Leveraging AI-enabled spearphishing to enhance cybersecurity

Palo Stacho
Study Directors: Dr. Carlo Pugnetti (HSLU), Palo Stacho (Cyberdise AG)
Study Partners: Hochschule Luzern (HSLU), CYBERDISE Awareness AG, GLB Group

ABSTRACT

This study examined the effectiveness of AI-enabled spearphishing in comparison with conventional phishing and normative training within a European context. Across a three-phase design with 539 participants, we measured both risk attitudes (via questionnaires) and risk behaviors (via phishing simulations). The research period was from November 2024 to June 2025.

Findings show that normative training significantly improved employees' risk attitudes, fostering greater awareness and responsibility, while risk exposure - particularly through AI/OSINT spearphishing - produced the strongest behavioral improvements, reducing susceptibility by ~60%. Conventional phishing was nearly as effective but more resource-intensive.

In contrast to U.S.-based studies, European OSINT environments provided fewer employee data points, reducing the realism of AI-generated phish.

The results confirm that training and exposure address different but complementary aspects of cybersecurity: training shifts mindsets, exposure changes actions. Effective organizational security programs should therefore integrate both approaches.

Keywords: AI-Phishing, OSINT-Phishing, AI-Phishing Exercises, OSINT-Phishing Exercises, AI-Awareness, OSINT-Awareness

I. INTRODUCTION

Spearphishing is one of the most successful attack vectors in cybersecurity. With the rise of AI tools, attackers can cheaply generate personalized campaigns. The study investigated whether AI-enabled spearphishing is effective in Europe and compared its impact to normative training and conventional phishing. It builds on prior work emphasizing the need to distinguish between risk attitude and risk behavior (Pugnetti et al., 2024).

II. BACKGROUND AND RELATED WORK

Cyber threats are increasing in cost and frequency (FBI, 2024). Awareness training has shown limited behavioral effects (Prümmer et al., 2024). AI-driven spearphishing can automate campaigns with minimal cost (Heiding et al., 2024), while deepfake-enabled multichannel attacks pose further risks (Masood et al., 2023). Distinguishing between risk attitude and risk behavior is crucial (Pugnetti et al., 2024).

III. RESEARCH QUESTIONS

1. Can AI-enabled spearphishing be equally effective in the European OSINT environment?
2. How effective is risk exposure compared to normative training in shaping risk behavior and risk attitude?

IV. METHODOLOGY

The study followed a three-phase design: baseline (Nov–Dec 2024), intervention (May 2025), and re-measurement (May–Jun 2025). 539 participants were

divided into four cohorts: Control, Training, Conventional Phishing, and AI/OSINT Phishing. Risk behavior was measured by visit and data-entry rates in phishing exercises. Risk attitude was measured with questionnaires.

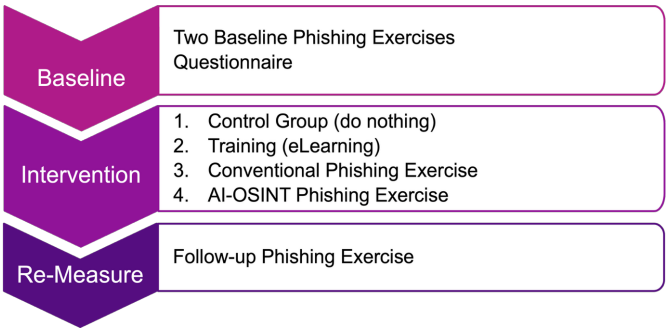

Figure 1: Experimental Study Setup

V. RESULTS

Baseline results showed and ~9% phished rates on average. Intervention results revealed that AI spearphishing reduced risky behavior most (~60%), conventional phishing ~45%, and training ~40%. Only training significantly improved risk attitudes. Re-measurement confirmed these patterns with significant group differences.

VI. DETAILED RESULTS

A. Baseline Phase

The baseline campaign consisted of two phishing campaigns and a questionnaire. Detailed phishing exercise results:

1. MS Office password expiration: 21.5% visited; 10.0% phished*
2. Employees' work clothing: 21.7% visited; 7.8% phished.

Questionnaire on cyber risk attitudes:

- We are a target for hackers
- Cybersecurity is everybody's responsibility
- I have to protect the company, etc.

B. Intervention Phase

- Convent. Phish 1: 22.3% visited; 7.4% phished*
- Convent. Phish 2: 25.0% visited; 14.2% phished
- AI Spearphish 1: 14.8% visited; 3.0% phished
- AI Spearphish 2: 9.6% visited; 3.7% phished

C. (Re-)Measure Phase

- Control Group: 20.4% visited; 10.9% phished*
- Training Group: 12.1% visited; 5.7% phished
- Convent. Phish Group: 10.9% visited; 6.6% phished
- AI/OSINT Phish. Group: 8.9% visited; 4.0% phished

Significant differences among the four intervention groups (p-values = 0.03 and 0.14).

The results of the re-measure phase for the control group were the same as those for the baseline phase.

**) of total, numbers not cumulative visited = site visited; phished = data entered*

VII. DISCUSSION

The study demonstrates a dual effect: training shifts attitudes, while exposure changes behavior. This supports Pugnetti et al. (2024). European OSINT limitations reduce AI phish realism compared to U.S. findings. Practical implications: combine both approaches for comprehensive security.

TABLE I
FINDINGS ON RISK ATTITUDE VS. RISK BEHAVIOR

Dimension	**Risk Attitude (mindset, perceptions)**	**Risk Behavior (observable actions)**
How measured	Questionnaire on statements such as: 1. We are a target for hackers 2. Cybersecurity is everyone's responsibility 3. I have to protect the company	Phishing simulations: visit rates and data entry ("phished") across baseline, intervention, and re-measurement phases
Effect of normative training	Significant improvement in attitudes — employees more strongly agreed with responsibility and awareness statements	Reduced susceptibility by ~40% (phished: ~11% → 6%)
Effect of conventional phishing exposure	No statistically significant change in attitudes	Reduced risky behavior by ~45% (phished: ~11% → 7%)

TABLE II
FINDINGS ON RISK ATTITUDE VS. RISK BEHAVIOR (CONTINUED)

Dimension	Risk Attitude (mindset, perceptions)	Risk Behavior (observable actions)
Effect of AI spearphishing exposure	No statistically significant change in attitudes	Strongest behavioral impact: ~60% reduction (phished: ~11% → 4%)
Control group	Small general improvement, not statistically significant	Behavior largely unchanged (phished ~11%)
Overall insight	Training is most effective for shaping mindsets — builds awareness and sense of responsibility	Exposure (especially AI spearphishing) is most effective for changing actions — strongly reduces susceptibility

VIII. CONCLUSION

AI spearphishing proved to be the most effective method for improving behavior, while training improved attitudes. Both are needed to strengthen organizational cybersecurity resilience:

1. AI phishes work best when it comes to improving behavior towards cyber risks and are also the cheapest.
2. Trainings drive the users mindset towards cybersecurity.
3. You need both: Phishing exercises and cybersecurity trainings.
4. The data obtained from European individuals through reconnaissance and the educational vulnerability profiles based on it are less rich than in comparable US studies. Regardless, the use of these profiles in AI/OSINT-based phishing simulations results in a 60% higher cybersecurity awareness than conventional phishing exercises.
5. A well-executed phishing simulation campaign having several attack emails has no more awareness effect after 5 months. Shorter exercise cycles are necessary to maintain a good cyber risk attitude.

IX. Implications for Practice and Product Development

Based on findings, CYBERDISE Awareness AG developed:

- AI-OSINT Reconnaissance
- Educational Vulnerability Profiles
- AI Phish Generator

These features have been rolled out in CYBERDISE V2.7.

- The finding, that e-learnings are more relevant as thought, the e-learning curriculum has also been updated, with a new version (CCC26) scheduled for 2025.

X. REFERENCES

[1] FBI. (2024). Internet Crime Report 2024.
[2] Heiding, F., Lermen, S., Kao, A., Schneier, B., & Vishwanath, A. (2024). Evaluating Large Language Models' Capability to Launch Fully Automated Spear Phishing Campaigns. arXiv:2412.00586v1.
[3] Masood, M., Nawaz, M., Malik, K. M., Javed, A., Irtaza, A., & Malik, H. (2023). Deepfakes generation and detection. Applied Intelligence, 53, 3974–4026. https://doi.org/10.1007/s10489-022-03766-z
[4] Prümmer, J., van Steen, T., & van den Berg, B. (2024). A Systematic Review of current cybersecurity training methods. Computers & Security, 136, 103585. https://doi.org/10.1016/j.cose.2023.103585
[5] Pugnetti, C., Björck, A., Schönauer, R., & Casián, C. (2024). Towards Diagnosing and Mitigating Behavioral Cyber Risks. Risks, 12(7), 116. https://doi.org/10.3390/risks12070116