Dec 12th, 12:00 AM

Phish Finders: Improving Cybersecurity Training Tools Using Citizen Science

Vinod Kumar Ahuja
University of Nebraska Omaha, vahuja@unomaha.edu

Holly K. Rosser
University of Nebraska Omaha, hrosser@unomaha.edu

Andrea Grover
University of Nebraska at Omaha, andreagrover@unomaha.edu

Matthew Hale
University of Nebraska at Omaha, mlhale@unomaha.edu

Phish Finders: Improving Cybersecurity Training Tools Using Citizen Science

Short Paper

Vinod Kumar Ahuja
University of Nebraska Omaha
vahuja@unomaha.edu

Holly Rosser
University of Nebraska Omaha
hrosser@unomaha.edu

Andrea Grover
University of Nebraska Omaha
andreagrover@unomaha.edu

Matthew Hale
University of Nebraska Omaha
mlhale@unomaha.edu

Abstract

Malicious web content includes phishing emails, social media posts, and websites that imitate legitimate sites. Phishing attacks are rising, and human-centered phishing risk mitigation is often an afterthought eclipsed by technical system-centric efforts like firewalls. Training tools can be deployed for combating phishing but often lack sufficient labeled training content. Using signal detection theory, this paper assesses the feasibility of using citizen science and crowdsourcing volunteers to label images for use in cybersecurity training tools. Crowd volunteer performance was compared to gold standard content and prior studies of Fortune 500 company employees. Findings show no significant statistical differences between crowd volunteers and corporate employees' performance on gold standard content in identifying phishing. Based on these findings, citizen scientists can be valuable for generating annotated images for cybersecurity training tools.

Keywords: Phishing, citizen science, cybersecurity, crowdsourcing

Introduction

Malicious web content takes on a variety of forms, including targeted and untargeted phishing emails, social media posts, and websites that emulate the look and feel of legitimate sites. As a cornerstone of multi-billion-dollar criminal profit schemes, phishing is rising, with millions of people victimized annually (Singh et al. 2016). A 2015 survey (Schupak 2015) found that 80% of respondents misidentified a phishing email as legitimate. Despite the significant scale of this problem, efforts to mitigate phishing risk mitigation by training users to better avoid dangerous content are often an afterthought eclipsed by technical system-centric efforts (Alabdan 2020; Howarth 2014) like firewalls and spam filters. In the current state of the art, human-centric risk mitigation focuses on email campaigns to "catch" risky users and then direct them to seminar-style training to learn more about phishing. While the former has been shown effective at estimating the likelihood of phishing success and identifying who needs training, the latter hasn't made significant inroads toward reducing phishing impacts in the long term (Alkhalil et al. 2021; Ghazi-Tehrani and Pontell 2021; Holdsworth and Apeh 2017). These results align with high-fidelity laboratory-based studies that collected think-aloud and eye-tracker data (Hale et al. 2017; Hefley et al. 2018), indicating that better human-centered prevention methods are a necessity for online safety and threat impact reduction.

More recently, training platforms that employ active learning techniques to better engage the user and help them improve their phishing assessment capabilities have emerged (Hale et al. 2017; KnowBe4 2022) to better address the problem. In these tools, users assess real-world content and determine if it is phishing or not. Facilitating active-learning requires an abundance of labeled phishing samples that users can review and interact with. Labels highlight specific areas in the content that should be prompting suspicion. Generating enough labeled data, which must also be relatively recent as "old" content is often treated with suspicion, is non-trivial and costly. Existing tooling continues to rely upon cybersecurity experts to view, assess, and label phishing samples (KnowBe4 2022). This approach is not scalable with the global shortage of experts, especially given the large amounts of labeled data needed for training tools. Notably, while machine learning seems like a sensible strategy for labeling data, it typically also requires a large human-labeled corpus for training, which could be generated through crowdsourcing.

Crowdsourcing has been used in other contexts for data generation, so it might work in phishing too. As a form of crowdsourcing that focuses on collecting and processing data (Law et al. 2017), citizen science is particularly notable for engaging the public as volunteers instead of paid microtask workers to generate labeled data (Bonney et al. 2014; Lukyanenko et al. 2020).

This paper explores phishing label data generation through crowdsourcing using adapted methods from citizen science. It evaluates the untrained volunteer crowd's ability to identify phishing on expert labeled content (i.e. gold standard) compared to a more well-studied population of users who have had anti-phishing training. The feasibility of crowdsourcing methods for labeling phishing cues in samples is directly tied to how well volunteers can identify phishing on expert labeled content. We also compared phishing detection among employees of a Fortune 500 company's IT department to crowd volunteers to further contextualize the results. While neither population should be expected to perform at the same level as cybersecurity experts, if untrained volunteers are able to perform comparably to trained employees then it suggests additional potential roles for volunteers in developing anti-phishing training. Specifically, our research question is: *How well do citizen science volunteers detect phishing threats on expert labeled content compared to corporate IT employees?*

To answer this question, we developed a citizen science project called Phish Finders to collect crowd data. Phish Finders asked volunteers to identify phishing tactics among a corpus of non-malicious and malicious phishing images. We used signal detection theory to measure how well citizen science volunteers in Phish Finders detected phishing cues compared to trained corporate IT employees. Signal detection theory emerged from WWII research on radar and radio signals as an explanation of the difference between signal and background noise. In a phishing context, signals are phishing cues that participants must identify, and background noise is other visual stimuli, such as a website or email styling. This study contributes a novel application of signal detection theory in IS research and empirical results indicating that citizen science volunteers and corporate employees perform similarly on phishing detection tasks.

Background

Crowdsourcing through Citizen Science

Crowdsourcing is a practice that outsources a task to a large number of people or volunteers (Howe 2006). Crowds can solve complex problems not suited for machine learning and can be used for high-volume tasks. For example, companies offer bug bounties to crowds to identify security vulnerabilities in their systems. The power of crowdsourcing lies in the collective wisdom of many individual and diverse participants (Blohm et al. 2013). Citizen science is a type of crowdsourcing that Information System (IS) research has started to incorporate to advance scientific work while generating large datasets and new discoveries (Levy and Germonprez 2017; Lukyanenko et al. 2020).

Phish Finders was developed on Zooniverse, an online citizen science platform with a thriving volunteer community that allows science teams to upload a corpus of images for analysis and tailor volunteers' response options to their research needs. It is commonly used for science tasks such as identifying wildlife species in photographs and spotting patterns in astrophysics data. Zooniverse projects collect demonstrably robust data by leveraging consensus across many volunteers (Hines et al. 2015; Kosmala et

al. 2016). Cumulatively, Zooniverse has involved millions of volunteers in research and has supported hundreds of studies. Citizen scientists have labeled images in a variety of disciplines, including astronomy, wildlife ecology, history, and physics (Hines et al. 2015), but to the best of our knowledge, it has not been focused on cybersecurity.

Despite wide adoption in diverse domains, researchers often harbor concerns over the quality of citizen science data generated by a heterogeneous group of nonprofessionals (Kosmala et al. 2016). To ensure that data produced by citizen science volunteers are valid and reliable for use in cybersecurity training tools, as a primary goal of this work, we analyzed the classification performance of corporate employees and citizen science volunteers using a signal detection theory lens.

Signal Detection Theory

Signal detection theory suggests that "the decisions are made against a background of uncertainty, and the goal of the decision-maker is to tease out the decision signal from background noise" (Anderson 2015, p. 1). Signal detection theory frames human decision-making behaviors as assessments of stimuli to identify relevant details from background noise. In phishing, the stimulus is online content, and the goal is to detect the presence or absence of phishing cues on the screen (Hautus et al. 2021). The stimuli can include signals like the presence of a lock icon in the browser address bar or trusted brand logos.

Signal detection theory has been used in a variety of domains with two types of stimuli called "noise" and "signal" in psychophysics (Hautus et al. 2021; Klein et al. 1997; Ye and van Raaij 2004). These two types of stimuli map to "nonmalicious" and "malicious" in cybersecurity. Noise is the nonmalicious background present in an image, whereas a signal is a malicious cue that needs to be identified by the participant (Hautus et al. 2021; Klein et al. 1997). Signal detection theory provides measures for assessing participants' performances according to their accuracy at detecting the signals from the background, i.e., their ability to identify phishing cues.

Evaluating performance at identifying signal and noise requires measuring how often participants correctly and incorrectly identify malicious and trustworthy content (Hautus et al. 2021; Klein et al. 1997; Ye and van Raaij 2004). For signals, we can describe accuracy as hit or miss, and for noise, a false alarm or correct rejection (Refer to table 1). A *hit*, *or true positive*, means a signal is correctly identified. If a participant correctly identifies a phishing cue, they accurately detect signals present in malicious content. Second, a *miss*, or *false negative,* means the participant fails to identify the presence of a signal. A miss can result in a participant falling victim to a phishing attack due to failure to identify the cues. Third, *false alarm*, or *false positive*, means that a participant identified a signal in trustworthy content when there wasn't any. False alarm can occur when participants are overly cautious and treat all stimuli as malicious or when quirks of content presentation are perceived as suspicious. Last, *correct rejection* or *true negative* means that there is no signal indicative of phishing, and the participant does not mark the image as containing a cue.

In signal detection, two factors can impact participants' performance: discriminability of the distribution and response criteria (Klein et al. 1997). The discriminability of the distribution explains how the data is distributed between the presence and absence of phishing cues, which impacts participants' responses. If participants perceive that the corpus contains primarily malicious images, then most of their responses will be "yes" when asked, "is there something phishy here" even when images do not have phishing content, leading to more false alarms. Similarly, if participants perceive that the corpus is primarily trustworthy content, then most of their responses will be "no" leading to a high ratio of misses. To address this factor in our research, we kept the distribution of images approximately equal between malicious and trustworthy content; future work could consider evaluating the impacts of presenting different ratios of malicious and trustworthy content. Participant performance can also be affected by their prior knowledge and training (Hautus et al. 2021), which cannot be controlled for in a citizen science context where recruiting volunteer participants requires minimal barriers to participation, in contrast to most paid crowdsourcing platforms where some level of worker screening is common.

		Classification by participants	
		Phishing Identified (Positive)	Phishing not identified (Negative)
Actual Data	Signal present (Phishing in stimuli)	Hit (True positive)	Miss (False negative)
	Signal absent (No phishing in stimuli)	False Alarm (False Positive)	Correct Rejection (True Negative)

Table 1. Signal detection classification matrix: Identifying phishing cues in image content.

Methods

This study was an observational experiment conducted within a fidelity-preserving synthetic environment, namely the crowdsourcing platform Zooniverse, in which images were presented in a nondescript web browser or email client. Volunteer performance was benchmarked and compared against data obtained from an earlier study of incentivized IT employees in a Fortune 500 company using a training tool called Cybertrust (Hale et al. 2017). In this prior study, a total of 34 HTML pages were generated and evaluated by cybersecurity experts and then used to train employees to recognize phishing cues. Employees were shown interactive emails and websites and asked to rate their trustworthiness. The content samples from the earlier study are used in this work as a "gold standard" for benchmarking participant performance based on a comparison to the evaluation by cybersecurity experts.

The Cybertrust study task was replicated in Phish Finders on the Zooniverse platform with some modifications due to platform capabilities. To structure the Phish Finders experiment, the 34 HTML pages from Cybertrust were converted into images. Four were used as examples in the training materials and were not used in the corpus for classification, and the remaining 30 images, containing 16 malicious and 14 trustworthy images, were retained as gold standard benchmarks. Another corpus of 1892 images containing 817 malicious and 1075 trustworthy images was generated from websites spanning various sectors commonly used in phishing attacks. These included banking, government, law enforcement, social networking, eCommerce, news, entertainment, and telecommunication sites. The websites were retrieved from the Internet Archive[1] and extracted as an image using a screenshot program. We also conducted an expert review to verify that the phishing techniques represented in the existing gold standard content were still relevant and present in the new corpus. Finally, a browser header was added for realism and to allow participants to identify suspicious domains in the browser address bar. As is typical of Zooniverse projects, training materials included a tutorial, field guide, and "help" button text; volunteers were not compelled to use any of this material.

In Cybertrust, participants viewed phishing and non-phishing images, rated them as trustworthy or not, and provided 5-point Likert ratings for the degree of perceived trustworthiness. After each view, Cybertrust participants were shown the correct response, with labeled bounding boxes around phishing content drawn by the experts as training to avoid such mistakes moving forward. Bounding boxes were shown around five malicious cues types (Hale et al. 2017, 2015): Spelling & Grammar, Malicious Links, Invalid Domain or Sender, and Appeals Used to Elicit Action - Authority, - Greed, and - Urgency. Phish Finders participants performed the same tasks, viewing content and identifying its perceived trustworthiness, but without receiving feedback. If Phish Finders participants labeled content as malicious, they were asked to annotate the image with color-coded bounding boxes to highlight phishing cues. If they observed any other indications of phishing besides the five cues listed above, they could select

[1] Internet Archive (https://archive.org/) captures and stores snapshots of various websites on the Internet at different points in time, and makes these data available for research use.

"Other Phishy Findings", draw a bounding box, and enter free text describing the issue. While the tasks completed by each population were different, both studies directly evaluated responses to the same gold standard content using the same set of labels, providing a reasonable foundation for comparison.

Data and Results

To assess their comparative performance, we evaluated the population-level responses of the Cybertrust participant and the Phish Finders volunteers, summarized in Table 1. The Phish Finders data was collected in January 2021. The data includes an anonymized participant id, session time (i.e., how long the participant engaged with the experiment in Zooniverse), the number of images classified, the labels and bounding boxes drawn, and Likert scale data described above. The Cybertrust data was cross-matched by content (stimuli) id for two types of analyses. A content-focused comparison examined population-level performance on the gold standard images. The overall performance of participants was examined across the full corpus of classifications in each study.

Description	**Cybertrust**	**Phish Finders**
Number of Participants	82 company employees	1825 volunteers
Average number of classifications per participant	31.56	16.15
Average session length	14.25 minutes	35.83 minutes
"gold standard" classifications	2,043 (participants only had access to gold standard and training images)	504 (randomly selected gold standard images presented at a rate of 10% of session images)
Total Classifications	2,588	29,489

Table 2: Descriptive Statistics for Cybertrust and Phish Finders Participants

Comparison of gold standard images (Content-focused)

Content-focused comparison evaluated performance on the same gold standard images for each group to assess the ability of trained corporate employees and untrained citizen science volunteers to identify phishing cues on the same images. We looked at all four types of signal detection outcomes and assessed the confidence of both groups in their assessments through the trustworthiness rating.

Hit rate, also called *sensitivity*, indicates whether participants correctly identified malicious image content where present (a true positive assessment). Comparing the performance of both the groups on the Wilcoxon rank sum test (W = 140, p-value = 0.6636), no significant difference between them was noted.

Miss rate, also called *victimization* rate, is the percentage of participants that incorrectly assess a malicious image as trustworthy, leading to victimization. The victimization rate for each of the 30 gold standard images was measured on both platforms as the number of times users missed the signal when a signal was present (aka, the content was malicious and users said it was legitimate). In Figure 1, the victimization rate for Phish Finders volunteers appears higher than the Cybertrust participants, but a Wilcoxon rank sum test shows similar victimization rates for the groups (W=140, p-value = 0.332).

False positive, or *false alarm*, indicates that content is trustworthy, but participants marked it as malicious. For both groups (refer to figure 1), each of the 14 trustworthy images were marked as malicious at least once. Cybertrust had a false positive rate of 30% (291 out of 961 total classifications), and Phish

Finders had a 40% rate (90 out of 225 total classifications). The Wilcoxon rank sum test (W=185.5, p-value = 1) indicates the false positive rates were not significantly different between the groups.

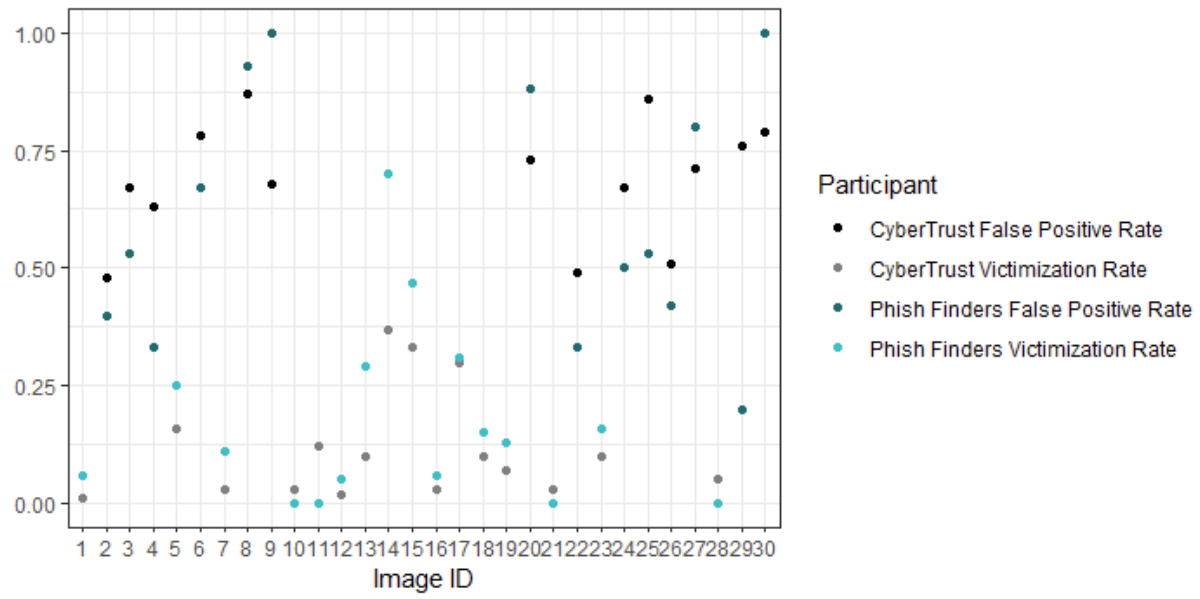

Figure 1: Victimization and False Positive Rate for Percentage of Population for Cybertrust and Phish Finders

Specificity identifies whether participants correctly identified trustworthy image content (i.e., a true negative assessment). With the Wilcoxon rank sum test (W = 76, p-value = 0.3229), no significant difference was noted between the groups.

Trustworthiness rating indicates the self-reported confidence of participants in detecting the phishing cues using a five-point Likert scale from very untrustworthy to very trustworthy. Cybertrust participants rated images as untrustworthy for 61% (798 not trusted/2043 total) of all classifications. Phish Finders volunteers rated the same images as untrustworthy in 63% (319 not trusted/504 total) of all classifications. The Wilcoxon rank sum test (W = 462, p-value = 0.8572) indicates no significant difference between the groups. These results suggest that in terms of how specific content was classified and rated, the citizen science volunteers and corporate employees performed similarly in identifying phishing cues.

Comparison of participant performance metrics

To compare the overall performance of participants in each group, irrespective of which images they classified, we calculated precision (ratio of hits to hits plus false alarms) and recall (ratio of hits to hits plus misses), and F1 score (harmonic mean of precision and recall) from the signal detection classification matrix (see Table 1). F1 scores assess overall performance in both the scenarios, namely presence of and absence of phishing cues. Precision and recall are plotted in Figure 2a, and F1 score and total images classified are plotted in Figure 2b. A higher F1 score indicates better overall accuracy.

From Figure 2, we see a group of Phish Finders volunteers with perfect precision and recall, or F1 scores of one. We examined the work of volunteers who had perfect scores and found that they classified only a few images, between one and five, yielding perfect scores with no false positives or false negatives. Hence, this finding is more an artifact of variable participation rates than evidence of truly expert classification.

To statistically compare the groups' performance, the normality assumption was analyzed for F1 scores, which indicated that data from Phish Finders is left skewed and not normal. To statistically check the normality assumption, the Shapiro Test was performed; results indicated that not all samples are normally distributed. Since the data is not normal, the Wilcoxon Rank Sum Test was used to statistically compare the accuracy score (F1) for both groups, with a result (W=50840, p-value = 0.1245), indicating no significant difference between the Cybertrust participant and Phish Finder volunteers. In aggregate, the untrained volunteers performed similarly to the trained participants.

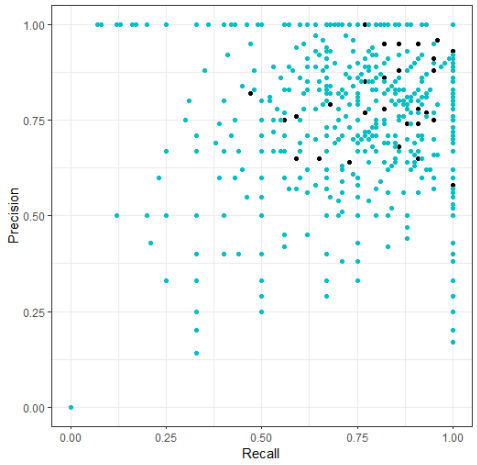

Figure 2a: Precision and Recall

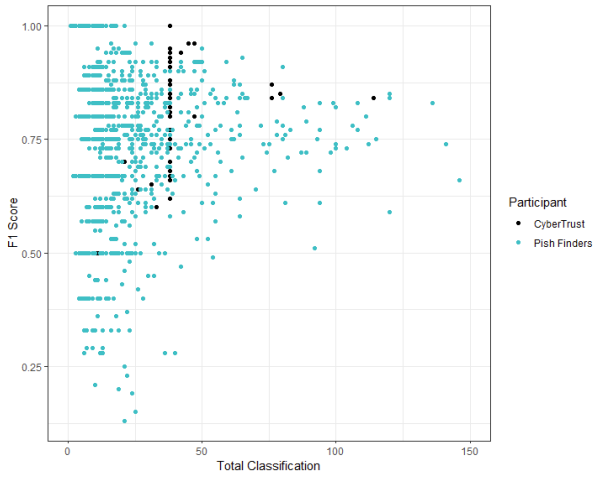

Figure 2b: Total Images Classified

Discussion

This study contributes empirical results showing that citizen science volunteers and corporate IT employees perform similarly on tasks focused on detecting phishing cues in gold standard content, and provides a novel application of signal detection theory in IS research. Signal detection theory offers a strong theoretical foundation for evaluating performance on labeling tasks, and may be useful for IS research focused on assessing the performance of crowd workers on tasks where a binary evaluation of accuracy can be made. This also means that citizen science volunteers could potentially support user testing during anti-phishing training tool development.

In Cybertrust, participants received phishing training as part of the study, with immediate feedback after 10 classifications, whereas Phish Finder's volunteers were able to skip the introductory tutorial and were not shown the results of their responses due to a lack of suitable platform functionality. The Phish Finders volunteers had less support and feedback, and therefore should not be expected to perform as well as the corporate employees. Our results showed no significant differences in the aggregate evaluations of the gold standard images, supporting the standard practice for such projects, which is to retain all data because prior studies have shown that one-off volunteers make meaningful contributions (Jackson et al. 2018). These results suggest, on multiple levels, that crowdsourcing can be an effective tool for developing annotated image corpuses focused on phishing, which opens up an interesting range of opportunities for human-centered cybersecurity research and training tool development.

Limitations and Future Work

This study's limitations include several challenges from using two platforms that are meaningfully similar but not identical. There were differences in task structure due to the Zooniverse platform capabilities, described earlier. While this complicated comparison across groups for the task as a whole, and for

annotations on specific cue types, we also believe that the methods used in this study may be advantageous for phishing victimization research, as it applied a novel strategy.

Another challenge was variability in the number of images classified by each individual. The variable number of items classified by Phish Finders volunteers added complexity to our analyses, as reported in our results. We could not directly assess the impact of the built-in training and feedback available to Cybertrust participants, as it was not available to Zooniverse volunteers, but this is a clear avenue for future work when platform functionality permits. The similarity of performance between populations also suggests that crowdsourcing could be used for testing novel anti-phishing training tools. We also expect that future work focusing on temporal analysis may provide insight into whether additional exposure and practice with identifying cues led to improved performance despite the lack of training.

Conclusions

We deployed a crowdsourcing project called Phish Finders on the citizen science platform Zooniverse for comparison to Fortune 500 employees on a similar task focused on identifying phishing cues in gold standard content. This paper contributes a comparison of the resulting data at the image level and at the participant level, with performance measures based on signal detection theory, simultaneously providing an example of the application of signal detection theory and citizen science methods in IS research. To answer our research question about the comparability of the two groups at detecting phishing, we found that participants in both groups had similar performance on most measures and the only meaningful difference in the outcomes may be attributable to the role of training and feedback, which should be further evaluated in future work. This result is consistent with prior findings that the collective performance of volunteers generates quality research data that can be on par with that of professionals (Kosmala et al. 2016). It also suggests that citizen science can provide valuable opportunities for cybersecurity research.

Acknowledgments

This work was supported by the University Committee on Research and Creative Activity (UCRCA), the University of Nebraska at Omaha (UNO). We thank Zooniverse volunteers for advancing this scientific research. We also thank the anonymous reviewers from the ICIS conference for their suggestions for improving this work.

References

Alabdan, R. 2020. “Phishing Attacks Survey: Types, Vectors, and Technical Approaches,” Future Internet (12:10), Multidisciplinary Digital Publishing Institute, p. 168. (https://doi.org/10.3390/fi12100168).

Alkhalil, Z., Hewage, C., Nawaf, L., and Khan, I. 2021. “Phishing Attacks: A Recent Comprehensive Study and a New Anatomy,” Frontiers in Computer Science (3). (https://www.frontiersin.org/articles/10.3389/fcomp.2021.563060).

Anderson, N. D. 2015. “Teaching Signal Detection Theory with Pseudoscience,” Frontiers in Psychology (6). (https://www.frontiersin.org/article/10.3389/fpsyg.2015.00762).

Blohm, I., Leimeister, J. M., and Krcmar, H. 2013. Crowdsourcing: How to Benefit from (Too) Many Great Ideas, p. 14.

Bonney, R., Shirk, J. L., Phillips, T. B., Wiggins, A., Ballard, H. L., Miller-Rushing, A. J., and Parrish, J. K. 2014. “Next Steps for Citizen Science,” Science (343:6178), American Association for the Advancement of Science, pp. 1436–1437. (https://doi.org/10.1126/science.1251554).

Ghazi-Tehrani, A. K., and Pontell, H. N. 2021. “Phishing Evolves: Analyzing the Enduring Cybercrime,” Victims & Offenders (16:3), Routledge, pp. 316–342. (https://doi.org/10.1080/15564886.2020.1829224).

Hale, M. L., Gamble, R., Hale, J., Haney, M., Lin, J., and Walter, C. 2015. “Measuring the Potential for Victimization in Malicious Content,” in 2015 IEEE International Conference on Web Services, , June, pp. 305–312. (https://doi.org/10.1109/ICWS.2015.49).

Hale, M., Walter, C., Lin, J., and Gamble, R. 2017. "A Priori Prediction of Phishing Victimization Based on Structural Content Factors," International Journal of Services Computing (5). (https://doi.org/10.29268/stsc.2017.5.1.1).

Hautus, M. J., Macmillan, N. A., and Creelman, C. D. 2021. Detection Theory: A User's Guide, (3rd ed.), New York: Routledge. (https://doi.org/10.4324/9781003203636).

Hefley, M., Wethor, G., and Hale, M. L. 2018. Multimodal Data Fusion and Behavioral Analysis Tooling for Exploring Trust, Trust-Propensity, and Phishing Victimization in Online Environments, , January 3. (https://doi.org/10.24251/HICSS.2018.108).

Hines, G., Kosmala, M., Swanson, A., and Lintott, C. 2015. "Aggregating User Input in Ecology Citizen Science Projects," in Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence, AAAI'15, Austin, Texas: AAAI Press, January 25, pp. 3975–3980.

Holdsworth, J., and Apeh, E. 2017. "An Effective Immersive Cyber Security Awareness Learning Platform for Businesses in the Hospitality Sector," in 2017 IEEE 25th International Requirements Engineering Conference Workshops (REW), , September, pp. 111–117. (https://doi.org/10.1109/REW.2017.47).

Howarth, F. 2014. "The Role of Human Error in Successful Security Attacks," Security Intelligence, , September 2. (https://securityintelligence.com/the-role-of-human-error-in-successful-security-attacks/, accessed May 4, 2021).

Howe, J. 2006. "The Rise of Crowdsourcing," Wired. (https://www.wired.com/2006/06/crowds/).

Jackson, C. B., Crowston, K., and Østerlund, C. 2018. "Did They Login? Patterns of Anonymous Contributions in Online Communities," Proceedings of the ACM on Human-Computer Interaction (2:CSCW), 77:1-77:16. (https://doi.org/10.1145/3274346).

Klein, B. D., Goodhue, D. L., and Davis, G. B. 1997. "Can Humans Detect Errors in Data? Impact of Base Rates, Incentives, and Goals," MIS Quarterly (21:2), p. 169. (https://doi.org/10.2307/249418).

KnowBe4. 2022. "Security Awareness Training | KnowBe4." (https://www.knowbe4.com, accessed April 26, 2022).

Kosmala, M., Wiggins, A., Swanson, A., and Simmons, B. 2016. "Assessing Data Quality in Citizen Science," Frontiers in Ecology and the Environment (14:10), pp. 551–560. (https://doi.org/10.1002/fee.1436).

Law, E., Gajos, K. Z., Wiggins, A., Gray, M. L., and Williams, A. 2017. "Crowdsourcing as a Tool for Research: Implications of Uncertainty," in Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing, CSCW '17, New York, NY, USA: Association for Computing Machinery, February 25, pp. 1544–1561. (https://doi.org/10.1145/2998181.2998197).

Levy, M., and Germonprez, M. 2017. "The Potential for Citizen Science in Information Systems Research," Communications of the Association for Information Systems (40:1). (https://doi.org/10.17705/1CAIS.04002).

Lukyanenko, R., Wiggins, A., and Rosser, H. K. 2020. "Citizen Science: An Information Quality Research Frontier," Information Systems Frontiers (22:4), pp. 961–983. (https://doi.org/10.1007/s10796-019-09915-z).

Schupak, A. 2015. "Majority of Americans Fall for Email Phishing Scams," CBS News. (https://www.cbsnews.com/news/majority-of-americans-fall-for-email-phishing-scams-cbs-intel-security-quiz/).

Singh, S., Dutta, S. C., and Singh, D. K. 2016. "Information Security and Its Insurance in the World of High Rise of Cybercrime Through a Model," in Proceedings of Fifth International Conference on Soft Computing for Problem Solving, Advances in Intelligent Systems and Computing, M. Pant, K. Deep, J. C. Bansal, A. Nagar, and K. N. Das (eds.), Singapore: Springer, pp. 93–98. (https://doi.org/10.1007/978-981-10-0451-3_10).

Ye, G., and van Raaij, W. F. 2004. "Brand Equity: Extending Brand Awareness and Liking with Signal Detection Theory," Journal of Marketing Communications (10:2), pp. 95–114. (https://doi.org/10.1080/1352726041000169379 4).