IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 12, Issue 9, September 2025

DOI: 10.17148/IARJSET.2025.12912

A Multi-Stage Behavioral Intervention Framework for Phishing Prevention in Remote Teams Using AI-Driven Contextual Nudges

Sujay S[1], Gresika N [2], Chitturi Naga Satyam [3]

Assistant Professor, Dept. of MCA, Surana College (Autonomous) Bengaluru, India[1]

PG Student, Dept. of MCA, Surana College (Autonomous) Bengaluru, India[2]

PG Student, Dept. of MCA, Surana College (Autonomous) Bengaluru, India[3]

Abstract: Phishing has emerged as one of the most persistent and evolving threats to cybersecurity, and its impact has grown even more severe with the widespread adoption of remote work. Employees working from home or hybrid environments often lack immediate IT assistance and rely heavily on digital platforms such as email, messaging apps, and cloud services. This makes them prime targets for attackers who exploit psychological triggers such as urgency, curiosity, and authority. Recent developments, including QR-code–based quishing and large language model (LLM) generated phishing emails, have enabled adversaries to create highly convincing messages that bypass traditional spam filters and deceive users at alarming rates. Studies shows that over 30% of participants in controlled experiments fall victim to these advanced phishing strategies, highlighting the limitations of conventional awareness programs and static filtering technologies. Although advanced artificial intelligence (AI) techniques, particularly hybrid models combining BERT and CNN, have achieved near 97.5% accuracy in classifying malicious content, these systems typically function in the background and do not provide real-time, user-facing assistance. This gap leaves individuals vulnerable at the precise moment when they must decide whether to trust or reject a suspicious message. Research in behavioral science has consistently demonstrated that contextual nudges, micro-prompts, and reinforcement mechanisms are more effective in shaping long-term secure practices compared to one-off training sessions or financial incentives. This paper proposes a multistage behavioral–AI framework designed specifically for remote workers to address this gap. The framework integrates four stages: Awareness Nudges, Micro-Actions, Reinforcement, and AI-Driven Contextual Alerts. Each stage complements the others—nudges capture attention, micro-actions promote critical reflection, reinforcement cultivates long-term secure habits, and contextual alerts provide AI-powered warnings at high-risk moments. Unlike prior approaches that address technical or behavioral dimensions in isolation, this framework merges them into a single, interactive model. By aligning advanced detection capabilities with human decision-making processes, the framework aims to reduce click-through rates, encourage proactive reporting, and foster a culture of security-conscious behavior among remote employees. In doing so, it provides not just a technical solution but also a sustainable strategy that adapts to evolving phishing threats while strengthening organizational resilience.

Keywords: Phishing Defense, AI-Guided Behavioral Nudges, Real-Time Security Alerts, QR-Code Phishing (Quishing), Reinforcement Mechanisms, Contextual Micro-Prompts, Hybrid BERT–CNN Detection Model.

I. INTRODUCTION

Phishing remains one of the most widespread and damaging cybersecurity threats, affecting organizations across industries regardless of size or sector. With the global shift toward remote and hybrid work models, the challenge has become even more complex. Remote employees, often operating outside enterprise-grade security perimeters, rely heavily on digital communication platforms such as email, messaging services, and collaboration tools. While these technologies improve flexibility and productivity, they also expand the attack surface for adversaries who specialize in social engineering. Unlike on-site employees, remote users may not have instant access to IT teams or peer support when confronted with suspicious communication, making them more susceptible to deception. Attackers have adapted quickly to this new environment, employing advanced strategies such as large language model (LLM)generated phishing emails and QR-code based quishing. Both methods have demonstrated alarming success rates in controlled studies, with a significant portion of participants deceived by content that mimics authentic organizational communication [1]. These developments show that phishing is no longer limited to poorly crafted emails riddled with spelling errors but has evolved into a sophisticated, AI-assisted operation capable of bypassing technical defenses and exploiting human behavior.

This work is licensed under a Creative Commons Attribution 4.0 International License

Traditional security mechanisms such as spam filters, antivirus tools and occasional training sessions play an important role, but they are no longer enough to stop today's advanced high personalized attacks. Filters can block some threats but miss certain dangerous emails, while training is done only once in a while, not at the exact moment users face a risky situation. AI based detection models, such as BERT combined with CNN, have shown strong results in identifying phishing attempts [9]. However, these tools work quietly in the background, and they do not guide the users on what to do when they encounter a suspicious message. On the other hand, behavioral studies show that giving users real-time nudges, small actions and continuous reminders can reduce risky behavior effectively. Nudges help users to notice warning signs, small pauses can also make them to think before clicking, and reinforcement mechanisms help to build safer online habits over time. These behavior-focused methods are rarely combined with technical security systems, which means organizations miss the chance to provide stronger, all around protection.

This paper introduces a multi-stage behavioral–AI framework specifically designed for remote workers. The framework consists of four interconnected stages: **(1) Awareness Nudges, (2) Micro-Actions, (3) Reinforcement, and (4) AI-Driven Contextual Alerts.** Together, these layers provide continuous, adaptive support at the precise moments when users are most vulnerable. The novelty of this approach lies in bridging the divide between advanced machine learning–based phishing detection and human decision-making psychology. The ultimate goal is to not only mitigate immediate phishing risks but also to establish a security-conscious culture among distributed teams. By aligning cutting-edge AI detection with user-centric behavioral interventions, this framework offers a scalable and future-ready solution that can evolve alongside emerging threats.

II. MOTIVATION AND SCOPE

A. Motivation

The transition to remote work has improved organizational flexibility but, at the same time, has expanded the risk surface for phishing attacks. Employees working remotely often lack direct IT support and rely heavily on digital platforms such as email and messaging services, which makes them easy targets for deception. Recent studies show that phishing attempts using LLM-generated emails and QR-code quishing have achieved success rates of more than 30% in controlled experiments [1], [3]. Although AI models like BERT combined with CNN deliver excellent detection accuracy [9], they generally operate silently in the background and do not provide timely assistance when a user is about to make a risky decision. In contrast, behavioral research emphasizes that embedded, real-time nudges and micro-prompts can significantly reduce the effectiveness of phishing campaigns [2], [5]. This contrast underlines the necessity of a framework that not only detects threats but also actively supports users at the most critical decision points. The purpose of this research is to bridge that gap by combining AI-driven detection mechanisms with behavior-focused interventions.

B. Scope

This study focuses only on phishing methods that use emails and QR codes(quishing) in remote work settings after 2023. These two attacking methods are highlighted because they are both very common and highly misleading. Email continues to be the main way for people to communicate at work, and QR codes have quickly become popular with attackers since they are convenient and can bypass many traditional security checks.

Other phishing approaches, such as smishing(via SMS), vishing(via phone calls), or large-scale infrastructure attacks, are not covered in this research. Although these are also serious threats, they function in different ways and involve different user behaviors compared to email and QR-code phishing. Including them would make the study too broad and reduce its depth. By keeping the focus on emails and QR codes, this research can provide a more detailed look at how AI-based detection systems and behavioral strategies can work together to protect remote workers. This focused approach ensures that recommendations are practical and useful for organizations dealing with today's phishing challenges.

III. PROBLEM DEFINITION

Remote employees are at high risk to phishing attacks because they often work without direct IT support and depend more on digital platforms for communication and daily tasks. Recent studies show that techniques like QR-code–based quishing and phishing emails generated by large language models (LLMs) are highly successful in deceiving users [1],[3].AI-based detection models, such as BERT and CNN combination have shown strong performance in spotting these threats[9]. However, they usually function quietly in the background and do not provide users with real-time guidance when they are about to take risky actions. On the other hand, behavioral nudges have been proven to improve user awareness and decision-making [2], but these are rarely integrated with technical detection systems. This gap creates a weakness in current defenses. Therefore, there is a need for real-time intervention framework that merges AI-driven detection with behavioral strategies, so that employees receive timely support at the very moment they are most likely to fall victim to phishing.

DOI: 10.17148/IARJSET.2025.12912

IV. PROPOSED FRAMEWORK

To bridge the gap between phishing detection and immediate user actions in remote work environments. This study introduces a four-stage behavioral-AI framework. This framework integrates real-time prompts, psychological indications, and intelligent alerts to guide users at the exact moment of critical security decisions. Unlike earlier studies that examined nudges [2], quishing prevention [1], or AI-driven classification methods [9] as standalone solutions, our framework integrates these elements into a single, real-time intervention model specifically designed for remote workers.

A. Stage 1: Awareness Nudges:

Brief, subtle inline notifications like "External sender—verify before clicking" show up in email clients or messaging software to increase user vigilance without interrupting workflow.

Example: A worker gets a message from it-support@securemail.net. A nudge shows up underneath the email subject "This is an external sender. Double-check authenticity before taking action." Evidence supporting this: Lain et al. [2] demonstrated inline nudges enhance vigilance more than standard training, so inspiring inclusion within our framework.

B. Stage 2: Micro-Actions:

Before performing potentially dangerous actions (such as clicking a link or scanning a QR code), users are presented with minimal, real-time prompts like "Do you trust this source?" or "Preview link destination?" Illustration: Raj hovers above a QR code he was sent through Slack. A micro-action popup: "QR source unknown — Check destination before scanning?" Evidence for support: Weinz et al. [1] discovered that users tend to click on QR-based phishing links without confirming them, so intervention at the point of decision is necessary. Supporting evidence: Lain et al. [2] found that lightweight reinforcements are more effective than annual awareness campaigns.

C. Stage 4: AI-Driven Contextual Alerts:

A hybrid detection engine based on BERT + CNN analyzes the tone, urgency, and structure of incoming messages. Context-aware alerts notify users about high-pressure language or suspicious content.

Example: A message reads: "Your account will be deactivated unless you act right away!" A pop-up warning appears in real-time: "Urgent tone used. Use caution." Evidence: Saha Roy et al. [10] suggested contextual NLP-based warnings; we extend this by integrating them with our end-user real-time feedback system.

V. REAL-WORLD SCENARIO: REMOTE WORK & THR PHISHING SURGE

The rapid shift to remote work during and after the pandemic significantly increased employee exposure to phishing attacks, especially in home settings where enterprise-grade security measures are absent. In mid-2023, a major U.S. energy company experienced a sophisticated QR-code–based phishing (quishing) attack. The attackers posed as Microsoft and distributed emails warning employees about the imminent expiration of multi-factor authentication (MFA). These messages contained QR codes that redirected unsuspecting users to a counterfeit Microsoft login portal. Ultimately, more than 100 corporate accounts were compromised, as employees were tricked by the urgency of the messages and bypassed existing email filters — particularly when accessing them on mobile devices outside the company's security perimeter.

At the same time, AI-generated phishing has emerged as a new and effective threat vector. In late 2023, IBM X-Force Red conducted a proof-of-concept where ChatGPT was instructed to generate phishing emails targeting employees of a global healthcare organization. Despite minimal input, the AI produced messages that closely mirrored the organization's internal communication style, including urgency and tone. The outcome was striking: approximately 11% of recipients interacted with the AI-crafted emails, a rate nearly identical to the 14% engagement observed with phishing emails designed by professional attackers. This shows that even short AI prompts can yield highly convincing phishing attempts, making them as effective as human-written attacks. It also demonstrates why traditional detection methods alone are insufficient and why real-time, behavior-based user interventions are necessary.

VI. HOW OUR BEHAVIORAL-AI FRAMEWORK FILLS THESE GAPS

Our framework is engineered to address actual issues faced by remote workers in phishing attacks:

1) Stage 1: AI Detection:

A robust AI system incorporating BERT and CNN models assists in detecting threats in real time by analyzing language patterns, urgency, and emotional tone—something simple filters cannot do.

2) Stage 2: Awareness Nudges:

Real-time reminders such as "External sender—verify before clicking" are displayed in emails or messaging apps, subtly reminding users to consider before acting.

3) Stage 3: Micro-Actions**:**
Little queries like "Are you sure about this link?" interrupt risky behavior, prompting users to pause and reflect on their next step.

4) Stage 4: Reinforcement:
After safe decisions, users receive immediate reward in the form of badges or quick quizzes. This makes them interactive and promotes secure behavior on a consistent basis, not only after training sessions.

TABLE 1
COMPARISON: TRADITIONAL VS. BEHAVIORAL–AI FRAMEWORK RESPONSE TO REMOTEWORK PHISHING

Parameter	Traditional Training	AI-Powered Monitoring
Surge in phishing from remote networks	Relies on basic network filters	AI-powered BERT + CNN model analyzes tone and urgency for accurate alerts
Lack of firewall protection outside office	Periodic security training, often delayed	Real-time nudges appear during risky interactions like email or QR scan
High volume of phishing messages confusing users	General awareness sessions conducted quarterly	Micro-interactions + reinforcement help users stay alert consistently

VII. LITERATURE REVIEW

- Lain et al. [2] investigated the impact of embedded phishing awareness mechanisms—such as nudges and contextual feedback—on users' decision-making. Their findings indicated that real-time cues had a significant impact on detection and user confidence in managing suspicious emails.
- Weinz et al. [1] examined upcoming phishing trends like quishing (QR-based phishing) and email generation using LLM. The results showed how more than 30% of users were tricked in controlled simulations, demonstrating the escalating advanced nature of phishing attacks.
- Saha Roy et al. [10] suggested a real-time phishing alert system based on contextual information, improving alert effectiveness by incorporating behavioral data—providing a bridge between technical and psychological countermeasures.

VIII. METHODOLOGY

This research takes a multi-step approach merging behavioral understanding with AI-driven detection for phishing mitigation in remote settings

A. Literature Selection Criteria:

We examined peer-reviewed articles and prominent conference proceedings from 2020 to 2025. Priority was given to research dealing with newly emerging phishing tactics (e.g., quishing, LLM-generated phishing), AI-based detection (such as BERT and CNN models), and behavioral cyber defense measures such as nudges, micro-actions, and reinforcement. Excluded were studies unrelated to phishing or not directed towards email- or QR-based threats.

B. Research Design and Approach:

• Literature Review: We analyzed prominent studies on phishing attack development, susceptibility of users, and AI-enabled classification. For instance, Lain et al. [2] showed the effectiveness of nudges in enhancing email security behavior, whereas Weinz et al. [1] emphasized the increasing menace of QR-code and LLM-based phishing.

• Threat Modeling: Taking cue from human-centric models such as SoK by Zhuo et al. [16], we identified the pivotal moments at which users tend to be susceptible to phishing—particularly within disconnected remote environments.

- Framework Design: A four-stage model was created:

Stage 1: Awareness Nudges Real-time messages within communication tools that encourage caution.

Stage 2: Micro-Actions Lightweight reminders prior to risky clicks to break up impulsive choices.

Stage 3: Reinforcement Positive reinforcement such as badges and mini-quizzes to reinforce safe habits.

Stage 4: AI Contextual Alerts A BERT + CNN system identifies phishing content through tone and intent analysis (drawing from methods by Al Subaiey et al. [9] and Saha Roy et al. [10]).

- Prototype Implementation: The system shall be implemented as a light browser plug-in or add-on integrated into email/chat services, merging UI prompts with backend AI computation.

- Evaluation Metrics**:**
  1. Phishing click-through rate (CTR)
  2. User reaction time to nudges and AI warnings
  3. treatment with reinforcement components (e.g., badges)
  4. truthfulness of phishing detection in real-time

- Experimental Setup: A 50–100 member user group will engage in remote work task simulation with phishing exposure. Control (no framework) and test groups (with framework) will be examined.

- Comparative Study**:** Basic spam filters and traditional awareness training like current tools will be compared to the system proposed.

- Feedback and Validation**:** Behavioral data and usability feedback will be obtained to evaluate the effects of alerts and nudges on decision-making.

C. Evaluated Metrics in Our Future Work:

Although a complete deployment and test by users are future work, we describe the evaluated metrics in our future work to determine whether the suggested behavioral-AI framework is effective once deployed.

TABLE 2
PLANNED EVALUATION METRICS FOR BEHAVIORAL–AI FRAMEWORK

Metric	**Description**	**Purpose**
Click-through Rate (CTR)	Percentage of users who click on phishing links in test scenarios	Measures effectiveness of nudges and alerts in preventing risky actions
Nudge Response Time	Time taken by users to respond to contextual prompts or inline warnings	Assesses real-time responsiveness and decision-making delay
Alert Accuracy	Precision and recall of BERT + CNN model in detecting phishing attempts	Evaluates AI performance in practical, dynamic environments
Engagement Rate	Frequency of user interaction with reinforcement tools (badges, quizzes)	Measures sustained behavioral engagement
False Positive/Negative Rates	Incorrectly flagged or missed phishing emails by the AI module	Ensures balance between safety and usability

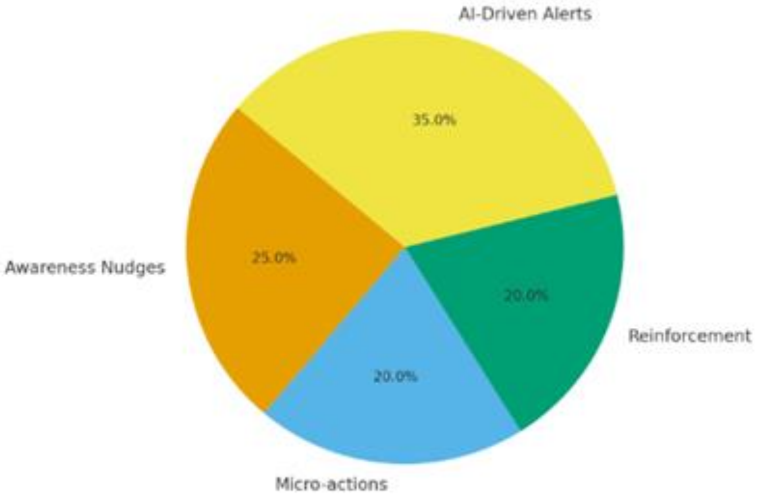

Figure 2.1 Contribution of Behavioral-AI Framework Stages

TABLE 3
ACCURACY OF DETECTION ALGORITHMS

Algorithm/Model	Accuracy (%)	Strengths	Limitations
Traditional Spam Filters	~85%	Simple, widely available, low processing overhead	Misses advanced phishing (LLM-generated, QR-based), high false negatives
CNN	~92%	Good at recognizing structural patterns in phishing emails and URLs	Less effective with context-heavy phishing messages
BERT	~95%	Excellent at capturing language semantics, tone, and intent	Computationally heavy, may struggle with multi-modal inputs like QR codes
Hybrid BERT + CNN	~97.5%	Combines semantic (BERT) + structural (CNN) features, high precision and recall	Resource intensive, requires large datasets and processing power

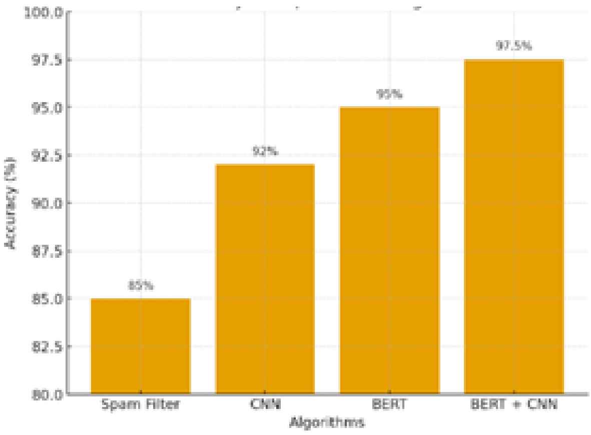

Figure 2.1 Accuracy Comparison of Algorithms

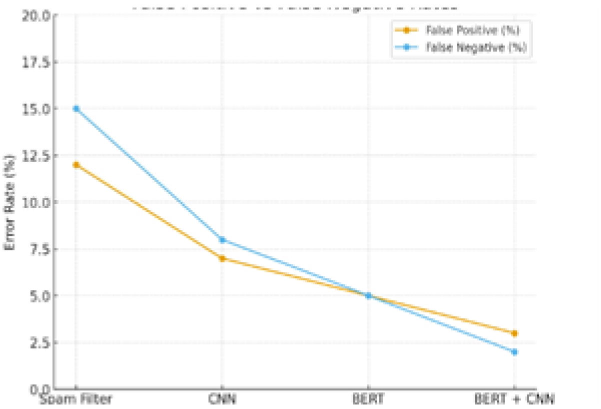

Figure 2.1 False Positive vs Negative Rates

IX. CHARTS AND GRAPHS

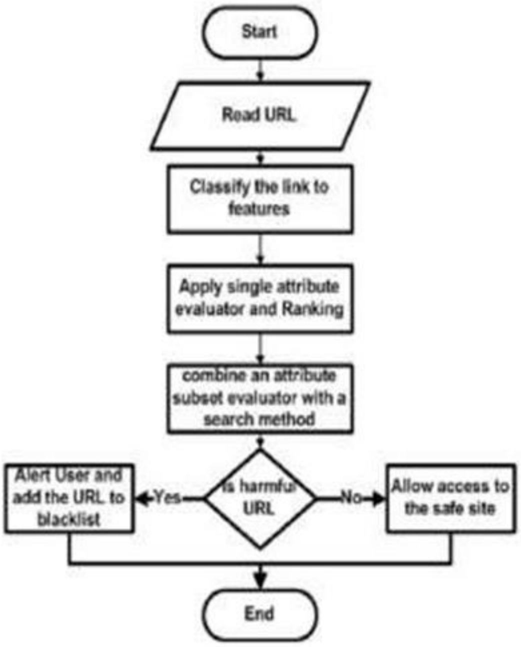
Start
Read URL
Classify the link to features
Apply single attribute evaluator and Ranking
combine an attribute subset evaluator with a search method
Is harmful URL
Yes
No
Alert User and add the URL to blacklist
Allow access to the safe site
End

Fig. 1. URL Evaluation Flowchart

This figure corresponds to Stage 4 of the framework (AI-Driven Contextual Alerts). It illustrates how the system evaluates URLs before action is taken, showing the sequence of classification and filtering steps that help detect potentially malicious links in real time.

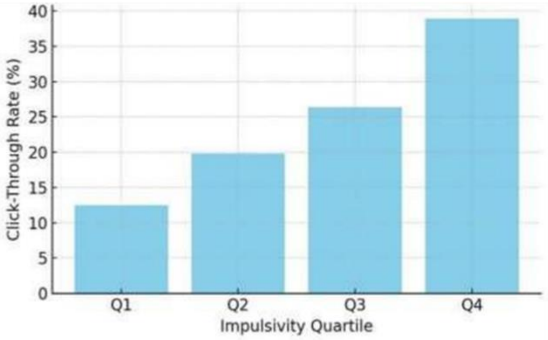

Fig. 2. Correlation Between User Impulsivity and Phishing Risk

This diagram validates the importance of Stages 2 and 3 (Micro-Actions and Reinforcement). It shows how users with higher impulsivity tend to have higher phishing click-through rates, underscoring the need for personalized nudges and repeated reinforcement to strengthen secure behavior. Impulsive users, which confirms the necessity of personalized nudges and reinforcement mechanisms.

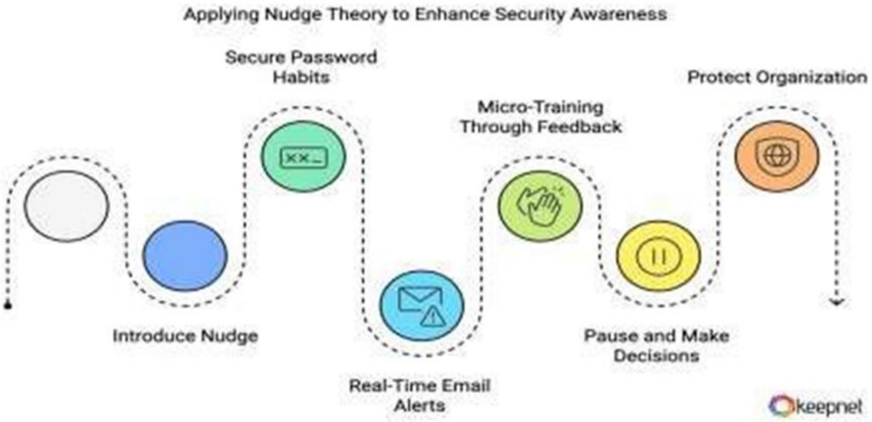

Fig. 3. Behavioral Security Flow with Nudge Theory

This workflow brings together nudges, decision pauses, and reinforcement loops to illustrate the end-to-end behavioral model. It demonstrates how users are guided from initial exposure to a suspicious message, through momentary hesitation, to corrective action reinforced by feedback.

X. FUTURE ENHANCEMENTS

Since phishing attacks are always changing, this system aims to keep defenses up to date by adjusting itself as needed. It also improves by learning how each person behaves and giving helpful reminders that match their habits and needs. For example, if a user keeps ignoring or reacting poorly to unclear warnings, the system will start using clearer alerts, helpful reminders based on the situation, or interactive messages to get their attention and help them respond better. ([2], [14]). The system can change its alerts based on the device, the time, or what the user has been doing. This way, it gives the right help at the right time. Using it for a while helps users stay aware without getting tired of too many alerts. These can mimic high-stress phishing attempts in a secure, isolated environment, allowing users to rehearse real-time responses and decisions without actual risk. For example, situations may include prompt payment demands from a simulated CEO, deceptive MFA expiration reminders, or harmful QR codes. With time, such experiential education has been found to improve memory recall and long-term sensitivity ([4], [13]).

Repeating these exercises every now and then, with progressively greater sophistication, can guarantee that fresh as well as veteran staff members stay immune to various phishing efforts. The platform should also develop as a mobile-first product, providing timely prods and micro learning within everyday workflows-even on mobile. It may incorporate brief training clips, interactive simulations, or simple reminders when employees open corporate accounts from mobile. This enables ongoing reinforcement and shielding, particularly for remote workers without around-the-clock IT oversight. It's important to design the system mainly for mobile devices because many employees use their phones for work communication. Showing helpful hints, quick training videos, and reminders right inside the phone's apps makes it easier for users to stay safe, even when they are not in the office. Security teams can also send instant alerts when a new phishing attack is found, helping users stay protected. To make learning about security fun, features like quizzes, badges, and leaderboards can be added. For example, employees can earn badges for spotting fake phishing emails or for always following safe habits. Also, by using real-time threat information, the system can adjust itself to new types of attacks like QR code scams or fake phone calls [1],[10]. Connecting the system with global cybersecurity networks helps it quickly recognize new threats and update protections within minutes. This way, the system stays strong and ready to protect against phishing attacks that keep changing.

CONCLUSION

This paper presents an integrated, behavior-oriented AI framework that combines psychological interventions with advanced detection techniques to safeguard remote users from phishing threats. Unlike traditional methods that mainly depend on the scheduled awareness sessions or reactive filtering tools, the proposed framework focuses on real-time and

DOI: 10.17148/IARJSET.2025.12912

adaptive engagement through four interconnected elements: Awareness Nudges, Micro Actions, Reinforcement, and AI-Driven Contextual Alerts. Each of these helps protect users from risks. Awareness nudges guide people by pointing out warning signs when they need to make a decision. Micro actions add short pauses that make users think before clicking on something risky or downloading files. Reinforcement helps users build good habits by giving them regular reminders. Finally, AI-driven alerts spot and warn about real dangers as they happen. All these together create a strong, easy-to-use system that helps keep users safe from phishing attacks better than older methods. This model is supported by earlier research showing the effectiveness of contextual nudges in shaping user behaviour and the accuracy of hybrid AI techniques such as BERT and CNN in detecting phishing content. However, the real strength of this framework lies for validations in workplace environments, where user feedback can be taken to refine its usability and effectiveness. Future developments could include personalized learning paths for individual users, mobile-first integration to protect employees who depend heavily on smartphones, gamified tools to sustain interest and motivation. By combining real-time threat intelligence with reinforcement learning, the system can remain resilient against emerging risks such as QR-code phishing or AI-powered deepfake voice scams. In the long run, this framework not only mitigates phishing threats but also promotes a security-conscious culture across organizations, a quality that is particularly critical for the modern remote workforce.

REFERENCES

[1]. M. Weinz, N. Zannone, L. Allodi, and G. Apruzzese, “The impact of emerging phishing threats: Assessing quishing and llm-generated phishing emails against organizations,” arXiv preprint arXiv:2505.12104, May 2025.

[2]. D. Lain, T. Jost, S. Matetic, K. Kostiainen, and S. Capkun, “Content, nudges and incentives: A study on the effectiveness and perception of embedded phishing training,” arXiv preprint arXiv:2409.01378, December 2024.

[3]. F. Heiding, S. Lermen, A. Kao, B. Schneier, and A. Vishwanath, “Evaluating large language models’ capability to launch fully automated spear phishing campaigns: Validated on human subjects,” arXiv preprint arXiv:2412.00586, November 2024.

[4]. X. Chen, M. Sacre, G. Lenzini, S. Greiff, V. Distler, and A. Sergeeva, “The effects of group discussion and role-playing training on self-efficacy,´ support-seeking, and reporting phishing emails,” arXiv preprint arXiv:2402.11862, February 2024. [Online].. Available: https://arxiv.org/abs/2402.11862

[5]. S. Zheng et al., “Checking, nudging or scoring? evaluating e-mail user security tools,” in USENIX Symposium on Usable Privacy and Security (SOUPS), 2023.

[6]. V. Distler et al., “The influence of context on response to spear-phishing attacks: an insitu deception study,” in ACM Conference, 2023.

[7]. D. Baltuttis, “Effects of visual risk indicators on phishing detection behavior: An eyetracking experiment,” Computers Security, 2024.

[8]. A. Singkeruang, S. Susanto, and N. Saeni, “Mitigating the risk of qushing threats using the security behavior intentions scale (sebis),” ResearchGate, 2025.

[9]. A. Al Subaiey et al., “Novel interpretable and robust web-based ai platform for phishing email detection,” arXiv preprint arXiv:2405.11619, 2024.

[10]. S. Saha Roy, C. Torres, and S. Nilizadeh, “Explain, don’t just warn! – a real-time framework for generating phishing warnings with contextual cues,” arXiv preprint arXiv:2505.06836, May 2025.

[11]. L. Stalans, “Predicting phishing victimization: Comparing prior victimization, cognitive, and emotional styles, and vulnerable or protective e-mail strategies,” Journal of Digital Forensics, Security and Law, 2023.

[12]. S. Williamson, “The era of artificial intelligence deception: Unraveling the complexities of false realities and emerging threats of misinformation,” Information (MDPI), 2024.

[13]. Z. Agha, “A systematic review on design-based nudges for adolescent online safety,” Computers in Human Behavior Reports, 2024.

[14]. J. Sumner, “Developing an artificial intelligence-driven nudge intervention to improve medication adherence: A human-centred design approach,” JMIR, 2023.

[15]. L. Huang, S. Jia, E. Balcetis, and Q. Zhu, “Advert: An adaptive and data-driven attention enhancement mechanism for phishing prevention,” arXiv preprint arXiv:2106.06907, 2021.

[16]. S. Zhuo, R. Biddle, Y. Koh, D. Lottridge, and G. Russello, “Sok: Human-centered phishing susceptibility,” arXiv preprint arXiv:2202.07905, 2022.

[17]. “A comprehensive survey: Evaluating the efficiency of artificial intelligence and machine learning techniques on cyber security solutions,” ResearchGate, 2024.

[18]. “Major energy company targeted in large qr code phishing campaign,” Cofense Blog, 2024.