OPEN ACCESS

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

AI-DRIVEN STATISTICAL ANALYSIS FOR CRIMINOLOGICAL CYBERSECURITY: MITIGATING WORKPLACE HARASSMENT AGAINST SINGLE WOMEN IN ENTERPRISE NETWORKS

Dr Anum Ali[1], Zaheema Iqbal[2], Dr. Muhammad Faisal Majeed[3], Mahtab Jamil Akhtar[4]

[1]*Department of Computer Science, Lahore Leads University, Pakistan*
[2]*National Defence University, Islamabad, Pakistan*
[3, 4]*Department of Oric, Lahore Leads University, Pakistan*

[*]***Corresponding Author:*** (dranumali.cs@leads.edu.pk)
DOI: (*https://doi.org/10.71146/kjmr983*)

Article Info

Abstract

The intersection of criminology, social dynamics, and enterprise cybersecurity represents a critical frontier in modern organizational management. As the digital workplace expands, malicious insider activities have evolved beyond traditional data theft to encompass interpersonal cyber stalking and workplace harassment, disproportionately affecting vulnerable demographics such as single women. This paper proposes a novel, interdisciplinary framework that leverages artificial intelligence and statistical analysis to detect and mitigate these highly specific behavioral anomalies within enterprise networks. By integrating advanced feature selection techniques, neurosymbolic artificial intelligence, and interpretable machine learning models, the proposed system translates criminological indicators of harassment into measurable network traffic anomalies. Furthermore, we discuss the practical implications of deploying such systems in small and medium-sized enterprises, alongside the ethical complexities of monitoring employee behavior. Ultimately, this research bridges the gap between technical intrusion detection systems and human-centric criminological analysis, offering a foundational blueprint for protecting targeted individuals in hyper-connected organizational environments.

Keywords:
Cyber security, criminology, women, work place.

Introduction

Workplace harassment represents a pervasive criminological issue that has increasingly migrated into the digital domain, creating complex challenges for modern enterprise environments. Single women, in particular, often face elevated risks of targeted cyber stalking, unauthorized surveillance, and digital intimidation from malicious insiders. In contemporary hyper-connected ecosystems, these perpetrators leverage corporate networks, Internet of Things (IoT) devices, and internal communication platforms to conduct their harassment. Consequently, the digital footprint of workplace harassment can be conceptualized as an insider threat, requiring robust cybersecurity solutions capable of identifying subtle, human-driven anomalies. While enterprise networks are crucial for organizational productivity, their complexity inadvertently provides a fertile ground for stealthy, malicious interpersonal behaviors that traditional security infrastructures are ill-equipped to handle.

The primary scope of this paper is to define digital workplace harassment as a distinct cybersecurity anomaly and to propose an AI-driven statistical analysis framework for its detection. We specifically focus on the criminological profiling of harassment against single women, translating psychological and behavioral stalking indicators into network traffic anomalies. The problem relies on the premise that harassers exhibit detectable digital patterns—such as repeated, unauthorized access to a specific colleague's digital workstation, unusual frequency of communications, or the anomalous querying of location-based IoT sensors. However, existing approaches in the cybersecurity domain are highly insufficient for addressing this specific sociological threat for several reasons. First, traditional intrusion detection systems focus primarily on external, financially motivated malware or large-scale data breaches, thereby completely ignoring the internal behavioral anomalies linked to localized interpersonal harassment (Kumar & Islam, 2024). Second, current threat detection mechanisms often lack the necessary interpretability required by Human Resources (HR) and criminological investigators, making it exceedingly difficult to prove malicious intent in a legal or corporate disciplinary context (Kumar & Islam, 2024).

To address these critical gaps, this paper introduces a specialized analytical approach combining behavioral criminology with advanced network security. The core contributions of this work are as follows:

- We propose a novel, interpretable artificial intelligence framework tailored specifically for the detection of internal cyber stalking and digital harassment patterns within enterprise networks.

- We integrate criminological behavioral indicators with advanced network feature selection and neurosymbolic AI, thereby bridging the operational gap between abstract human malice and quantifiable network anomalies.

Related Work

Enterprise Network Threat Detection and Interpretability

The first category of related literature focuses on the detection of threats within complex enterprise and industrial networks using advanced machine learning. Modern enterprise systems frequently face data scarcity and high false-positive rates when utilizing deep learning for threat detection, particularly where privacy concerns restrict data collection (Kumar & Islam, 2024). To combat this, researchers have successfully developed interpretable cyber threat detection systems utilizing bidirectional gated recurrent units, attention mechanisms, and Shapley additive explanations (Kumar & Islam, 2024). While these systems demonstrate immense strength in providing understandable outputs for security analysts, their primary weakness lies in their almost exclusive orientation toward equipment sabotage or corporate espionage. Compared to our work, these

existing interpretable models lack a criminological dimension; our research actively adapts these explainable AI techniques to specifically identify the behavioral signatures of human-to-human harassment.

AI and Adaptive Security Frameworks

The second critical area of research involves the use of adaptive security frameworks and transfer learning to combat evolving cyber threats. For instance, in IoT-enabled sectors and 5G networks, autonomous adaptive security frameworks utilize closed feedback loops and advanced analytics to dynamically respond to system vulnerabilities (Abie & Pirbhulal, 2024). Similarly, the use of Neurosymbolic AI and transfer learning has been shown to drastically improve network intrusion detection by combining the data-processing power of neural networks with the logical reasoning of symbolic AI (Tran et al., 2025). Conversely, adversaries are also utilizing reinforcement learning to create stealthy, AI-powered malware that dynamically adapts its behavior to evade detection (Assen et al., 2023). The strength of these adaptive systems is their dynamic responsiveness, yet they predominantly target automated malware or ransom ware (Assen et al., 2023). Our approach distinguishes itself by leveraging the logical rule-based strengths of Neurosymbolic AI (Tran et al., 2025) to encode criminological definitions of stalking, thereby tracking human adversaries who continuously adapt their harassment tactics.

Human-Centric Cybersecurity and Organizational Adoption

The final category encompasses literature addressing the human and social aspects of cybersecurity, particularly regarding how different demographics and organizational sizes interact with security protocols. Studies examining parents' perceptions of children's cybersecurity highlight the critical need for solutions that address the specific vulnerabilities of distinct user groups based on social dynamics (Quayyum et al., 2021). Furthermore, research on Small and Medium-sized Enterprises (SMEs) demonstrates that developers of cybersecurity solutions often fail to understand the specific contextual requirements of smaller organizations, leading to poor adoption rates (Shojaifar et al., 2020). The strength of this sociological research is its emphasis on user requirements and human vulnerability (Quayyum et al., 2021) (Shojaifar et al., 2020). However, it generally neglects the internal victimization of employees, such as single women facing persistent harassment. This work builds upon the human-centric security paradigm by explicitly designing a detection framework that meets the operational constraints of SMEs (Shojaifar et al., 2020) while addressing the specific criminological threat of workplace gender-based harassment.

Method/Approach

The Harass Detect-AI Framework

To effectively map criminological indicators of single women's workplace harassment to observable network traffic, we propose a structured, multi-modular framework termed "Harass Detect-AI." The first module of this framework is dedicated to behavioral feature extraction and optimization. The use of machine learning models requires high-quality, non-redundant data; therefore, we apply rigorous feature selection methods such as Information Gain, the Chi-Squared Test, and Recursive Feature Elimination (Silva et al., 2024). In the context of harassment, these selected features are not standard malware signatures, but rather user-interaction metrics. This includes the frequency of internal directory searches, timestamp anomalies indicating after-hours surveillance of a target's workstation, and unauthorized attempts to interface with localized IoT devices near the victim. By identifying the most impactful features, we can increase the computational efficiency of the anomaly detection process while preserving accurate generalization (Silva et al., 2024).

The second module focuses on analytical processing using Neurosymbolic Artificial Intelligence (NSAI) (Tran et al., 2025). We implement an NSAI architecture that fuses neural network pattern recognition with symbolic, rule-based logic derived from criminology. The neural component monitors the optimized network traffic streams for statistical deviations from the baseline behaviors of the suspected harasser. Concurrently, the symbolic component applies rigid logical constraints based on legal and criminological definitions of workplace harassment—such as the requirement for repeated, unwanted contact over a specific temporal threshold. A critical design choice in this module is the integration of Shapley additive explanations and decision tree techniques to ensure high interpretability (Kumar & Islam, 2024). This rationale is essential because the output must be legally and administratively actionable; HR professionals must be able to understand precisely which digital behaviors triggered the harassment alert without requiring deep technical expertise.

Hypothetical Evaluation Plan

Because real-world datasets depicting the exact network traces of workplace harassment against single women do not exist—largely due to strict privacy laws and the anonymization protocols that currently strip away valuable behavioral ground truth (Laprade et al., 2020)—we propose a comprehensive hypothetical evaluation plan. We plan to utilize a generative adversarial network to create synthetic, high-fidelity enterprise industrial network data (Kumar & Islam, 2024), simulating a standard SME environment with 500 active users. Into this synthetic baseline, we will inject parameterized "harassment traffic," modeled after documented criminological case studies of cyber stalking.

1. **Baseline Generation: ** Simulate normal collaborative network traffic, including localized IoT interactions and standard file-sharing activities.

2. **Anomaly Injection: ** Introduce specific vectors of digital harassment, such as stealthy, repeated querying of a simulated single female employee's endpoint and access logs.

3. **Performance Benchmarking: ** Deploy the Harass Detect-AI framework against traditional deep learning intrusion detection models.

4. **Metric Analysis: ** We will evaluate the models based on Precision, Recall, F1-score, and computational training time (Silva et al., 2024). We hypothesize that our NSAI approach will yield a significantly higher Precision rate compared to standard models, as the symbolic logic will filter out false positives caused by legitimate, non-malicious collaboration.

Discussion

Practical Implications and Deployment

Deploying Harass Detect-AI carries significant practical implications, particularly for Small and Medium-sized Enterprises. SMEs represent a massive portion of the economy but frequently lack the specialized HR and IT resources to systematically manage cybersecurity and internal behavioral threats (Shojaifar et al., 2020). By utilizing an interpretable, neurosymbolic approach and our proposed framework can be integrated as an overlay onto existing enterprise network infrastructure without requiring a dedicated team of cybersecurity analysts. The clear, logical outputs generated by the system provide organizational leadership with empirical, statistical evidence of digital harassment. This capability is paramount for protecting vulnerable demographics, ensuring that claims of workplace harassment are supported by immutable network data rather than relying solely on subjective, human testimony.

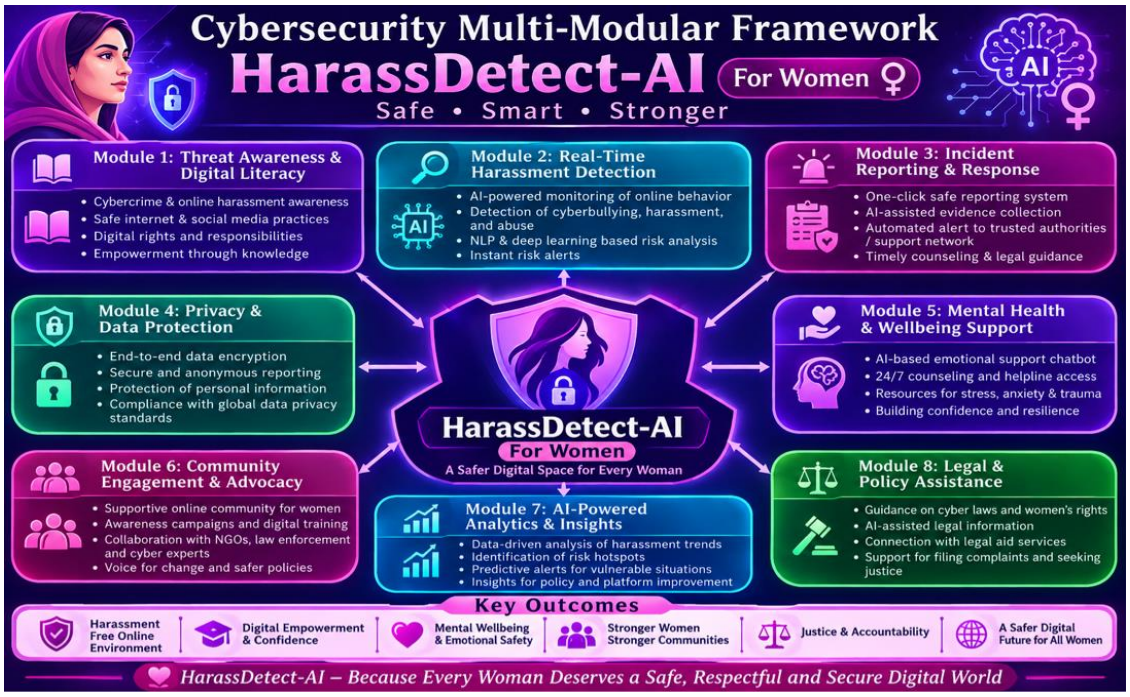

Fig 1: Cybersecurity multi-modular framework termed Harass Detect-AI

Limitations and Failure Modes

Despite its robust design, the proposed framework faces several critical limitations and potential failure modes.

- First, there is a substantial risk of false positives when legitimate, intensive professional collaboration mathematically mimics the network footprint of digital stalking. For example, two employees working closely on a high-pressure project may generate traffic frequency anomalies that the system misclassifies as harassment.

- Second, technologically sophisticated harassers may deploy evasion tactics similar to those used by modern malware. By utilizing AI-powered techniques to dynamically adapt their encryption or by utilizing out-of-band, decentralized communication channels (Assen et al., 2023); harassers could easily bypass internal network monitoring.

- Third, the system's efficacy relies heavily on the quality of synthetic training data. As noted, current cybersecurity datasets often lack realistic ground truth or are highly anonymized (Laprade et al., 2020), making it exceptionally difficult to train models that perfectly reflect the nuanced reality of human-to-human harassment.

Ethical Considerations and Risks

The deployment of an AI-driven behavioral monitoring system inherently introduces severe ethical considerations.

- The most prominent risk is the potential violation of employee privacy. Continuous, granular monitoring of internal network queries and IoT interactions could lead to an invasive surveillance culture, damaging employee morale and violating regional data protection regulations.

- Furthermore, there is a pronounced risk of algorithmic bias. If the logical rules or the synthetic data inadvertently profile specific cultural communication styles or neurodivergent working patterns as "anomalous"

or "harassing," the system could generate discriminatory outcomes, leading to wrongful accusations and severe reputational damage.

Future Work

Future research must expand the technological and sociological boundaries of this framework.

- One critical pathway is the integration of hyper dimensional Computing (HDC). Given the massive scale of enterprise network data, HDC has shown significant promise in efficiently processing high-dimensional data to identify complex unknown attack patterns (Ghajari et al., 2025). Applying HDC could drastically reduce the computational overhead of monitoring thousands of internal communication nodes for micro-anomalies related to harassment.

- Additionally, exploring Quantum Machine Learning (QML) presents a frontier for securing the integrity of the detection system itself. As quantum computational power becomes more accessible, integrating QML could secure the IDS against tampering by highly privileged insider threats—such as rogue IT administrators who may be the perpetrators of the harassment (Abreu et al., 2024).

Conclusion

The persistence of digital workplace harassment against single women demands a paradigm shift in how we conceptualize enterprise network security. This paper has outlined an interdisciplinary approach that reframes targeted interpersonal cyber stalking as a critical internal network anomaly. By harnessing the logical reasoning of neurosymbolic artificial intelligence and the transparency of interpretable machine learning models, the proposed Harass Detect-AI framework provides a structured methodology for identifying malicious human intent hidden within vast streams of corporate network traffic. Moving away from a strictly malware-centric view of cybersecurity allows organizations to leverage statistical analysis for profound social protection.

Ultimately, addressing the criminological realities of the modern workplace requires a synthesis of sociological understanding and advanced technical infrastructure. While limitations regarding data scarcity, potential false positives, and ethical privacy concerns remain, the utilization of AI for behavioral network analysis offers a highly promising protective mechanism. Future interdisciplinary research must continue to refine these algorithms, ensuring that as enterprise networks grow in complexity, the safety and digital well-being of vulnerable employee demographics remain a central, automated priority.

References

Kumar, Prabhat, & Islam, A. K. M. Najmul (2024). Interpretable Cyber Threat Detection for Enterprise Industrial Networks: A Computational Design Science Approach. https://arxiv.org/pdf/2409.03798v1

Abie, Habtamu, & Pirbhulal, Sandeep (2024). Autonomous Adaptive Security Framework for 5G-Enabled IoT. https://arxiv.org/pdf/2406.03186v1

Tran, Huynh T. T., Sander, Jacob, Cohen, Achraf, Jalaian, Brian, & Bastian, Nathaniel D. (2025). Neurosymbolic AI Transfer Learning Improves Network Intrusion Detection. https://arxiv.org/pdf/2509.10850v1

Assen, Jan von der, Celdrán, Alberto Huertas, Luechinger, Janik, Sánchez, Pedro Miguel Sánchez, Bovet, Gérôme, Pérez, Gregorio Martínez, & Stiller, Burkhard (2023). RansomAI: AI-powered Ransom ware for Stealthy Encryption. https://arxiv.org/pdf/2306.15559v1

Quayyum, Farzana, Bueie, Jonas, Cruzes, Daniela S., Jaccheri, Letizia, & Vidal, Juan Carlos Torrado (2021). Understanding parents' perceptions of children's cybersecurity awareness in Norway. https://arxiv.org/pdf/2108.02512v1

Shojaifar, Alireza, Fricker, Samuel A., & Gwerder, Martin (2020). Elicitation of SME Requirements for Cybersecurity Solutions by Studying Adherence to Recommendations. https://arxiv.org/pdf/2007.08177v1

Silva, Miguel, Vitorino, João, Maia, Eva, & Praça, Isabel (2024). Efficient Network Traffic Feature Sets for IoT Intrusion Detection. https://arxiv.org/pdf/2406.08042v1

Laprade, Craig, Bowman, Benjamin, & Huang, H. Howie (2020). PicoDomain: A Compact High-Fidelity Cybersecurity Dataset. https://arxiv.org/pdf/2008.09192v1

Ghajari, Ghazal, Ghimire, Ashutosh, Ghajari, Elaheh, & Amsaad, Fathi (2025). Network Anomaly Detection for IoT Using Hyper dimensional Computing on NSL-KDD. https://doi.org/10.1109/SATC65530.2025.11136944

Abreu, Diego, Rothenberg, Christian Esteve, & Abelem, Antonio (2024). QML-IDS: Quantum Machine Learning Intrusion Detection System. https://doi.org/10.1109/ISCC61673.2024.10733655