Immagine che contiene appendiabiti, design

Descrizione generata automaticamente con attendibilità media

**Deliverable D1.**

**Literature review**

This deliverable reports on the study of the literature review on existing solutions for human vulnerability assessment and mitigation. The output of the review is a user model for defining the sensible human factors, the best approaches to profile users’ vulnerabilities and train users against cyber-attacks, as well as the existing no-code approaches.

# Literature review on existing solutions for human vulnerability assessment and mitigation

## Introduction

In today's digital society, organizations in all sectors face increasingly complex cybersecurity challenges. Increasing dependence on digital infrastructure, the spread of connected devices, and the widespread use of online applications and services have heightened the importance of protecting sensitive information and critical systems from potential cyber threats.

Although efforts to strengthen the security of information systems are significant, a crucial aspect often overlooked emerges: the human factor. Cyber-attacks can exploit technical vulnerabilities in systems, but attackers often aim to exploit human weaknesses to gain unauthorized access or compromise information security.

These vulnerabilities can result from several factors, including:

* Manipulation, such as deception or coercion of authorized employees or users.
* Negligence, such as failure to take basic safety measures.

Users, by their actions, represent a major source of vulnerabilities in digital security. For example, according to a Verizon study, 74% of cybersecurity breaches are caused by human error, which includes social engineering attacks, mistakes or misuse [1]. To effectively address cybersecurity challenges, it is essential to understand how humans interact with the digital environment thoroughly. It is necessary to identify possible weaknesses and assess how these can be exploited by malicious attackers. This is crucial to developing effective strategies and countermeasures to protect organizations' information and digital assets.

To gain an in-depth understanding of methodologies for assessing risk to digital security caused by humans, the Systematic Literature Review (SLR) was adopted. This methodology provides a comprehensive and objective picture of existing practices, as it offers a systematic approach that allows for a thorough and unbiased analysis of the evidence in the literature.

SLR is particularly well suited to synthesizing a wide range of studies, providing a consolidated analysis of the different methodologies used to assess human vulnerabilities. This approach overcomes possible limitations of individual studies, helping to outline a completer and more accurate overview of current practices.

The use of SLR aims to offer a thorough and comprehensive view of the methodologies employed in the field of digital security, thus providing a solid foundation for understanding the current landscape and future directions of development.

## Planning and conducting the SLR

A Systematic Literature Review (SLR) was conducted using a reproducible and in-depth approach, with the main objective of systematically analyzing the methodologies currently employed by companies to assess digital security risk caused by humans. According to the methodology proposed by Kitchenham, an SLR requires three phases: planning, conducting, and reporting[2] This section will outline the first two phases of SLR in detail, while the next section will delve into the final phase, which is the reporting of the results obtained.

### Planning for SLR

SLR planning includes the following activities[2]: 1) formulation of the research question; 2) definition of search strings; 3) selection of data sources; and 4) definition of inclusion criteria. Below, we give the details of each activity.

#### Formulation of the research question

The research question formulated for this study focuses on investigating the current state of research regarding the assessment of human-caused vulnerabilities in digital security. The research question is as follows, "What state-of-the-art solutions exist to assess human-caused vulnerabilities in digital security?" The answer to this question will provide an in-depth understanding of how the literature addresses this problem.

#### Defining Search Strings

This section defines the search strings used for the literature review. The strings were formulated by taking the key concept of "human vulnerability assessment" as a starting point and linking it to the field of "cyber security." In order to ensure comprehensive coverage of the literature, it was decided to also include linkage to human behavior, particularly user behavior, as well as concepts related to cyber security attacks such as "social engineering" (including, for more detail, the concept of "phishing") and "insider threat." The resulting search strings used to query search engines were:

*human vulnerability assessment cybersecurity*

*("human behavior" OR "human behavior") cybersecurity*

*social engineering awareness assessment*

*user behavior cybersecurity policies*

*unintentional insider threat assessment*

*phishing awareness assessment*

#### Definition of inclusion criteria

This step concerns the final selection of relevant publications based on 3 inclusion criteria:

* Peer reviewed, meaning the article is the result of an expert review process.
* Written in English.
* Focused on a methodology for assessing risk caused by human vulnerability in the digital environment. The relevance of each publication is assessed by analyzing the title, abstract and, if necessary, the introduction.

### Conduct the SLR

After the initial planning phase, the literature review is carried out. Following Kitchenham's guidance, two main activities were carried out: performing the literature review and data synthesis [2], which are described in the following sections.

Immagine che contiene testo, schermata, diagramma, Parallelo

Descrizione generata automaticamente

### Performing the literature review

This activity follows the process shown in Figure 1, which consists mainly of two steps:

Step 1-Searching the digital library: a search was conducted in the Google Scholar digital library using the search strings described in Section 2.1;

Phase 2 - Backward and Forward Snowballing: references and citations of publications resulting from the previous phase were checked, as well as publications that cited publications from Phase 1 [3].

The initial search of the digital library yielded a total of twenty-one potentially relevant publications. After checking for duplicates, a dataset of eighteen publications was obtained. In the end, this phase resulted in a total of eighteen publications. Each publication was then analyzed by examining the abstract, introduction, and conclusion, considering the other inclusion criteria defined in the previous section. After this step, eight publications were excluded, resulting in a final dataset of ten publications. Phase two allowed additional publications to be retrieved, bringing the final set to twelve publications

Immagine che contiene schermata, linea, Diagramma, Rettangolo

Descrizione generata automaticamente

Figure 1. Distribution of publications by year.

### Summary of data

The twelve publications resulting from the research phases are listed in Table 1. The distribution of selected publications according to their year of publication is shown in Figure 2.

## Analysis and description of the SLR results

In this section, the results of the literature review conducted to answer the research question proposed in Section 2.1 are presented. Based on an in-depth analysis of selected publications, a dimension was identified that characterizes methodologies for assessing risk to digital security caused by humans. This dimension provides an overview of the state of the art and can be used as a starting point for designing new methodologies. Table 1 summarizes the dimension and the positioning of publications in this dimension.

|  |  |  |
| --- | --- | --- |
| **Reference** | **Publication Title** | **Methodology of collection of the data** |
| [4] | Leveraging human factors in cybersecurity: an integrated methodological approach | psychometric questionnaire |
| [5] | An integrated cybernetic awareness strategy to assess cybersecurity attitudes and behaviors in school context | psychometric questionnaire |
| [6] | Measuring the human factor of cyber  security | simulation of phishing attacks |
| [7] | An Instrument to Measure Human Behavior Toward Cyber Security.  Policies | psychometric questionnaire |
| [8] | Design of Security Training System for  Individual Users | simulation of phishing attacks |
| [9] | Developing a simulated intelligent instrument to measure user behavior  toward cybersecurity policies | psychometric questionnaire + gamification |
| [10] | Measuring the Information Security Awareness Level of Government Employees Through Phishing  Assessment | Psychometric questionnaire + phishing attack simulation |
| [11] | Assessment of Employee Susceptibility to Phishing Attacks at U.S. Health Care  Institutions | simulation of phishing attacks |
| [12] | CyberPhishing: A Game-Based  Platform for Phishing Awareness Testing | simulation of phishing attacks  + gamification |
| [13] | Assessment of End-User Susceptibility to Cybersecurity Threats in Saudi  Arabia by Simulating Phishing Attacks | simulation of phishing attacks |
| [14] | Determining employee awareness using the Human Aspects of Information Security Questionnaire (HAIS-Q) | psychometric questionnaire |
| [15] | The Human Aspects of Information Security Questionnaire (HAIS-Q): Two further validation studies. | psychometric questionnaire |

The resulting dimension is described in the following paragraphs. For the dimension, the classification of publications will be discussed.

### Dimension: Data collection methodology.

Within the analyzed publications, three distinct approaches to data collection regarding human behaviors in digital environments emerged. In the following paragraphs, a detailed exposition of the techniques will be made, highlighting the peculiarities of each and providing a clear and comprehensive overview of the methodological process adopted. This insight will enable the complexity and richness of the s trategies used to acquire crucial information regarding the assessment of risk caused by humans in digital contexts.

**Psychometric questionnaires**. The use of psychometric questionnaires emerges as a quantitative approach of significant relevance to assess users' adherence to security policies in digital environments.

In the context of the reviewed publications [4][10], one particular survey instrument that has emerged is the Human Aspects of Information Security Questionnaire (HAIS- Q) by Parson et al.[14]. This questionnaire is based on the KAB model, a behavioral theory that connects three key elements: knowledge (Knowledge), attitude (Attitude) and behavior (Behavior). The KAB model assumes that as knowledge on a given topic increases, an individual will develop a more positive attitude and be more likely to translate that attitude into consistent behaviors.

In the context of digital security, the goal of applying the KAB model is to identify factors that influence employees' information security behavior. Specifically, HAIS-Q leverages the KAB model to examine the relationship between employees' policy knowledge, attitudes and behaviors, thus providing a comprehensive framework for targeted and effective interventions in corporate information security.

The HAIS-Q provides an opportunity to explore the KAB model through seven user scenarios, which we can define as "areas of interest" (AI), representing specific contexts in which users interact with products or systems. These areas of interest cover (AI1) password management, (AI2) e-mail usage, (AI3) Internet usage, (AI4) mobile computing, (AI5) social networking, (AI6) incident reporting, and (AI7) information management.

The questionnaire has a total of 63 items, divided into three sets of items corresponding to the KAB components. The first set of 21 items measures the "Knowledge" of employees, the second with 21 items deals with "Attitude," and the third with 21 items deals with employee-reported "Behavior." Each of the 21 items within each group covers the 7 "areas of interest," with 3 topics for each, thus enabling the detailed analysis of each area of interest through the KAB components. This approach results in the investigation of each "area of interest" through three different perspectives, "Knowledge," "Attitude," and "Behavior.".

The HAIS-Q questionnaire adopts the Likert scale as the primary tool for collecting participants' responses. Specifically, the Likert scale requires participants to assign a score from 1 to 5 to each item, where 1 represents "strongly agree" and 5 represents "strongly disagree," based on the degree of agreement or disagreement with the proposed statement. This statement can have a favorable or unfavorable valence with respect to the topic being evaluated. In this context, it is important to note that 32 of the 63 questions are worded with a negative connotation, identified in Figures 3 and 4 by the use of the circumflex accent (^) at the end of the item description.

In order to ensure consistency in data analysis using the Likert scale, the scores associated with the unfavorably worded items are reversed before summation ensuring consistent alignment of scores, thus simplifying comparison and overall understanding of the responses provided by participants.

The responses to the questionnaire were thoroughly analyzed using three different research methodologies:

HAIS-Q Reliability Analysis using Cronbach's Alpha [4][14]: this method was used to assess the internal consistency of the HAIS-Q questionnaire by examining how faithfully the questionnaire items measured the same construct. A higher value of Cronbach's alpha indicates greater internal consistency among items, reinforcing the validity of the questionnaire.

Descriptive Analysis (Mean, Standard Deviation and Median) of the Sample on the Components of the KAB Model in the Different Areas of Interest[4]: measures of central tendency such as the mean and median provide an indication of the typical value of responses, while the standard deviation measures the dispersion of the data around this central tendency. This analysis provides a clear and understandable overview of the characteristics of the sample.

Repeated-Measures Analysis of Variance with General Linear Model [4]: repeated-measures analysis of variance compares the averages of responses across AIs to determine whether there are statistically significant differences. The general linear model is used to handle various interactions between variables. This analysis helps to specifically identify which Areas of Interest show significant differences, thus contributing to a detailed understanding of the dynamics in the sample.

The implementation of these three methodologies contributed to a comprehensive and rigorous evaluation of the questionnaire responses, providing a solid basis for interpreting the results and trends that emerged in the research conducted.

In the reviewed publications, a methodology designed to assess attitudes and risk behaviors related to cybersecurity in the school setting was identified [5]. This integrated approach has three basic steps:

(1) the assessment of attitudes and behaviors through the use of two questionnaires, (2) the creation of a self-diagnosis application to assess students' level of knowledge about cybersecurity, and (3) the implementation of a lesson plan to raise students' awareness of cybersecurity-related attitudes and behaviors.

Focusing on the first phase, two questionnaires were developed, named "School Cybersecurity Behaviors (CsB-S)" and "School Cybersecurity Attitudes (CsA-S). The final version of these questionnaires were evaluated by experts in cybersecurity and cyber awareness, thus ensuring their relevance and validity in the school context.

Attitudes were assessed using the CsA-S questionnaire, consisting of twenty-five questions with responses rated on a Likert scale of 1 to 4 points, where: 1 = Totally agree, 2 = Agree,3 = Disagree, 4 = Totally disagree.

Immagine che contiene testo, carta, Carattere, lettera

Descrizione generata automaticamente

Figure 2. Cybersecurity Attitudes in School (CsA-S)

Similarly, the CsB-S questionnaire assessed behaviors through twenty questions with responses rated on a seven-point Likert scale, where 1 = Never,7 = Everyday.

Immagine che contiene testo, Carattere, schermata, numero

Descrizione generata automaticamente

Figure 3. Cybersecurity Behaviors in School (CsB-S).

In the course of this research, statistical analysis of the collected data was conducted through the use of IBM SPSS software, a recognized application for statistical analysis. In order to examine differences in scores between specific demographic groups, targeted statistical tests were employed. In particular, the use of t-tests was adopted to compare score averages between groups of different genders and school grades. Contextually, nonparametric Mann-Whitney tests were employed to assess differences in questionnaire scores across grades.

In order to ensure internal consistency of the measurement scales in the questionnaires, analysis was conducted using Cronbach's alpha, following an approach similar to that taken by Parsons et al. in the HAIS-Q previously described[14].

An alternative approach, focusing on human behavior and its crucial role in information security, is the questionnaire developed by Alissa et al. [7]. This instrument aims to assess human behavior in relation to information security policies within a natural context. Its distinguishing feature is that it offers people the opportunity to act normally, without feeling obliged to respond perfectly, even in the presence of strict cybersecurity policies.

The approach of Alissa et al. [7] involves the selection of measures and then for each measure specific policies to assess human behavior, claiming that these measures are relevant indicators of how people behave in relation to cybersecurity policies. In the article, the crucial importance of wording the questions appropriately is emphasized to obtain an accurate measurement of behavior. This implies the fundamental assumption that the effectiveness of the questionnaire depends on the quality and appropriateness of the questions asked during the assessment.

In the case of Alissa et al. [7], the selection of specific measures was based on the literature reviewed, considering the most discussed and relevant ones:

The first measure is passwords, specified based on the consensus research project of the System Admin, Audit, Network, Security (SANS) Institute, which ranked passwords as a critical component and important aspect of information and computer security. It provides the user with guidelines and best practices for creating secure passwords.

The second measure deals with e-mail, helping to protect data and providing guidance to users on the safe use of their inbox.

The third measure is identity, based on scientific studies that specify human behavior in the case of identity loss, providing the user with rules to avoid identity theft.

Fourth, sensitive information and how the individual behaves on their behalf are analyzed, giving the user the opportunity to avoid exposure of sensitive data.

The fifth measure addresses physical/resource security and the individual's behavior to protect their resources from unauthorized access by providing workstation-specific guidance

In this methodology, all measures are used to select specific policies on which questions asked of respondents will be based, structured as direct questions with affirmative or negative answers, agreement/disagreement, or open-ended questions. However, it is important to note that Alissa et al.[6] do not provide a method for analyzing the results. The following will present the policies with the attached processed questions.

Immagine che contiene testo, schermata, Carattere, numero

Descrizione generata automaticamente

Figure 4. Policies and questions regarding identity measurement.

### Phishing Attack Simulation.

Simulation of phishing attacks represents a methodology based on exposing users to simulated attempts at social engineering, commonly known as phishing. These attacks are designed to mislead people through manipulation of human psychology, aiming to push them to perform actions that can compromise their security through electronic communication channels. The ultimate goal of such attacks is to exploit individuals' vulnerabilities in order to gain unauthorized access to sensitive information or systems [8].

The implementation of phishing simulations makes it possible to assess users' susceptibility to being deceived, for example, by clicking on malicious links in unsafe e- mails. Such an approach provides an in-depth analysis of users' reaction to similar threats, revealing their ability to recognize and handle potentially dangerous situations online. The ways to test this approach vary depending on the specific attacks to which you wish to expose the user.

Lim et al.[8] developed an innovative evaluation and training system that is based on the fundamental assumption that trainees can acquire skills when exposed to phishing simulations over time. The goal of the system is to offer security training targeting phishing and SMSishing attacks to end users. This system has three key components: the trainee, the Center System, and the Monitoring and Reporting System.

The trainee, represents the individual undergoing the security training, is exposed to simulated social engineering attacks via e-mail and SMS on their devices, such as PCs or smartphones. In the absence of simulation information, trainees might react by opening the e-mail or accessing the website through the virtual phishing URL included in the e-mail or SMS.

Center Systems takes on the role of managing the virtual phishing site and conducting simulated attacks against trainees. This component creates a realistic environment that facilitates the learning process to phishing attacks. Its features include an e-mail system, a virtual web phishing system, a SMS and a Content System, the latter used to manage the content that will later be used within the e-mail or SMS.

The e-mail system, within the Center System, uses a receipt notification function to alert when a trainee opens an e-mail. SMSishing content consists of messages that can trick users, such as invitations to weddings or annual events. The address of the virtual phishing site is presented as a shortened URL in order not to be recognized as a phishing site through the URL and to masquerade as if it were a financial institution or a news site containing social issues.

Immagine che contiene testo, diagramma, schermata, linea

Descrizione generata automaticamenteThe Monitoring and Reporting System, an integral part of the security training system, is designed to manage and collect information on the entire system, including the Center System component. This system records and analyzes data such as e-mail and SMS transfer records, recipient e-mail reading, and virtual phishing site access.

The information collected is leveraged to generate reports presented in a Web environment, facilitating analysis by security managers. The monitoring and reporting system makes use of a database structure to store and manage the collected data, providing statistics and analysis on trainee behavior.

Analysis results are made accessible in a Web environment, facilitating review and interpretation of data by security managers. Additional system features include real-time monitoring of users' e-mail readings, virtual phishing site access, and the ability to configure and manage the table of contents.

To implement and evaluate the proposed system, Lim et al.[8] used a server environment with the following setup: a Web server implemented with Apache for the Monitoring and Reporting System, MySQL as the database, and PHP as the scripting language. The mail server was built using the open source Sendmail. During testing, a virtual phishing e-mail was sent to a specific sample of participants, evaluating user responses and actions to measure the effectiveness of the proposed training system.

The results of the tests conducted by Lim et al.[8] have been reported in detail. In the first and second tests, which involved 481 participants, it was observed that 47% read the phishing e-mail, while in the second test this percentage increased to 32%. In the third and fourth tests, which involved 1045 participants, the number of people who read the e-mail doubled to 52% in the third test and then decreased to 33% in the fourth test. The analysis also showed changes in click rates on the phishing link, going from 16% in the first test to 12% in the fourth test.

This detailed data provides a solid basis for accurate assessment of each user's attitude. It also enables understanding of the effects of the proposed training and demonstrates its ability to improve users' awareness and actions in response to phishing threats.

Similarities can be seen in another tool called the "Phony Phish System," developed by Bowen et al. [6], is composed of several components, each with a specific role in performing the experiments aimed at assessing the security of organizations against e-mail phishing attacks. These components include:

Crawler Module: Used to collect a list of target identities needed for experiments.

E-mail Generator: Integrates various components of the system and manages the sending of e-mails for experiments. This module processes actual e-mails, making changes to entity names and using MD5 hashes to ensure anonymity of user identity information.

Emails that can be generated include several types:

E-mail with internal URLs: internal URLs refer to Web addresses or paths that are within the same network or system of a specific organization or company. The goal is to test how many people verify the recipient's address before opening the e-mail.

E-mail with External URLs External URLs refer to Web addresses or paths that are outside an organization's specific network or system. The goal is to attract users by deception.

Credential Forms: These e-mails contain links to forms that ask users to enter their credentials, for the purpose of assessing how many users are willing to expose their credentials. Credentials are not stored.

Beacon Documents: These emails contain PDF attachments that emit a beacon to the servers when opened. The beacons are designed so that each user issues a unique response, enabling tracking.

Web Application: Plays a crucial role in the Phony Phish System by facilitating the collection of user responses to phishing emails. It allows researchers to monitor users' interactions, such as clicking links or filling out forms, and collect data on their behavior and susceptibility to phishing attacks.

Immagine che contiene schermata, diagramma, linea, testo

Descrizione generata automaticamente

Each component of the Phony Phish System contributes specifically to corporate security assessment, providing a comprehensive approach to measuring susceptibility to phishing attacks and assessing an organization's overall security level.

To test such a system, Bowen et al.[6] conduct two rounds of experiments by sending 500 e-mails for each of the four types of decoy e-mails, with the goal of assessing the propensity of users to fall victim to such attacks. The choice of sample size was based on standard statistical techniques, not well specified ensuring significance in selecting a parameter of choice from the approximately 70,000 IDs in the Columbia University directory, where the test was conducted.

The experiments showed consistency in results between the two cycles, underscoring the possibility of training users to be cautious in the face of suspicious e-mails, although sometimes a significant number of test iterations are required. From the data collected shows that, on average, participants taking the longest time to learn took at most four iterations.

An interesting observation concerns users' response to e-mails that appear to come from internal sources but actually have external sender addresses. Users seem to be less inclined to respond to such e-mails, suggesting greater awareness. However, the difference in e-mail content could affect this trend, highlighting the need for further investigation.

The results were also analyzed from a role-specific perspective, revealing that students are more susceptible to phishing emails than university staff. This could be attributed to less uniformity in adherence to Columbia University's security policies.

Additional considerations include the fact that a significant number of users entered their credentials into the fake forms, with non-academic staff being more vulnerable than academic staff.

Another methodology, without resorting to the use of facilitator systems for creating phishing e-mails, concerns direct phishing campaigns. These are defined as e-mails with specific content sent to a group of employees [11]. According to the research conducted by Gordon et al. [11], Ikhsan et al. [10], and Aljeaid et al. [13], simulated phishing campaigns are crafted to resemble authentic phishing e-mails as closely as possible. If one is clicked on, the real-time opportunity is used to impart a brief educational session on phishing.

Because the likelihood of clicking on different phishing e-mails may vary according to content, a classification was made by Gordon et al. [11] of e-mails into one of the following three categories: office-related, personal, or information technology (IT)- related. To conduct a statistical analysis, Gordon et al. [13] recommend considering the date of the phishing campaign, the campaign number, the number of emails sent, the number of emails clicked, and the content of the emails. These metrics provide a comprehensive and detailed picture for assessing the effectiveness of direct phishing campaigns and identifying any areas for improvement in user awareness and response to such threats.

Gordon et Al.[11] implemented the simulated phishing campaigns involving six U.S. healthcare institutions as a convenience sample. The campaign included 101 phishing simulations, during which a total of 2,975,019 emails were sent between August 1, 2011 and April 2018. The results showed an overall click-through rate of 14.2 percent, indicating that approximately 1 in 7 of the simulated emails were clicked by employees of the institutions involved. Further analysis revealed significant variations in click rates among institutions, with substantial differences in email categories and seasons.

Other types of less conventional methodologies emerge from the study conducted by Aljeaid et al. [13], which aims to assess the level of security knowledge and awareness by simulating other phishing attack techniques. One such simulated attack is clone phishing: in this mode, "phishers" imitate a legitimate website by replicating its design, structure, logos and images. These simulated sites usually induce the user to log in to a system with the intent of stealing information and compromising the security of the local computer, subsequently redirecting the user to malware-infected pages [13]. The first phase of this procedure involves cloning a Web site through various available tools designed to facilitate the creation of new sites using existing templates. Next, the abuse technique of "domain squatting" was adopted through the purchase of a "combosquatting" domain name. "Combosquatting" consists of combining registered trademarks of the legitimate DNS with additional words in order to deceive victims.

In the next step, if users click on any electronic portal, they are redirected to an alert page. Here, they are asked to fill out a questionnaire that investigates demographic information, visual signs that attract attention when checking the legitimacy of counterfeit websites, the level of knowledge about cybersecurity, and the usefulness of the experimental method in improving awareness about phishing.

Additional tool tested by Aljeaid et al. [13] is based on a simulation of social network- based phishing (SNP) attacks: a form of online deception in which a scammer impersonates a trusted person or legitimate organization through social media, such as Twitter, Instagram, WhatsApp and LinkedIn, in order to induce the victim to reveal valuable information[13]. In his study, Alijead et al. [13] divided the simulation cycle into three phases:

The simulation cycle was divided into three phases:

1. Planning:

* Identification of the sample of victims.
* Specification of valuable information to be stolen through phishing.
* Determination of phishing techniques to be used.
* Creating a reliable-looking web page.

1. Collection:

* Victims send their sensitive information.
* Analysis of the information obtained.

1. Fraud:

* Successful social network-based phishing (SNP) attack.

Aljead et al.[13] apply the simulation cycle by creating a scam website for a fictitious car company called "Madar," distributed on social networking platforms such as WhatsApp, Telegram and Facebook.

The planning of the attack included attracting visitors through the promise of a prize drawing on the anniversary of "Madar." Participants were asked to provide personal information, including sensitive details such as national identity card numbers. To ensure an ethical approach, participants had to explicitly agree to the terms and conditions, waiving any rights to the information provided.

The data collection phase showed that about 47 percent of visitors were exposed to phishing attack, with a significant number of victims among college students. The simulation carried out also included an educational phase, warning participants of the attack and providing information on preventive measures.

### Gamification

Gamification represents a technique used to enhance user engagement and motivation by introducing ludic elements into non-strictly ludic contexts[9]. In the analysis of publications conducted, its use was highlighted, both in the context of psychometric questionnaires and in simulations of phishing attacks.

Alissa et al., in their in-depth study, extend the progress of their questionnaire designed to assess human behaviors regarding cybersecurity policies [7]. They convert the questionnaire, created through selected policies, into adaptive user scenarios. The goal of this transformation is to address the problem of unreliable responses typically associated with traditional questionnaires. To this end, they develop a simulated intelligent tool that leverages gamification[9] to increase user engagement and motivation, resulting in more realistic and meaningful responses.

The interactive questionnaire is implemented through web programming languages, including HTML5/CSS and JavaScript, which allow for an engaging and interactive visual representation. It consists of ten scenes, which address specific contexts related to the policies analyzed by Alissa et al.[7]. The ten scenes presented in their study are briefly described below:

Sending a Resume and Job Application:

A potential candidate sends his resume to a company asking for a job. The secretary must decide whether to open the file immediately, check it before downloading, or open it later, thus affecting information security.

Acceptance of a Job and Password Management:

A new employee accepts a job and is given a default password. She must decide whether to keep it or change it, facing the evaluation of the password strength.

Password storage:

After entering a new password, the user must decide whether to store it in the mind, save it on the phone, or leave it open for all to see, affecting password security.

Preparation of a Meeting and Submission of a Report:

An employee is asked to prepare a presentation for a meeting and must send a report via e-mail. He must decide when to send the e-mail with respect to the meeting.

Closing the Computer at the End of the Day:

After a day's work, the user must decide how to turn off the computer: via a shutdown button, by pressing the power button, or by removing the power cord.

Office Closure and Physical Security:

Before leaving the office, the user must decide whether to lock the door, ensuring the physical security of IT tools.

Personal Item Loss Experience:

A colleague takes advantage of an employee's absence, takes a picture of the credit card, and tries to access the bank account. The user must decide how to respond to the situation.

Managing a Marketing Email:

The user receives a marketing e-mail and must decide whether to open, block, or ignore it, facing possible phishing threats.

Submission of a Sensitive Report:

A supervisor asks the user to send a critical report. The user must decide how to protect the content of the e-mail, choosing between sending it immediately, asking a colleague to send it, or postponing sending it.

Discussion on Security and Backup Practices:

During a conversation, a colleague asks if the user installs antivirus protections and software updates and if the user follows the company's weekly backup policy. The user must answer these questions by reflecting the cybersecurity practices adopted.

These ten scenes cover a variety of realistic situations, evaluating users' decisions and actions in diverse cybersecurity-related contexts. The diversity of scenarios provides a comprehensive framework for evaluating human behavior in response to specific cybersecurity challenges. This innovative approach aims to overcome the limitations associated with traditional data collection methods, offering a more dynamic and adherent perspective on real user behaviors in cybersecurity-related contexts.

Immagine che contiene schermata, Viso umano, testo, design

Descrizione generata automaticamente

Figure 5. Interactive questionnaire scenes

Hale et al. [12] propose an innovative platform called CyberPhishing, which is game- based and combines the realism of "in the wild" approaches with the training capabilities of testing approaches. Such a platform is designed to educate users about phishing attacks through a simulation game, providing specific feedback based on content misclassification and tailoring training to individual deficiencies. The platform consists of three basic elements: e-mail, web browsing, and social media. All of these are integrated into a game scenario and are mediated by the simulation modules, while the game engine reacts to user actions, guiding the development of the game scenario and collecting usage data for analysis.

Email simulation in CyberPhishing presents users with email content, including malicious phishing messages and legitimate emails, within an HTML-enabled interface. This simulation goes beyond simple display, including various elements such as fake links, images with fake links (visible on mouseover), attachments, and javascript simulation features. This simulation replicates standard e-mail interfaces, such as Gmail, allowing users to interact with e-mails and use similar functionality. Users can click on items in the inbox to view them in full, providing a realistic e-mail browsing experience. It is crucial to emphasize that all clickable links or pop-ups in the simulation do not direct users to actual malicious sites, ensuring user safety.

Immagine che contiene testo, schermata, linea, Carattere

Descrizione generata automaticamente

Figure 6. Simulated mailbox interface.

Immagine che contiene testo, schermata, Pagina Web, Sito Web

Descrizione generata automaticamente

Figure 7. Simulated e-mail interface.

The web browser simulation is an essential component of CyberPhishing and provides users with the opportunity to experience realistic browsing scenarios while learning to identify phishing indicators in web content. Designed to replicate the functionality of a real Web browser, this simulation allows users to navigate Web sites within the simulation environment. It also includes examples of malicious counterfeit websites, where users can identify malicious links and URLs other than legitimate ones.

Immagine che contiene testo, Sito Web, Pubblicità online, schermata

Descrizione generata automaticamente

Figure 8. Simulated web browser interface.

The social media simulation in CyberPhishing recreates a Twitter-like interface, allowing users to interact with simulated social media content. This simulation, known as "My BVerse" in the role-playing game scenario, provides users with a realistic environment in which to evaluate social media content, including potential phishing attempts. The interface displays contextual clues, such as usernames and user names, similar to typical Twitter displays.

Immagine che contiene testo, schermata, Pagina Web, Sito Web

Descrizione generata automaticamente

Figure 9. Simulated social network interface "My Bverse".

## Conclusions

This systematic literature review focused on methodologies for assessing the risk associated with digital security arising from human behavior. Emerging results indicated that the most widely adopted methodologies include psychometric questionnaires and simulated phishing attacks.

Psychometric questionnaires, tools based on the analysis of cognitive and behavioral factors, focus on assessing users' knowledge, attitudes and behaviors regarding cybersecurity. Despite their relatively simple and inexpensive implementation, it is important to note that these questionnaires may not be completely effective in measuring users' actual behaviors in the face of threats because they rely on users' self- assessment, potentially leading to responses that are compliant but not reflective of reality or totally incorrect due to loss of user interest.

Simulations of phishing attacks, on the other hand, allow users' susceptibility to such threats to be assessed by creating simulated scenarios of fraudulent e-mails or websites. This methodology offers a more accurate measurement of actual user behavior. It should be noted that simulations of phishing attacks can involve higher costs and greater complexity in implementation, as well as raise ethical issues related to the need to keep user exposure to simulations secret in order to capture authentic behaviors.

During the analysis carried out within the following study, the application of gamification in the context of risk assessment associated with digital security was also examined. Gamification, based on the integration of playful elements, emerges as an effective tool in arousing user interest and maintaining it throughout the assessment process. This contribution may translate into greater accuracy of the results obtained. It should be noted, however, that at present this methodology holds particular importance in the risk mitigation and prevention phases, as it stimulates deeper learning on the part of users.

The choice of risk assessment methodology should be guided by a careful assessment of context-specific needs and available resources. While psychometric questionnaires can provide a general view of users' knowledge and attitudes, simulations of phishing attacks provide a more realistic measurement of actual behaviors. Gamification, on the other hand, represents a promising option for engaging users, but its implementation should be carefully considered in relation to the evaluation objectives it is intended to pursue.

The search fo papers in the systematic literature review is affected by the search strings used. A feasible improvement would be to modify the search strings to obtain a broader view of existing solutions to assess human-caused vulnerabilities in cybersecurity.

## References

[1] Verizon, 'DBIR 2023 Data Breach Investigations Report About the cover', 2023. [Online]. Available: https://www.verizon.com/business/resources/reports/dbir/

[2] B. Kitchenham, 'Procedures for Performing Systematic Reviews,' 2004. [Online].

Available: https://www.researchgate.net/publication/228756057

[3] C. Wohlin, 'Guidelines for Snowballing in Systematic Literature Studies and a Replication in Software Engineering', 2014, doi: 10.1145/2601248.2601268.

[4] A. Pollini et al, 'Leveraging human factors in cybersecurity: an integrated methodological approach', Cognition, Technology and Work, vol. 24, no. 2, pp. 371-390, May 2022, doi: 10.1007/s10111-021-00683-y.

[5] M. Antunes, C. Silva, and F. Marques, 'An integrated cybernetic awareness strategy to assess cybersecurity attitudes and behaviors in school context', Applied Sciences (Switzerland), vol. 11, no. 23, Dec. 2021, doi: 10.3390/app112311269.

[6] B. M. Bowen, R. Devarajan, and S. Stolfo, 'Measuring the human factor of cyber security', 2011 IEEE International Conference on Technologies for Homeland Security, HST 2011, pp. 230-235, 2011, doi: 10.1109/THS.2011.6107876.

[7] K. A. Alissa et al, 'An Instrument to Measure Human Behavior Toward Cyber Security Policies', 21st Saudi Computer Society National Computer Conference, NCC 2018, Dec. 2018, doi: 10.1109/NCG.2018.8592978.

[8] I. kwon Lim, Y. G. Park, and J. K. Lee, 'Design of Security Training System for Individual Users', Wirel Pers Commun, vol. 90, no. 3, pp. 1105-1120, Oct. 2016, doi: 10.1007/s11277-016-3380-z.

[9] K. A. Alissa et al, 'Developing a Simulated Intelligent Instrument to Measure User Behavior toward Cybersecurity Policies', 2021.

[10] M. G. Ikhsan and K. Ramli, 'Measuring the Information Security Awareness Level of Government Employees Through Phishing Assessment'.

[11] W. J. Gordon et al, 'Assessment of Employee Susceptibility to Phishing Attacks at US Health Care Institutions', JAMA Netw Open, vol. 2, no. 3, Mar. 2019, doi: 10.1001/jamanetworkopen.2019.0393.

[12] M. L. Hale, R. F. Gamble, and P. Gamble, 'CyberPhishing: A game-based platform for phishing awareness testing', in Proceedings of the Annual Hawaii International Conference on System Sciences, IEEE Computer Society, Mar. 2015,

pp. 5260-5269. doi: 10.1109/HICSS.2015.670.

[13] D. Aljeaid, A. Alzhrani, M. Alrougi, and O. Almalki, 'Assessment of end-user susceptibility to cybersecurity threats in saudi arabia by simulating phishing attacks', Information (Switzerland), vol. 11, no. 12, pp. 1-19, Dec. 2020, 10.3390/info11120547.

[14] K. Parsons, A. McCormac, M. Butavicius, M. Pattinson, and C. Jerram, 'Determining employee awareness using the Human Aspects of Information Security Questionnaire (HAIS-Q)', Comput Secur, vol. 42, pp. 165-176, 2014, doi: 10.1016/j.cose.2013.12.003.

[15] K. Parsons, D. Calic, M. Pattinson, M. Butavicius, A. McCormac, and T. Zwaans, 'The Human Aspects of Information Security Questionnaire (HAIS- Q): Two further validation studies', Comput Secur, vol. 66, pp. 40-51, May 2017, doi: 10.1016/j.cose.2017.01.004.