Personalized Guidelines for Design, Implementation and Evaluation of Anti-phishing Interventions

Orvila Sarker
University of Adelaide
Adelaide, Australia
orvila.sarker@adelaide.edu.au

Sherif Haggag
University of Adelaide
Adelaide, Australia
sherif.haggag@adelaide.edu.au

Asangi Jayatilaka
University of Adelaide
Adelaide, Australia
asangi.jayatilaka@adelaide.edu.au

Chelsea Liu
University of Adelaide
Adelaide, Australia
chelsea.liu@adelaide.edu.au

***Abstract*—Background: Current anti-phishing interventions, which typically involve one-size-fits-all solutions, suffer from limitations such as inadequate usability and poor implementation. Human-centric challenges in anti-phishing technologies remain little understood. Research shows a deficiency in the comprehension of end-user preferences, mental states, and cognitive requirements by developers and practitioners involved in the design, implementation, and evaluation of anti-phishing interventions. Aims: This study addresses the current lack of resources and guidelines for the design, implementation and evaluation of anti-phishing interventions, by presenting personalized guidelines to the developers and practitioners. Method: Through an analysis of 53 academic studies and 16 items of grey literature studies, we systematically identified the challenges and recommendations within the anti-phishing interventions, across different practitioner groups and intervention types. Results: We identified 22 dominant factors at the individual, technical, and organizational levels, that affected the effectiveness of anti-phishing interventions and, accordingly, reported 41 guidelines based on the suggestions and recommendations provided in the studies to improve the outcome of anti-phishing interventions. Conclusions: Our dominant factors can help developers and practitioners enhance their understanding of human-centric, technical and organizational issues in anti-phishing interventions. Our customized guidelines can empower developers and practitioners to counteract phishing attacks.**

***Index Terms*—human factor, personalized guidelines, phishing education, phishing training, phishing awareness, phishing intervention**

I. Introduction

Phishing is a form of cyber attacks committed by using fraudulent and deceitful communication techniques, such as emails or messages, to entrap users into providing sensitive personal information, such as passwords, credit card details, or social security numbers [1]. Automated detection techniques can provide a line of defence against phishing (e.g., [2]), however, ultimately end users serve as the final safeguard. The abundance of successful recent phishing attacks indicates that further efforts are needed to enhance end users' phishing education, training, and awareness (PETA) [3], [4].

Empirical investigations have demonstrated that tailored design, implementation, and evaluation of phishing interventions can be efficacious in helping users recognize and mitigate phishing hazards [5]–[8]. A phishing intervention refers to any anti-phishing system, software, tool, or framework that helps users deal with a phishing attack and requires user intervention [9]. The cognitive needs and mental status of individual end users play an important role in determining the effectiveness of phishing intervention and, consequently, ought to take into consideration by the developers and practitioners during the design, implementation and evaluation process of phishing intervention [10]. However, evidence shows that developers often neglect end-users' decision-making processes and cognitive limitations in the design, implementation, and evaluation of phishing interventions. This leads to human-centric weaknesses or usability issues, rendering the end-users susceptible to phishing attacks [11]–[13]. In order to ensure the efficacy and usefulness of anti-phishing interventions, their key features such as the content and methods of anti-phishing intervention must be tailored to the needs of individual users [14]–[16].

Motivation: The effectiveness of phishing education, training, and awareness (PETA) interventions significantly depends on decisions made by various developers and practitioners involved in the design, implementation, and evaluation of these interventions. Studies document numerous examples of failures in such decision-making: some email providers do not use Simple Mail Transfer Protocol (SMTP) authentication mechanisms, thus allowing the attackers to send emails from spoofed email addresses [17]; many web developers of e-commerce enterprises fail to employ Secure Socket Layer (SSL) to secure their login page [12]; organizations' security officers often conduct phishing training without following any formal policies [18] and use unrealistic or irrelevant email templates [11], [19], [20]; developers design interventions with complex user interfaces that require specialized knowledge to install and utilize [21]–[24]. These examples highlight the need

978-1-6654-5223-6/23$31.00 © 2023 Crown

for a structured set of guidelines provided to developers and practitioners to help them comprehend the diverse range of needs and challenges faced by end-users. However, currently, there is a limited availability of such guidelines for developers and practitioners. Moreover, most existing resources are intended for a singular group of practitioners or a particular class of cyber security interventions but are not specific to phishing interventions. For instance, Lujo et al. [25] and Lynsay et al. [26] reported guidelines for browser security warning design (this is a type of phishing intervention), and Mirium et al. [27] provided guidelines for the development of cyber security games (not directly related to phishing). Consequently, the target user group of these guidelines are primarily designers or developers of phishing interventions, with limited relevance to other practitioners such as organization managers or cyber security officers. Guidelines for IT security management (e.g., proposed by Pooya et al. [28] and Sonia et al. [29]) do not offer insights specific to phishing prevention. Overall, current guidelines and best practices for anti-phishing interventions (particularly on their design, implementation, and evaluation) are scattered around various academic and grey literature studies and not presented to practitioners in an easily accessible and personalized format.

In light of these needs and challenges faced by developers and cyber security practitioners, we aim to investigate the following research question in our study: *what guidance can be provided to support the developers and practitioners in addressing usability issues in anti-phishing interventions from an individual, technical and organizational perspective?* Our study has devised a set of guidelines by synthesizing the existing literature to aid practitioners to obtain a more holistic understanding of end-user needs and experiences, in order to enhance the effectiveness of the design, implementation, and evaluation of anti-phishing interventions.

Contributions: • From the existing literature, we have identified *22 dominant factors* which impact the effectiveness of anti-phishing interventions. These include 15 human factors (such as age, complacency, and educational qualification), 4 technical factors (such as device type and gamer type) and 3 organizational factors (such as organizational position and working hours). Our presented dominant factors can assist practitioners to attain a deeper understanding of the important determinants of the success of phishing interventions. • We have reported *41 guidelines* on the design, implementation, and evaluation of anti-phishing interventions, which are systematically compiled based on recommendations derived from a comprehensive Multi-vocal Literature Review (MLR) of 69 studies, thus making the guidelines the first of its kind in terms of its comprehensiveness and breadth of coverage. Our devised guidelines can be a useful resource to aid practitioners to improve the efficacy of the design, implementation and evaluation of anti-phishing interventions.

II. Methodology

Our methodology, summarized in Fig 1, consists of five steps, as detailed in this Section.

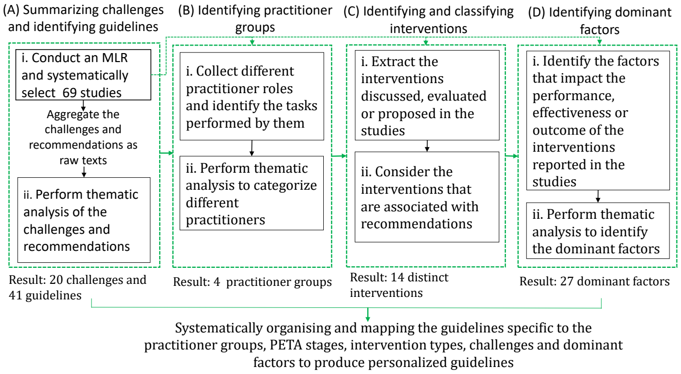

(A) Summarizing challenges and identifying guidelines
(B) Identifying practitioner groups
(C) Identifying and classifying interventions
(D) Identifying dominant factors
i. Conduct an MLR and systematically select 69 studies
Aggregate the challenges and recommendations as raw texts
ii. Perform thematic analysis of the challenges and recommendations
i. Collect different practitioner roles and identify the tasks performed by them
ii. Perform thematic analysis to categorize different practitioners
i. Extract the interventions discussed, evaluated or proposed in the studies
ii. Consider the interventions that are associated with recommendations
i. Identify the factors that impact the performance, effectiveness or outcome of the interventions reported in the studies
ii. Perform thematic analysis to identify the dominant factors
Result: 20 challenges and 41 guidelines
Result: 4 practitioner groups
Result: 14 distinct interventions
Result: 27 dominant factors
Systematically organising and mapping the guidelines specific to the practitioner groups, PETA stages, intervention types, challenges and dominant factors to produce personalized guidelines

Fig. 1: Methodology of this study

A. Summarizing challenges and identifying guidelines

1) Conduct an MLR and systematically select 69 studies: We conduct an MLR with the aim of identifying the challenges and recommendations concerning anti-phishing interventions. As anti-phishing interventions are inherently industry-oriented, we incorporate the important perspectives of practitioners by including grey literature studies. Our MLR adheres to the protocols outlined by Kitchenham and Charters [30] and Garousi et al. (2019) [31].

In our study selection process, we utilized the Scopus [32] database, as it offers greater breadth and depth of academic literature compared to other databases [33], [34]. We ran a pilot study with ACM digital library and IEEE Xplore digital library to ensure that Scopus is comprehensive. Following a thorough examination of the search keywords used in prior review papers in this field (e.g., [35], [36]) and refinement through conducting pilot searches, we use the search keywords "aware*" or "interven*" or "nudge*" or "warn*" or "protect*" or "security indicators" or "alert*" along with the keyword "phish*" to collect the academic studies. To ensure the quality of the collected studies, we only included studies that had a CORE [37] rank of A*, A and B. The CORE ranking system aims to ensure high-quality standards and rigorous peer review processes for selected journals and conferences [38]. We omitted papers that had a CORE rank B and were published before 2012. The reason for this is that our pilot study revealed that a number of CORE B papers that were published prior to 2012 proposed client-side anti-phishing tools without conducting a real-world evaluation of their usability (e.g., Passpet [39]). We excluded studies that have provided automated anti-phishing solutions (defined in [40]) and research that were not written in English, short papers (less than 6 pages), and literature survey papers.

To collect the grey literature studies we employed Google as our search database. Google serves as a widely accepted and utilized search engine for gathering grey literature study [41]–[43]. For the grey literature study collection, we used the search terms "education", "training", "awareness" with the term "phish*". The rationale for employing an alternative search query distinct from the one employed in the academic study stems from the search methodology employed by

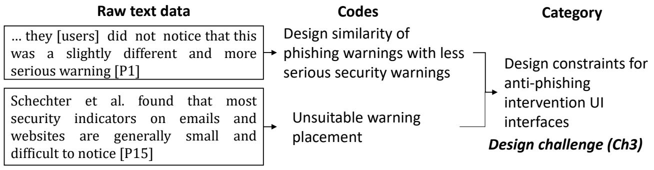
Google; Google conducts searches using the specified search terms across its entire index of webpages [44]. Hence, apart from the duplicated webpages and academic studies that have already been identified in Scopus, our pilot study using the identical search string as employed in academic studies yielded numerous extraneous results. Therefore, we used different search terms to obtain more relevant results following the suggestions by Garousi et. al [31].

The grey literature study collection process concluded on Google's 16th page as no new or redundant information was identified, as recommended by Garousi et al [31]. To ensure the quality of the grey literature studies, we assessed the publication's authority, methodology, presence of reliable references, the date of publication, the novelty of the article, and the article's outlet type as suggested by Garousi et al [31]. More information on our literature selection process, including the search string, inclusion/exclusion criteria, quality assessment criteria, and the list of selected studies in the academic and grey literature, is provided in our online appendix [45].

After applying the inclusion/exclusion criteria and data quality assessment, we collected a total of 69 studies, including 53 academic and 16 grey literature studies. We use the symbol P[*] to denote the studies throughout the rest of the paper.

2) Perform thematic analysis of the challenges and recommendations: To identify the challenges and recommendations documented in the literature, we utilized thematic analysis - a standard data analysis method for qualitative data - to process the raw textual data (challenges and recommendations). In particular, we adhered to the process outlined by Braun and Clarke [46] by using Nvivo - a tool designed for qualitative data analysis [46]. After extracting the textual data into an Excel spreadsheet, we imported the data into Nvivo to perform open coding, which involves breaking down the data into smaller components and assigning labels to each component [47]. This process was conducted iteratively, with codes generated in the initial stage being modified and updated in later stages. Examples of the data analysis process for challenge and guideline identification are shown in Fig. 2 and Fig. 3 respectively.

From the analyzed data, we identified 20 challenges pertaining to the design, implementation, and evaluation phases of phishing education, training, and awareness interventions. We classify the challenges into three broad categories: first, design challenges relate to the concept, functionality, feature, or user interface of anti-phishing interventions. Examples of design challenges include *inconsistent UI design across browsers and mobile devices* [P22, P49], *unsuitable warning placement* [P2, P3, P5, P7, P11, P15, P25], and *complex interface and configuration in the game design* [P28]. Second, implementation challenges relate to intervention automation, deployment, and adoption (e.g., *the interdependancy between different factors and platforms in implementing effective anti-phishing measures* [P23, P38, P50]). Third, evaluation challenges relate to the measurement of usability and effectiveness of anti-phishing interventions and quantification of training outcomes. As developers and practitioners are the primarily targeted user

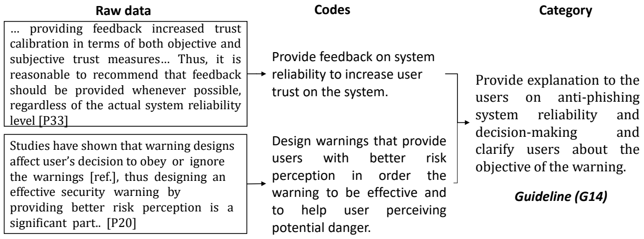

Raw text data
Codes
Category

... they [users] did not notice that this was a slightly different and more serious warning [P1]

Schechter et al. found that most security indicators on emails and websites are generally small and difficult to notice [P15]

Design similarity of phishing warnings with less serious security warnings
Unsuitable warning placement
Design constraints for anti-phishing intervention UI interfaces
Design challenge (Ch3)

Fig. 2: Identification of the challenges using thematic analysis



... providing feedback increased trust calibration in terms of both objective and subjective trust measures... Thus, it is reasonable to recommend that feedback should be provided whenever possible, regardless of the actual system reliability level [P33]

Studies have shown that warning designs affect user's decision to obey or ignore the warnings [ref.], thus designing an effective security warning by providing better risk perception is a significant part.. [P20]

Provide feedback on system reliability to increase user trust on the system.
Design warnings that provide users with better risk perception in order the warning to be effective and to help user perceiving potential danger.
Provide explanation to the users on anti-phishing system reliability and decision-making and clarify users about the objective of the warning.
Guideline (G14)

Fig. 3: Identification of guidelines using thematic analysis

groups of our personalized guidelines, we excluded challenges relevant exclusively to researchers (i.e., "Ch17 - limited demographic consideration in the study settings" [P1, P13, P14]).

We collected several suggestions or recommendations documented in the literature such as "feedback should be provided whenever possible, regardless of the actual system reliability level [P33]", "designing an effective security warning by providing better risk perception is a significant part [P20]". We performed the thematic analysis of these recommendations to report 41 guidelines.

B. Identifying practitioner groups

To report personalized guidelines for anti-phishing intervention developers and practitioners, it is important to identify their specific needs and responsibilities involved in the design, implementation, and evaluation processes of anti-phishing interventions. To derive a mapping between practitioner groups and their roles and responsibilities, following existing studies in other domains (e.g., [48]–[50]) we categorize different practitioner groups based on their functional roles as documented in the literature. From the literature we first gathered information on various practitioner roles, specifically 28 different groups, such as browser developer [P3, P8], platform designer [P4], designers of anti-phishing applications [P18], game developer [P36], information security officer [P27], chief information security manager [P34], and cyber security practitioners and decision makers [P53]. We then carefully investigated the responsibilities performed by each practitioner groups based on the recommendations collected from the MLR. For example, from the following textual data "*phishing campaign managers need to organize multiple successive simulation/training cycles to cultivate a phishing awareness culture [P53]*", we infer that one of the responsibilities of *phishing campaign managers* is to execute and evaluate anti-phishing education and training programs. Fig. 4 shows an example of how we identified the responsibilities of different practitioner groups from raw textual data in the literature.

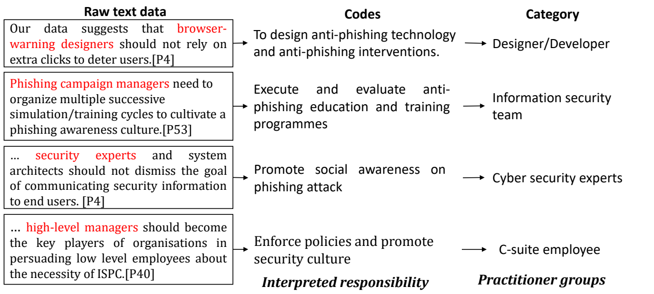

Raw text data
Codes
Category
Our data suggests that browser-warning designers should not rely on extra clicks to deter users.[P4]
To design anti-phishing technology and anti-phishing interventions.
Designer/Developer
Phishing campaign managers need to organize multiple successive simulation/training cycles to cultivate a phishing awareness culture.[P53]
Execute and evaluate anti-phishing education and training programmes
Information security team
... security experts and system architects should not dismiss the goal of communicating security information to end users. [P4]
Promote social awareness on phishing attack
Cyber security experts
... high-level managers should become the key players of organisations in persuading low level employees about the necessity of ISPC.[P40]
Enforce policies and promote security culture
C-suite employee
Interpreted responsibility
Practitioner groups

Fig. 4: Categorization of different practitioner groups using thematic analysis

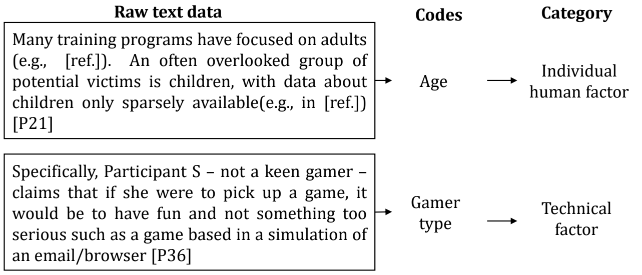

Raw text data
Codes
Category
Many training programs have focused on adults (e.g., [ref.]). An often overlooked group of potential victims is children, with data about children only sparsely available(e.g., in [ref.]) [P21]
Age
Individual human factor
Specifically, Participant S – not a keen gamer – claims that if she were to pick up a game, it would be to have fun and not something too serious such as a game based in a simulation of an email/browser [P36]
Gamer type
Technical factor

Fig. 5: Dominant factor identification from raw text data

Based on these identified responsibilities of each practitioner groups, we then categorized all practitioners involved in anti-phishing interventions into four major categories, namely *designer/developer*, *information security team*, *cyber security experts*, and *C-suite employees*. Table I summarizes the responsibilities of different practitioner groups. A fraction of the responsibilities overlap across different practitioner groups. For example, a designer/developer whose main responsibility is to *design* anti-phishing tools may also *evaluates* an anti-phishing software [P22, P57, P61, P66, P67], because a designer/developer may need to test the usability of interventions before finalizing the design. However, as shown in Table I, evaluating anti-phishing interventions is mainly carried out by organizations' information security teams.

C. Identifying and classifying interventions

For each study collected in our MLR, we carefully examined the introduction, methodology, and results sections in search of any interventions that were suggested, debated, or evaluated. Consequently, in order to tailor the guidelines to each type of intervention, we only included those treatments that had pertinent recommendations reported in the research. We classified the interventions into three types - education, training, or awareness - based on characteristics such as their intended goal, presentation, and method of delivery of anti-phishing information to users (the terms phishing education, training and awareness are defined in our online appendix [45]). We identified 2 types of phishing education: *anti-phishing instructions* and *educational games*. Identified phishing training interventions are: *phishing simulation and embedded training*, *phishing training game*, *narrative-based training*, *instructor-based training*, *information and guidance-based training*. For phishing awareness interventions, we found *email client phishing indicators*, *browser SSL warnings*, *browser EV certificate warning*, *browser security toolbar*, *browser phishing warning*, *QR code scanner phishing warnings* and *Interactive custom phishing indicator*. Please refer to our online appendix [45] for the definitions of each intervention.

D. Identifying dominant factors

From the raw texts collected on the challenges and recommendations relating to anti-phishing interventions (as discussed in Section II-A2), we investigated the main *reasons* for poor outcomes of existing anti-phishing interventions, as well as areas of improvement suggested by the authors to achieve better user experience. Based on this synthesized information, we identify and coin the term *dominant factors*, which refers to the individual, technical or organisational factors that may either enhance or impede the overall outcome of anti-phishing interventions. These factors are called *dominant* as these factors were argued to influence the outcome of the anti-phishing interventions after empirical evaluation with the users in the results and discussions of our collected studies.

We adopt the terminology utilized by prior researchers [35], [51], [52] to designate the dominant factors identified in our investigation. As an example of identifying a dominant factor, from the textual data *"staff may expect to learn more from experts while [college] students may expect to learn more from their peers [P48]"*, we derived an understanding that, according to the authors, the education qualifications of users have a significant bearing in determining their preference for the type of training methods and their effectiveness. This information indicates that designers could consider the educational qualifications of the end-users to improve user experience with the phishing intervention. Accordingly, the dominant factor identified here is users' *educational qualification*. Fig. 5 illustrates an example of dominant factor identification from the raw textual data.

Fig. 6 depicts an example of the interconnection among challenges, guidelines, practitioner groups, interventions, and dominant factors derived from the raw text data. To determine the interconnection between challenges and interventions, we identify the interventions discussed in the study and the limitations mentioned within those interventions. Similarly, in establishing the interconnection among guidelines, interventions, and practitioner groups, we search for recommendations that are directed towards practitioners to enhance the outcomes of specific interventions.

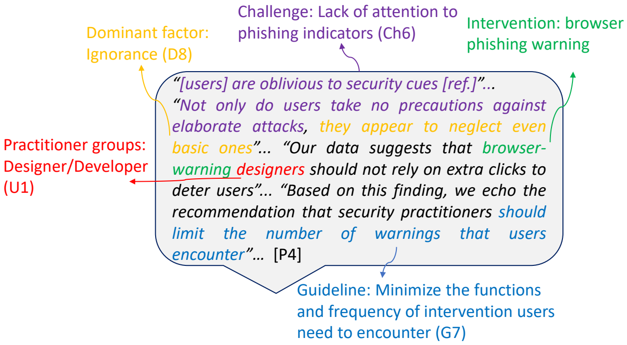
III. Research Findings

A. Dominant factors impacting anti-phishing interventions

We analyzed the human factors discussed by Dupont [51] and refined by Desolda et al. [35] in the context of phishing attacks. A total of 8 human-oriented factors identified from our textual data (complacency, distraction, lack of communication, pressure, lack of knowledge, lack of resources, fatigue, and norms) confirm the factors documented by Dupont. A number

TABLE I: Identified practitioner groups and their responsibilities

No	Practitioners	Responsibilites/Activities
U1	Designer/ Developer	• Design and deploy anti-phishing technology and anti-phishing interventions [P4, P19, P54, P60, P61, P65]. • Conduct usability testing of anti-phishing intervention to improve the design [P22, P57, P61, P66, P67].
U2	Information security team	• Implement the anti-phishing technology and anti-phishing solutions within the organization [P26, P50, P53 P61, P63, P67]. • Execute and evaluate anti-phishing education and training programmes [P30, P54, P59, P61, P62, P69].
U3	Cyber security experts	• Make decisions on the appropriate elements and aspects to be included in the anti-phishing interventions [P13, P41, P42], promote social awareness on phishing attack [P4, P21].
U4	C-suite employee	• Enforce policies to educate and train employees against phishing attack [P11, P38, P40 P50, P53, P54, P57, P67] and to promote security culture [P21, P26, P38, P40, P50, P56, P57, P59, P60, P61, P67, P68, P69] within the organization. • Collaborate with the organization's security team to adopt strong anti-phishing measures and prepare and execute a phishing incident response plan [P53, P56, P57, P60, P61].



Dominant factor: Ignorance (D8)
Challenge: Lack of attention to phishing indicators (Ch6)
Intervention: browser phishing warning
Practitioner groups: Designer/Developer (U1)
Guideline: Minimize the functions and frequency of intervention users need to encounter (G7)

"[users] are oblivious to security cues [ref.]"... "Not only do users take no precautions against elaborate attacks, they appear to neglect even basic ones"... "Our data suggests that browser-warning designers should not rely on extra clicks to deter users"... "Based on this finding, we echo the recommendation that security practitioners should limit the number of warnings that users encounter"... [P4]

Fig. 6: Example interconnection among challenges, guidelines, practitioner groups, interventions and dominant factors

of additional factors have emerged from our textual data, prompting us to search the literature for examples of their manifestation in order to categorize these factors. Following the approach by a study in the field of software engineering by Dulaji et al. [52], we were able to identify and name the other previously unexplored factors, which make a significant contribution of new knowledge to this field.

We identified 22 dominant factors to underscore the significance of incorporating user needs and preferences in the design, implementation, and evaluation of anti-phishing interventions. These dominant factors emphasize that neglecting user's requirements and inclinations can hinder practitioners' efforts in providing tailored design, implementation, and assessment of a system, which in turn results in compromised usability and suboptimal outcomes. Table II details our 22 identified dominant factors and how their inclusion or absence can influence the outcome of anti-phishing interventions. We grouped our dominant factors into three categories: individual human factors, technical factors, and organizational factors.

1) Individual human factors: Based on the analysis of selected studies, we identified that various demographic characteristics of individual users require greater attention from developers and practitioners to enhance the efficacy of anti-phishing interventions across different user groups. According to the literature, developers and practitioners need to take into consideration a number of user demographic characteristics, including *age* and *educational qualification*, in order to enhance the efficacy of anti-phishing interventions across different user groups.

In addition to demographic characteristics, our additional dominant factors also relate to cognitive functioning and limitations of individual users in order to provide a comprehensive picture of user-level characteristics. Specifically, *knowledge decay*, *distraction*, *lack of attention* and *lack of motivation* all constitute individual limitations that can reduce user's capacity to effectively identify or counteract phishing attacks. For example, studies have shown that the knowledge acquired by individuals through phishing training tends to degrade or dissipate over time [P7, P13, P21, P31, P34]. Again, phishing warnings frequently pass unnoticed as users become diverted by concurrent tasks or are incapable of maintaining attention across multiple stimuli [P13, P14, P41].

Our investigation reveals that certain personality traits or user characteristics, such as *complacency* and *optimism bias*, may lead to users disregarding anti-phishing interventions. For example, users tend to overestimate the efficacy of their organization's anti-phishing measures [P7] and exhibit over-reliance on website content that is visually attractive [P2, P5, P8, P11, P49]. It is noteworthy to mention that our dominant factors represent distinct attitudes or mindsets. For example, complacency involves a belief that things are good enough as they are and that there is no need for further effort or change [53] which may cause complacent individuals to overlook potential phishing risks. Conversely, optimistic individuals feel less likely to experience cyber attacks which may cause them to share passwords, visit untrustworthy websites and so on [54].

An absence of tailored design, implementation, and evaluation of anti-phishing interventions by taking into consideration user needs and preferences could result in overwhelming the user and causing *security fatigue*. For example, fatigue can result from frequent exposure to warning and risk notifications, thereby reducing their effectiveness [P4, P13, P14, P17, P18, P26]. Similarly, excessive training can lead to training fatigue [P34, P52, P53, P58, P60, P61, P62, P69]. The complexity inherent in software installation procedures, as well as the intricacy of the process for reporting phishing incidents, may lead to *lack of user motivation*.

2) Technical factors: The effectiveness of anti-phishing warnings is greatly influenced by the *type of devices* utilized by the users.

TABLE II: Dominant factors and their impact on anti-phishing interventions

No	Factors	Impact	#
		Individual human factors	
D1	Age	• Children aged 8-13 require specialized phishing educational intervention as they are biologically less attentive [P21, P35]. • Teenagers tend to make decisions quickly without considering the consequences, and are more susceptible to being persuaded by urgency and panic-inducing phishing emails [P35]. • Older employees have relatively bad training outcomes as they prioritize maintenance over growth [P40]. • Age 18-25 are vulnerable to phishing attacks [P6].	4
D2	Complacency	• Users' overconfidence in the appealing web content leads them to disregard phishing warnings [P2, P5, P8, P11, P49]. • Users' prior experience with websites results in overconfidence, causing them to disregard phishing warnings [P1, P2, P5, P11, P13, P25, P44]. • Users over-rely on site reputation and trust the warning [P14]. • Users are overconfident about their ability to detect phishing [P43, P44, P45], over-trust on their organizational technical phishing solutions [P7].	14
D3	Confusion	• User confusion arises due to similarity in domain names [P45], webhosting [P45, P49], distinct warning design patterns among vendors [P25, P49] and conflicting information present in the anti-phishing guidelines [P42]. • Users become confused about the purpose of a received training email [P5].	5
D4	Curiosity	• Users click on the phishing link out of curiosity [P2, P25].	2
D5	Distraction	• Users are distracted by other tasks as security is not their main concern [P13, P14, P41]. • Individuals are unable to focus on multiple things simultaneously (e.g., noticing on phishing warning while doing online shopping) [P13].	3
D6	Educational Qualification	• Phishing stories from a peer is an effective method of training for college students [P48]. • University staffs learn better from facts from an expert-based training method [P48]. • Compared to bachelor's degree, users having master's and PhD degrees are more confident in detecting phishing [P52].	2
D7	Knowledge decay	• The knowledge gained by users during phishing training tends to dissipate over time [P7, P13, P21, P31, P34].	5
D8	Ignorance	• Users failed to look at anti-phishing interventions [P7, P13, P17, P28], ignored as web content looked legitimate [P2] and when received a high frequency of warnings [P4].	6
D9	Lack of communication	• Before designing and implementing anti-phishing software, users' interests and needs are not well investigated [P16].	1
D10	Lack of motivation	• Users are not motivated enough to install anti-phishing software on their devices [P31, P37], show unwillingness to report phishing due to a complicated reporting process [P50, P58, P63] and do not find the training and educational material interesting [P10, P19, P28].	8
D11	Lack of trust	• Users do not trust anti-phishing warnings due to limited accuracy of anti-phishing tools [P1, P11].	2
D12	Optimism bias	• Optimistic users tend to be less conscious as they believe that negative events only happen to others [P13].	1
D13	Perceived vulnerability and severity	• An individual's heightened understanding of the consequences of phishing attacks enhances their resistance to these types of attacks [P40].	1
D14	Pressure	• Phishing incident response by IT staff gets delayed due to the reception of a high volume of phishing reports [P50]. • An individual receiving a high volume of emails is more susceptible to phishing attacks [P26].	2
D15	Fatigue	• Providing comprehensive instruction could result in overwhelming the user [P13]. • Frequent exposure to warning causes warning fatigue [P4, P13, P14, P17, P18, P26]. • Frequent risk notifications and excessive training result in training fatigue [P34, P53, P58, P60, P61, P62, P69].	13
		Technical factors	
D16	Device type	• Individuals who rely on mobile devices are at a higher risk, as phishing signs are obscured or not fully visible on the small screens of mobile devices [P49].	1
D17	Gamer type	• A casual player is unsatisfied with playing a phishing game that is designed for serious gamers, and conversely, a serious gamer is unfulfilled playing a phishing game that is intended for casual players [P36].	1
D18	Lack of knowledge	• Users do not understand anti-phishing warnings due to lack of knowledge about security and security indicators [P1, P4, P5, P6, P7, P8, P9, P10, P11, P13, P14, P17, P20, P21, P28, P35, P39, P46, P47, P49].	20
D19	Lack of resource	• User do not have enough infrastructure support when they work from home [P6]. • Absence of abstractness in the anti-phishing recommendations and lack of advanced anti-phishing tools reduces users' self-efficacy [P42]. • Users do not receive training emails due to emails being in the spam folder [P28].	3
		Organizational factors	
D20	Organizational position	• Employees in a higher position in an organization are more vulnerable regardless of the phishing training or punishment [P40].	1
D21	Social influence	• People trust others' phishing stories as they perceive this information as trustworthy [P15]. • Observing others share information results in heightened levels of disclosure [P13]. • The motivation, self-efficacy, and cognitive ability of employees are impacted by the social relationships within and surrounding the organization [P26, P40].	4
D22	Norms	• Organization's procedural measures (e.g., security policies, standards and guidelines) have a beneficial effect on raising security consciousness [P38].	1

Research has shown that web developers tend to avoid adding phishing indicators to mobile browsers to save space for web content [P16]. Based on a research report by [P36], it is evident that anti-phishing educational games fail to tailor their content to the specific interests of individual user groups. To elaborate, the games designed for serious gamers do not meet the expectations of casual players, and vice versa. Consequently, designers of anti-phishing games ought to take into account the *type of gamers* as a factor when creating an educational game that can cater to the unique requirements of both casual and serious gamers.

According to multiple studies, the challenge faced by users with limited technical knowledge is attributable to their insufficient familiarity with security indicators and tools, as well as the complexity of the requirements of third-party tools [P1, P4, P5, P6, P7, P8, P9, P10, P11, P13, P14, P17, P20, P21, P28, P35, P39, P46, P47, P49]. This highlights the significance of incorporating the needs of novice users into the design process. Additionally, the effectiveness of users in detecting and preventing phishing attacks is reduced by a lack of resources, such as infrastructure support and advanced anti-phishing tools, and the absence of abstractness in anti-phishing recommendations [P6, P42].

3) Organizational factors: We identified several organizational factors from the literature that could be incorporated to enhance the effectiveness of the anti-phishing interventions. For example, according to research, *higher-positioned* employees in an organization are more susceptible to phishing attacks, regardless of their previous training or negative experience of being victims of phishing attacks [P40].

The literature shows that the implementation of security policies, standards, and guidelines in an organization is beneficial to increasing phishing awareness among employees. Additionally, the positive impact of organizational *norms* and *normative beliefs* on employees is observed in their exercise of greater caution, when opening potentially harmful phishing emails [P38]. Employees' motivation, self-efficacy, and cognitive ability are affected by social relationships [P26, P40]. For example, according to several studies [P13, P15], individuals tend to trust and share phishing stories based on the perceived trustworthiness of the information and the influence of social relationships within and around their organization.

B. Guidelines for the design, implementation and evaluation of anti-phishing interventions

We present our guidelines and the rationale underpinning each guideline in Table III. In the following paragraphs, we briefly discuss the guidelines personalized to different practitioner groups, intervention stages, intervention types, challenges and dominant factors in anti-phishing interventions.

• Among the 41 guidelines, 20 are mapped as relevant to our first user group (U1) consisting of designers and developers. This user group has the highest number of guideline relevant to them. These guidelines (G1 to G11, G13, G14, G16, G17, G19, G21 to G23, G27) include recommendations on interface design, placement of the phishing indicator, intervention content design, user engagement strategies, attention-drawing techniques, and enhancements for existing and future intervention designs.

• Our second user group (U2) consisting of *information security teams* of the organizations have a total of 19 guidelines (G12, G15, G16, G18, G20, G24, G25, G28, G29, G31 to G38, G40, G41) mapped to them as potentially applicable. These guidelines are intended to assist these cyber security professionals in conducting effective phishing education and training sessions, implementing measures to reinforce the organization's security, enhancing the speed and efficiency of phishing incident response and reporting procedures, as well as improving the educational resources made available to users to better cope with the threat of phishing.

• The guidelines G25 and G26 have been devised for *cyber security experts* (U3)), with a primary focus on conducting an investigation on the anti-phishing solution before adoption, as well as implementing protocols to ensure organizational adherence to a standardised email and anti-phishing webpage template.

• The guidelines G26, G27, G30, and G39 have been reported specifically for *C-suite employees* (U4) within organizations. While these guidelines are primarily intended for use by C-suite employees, certain guidelines, such as G28 and G29, are also relevant to the organization's security team (U2). These guidelines emphasize the importance of collaboration between the C-suite and the security team to develop policies that meet the needs of employees and provide better support for victims of phishing attacks.

• The user group U1 is mapped to guidelines for all three stages of phishing intervention, whereas the user group U2 are linked to guidelines solely for the implementation and evaluation stages. In contrast, user groups U3 (cyber security expert) and U4 (C-suite employee) have guidelines exclusively for the implementation stage of phishing intervention.

• Across the three stages of anti-phishing interventions, we provide 19 guidelines (G1 to G11, G13, G14, G16, G17, G19, G21, G22, and G27) for the design stage, 17 guidelines (G12, G15, G16, G18, G20, G23 to G28, G30 to G32, G35 to G37, and G39) for the implementation, and 8 guidelines (G13, G16, G29 to G33) for the evaluation of anti-phishing interventions. It is noteworthy that some guidelines are applicable to multiple different stages. For example, G16 is applicable to the design, implementation, and evaluation stages. We arrived at this recommendation based on several studies (e.g., P7, P13, P15, and P16) that suggest incorporating users' preferences in the design (e.g., the layout of anti-phishing intervention), implementation (e.g., training methods used to educate users), and evaluation (e.g., email templates used to assess users' phishing knowledge) of phishing interventions.

• The guidelines presented in this study have the potential to address several challenges in the design, implementation, and evaluation of existing anti-phishing interventions. For example, these guidelines can be particularly useful in addressing design constraints specific to anti-phishing warning user interface (e.g., G7, G8, G9), improving content-related issues for phishing education and training (e.g., G4, G5, G6), mitigating issues related to the deployment and adoption of anti-phishing technologies (e.g., G18 and G7), overcoming limitations associated with existing anti-phishing planning, policies, and guidelines (e.g., G12, G24, G25).

TABLE III: Guidelines for anti-phishing interventions

No	Guidelines	Rationale
G1	Remove deceptive user interface elements for unverified emails and incorporate an alert icon within the email client to indicate potentially fraudulent emails.	• Disabling misleading UI elements (e.g., profile photo, email history) for unverified sender addresses will reduce user confusion [P16]. • Placing a security indicator for unverified email delivered to the user acts as a forcing function for the sender domain to configure their SPF/DMRC/DKIM correctly [P7, P16].
G2	Clearly display the underlying URL of a suspicious link in the email client	• Clearly displaying the underlying URL of a suspicious link in the email client (link-focused warning) make it easier for users to notice where the links' actual destination [P25].
G3	Incorporate progressive disclosure in the design and add a learn more button.	• Progressive design and learn more buttons help to facilitate general advice, satisfy user curiosity, and support user investigations [P4, P5, P25, P51].
G4	Use visual examples and explanations and avoid technical jargon in the content.	• Avoiding technical details in the content can make them understandable to non-expert users [P1]. • Integrating visual examples and explanations on phishing cues presented helps users memorize and understand better [P42].
G5	Present abstract information and leverage situated learning in the content.	• Leveraging situated learning in the design can make the intervention interesting and engaging, and also improves learning outcomes [P5, P10, P19, P28, P34, P36, P37, P61, P62]. • Too much information in the content can be unappealing to inexperienced users [P1, P5, P13, P18, P41]. • Adopting situated learning is beneficial as learning science suggest that simply asking users to follow some advice would not be helpful [P5].
G6	Introduce varieties in the content and keep the information up to date.	• Including varieties in the content can help users tackle new and emerging phishing attacks [P19, P57, P58, P59, P61, P65].
G7	Minimize the functions and frequency of intervention users need to encounter.	• Limiting the frequency of the warnings reduce warning fatigue [P4]. • Minimum number of functionalities in the game can help finish the game easily, easy for users to remember when functionalities are less [P10].
G8	Design phishing warnings differently from standard warnings.	• Variation in the design increases the likelihood for users to read it, ensures they are taken seriously and prevent habituation [P1, P2, P14].
G9	Make the critical information easily accessible and visible to the users.	• To make users easily notice the warnings [P1, P4, P8, P25], increase warning adherence [P25] and to impose forced attention [P8, P25].
G10	Create uniform phishing indicators across different browsers and mobile interfaces.	• This will reduce the susceptibility of mobile device users [P16].
G11	Provide users clear choices and actionable items to proceed.	• Active interruption and actionable items minimize the user's workload, are naturally noticeable and users can use their time efficiently [P1, P2, P4, P5, P7, P20, P22, P24, P25 P41, P43, P44]
G12	Offer intervention immediately after users fall for phishing.	• Avoiding delay in displaying warnings minimizes users' confusion [P5]. The right timing of training intervention provides instant education [P2].
G13	Perform usability tests and collect user feedback.	• Collecting users' feedback from usability testing can improve future intervention design [P18, P22, P57, P61, P66, P67].
G14	Provide an explanation to the users on anti-phishing system reliability and decision-making and clarify users about the objective of the intervention.	• Feedback on the anti-phishing system increases users' trust [P7, P8, P11, P14, P33, P39, P43], helps users perceive potential danger [P20], increases user understanding and improves user ability to detect phishing [P18, P39]. • Making it clear to the users why they have displayed the intervention or not taken to the website to avoid their confusion [P5,P14].
G15	Use both technical and human-centric defence mechanisms to cope with phishing.	• Prevent user's over-reliance on technology, provide additional defence in detecting unpredictable, highly dynamic, and increasingly sophisticated phishing attacks [P3, P5, P12, P17, P18, P26, P27, P28, P38, P41, P51, P53, P57, P58, P59]. • Educating users about the security properties of different interventions remove their misunderstanding that leads to mistake [P14]. • Training all individual who has access to the organization increase the organization's robustness [P53]. • Human-centric defence mechanisms organized by C-suit employees can help low-level employees in the organization to learn about phishing [P21, P38, P40, P56, P57, P59, P61, P67, P68, P69].
G16	Personalize the intervention style and medium based on the target user's demographic.	• Personalized phishing training can take into account user's preferences (e.g., individual preferred training method [P15, P21], content relevant to the organization [P16, P58], roles and responsibilities [P40, P53, P58, P60], age [P21, P35]) to ensure users receive targeted education and training [P7, P13, P15, P16, P21, P26, P35, P36, P40, P48, P52, P53, P57, P58, P59, P60, P61, P62, P64, P66, P67].
G17	Consider the decision-making process and vulnerabilities of humans in the design.	• Taking into account the vulnerabilities and decision-making processes of the user (e.g., users' misconceptions and perspectives [P11], perceived threat [P9]) increases the effectiveness of anti-phishing interventions for end users and assist to develop the tailored approach [P4, P6, P7, P9, P11, P18, P24].
G18	Configure IT system for phishing training.	• Preparing IT system to avoid simulated email being filtered by technical filters helps users being missed for training [P69]. • Verifying if inventory management software is utilizing scanning, analysis, or probing techniques help detect abnormally high levels of external IP addresses [P54].
G19	Design visually distinct user-friendly URL bar.	• Noticeable and consistent URL bar helps users differentiate legitimate and malicious domains easily [P2, P8, P46].
G20	Use automated platforms and improved tools for phishing training, incident management and reporting.	• Automated approaches help to better support managing complex situations, delivering personalized content and threat identification [P61, P63, P67, P50].

Continued on next page

Guidelines for anti-phishing interventions – continued from previous page

No	Guidelines	Rationale
G21	Disable JavaScript on login forms when a form element is in focus.	• Deactivating JavaScript on webpages every time the focus is put on a form element prevents the attacker from capturing the keystrokes or initiating timing attacks [P16, P22, P23].
G22	Explain the capabilities and effectiveness of the deployed anti-phishing solution clearly to the users.	• Reliable trust signals to the users can prevent over-trust and over-reliance on the deployed anti-phishing solutions [P11]. • Utilizing interactive error messages to elucidate the purpose of a website can deter users from engaging in destructive actions [P43, P44].
G23	Use email authentication protocols to encrypt emails and filter out incoming malicious emails.	• To achieve better resiliency [P18,P51] and to make more informed decision [P16, P27] on the incoming emails.
G24	Send pre-notification to the users before conducting phishing training, however, perform random phishing training.	• Sending pre-notification to the participants prevents discomfort [P30, P69]. • Emphasising on the anonymity of phishing training can reduce the effect of prairie dogging and estimate of organization's likelihood to fall victim to phishing [P59, P61, P62, P69].
G25	Conduct prior investigation before adopting anti-phishing tools, identify most vulnerable group and determine priority topics.	• Perform prior research and analyze the reviews on tool vendors to select the right tool [P26, P61] • Identifying vulnerable users can help reduce training time and efforts [P26]. • Teaching everything or huge amount of information can cause security fatigue [P13].
G26	Follow a consistent template for organizational emails and create a standard template for anti-phishing webpages.	• A consistent email structure helps employees to notice the discrepancies in phishing emails easily [P41]. • A standardized template for anti-phishing webpages reduces inconsistency helps avoid confusion and helps web-designer implement their anti-phishing tools easily [P42].
G27	Introduce a user-friendly, built-in phishing reporting tool within the client system. Develop a formal procedure to handle phishing reports.	• Having a formal procedure placed makes it convenient to handle phishing reports [P50]. • An in-client phishing incident reporting tool makes phishing reporting easier [P58, P63].
G28	Get employees' feedback to modify the organization's policy.	• Obtain staff's feedback after phishing simulation to modify the organization policy accordingly to meet staff's needs [P50].
G29	Deploy help-desk and victim support for users.	• Deploying post simulation help desk support allows further users' investigations [P51]. • Deploying help-desk support can assist external users in determining the authenticity of an email sent from the organization [P51]. • Add a victim support option in the anti-phishing webpages can help users to fix potential problems [P42].
G30	Create a structured policy and documentation. Regularly assess and manage phishing awareness efforts.	• Appropriate policy and documentation ensure that all the employees adapt themselves to security countermeasures and requirements [P26, P38, P60]. • Continuous measurement, improved management and policy making helps to achieve improved phishing defence [P11, P38, P40, P50, P53, P54, P57, P67].
G31	Conduct phishing simulation with embedded training.	• Assist the organization's security team in practicing the handling and response to simulated phishing incidents to enhance preparedness for real phishing attacks [P53, P56, P57, P60, P61]. • Embedding learning content with phishing simulation provides education on demand [P5, P7, P12, P27, P53, P56, P57, P58, P59, P60, P61, P67, P68, P69].
G32	Conduct phishing simulation that adheres to the guidelines of the data privacy policy appropriate to the region.	• Data privacy policy-compliant phishing training protects participants sensitive information, hence reducing data breaches [P26, P69].
G33	Provide users immediate feedback on their performance.	• Users feel motivated if instant corrective feedback is provided after testing and evaluating their phishing knowledge in their regular environment [P7, P10, P31].
G34	Use realistic and equally difficult training emails. Use challenging questions to test phishing knowledge.	• Realistic and equally difficult email helps to test the persistence of the training's effect [P7]. • An extensive test with challenging questions reduce repetitive training costs and can help avoid the ceiling effect [P21].
G35	Implement progressive and self-adaptive phishing training.	• Dynamic and self-adaptive phishing training improve user sensitivity to deception cues [P24, P63, P64, P66].
G36	Adopt video and interactive education and training materials.	• Video and interactive training are more effective as users do not need refreshment very quickly [P5, P11, P19, P34, P36]
G37	Utilize the expertise of external service providers to aid in phishing knowledge assessment and awareness material development.	• Leveraging external service providers can support better phishing knowledge assessment and awareness material development [P54, P60].
G38	Choose evaluation metrics and baselines that are useful and relevant.	• Click-through rate should be normalized based on the persuasiveness of the training template to produce a sound analysis and evaluation [P32, P54, P56, P58, P59, P60, P61, P68].
G39	Train users how to report phishing and reward secure behaviour.	• Training users on how to report phishing incidents and explaining the benefits of reporting can help to establish a phishing reporting culture [P26, P50, P58, P60, P69]. • Rewarding employees for their secure behaviour can motivate and encourage them to perform better [P30, P61, P66].
G40	Conduct multiple cycles of follow-up training.	• Help to assess users' short-term and long-term knowledge retention after training [P26, P31, P52, P54, P57, P58]. • Repetitive training in a short period helps users learn a second time if they had difficulty understanding in the first time [P5, P7, P24, P27, P34, P53, P56, P57, P62, P67, P68, P69]. • Follow-up training (for children) to counter knowledge decay of the ability to identify phishing [P21].
G41	Avoid frequent reminders and over-training and keep the reminders short and simple.	• Avoiding frequent risk notifications and over-training reminders can reduce training fatigue [P34, P52, P53, P58, P60, P61, P62, P69]. • Including a lower bound of information in the reminder measures can reduce security fatigue [P34].

• Among our 3 main intervention types (i.e., education, training and awareness), our analysis has yielded a set of 9 guidelines for education interventions (G5, G7, G16, G17, G25, G26, G29, G33, G36), 27 guidelines for training interventions (G5, G6, G7, G11 to G16, G18, G20, G24, G25, G27, G28, G30, G31 to G39), and 19 guidelines for awareness interventions (G1 to G5, G7 to G11, G13 to G17, G19, G21 to G23).

• Some of our challenges (i.e., "Ch5 - performance limitations of anti-phishing tools", "Ch18 - insufficient usability and effectiveness evaluation of phishing interventions", "Ch19 - lack of sophisticated quantification of phishing training outcome") do not exhibit any discernible dominant factors. As a result, the guidelines G13 and G38, which are intended to address these challenges, are not linked to any particular dominant factor.

IV. Threat to validity

The guidelines we compiled may not be *comprehensive* because potential design, implementation and evaluation principles are not always explicitly articulated [40]. The presented guidelines have been formulated with specific regard to the collected study context being investigated, and therefore, their *generalizability* may be limited. As a means of substantiating these standpoints, we posit that the 69 studies used in this research were collected through a rigorous process of quality assessment. Most of the guidelines that we have formulated are supported by more than one study in the literature, which underscores their applicability in contexts that are different from the collected study context. This is because these studies have involved diverse user types, varying sample sizes, different intervention types, and other variables. Additionally, our analysis encompassed industry reports [e.g., P58] and case studies with various organizations [e.g., P63, P64, P65]. The inclusion of grey studies facilitates the mitigation of bias that stems from a proclivity to publish studies that report favorable results exclusively [55].

Regarding the *representativeness* of the data used in this study, the collection of textual data is restricted to 53 academic and 16 grey studies which have been identified in literature searches. We recognize that future research could broaden the scope of the searches and analysis by including additional data sources such as interviews and surveys [56]. In order to validate and strengthen the usability and effectiveness of our guidelines, we plan to conduct a semi-structured interview study with developers and cyber security practitioners in our future work.

As multiple researchers were involved in this study, in order to minimise *researcher bias*, various activities (e.g., study selection, data search, extraction, analysis, and synthesis) were conducted in accordance with a well-defined research protocol, following the established guidelines proposed by Kitchenham and Charters [30] and Garousi et al. [31]. The research protocol was modified and updated by conducting a pilot study of randomly selected 10 studies. The first author collected 90% of the data (62 out of 69), while the third author extracted the remaining 10% (7 out of 69). All data were shared in a collaborative folder and cross-checked by each author and any issues or disagreements were resolved in weekly research meetings among the authors.

While employing thematic analysis permits the data analysis to be grounded in the textual data collected from academic and grey literature studies, there is a threat of *subjectivity* of the data analysis [57]. To alleviate this threat, we discussed the issues and concerns of the emergent findings throughout the study in the weekly meetings. Throughout the iterative and intertwined rounds of data collection and analysis, the first author led the data analysis with support from other researchers who acted as validators at each stage.

V. Conclusion

Current anti-phishing interventions encounter several obstacles, such as poor user interface design, lack of engaging and interesting content, incomplete or outdated anti-phishing instructions, flawed anti-phishing training implementation, and deficient anti-phishing policies within organizations. The usability issues that arise from the current one-size-fits-all anti-phishing interventions can be attributed to a need for greater awareness among developers and practitioners of end-user requirements and preferences. There is a current lack of available personalized guidelines to assist developers and practitioners for this purpose.

To address this gap, in this study, we first identified 22 dominant factors, consisting of 15 individual, 4 technical, and 3 organizational factors that impact the effectiveness and outcomes of anti-phishing interventions from 53 academic and 16 grey literature studies. We then present 41 guidelines to aid developers and practitioners in addressing these issues within current anti-phishing intervention design, implementation, and evaluation. Our guidelines are for four distinct practitioner groups: designers/developers, information security teams, cyber security experts, and C-suite employees. We offered guidelines for 14 different types of interventions within phishing education, training, and awareness, and for overcoming 19 different challenges that may arise in such interventions. Our personalized guidelines aim to improve the effectiveness of current anti-phishing software development, deployment, and assessment practices. By reporting these guidelines to address the needs of anti-phishing practitioners, we aim to contribute to the ongoing efforts to mitigate the threat of phishing attacks. In the future work, we aim to investigate the difference

Acknowledgment

The work has been supported by the Cyber Security Research Centre Limited whose activities are partially funded by the Australian Government's Cooperative Research Centres Programme.

References

[1] A. Jenkins, N. Kokciyan, and K. E. Vaniea, "Phished: Automated contextual feedback for reported phishing," in *18th Symposium on Usable Privacy and Security*, Usenix, 2022.

[2] M. M. Alani and H. Tawfik, “Phishnot: A cloud-based machine-learning approach to phishing url detection,” *Computer Networks*, p. 109407, 2022.

[3] R. Wash, N. Nthala, and E. Rader, “Knowledge and capabilities that non-expert users bring to phishing detection,” in *Seventeenth Symposium on Usable Privacy and Security (SOUPS 2021)*, pp. 377–396, 2021.

[4] APWG, “Keeping up with phishing.” https://apwg.org/trendsreports/, 2022.

[5] B. Kaiser, J. Wei, E. Lucherini, K. Lee, J. N. Matias, and J. Mayer, “Adapting security warnings to counter online disinformation,” in *30th USENIX Security Symposium (USENIX Security 21)*, pp. 1163–1180, 2021.

[6] N. A. G. Arachchilage, S. Love, and K. Beznosov, “Phishing threat avoidance behaviour: An empirical investigation,” *Computers in Human Behavior*, vol. 60, pp. 185–197, 2016.

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

[8] P. Kumaraguru, J. Cranshaw, A. Acquisti, L. Cranor, J. Hong, M. A. Blair, and T. Pham, “School of phish: a real-world evaluation of anti-phishing training,” in *Proceedings of the 5th Symposium on Usable Privacy and Security*, pp. 1–12, 2009.

[9] A. Franz, V. Zimmermann, G. Albrecht, K. Hartwig, C. Reuter, A. Benlian, and J. Vogt, “Sok: Still plenty of phish in the sea—a taxonomy of user-oriented phishing interventions and avenues for future research,” in *Seventeenth Symposium on Usable Privacy and Security (SOUPS 2021)*, pp. 339–358, 2021.

[10] L. Li, E. Berki, M. Helenius, and S. Ovaska, “Towards a contingency approach with whitelist-and blacklist-based anti-phishing applications: what do usability tests indicate?,” *Behaviour & Information Technology*, vol. 33, no. 11, pp. 1136–1147, 2014.

[11] S. Egelman, L. F. Cranor, and J. Hong, “You’ve been warned: an empirical study of the effectiveness of web browser phishing warnings,” in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 1065–1074, 2008.

[12] M. Wu, R. C. Miller, and S. L. Garfinkel, “Do security toolbars actually prevent phishing attacks?,” in *Proceedings of the SIGCHI conference on Human Factors in computing systems*, pp. 601–610, 2006.

[13] M. Alsharnouby, F. Alaca, and S. Chiasson, “Why phishing still works: User strategies for combating phishing attacks,” *International Journal of Human-Computer Studies*, vol. 82, pp. 69–82, 2015.

[14] C. Reuter, L. L. Iacono, and A. Benlian, “A quarter century of usable security and privacy research: transparency, tailorability, and the road ahead,” 2022.

[15] F. B. Salamah, M. A. Palomino, M. Papadaki, and S. Furnell, “The importance of the job role in social media cybersecurity training,” in *2022 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW)*, pp. 454–462, IEEE, 2022.

[16] J.-W. Bullee and M. Junger, “How effective are social engineering interventions? a meta-analysis,” *Information & Computer Security*, 2020.

[17] H. Hu and G. Wang, “End-to-end measurements of email spoofing attacks.,” in *USENIX Security Symposium*, pp. 1095–1112, 2018.

[18] K. Althobaiti, A. D. Jenkins, and K. Vaniea, “A case study of phishing incident response in an educational organization,” *Proceedings of the ACM on Human-Computer Interaction*, vol. 5, no. CSCW2, pp. 1–32, 2021.

[19] D. D. Caputo, S. L. Pfleeger, J. D. Freeman, and M. E. Johnson, “Going spear phishing: Exploring embedded training and awareness,” *IEEE Security & Privacy*, vol. 12, no. 1, pp. 28–38, 2013.

[20] A. Burns, M. E. Johnson, and D. D. Caputo, “Spear phishing in a barrel: Insights from a targeted phishing campaign,” *Journal of Organizational Computing and Electronic Commerce*, vol. 29, no. 1, pp. 24–39, 2019.

[21] G. CJ, S. Pandit, S. Vaddepalli, H. Tupsamudre, V. Banahatti, and S. Lodha, “Phishy-a serious game to train enterprise users on phishing awareness,” in *Proceedings of the 2018 annual symposium on computer-human interaction in play companion extended abstracts*, pp. 169–181, 2018.

[22] Z. A. Wen, Z. Lin, R. Chen, and E. Andersen, “What. hack: engaging anti-phishing training through a role-playing phishing simulation game,” in *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, pp. 1–12, 2019.

[23] C. Marforio, R. Jayaram Masti, C. Soriente, K. Kostiainen, and S. Čapkun, “Evaluation of personalized security indicators as an anti-phishing mechanism for smartphone applications,” in *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, pp. 540–551, 2016.

[24] K. Althobaiti, N. Meng, and K. Vaniea, “I don’t need an expert! making url phishing features human comprehensible,” in *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, pp. 1–17, 2021.

[25] L. Bauer, C. Bravo-Lillo, L. F. Cranor, and E. Fragkaki, “Warning design guidelines (cmu-cylab-13-002),” 2013.

[26] L. A. Shepherd and K. Renaud, “How to design browser security and privacy alerts,” *arXiv preprint arXiv:1806.05426*, 2018.

[27] M. Gáliková, V. Švábenskỳ, and J. Vykopal, “Toward guidelines for designing cybersecurity serious games,” in *Proceedings of the 52nd ACM Technical Symposium on Computer Science Education*, pp. 1275–1275, 2021.

[28] P. Jaferian, D. Botta, F. Raja, K. Hawkey, and K. Beznosov, “Guidelines for designing it security management tools,” in *Proceedings of the 2nd ACM Symposium on Computer Human interaction For Management of information Technology*, pp. 1–10, 2008.

[29] S. Chiasson, P. van Oorschot, and R. Biddle, “Even experts deserve usable security: Design guidelines for security management systems,” in *SOUPS Workshop on Usable IT Security Management (USM)*, pp. 1–4, 2007.

[30] B. Kitchenham and S. Charters, “Guidelines for performing systematic literature reviews in software engineering,” *Technical report, EBSE Technical Report EBSE-2007-01*, 2007.

[31] V. Garousi, M. Felderer, and M. V. Mäntylä, “Guidelines for including grey literature and conducting multivocal literature reviews in software engineering,” *Information and Software Technology*, vol. 106, pp. 101–121, 2019.

[32] “Scopus.” https://www.scopus.com/.

[33] M. Zahedi, M. Shahin, and M. A. Babar, “A systematic review of knowledge sharing challenges and practices in global software development,” *International Journal of Information Management*, vol. 36, no. 6, pp. 995–1019, 2016.

[34] M. Shahin, M. A. Babar, and M. A. Chauhan, “Architectural design space for modelling and simulation as a service: a review,” *Journal of Systems and Software*, vol. 170, p. 110752, 2020.

[35] G. Desolda, L. S. Ferro, A. Marrella, T. Catarci, and M. F. Costabile, “Human factors in phishing attacks: A systematic literature review,” *ACM Computing Surveys (CSUR)*, vol. 54, no. 8, pp. 1–35, 2021.

[36] S. Baki and R. Verma, “Sixteen years of phishing user studies: What have we learned?,” *arXiv preprint arXiv:2109.04661*, 2021.

[37] “Core.” https://www.core.edu.au/.

[38] “Core conference rankings 2021: Process followed and data considered.” https://drive.google.com/file/d/1bKa40nheaQ3zfuXu3jSpKIw5TnhK9USR/view.

[39] K.-P. Yee and K. Sitaker, “Passpet: convenient password management and phishing protection,” in *Proceedings of the second symposium on Usable privacy and security*, pp. 32–43, 2006.

[40] D. Jampen, G. Gür, T. Sutter, and B. Tellenbach, “Don’t click: towards an effective anti-phishing training. a comparative literature review,” *Human-centric Computing and Information Sciences*, vol. 10, no. 1, pp. 1–41, 2020.

[41] V. Garousi, M. Felderer, and T. Hacaloğlu, “Software test maturity assessment and test process improvement: A multivocal literature review,” *Information and Software Technology*, vol. 85, pp. 16–42, 2017.

[42] C. Islam, M. A. Babar, and S. Nepal, “A multi-vocal review of security orchestration,” *ACM Computing Surveys (CSUR)*, vol. 52, no. 2, pp. 1–45, 2019.

[43] B.-J. Butijn, D. A. Tamburri, and W.-J. v. d. Heuvel, “Blockchains: a systematic multivocal literature review,” *ACM Computing Surveys (CSUR)*, vol. 53, no. 3, pp. 1–37, 2020.

[44] G. Cascavilla, D. A. Tamburri, and W.-J. Van Den Heuvel, “Cybercrime threat intelligence: A systematic multi-vocal literature review,” *Computers & Security*, vol. 105, p. 102258, 2021.

[45] “Supplementary material for this study.” https://figshare.com/s/f8f2a125177c414b3529.

[46] V. Braun and V. Clarke, “Using thematic analysis in psychology,” *Qualitative research in psychology*, vol. 3, no. 2, pp. 77–101, 2006.

[47] A. Sbaraini, S. M. Carter, R. W. Evans, and A. Blinkhorn, “How to do a grounded theory study: a worked example of a study of dental practices,” *BMC medical research methodology*, vol. 11, no. 1, pp. 1–10, 2011.

[48] U. Bhatt, A. Xiang, S. Sharma, A. Weller, A. Taly, Y. Jia, J. Ghosh, R. Puri, J. M. Moura, and P. Eckersley, “Explainable machine learning in deployment,” in *Proceedings of the 2020 conference on fairness, accountability, and transparency*, pp. 648–657, 2020.

[49] S. R. Hong, J. Hullman, and E. Bertini, “Human factors in model interpretability: Industry practices, challenges, and needs,” *Proceedings of the ACM on Human-Computer Interaction*, vol. 4, no. CSCW1, pp. 1–26, 2020.

[50] R. Tomsett, D. Braines, D. Harborne, A. Preece, and S. Chakraborty, “Interpretable to whom? a role-based model for analyzing interpretable machine learning systems,” *arXiv preprint arXiv:1806.07552*, 2018.

[51] G. Dupont, “The dirty dozen errors in maintenance,” in *The 11th symposium on human factors in maintenance and inspection: Human error in aviation maintenance*, 1997.

[52] D. Hidellaarachchi, J. Grundy, R. Hoda, and I. Mueller, “The influence of human aspects on requirements engineering-related activities: Software practitioners’ perspective,” *ACM Transactions on Software Engineering and Methodology*, 2022.

[53] J. K. Nwankpa and P. M. Datta, “Remote vigilance: The roles of cyber awareness and cybersecurity policies among remote workers,” *Computers & Security*, vol. 130, p. 103266, 2023.

[54] K. M. Alnifie and C. Kim, “Appraising the manifestation of optimism bias and its impact on human perception of cyber security: A meta analysis,” *Journal of Information Security*, vol. 14, no. 2, pp. 93–110, 2023.

[55] F. Kamei, G. Pinto, I. Wiese, M. Ribeiro, and S. Soares, “What evidence we would miss if we do not use grey literature?,” in *Proceedings of the 15th ACM/IEEE International Symposium on Empirical Software Engineering and Measurement (ESEM)*, pp. 1–11, 2021.

[56] S. Hermawati and G. Lawson, “Establishing usability heuristics for heuristics evaluation in a specific domain: Is there a consensus?,” *Applied ergonomics*, vol. 56, pp. 34–51, 2016.

[57] R. N. Rajapakse, M. Zahedi, and M. A. Babar, “An empirical analysis of practitioners’ perspectives on security tool integration into devops,” in *Proceedings of the 15th ACM/IEEE International Symposium on Empirical Software Engineering and Measurement (ESEM)*, pp. 1–12, 2021.