OPEN ACCESS

Kashf Journal of Multidisciplinary Research
Vol: 03 - Issue 01 (2026)
P-ISSN: 3007-1992 E-ISSN: 3007-200X
https://kjmr.com.pk

# CYBER SECURITY ESSENTIALS FOR SOCIAL WORKERS: MITIGATING AI-DRIVEN SOCIAL ENGINEERING THROUGH SURVEY-BASED BEHAVIORAL INTERVENTIONS

***Dr Anum Ali[1], Muhammad Aslam[2]***

[1]*Schaefer School of Engineering, Stevens Institute of Technology, New York, USA.*
*Al-Khwarizmi Institute of Computer Science (KICS), UET, Pakistan.*

[2]*University of Engineering and Technology, Lahore, Pakistan.*

[*]***Corresponding Author:*** (aali17@stevens.edu)
***DOI:*** (*https://doi.org/10.71146/kjmr962*)

**Article Info**



**Abstract**

The rapid evolution of artificial intelligence has introduced sophisticated cyber security threats, particularly in the form of AI-driven social engineering. Social workers, who handle highly sensitive and confidential client data, are increasingly targeted by these advanced attacks despite often lacking extensive technical training. This paper explores the intersection of human behavior, survey data analytics, and cyber security to propose a comprehensive framework designed specifically for the social work sector. By leveraging hypothetical survey data to identify psychological vulnerabilities and utilizing AI-simulated phishing campaigns for targeted education, this study outlines a method to transition cyber security awareness from passive compliance to active behavioral change. Ultimately, this research demonstrates that mitigating modern cyber threats in non-technical domains requires a multidisciplinary approach that prioritizes human-centric training alongside technological safeguards.

**Keywords:** *cyber security, social engineering, Survey-Informed Behavioral Intervention (SIBI), identity tracking.*

## Introduction

Cyber security plays an indispensable role in safeguarding modern information infrastructure, as protecting sensitive data has become one of the most critical challenges of the digital age (Reddy & Reddy, 2014). In the context of social services, professionals routinely collect, store, and transmit highly confidential records concerning vulnerable populations, making their databases lucrative targets for malicious actors. Unfortunately, the human element remains a highly exploitable aspect of cyber security assurance, as human behavior frequently dictates the success or failure of security protocols (Evans et al., 2016). Recently, the proliferation of artificial intelligence has enabled attackers to automate and personalize social engineering tactics, such as spear-phishing and deepfake audio impersonations, which can easily deceive individuals who lack specialized technical training. Consequently, there is an urgent need to understand how these technologies impact social workers and to develop robust, human-centric defenses.

The primary problem addressed in this paper is the vulnerability of social workers to AI-driven social engineering attacks and the subsequent need for tailored, data-driven security interventions. The scope of this research encompasses the analysis of human behavior through survey data, the identification of specific psychological manipulation tactics employed by AI systems, and the design of an educational framework to build resilience among non-technical personnel. While technological firewalls and automated traffic analytics can prevent many technical intrusions, they cannot effectively stop an employee from willingly handing over credentials when manipulated by a highly convincing, AI-generated pretext (Miao et al., 2018). Therefore, securing the social work sector requires addressing the human user as the primary perimeter of defense.

Despite decades of research into information security, existing approaches to mitigating human-centric vulnerabilities remain highly insufficient for several reasons. First, much of the prevailing literature and industry practice relies on overly technocentric focuses, which neglect non-technical issues such as human psychology, targeted education, and localized policy (Dang et al., 2024). Second, traditional cyber security awareness campaigns frequently fail to achieve actual behavioral change because they rely heavily on "fear appeals" rather than providing actionable, comprehensible advice that motivates users (Bada et al., 2019). These generic training models do not account for the specific high-stress, empathy-driven nature of social work, leaving these professionals uniquely susceptible to emotional manipulation.

To address these critical gaps, this paper proposes a novel, sector-specific methodology for cyber security training. The primary contributions of this paper are as follows:

- We introduce a structured behavioral framework that utilizes psychometric survey data to baseline the specific social engineering vulnerabilities of social workers.
- We present a hypothetical evaluation plan integrating AI-simulated phishing attacks tailored to the empathy-driven workflows of social services, demonstrating how adaptive training can

sustainably alter security behaviors.

## Related Work

### Human Behavior and Cyber Security Awareness

The first category of related work focuses on the intersection of human behavior and information security assurance. Research demonstrates that despite the long-standing application of technical security practices in industry, numerous breaches continue to occur due to human error and manipulation (Evans et al., 2016). Furthermore, studies into cyber security awareness campaigns reveal that merely providing information about risks is insufficient; individuals must be psychologically motivated to change their attitudes and intentions (Bada et al., 2019). A major weakness of current behavioral models is their generalized nature, which fails to account for the unique occupational psychological profiles of specific groups, such as the inherent empathy and urgency found in social workers. The present work builds upon these foundational psychological insights but diverges by specifically correlating survey-based behavioral data with customized intervention strategies tailored to the social care sector.

### Non-Technical Challenges in Information Systems

A second distinct area of literature examines cyber security through a non-technical perspective, emphasizing organizational and human factors over algorithmic defenses. It has been established that non-technical issues—namely education, awareness, policy, standards, and human risk management—are equally as critical as technical solutions in fields like energy informatics and broader critical infrastructure (Dang et al., 2024). While this perspective provides a strong conceptual strength by broadening the scope of security, a significant limitation is the lack of empirical frameworks that operationalize these concepts for localized, non-technical workforces. Our research addresses this limitation by adapting these non-technical pillars into a structured, executable training pipeline that specifically empowers social workers to recognize and deflect social engineering.

### AI and Emerging Disruptive Cyber Threats

The third category of relevant literature addresses the macro-level impact of emerging technologies and potential "game-changers" in the cyber security landscape. The continuous emergence of new technologies changes the face of cyber security, introducing disruptive paradigms that outpace traditional defensive measures (Reddy & Reddy, 2014). Specifically, the integration of advanced mathematics, statistics, and artificial intelligence into both offensive and defensive cyber operations has been identified as a highly disruptive shift that will dictate research priorities for the foreseeable future (Meza et al., 2009)(Kott et al., 2015). While these studies excel at forecasting high-level strategic threats, they often overlook the micro-level tactical implications for non-technical end-users interacting with AI-generated deceptions. This paper

bridges that gap by directly addressing how AI acts as an offensive social engineering tool against human targets and proposing a localized, data-driven defense mechanism.

## Method/Approach

### Structured Framework and Pipeline

To effectively protect social workers against AI-enhanced social engineering, we propose the Survey-Informed Behavioral Intervention (SIBI) framework. This framework operates on the premise that effective training must be continuously adaptive and grounded in empirical data regarding the target demographic's psychological predispositions. The pipeline consists of three primary modules, designed to be implemented iteratively within a social service organization.

The SIBI pipeline is structured as follows:

1. **Survey Data Collection & Vulnerability Profiling:** Employees complete a comprehensive psychometric survey designed to measure their susceptibility to various social engineering triggers, such as authority, urgency, and empathy.
2. **AI-Simulated Attack Generation:** Based on the vulnerability profiles, an AI engine generates highly targeted, benign social engineering attacks (e.g., simulated phishing emails claiming a child welfare emergency) tailored to the specific psychological weaknesses of different employee cohorts.
3. **Psychologically Tailored Remediation:** Employees who fall victim to the simulated attacks are immediately directed to micro-training modules that deconstruct the specific emotional manipulation tactic used, avoiding fear-based messaging in favor of constructive empowerment.

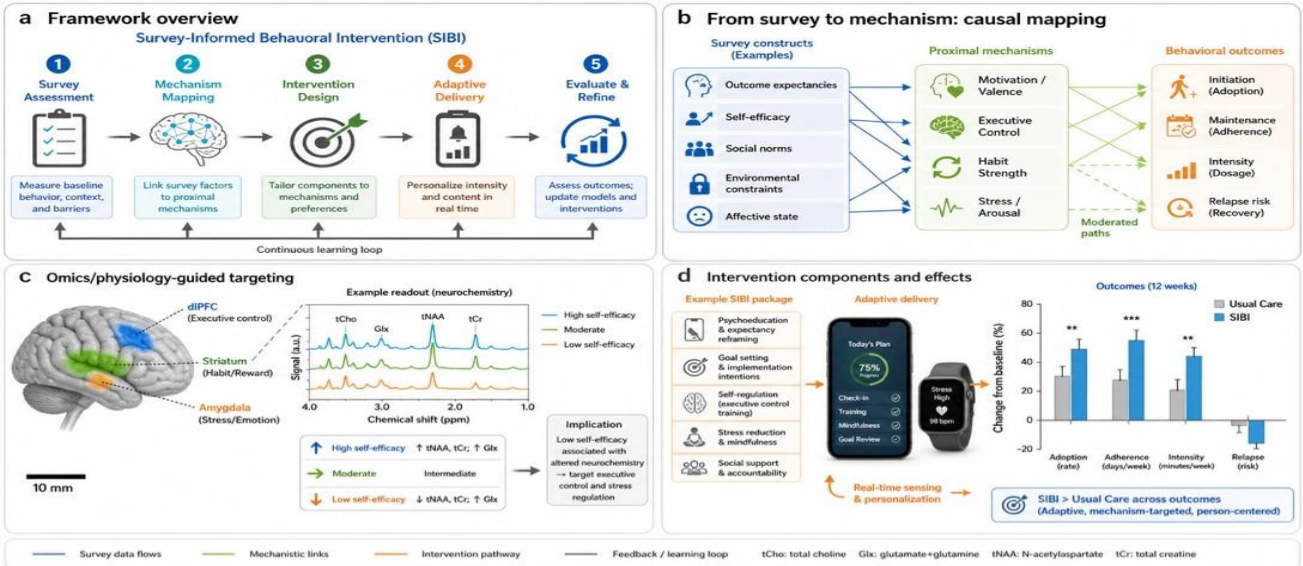


Fig:1 SIBI framework cybersecurity in social engineering Key Design Choices and Rationale

The design of the SIBI framework is heavily influenced by the necessity of moving away from generic, compliance-based security training. We utilize survey data as the foundational step because understanding how people perceive risks is an absolute prerequisite to creating effective, behavior-changing awareness campaigns (Bada et al., 2019). Social workers, by the nature of their profession, are trained to respond rapidly to crises and exhibit high levels of empathy, making them naturally susceptible to urgency-based and sympathy-based lures. By employing AI to generate the simulated attacks, the framework realistically mirrors the contemporary threat landscape where adversaries use machine learning to automate highly persuasive, customized spear-phishing campaigns. This ensures that the training environment accurately reflects the sophistication of modern threats, moving beyond easily identifiable, traditional phishing attempts (Kott et al., 2015).

**Evaluation Plan**

To validate the proposed SIBI framework, we outline a hypothetical evaluation plan utilizing a synthetic dataset of 500 social workers across various municipal agencies. The evaluation will follow a pre-test/post-test experimental design over a six-month period. Initially, all participants will undergo a baseline assessment featuring untargeted, generic phishing simulations to establish a baseline click-rate. Following this, the experimental group (n=250) will be subjected to the SIBI pipeline, receiving AI-generated, psychologically tailored simulations based on their survey data, while the control group (n=250) will receive standard, non-tailored cyber security videos. Success will be measured by comparing the reduction in the susceptibility rate (i.e., the percentage of users who compromise their credentials) between the two groups during a final, unannounced testing phase featuring novel AI-generated zero-day lures.

**Discussion**

**Practical Implications and Deployment**

The deployment of the SIBI framework carries significant practical implications for organizational security management in the public sector. Integrating this framework requires a shift in how administrative bodies allocate resources, moving funding away from purely technical defense mechanisms toward continuous human-centric education. Because social services often operate under strict budgetary constraints, the automated nature of the AI-simulated attacks ensures that the system can scale efficiently without requiring a large dedicated cyber security team. Furthermore, as data protection regulations become increasingly stringent, implementing empirically backed, adaptive training protocols can assist social work agencies in maintaining compliance while genuinely reducing their operational risk profile. Ultimately, fostering this level of awareness transforms employees from being the weakest link into an active, resilient layer of the organization's defense architecture (Dang et al., 2024).

**Limitations and Failure Modes**

Despite its potential, the proposed approach presents several notable limitations and potential failure modes.

- First, self-reported survey data is inherently subject to response bias, as employees may answer psychometric questions based on how they believe management wants them to respond rather than reflecting their true behavioral tendencies.
- Second, while the AI attack simulations can mimic current social engineering strategies, they may fail to capture entirely novel, unforeseen manipulation tactics, leaving workers vulnerable to zero-day behavioral exploits.
- Third, the social work sector is notorious for high staff turnover and severe occupational burnout, meaning that continuous security training could impose an unsustainable cognitive burden on an already over-stressed workforce.

**Ethical Considerations and Risks**

Deploying simulated social engineering attacks within a highly empathetic and stressful profession introduces substantial ethical considerations.

- Primarily, simulating emergencies—such as fabricating a scenario involving an at-risk client to test a social worker's security compliance—can cause undue psychological distress and erode trust between the staff and organizational leadership.
- Additionally, profiling the psychological vulnerabilities of employees via survey data creates a risk of misuse; if this data is not strictly anonymized, management could unethically use an employee's "security risk score" for punitive measures, performance degradation, or termination rather than educational support.

**Future Work**

To further refine the mitigation of AI-driven social engineering in non-technical sectors, several avenues of future research must be explored.

- First, researchers should conduct longitudinal empirical studies spanning several years to determine the long-term decay rate of cyber security awareness in high-stress professions, evaluating how frequently the SIBI interventions must be repeated to maintain efficacy.
- Second, future work should explore the crossover potential of this framework into other non-technical public sector domains, such as healthcare nursing and emergency dispatch, to determine if specific empathetic occupational profiles share universal security vulnerabilities.

**Conclusion**

The intersection of social work and cyber security represents a critical vulnerability in the modern

data ecosystem, largely due to the highly sensitive nature of the information involved and the empathetic disposition of the workforce. As artificial intelligence continues to lower the barrier for executing highly persuasive, automated social engineering attacks, traditional, technically rigid defense mechanisms and fear-based awareness campaigns are no longer sufficient to protect organizational data. This paper has proposed a comprehensive, survey-informed framework that specifically targets the psychological vulnerabilities of non-technical staff through tailored AI-simulated interventions.

By prioritizing human behavior and continuous, adaptive education, social service organizations can effectively counter the manipulative tactics employed by modern adversaries. While ethical implementation and resource management remain challenging considerations, the transition toward empirical, psychologically grounded security training is absolutely necessary. Ultimately, recognizing and addressing the human element as a core component of cyber security assurance will ensure that social workers can safely navigate the digital landscape while continuing to provide essential services to vulnerable populations.

**References**

Reddy, G. Nikhita, & Reddy, G. J. Ugander (2014). *A Study Of Cyber Security Challenges And Its Emerging Trends On Latest Technologies*. International Journal of Engineering and Technology - UK ISSN: 2049-3444, Volume 4 No.1 January 2014. https://arxiv.org/pdf/1402.1842v1

Evans, Mark, Maglaras, Leandros A., He, Ying, & Janicke, Helge (2016). *Human Behaviour as an aspect of Cyber Security Assurance*. https://arxiv.org/pdf/1601.03921v1

Miao, Yuantian, Ruan, Zichan, Pan, Lei, Wang, Yu, Zhang, Jun, & Xiang, Yang (2018). *Automated Big Traffic Analytics for Cyber Security*. https://arxiv.org/pdf/1804.09023v1

Dang, Duong, Vartiainen, Tero, & Mekkanen, Mike (2024). *Cyber Security in Energy Informatics: A Non-technical Perspective*. https://arxiv.org/pdf/2405.01867v1

Bada, Maria, Sasse, Angela M., & Nurse, Jason R. C. (2019). *Cyber Security Awareness Campaigns: Why do they fail to change behaviour?*. International Conference on Cyber Security for Sustainable Society, 2015. https://arxiv.org/pdf/1901.02672v1

Meza, Juan, Campbell, Scott, & Bailey, David (2009). *Mathematical and Statistical Opportunities in Cyber Security*. https://arxiv.org/pdf/0904.1616v1

Kott, Alexander, Swami, Ananthram, & McDaniel, Patrick (2015). *Six Potential Game-Changers in Cyber Security: Towards Priorities in Cyber Science and Engineering*. https://arxiv.org/pdf/1511.00509v1