Improving Cyber Risk Behavior through AI-Enabled Spearphishing – A Comparative Analysis

Carlo Pugnetti [1,*] and Palo Stacho [2]

[1] Institute of Financial Services Zug IFZ, Lucerne School of Business, Suurstoffi 1, 63434 Rotkreuz, Switzerland; carlo.pugnetti@hslu.ch
[2] Cyberdise AG, Poststrasse 26, 6300 Zug, Switzerland; palo.stacho@cyberdise.io
* Correspondence: carlo.pugnetti@hslu.ch

Abstract: Employee risk behavior is a critical component of a company's vulnerability, as well as of prevention of and response to cyber risks. Companies have traditionally impacted this behavior through communication, normative training and phishing exercises. However, the relative impact of training vs. experienced risk exposure is not well understood. This paper investigates and compares the effects of normative training, conventional phishing exercises and AI-enabled spearphishing exercises on cyber risk behavior and attitude. We observe actual rather than imputed risk behavior in Western European business setting to find that, while all measures improve risk behavior, exposure to spearphishing attacks is especially impactful. Additionally, it confirms a connection between training and risk attitude, but only a weak connection between attitude and behavior in a cyber context. The results also confirm both the feasibility and effectiveness of AI-enabled spearphishing in a Western European data and privacy environment. Our study strongly suggests that spearphishing attacks will become common also in settings with strong privacy laws and relatively poor availability of OSINT personal data, and that companies will need to incorporate spearphishing exercises and other measures to personalize risk messages in their cyber defenses in order to counteract and mitigate these developments.

Keywords: Cyber Security; Cyber Risk; Artificial Intelligence; Risk Behavior; Risk Exposure; Risk Mitigation; Cyber Security Awareness; Security Culture.

Academic Editor: Firstname Lastname

Received: date
Revised: date
Accepted: date
Published: date

Citation: To be added by editorial staff during production.

1. Introduction

Cyber attacks are a significant and growing threat to company operations. In the US, for example, the FBI reports some 860,000 complaints due to internet crime for losses of US$ 16.6 bn in 2024, and representing an increase of 33% with respect to 2023 (FBI 2025). The picture is consistent across many countries, with Switzerland for example reporting some 63,000 cyber incidents in 2024, for an increase of more than 27% with respect to 2023 (NCSC 2025). The World Economic Forum ranks cyber risks on fourth place in the short term, and on eighth place in the medium term (WEF 2024), while two recent studies in Switzerland rank Cybersecurity as the top technology-related topic for both the insurance and the banking industries; in both cases ahead of AI (IFZ 2025, Blattmann et al. 2025). Minimizing the threat, therefore, both in terms of minimizing the probability of occurrence as well as mitigating the consequences of these attacks is a critical issue for organizations across the world. In addition to an up-date IT infrastructure and processes, a significant component of this

defense is the behavior of employees across the organization. These employees can both open avenues of attack and constitute a significant resource for mitigating responses depending on their risk behavior.

Several personality traits and decision-making styles have been linked with cyber risk behavior. However, these findings are non-actionable since both drivers are difficult to influence. Risk attitude on the other hand can be more readily impacted, and several training programs have aimed to do so. These programs have tended to focus on imparting knowledge, and the link between knowledge and action has been proven to be tenuous (Chaudhary et al. 2023). Most mitigation efforts also include phishing exercises to expose employees to simulated threats and practice behavior. These efforts have been successful; however, they are costly and improvements tend to be short-lived if not repeated (Reinheimer et al. 2020). Further, new technologies leveraging AI are starting to appear, creating an increasingly dynamic threat environment (Heiding et al. 2024).

Against this backdrop we aim to further the research and improve cybersecurity by investigating the relative effectiveness of different cyber risk mitigation strategies aimed at employee behavior, and address the following research question:

How do normative training, conventional phishing, and AI-spearphishing impact employees' cyber risk behavior and their attitude towards cyber risks?

Two broader themes in this context are a) the effects of exposure to risk vs. normative cybersecurity training, as well as b) the feasibility and risk posed by AI-enabled spearphishing campaigns in a Western European setting. An additional important contribution is using observed rather than imputed risk behavior to measure these effects.

2. Review of Current Literature

Cyber risk behavior and its drivers have been investigated in several studies. These efforts, however, can be further improved and extended. At a fundamental level, Hadlington (2017) links self-assessed risky cybersecurity behavior to internet addiction, impulsivity and attitude towards cyber risks, while Halevi et al. (2015) find that women and more conscientious subjects are more likely to respond to spearphishing attacks. These results, however, are not supported by Gratian et al. (2018). In their analysis of the effects of demographic factors, personality traits, risk-taking preferences, and decision-making styles on four dimensions of cybersecurity behavior intentions, gender was shown to have a significant impact across factors, with women significantly less likely to engage in risky behavior. Ethical and health/Ssafety risk-taking was most readily linked to cyber risk behavior, as was an avoidant decision-making style. A rational decision-making style, on the other hand was linked to less risky behavior, as was the conscientiousness personality trait. The latter, however, not consistently. The perception of risk and its influence on cyber security behavior have also been investigated. Gillam and Foster (2020) investigate self-reported cybersecurity behavior leveraging Protection Motivation Theory PMT (Rogers 1975) and the subsequent Technology Threat Avoidance Theory TTAT (Liang and Xue 2009). In a survey of 184 working US adults, they find that cybersecurity behavior has significant predictive associations with perceived susceptibility, cost, and self-efficacy, but that these factors explain less than 10% of the variance. Debb and McClennan (2021), on the other hand, leverage PMT in a survey of 612 US college students to find that self-reported risk behavior correlates with perceived vulnerability, and that perceived vulnerability is in turn influenced by self-assessed computer skills, internet skills, and security efficacy, and also by prior experiences, as well as perceived benefits and severity. While insightful, these efforts focus on factors that cannot be easily influenced. Cybersecurity theories and models are not yet sufficiently

developed and integrated, most studies concentrate on college students, and the impact is measured in terms of self-assessment or intention of risk behavior (Alsharida et al. 2023, Almansoori et al. 2023). Kannelønning and Katsikas (2023) similarly observe that only one article in their literature review uses only objective measures to assess cybersecurity behavior. To address these issues, building on previous research (Pugnetti and Bekaert 2018, Pugnetti and Casián 2021) Björck et al. (2024) and Pugnetti et al. (2024) develop a diagnostic to quantify employee attitudes towards cyber risks along seven dimensions, link these to observed risk behavior and suggest a methodology using risk communication frameworks to influence risk attitude and behavior.

Phishing constitute one of the most common attack vectors. Alkhalil et al. (2021) catalogue different types of phishing attacks and survey available countermeasures. They identify three lines of defence: first, human-based solutions educating users to recognize and avoid phishing emails. Second, technical solutions to prevent the threat from materializing at the user's device. Third, law enforcement as a deterrent control. Specifically for the human component, they suggest training to raise awareness, using mock phishing attacks to test vulnerabilities and assess own knowledge, and gamifying phishing recognition training. Similarly, Naqvi et al. (2023) survey the literature to understand mitigation strategies against phishing attacks. Out of the 248 studies investigated, only 33 took a human-centric approach, while the rest concentrated on technical solution with more than half leveraging machine learning. Varshney et al. (2024) note the continuing importance of the phishing threat and develop comprehensive lists of attack types, social and cognitive factors behind phishing, anti-phishing schemes, as well as anti-phishing organizations and laws. While the factors leading to successful phishing attacks listed are all social and cognitive, the solutions investigated are purely technical. Desolda et al. (2021) investigate extant literature specifically on the topic of human factors in phishing attacks to identify a complex and rapidly evolving threat environment combining several mediums, vectors, and technical approaches targeting twelve human factors ranging from complacency to stress contributing to cyber behavioral vulnerabilities. They identify four types of solutions to address these vulnerabilities: a) improving the user interface to inform of potential danger, b) changing attitude, behavior and psychological aspects, for example improving employee attitudes towards security policies, c) focus on knowledge, education, and training, and d) further develop frameworks, models and taxonomies to understand human factor-related issues. Vulnerabilities are also anchored in common misconceptions about phishing among users. Mossano and Volkamer (2025) catalog 14 such misconceptions and link them to either lack of technical knowledge or confusion as overarching themes. However, the actual dynamics driving human decisions when confronted with phishing messages is not yet sufficiently understood, and Gallo et al. (2024) develop a system to collect and analyze detailed user behavior in this context with the long-term goal of tailoring mitigating measures to type of attack and individual characteristics. Thus, human factors are recognized as an important component of successful cyber attcks and need to be addressed in order to improve security.

Lain et al. (2022) conduct a long-term, large-scale experiment (14,000 participants over 15 months) to understand the susceptibility to phishing attacks in large organizations. The results confirm and validate several previous results regarding age and computer skills, but not gender, and the effectiveness of warnings delivered with suspicious emails, but that more complicated warnings are not more effective than simpler ones. In addition, they call into question the effectiveness of the current practice of embedding training in phishing exercises. Phishing education, training and awareness is a complex endeavour, and Sarker et al. (2024) identify from the literature 20 challenges and 23 critical success factors to design, implement and improve such a program. Programs aiming at raising cybersecurity awareness have typically only managed to raise employee knowledge. This is in large part because these programs are conceived as normative training exercises rather than with the goal of

impacting behavior (Chaudhary et al. 2023). Nonetheless, most studies on cybersecurity training reported a positive impact. Prümmer et al. (2024) note several critical observations on these studies, including reliance on common sense rather than on a solid theoretical underpinning, working with small sample sizes, and in most cases testing intention and knowledge rather than behavior. While knowledge is an important first step, behavioral change is necessary, and the two tend to be only loosely linked (Zwilling et al. 2022). Chaudhary (2024) identifies seven principles for drafting cybersecurity messages to drive behavioral change. To this end, companies routinely conduct exercises to train employees to recognize simulated phishing attacks. Hillman et al. (2023) evaluate the effectiveness of these efforts in a controlled experiment of financial institutions. Using three waves of phishing attacks with embedded training over a six-months period, they observed a decrease in the click rate from 25% to 7% and an improvement in the reporting rate from 8% to 18%. The phrasing used in the message did not impact these results; however, more personalized messages were more successful. Thus, phishing exercises are effective and, while advanced technologies can mitigate attacks, employee awareness and behavior will continue to play a critical role. Braun et al. (2025) compare the results of phishing simulations in 36 companies over three years to determine that the threat posed by phishing attacks depends on a number of factors. First, there are significant differences between sectors and individual companies, and among departments in the same company, meaning that results are likely not generalizable. However, personalized phishes, phishes using a "don't miss out" message, and phishes received in stressful periods were most likely to generate dangerous engagement. In addition to this complexity, the improvement these programs provide can be fleeting and lasting in general at most six months (Reinheimer at al. 2020, Berens et al. 2022). This need for repeated, ongoing training independently of technology developments leads to high costs for security awareness campaigns, most of which are in the form of employee time and attention, and are thus not captured in traditional cost estimates (Brunken et al. 2023). Thus, while phishing exercises are more effective than trainings, improvements vary widely and are temporary, indicating the opportunity for further improvements.

The accelerating pace of AI, linked with a broader availability of personal data on open platforms and compromised closed platforms is opening the door for new and potentially more devastating attack vectors. Hazell (2023) pilots the creation of spearphishing attacks using Large Language Models (LLMs) on British members of parliament based on publicly available information to find these attacks realistic and cost effective. Heiding et al. (2024) train AI agents on students' publicly available information to generate truly worrisome results. Useful and accurate information was available in 88% of the cases, and the AI generated spearphishing attacks that were equally as effective as tailored attacks generated by human experts, and for a fraction of the costs. In addition, current phishing detection software based on keyword and pattern recognition is not able to identify these emails as malicious. Both authors suggest the further development of defensive AI tools. The threat environment is likely to further evolve with the inclusion of visual and audio deepfakes to mimic trusted actors. The costs and data required to train these models have dropped significantly and real-time deepfakes have begun to appear. Automatic detection of deepfakes is challenging (Masood et al. 2023). Thus, technological advances will further develop the capability of malicious actors to take advantage of human vulnerabilities, further raising the importance of human-based mitigation mechanisms supported by technical mitigation efforts.

3. Methodology

Our experiment investigates the relative effects of normative training, conventional phishing and AI-enabled spearphishing on employee risk behavior and risk attitude. The experiment took place with the support of GLB Genossenschaft, a construction company

based in the Canton of Berne in Switzerland, between the end of 2024 and mid 2025. All employees with access to a computer and email were assigned at random to one of four intervention groups. These groups were then updated in the course of the year as new employees joined the company and others left, so that the final groups that experienced the entire process were slightly different in size. The group sizes, project plan, and interventions are summarized in Figure 1.

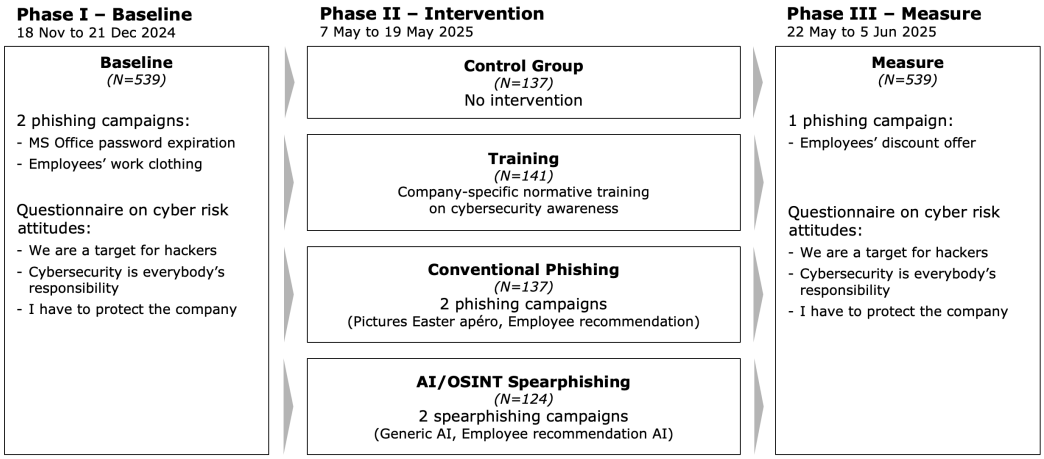

Figure 1. Experimental setup

All employees were exposed to a baseline phishing campaign consisting of two conventional phishes, the first warning of the expiration of the MS Office password, the second offering discounted work clothing for employees. The phishing was followed by a diagnostic questionnaire focusing on three of the dimensions describing attitudes towards cyber risks identified by Björck et al. (2024) and investigated by Pugnetti et al. (2024), namely: 1. We are a target for hackers, 2. Cybersecurity is everybody's responsibility, and 3. I have to protect the company. These were tested following the same reduced set of five questions per dimension. This baseline phase occurred between November 18th and December 21st, 2024.

Following a relatively lengthy time period of capability development and testing to ensure a credible spearphishing campaign based on the relatively scarce OSINT data available on the employees, the actual intervention occurred between May 7th and May 19th, 2025. The control group had no interaction with the project team and was naturally exposed to company communication and broader political events. The second group repeated an existing training program on cybersecurity awareness. A third group was exposed to two conventional phishing campaigns, the first promising access to the pictures taken at a recent company-wide event for Easter, and the second requesting information on colleague applying to a job within the company. Both campaigns were well-crafted, including the correct company logo, link to actual company events, the correct name of the hiring manager and an almost legitimate url address for the landing page. The last group was exposed to an automatically generated spearphishing campaign leveraging publicly available information on the individual employees. The first AI-spearphish chose among available information to generate a message of potential interest to the target, for example using residence(s), hobbies, or specialization. The second again asked for an employee recommendation, but the message was generated automatically from company data. While the messages were personally targeted, the landing page was generic, for example without the company logo and minimal additional information about the purported offer. Possibly because of these differences, employee responses to these campaigns were markedly different. Employees visited the sites from the conventional phishing campaign in 22.3% and 25.0% of the cases, and entered login information in 7.4% and

14.2% of the cases, respectively. Employees that were part of the spearphishing campaign were considerably more risk averse, visiting the website in 14.8% and 9.6% of the cases, and entered login information in 3.0% and 3.7% of the cases, respectively.

Finally, all employees were once again phished to measure the effect, if any, of the previous interventions. All employees were subsequently again surveyed to measure any changes to risk attitude, again along the same three dimensions and following the same process as in the baseline phase. These measurements occurred between May 22nd and June 5th, 2025.

4. Results

The results of the baselining, intervention on and measurement of risk behavior are summarized in Table 1 for each of the four experimental groups. The result "visited" refers to employees clicking on the link, while "phished" means that employees entered login information. The baseline results were consistent across all groups, hovering at around 21% of employees visiting the malicious site, while 10% and 8% entered login information respectively. While these results are objectively high, they are in line with existing literature and practical experience. The results in the Measure phase, on the other hand, were significantly different among the four experimental groups. The control group shows a behavior similar to the baseline, with a 20% visited rate and 10% phished rate. The training group is significantly less likely to visit the malicious site and enter data (approximately 12% and 5.5% respectively). The group experiencing conventional phishing behaved similarly (approximately 11% and 6.5% respectively), while the group exposed to AI-enabled spearphishing exhibited an even more risk-averse behavioral impact, with approximately 9% visiting the malicious site and 4% entering login data. The differences between the groups are statistically significant for visiting the site, and just above the 0.1 threshold for entering data.

Table 1. Observed risk behavior

			Control	**Training**	**Conventional Phish**	**AI Spearphish**	
		N	137	141	137	124	
Baseline							*p*-values
Phish 1	Visited	%	22.6	21.3	19.7	22.6	*0.93*
	Phished	%	10.2	9.2	9.5	11.3	*0.95*
Phish 2	Visited	%	19.0	25.5	19.0	23.4	*0.46*
	Phished	%	8.0	7.8	5.8	9.7	*0.72*
Measure							*p*-values
Phish	Visited	%	20.4	12.1	10.9	8.9	*0.03*
	Phished	%	9.8	5.4	6.6	3.9	*0.14*
Delta Measure vs. Baseline (*p*-values)							
	Visited		*0.76*	*0.01*	*0.10*	*<0.01*	
	Phished		*0.70*	*0.53*	*0.47*	*0.09*	

An additional analysis providing insight into the result is the difference between baseline phase and measurement phase results for each one of the experimental groups. The results for the control group were consistent across the two phases although the phishing exercises were different. The results for the other groups, on the other hand, are different between the two phases. The differences are statistically significant for all groups

for visiting the website, but only the group that was spearphished shows statistically significant improvements for entering login data. Thus, all interventions have observable, statistically significant effects. These results are more visible for the initial response to malicious emails, and less so when faced with security-relevant login information. A working hypothesis is that the behavior of people who are likely to fully fall for phishing attacks are generally more difficult to influence. On the other hand, employees who are generally more aware of cyber risks can be more readily reminded to consider the source of questionable emails and act accordingly.

An additional lens through which we can understand the dynamics of this behavioral change is to observe the responses to the interventions themselves. In our experiment, employees fell for AI-enabled spearphishing messages significantly less frequently than for the traditional, manually developed phishing attacks. This is likely because of quality issues in OSINT data and of the quality of the landing page. Thus, only some 3-4% of respondents entered login data following the spearphishing attacks vs. the 7-14% for the traditional phishing attacks. In spite of this apparent lack of sophistication, AI-spearphishing attacks were more effective at influencing employee risk behavior. Indeed, a potential interpretation is that this lack of sophistication may be an unintended driver of this effectiveness. The authors' working hypothesis is that respondents realize that they are being targeted as individuals, based on some kind of personal information, rather than just as an employee of the company. This realization triggers a more significant behavioral change.

Additionally, we investigated attitudes towards cyber risks along three dimensions, as shown in Table 2, where higher results indicate higher awareness on a scale from 1, fully disagree, to 7, fully agree. Generally, risk attitudes improved for all experimental groups. However, only the improvements for the training group are statistically significant. These results challenge previous research insights (e.g., Pugnetti et al. 2024), and potentially indicates different mediating effects for training vs. risk exposure to impact risk behavior. Thus, while it can be hypothesized that risk attitude impacts risk behavior in the training group, the same cannot be said for the groups exposed to conventional phishing or spearphishing. These groups maintain the same attitude towards cyber risks but alter their behavior significantly. Thus, two different mechanisms for behavioral change may be activated and combining the interventions into a coherent cybersecurity program may deliver additive or perhaps even super-additive results.

Table 2. Measured attitudes towards cyber risks.

		Control	**Training**	**Conven. Phish**	**AI Spearph.**	
	N	137	141	137	124	
Baseline						*p*-values
We are a target for hackers		5.26	5.34	5.64	5.62	*0.46*
Cybersecurity is everybody's responsibility		5.99	5.85	6.04	6.24	*0.28*
I have to protect the company		6.14	6.14	6.14	6.39	*0.42*
Measure						*p*-values
We are a target for hackers		5.73	5.82	5.74	5.89	*0.94*
Cybersecurity is everybody's responsibility		6.21	6.24	6.10	6.35	*0.61*
I have to protect the company		6.26	6.45	6.33	6.45	*0.49*
Delta Measure vs. Baseline		*p*-values				
We are a target for hackers		*0.13*	*0.05*	*0.74*	*0.38*	
Cybersecurity is everybody's responsibility		*0.26*	*0.02*	*0.78*	*0.62*	
I have to protect the company		*0.40*	*0.04*	*0.30*	*0.77*	

5. Conclusions and Next Steps

The flexibility and proactive support of our partners in this research effort allowed the study to investigate the impact of three different intervention measures with a sufficiently large sample size and compare the results to a control group in an identical cultural and commercial environment. This allowed us to generate five important insights.

Behavior and attitude are distinct

The first critical insight is the confirmation that behavioral change and attitude towards cyber risks are different and distinct. While training impacted risk attitude, it impacted behavior least among the measured interventions. The converse is also true: risk exposure, especially to personalized spearphishing exercises impacted behavior most but did not impact attitude in a significant way. This result agrees with several findings in the current literature (e.g., Chaudhary et al. 2023), which show knowledge about cyber risks and appropriate risk-averse behavior are not always connected. In practice, however, most cyber awareness training programs are designed to impart knowledge and not directly change behavior. This is a fundamental gap in cyber defense and prevention that will need to be filled, and our study provides an indication of just how much vulnerability can be reduced by doing so.

Spearphishing exercises as an important preventive cyber defense tool

Perhaps the most striking result of the study is the large behavioral change effected by spearphishing exercises. While the reasons for this effect are not immediately clear from our study, we postulate that this is due to the raised awareness with the recipient that their personal information is available somewhere, and that it is being used for nefarious purposes. This is a fundamentally different experience than receiving a more general message about company events. This effect can and should be leveraged in risk mitigation exercises.

AI/OSINT spearphishing is an emerging threat

Despite the relative dearth of OSINT data and the specialized profile of the partner company and therefore of its employees, it was possible in a relatively short time to develop successful spearphishing attacks. The relative low click and data entry rate achieved can be clearly improved with more sophisticated messaging and landing pages. These factors, combined with the inherent scalability and therefore cost-effectiveness of AI tools indicates that spearphishing will be an emerging threat also in relatively less data and target-rich environments in Central and Western Europe.

Traditional cyber defense programs

While AI-spearphishing provides the largest behavioral improvements, normative training and traditional phishing exercises are also associated with a positive impact. Thus cybersecurity awareness programs should continue to integrate both of these elements. A question not addressed in our study pertains to the optimal combination of all elements in a coherent cyber defense program. Sequence, frequency and contents need to be optimized considering impact, costs, complementarity and employee engagement. Thus, the solution may potentially be company-specific rather than general.

Limited OSINT data availability

Compared with reviewed studies in the United States, the quality of the data we were able to collect was significantly inferior. While Heiding et al. (2024), for example was able to find high quality data for almost 90% of subjects, we estimate our success rate was

around 50%. The most common issue was the case of homonyms, where employees shared a name with people with a higher online profile. We corrected the issue by including "Switzerland" as a filtering criterion to improve data quality. This indicates an additional venue of risk mitigation for companies, educating and guiding employees' online footprint. While companies will not be able to fully control employee online presence, except perhaps for a few security-critical exceptions, a reduction and curation of available data should help at least partially defuse the emerging threat of AI spearphishing.

5.1 Implications for CISO / Cybersecurity officers

The insights above can be leveraged immediately in practice to improve cyber defenses. First, programs should include AI-enabled spearphishing exercises in the portfolio of activities. Even in situations where employees do not keep a significant online profile, it should be possible to generate credible messages. The engines to do so are available from vendors or can be developed in-house, and they are scalable and highly effective at improving behavior. Second, programs should continue to incorporate a variety of exercises, including normative training, phishing exercises, and general communication from senior executives to emphasize the importance of the issue. This is important in order to keep awareness high, influence attitude towards cyber risks, and impact behavior over longer time horizons. Third, while not directly linked to the results of the study programs should link into other current activities. Thus, for example, employees should be activated to detect emerging threats, for example by reporting suspicious messages with a reporting button, praire-dogging to alert colleagues and activate colleagues in the IT department with any observations, questions and concerns. Fourth, cybersecurity programs need to be tailored to each company's culture and work environment. Thus, the content and tone of the messages needs to reflect the environment in which the exercises are conducted as well as the tools and working and security guidelines of the company.

5.2 Implications for future research

While the results provide useful and actionable insights, a number of questions remain open for future research. First, the study investigated single interventions, and a critical open question will be how to integrate multiple such interventions into a program. Optimal sequence, frequency and contents will need to be investigated. Second, it would be useful to test additional interventions. For example, risk communication interventions and interventions leveraging multichannel and deepfake exercises will need to be tested. Third, the rate of decay of behavioral change will need to be investigated. While current research shows a complete effectiveness decay after six months, it would be interesting to see if personalized, spearphishing messages are effective over a longer time period. Fourth, we encountered organizational pushback towards the end of the study, primarily due to employee fatigue. Outside of the project, the authors have encountered cases of malicious noncompliance, where employees entered data in phishing emails on purpose out of frustration with cyber security training. Thus, an important question is the improvement of employee motivation and cooperation, and exercises will need to be designed also with this question in mind. A fifth and last area for additional research is the open questions of how training and risk exposure effect behavior. While training impacts attitude and thus in turn behavior, risk exposure impacts behavior without a tangible effect on attitude. A potential explanation is a spurious result in our data set; however, a more intriguing explanation could link to a different mediating mechanism. This should be analyzed in a new research effort.

Author Contributions: Conceptualization, C.P. and P.S.; methodology, C.P; software, P.S.; formal analysis, C.P.; resources, P.S.; data curation, P.S.; writing—original draft preparation, C.P.;

writing—review and editing, C.P. and P.S.; visualization, C.P.; funding acquisition, C.P. and P.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded in part by Innosuisse - Schweizerische Agentur für Innovationsförderung, grant Innocheck 73275.1 INNO-ICT.

Data Availability Statement: Data available upon request from the authors.

Acknowledgments: We gratefully acknowledge the support of GLB Genossenschaft, in particular I.A. and T.K., as well as the important contribution of S.M.

Conflicts of Interest: The authors declare no conflicts of interest.

References

(Alkhalil et al. 2021) Alkhalil, Zainab, Chaminda Hewege, Liqaa Nawaf, and Imtiaz Khan. 2021. Phishing Attacks: A Recent Comprehensive Study and New Anatomy. *Frontiers in Computer Science*, 3, 563060. doi: 10.3389/fcomp.2021.563060.

(Almansoori et al. 2023) Almansoori, Afrah, Mostafa al-Emran, and Khaled Shaalan. 2023. Exploring the Frontiers of Cybersecurity Behavior: A Systematic Review of Studies and Theories. *Applied Sciences*, 13, 5700. doi: 10.3390/app13095700.

(Alsharida et al. 2023) Alsharida, Rawan, Bander Ali Saleh Al-rimy, Mostafa Al-Emran, and Anazida Zainal. 2023. A systematic review of multi perspectives on human cybersecurity behavior. *Technology in Society*, 73, 102258. doi: 10.1016/j.techsoc.2023.102258.

(Berens et al. 2022) Berens, Benjamin, Katerina Dimitrova, Mattia Mossano, and Melanie Volkamer. 2022. Phishing awareness and education – When to best remind? In *Workshop on Usable Security and Privacy (USEC)*. Network and Distributed System Security (NDSS) Symposium 2022. doi: 10.14722/usec.2022.23075.

(Björck et al. 2024) Björck, Albena, Carlo Pugnetti, and Carlos Casián. 2024. Communicating to Mitigate Behavioral Cyber Risks: The Case of Employee Vulnerability. In *Communicating Risk and Safety*. Edited by Timothy L. Sellnow and Deanna D. Sellnow. Berlin and Boston: De Gruyter Mouton, pp. 585–606. doi: 10.1515/9783110752427.

(Blattmann et al. 2025) Blattmann, Urs, Thomas Fischer, Joel Ettlin. 2025. IFZ Studie Bank-IT und Sourcing 2025. Hochschule Luzern Wirtschaft, Rotkreuz, Switzerland. ISBN: 978-3-907379-51-6. Available at: https://blog.hslu.ch/bankingservices/.

(Braun et al. 2025) Braun, Oskar, Jan Hörnemann, Norbert Pohlmann, Tobias Urban, and Matteo Grosse-Kampmann. 2025. Different Seas, Different Phishes – Large-Scale Analysis of Phishing Simulations Across Different Industries. *ACM Asia Conference on Computer and Communication Security (ASIA CCS '25).* August 25-29, Hanoi, Vietnam. doi: 10.1145/3708821. 3733905.

(Brunken et al. 2023) Brunken, Lina, Annalina Buckmann, Jonas Hielscher, and M. Angela Sasse. 2023. To Do This Properly, You Need More Resources: The Hidden Costs of Introducing Simulated Phishing Campaigns. *In 32nd USENIX Security Symposium (USENIX Security 23),* August 9-11, Anaheim, CA, USA, pp. 4105-4122. ISBN: 978-1-939133-37-3.

(Chaudhary et al. 2023) Chaudhary, Sunil, Vasileios Gkioulos, and Sokratis Katsikas. 2023. A quest for research and knowledge gaps in cybersecurity awareness for small and medium-sized enterprises. *Computer Science Review*, 50, 100592. doi: 10.1016/j.cosrev.2023.100592.

(Chaudhary 2024) Chaudhary, Sunil. 2024. Driving behaviour change with cybersecurity awareness. *Computers & Security*, 142, 103858. doi: 10.1016/j.cose.2024.103858.

(Debb and McClennan 2021) Debb, Scott, and Marnee McClennan. 2021. Perceived Vulnerability as a Determinant of Increased Risk for Cybersecurity Risk Behavior. *Cyberpsychology, Behavior, and Social Networking*, 24, 9. doi: 1089/cyber.2021.0043.

(Desolda et al. 2021) Desolda, Giuseppe, Lauren Ferro, Andrea Marrella, Tiziana Catarci, and Maria Francesca Costabile. 2021. Human Factors in Phishing attacks: A Systematic literature Review. *ACM Computing Surveys*, 54, 8, 173. doi: 10.1145/3469886.

(FBI 2025) Federal Bureau of Investigation. 2025. Internet Crime Report 2024. Available at: https://www.ic3.gov/AnnualReport/Reports/2024_IC3Report.pdf.

(Gallo et al. 2024) Gallo, Luigi, Danilo Gentile, Saverio Ruggiero, Alessio Botta, and Giorgio Ventre. 2024. The human factor in phishing: Collecting and analyzing user behavior when reading emails. *Computers & Security*, 139, 103671. doi: 10.1016/j.cose.2023.103671.

(Gillam and Foster 2020) Gillam, Andrew, and W. Tad Foster. 2020. Factors affecting risky cybersecurity behaviors by U.S. workers: An exploratory study. *Computers in Human Behavior*, 108, 106319. doi: 10.1016/j.chb.2020.106319.

(Gratian et al. 2018) Gratian, Margaret, Sruthi Bandi, Michel Cukier, Josiah Dykstra, and Amy Ginther. 2018. Correlating human traits and cybersecurity behavior intentions. *Computers & Security*, 73, 345-358. doi: 10.1016/j.cose.2017.11.015.

(Hadlington 2017) Hadlington, Lee. 2017. Human factors in cybersecurity; examining the link between Internet addiction, impulsivity, attitudes towards cybersecurity, and risky cybersecurity behaviours. *Heliyon 3*, 7. doi: 10.1016/j.heliyon.2017.e00346.

(Halevi et al. 2015) Halevi, Tzipora, Nasir Memon, Odev Nov. 2015. Spear-Phishing in the Wild: A Real-World Study of Personality, Phishing Self-efficacy and Vulnerability to Spear-Phishing Attacks. doi: 10.2139/ssrn.2544742.

(Hazell 2023) Hazell, Julian. 2023. Large Language Models Can Be Used to Effectively Scale Spear Phishing Campaigns. *arXiv preprint arXiv:2305.06972*.

(Heiding et al. 2024) Heiding, Fred, Simon Lermen, Andrew Kao, Bruce Schneier, and Arun Vishwanath. 2024. Evaluating Large Language Models' Capability to Launch Fully Automated Spear Phishing Campaigns: Validated on Human Subjects. *arXiv preprint arXiv:2412.00586*.

(Hillman et al. 2023) Hillman, Doron, Yaniv Harel, and Eran Toch. 2023. Evaluating organizational phishing awareness training on an enterprise scale. *Computers & Security*, 132, 103364. doi: 10.1016/j.cose.2023.103364

(IFZ 2025) Institute of Financial Services Zug. 2025. IFZ Versicherungsstudie 2025. Edited by Florian Schreiber. Hochschule Luzern Wirtschaft, Rotkreuz, Switzerland. Available at: https://www.hslu.ch/de-ch/hochschule-luzern/forschung/projekte/detail/?pid=5926.

(Kannelønning and Katsikas 2023) Kannelønning, Kristian, and Sokratis Katsikas. 2023. A systematic literature review of how cybersecurity-related behavior has been assessed. *Information & Computer Security*, 31, 463-477. doi: 10.1108/ICS-08-2022-0139.

(Lain et al. 2022) Lain, Daniele, Kari Kostiniainen, and Srdjan Čapkun. 2022. Phishing in Organizations: Findings from a Large-Scale and Long-Term Study. *2022 IEEE Symposium on Security and Privacy (SP)*. doi: 10.1109/SP46214.2022.9833766.

(Liang and Xue 2009) Liang, Huigang, and Yajiong Xue. 2009. Avoidance of information technology threats: A theoretical perspective. *MIS Quarterly*, 33, 71-90.

(Masood et al. 2022) Masood, Momina, Mariam Nawaz, Khalid Mahmood Malik, Ali Javed, Aun Irtaza, Hafiz Malik. 2022. Deepfakes generation and detection: state-of-the-art, open challenges, countermeasures, and way forward. *Applied Intelligence*, 53, 3974-4026. doi: 10.1007/s10489-022-03766-z.

(Mossano and Volkamer 2025) Mossano, Mattia, and Melanie Volkamer. 2025. Literature Review: Misconceptions About Phishing in *International Symposium on Human Aspects of Information Security and Assurance*, 215-228. Springer, Cham, 2025. doi: 10.1007/978-3-72559-3_15.

(Naqvi et al. 2023) Naqvi, Bilal, Kseniia Perova, Ali Farooq, Imran Makhdoom, Shola Oyedeji, and Jari Porras. 2023. Mitigation strategies against the phishing attacks: A systematic literature review. *Computers & Security*, 132, 103387. doi: 10.1016/j.cose.2023.103387.

(NCSC 2025) National Cyber Security Centre. 2025. Cybersecurity – Situation in Switzerland and internationally. Semi-Annual Report 2024/II (July-December). Available at: https://www.ncsc.admin.ch/ncsc/en/home/dokumentation/berichte/lageberichte/halbjahresbericht-2024-2.html.

(Pugnetti et al. 2024) Pugnetti, Carlo, Albena Björck, Reto Schönauer, and Carlos Casián. 2024. Towards Diagnosing and Mitigating Behavioral Cyber Risks. *Risks*, 12, 116. doi: 10.3390/risks12070116.

(Pugnetti and Bekaert 2018) Pugnetti, Carlo, and Xavier Bekaert. 2018. *A Tale of Self-Doubt and Distrust. Onboarding Millennials: Understanding the Experience of New Insurance Customers*. ZHAW School of Management and Law, ISBN 978-3-03870-021-0.

(Pugnetti and Casián 2021) Pugnetti, Carlo, and Carlos Casián. 2021. *Cyber Risks and Swiss SMEs: An Investigation of Employees' Attitudes and Behavioral Vulnerabilities*. ZHAW School of Management and Law. doi: 10.21256/zhaw-21478.

(Prümmer et al. 2024) Prümmer, Julia, Tommy van Steen, and Bibi van den Berg. 2024. A systematic review of current cybersecurity training methods. *Computers & Security*, 136, 103585. doi: 10.1016/j.cose.2023.103585.

(Reinheimer et al. 2020) Reinheimer, Benjamin, Lukas Aldag, Peter Mayer, Mattia Mossano, and Reyhan Duezguen. 2020. An Investigation of phishing awareness and education over time: When and how to best remind users in *Sixteenth Symposium on Usable Privacy and Security (SOUPS 2020)*, 259-284. ISBN 978-1-939133-16-8.

(Rogers 1975) Rogers, Ronald. 1975. A protection motivation theory of fear appeals and attitude change. *The Journal of Psychology*, 91, 93-114. doi: 10.1080/00223980.1975.9915803.

(Sarker et al. 2024) Sarker, Orvila, Asangi Jayatilaka, Sherif Haggag, Chelsea Liu, and M. Ali Babar. 2024. A Muti-vocal Literature Review on challenges and critical success factors of phishing education, training and awareness. *The Journal of Systems & Software*, 208, 111899. doi: 10.1016/j.ss.2023.111899.

(Varshney et al. 2024) Varshney, Gaurav, Rahul Kumawat, Vijay Varadharajan, and Uday Tupakula. 2024. Anti-phishing: A comprehensive perspective. *Expert Systems With Applications*, 238, 122199. doi: 10.1016/j.eswa.2023.122199.

(WEF 2024) World Economic Forum. 2024. The Global Risk Report 2024 – Insight report. Available at: https://www.weforum.org/publications/global-risks-report-2024/.

(Zwilling et al. 2022) Zwilling, Moti, Galit Klien, Dušan Lesjk, Łukasz Wiechetek, Faith Cetin, and Hamdullah Nejat Basim. 2022. Cyber Security Awareness, Knowledge and Behavior: A Comparative Study. *Journal of Computer Information Systems*, 2022, 62, 1, 82-87. doi: 10.1080/08874417.2020.1712269.