

LMU Munich  
Frauenlobstraße 7a  
D-80337 München  
Institut für Informatik

Bachelorarbeit

# **Alerting Users to Phishing Threats: A Comprehensive Evaluation of Warning Techniques**

Yunus Emre Yavuz

yunus.yavuz@campus.lmu.de

<b>Course of Study:</b>	Informatik
<b>Examiner:</b>	Univ.-Prof. Dr. Florian Alt
<b>Supervisor:</b>	Dr. Verena Distler, Felix Dietz, M.Sc.

**Commenced:** 20.01.2024

**Completed:** 17.06.2024

“Alerting Users to Phishing Threats: A Comprehensive Evaluation of Warning Techniques”  
June 11, 2024

© 2024 Yunus Emre Yavuz

This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 License (CC BY-SA 4.0):  
<http://creativecommons.org/licenses/by-sa/4.0/>



Typesetting: PDF- $\text{\LaTeX}$  2 $\epsilon$

## Kurzfassung

Phishing-Angriffe sind trotz diverser Fortschritte im Bereich der Cybersicherheit nach wie vor eine große Bedrohung. Diese Arbeit befasst sich mit der Verbesserung der Erkennung von und Reaktion auf Phishing durch die Gestaltung und Implementierung visueller Warnungen in E-Mail-Clients. Über einen Zeitraum von zwei Wochen wurde eine Mixed-Methods-Studie mit 16 Teilnehmern durchgeführt, in der Eye-Tracking-Technologie und qualitatives Feedback integriert wurden, um die Interaktionen mit verschiedenen Phishing-Warnungen in Mozilla Thunderbird zu bewerten.

Die Studie ergab, dass Warnungen, die direkt in den E-Mail-Inhalt integriert sind, das Engagement der Benutzer und die Reaktionszeiten im Vergleich zu peripheren Warnungen deutlich verbessern. Eye-Tracking-Daten zeigten, dass unmittelbare und prominent platzierte Warnungen effektiver sind, während periphere Warnungen tendenziell erst später wahrgenommen werden, was ihre Effektivität in realen Szenarien verringern könnte. Qualitatives Feedback unterstrich die Bedeutung von Klarheit, Kontext und lehrreichen Inhalten in den Warnhinweisen, um das Verständnis der Benutzer zu verbessern und Phishing-Angriffe zu verhindern. Diese Ergebnisse legen nahe, dass die Gestaltung von Phishing-Warnungen erhebliche Auswirkungen auf die Cybersicherheit haben kann. Durch die Konzentration auf benutzerzentrierte Warnhinweise, die sowohl auffällig als auch informativ sind, können die Abwehrmechanismen von Phishing-Angriffen erheblich gestärkt werden. Diese Forschungsarbeit leistet einen Beitrag zum breiteren Feld der Cybersicherheit, indem sie evidenzbasierte Empfehlungen für die Gestaltung effektiver Phishing-Warnungen liefert, die darauf abzielen, die Prävalenz und die Auswirkungen von Phishing zu reduzieren.



## **Abstract**

Phishing attacks remain a significant threat despite advancements in cybersecurity. This thesis addresses the enhancing of user detection and response to phishing through the design and implementation of visual warnings within email clients. Over a two-week period, a mixed-methods study involving 16 participants was conducted, integrating eye tracking technology and qualitative feedback to assess interactions with various phishing warning designs in Mozilla Thunderbird.

The study found that warnings directly integrated within the email content significantly improve user engagement and response times compared to peripheral warnings. Eye tracking data revealed that immediate and prominently placed warnings are more effective, while peripheral warnings tend to be noticed later, potentially reducing their effectiveness in real-world scenarios. Qualitative feedback highlighted the importance of clarity, context, and educational content in warnings to enhance user understanding and prevent phishing attacks.

These findings suggest that the design of phishing warnings can have substantial implications for cybersecurity. By focusing on user-centric warning designs that are both noticeable and informative, phishing defense mechanisms can be significantly strengthened. This research contributes to the broader cybersecurity field by providing evidence-based recommendations for designing effective phishing warnings, aiming to reduce the prevalence and impact of phishing attacks across digital platforms.



# Contents

<b>1</b>	<b>Introduction</b>	<b>13</b>
<b>2</b>	<b>Background and Related Work</b>	<b>15</b>
2.1	Phishing Definition and Techniques . . . . .	15
2.2	Technical Detection Strategies . . . . .	16
2.3	User Education and Training . . . . .	16
2.4	Social Engineering and the Human Factor . . . . .	17
<b>3</b>	<b>Methodology</b>	<b>19</b>
3.1	Research Design . . . . .	19
3.2	Research Setting . . . . .	19
3.3	Recruitment of Participants . . . . .	19
3.4	Material . . . . .	19
3.5	Procedure . . . . .	21
3.6	Data Analysis . . . . .	23
<b>4</b>	<b>Design and Implementation</b>	<b>25</b>
4.1	Warning Types . . . . .	25
4.2	Visual Design Strategy and Rationale . . . . .	28
4.3	Iterative Design Adjustments . . . . .	28
4.4	Technical Background . . . . .	29
<b>5</b>	<b>Results</b>	<b>33</b>
5.1	Questionnaire Data . . . . .	33
5.2	Interview Results . . . . .	34
5.3	Eye Tracking Results . . . . .	36
<b>6</b>	<b>Discussion</b>	<b>39</b>
6.1	Interpretation of Qualitative Feedback . . . . .	39
6.2	Interpretation of Eye Tracking Results . . . . .	41
6.3	Limitations . . . . .	42
<b>7</b>	<b>Conclusion</b>	<b>45</b>
	<b>Bibliography</b>	<b>47</b>
<b>A</b>	<b>Appendix</b>	<b>51</b>
A.1	GitHub Repository . . . . .	51
A.2	Study Protocol . . . . .	51
A.3	Interview Questions . . . . .	53
A.4	Warning Designs . . . . .	54
A.5	QDA Codebook . . . . .	56
A.6	Eye Tracking Heatmap . . . . .	57





# List of Figures

3.1	Participants' environment in Mozilla Thunderbird. . . . .	20
3.2	An example for a mock phishing email. . . . .	20
3.3	Summary of the study procedure. . . . .	22
3.4	Graphic illustration of our QDA codes, generated in MAXQDA 2024. . . . .	23
3.5	AOI zones in Tobii Pro Lab. . . . .	24
4.1	Generic Banner . . . . .	25
4.2	Tooltip Warning on Link Hover . . . . .	26
4.3	Signature-Specific Warning . . . . .	27
4.4	Greeting-Specific Warning . . . . .	27
4.5	Mock phishing page when participants click on links. . . . .	30
5.1	Distribution of Affinity for Technology Interaction among participants. . . . .	34
5.2	Time to first fixation for the side AOI. . . . .	37
5.3	Comparison of the average fixation duration for center and side AOIs. . . . .	37
A.1	Generic Banner (Version 1) . . . . .	54
A.2	Tooltip Warning on Link Hover (Version 1) . . . . .	54
A.3	Signature-Specific Warning (Version 1) . . . . .	55
A.4	Greeting-Specific Warning (Version 1) . . . . .	55
A.5	Accumulated eyetracking heatmap, generated in Tobii Pro Lab . . . . .	57



List of Tables

5.1 Demographic data collected in the survey. . . . . 33

6.1 Overview of warning types and user feedback. . . . . 41



# 1 Introduction

In today's digital age, the internet is integral to daily activities such as communication, transactions, and accessing information. Yet, this increased dependency brings heightened security risks, especially from phishing attacks. Phishing, a prevalent cyber threat, involves deceiving individuals into revealing sensitive information by masquerading as a trustworthy entity in digital communications. There is a vast amount of phishing vectors, such as SMS, social media or email, with the latter being the most common medium for phishing attacks [30]. These attacks are used to direct victims to fraudulent websites where they are coerced into providing personal and financial details, allowing attackers unauthorized access to accounts [27]. The impact of phishing extends beyond individual victims to organizations worldwide, causing significant financial losses and eroding trust in digital systems. In 2023, the Anti-Phishing Working Group (AWPG) observed approximately 4.9 Million phishing attacks [13]. Despite the deployment of various anti-phishing technologies, the human factor remains a critical vulnerability. The effectiveness of phishing warnings, therefore, is paramount in the cybersecurity defense arsenal. These warnings, if designed and implemented effectively, can play a crucial role in educating users and averting the success of phishing attacks.

In response to the escalating threat of phishing attacks, numerous efforts have been undertaken by cybersecurity experts and organizations to mitigate the risks. These include the development of algorithms for threat detection [24] and the implementation of user awareness programs [14]. Users' ability to detect and respond to phishing attempts is critical, prompting a shift toward developing more user-centered security solutions. These solutions emphasize visual warnings that act as direct interventions, alerting users to potential threats. However, the effectiveness of such warnings hinges on several factors that this study aims to explore and understand.

For our study, we developed mock-up visual warnings to simulate real alerts in an email client and expanded their functionality with a custom plugin for Mozilla Thunderbird, a widely-used open-source email client chosen for its adaptability. This plugin enhances the visual presentation of phishing warnings, utilizing common design principles to capture users' attention more effectively.

We created a unique set of emails incorporating both regular communications and mock phishing scenarios to simulate a realistic email environment, providing a robust platform for our user study. Using eye tracking technology and detailed user feedback, we explored how users perceive, understand, and react to various visual warning strategies. This investigation offers insights into the design of more effective phishing warnings and highlights the importance of user interface design in cybersecurity. By focusing on user-centric design elements, our study aims to improve how visual cues are used to strengthen defenses against phishing attacks. Our research is structured around the following research questions:

- RQ1: How do users' perceptions differ in response to various phishing warnings in email interfaces, and which type is perceived as the most effective? *Here, "effective" refers to the ability of the warnings to increase user awareness, prompt appropriate user actions to mitigate risk (e.g., not clicking on malicious links), and enhance recognition of phishing threats in future encounters.*
  - How do users perceive the helpfulness of each warning type?
  - Which warning design do users prefer and why?
- RQ2: What common gaze patterns emerge among users when scanning email content, and how can these patterns help with the strategic placement of phishing warnings to maximize user attention?

To address these research questions we have laid out the following objectives for our study:

1. *Evaluating Visual Warning Strategies:* We aim to examine the effectiveness of various visual elements (banner placement, animations) in warning designs in terms of their immediate recognizability and understandability.
2. *Analyzing Attention Distribution:* Using eye tracking technology, we plan to identify which visual elements attract users' attention and how these elements influence perception and behaviour.
3. *Assessing User Reactions:* Through interviews with participants, we intend to capture their subjective evaluations and opinions regarding the different warning types. This includes exploring user trust and their perception on the warnings helpfulness.
4. *Establishing Best Practices:* Based on the collected data and user feedback, we seek to identify design principles that maximize the effectiveness of phishing warnings.

In summary, our research reveals that effectively designed phishing warnings can significantly influence users' abilities to recognize and respond to phishing attempts. Eye tracking data confirmed that warnings integrated within the email body are noticed and acted upon more swiftly than those placed peripherally or appearing with a delay. Qualitative feedback underscored the importance of clear, contextual information within these warnings, enhancing users' understanding of potential threats.

The implications of these results are substantial, advocating for a shift in how cybersecurity defenses are structured. As phishing tactics evolve, the need for warnings that seamlessly integrate into users' workflows and effectively communicate risks becomes increasingly critical. By focusing on the development of user friendly and informative warning systems, we can better equip individuals to combat phishing threats, potentially reducing the success rate of such attacks across various platforms.

## 2 Background and Related Work

In the dynamic field of cybersecurity, understanding the multifaceted nature of phishing attacks is essential for developing effective defensive strategies. This section provides a comprehensive overview of the key areas of research that inform our study on phishing warnings. By exploring the psychological underpinnings of social engineering, the evolution of phishing detection technologies, and the critical role of user education, we aim to contextualize our research within the broader discourse on combating phishing. Each of these areas offers vital insights that directly influence the design and implementation of phishing warnings, ensuring they are not only technically sound but also psychologically effective. This background forms the basis for our approach to enhancing how users recognize and respond to phishing threats.

### 2.1 Phishing Definition and Techniques

Phishing is a deceptive practice where cybercriminals use fraudulent communication, typically via email, to trick individuals into revealing personal information, such as passwords and credit card numbers. This cyber threat exploits the trust of users through cleverly disguised messages that mimic legitimate sources. Understanding the array of phishing techniques is essential for developing effective countermeasures and for training users to recognize and avoid these threats.

As detailed by Alkhalil et al. [3], phishing attacks employ a combination of technical and psychological tactics to extract sensitive information from victims. These attacks are not uniform but vary significantly in their approach and execution. Alkhalil et al. identify several prevalent phishing techniques:

- *Email Phishing*: The primary method where attackers send fraudulent emails mimicking legitimate sources to steal personal information. Visual warnings for this technique should clearly differentiate genuine communications from phishing attempts, for instance by highlighting suspicious elements directly within the email interface.
- *Spear Phishing and Whaling*: Highly targeted forms that use personalized information to increase the attack's credibility.
- *Vishing (Voice Phishing) and SMiShing (SMS Phishing)*: Phishing via phone calls and SMS that exploit the user's trust in mobile communications.
- *Website Forgery*: Creating counterfeit websites that replicate genuine ones to capture personal data.
- *Social Media Phishing*: Using social platforms to trick users into revealing sensitive data through deceptive posts or messages.

This classification underscores the importance of tailored defensive strategies that address the specific vectors and tactics of each phishing type, enhancing both technological defenses and user training programs.

### 2.2 Technical Detection Strategies

Phishing detection technologies are essential tools, designed to identify and mitigate phishing attempts across various digital platforms. Among these technologies, email client spam filters, with phishing emails being a subset of spam mails, represent a fundamental line of defense, filtering out potential phishing emails based on specific criteria such as sender reputation [22] or message content [5].

In their paper, Bhomwick et al. [5] delve into the evolution of machine learning techniques in the realm of email spam filtering. They chart the journey from initial heuristic-based approaches to the adoption of sophisticated machine learning models, capable of dynamically adapting to the continually evolving strategies employed by spammers. The paper not only highlights the effectiveness of these techniques but also points towards future directions for research in improving spam detection mechanisms.

Another approach in phishing detection is explored by Md. Abu Ashraf Siddiq et al. [23], utilizing deep learning techniques. Their study, employing neural networks, showcases the potential of deep learning in accurately identifying phishing websites. This highlights deep learning's pivotal role in cybersecurity, especially in distinguishing malicious sites, and emphasizes its contribution to advancing phishing detection systems.

While technological solutions are crucial, their effectiveness ultimately depends on the users' ability to understand and respond to the warnings they generate. This interdependency brings our research into focus, aiming to optimize visual warnings to ensure they not only capture attention but also convey urgency effectively, prompting appropriate user actions in the face of potential threats.

### 2.3 User Education and Training

Understanding the mechanisms behind phishing attacks and the reasons users fall prey to them is essential in developing effective countermeasures. A key vulnerability is the general lack of cybersecurity awareness among users, rendering them susceptible to manipulation by attackers. In their paper Dhamija et al. [9] highlight this issue, pointing out the critical need for enhanced user understanding to mitigate vulnerability to phishing shemes.

To adress this gap, various training methods have been explored, aiming to improve user ability to recognize and resist phishing attempts. Research by Sheng et al. [21] showcases an innovative approach in phishing education through the 'Anti-Phishing Phil' game. This research demonstrates that integrating educational content with interactive gameplay can significantly increase the efficacy of cybersecurity training. Such engaging methods not only educate but also actively involve users in the learning process, making the lessons more memorable and practical.

Further emphasizing the importance of user education, Kumaguru et al. [16] investigated the impact of training on users' abilites to distinguish between legitimate and fraudulent websites. Their findings underscore the effectiveness of well-crafted educational programs in empowering users to safeguard themselves against phishing threats.

Additionally, Wash's article [29] on how IT experts detect phishing emails provides a unique perspective on the cognitive strategies employed by experienced professionals. By interviewing IT experts, Wash explains a three-stage process these professionals use: sensemaking, suspicion, and decision-making, when evaluating the legitimacy of an email. This approach underscores the depth of expertise and vigilance required to effectively identify phishing attempts, highlighting the potential benefits of training programs that enhance these cognitive skills among general users.



Integrating these insights, our study leverages the principles of effective education and cognitive engagement in designing visual warnings that are not only noticeable but also informative. By enhancing user understanding of what phishing attempts look like, we aim to elevate the baseline of cybersecurity awareness, thereby reducing the efficacy of phishing attacks.

## 2.4 Social Engineering and the Human Factor

Social engineering is described as an attack type where the assailant exploits human vulnerabilities through social interaction to compromise cybersecurity goals like confidentiality, integrity, and availability of cyber elements. Unlike traditional attacks that may require overcoming technical barriers, social engineering attacks manipulate human psychology to bypass security measures. It's particularly challenging to defend against because it preys on innate human traits, which can often be turned into security vulnerabilities by skilled attackers [28].

Exploring vulnerability in social networks, Albladi et al. [2] analyze factors that predict individuals' vulnerability to social engineering in social networks. It proposes a model incorporating user characteristics such as level of involvement, social network experience, and perceptual factors like risk perception and competence. This research aims to identify the most vulnerable users for targeted training programs, contributing to safer social networking practices.

Highlighting persuasive email techniques, Koddebusch [15] focuses on the specific persuasive techniques employed in phishing emails, such as invoking authority or urgency, which effectively manipulate users' decision-making processes. This manipulation illustrates how cyber attackers exploit psychological triggers to induce actions that users might otherwise consider risky or unnecessary. Koddebusch's findings provide empirical evidence that these persuasion strategies significantly increase the success rate of phishing attacks, underscoring the critical need for awareness and training that can help individuals recognize and resist such tactics.

In their research, Abdrabou et al. [1] investigate how phishing emails affect user behavior by tracking their gaze and mouse movements during an email sorting task. The study aims to understand if these physiological and behavioral cues can serve as indicators of exposure to phishing attacks, which are often designed to manipulate users into revealing sensitive information. Their findings suggest that subtle changes in behavior, such as eye movements or mouse dynamics, could potentially be used to identify when users are interacting with phishing content.

Distler's study [10] examines the impact of various contextual factors on individuals' responses to phishing within an organizational setting. Using an in-situ deception methodology with 14 participants in their real work environments, the study reveals that contexts significantly influence the likelihood of falling for phishing emails. Distler's findings underscore the importance of contextualized training and proactive organizational policies to enhance cybersecurity. The research also underscores the role of language and writing style in the detection of phishing mails, revealing that unexpected variations in these elements can alert recipients to potential phishing attempts.

Building on Distler's findings, our study also introduces and evaluates targeted phishing warnings that focus on anomalies in greeting patterns and email signatures—common areas where deviations from normative language use might occur. By flagging unusual language patterns that recipients have learned to associate with phishing, as evidenced in Distler's study, these warnings aim to enhance the detection capabilities of email users, potentially increasing their vigilance and reducing the likelihood of successful phishing attacks.



## **3 Methodology**

### **3.1 Research Design**

Our study adopts a mixed-methods approach, integrating eye tracking data and qualitative interviews to examine how users perceive and interact with phishing warnings in email interfaces. The quantitative component makes use of eye tracking to objectively measure user attention and gaze patterns across various warning designs. The qualitative segment, through semi-structured interviews, investigates participant perspectives, enriching the data with personal insights and preferences. This dual-faceted approach aims to reveal not only immediate visual engagements but also deeper user experiences, driving towards more effective phishing warning implementations.

### **3.2 Research Setting**

Our research was conducted in a controlled lab environment at the University of the Bundeswehr. Upon arrival, participants were welcomed and escorted to a dedicated workstation, which was arranged to facilitate the study's requirements. This workstation was equipped with a computer system that had the email client, Mozilla Thunderbird, preloaded and configured for immediate use. An eye tracking device was connected to the computer to record and analyze participants' gaze patterns as they interacted with the email client. The setting was optimized to mimic a typical user experience while ensuring precise data capture for the study's analysis.

### **3.3 Recruitment of Participants**

For the recruitment of participants for this study, we utilized multiple communication channels within the informatics community of the University. The study was advertised in group chats on popular platforms such as Discord, Whatsapp and Slack, targeting a diverse range of participants. Additionally, we utilized the LMU newsletter to reach a broader audience within the university setting. There were no specific prerequisites for participants, allowing for a wide demographic range. The study was open to all individuals aged 18 and above. Participants received a 10 EUR compensation for their participation.

### **3.4 Material**

#### **3.4.1 Technical Equipment and Tools**

Central to the empirical aspect of this study is the integration of eye tracking technology, specifically utilizing the Tobii Pro Spark eye tracker in conjunction with the Tobii Pro Lab software. This setup plays a crucial role in objectively capturing where participants' attention is focused as they interact with email content and phishing warnings. By recording gaze patterns we gain valuable insights into the reactions to different warning designs. The Tobii Pro Lab software allows us the seamless recording of the participants' screen

3 Methodology

alongside their gaze movements and patterns, facilitating a comprehensive analysis of the captured data. Audio recording was conducted using a Google Pixel smartphone, which features built-in auto transcription software, providing us with a reliable method for capturing and transcribing participant feedback during the study. Each transcript was additionally manually reviewed to ensure the accuracy of the data and to filter out irrelevant or personal information.

3.4.2 Emails

To establish a genuine email environment for the study, we created a suite of 30 emails to mirror the typical inbox of a professional workspace. This collection included routine business correspondences, such as project updates, meeting invitations, and newsletters, mixed with 9 simulated phishing emails. The inclusion of one false negative, a phishing email not recognized as such, further emulates the realistic challenges users face. These simulations employed prevalent phishing strategies observed in actual attacks, such as soliciting confidential information, fraudulent alerts, and deceptive links to malicious websites.

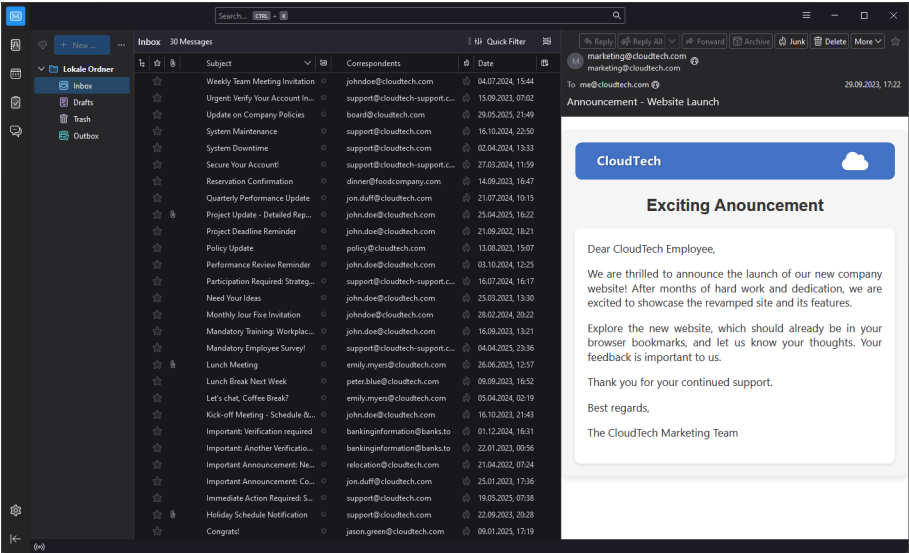


Figure 3.1: Participants' environment in Mozilla Thunderbird.

Urgent: Verify Your Account Information Now!

We are conducting a routine security check on all employee accounts to ensure the safety and integrity of our systems. Unfortunately, we have encountered a problem with your account information.

To prevent any disruptions to your access, we kindly ask you to verify your account details immediately by clicking on the following link:

[Verify Account](#)

Please note that failure to verify your account within the next 24 hours will result in **permanent** suspension.

Thank you for your cooperation in this matter.

Sincerely,

Cloudtech Ltd.

Figure 3.2: An example for a mock phishing email.

## 3.5 Procedure

### 3.5.1 Consent, Demographics and ATI-S

Before beginning with the study tasks, participants are required to sign a consent form that includes permissions for screen and audio recording as well as eye tracking. Following consent, they are asked to provide demographic information such as age, gender, educational background, and professional experience. Additionally, participants fill out the short version of the Affinity for Technology Interaction Scale (ATI-S)<sup>1</sup>. This scale assesses individuals' comfort levels and predispositions towards technology, which helps to illuminate how these factors may influence their perceptions and interactions with phishing warnings.

We opted to use the short version of the ATI scale due to its ability to efficiently measure technological affinity without imposing extensive questionnaire fatigue on participants. The shorter scale offers a balance between comprehensive assessment and participant convenience, ensuring capture of data while maintaining high engagement levels.

### 3.5.2 Briefing and Setup

Participants receive a briefing on the eye tracking process and a brief overview of Mozilla Thunderbird if they are not already familiar with its interface. The eye tracking calibration is conducted next to ensure accurate data capture. Audio recording is also enabled to capture the participants' verbal feedback throughout the study, facilitating later analysis of their qualitative responses. Participants are informed that they are welcome to ask questions at any time during the study to clarify any doubts or concerns they may have.

### 3.5.3 Phase One: Sorting Task and Collection of Eyetracking Data

Phase One serves as the quantitative backbone of our study, utilizing eye tracking technology to gather gaze data as participants engage with a series of emails. This phase is designed to simulate a real-world working environment to objectively assess how participants interact with our phishing warnings in a typical workflow.

#### Task Setup and Environment

Participants are cast in the role of an employee at the fictional company "CloudTech". This setup is intended to create a realistic email environment. Upon commencement they are provided with a folder containing a mix of emails, including routine communications from fictional colleagues and company announcements. Within these emails there are several phishing attempts, though participants are not made aware of these inclusions to avoid biasing their responses.

#### Instructions and Procedure

To simplify the email management task and ensure clarity, participants are instructed to discard any email into the trash folder that they suspect or deem unnecessary, without the need to further categorize or mark emails. This is designed to streamline the sorting process, allowing participants to focus more on the content and less on organizational tasks. During the sorting process, participants are encouraged to verbalize their thoughts and decisions. This "think aloud" method is crucial for understanding the decision-making process behind their actions.

---

<sup>1</sup><https://ati-scale.org/>

#### Observer Role and Time Allocation

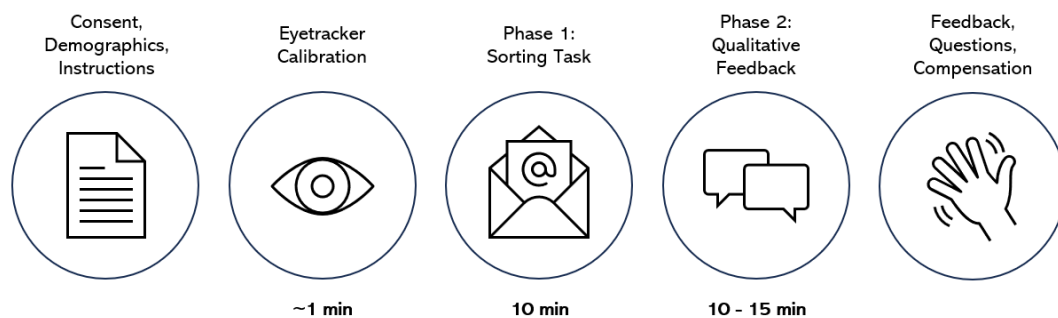
The observer's role during this phase is to listen and monitor the explanations provided by the participants, intervening minimally (e.g. when questions arise) to avoid influencing their natural behavior. Each participant is allocated approximately 10 minutes to complete the task.

#### 3.5.4 Phase Two: Gathering Qualitative Feedback

Following the sorting task, participants undergo a detailed semi-structured interview. They are asked about their subjective perceptions, their strategies for identifying phishing attempts and their opinions on the clarity and effectiveness of different warning designs. This qualitative data enriches the study's findings by adding personal context to the observed behaviours and eyetracking metrics. The interview questions cover several key areas:

- **General Perception:** Inquiry into how participants perceived the warnings overall, including any elements that particularly stood out to them and the influence of the warnings on their decision making process.
- **Effectiveness:** Evaluation of the warnings' effectiveness presented during the study, identifying any warnings perceived as particularly helpful or less so. This includes a comparison between the different warnings.
- **Design:** Assessment of the warnings' design aspects, such as color, placement, and animations, and their impact on the users' attention.
- **Personal Behaviour:** Discussion around any new information noticed in the warnings that hadn't been considered before and typical behaviours when encountering phishing emails.
- **Feedback:** Suggestions for improving future iterations of the study and an opportunity for participants to raise any further questions or comments.

The full study protocol and questioning framework is provided in appendix A.2 and A.3.



**Figure 3.3:** Summary of the study procedure.

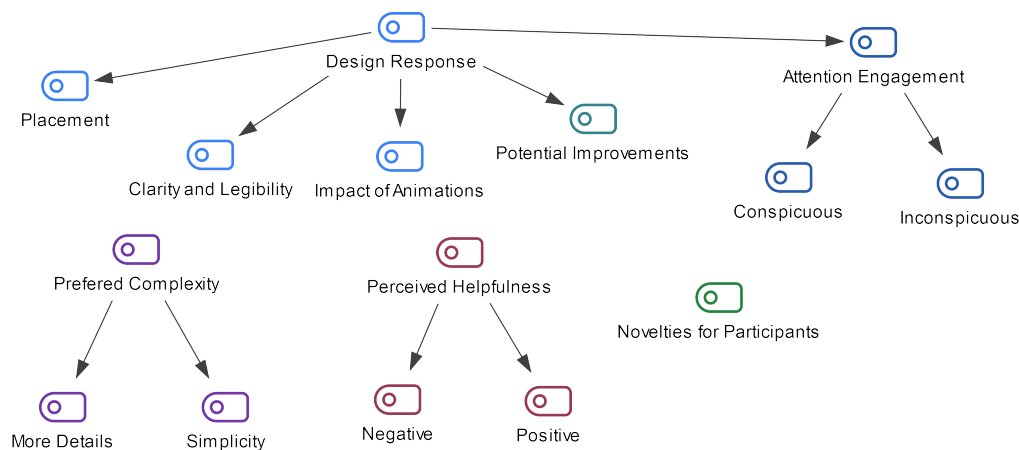
## 3.6 Data Analysis

### 3.6.1 Qualitative Data Analysis

To enhance our understanding of participants' perceptions regarding different phishing warning designs, we employed a Qualitative Data Analysis (QDA) approach. Following established methodologies outlined in literature [20], our method involved a hybrid coding strategy, combining both deductive and inductive techniques. Initially we established a set of codes based on our research questions and theoretical framework (deductive). As we analyzed the data, we expanded and refined this set to include new themes and patterns that emerged from participant feedback (inductive). This approach ensured that our analysis remained both structured and adaptive to the insights yielded by the study.

To facilitate this thorough and efficient exploration of qualitative data, we utilized MAXQDA2024. This software provided robust tools for organizing, coding, and analyzing the interview data, making the complex process of qualitative analysis more manageable and systematic. The project files are available in our GitHub repository, linked in appendix A.1.

Our coding framework captures the key dimensions of our research. Each main category and sub-category focuses on different aspects of how users interact with and perceive the visual and functional elements of phishing warnings. A detailed list of codes with explanations is provided in appendix A.5.



**Figure 3.4:** Graphic illustration of our QDA codes, generated in MAXQDA 2024.

### 3.6.2 Eyetracking Data Analysis

To evaluate the eye tracking data, we formulated the following hypotheses:

- H1** Delayed slide-in animations of warnings will disrupt typical email reading patterns and draw more attention.
- H2** Warnings placed on the side of the email interface will be noticed later compared to those within the email body or at the top.
- H3** Content integrated directly into the email body (e.g. Greeting or Signature Warning) will hold user attention for longer periods.

### 3 Methodology

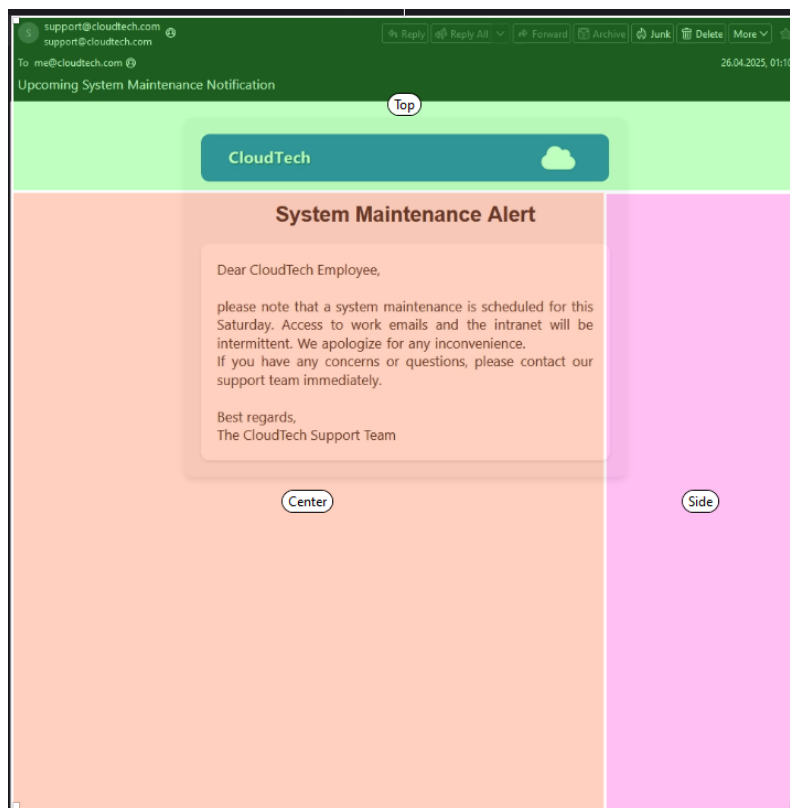
To analyze the eye-tracking data, we used the Tobii Pro Lab software, which provided comprehensive tools for visualizing and interpreting gaze patterns. Our analysis involved several key tasks:

1. **Defining Areas of Interest (AOI):** We defined specific AOIs within the email interface. This allowed us to quantify metrics such as fixation durations and frequencies for specific areas. The AOIs included different parts of the email where warnings were placed (see figure 3.5).
2. **Establishing Times of Interest (TOI):** To focus on critical moments of interaction, we defined TOIs. These were specific instances when participants opened emails containing warnings.
3. **Visualization and Gaze Path Analysis:** Using Tobii Pro Lab's visualization tools, we observed and analyzed gaze points. This was crucial for Hypothesis 1, as it allowed us to see how animations affected reading patterns.

To align with our hypotheses, we collected the following eye tracking metrics:

1. **Time to First Fixation:** This metric measures the time it takes for participants to fixate on a specific AOI after the email is presented. It helps us understand how quickly each warning design draws attention compared to others.
2. **Average Fixation Duration:** By averaging the duration of all fixations within each AOI, we can assess the overall engagement level with different components of the email interface. Fixation durations range from 150-300ms, with higher durations meaning a higher level of engagement [25].

These methods and metrics allowed us to rigorously test our hypotheses and evaluate the effectiveness of different phishing warning designs in capturing user attention.



**Figure 3.5:** AOI zones in Tobii Pro Lab.



# 4 Design and Implementation

## 4.1 Warning Types

For our study we designed four primary types of phishing warnings, each having two different variations. These variations aim to explore different aspects of user interaction and attention.

### 4.1.1 Group 1: General Alerts

**Generic Banner:** A broad, non-specific warning that does not target any specific content within the email but serves as a general alert to the possibility of phishing.

- Variant 1: The banner slides in from the top. This placement tests the immediate visibility and impact of warnings when users first open an email.
- Variant 2: The banner is positioned on the right side of the email body. This alternative aims to assess how peripheral placement affects user noticeability and response compared to more traditional top placement.

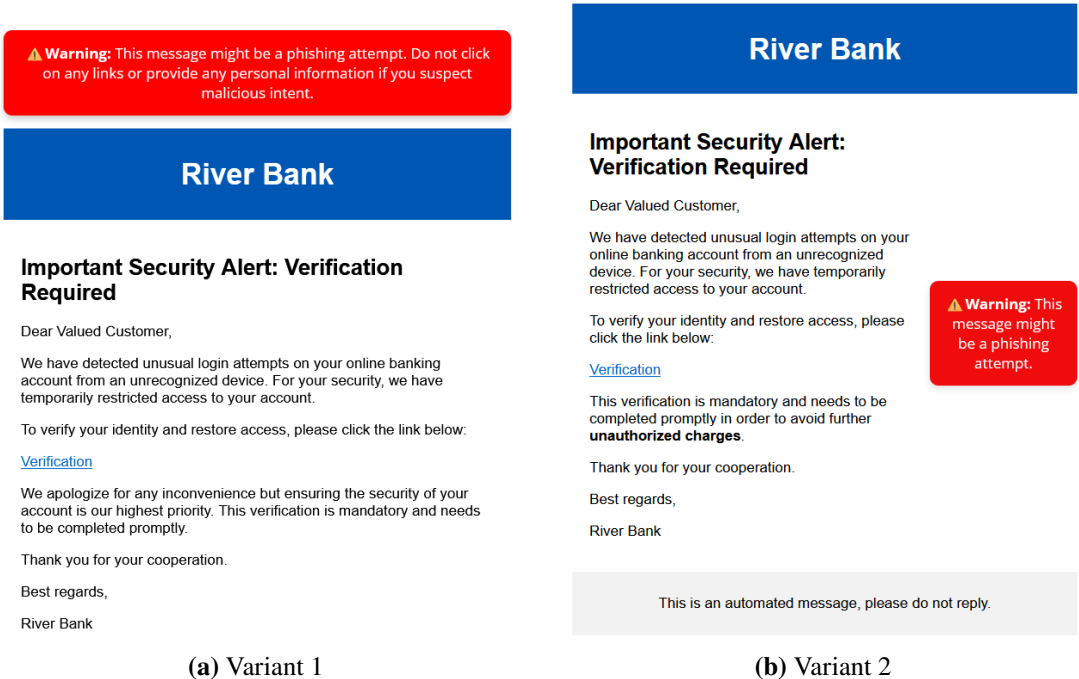


Figure 4.1: Generic Banner

4.1.2 Group 2: Context-Aware Warnings

**Tooltip Warning on Link Hover:** This warning appears when a user hovers over a link, providing immediate contextual information about the potential danger of the link, making it context-aware as it is specifically related to the actions the user is about to take (i.e., clicking a link). A tooltip appears, warning the user and preventing them from clicking the link. After a 3 second countdown, the tooltip moves to the side, making the link clickable at the users own risk.

This approach aligns closely with the research conducted by Petelka et al. [19], which emphasizes the effectiveness of proximity and user interaction with link-focused phishing warnings in reducing susceptibility to phishing attacks. In this thesis, the tooltip warning activates when a user hovers over a link, immediately providing contextual information and temporarily blocking access to the link, thereby forcing the user to recognize and process the warning before proceeding. This method mirrors Petelka et al.’s findings that such forced interaction can significantly heighten user caution and awareness. Moreover, this study further explores the impact of varying levels of informational detail in the tooltips, assessing how these variations influence user perception.

- Variant 1: Includes more detailed information within the banner, such as the actual URL. This version tests whether providing additional context within the warning influences user decisions about clicking on links.
- Variant 2: Presents less information, focusing on a simple cautionary message. The aim here is to evaluate the effectiveness of minimalistic warnings in deterring harmful interactions.

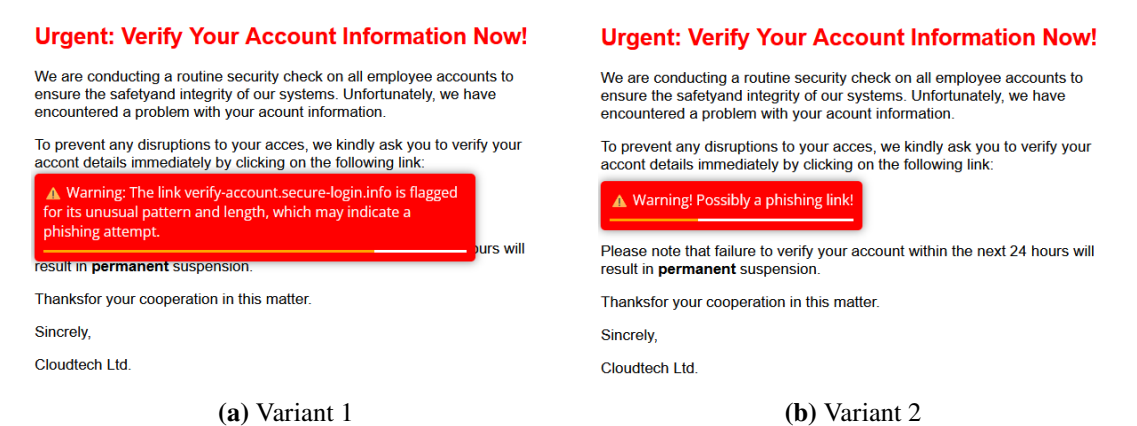


Figure 4.2: Tooltip Warning on Link Hover

4.1.3 Group 3: Content-Specific Warnings

**Signature-Specific Warning:** This warning is triggered by anomalies in the email’s signature block which may indicate that the message is not from a legitimate source. It focuses on deviations from standard or expected signature formats commonly seen in regular correspondence with the sender.

- Variant 1: Accompanied by proof from past emails, offering a direct comparison to highlight discrepancies. This tests the baseline effectiveness of alerting users to potential signature anomalies with additional context.
- Variant 2: Presents less information, focusing on a simple cautionary message next to the signature. The aim here is to evaluate the effectiveness of minimalistic warnings in deterring harmful interactions.

Dear Team,

Please find the agenda for our weekly meeting attached.

Your input on item 3 would be particularly valuable. Looking forward to our productive discussion.

Best regards,

John Doe  
Head of HR  
CloudTech Inc.  
johndoe(at)cloudtech.com  
Phone: (321) 654-0987



**Warning: Potential phishing attempt!**  
This signature differs from the usual signature. Previous emails from this address have used the following signature:

John Doe  
R&D Department Head  
CloudTech Innovations  
johndoe(at)cloudtech.com  
Phone: (123) 456-7890

(a) Variant 1

Dear Team,

Please find the agenda for our monthly meeting attached.

Your input on item 3 would be particularly valuable. Looking forward to our productive discussion.

Best regards,

John Doe  
Head of HR  
CloudTech Inc.  
johndoe(at)cloudtech.com  
Phone: (321) 654-0987



**Warning: Potential phishing attempt!**  
This signature differs from the usual signature.

(b) Variant 2

**Figure 4.3: Signature-Specific Warning**

**Greeting-Specific Warning:** Similar to the previously described signature warning, this alert focuses on irregularities in the email's greeting. It is designed to catch attention when the greeting does not match the usual formats, suggesting that the email could be a phishing attempt.

- Variant 1: Includes "proof" from past emails, similar to the signature warning variation. Additionally this warning banner is interactive, as the user can click on the corresponding email, opening it in a new tab. This is designed to see if direct comparisons with previous legitimate emails help users to better understand why the email might be a phishing attempt.
- Variant 2: Without additional context, focusing on a simple cautionary message next to the greeting. This variation assesses users' ability to detect phishing without the aid of explicit historical comparisons.

#### Participation Required: Strategic Employee Survey

Hello,

**Warning: Potential phishing attempt!**  
This greeting is unusual for messages from CloudTech. For reference, see past greetings used in communications:

- 01/01/24: "Dear CloudTech Employee, ..."
- 01/12/23: "Dear CloudTech Employee, ..."
- 10/11/23: "Dear CloudTech Employee, ..."

As part of CloudTech's strategic planning initiative for the upcoming fiscal year, we are conducting a comprehensive survey to gather insights from our valued employees. Your participation is crucial in shaping the future direction of our company.

To contribute, please access the survey through the following link:

[Survey](#)

(a) Variant 1

#### Participation Required: Strategic Employee Survey

Hello,

**Warning: Potential phishing attempt!**  
This greeting is unusual for messages from CloudTech.

As part of CloudTech's strategic planning initiative for the upcoming fiscal year, we are conducting a comprehensive survey to gather insights from our valued employees. Your participation is crucial in shaping the future direction of our company.

To contribute, please access the survey through the following link:

[Survey](#)

(b) Variant 2

**Figure 4.4: Greeting-Specific Warning**

### 4.2 Visual Design Strategy and Rationale

We followed a visual design strategy for the phishing warnings developed in our study, aiming to merge aesthetics with functionality to create an effective warning mechanism. Building upon established research [31], our design approach was crafted to enhance the salience, recognizability, and comprehensibility of warnings. We focused on several key elements identified in the literature as critical for maximizing the effectiveness of visual warnings: color usage, typography, and the inclusion of universally recognized symbols. These elements were chosen based on their proven ability to capture attention quickly and convey urgency effectively, ensuring that users could not only notice but also understand the warnings promptly and act accordingly.

#### 4.2.1 Color Scheme and Typography

The choice of a vivid shade of red (hex: #ff4d4d) for the background of our warnings leverages its psychological impact, as this decision is rooted in research indicating that red can trigger heightened alertness and caution [17, 31]. The addition of a contrasting yellow triangle, serving as a pictorial symbol universally recognized for caution, further enhances the visibility and impact of the warnings. This use of universally understood symbols is supported by literature emphasizing the importance of immediate comprehension in warning design [31].

Rounded corners were introduced to the warning design to make the visual alerts appear more modern and less aggressive, maintaining user comfort without sacrificing visibility. This design choice aligns with contemporary design trends that favor user-friendly interfaces.

The font choice, Segoe UI, was selected for its clarity and the decision to use bold fonts aims to further enhance readability.

#### 4.2.2 Animation and Timing

Each warning features a sliding animation that initiates one second after the email is opened. This deliberate delay and the dynamic entrance of the warnings were designed to distinguish them clearly from the static email content. By introducing the warnings as distinct, dynamic elements, we aimed to disrupt the user's initial scan of the email content, drawing their focus directly to the alert and reinforcing the perception of the warnings as separate and urgent. Research supports this approach, showing that animations can significantly enhance the allocation of attentional resources, making such dynamic elements more effective at capturing and holding viewer attention [7]. This strategic use of animations is therefore critical in ensuring that the warnings are not only noticed but also perceived as immediate and pressing alerts.

### 4.3 Iterative Design Adjustments

Over the course of the study, feedback from participants highlighted several areas for improvement in our warning designs. This led to an iterative design process, where modifications were made to enhance clarity and legibility. The new warnings were implemented from the twelfth participant onwards. This allowed us to compare the effectiveness of the original and modified designs directly within the same study framework. Here are the key adjustments made:

- **Color and Typography:** We intensified the shade of red used for the warning backgrounds to a deeper, more vivid tone (hex: #f00c0c). The font was changed to 'Open Sans' and we also used more bold fonts. This resulted in a generally more legible design.

- **Animation Timing:** We removed the delay in the animations, allowing warnings to appear immediately as the email is accessed.
- **Greeting-Specific Warning:** The buttons that link to previous emails for comparison have been updated to more clearly indicate that they are interactive.

The figures in section 4.1 already incorporate these changes. An overview of our first iteration of the warning types is provided in appendix A.4.

## 4.4 Technical Background

### 4.4.1 Email Client

Our experimental setup was anchored in Mozilla Thunderbird, selected for its adaptability in customizing email functions via plugins. We utilized a portable installation of Thunderbird<sup>1</sup>, which facilitated an easy setup and ensured portability between setups. Custom emails were directly placed into the local folder of Thunderbird, eliminating the need for participants to log into an actual email account, thereby simplifying the process and enhancing security. This approach also simplified the process of resetting the setup after each participant, ensuring consistency and efficiency across sessions.

### 4.4.2 Emails and Custom Plugin

To facilitate the integration of phishing warnings into the emails, the warnings were directly embedded into the HTML structure of each email. This integration was aimed at replicating how an email client's security system might naturally react to detecting potential threats.

Additionally, a custom plugin was developed to enhance the functionality of the phishing warnings embedded within the email content. Utilizing Mozilla Thunderbird's WebExtensions API<sup>2</sup> and JavaScript code, this plugin was engineered to introduce dynamic visual cues, such as animations and interactive elements. The core functionality of the plugin centers around its capability to inject JavaScript code directly into the HTML of each email. This injection process is crucial as it allows for real-time modification of the email content, thereby enabling animations that draw the user's attention to specific areas of concern and interactive features that provide additional context.

#### Check Email Subject

As we only want our warnings to appear for a predefined set of emails, the plugin first checks the subject line of the email that is currently viewed. If it detects one of our predefined phishing emails, the plugin dynamically injects the necessary scripts into the email's HTML content. This method ensures that the interactive warnings and animations are only activated when relevant.

#### Event Listeners

For our hover-triggered tooltips on suspicious links we added an event listener to the hyperlink, which listens for mouse events and triggers the appearance of the tooltip when necessary.

---

<sup>1</sup><https://www.thunderbird-mail.de/lexicon/entry/44-portable-thunderbird/?l=2>

<sup>2</sup><https://webextension-api.thunderbird.net/>

### Animations

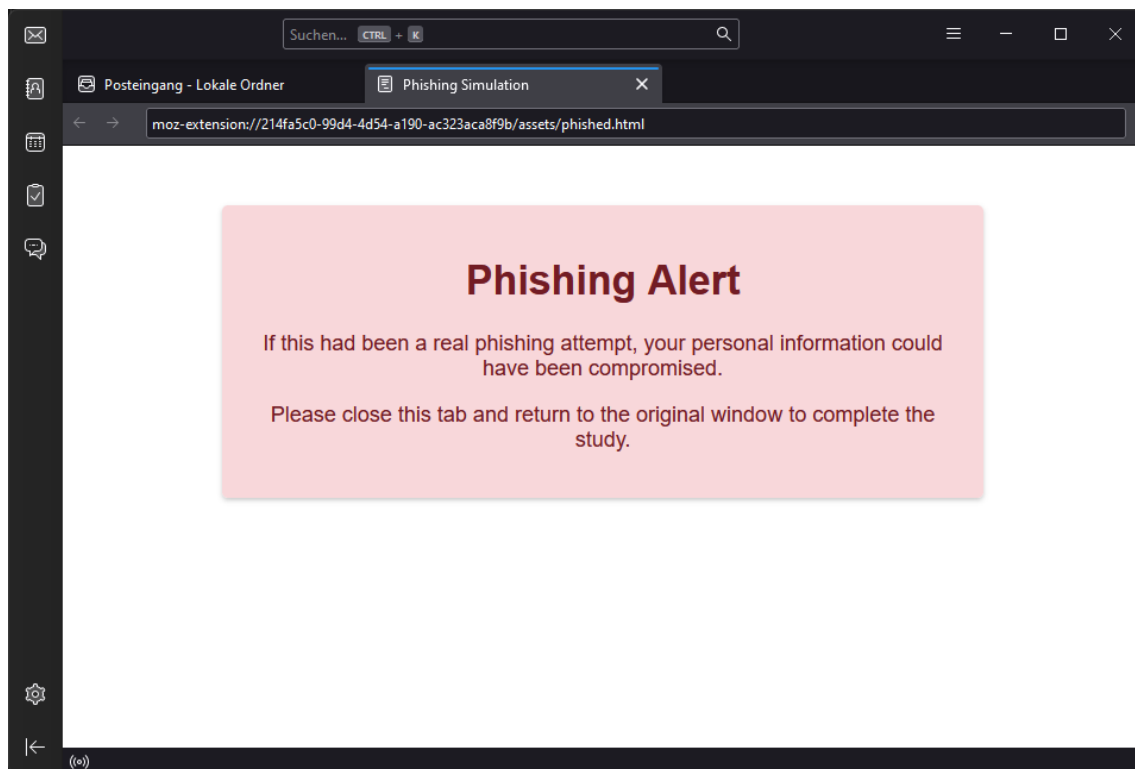
For animations we used the CSS @keyframes at-rule<sup>3</sup>. Doing this via the plugin was strictly necessary as Mozilla Thunderbird does not natively support animations embedded in the HTML code of emails.

### Interactive Elements

This plugin also makes the functionality of variant 1 of our greeting-specific warning (4.4a) possible. Within the HTML of the email, we embedded a list that represents past emails. The plugin expands each list item with an interactive capability: clicking any list entry opens a new tab displaying the content of the corresponding past email. Additionally the plugin handles the styling of the buttons. It applies a green background color to indicate interactivity and sets the cursor to a pointer, visually cueing users that the entries are clickable.

### Mock Phishing Page

We implemented a feature in the plugin to handle cases when users click on links within our mock phishing emails. Upon doing so, our plugin redirects them to a mock phishing page, designed to simulate the potential consequences of clicking on a phishing link. In terms of the technical implementation, the feature operates by attaching an event listener to each hyperlink within the mock phishing emails. When a link is clicked, rather than following the original URL, the event listener intercepts this action and instead directs the user to our predefined HTML page. This workaround was necessary as Mozilla Thunderbird restricts plugins from directly embedding links in emails for security purposes.



**Figure 4.5:** Mock phishing page when participants click on links.

---

<sup>3</sup><https://developer.mozilla.org/en-US/docs/Web/CSS/@keyframes>

### 4.4.3 Python Script to Shuffle Emails

Additionally, a Python script was utilized to shuffle the date of each email. As the emails were sorted by their corresponding dates, this allowed us to randomize the order of the emails for each participant. This step was critical to minimize any sequence bias and ensure that each participant's exposure to the emails was unique, thereby providing a more robust measure of the warnings' effectiveness across varied contexts. The full source code of the plugin, the Python script, and the emails, will be provided in appendix A.1, accessible through a GitHub repository.





## 5 Results

### 5.1 Questionnaire Data

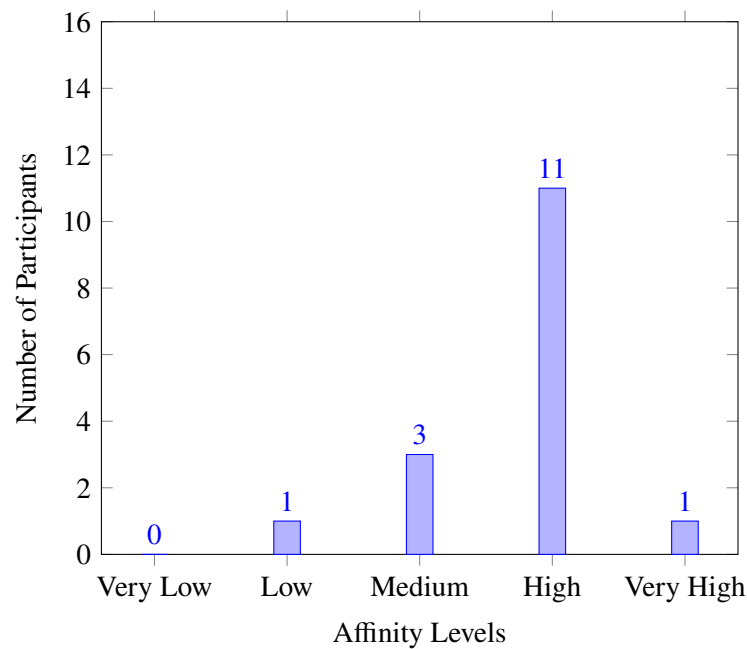
#### 5.1.1 Demographics

Demographics	Total (N=16)	%
<b>Age</b>		
20 - 29	13	81.25
30 - 39	2	12.50
40+	1	6.25
<b>Gender</b>		
Male	10	62.50
Female	6	37.50
Non-binary	0	0.00
Prefer not to tell	0	0.00
<b>Education</b>		
Doctorate degree	1	6.25
Graduate degree	2	12.50
Undergraduate degree	9	56.25
High school diploma	4	25.00
Secondary education	0	0.00
Other	0	0.0
<b>Profession</b>		
Student	13	87.5
Research Assistant	1	6.25
Retiree	1	6.25

**Table 5.1:** Demographic data collected in the survey.

The demographic data from our study primarily shows a young participant base with 81.25% aged between 20-29, primarily students (87.5%). The gender distribution was majority male (62.50%) with the remainder female. Most participants had at least some university education, with 56.25% holding an undergraduate degree and a smaller number having graduate degrees or a high school diploma. This generally younger demographic was anticipated, as the recruitment strategies employed were focused on platforms commonly frequented by younger audiences, such as university chat groups on Discord, WhatsApp, and Slack, as well as through the LMU newsletter.

### 5.1.2 ATI-S Scale



**Figure 5.1:** Distribution of Affinity for Technology Interaction among participants.

A significant aspect of the participant profile is their technological affinity, as indicated by the Affinity for Technology Interaction Short Scale (ATI-S) values. This data revealed a pronounced tendency towards technology, with 12 participants scoring high or very high in technology affinity. This implies that the majority of participants were not only familiar with but also comfortable using technology, which is a crucial consideration in the context of phishing awareness and the effectiveness of warnings.

## 5.2 Interview Results

### Design Response

**Clarity and Legibility:** Participants generally found the color red effective in drawing attention, with one noting *"The red is just very catchy"* (P9), highlighting the visual impact of vibrant colors. The legibility of warnings was often a concern, as evidenced by comments such as *"It's the white on red what makes it difficult to read at times."* (P9), indicating difficulties with text contrast.

**Impact of Animations:** Animations were noted for their effective attention-grabbing capabilities. For instance, one participant mentioned, *"When the text field moves in like this, your eyes automatically move there."* (P2), showing how animations guide viewers' focus. A participant remarked on the possible negative impact of frequent animations, stating, *"Moving warnings are more noticeable than ones that just appear on the screen. But if you put moving warnings on all phishing mails, it might be overwhelming and tire the eyes and mind. A balance is necessary. Maybe being too flashy, could even lead to desensitization over time."* (P5) Some participants also noted that the animations were too slow to appear, *"It would be better if the warnings appeared immediately upon opening the email instead of having a delay"* (P5), which could lead to critical oversights.

**Placement:** Participants commonly observed that warnings placed on the side of the email were less noticeable, often residing in their peripheral vision. As one participant expressed, *"I actually found this one a bit inconspicuous because you have the email body here and the warning there. It reminds me of those advertisements on websites placed on the side of the page."* (P1) Conversely, warnings integrated directly into the email body proved to be more conspicuous. Participants reported that these warnings grabbed their attention immediately upon opening the email, as they were in the direct line of sight while reading the email content.

### **Perceived Helpfulness**

**Context-Aware Warnings (Tooltip on Hover Over Link):** This type of warning was generally seen as highly effective because it actively prevents clicks on potentially malicious links. A participant mentioned, *"I believe the one covering the link was the most effective because it immediately prevented interaction."* (P6).

**Content-Specific Warnings (Signature and Greeting Warnings):** This type of warning was appreciated for its specificity, noting that these kind of details are easily overlooked in phishing emails: *"I really liked the signature warnings. That could be something that might not catch the user's attention at all."* (P5). A common point of criticism was, that this kind of warning is often not applicable, as not every email, especially in a non-business context, ends with a signature or uses a consistent greeting.

**General Alerts:** This type of warning received mixed feedback due to their positioning and the lack of detailed information they provided. Participants noted its visibility issues and expressed a preference for warnings that include explanatory details about the alert: *"Those that just flew in and had standard text might be the weakest."* (P6).

### **Preferred Level of Detail**

A key theme that emerged from the interviews was a clear division in preference for either more detailed warnings or simpler alerts. Participants who favored more details appreciated the additional context and explanations provided in the warnings, which helped them understand why an email was flagged as suspicious. One participant stated, *"I preferred the longer ones, because it teaches me how to recognize future phishing attempts."* (P11), emphasizing the educational aspect. Another point mentioned was the fact that a longer explanation might make it easier and more helpful for non tech-savvy people. The Greeting Warning with interactive buttons presented a unique challenge during the study. While intended to enhance user engagement by providing direct context and proof, many participants either did not realize the buttons were clickable or hesitated to use them. This hesitation stemmed from a fear that interacting with the buttons might inadvertently lead them into a phishing trap. One participant expressed, *"I wouldn't like to click on those buttons if I would get such a warning. I don't really know if I can trust it."* (P16).

Conversely, some participants (P5, P6) preferred simple warnings, valuing the straightforwardness and immediacy of the message without additional information. This preference indicates that for some users, efficiency and speed in email processing are prioritized over detailed educational content.

### **Novelties for Participants**

The qualitative data analysis underscored that many participants found the greeting and signature-specific warnings to be novel features. These aspects of the phishing warnings introduced new elements that participants had not previously encountered or considered significant in identifying phishing emails. As one participant reflected, *"I have not seen such tailored warnings in phishing emails, which recognize a specific pattern and tell you why it's a phishing attempt. That is very new to me."* (P13).

### **Refining Warning Design: Incorporating Feedback**

As outlined in section 4.3, modifications were made to the phishing warning designs beginning with participant 12. These changes were positively received, as indicated by subsequent participant feedback. Post-adjustment, there were no further complaints regarding the readability of the text or the timing of the animations, except for a single comment about the red tone being potentially too strong (P15). Participants appreciated the quicker responsiveness and greater clarity of the warnings, which effectively addressed the initial concerns raised about legibility and the timing of animation effects.

## **5.3 Eye Tracking Results**

During the data analysis phase, it was noted that the eye tracking data for one participant (P14) was significantly lower in sample size compared to others. This anomaly could be attributed to improper seating position or a flawed calibration process. To maintain the integrity and reliability of the analysis, this participant's data was excluded, resulting in a total of 15 eye tracking samples being considered for the final evaluation.

A full export of our raw eyetracking data is provided in the GitHub repository linked in appendix A.1.

### **H1: Delayed Animation's Influence on Reading Pattern**

Eye tracking data indicated no significant disruption in reading patterns due to delayed animations in phishing warnings. Observing the gaze patterns showed that the onset timing of animations, whether immediate or delayed, did not alter the attention flow significantly. Users exhibited similar engagement with immediate and delayed warning presentations, suggesting a uniform approach to processing these visual cues.

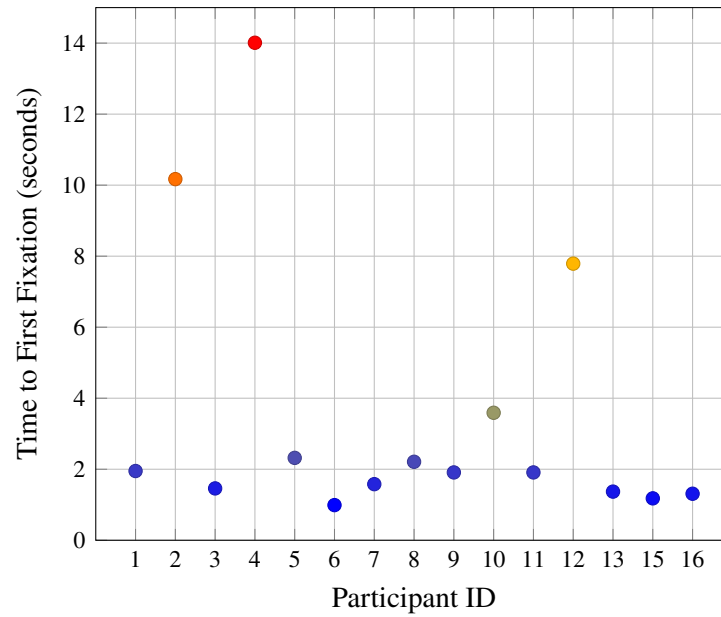
### **H2: Visibility of Side-Placed Warnings**

The eye tracking analysis showed that warnings positioned on the side of the email interface captured attention quickly but consistently later than other warnings. The focus of gaze primarily remained on the other areas of the emails, with peripheral warnings drawing attention only after the central and top content had been reviewed. This observation was consistent across the study's participants, supporting the hypothesis that side-placed warnings are less immediately noticeable.

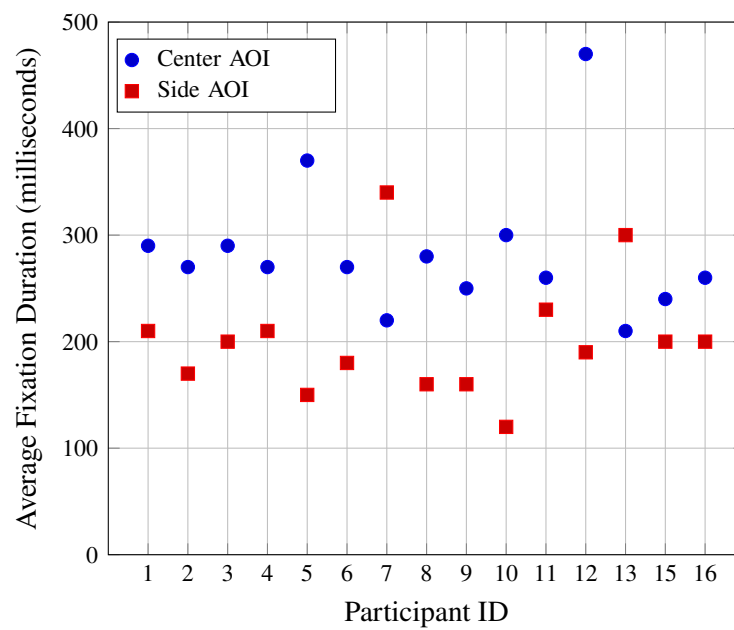
Eye tracking data (figure 5.2) supports this finding, showing that the median time taken for participants to first fixate on the Side AOI was approximately 2 seconds. To put this in contrast we compared this to the Top AOI data, where we recorded a notably quicker median time to first fixation of approximately 1.2 seconds.

### **H3: Sustained Engagement of Body-Integrated Warnings**

Data strongly supported the hypothesis that warnings integrated directly into the email body were engaged with more thoroughly by users than those placed at the side. The analysis of average fixation durations (figure 5.3) revealed that engagement times for the Center AOI were generally longer across participants. However, there were notable exceptions for participants 7 and 13, where the fixation duration on the Center AOI was shorter compared to the other participants.



**Figure 5.2:** Time to first fixation for the side AOI.



**Figure 5.3:** Comparison of the average fixation duration for center and side AOIs.



## 6 Discussion

### 6.1 Interpretation of Qualitative Feedback

#### 6.1.1 Emerging Themes and Best Practices

The qualitative feedback analysis illuminates several emerging themes that contribute to the effectiveness of phishing warnings. Reflecting on these, we can delineate best practices for designing phishing warning systems that align with user expectations.

##### **Attention-Grabbing without being Overwhelming**

A recurring theme is the delicate balance between capturing users' attention and avoiding sensory overload. Warnings that are too subtle may go unnoticed, while those that are too aggressive could be dismissed as annoyances. Best practices suggest employing moderate animations and placing warnings within the users' immediate visual field. For example, one participant noted, *"When the text field moves in like this, your eyes automatically move there."* (P2) indicating the value of animated cues. However, animations should not be excessively flashy or complex, as they risk desensitizing users over time. Supporting this, a study on the effects of animation in online contexts found that while animations increase visibility and engagement, they must be balanced to avoid overshadowing important content or overwhelming the viewer [8].

##### **Clarity and Contextual Information**

Clarity in warning messages is essential. Participants expressed a preference for warnings with a clear and legible typography set against a contrasting background, improving readability. The use of the color red was frequently mentioned as effective due to its association with caution and urgency.

Moreover, providing contextual information within the warning, such as the reason an email was flagged and highlighting subtle language cues, such as the greeting, can enhance users' understanding and prompt more informed action. As one user aptly put it, *"I preferred the longer ones, because it teaches me how to recognize phishing attempts."* (P11), highlighting the educative function.

This emphasis on context-rich warnings aligns with findings from other studies [4, 6], which also highlight the importance of explanatory warnings in phishing emails, thus not only alerting users to immediate threats but also equipping them with knowledge that could prevent future phishing attacks. Such strategies are crucial for improving user understanding and ensuring that warnings lead to informed actions and thereby enhancing the overall effectiveness of phishing warnings.

##### **Trust and Interactivity**

The trustworthiness of warnings is key; users must feel confident interacting with them. This includes a hesitance to engage with clickable elements within the warnings due to fear of intensifying a potential threat. Best practices would therefore involve clearly distinguishing warning elements from potential phishing mechanisms and ensuring that interactive components, when used, are immediately recognizable as safe and official parts of the email client interface.

##### **Incorporating User Feedback for Iterative Improvement**

The inclusion of user feedback into the design process is not only beneficial but essential for the development of effective phishing warnings. Adjustments made in response to participant insights, such as enhancing color contrast to improve legibility, exemplify how iterative design based on user experience can significantly help

in designing effective security measures. As highlighted in the article by Grobler et al. [12] actively engaging users in refining security solutions helps tailor these systems to better meet their needs and expectations, thereby increasing their efficacy and user adoption rates. Best practices would therefore advocate for continuous user testing and refinement, ensuring that warnings adapt over time to align with evolving user behaviors and phishing tactics. This proactive approach ensures that security measures remain robust, user-friendly, and effective in real-world scenarios.

### 6.1.2 Summary: Users' Perceptions of Phishing Warnings

To address RQ1 regarding user perceptions and the effectiveness of various phishing warnings in email interfaces, we analyzed qualitative feedback from participants. The effectiveness was assessed based on the ability of warnings to enhance user awareness, encourage appropriate actions to mitigate risks, such as not clicking on malicious links, and improve recognition of phishing threats in future encounters.

#### Perceived Helpfulness and Effectiveness of Warning Types

1. **Context-Aware Warnings (Tooltip on Hover Over Link):** Participants found context-aware warnings extremely effective. These warnings prevent direct interaction with malicious links, thereby immediately mitigating risk. This direct intervention makes them highly valuable in protecting users in real-time.
2. **Content-Specific Warnings (Signature and Greeting Warnings):** These warnings are appreciated for their specificity and relevance, particularly in environments where consistency is expected, such as professional emails. They provide clear and relevant context by highlighting anomalies in expected email patterns, offering a direct comparison with typical content. Their educational value is significant as they enhance users' ability to identify subtle signs of phishing attempts, which might otherwise be overlooked.
3. **Generic Banners:** Generic banners received the least favorable feedback. Their effectiveness was often questioned due to poor placement and lack of detailed information, which led to them being easily ignored. Especially variant 2 (side placed banner) was criticized for being too inconspicuous. Participants compared its noticeability to peripheral advertisements on websites, which are often disregarded due to desensitization. This feedback highlights the importance of not only the placement of warning banners but also the inclusion of specific reasons why an email might be flagged as potentially dangerous, enhancing user understanding and response to phishing threats.

#### Preferred Warning Design

From the feedback, it's clear that users prefer warnings that are not only visible but also informative. Warnings that blend seamlessly into the user's workflow in the email interface, such as those integrated within the primary content, are particularly effective. These warnings grab attention quickly, provide essential information, and are directly interacted with, enhancing both immediate and long-term phishing threat recognition.



Warning Type	Key Feature	User Feedback
Context-Aware	Actively prevents clicking on links	Very positive; users felt it was very effective and helpful, appreciating the proactive nature of the warning.
Content-Specific	Highlights anomalies in the language	Positive; users appreciated their specificity and educational value, noting they felt more informed about phishing tactics.
General	Broad, non-specific warning	Less favorable; often overlooked, reported by users as too generic and lacking actionable details.

**Table 6.1:** Overview of warning types and user feedback.

## 6.2 Interpretation of Eye Tracking Results

### 6.2.1 Impact of Delayed Animations

Our hypothesis that delayed animations would disrupt the typical email reading pattern and draw more attention was not supported by the data. The implications suggest that while animations serve as useful attention-grabbing tools, their effectiveness does not necessarily benefit from a delayed presentation. Aligning with the qualitative interview results, it appears that ensuring the visibility of warnings immediately upon email viewing is more crucial for capturing attention and facilitating prompt user reactions.

### 6.2.2 Effective Warning Placement

The data strongly supports that warnings placed within the central viewing area of an email are engaged with more extensively. This is evidenced by generally longer average fixation durations on the Center AOI. These findings underscore the efficacy of aligning warning placements with natural reading behaviors, thereby enhancing the visibility and impact of security warnings.

However, the analysis revealed notable deviations for participants 7 and 13, who exhibited shorter fixation durations on the Center AOI compared to other participants. This variability in user interaction suggests that while centrally-placed warnings are generally effective, their impact can vary significantly among users due to individual differences in reading habits or the visual salience of the warnings.

The extended engagement with central warnings highlights their effectiveness in maintaining user attention, which is crucial for ensuring that warnings are not only noticed but also sufficiently processed. This is vital for the effective communication of security-related information. Based on these insights, designers are encouraged to position crucial information and alerts within the central visual field to maximize their impact. Furthermore, understanding the reasons behind the shorter fixation times for specific participants could inform more tailored design approaches, ensuring that warnings are effective across a diverse user base.

In contrast, warnings positioned on the periphery of the email interface, such as on the sides, were consistently noticed later and engaged with less extensively. Although our data indicates only a slight delay in detection, this brief period can be critical, particularly during a hectic workday when users may be under pressure or multitasking. This oversight provides a window during which users could inadvertently interact with deceptive links or other malicious content. Moreover, the presence of outliers where the time to first fixation was exceedingly long highlights the potential severity of this issue.

Furthermore, the average fixation duration on these side-placed warnings was also notably lower than that

for central warnings, suggesting that even when the warning is noticed, the engagement is less intensive. This lack of deep engagement suggests that the warnings may not be registering effectively with users. This observation aligns with participant feedback, where many reported not noticing the side-placed warnings at all, despite eye tracking data confirming visual detection.

This phenomenon, supported by the comments of two participants (P1, P9), reinforces the argument that peripheral warnings may be less effective due to a learned tendency to disregard non-central elements. The reduced fixation durations further corroborate the likelihood that users engage only superficially with these warnings when they do notice them, which could seriously undermine their effectiveness in security-sensitive environments.

### 6.2.3 Summary: Gaze Patterns and Warning Placement

RQ2 explores how common gaze patterns among users can inform the strategic placement of phishing warnings. The analysis confirms a predominant focus on the central area of the email interface, a natural result of typical reading behaviors. This focal tendency significantly benefits the placement of warnings directly within this central area, ensuring they are both quickly noticed and more actively engaged with.

These insights provide a robust foundation for future email interface designs, suggesting that to maximize the impact of phishing warnings, designers should prioritize central visual placement of critical security cues. This approach will significantly improve defense mechanisms against phishing attacks, ensuring that warnings are not only visible but also integrated into the flow of natural user interactions with email content.

## 6.3 Limitations

### 6.3.1 Participant Diversity and Sample Size

The majority of our participants were drawn from a similar demographic group, predominantly consisting of university students and young adults within the same age range and academic environment. Moreover, a significant portion of participants demonstrated a high Affinity for Technological Interaction (ATI), as illustrated in figure 5.1. This heightened technological comfort and skill among our sample may not accurately represent the broader population's varying levels of technological proficiency. The lack of diversity in our sample, both demographically and in terms of technological affinity, could limit the generalizability of our findings across a broader population and potentially introduce biases in the evaluation of our phishing warnings. For example, older adults might interpret and respond to visual cues differently due to variations in technological fluency and cognitive processing speeds, which may differ markedly from the responses observed in younger, more technologically adept participants [26].

Additionally, the iterative design adjustment made based on participant feedback later in the study highlighted another significant limitation related to sample size. After implementing changes to the warning designs, it was planned to test these modifications with eight additional participants. Unfortunately, only four additional participants were able to participate in the study. This smaller-than-expected sample size of participants might not provide a sufficiently robust dataset to draw definitive conclusions about the effectiveness of the redesigned warnings.

### 6.3.2 Study Setup and Ecological Validity

The use of a controlled lab environment to conduct this study poses significant limitations regarding ecological validity. Participants were aware that their actions were being observed, potentially influencing their behavior—a phenomenon known as the Hawthorne effect [18]. This awareness can alter natural responses and may not accurately reflect their behavior in less controlled environments.

Furthermore, the artificial nature of the study environment can significantly impact participants' risk perception [11]. Since the risk to personal data is non-existent in a simulated phishing attack, participants' reactions to warnings and their decision-making processes might not truly reflect their actions in a real threat scenario. This discrepancy can skew results related to actual risk perceptions and behaviors, potentially leading to overestimations of the effectiveness of warning systems in genuine contexts.



## 7 Conclusion

This thesis has extensively investigated the impact of various visual warning designs on user responses to phishing threats in email clients. The use of eye tracking technology and qualitative feedback has provided a deep understanding of how different warning strategies affect user behaviour and engagement.

The research conclusively demonstrates that integrating warnings directly within the email content markedly enhances both the immediacy of user reactions and their overall engagement with the warnings. These findings underline the importance of prominent, context-rich warnings that are easily visible and provide actionable information to users.

Furthermore, the study highlights the crucial role of educational content in warnings. Clear, informative warnings not only alert users to immediate threats but also educate them, enhancing their ability to identify similar threats independently in the future. This dual function of warnings is vital in the ongoing battle against phishing, as well-informed users are the first line of defense in cybersecurity.

However, the study also highlights significant opportunities for expansion in both scope and realism for future research. A crucial first step is broadening the participant pool to include a more diverse range of ages, professions, and educational backgrounds, and increasing sample sizes to ensure findings are robust and generalizable. This will enhance our understanding of how different user groups perceive and react to phishing threats, supporting the development of more universally effective defensive measures. Additionally, to significantly boost ecological validity, future studies could integrate phishing detection research into real-world user environments. This advancement might entail the development and application of phishing detection algorithms capable of operating effectively within users' naturalistic settings. Such a move would not only provide more authentic data but also allow for a robust assessment of phishing warning effectiveness under typical usage conditions.

Looking forward, the rapid advancement of artificial intelligence and large language models opens new frontiers for phishing defense. These technologies could be leveraged to develop more adaptive, personalized phishing detection and prevention strategies, enhancing the precision and responsiveness of warnings. For instance, AI could analyze user behavior to tailor warnings to individual risk profiles or detect subtle phishing attempts that evade traditional detection methods. This proactive and technologically innovative approach is crucial in an era where cyber threats are continually evolving, thus ensuring the digital safety of users globally. As cybercriminals become more sophisticated, integrating cutting-edge technologies into cybersecurity measures will be essential for staying ahead of threats and safeguarding digital interactions.



# Bibliography

- [1] Yasmeeen Abdrabou, Felix Dietz, Ahmed Shams, Pascal Knierim, Yomna Abdelrahman, Ken Pfeuffer, Mariam Hassib, Florian Alt. *Revealing the Hidden Effects of Phishing Emails: An Analysis of Eye and Mouse Movements in Email Sorting Tasks*. 2023. arXiv: 2305.17044 [cs.HC].
- [2] Samar Muslah Albladi, George R. S. Weir. “Predicting individuals’ vulnerability to social engineering in social networks”. In: *Cybersecurity* 3.1 (Mar. 2020), p. 7. issn: 2523-3246. DOI: 10.1186/s42400-020-00047-5. URL: <https://doi.org/10.1186/s42400-020-00047-5>.
- [3] Zainab Alkhalil, Chaminda Hewage, Liqaa Nawaf, Imtiaz Khan. “Phishing Attacks: A Recent Comprehensive Study and a New Anatomy”. In: *Frontiers in Computer Science* 3 (2021). issn: 2624-9898. DOI: 10.3389/fcomp.2021.563060. URL: <https://www.frontiersin.org/articles/10.3389/fcomp.2021.563060>.
- [4] Joseph Aneke, Carmelo Ardito, Giuseppe Desolda. “Help the User Recognize a Phishing Scam: Design of Explanation Messages in Warning Interfaces for Phishing Attacks”. In: *HCI for Cybersecurity, Privacy and Trust: Third International Conference, HCI-CPT 2021, Held as Part of the 23rd HCI International Conference, HCII 2021, Virtual Event, July 24–29, 2021, Proceedings*. Berlin, Heidelberg: Springer-Verlag, 2021, pp. 403–416. ISBN: 978-3-030-77391-5. DOI: 10.1007/978-3-030-77392-2\_26. URL: [https://doi.org/10.1007/978-3-030-77392-2\\_26](https://doi.org/10.1007/978-3-030-77392-2_26).
- [5] Alexy Bhowmick, Shyamanta Hazarika. “Machine Learning for E-mail Spam Filtering: Review, Techniques and Trends”. In: (June 2016).
- [6] Paolo Buono, Giuseppe Desolda, Francesco Greco, Antonio Piccinno. “Let warnings interrupt the interaction and explain: designing and evaluating phishing email warnings”. In: *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems*. CHI EA ’23. Hamburg, Germany: Association for Computing Machinery, 2023. ISBN: 9781450394222. DOI: 10.1145/3544549.3585802. URL: <https://doi.org/10.1145/3544549.3585802>.
- [7] Muller Y.M. Cheung, Weiyin Hong, James Thong. “Effects of Animation on Attentional Resources of Online Consumers”. In: *Journal of the Association of Information Systems* 18 (Aug. 2017), pp. 605–632. DOI: 10.17705/1jais.00464.
- [8] Muller Y.M. Cheung, Weiyin Hong, James Thong. “Effects of Animation on Attentional Resources of Online Consumers”. In: *Journal of the Association of Information Systems* 18 (Aug. 2017), pp. 605–632. DOI: 10.17705/1jais.00464.
- [9] Rachna Dhamija, J. D. Tygar, Marti Hearst. “Why phishing works”. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI ’06. Montréal, Québec, Canada: Association for Computing Machinery, 2006, pp. 581–590. ISBN: 1595933727. DOI: 10.1145/1124772.1124861. URL: <https://doi.org/10.1145/1124772.1124861>.

- [10] Verena Distler. “The Influence of Context on Response to Spear-Phishing Attacks: an In-Situ Deception Study”. In: *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. CHI ’23. Hamburg, Germany: Association for Computing Machinery, 2023. ISBN: 9781450394215. DOI: 10.1145/3544548.3581170. URL: <https://doi.org/10.1145/3544548.3581170>.
- [11] Simson L. Garfinkel, Heather Richter Lipford. “Usable Security: History, Themes, and Challenges”. In: *Usable Security*. 2014. URL: <https://api.semanticscholar.org/CorpusID:39939011>.
- [12] Marthie Grobler, Raj Gaire, Surya Nepal. “User, Usage and Usability: Redefining Human Centric Cyber Security”. In: *Frontiers in Big Data* 4 (Mar. 2021). DOI: 10.3389/fdata.2021.583723.
- [13] Anti-Phishing Working Group. “Phishing activity trends report 2nd quarter 2023”. In: 2023. URL: [https://docs.apwg.org/reports/apwg\\_trends\\_report\\_q2\\_2023.pdf](https://docs.apwg.org/reports/apwg_trends_report_q2_2023.pdf).
- [14] Daniel Jampen, Gürkan Gür, Thomas Sutter, Bernhard Tellenbach. “Don’t click: towards an effective anti-phishing training. A comparative literature review”. In: *Human-centric Computing and Information Sciences* 10.1 (Aug. 2020), p. 33. ISSN: 2192-1962. DOI: 10.1186/s13673-020-00237-7. URL: <https://doi.org/10.1186/s13673-020-00237-7>.
- [15] Michael Koddebusch. “Exposing the Phish: The Effect of Persuasion Techniques in Phishing E-Mails”. In: *DG.O 2022: The 23rd Annual International Conference on Digital Government Research*. dg.o 2022. Virtual Event, Republic of Korea: Association for Computing Machinery, 2022, pp. 78–87. ISBN: 9781450397490. DOI: 10.1145/3543434.3543476. URL: <https://doi.org/10.1145/3543434.3543476>.
- [16] Ponnurangam Kumaraguru, Steve Sheng, Alessandro Acquisti, Lorrie Cranor, Jason Hong. “Teaching Johnny not to fall for phish”. In: *ACM Trans. Internet Techn.* 10 (May 2010). DOI: 10.1145/1754393.1754396.
- [17] Michal Kuniecki, Joanna Pilarczyk, Szymon Wichary. “The color red attracts attention in an emotional context. An ERP study”. In: *Frontiers in Human Neuroscience* 9 (Apr. 2015), p. 212. DOI: 10.3389/fnhum.2015.00212.
- [18] Jim McCambridge, John Witton, Diana Elbourne. “McCambridge J, Witton J, Elbourne DRSystematic review of the Hawthorne effect: new concepts are needed to study research participation effects. J Clin Epidemiol 67: 267-277”. In: *Journal of clinical epidemiology* 67 (Nov. 2013). DOI: 10.1016/j.jclinepi.2013.08.015.
- [19] Justin Petelka, Yixin Zou, Florian Schaub. “Put Your Warning Where Your Link Is: Improving and Evaluating Email Phishing Warnings”. In: *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. CHI ’19. Glasgow, Scotland Uk: Association for Computing Machinery, 2019, pp. 1–15. ISBN: 9781450359702. DOI: 10.1145/3290605.3300748. URL: <https://doi.org/10.1145/3290605.3300748>.
- [20] Stefan Rädiker, Udo Kuckartz. “Analyse qualitativer Daten mit MAXQDA”. In: Springer Fachmedien Wiesbaden GmbH, 2018. ISBN: 978-3-658-22094-5. DOI: <https://doi.org/10.1007/978-3-658-22095-2>. URL: <https://doi.org/10.1007/978-3-658-22095-2>.



- [21] Steve Sheng, Bryant Magnien, Ponnurangam Kumaraguru, Alessandro Acquisti, Lorrie Faith Cranor, Jason Hong, Elizabeth Nunge. “Anti-Phishing Phil: the design and evaluation of a game that teaches people not to fall for phish”. In: *Proceedings of the 3rd Symposium on Usable Privacy and Security*. SOUPS '07. Pittsburgh, Pennsylvania, USA: Association for Computing Machinery, 2007, pp. 88–99. ISBN: 9781595938015. DOI: 10.1145/1280680.1280692. URL: <https://doi.org/10.1145/1280680.1280692>.
- [22] Wenxuan Shi, Maoqiang Xie. “A Reputation-based Collaborative Approach for Spam Filtering”. In: *AASRI Procedia* 5 (2013). 2013 AASRI Conference on Parallel and Distributed Computing and Systems, pp. 220–227. ISSN: 2212-6716. DOI: <https://doi.org/10.1016/j.aasri.2013.10.082>. URL: <https://www.sciencedirect.com/science/article/pii/S2212671613000838>.
- [23] Md. Abu Ashraf Siddiq, Mohammad Arifuzzaman, M. S. Islam. “Phishing Website Detection using Deep Learning”. In: *Proceedings of the 2nd International Conference on Computing Advancements*. ICCA '22. Dhaka, Bangladesh: Association for Computing Machinery, 2022, pp. 83–88. ISBN: 9781450397346. DOI: 10.1145/3542954.3542967. URL: <https://doi.org/10.1145/3542954.3542967>.
- [24] Kutub Thakur, Md Liakat Ali, Muath A. Obaidat, Abu Kamruzzaman. “A Systematic Review on Deep-Learning-Based Phishing Email Detection”. In: *Electronics* 12.21 (2023). ISSN: 2079-9292. DOI: 10.3390/electronics12214545. URL: <https://www.mdpi.com/2079-9292/12/21/4545>.
- [25] Tom Tullis, Bill Albert. “Chapter 7 - Behavioral and Physiological Metrics”. In: *Measuring the User Experience (Second Edition)*. Ed. by Tom Tullis, Bill Albert. Second Edition. Interactive Technologies. Boston: Morgan Kaufmann, 2013, pp. 163–186. ISBN: 978-0-12-415781-1. DOI: <https://doi.org/10.1016/B978-0-12-415781-1.00007-8>. URL: <https://www.sciencedirect.com/science/article/pii/B9780124157811000078>.
- [26] “Visual Information Processing in Young and Older Adults.” In: *Frontiers in Aging Neuroscience* 11 (2019). DOI: 10.3389/FNAGI.2019.00116.
- [27] Jingguo Wang, Tejaswini C. Herath, Rui Chen, A. Vishwanath, H. Rao. “Research Article Phishing Susceptibility: An Investigation Into the Processing of a Targeted Spear Phishing Email”. In: *IEEE Transactions on Professional Communication* 55 (2012), pp. 345–362. DOI: 10.1109/TPC.2012.2208392.
- [28] Zuoguang Wang, Hongsong Zhu, Peipei Liu, Limin Sun. “Social engineering in cybersecurity: a domain ontology and knowledge graph application examples”. In: *Cybersecurity* 4.1 (Aug. 2021), p. 31. ISSN: 2523-3246. DOI: 10.1186/s42400-021-00094-6. URL: <https://doi.org/10.1186/s42400-021-00094-6>.
- [29] Rick Wash. “How Experts Detect Phishing Scam Emails”. In: *Proc. ACM Hum.-Comput. Interact.* 4.CSCW2 (Oct. 2020). DOI: 10.1145/3415231. URL: <https://doi.org/10.1145/3415231>.
- [30] Suzanne Widup, Marc Spitler, David Hylander, Gabriel Bassett. “2018 Verizon Data Breach Investigations Report”. In: Apr. 2018.

- [31] Michael S Wogalter, Vincent C Konzola, Tonya L Smith-Jackson. “Research-based guidelines for warning design and evaluation”. In: *Applied Ergonomics* 33.3 (2002). Fundamental Reviews in Applied Ergonomics 2002, pp. 219–230. ISSN: 0003-6870. DOI: [https://doi.org/10.1016/S0003-6870\(02\)00009-1](https://doi.org/10.1016/S0003-6870(02)00009-1). URL: <https://www.sciencedirect.com/science/article/pii/S0003687002000091>.

All links were last followed on June 11, 2024.

# A Appendix

## A.1 GitHub Repository

 [github.com/yunuseyyz/Bachelorthesis\\_Phishing](https://github.com/yunuseyyz/Bachelorthesis_Phishing)

This public repository contains:

- Email Samples
- Thunderbird Plugin Sourcecode
- Python Script Sourcecode
- Eyetracking Data Exports
- MAXQDA2024 Project Files
- Transcripts (*Note: Transcripts have been manually reviewed. Irrelevant and personal information have been removed.*)

## A.2 Study Protocol

### Pre-Setup

- Software Setup: Set up Mozilla Thunderbird on the participant's computer, installing the necessary addon that will be used during the study. Ensure that all settings are optimal for the tasks ahead. Install Tobii Pro Lab for extensive eye tracking tools.
- Seating Arrangement: Arrange for a non-rolling chair to minimize movement and ensure more accurate eye tracking data.
- Eye Tracker: Attach and setup the eye tracker.
- Moderator's Setup (optional): Arrange a second screen for the moderator in Tobii Pro Lab to monitor and observe the eye tracking data in real-time.

### Welcome and Consent

- Introduction to the Study:

*"Hello and welcome! Thank you for participating in our study today. I will be guiding you through the process and the session will last about 30 minutes. You will get a small task where you need to sort through a fictional inbox, I will explain in detail when we get there. While doing that we will have an eye tracker running, which will track and record your eye movements. After the sorting task I will ask you a couple of questions."*

- Consent Form:

*"Before we begin, please take a moment to read this consent form. If everything is clear, please sign the form at the bottom."*

### **Before Beginning**

- Eye Tracker Calibration and Audio Recording

*"Our first step will be to calibrate the eye tracker for a more accurate data capture. Just follow the instructions on the screen. This will take around 30 seconds. (...) "Thank you for that. Now we can begin with the task. I will also start the audio recording now."*

### **Phase 1: Email Sorting (Approximately 10 Minutes)**

- Task Description:

*"In this part, you will act as an employee at 'Cloudtech'. Your task is to sort through your emails as you would in your work inbox. You only need to decide whether to delete each email or keep it. No other actions are necessary. "As you sort through the emails, please try to verbalize your thoughts. Say out loud what you are thinking about each decision. This helps us understand your reasoning. Feel free to open any links or attachments in the emails as you see fit. Interact with the emails as you would normally. Let me know when you are done."*

- Task Finished:

*"Alright, thank you. Now we will get to the second phase of the study."*

### **Part 2: Detailed Interview (Approximately 10 Minutes)**

- Debriefing:

*"Before we move on to the interview I will debrief you on what this study is about. Our study is about how users interact and engage with different types of visual warnings in email clients. For that we first collect eyetracking data for an objective view. After that we conduct a small semi-structured interview to ask you about your subjective experiences with the warnings. Before we move on, I will briefly show you through the warnings again. (...) Alright now to the questions."*

- Interview:

*See appendix A.3.*

### **Conclusion of the Session**

- Compensation:

*"Alright. Thanks a lot for participating! You will be compensated with 10€ via PayPal or bank transfer. I will send you a link, where you can enter your payment information. Alternatively you can opt for 1 MMI point, if you are a Media Informatics Student."*

- Thank You and Farewell:

*"Your input is very valuable to our research. Thank you again for your time today. Have a great day!"*

## A.3 Interview Questions

These are the questions posed to the participants in the second phase. Note that the interview was semi-structured so deviations and follow up questions were possible.

### **General Perception:**

- How did you generally perceive the warnings? Did they immediately stand out?
- Follow-up questions, e.g., if "Good", ask why exactly good or bad.
- Were there elements in the warnings that particularly stood out to you? What caught your attention the most?
- What role did the warning notices play in your decisions?

### **Effectiveness of the Warnings:**

- How do you assess the effectiveness of the various warning notices presented to you during the study?
- Were there any warning notices you found particularly helpful or less helpful?
- Each warning notice had 2 versions (simple/detailed). Which version do you prefer, and why? (Show the warnings again)

### **Design of the Warning Notices:**

- How do you evaluate the design of the warning notices (e.g., color, placement, animations)? Did these influence your attention?
- How do you assess interactive elements in warning notices (e.g., greeting warning)?
- Do you have any suggestions for improving the design of the warning notices?

### **Personal Behavior:**

- Were there any pieces of information in the warnings that you hadn't paid attention to before?
- How do you usually handle emails once there's a suspicion of phishing? Or what specifically do you look at once there's a suspicion?

### **Feedback:**

- Do you have any suggestions on how we could improve the study in the future?
- Do you have any other questions or comments?

## A.4 Warning Designs

Below is an overview of version 1 of each warning type.

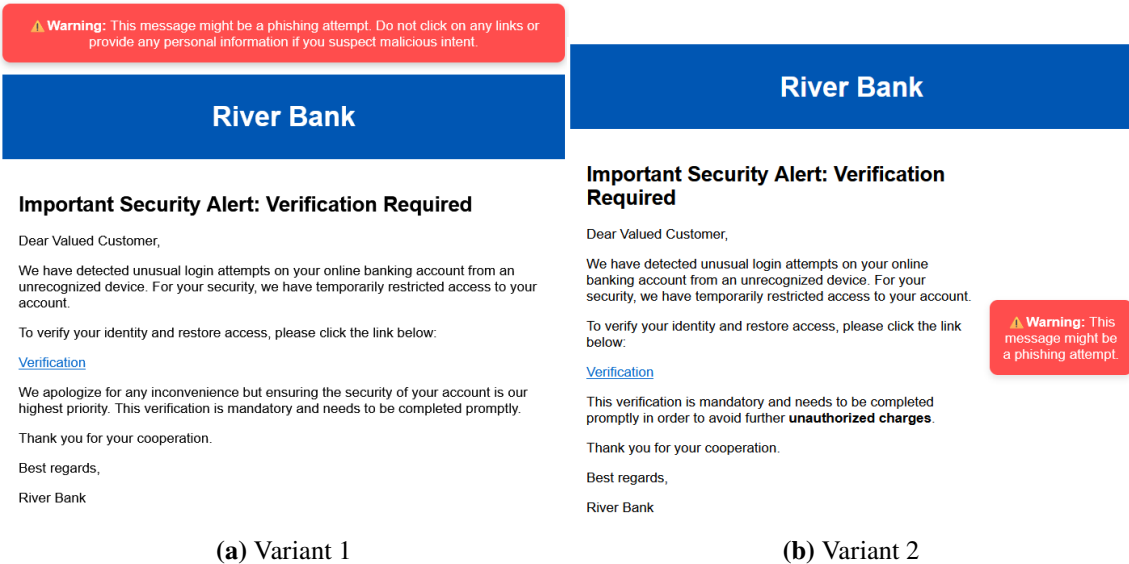


Figure A.1: Generic Banner (Version 1)

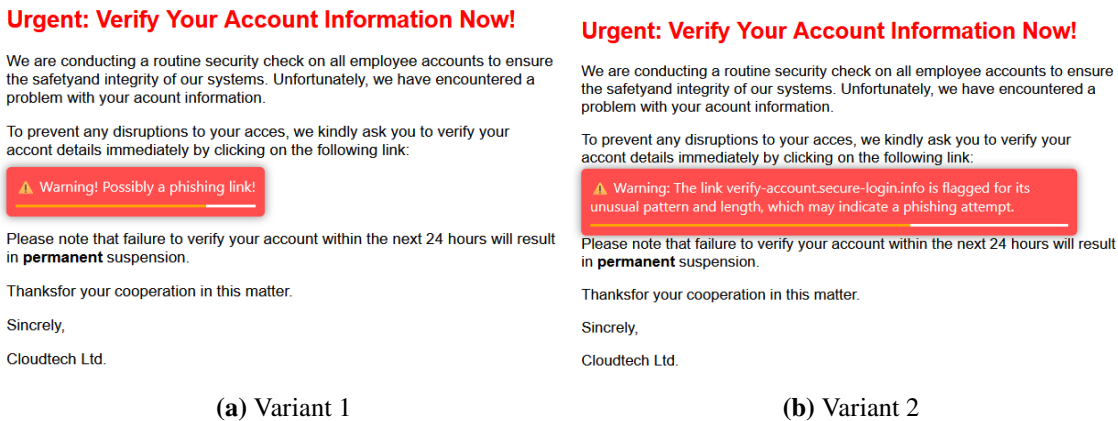
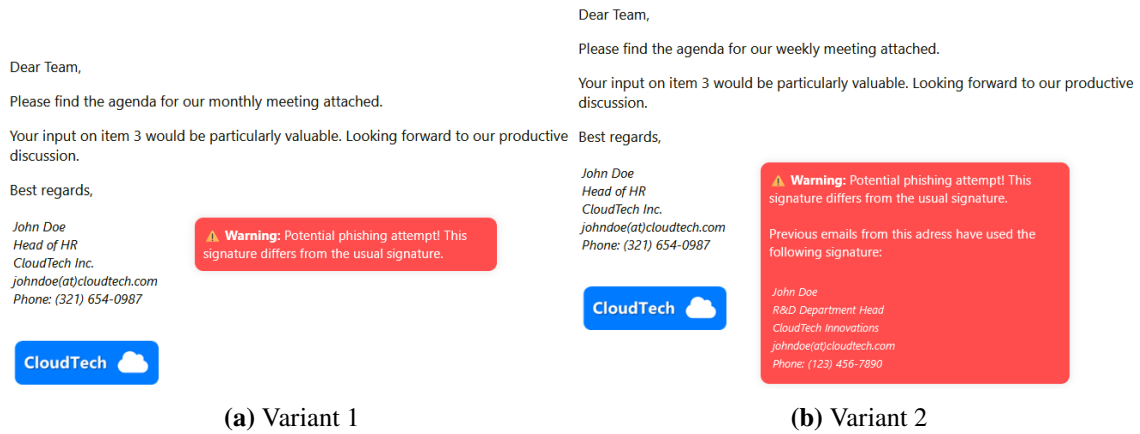


Figure A.2: Tooltip Warning on Link Hover (Version 1)



**Figure A.3: Signature-Specific Warning (Version 1)**



**Figure A.4: Greeting-Specific Warning (Version 1)**

## A.5 QDA Codebook

Below is the list of codes developed and utilized during the qualitative data analysis process of our study.

- Design Response: This code looks into the overall aesthetic and functional reception of the warning designs by the participants.
  - Clarity and Legibility: Evaluates how easily the warnings can be read and understood.
  - Impact of Animations: Assesses how motion and dynamic elements in the warnings affect user attention and understanding.
  - Placement: Considers the effectiveness of where warnings are located within the email interface.
  - Potential Improvements: Gathers participant suggestions on how to enhance the warning designs.
  - User Attention: Analyzes how conspicuous or inconspicuous these warnings are to the users.
    - \* Conspicuous: Warnings that are immediately noticeable.
    - \* Inconspicuous: Warnings that tend to be overlooked.
- Novelties for Participants: Identifies aspects of the phishing warnings that were particularly novel or unexpected to the participants.
- Perceived Helpfulness: Explores participants' perception on each warnings helpfulness.
  - Negative: Participant feedback that highlights shortcomings or ineffective aspects of the warnings.
  - Positive: Positive reactions that underscore the effectiveness and utility of the warnings.
  - Preferred Complexity: Explores participants' preferences for the complexity of the information provided in the warnings.
    - \* More Details: Preference for detailed and informative warnings.
    - \* Simplicity: Preference for straightforward, minimal warnings.



## A.6 Eye Tracking Heatmap

Included here is an eye tracking heatmap using the gaze data of all participants. Although not directly referenced in the main findings, this heatmap may serve as a useful resource for other researchers interested in the visual patterns of user interaction. It also serves an educational purpose, illustrating the application of eye tracking technology in research.



**Figure A.5:** Accumulated eyetracking heatmap, generated in Tobii Pro Lab



### **Declaration**

I hereby declare that the work presented in this thesis is entirely my own and that I did not use any other sources and references than the listed ones. I have marked all direct or indirect statements from other sources contained therein as quotations. Neither this work nor significant parts of it were part of another examination procedure. I have not published this work in whole or in part before. The electronic copy is consistent with all submitted copies.

---

place, date, signature