easy, medium, or hard). We independently assessed these and agreed that their difficulty levels aligned with the NIST Phish Scale's "Phishing Cues" dimension. Then, for the premise alignment dimension, the lead author classified the templates based on their contextual familiarity to users [11]. We distinguished three levels of familiarity: common (regularly encountered work-related platforms at ACME Corp.),

Table 1: Phishing Template Difficulty Estimate - Assessment of phishing email templates based on vendor difficulty scores and system familiarity, resulting in a combined difficulty score. Here we combine the vendors assessment of difficulty and add the likelihood that a user will use the particular system to come up with a total score.

Lure Theme	Vendor Score	System Use	Total Score
Attachment - Word	Hard (3)	High (3)	6
Vulnerabilities ǀ Office	Medium (2)	High (3)	5
Attachment - XLS	Easy (1)	High (3)	4
Loom - Recording Shared	Hard (3)	Low (1)	4
Device non-compliant - Microsoft	Medium (2)	Medium (2)	4
Fax via Voicemail Office	Easy (1)	Medium (2)	3
Email not delivered ǀ Office	Easy (1)	Medium (2)	3
Email not delivered ǀ Google	Easy (1)	Low (1)	2
Marketing: Collab request	Easy (1)	Low (1)	2

occasional (tools that ACME Corp. employees might recognize but that are not company supported), and unfamiliar (services not likely to be relevant to this population).

After this, we had a 3x3 rating system for PhishVendor's email templates. We assigned each template a score by summing the cue and premise alignment dimensions, assigning a 1 for easier lures and a 3 for harder lures on each dimension. Thus the minimum difficulty score was a 2 (1+1) and the maximum a 6 (3+3).

Table 1 summarizes the phishing lures we used.

3.4 Pilot Study

We conducted two pilots to assess our technical infrastructure and obtain feedback from ACME Corp. management.

The first pilot involved 100 randomly selected employees. We gave them no training and sent three phishing lures (easy, medium, and hard). We confirmed that our email delivery and data collection systems were working.

The second pilot used 400 randomly selected employees. These were assigned training before undergoing a second, more comprehensive phishing campaign with nine phishing templates. This pilot phase allowed us to evaluate PhishVendor's training modules and give them feedback for refinement, as well as guided our selection of contextually appropriate phishing lures.[1] These preliminary data generally supported our hypotheses. We continued to identify and resolve minor technical glitches.

3.5 Treatments for Main Study

For the main study, we used a between-subjects design with a control and two treatments.

- *Control.* This group received no training from PhishVendor. We sent each of these subjects 1 lure from our set of 9 phishing lures.
- *Treatment 1: Training.* This group received PhishVendor's lecture-based training. We then deployed the lures in the same manner as the Control group.
- *Treatment 2: Training+.* This group received PhishVendor's lecture-based training plus PhishVendor's phishing simulation. We then deployed the lures in the same manner as the Control group.

The sizes of the groups are summarized in Table 2.

Table 2: Demographic breakdown of the study participants by treatment group.

	Total	Trained	Trained+	Control
Total	4755.0	2406.0	1866.0	483.0
Male	1841.0	1008.0	650.5	182.5
Female	2914.0	1398.0	1215.5	300.5

Metrics Training effectiveness was assessed by measuring the phishing email detection rate, click-through rate, and reporting rate across groups [52].

Longitudinal consideration All trainings were completed in the same two-week window. Phishing lures were then sent to subjects within a subsequent three-month window. This study is ongoing; our full population is ∼15,000 and report only the results from subjects who have already received the phishing lures.

Confounds This study was performed as part of ACME Corp.'s annual, legally-mandated security training. Many employees, though not all, had participated in previous annual security trainings. All users were also offered non-mandatory security awareness material during ACME Corp.'s annual Cyber Security Awareness Month, which occurred shortly before our study began.

[1] As one effect of our real-world context, ACME Corp. management instructed us to omit the most effective lures (human resources and child-in-danger) to minimize employee distress.

3.6 Data Analysis

To preserve participant privacy while enabling demographic analysis, user data was anonymized using a two-step process:

3.6.1 Privacy Protection Measures

To preserve participant privacy while enabling demographic analysis, we implemented the following measures:

1. **Email Anonymization:** Email addresses were hashed using SHA-256 with cryptographically secure salt to ensure irreversible anonymization while maintaining unique user tracking.

2. **Gender Inference:** First names were analyzed using the gender-guesser library, mapping to predefined neutral stand-in names, enabling demographic analysis without exposing personal identities [50].

This privacy-preserving approach adhered to ethical research standards while allowing us to explore variations in phishing susceptibility across different user demographics [53].

3.6.2 Statistical Analysis Framework

We employed the following statistical methods to evaluate training efficacy and test our hypotheses:

- **Two-Way ANOVAs:** To test H1, H2, H3, and H4, we conducted a series of two-way ANOVAs. Separate ANOVAs were performed for three dependent variables: email open rate (OpenedRate), click-through rate (ClickedRate), and reporting rate (ReportedRate). The independent variables were Intervention (Control, Trained Only, Trained+Exercise) and Difficulty (Low, Medium, High), derived from our approximation of the NIST Phish Scale. We examined both main effects and the interaction effect between Intervention and Difficulty.
- **Logistic Regression:** To further investigate the factors influencing user behavior, we developed logistic regression models for each of the three outcome variables (OpenedRate, ClickedRate, ReportedRate). The predictor variables included Intervention, Difficulty, and their interaction.

All statistical analyses were conducted using Python's statistical libraries (specifically mentioning the libraries used, e.g., 'statsmodels', 'scipy.stats') with a significance threshold of $\alpha = 0.05$. Effect sizes (e.g., η^2 for ANOVA, odds ratios for logistic regression) were calculated to assess the practical significance of the findings. For inferential statistics (ANOVAs and logistic regressions), we only used data from after the training intervention, allowing for comparison between treatments. The baseline data, while useful for characterizing pre-training behavior, was not included in inferential comparisons between groups.

4 Main study

4.1 Control Group - Blind Phishing Baseline

Before implementing any training interventions or employing NIST Phish Scale metrics, a preliminary "single-blind" phishing campaign was conducted. In this initial phase, participants received phishing emails without advance notice or any prior training, allowing for the collection of baseline data on user susceptibility and reporting behaviors. This baseline serves as a critical benchmark against which subsequent training effectiveness and the influence of varying phishing difficulty levels can be measured.

5 Results

5.1 Overview of Results

Overall, our findings revealed significant effects of both phishing lure difficulty and training modality on user behavior. As detailed below, the results provide support for the predictive validity of the NIST Phish Scale and demonstrate the positive impact of interactive training on phishing reporting rates. We found no significant interaction effect between training type and lure difficulty.

In our baseline scenario, only 8.3% of phishing emails were reported. When employees underwent traditional training, the reporting rate increased modestly to 9.1%.

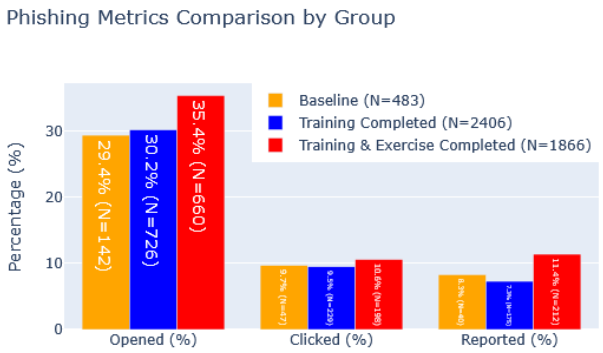

Figure 2: Overall Metrics Post Training: Comparison of open rate, click rate, and reporting rate across the three experimental groups after the training intervention. *N* users in the corner. *N* email interactions on the bars.

However, when we augmented the training with an interactive exercise, the reporting rate jumped to 11.4%—a remarkable 37% increase over the baseline reporting rate and a 25% improvement over traditional training alone. This significant boost in reporting not only demonstrates that a multimodal approach can enhance user vigilance, but it also suggests that interactive components are especially effective at engaging users and encouraging proactive security behaviors. The data strongly indicate that training interventions that include

hands-on, interactive elements can substantially improve the overall detection and reporting of phishing attempts, thereby strengthening an organization's first line of defense.

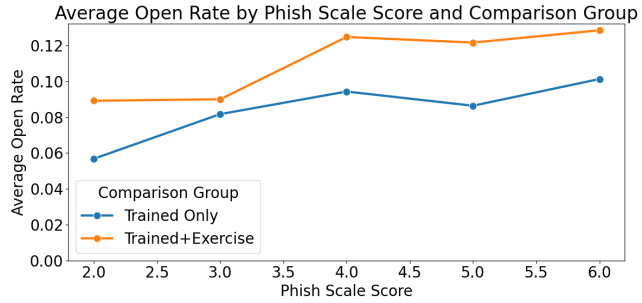

Figure 3: Open Rate: Average open rate for phishing emails, categorized by Phish Scale score and experimental group.

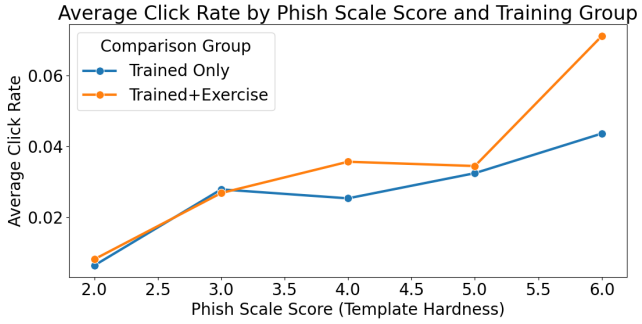

Figure 4: Click Rate: Average click rate for phishing emails, categorized by Phish Scale score and experimental group.

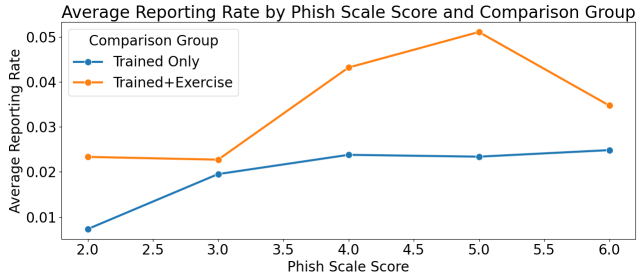

Figure 5: Reporting Rate: Average reporting rate for phishing emails, categorized by Phish Scale score and experimental group

5.2 Hypothesis 1: Phish Scale and Click-Through Rates

H1 predicted that phishing emails with higher difficulty, as measured by our approximation of the NIST Phish Scale, would exhibit higher click-through rates than those with lower difficulty. These data are visualized in Figure 4 for the two treatment groups. Visual inspection suggests a general increase in click rate as phishing lure hardness goes up, supporting the hypothesis. See Table 3.

Overall Correlations: Correlation between Click Rate and Phish Scale Score: 0.65 Correlation between Click Rate and Vendor Hardness: 0.46 Correlation between Click Rate and In Use Score: 0.54

To test this formally, we conducted a one-way ANOVA with `Clicked_Flag` as the dependent variable and `phish_scale_category` (Low, Medium, High) as the independent variable. The results revealed a significant main effect of Difficulty ($F(2, 12086) = 15.65$, $p < 0.001$, $\eta^2 = 0.003$).

Post-hoc analyses (Tukey's HSD) revealed that click-through rates were significantly higher for High difficulty lures compared to both Medium ($p = 0.010$) and Low difficulty lures ($p < 0.001$). Medium difficulty lures also had significantly higher click rates than Low difficulty lures ($p = 0.021$).

Therefore, **H1 was supported**. The click-through rate increased with increasing phishing lure difficulty, as predicted by the NIST Phish Scale.

5.3 Hypothesis 2: Overall Training Impact

H2 posited that training would have a significant impact on the detection and reporting of phishing emails. The pertinent data are visualized in Figure 2. Note the increase in opening, clicking, and reporting in both treatment groups as compared to the baseline. The first two activities are undesirable, but the final activity (reporting) is the ultimate indicator of whether or not the employee determined there was a phishing attempt in progress.

Therefore, **H2 was partially supported**. Training significantly impacted all of the behaviors, although not necessarily as one would expect.

5.4 Hypothesis 3: Interactive Training and Reporting Rates

H3 stated that the inclusion of interactive components (Trained+Exercise) would lead to significantly higher reporting rates compared to traditional lecture-based training (Trained Only).

To test this, a two-way ANOVA was conducted with `Reported_Flag` as the dependent variable, and `Comparison_Group_Clean` and `phish_scale_category` as independent variables. The ANOVA revealed a significant main effect of `Comparison_Group_Clean` ($F(1, 12088) = 18.98$, $p < 0.001$, $\eta^2 = 0.002$).

Post-hoc comparisons (Tukey's HSD) showed that the `Trained_Exercise` group had a significantly higher reporting rate than the `Trained_Only` group ($p < 0.001$, mean difference = 0.0136).

Therefore, **H3 was supported**. The inclusion of the interactive exercise significantly increased reporting rates compared to traditional training alone. However, note that the effect

Table 3: Template Statistics with Score Comparisons

Template	Total Sent	Click Rate (%)	Phish Scale Score
Attachment - Word	385	22.0779	6
Vulnerabilities | office	379	11.8734	5
Attachment - xls	391	9.46292	4
Device non-compliant - Microsoft	347	11.5274	4
Loom - Recording Shared	393	9.66921	4
Email not delivered | office	366	1.36612	3
Fax via Voicemail Office	355	22.5352	3
Email not delivered | Google	394	2.03046	2
From Marketing: Collab request	356	1.68539	2

size (as can be seen visually in Figure 2) is relatively small, affecting the mean by about 1%.

Baseline (N=483)

Metric	Value
Total Sent	483
Opened (%)	29.4% (N=142)
Clicked (%)	9.7% (N=47)
Reported (%)	8.3% (N=40)

Training Completed & Phished (N=2406)

Metric	Value
Total Sent	2406
Opened (%)	30.2% (N=726)
Clicked (%)	9.5% (N=229)
Reported (%)	7.3% (N=175)

Training & Exercise Completed & Phished (N=1866)

Metric	Value
Total Sent	1866
Opened (%)	35.4% (N=660)
Clicked (%)	10.6% (N=198)
Reported (%)	11.4% (N=212)

5.5 Hypothesis 4: Interaction Effect of Training and Difficulty

H4 predicted an interaction effect between training modality and phishing lure difficulty, such that the benefit of interactive training on reporting rates would be more pronounced for easier phishing lures.

The two-way ANOVA (described in H3) also tested for the interaction between training modality and lure difficulty. The interaction effect was not significant ($F(2, 12088) = 2.01$, $p = 0.134$).

Therefore, **H4 was not supported**. The effect of training on reporting rates did not significantly differ across the difficulty levels of the phishing lures.

6 Discussion

6.1 Insights for Tailoring Training

The initial baseline suggests several opportunities for refining the training approach:

- **Emphasizing Reporting Mechanisms:** Given the low reporting rates, training content should underscore the significance of reporting suspicious emails. Clear instructions and examples may motivate users to take a more active role in defending the organization.
- **Introducing Foundational Awareness Early:** Even brief, introductory training modules can equip users with the conceptual tools necessary to identify basic phishing indicators, potentially reducing subsequent CTR.
- **Establishing a Feedback Loop:** Highlighting the organizational impact of user reports—such as sharing scenarios where reported emails led to enhanced defenses—may reinforce positive reporting behaviors and cultivate a sense of collective security responsibility.

6.2 Training Used in Our Study

Recent studies have demonstrated that training does not affect the assessment of user interaction with simulated phishing emails. That intervention, similar to ours, has had no effect. Our training at least appears to have affected the reporting rate and some minor impact on users clicking on the links inside of our simulated phishing emails, which is not a desired outcome of the training. It would be appropriate to discuss the composition of our training. Our training consisted of 15 brief (1-3 minute) videos, five of which were produced by the in-house communications team and were specific to our organization; the other 10 were selected from the library provided by our

vendor, followed by a brief 10-question assessment. The nature of our organization is quite acquisitive, and it is worth noting that many of the colleagues receiving training may have been exposed to little phishing/cybersecurity awareness training before this. Training was communicated to the user population first through mass email from executive leadership and blog posting on the company's intranet site and followed a October Cybersecurity Awareness Month campaign that involved weekly email communications and videos posted similarly. We should also mention that although our training provider does provide user education as part of the phishing simulation, the number of users who follow through on the instructions in those email trainings is minuscule as they ask users to report them. The number of users who open, click, and report is minuscule.

6.3 KnowBe4 Effect

The phish scale could potentially enhance the success of cybersecurity awareness training programs by creating a more equitable environment [12]. Education can be challenging, and cybersecurity awareness professionals may succumb to the temptation of focusing solely on the test, particularly if that is how their performance is assessed. There is the potential to set up a perverse incentive system where awareness managers may unintentionally create difficult phishing tests to begin with and gradually ease off on the difficulty, creating the appearance of improvement in user awareness when what was actually happening was that the simulated phishing attacks were getting easier [8]. The phish scale or other more concrete metrics may be able to prevent even the accidental implementation of such a system.

The real metric organizations are actually being graded against is their users' interaction with real phishing attacks sent by adversaries [54]. Future research could look into how people report real phishing attempts changing over time [3]. However, this will need to be tightly integrated with helpdesk systems, email security tools, and an understanding of how many phishing attacks are happening against the organization at any given time. These challenges are complex, yet they are not insurmountable [55].

6.4 Phishing detection a cue-less world

This study is based, in part, on the idea that there are observable "cues" [56]—such as bad grammar, malicious links, and URLs or sender addresses that don't match—that can be used by the human recipient of that email to determine if they are under attack. As with much phishing detection and prevention research (and advice), we look to these cues to indicate that there is a problem. However, the rapid advancement of large language models (LLMs) [57] and generative AI tools fundamentally undermines this approach. As shown by the increasing sophistication of AI-generated phishing content (deepfakes and personalized, contextually relevant messages), attackers can now create attacks that look exactly like real communications, base them on open-source intelligence about the target and do that at potentially very large scales. The ability for AI to replicate the way people write, to create realistic audio and video, and to even generate full phishing websites en masse makes cue-based analysis more and more useless. This poses a significant challenge for future work and systems that rely on this type of metric.

Future work should look at what organizations should do in the face of these new threats [3]. Reports of real-time generative AI video impersonation attacks already exist. Proving the provenance of media is a problem that is being looked at. Adobe and others are developing technologies to allow users to identify the provenance of media based on cryptographic signing and metadata. The application of similar technologies to real-time communication could be next. However, we must consider the human factors, and particularly UI/UX components, to ensure the ease of use, or suffer the fate of similar technology designed for emails (PGP, GPG, etc.) which has long been available but seldom adopted.

AI-powered phishing automation will force us to reassess not only the threat but also how effective our training is [57]. As "cues" are no longer reliable guides to malicious content [56], security awareness training programs emphasizing traditional markers symptomatic of malicious intent will be less successful [58]. Instead, there needs to be a stronger focus on contextual awareness, critical thinking, and fact-checking processes [59]. These skills are notoriously challenging to teach [60]. Our findings also point the way to new research that looks into other ways to rate the difficulty of phishing lures, not just based on obvious signs, but also on deeper factors such as the plausible nature of requests, the ability to manipulate emotions, or the sophistication of the social engineering techniques used [54]. Being able to produce perfect phishing emails may make some of the current best practice advice obsolete [61].

6.5 The expanding attack surface

This analysis, as much of the current literature on message-based social engineering attacks, focuses on email-based phishing. We would be remiss not to acknowledge that the human attack surface is expanding and encompasses all the ways humans communicate [5]. Phishing is no longer confined to electronic mail messages but also includes other communication channels; wherever people are to be found, other people will try to take advantage of them. Phishing now comes through SMS (smishing) [62], voice calls (vishing), social media platforms [63], messaging apps, and even QR codes [64]. Security awareness training can no longer focus solely on email messages and business email compromise [65]. Situational awareness can no longer be boiled down to whether or not a communication is rife with typos. Organizations must

expand user or security awareness training to cover the diverse attack surface that the human factor now represents [3]. While improvements in authentication technologies such as multifactor authentication and password replacements like passkeys and phishing-resistant authentication mechanisms such as Microsoft's Windows Hello are making attacks more challenging [66], the underlying "infinity day" vulnerability that is the human behind the screen remains [41].

6.6 The Death of Digital Trust

With the rise of sophisticated attacks using AI-powered deepfakes, we face a fundamental challenge that extends beyond the simple threat of phishing [57]. Potentially, this leads to the erosion of trust in every form of electronically mediated communication [67]. As generative AI models have become increasingly proficient at duplicating human language, at creating convincing audio and video deepfakes, and allowing for the personalization of attacks on a grand scale, the very foundations of digital interaction are becoming threatened [61]. We're fast approaching the point when, as the old saying goes, we'll no longer be able to "believe our lying eyes" – or ears, for that matter.

We've typically been able to rely on cues (in fact, the Phish Scale depends on it) to allow us to assess the authenticity of our digital interactions, at least on some level [56]. Where do we turn when every email perfectly represents and replicates the writing style of your colleagues? When a voice on the phone is a perfect simulation of your chief executive officer? When your video conference is full of participants that appear to be individuals you know well but are, in fact, AI-generated constructs [68]? Implications that formerly lived in the realm of science fiction now extend into the reality of modern cybersecurity awareness training. This erosion of trust has the potential to disrupt business operations, damage interpersonal relationships, and undermine the very fabric of online society [3]. If we can no longer trust the apparent source or content of digital messages, how can we conduct business, collaborate on projects, or even maintain casual conversations with any degree of confidence?

An erosion of trust this great will disrupt business operations, damage interpersonal relationships, and undermine the fabric of our online society [54]. If we can no longer trust the apparent source or content of digital messages, even more than presently, how can we conduct business, or collaborate on projects, or even maintain casual conversations with any degree of confidence with whom we are interacting? However, the technical challenges are significant, and the human factors are even more daunting [58]. Any such system must be seamlessly integrated into existing communication platforms and be virtually transparent to the end-user. Complexity, as evidenced by the limited adoption of email encryption technologies like PGP and GPG, is the enemy of usability and, ultimately, of security [66].

A pair of extreme potential solutions emerge. A knee-jerk reaction might be to retreat back to in-person interactions for any sort of sensitive communication, to only conduct private conversations face-to-face, or to transact business in the flesh [14]. Unfortunately, in this globalized, digitally-driven world, this solution seems entirely impractical. The second, while more plausible, lies in developing robust, technologically-enforced systems for verifying identity and the integrity of communication and participants [69].

Solutions that emerge here will have to leverage cryptographic principles, similar to those used to secure communications between web browsers and websites, to establish trust in new online transactions [70]. Making use of digital signatures, verifiable credentials, and potentially even distributed ledger technologies could play a role in creating systems where the provenance and authenticity of communications and the participants could be independently verified [6]. This solution imagines a future where every email, every voice call, and every video conference participant is using some form of cryptographically-verified digital passport, which attests to both their identity and to the integrity of the communication.

The development and widespread adoption of such trust-enhancing technologies will require a concerted effort from researchers, industry stakeholders, and policymakers [71]. The stakes are high: the future of digital communication, and perhaps even the future of trust itself, hangs in the balance. Future research must prioritize not only the technical aspects of these solutions but also their usability and societal implications [72]. This is no longer simply a cybersecurity problem; it is a fundamental challenge to how we interact and build relationships in the digital age.

7 Conclusions

This work contributes to a growing body of research [12, 43, 73] questioning [7, 11] the value of traditional user training [74]. Our findings reinforce concerns that conventional awareness approaches may be insufficient against increasingly sophisticated phishing techniques [44]. By quantifying phishing difficulty through standardized metrics like the NIST Phish Scale, organizations can better evaluate training effectiveness and target specific vulnerabilities [41].

The integration of interactive training elements showed measurable improvements in reporting behaviors, suggesting that engagement-focused approaches may yield better outcomes than passive instruction methods [59]. As phishing techniques evolve with AI capabilities [61], our work provides insights for human risk professionals tasked with fortifying the "human firewall" against increasingly sophisticated attacks [3].

This research provides compelling evidence for adopting difficulty-based metrics in large-scale cybersecurity education initiatives and continued research into what we need to keep

humans secure in an era where traditional cues of deception are rapidly disappearing [56, 57].

References

[1] Ponnurangam Kumaraguru, Young Rhee, Alessandro Acquisti, Lorrie Faith Cranor, Jason Hong, and Elissa Nunge. Protecting people from phishing: The design and evaluation of an embedded training email system. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 905–914, 2007.

[2] Kostiainen K. \u Capkun S. Lain, D. Phishing in organizations: Findings from a large-scale and long-term study. In *2022 IEEE Symposium on Security and Privacy (SP)*, pages 842–859, 2022.

[3] Zainab Alkhalil, Chaminda Hewage, Liqaa Nawaf, and Imtiaz Khan. Phishing Attacks: A Recent Comprehensive Study and a New Anatomy. *Frontiers in Computer Science*, 3:563060, March 2021. https://www.frontiersin.org/articles/10.3389/fcomp.2021.563060/full.

[4] Frank L. Greitzer, Wanru Li, Kathryn B. Laskey, James Lee, and Justin Purl. Experimental Investigation of Technical and Human Factors Related to Phishing Susceptibility. *ACM Transactions on Social Computing*, 4(2):1–48, June 2021. https://dl.acm.org/doi/10.1145/3461672.

[5] Ahmed Aleroud and Lu Zhou. Phishing environments, techniques, and countermeasures: A survey. *Computers & Security*, 68:160–196, 2017.

[6] Uta Menges and Annette Kluge. Contrasting and Synergizing CISOs' and Employees' Attitudes, Needs, and Resources for Security Using Personas. In *2024 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW)*, pages 456–472, Vienna, Austria, July 2024. IEEE. https://ieeexplore.ieee.org/document/10628768/.

[7] Grant Ho, Ariana Mirian, Elisa Luo, Khang Tong, Euyhyun Lee, Lin Liu, Christopher A Longhurst, Christian Dameff, Stefan Savage, and Geoffrey M Voelker. Understanding the Efficacy of Phishing Training in Practice. *IEEE Symposium on Security and Privacy*, 2025.

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

[9] C. I. Canfield, B. Fischhoff, and A. Davis. Better watch out: Comparing phishing metacognition with real emails. *Learning and Metacognition*, 14(3):343–362, 2019. https://doi.org/10.1007/s11409-019-09197-5.

[10] Deanna D. Caputo, Shari Lawrence Pfleeger, Jesse D. Freeman, and M. Eric Johnson. Going Spear Phishing: Exploring Embedded Training and Awareness. *IEEE Security & Privacy*, 12(1):28–38, January 2014. http://ieeexplore.ieee.org/document/6585241/.

[11] Hossein Abroshan, Jan Devos, Geert Poels, and Eric Laermans. Phishing Happens Beyond Technology: The Effects of Human Behaviors and Demographics on Each Step of a Phishing Process. *IEEE Access*, 9:44928–44949, 2021. https://ieeexplore.ieee.org/document/9380285/.

[12] Michelle Steves, Kristen Greene, and Mary Theofanos. Categorizing human phishing difficulty: A Phish Scale. *Journal of Cybersecurity*, 6(1):tyaa009, January 2020. https://academic.oup.com/cybersecurity/article/doi/10.1093/cybsec/tyaa009/5905453.

[13] Ankit Kumar Jain and B.B. Gupta. A survey of phishing attack techniques, defence mechanisms and open research challenges. *Enterprise Information Systems*, 16(4):527–565, April 2022. https://www.tandfonline.com/doi/full/10.1080/17517575.2021.1896786.

[14] Orvila Sarker, Asangi Jayatilaka, Sherif Haggag, Chelsea Liu, and M. Ali Babar. A Multi-vocal Literature Review on challenges and critical success factors of phishing education, training and awareness. *Journal of Systems and Software*, 208:111899, February 2024. https://linkinghub.elsevier.com/retrieve/pii/S0164121223002947.

[15] M. Almseidin, A. M. Abu Zuraiq, M. Al-kasassbeh, and N. Alnidami. Phishing detection based on machine learning and feature selection methods. *International Journal of Interactive Mobile Technologies*, 13(12):71–183, 2019. https://doi.org/10.3991/ijim.v13i12.11411.

[16] Md Meraj Ansari, Amrutanshu Panigrahi, Geethamanikanta Jakka, Abhilash Pati, and Krutikanta Bhattacharya. Prevention of Phishing attacks using AI Algorithm. In *2022 International Conference on Omni-Layer Intelligent Systems (COINS)*, 2022. https://dx.doi.org/10.1109/ODICON54453.2022.10010185.

[17] Thulasi Bikku, Mude Nikitha, Anjali Vajja, K. Harshitha, and Jhansi Rani. Optimized Machine Learning Algorithm to classify Phishing Websites. In

2022 IEEE International Conference on Electronics, Computing and Communication Technologies (CONECCT), 2022. https://dx.doi.org/10.1109/ICEARS53579.2022.9752223.

[18] Ruth Korede Ayeni, Ayodele Ariyo Adebiyi, Julius Olatunji Okesola, and Ennmanuel Igbekele. Phishing Attacks and Detection Techniques: A Systematic Review. In *2024 International Conference on Science, Engineering and Business for Driving Sustainable Development Goals (SEB4SDG)*, pages 1–17, Omu-Aran, Nigeria, April 2024. IEEE. https://ieeexplore.ieee.org/document/10630203/.

[19] G. Gopika, M. Sreekrishna, Katika Karthik, and C. Reddy. Privacy Preserving Secure and Efficient Detection of Phishing Websites Using Machine Learning Approach. In *2023 International Conference on Electronics, Computing and Communication Technologies (CONECCT)*, 2023. https://dx.doi.org/10.1109/ICECAA58104.2023.10212349.

[20] A. A and P. K. Towards the Detection of Phishing Attacks. In *2020 4th International Conference on Trends in Electronics and Informatics (ICOEI)*, 2020. https://dx.doi.org/10.1109/ICOEI48184.2020.9142967.

[21] Emtethal K. Alamri, Abdullah M. Alnajim, and Suliman A. Alsuhibany. Investigation of Using CAPTCHA Keystroke Dynamics to Enhance the Prevention of Phishing Attacks. *Future Internet*, 14(3):82, 2022. https://dx.doi.org/10.3390/fi14030082.

[22] Ponnurangam Kumaraguru, Young Rhee, Steve Sheng, Sayon Hasan, Alessandro Acquisti, Lorrie Faith Cranor, and Jason Hong. Getting users to pay attention to anti-phishing education: Evaluation of retention and transfer. In *Proceedings of the Anti-Phishing Working Groups 2nd Annual eCrime Researchers Summit*, pages 70–81, 2007.

[23] Darem A. Abawajy J. Alhashmi, A.A. Taxonomy of cybersecurity awareness delivery methods: A countermeasure for phishing threats. *International Journal of Advanced Computer Science and Applications*, 12(10):29–35, 2021.

[24] U.S. Department of Health and Human Services. Security standards: Technical safeguards - hipaa security rule, 2023. https://www.hhs.gov/hipaa/for-professionals/security/laws-regulations/index.html.

[25] European Parliament and Council of the European Union. General data protection regulation (gdpr) - article 39: Tasks of the data protection officer, 2016. https://gdpr-info.eu/art-39-gdpr/.

[26] Payment Card Industry Security Standards Council. Payment card industry data security standard (pci dss) v4.0, 2023. https://www.pcisecuritystandards.org/document_library.

[27] International Organization for Standardization (ISO). Iso/iec 27001:2022 - information security, cybersecurity and privacy protection, 2022. https://www.iso.org/standard/27001.

[28] National Institute of Standards and Technology (NIST). Framework for improving critical infrastructure cybersecurity, version 2.0, 2024. https://www.nist.gov/cyberframework.

[29] SANS Institute. Security awareness training - phishing awareness solutions, 2025. https://www.sans.org/security-awareness-training/products/security-awareness-solutions/phishing/.

[30] Kanchan Patil and Sai Rohith Arra. Detection of Phishing and User Awareness Training in Information Security: A Systematic Literature Review. In *2022 International Conference on Intelligent Computing and Networking (ICICN)*, pages 1–8. IEEE, 2022.

[31] Calic D. Delfabbro P. Reeves, A. "Get a red-hot poker and open up my eyes, it's so boring": Employee perceptions of cybersecurity training. *Computers & Security*, 106:1–13, 2021.

[32] Infosec Institute. Infosec iq phishing simulations - security awareness training, 2025. https://www.infosecinstitute.com/iq/.

[33] Paula Bitrián, Isabel Buil, Sara Catalán, and Dominik Merli. Gamification in workforce training: Improving employees' self-efficacy and information security and data protection behaviours. *Journal of Business Research*, 179:114685, June 2024. https://linkinghub.elsevier.com/retrieve/pii/S0148296324001899.

[34] Matthew Canham, Clay Posey, and Michael Constantino. Phish Derby: Shoring the Human Shield Through Gamified Phishing Attacks. *Frontiers in Education*, 6:807277, January 2022. https://www.frontiersin.org/articles/10.3389/feduc.2021.807277/full.

[35] ESET. Cybersecurity awareness training - gamified learning modules, 2025. https://www.eset.com/us/business/cybertraining/.

[36] Terranova Security. Phishing simulation platform - just-in-time training, 2025. https://www.phishlabs.com/services/security-awareness-training.

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

[38] Kathryn Coronges, R. Dodge, C. Mukina, Zachary Radwick, Joseph Shevchik, and E. Rovira. The Influences of Social Networks on Phishing Vulnerability. In *2012 45th Hawaii International Conference on System Sciences*, pages 5390–5399, 2012. https://dx.doi.org/10.1109/HICSS.2012.657.

[39] Murad Aburrous, M Anwar Hossain, Keshav Dahal, and Fadi Thabtah. Intelligent phishing detection system for e-banking using fuzzy data mining. In *2010 International Conference for Internet Technology and Secured Transactions*, pages 1–6. IEEE, 2010.

[40] Devdatta Akhawe and Adrienne Porter Felt. Alice in warningland: A large-scale field study of browser security warning effectiveness. In *22nd USENIX Security Symposium (USENIX Security 13)*, pages 257–272, 2013.

[41] S. M. Albladi and G. R. S. Weeir. Features of users influencing social engineering attacks in social networks. *Human-centered information sciences and computing*, 8(1):1–24, 2018. https://doi.org/10.1186/s13673-018-0128-7.

[42] C. Horton, C. McReynolds, B. Doran, V. Johnson, and G. Veletsianos. Increasing cybersecurity awareness and engagement: Evaluating an escape room approach. 2018(1):3.

[43] Daniele Lain, Kari Kostiainen, and Srdjan Capkun. Phishing in Organizations: Findings from a Large-Scale and Long-Term Study. http://arxiv.org/abs/2112.07498, December 2021.

[44] Ayman El Aassal and Rakesh M. Verma. Spears Against Shields: Are Defenders Winning the Phishing War? In *Proceedings of the 4th ACM Workshop on Security Information Workers*, pages 3–10, 2019. https://dx.doi.org/10.1145/3309182.3309191.

[45] Angelo Carella, Maxim Kotsoev, and Tudor M Truta. Impact of security awareness training on phishing click-through rates. In *2017 IEEE International Conference on Big Data (Big Data)*, pages 4458–4466. IEEE, 2017.

[46] Aggarwal P. Rajivan P. Gonzalez C. Singh, K. Cognitive elements of learning and discriminability in anti-phishing training. *Computers & Security*, 127:103105, 2023.

[47] Alex Sumner, Xiaohong Yuan, Mohd Anwar, and Maranda McBride. Examining Factors Impacting the Effectiveness of Anti-Phishing Trainings. *Journal of Computer Information Systems*, 62(5):975–997, September 2022. https://www.tandfonline.com/doi/full/10.1080/08874417.2021.1955638.

[48] Anne Clara Tally, Jacob Abbott, Ashley M Bochner, Sanchari Das, and Christena Nippert-Eng. Tips, Tricks, and Training: Supporting Anti-Phishing Awareness among Mid-Career Office Workers Based on Employees' Current Practices. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, pages 1–13, Hamburg Germany, April 2023. ACM. https://dl.acm.org/doi/10.1145/3544548.3580650.

[49] Ingolf Becker, Simon Parkin, and M. Angela Sasse. Finding Security Champions in Blends of Organisational Culture. In *Proceedings 2nd European Workshop on Usable Security*, Paris, France, 2017. Internet Society. https://www.ndss-symposium.org/wp-content/uploads/2018/03/eurousec2017_07_Becker_paper.pdf.

[50] S. Barth, M. D. T. de Jong, M. Junger, P. H. Hartel, and J. C. Roppelt. Putting the privacy paradox to the test: Online privacy and security behaviors among users with technical knowledge, privacy awareness, and financial resources. *Telematics and Informatics*, 41:55–69, 2019. https://doi.org/10.1016/j.tele.2019.03.003.

[51] John M. Blythe and Lynne Coventry. Costly but effective: Comparing the factors that influence employee anti-malware behaviours. *Computers in Human Behavior*, 87:87–97, October 2018. https://linkinghub.elsevier.com/retrieve/pii/S0747563218302498.

[52] Eric Amankwa, Marianne Loock, and Elmarie Kritzinger. Establishing information security policy compliance culture in organizations. *Information & Computer Security*, 26(4):420–436, October 2018. https://www.emerald.com/insight/content/doi/10.1108/ICS-09-2017-0063/full/html.

[53] Matthew Canham, Shanée Dawkins, and Jody Jacobs. Not All Victims Are Created Equal: Investigating Differential Phishing Susceptibility. In Dylan D. Schmorrow and Cali M. Fidopiastis, editors, *Augmented Cognition*, volume 14694, pages 3–21. Springer Nature Switzerland, Cham, 2024. https://link.springer.com/10.1007/978-3-031-61569-6_1.

[54] Ryan Wright, Steven Johnson, and Brent Kitchens. Phishing Susceptibility in Context: A Multilevel Information Processing Perspective on Deception Detec-

tion. *MIS Quarterly*, 47(2):251–270, 2023. https://dx.doi.org/10.25300/misq/2022/16625.

[55] S. Furman, M. Theofanos, Yee-Yin Choong, and Brian C. Stanton. Basing Cybersecurity Training on User Perceptions. *IEEE Security & Privacy*, 10(2):60–63, 2012.

[56] Morrison B.W. Ingrey K. Wiggins M.W. Bayl-Smith P. Morrison N. Ackerley, M. Errors, irregularities, and misdirection: Cue utilisation and cognitive reflection in the diagnosis of phishing emails. *Australasian Journal of Information Systems*, 26:1–21, 2022.

[57] Hafzullah İŞ. LLM-Driven SAT Impact on Phishing Defense: A Cross-Sectional Analysis. In *2024 12th International Symposium on Digital Forensics and Security (ISDFS)*, pages 1–5, San Antonio, TX, USA, April 2024. IEEE. https://ieeexplore.ieee.org/document/10527274/.

[58] Valenzuela C. Plate O. Tanvir T. Auton-J.C. Bayl-Smith P. Wiggins M.W. Sturman, D. The role of cue utilization in the detection of phishing emails. *Applied Ergonomics*, 106:1–13, 2023.

[59] Joakim Kävrestad, Allex Hagberg, Marcus Nohlberg, Jana Rambusch, Robert Roos, and Steven Furnell. Evaluation of Contextual and Game-Based Training for Phishing Detection. *Future Internet*, 14(4):104, 2022.

[60] Jamaine Mungo. Self-paced cybersecurity awareness training educating retail employees to identify phishing attacks. *Journal of Cyber Security Technology*, 8(2):71–119, April 2024. https://www.tandfonline.com/doi/full/10.1080/23742917.2023.2244210.

[61] Ashwinee Panda, Christopher A. Choquette-Choo, Zhengming Zhang, Yaoqing Yang, and Prateek Mittal. Teach LLMs to Phish: Stealing Private Information from Language Models. http://arxiv.org/abs/2403.00871, March 2024.

[62] Sarah Tabassum, Cori Faklaris, and Heather Richter Lipford. What Drives SMiShing Susceptibility? A U.S. Interview Study of How and Why Mobile Phone Users Judge Text Messages to be Real or Fake.

[63] Katherine R. Garcia, Jeremiah Ammons, Xiangrui Xu, and Jing Chen. Phishing in Social Media: Investigating Training Techniques on Instagram Shop. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 67(1):1850–1855, September 2023. https://journals.sagepub.com/doi/10.1177/21695067231192588.

[64] I. Ghafir, J. Saleem, M. Hammoudeh, H. Faour, V. Prenosil, S. Jaf, S. Jabbar, and T. Baker. Security threats to critical infrastructure: The human factor. *Journal of Supercomputing*, 74(10):4986–5002, 2018. https://doi.org/10.1007/s11227-018-2337-2.

[65] Loukas G. Gan D. Heartfield, R. You are probably not the weakest link: Towards practical prediction of susceptibility to semantic social engineering attacks. *IEEE Access*, 4:6910–6928, 2016.

[66] Dirk-jan Mollema. Phishing the Phishing Resistant Phishing for Primary Refresh Tokens in Microsoft Entra.

[67] Fred Heiding, Simon Lermen, Andrew Kao, Bruce Schneier, and Arun Vishwanath. Evaluating Large Language Models' Capability to Launch Fully Automated Spear Phishing Campaigns: Validated on Human Subjects. http://arxiv.org/abs/2412.00586, November 2024.

[68] Daniel Nahmias, Gal Engelberg, Dan Klein, and Asaf Shabtai. Prompted Contextual Vectors for Spear-Phishing Detection. http://arxiv.org/abs/2402.08309, February 2024.

[69] Peter K. K. Loh, Aloysius Z. Y. Lee, and Vivek Balachandran. Towards a Hybrid Security Framework for Phishing Awareness Education and Defense. *Future Internet*, 16(3):86, March 2024. https://www.mdpi.com/1999-5903/16/3/86.

[70] Giuseppe Desolda, Lauren S. Ferro, Andrea Marrella, Tiziana Catarci, and Maria Francesca Costabile. Human Factors in Phishing Attacks: A Systematic Literature Review. *ACM Computing Surveys*, 54(8):1–35, November 2022. https://dl.acm.org/doi/10.1145/3469886.

[71] Simon Parkin. "'What Keeps People Secure is That They Met The Security Team': Deconstructing Drivers And Goals of Organizational Security Awareness". https://data.4tu.nl/datasets/9dc01aa6-8274-43f4-b137-6d185e7008d1, February 2024.

[72] Luigi Gallo, Danilo Gentile, Saverio Ruggiero, Alessio Botta, and Giorgio Ventre. The human factor in phishing: Collecting and analyzing user behavior when reading emails. *Computers & Security*, 139:103671, April 2024. https://linkinghub.elsevier.com/retrieve/pii/S0167404823005813.

[73] Marco De Bona and Federica Paci. A real world study on employees' susceptibility to phishing attacks. In *Proceedings of the 15th International Conference on Availability, Reliability and Security*, pages 1–10, Virtual Event Ireland, August 2020. ACM. https://dl.acm.org/doi/10.1145/3407023.3409179.

[74] Nina Marshall, Daniel Sturman, and Jaime C. Auton. Exploring the evidence for email phishing training: A scoping review. *Computers & Security*, 139:103695, April 2024. https://linkinghub.elsevier.com/retrieve/pii/S0167404823006053.