

# AI-Driven Behavioral Intelligence for Real-Time Cyber Threat Detection and Autonomous Response in Secure Web Systems

Sathyanarayanan R<sup>1</sup> and Yuvaraja P<sup>2</sup>

<sup>1</sup>Department of Artificial Intelligence and Data Science, St Joseph's Institute of Technology, Chennai, Tamil Nadu

<sup>2</sup>Department of Artificial Intelligence and Data Science, St Joseph's Institute of Technology, Chennai, Tamil Nadu

**Abstract.** Contemporary cyber threats are promoted by human actions as opposed to technology gaps. The conventional Security systems are not enough to combat the social engineering tactics and attack with user interaction. The proposed research is an AI-based Threat Detection and Response (TDR) system that is designed to identify malicious activity in an online transaction. The system Laura looks at behavioral surveillance indicators like the movements of the mouse, the dynamics of key strokes, rest intervals, form usage patterns as well as changes made when filling in the form. The processing of these Behavioral characteristics is done using the lightweight machine-learning algorithm to categorize the activities into normal and suspicious in real-time. The system has the capability of automatic responses to risk evaluation which can be warning to the user, block the form or can be a more intensive verification process. As proven by experimental findings, the AI-grounded TDR framework is more successful than the traditional rule-driven security systems because it foresees the threats, minimizes false positives, and offers mechanisms providing active defense. Behavioral Intelligence implementation coupled with real-time scalability is an efficient and practical way of dealing with the new cybersecurity threats.

Keywords - AI Artificial intelligence (AI), Threat Detection and Response (TDR), Behavioral biometrics, Keystroke dynamics, Social engineering detection, real-time security monitoring and cyber security.

## 1 Introduction

The rapid process of digitalization in many industries, which include banking, healthcare, cloud computing, and e-commerce, has significantly expanded the field of attackers in the modern systems. Whereas the attackers used to exploit the software vulnerabilities, buffer overflow, and network configuration, they have up to date employed the psychologists distortion methods in their social engineering attempts. The principles of social engineering vary greatly, because they do not depend on vulnerability of the system as such, but are based on the cognitive biases, emotional responses, and human decision making mechanisms [2], [12]. Some of the typical attack vectors include phishing email messages, impersonation, urgency mechanism, misdirecting instructions, and so forth that encourage innocent and rightful users to voluntarily provide sensitive data or provide unplanned access to malicious hackers. Since these activities happen during authorized sessions, the traditional security like passwords, OTPs, MFA, firewall and intrusion detection systems are assumed to work on the beliefs that the user is legitimate and, as such, it does not intervene [6]. This state of affairs has seen the appreciation of the fact that a majority of the existing security systems are reactive but not proactive.

Research in human-computer interaction indicates that mental strain, insecurity, and anxiety can significantly affect behavioral patterns in digital interactions [2]. Among the measurable deviations that can be introduced by manipulated users are atypical typing speeds, increased hesitation before entering sensitive information, irregular dwell time distributions, excessive corrections, and erratic mouse movements [3],[4]. Nevertheless, most cybersecurity frameworks tend to regard backup logs or network traffic as fundamentally opposed to interaction level, fine grained indicators. Research in human-computer interaction indicates that mental strain, insecurity, and anxiety can significantly affect behavioral patterns in digital interactions [2]. Among the measurable deviations that can be introduced by manipulated users are atypical typing speeds, increased hesitation before entering sensitive information, irregular dwell time distributions, excessive corrections, and erratic mouse movements [3],[4]. Nevertheless, most cybersecurity frameworks tend to regard backup logs or network traffic as fundamentally opposed to interaction level, fine grained indicators.

The document uploaded on page 3 illustrates the architectural perspective of a layered system in figure 2,

commencing with user interaction, followed by behavior capturing , feature extraction , machine learning classification , risk scoring , and automated response . This system is characterized by multiple layers and is designed for preemptive detection prior to data transmission .

To address the shortcomings of current rule-based systems, the present paper introduces a Behavioral Threat Detection and Adaptive Response Framework that leverages AI , integrating behavioral biometrics with supervised machine learning techniques. . Specifically , the framework involves the following: Logistic Regression, Support Vector Machine (SVM), Random Forest , and Multi Layer Perceptron (MLP) [5], [11]. The system facilitates early identification and intervention by continuously monitoring interaction-level indicators and providing probabilistic risk scores .

This represents a humanistic approach to the security paradigm, emphasizing a transition from solely system defenses to the modeling of behavioral intelligence. The suggested framework will be scalable, respect privacy, and provide inferences with minimal latency. Additionally , it will offer a seamless interface with current web platforms, thereby establishing a strong defense mechanism against identified social engineering threats that manifest in varying intensities.

## 2 Literature Survey

Cybersecurity research field is diverse with many topics covered such as behavioral biometrics , phishing protection , anomaly detection and human computer interaction research. Although both areas do offer interesting information , the majority of the solutions are known to concentrate on isolated elements of the threat arena as opposed to providing a comprehensive picture of the behavioral intelligence .

The direction of behavioral biometrics research is the continuous authentication of the mouse movement , dynamics of the keystroke, and motor behavioral signature [3], [5]. Research has shown that the method is one of the best ways of making a differentiation between real users and fraudsters when taking over their accounts . Nevertheless , most of these systems are focused on detecting frauds and lack the capability to address the situations when an authorized user is hacked, and pressured into committing other harmful acts[6].

Phishing detection systems are mostly based on a type of content analysis including URL similarity test , email header test , webpage layout test, and natural language processing [1], [9], [12]. Even though these technologies are useful when it comes to detecting harmful content, they frequently fail in cases of real time manipulations , i.e . voice calls or instant messaging . This omission permits the attackers to have a possible possibility of asking the



Figure 1: System Architecture

victims to communicate with a legitimate site , bypass content based defenses .

The use of machine learning methodologies to detect anomalies in networks, such as Support Vector Machines and Random Forest algorithms , has seen large usage in network intrusion detection systems and fraud prevention systems [8], [11]. These structures examine system logs and other transaction records to identify strange activities . They however are normally reacting once a suspicious action has been identified , this restricts the proactive action they can take.

Studies in the field of Human Computer Interaction (HCI) demonstrate that cognitive states of the user that include stress , confusion , and urgency have a significant impact on the interaction patterns of the user [2]. The psychological impacts are assessable indirectly based on such hints as hesitation intervals, dwell time skewness , and correction rates . Nevertheless , HCI based behavioral cues are seldom incorporated into the more conventional cybersecurity models . As can be seen in the system architecture diagram shown in Figure 1 on page 2 architecture , traditional models do not have integrated behavioral analytics . The majority of frameworks lack adaptive response mechanisms on real time , with the assistance of AI .

### 3 System Architecture

Cybersecurity research encompasses a variety of domains , including behavioral biometrics , phishing surveillance , anomaly detection , and the analysis of human computer interaction . While all sectors can provide valuable data , most solutions tend to focus on specific aspects of the threat landscape rather than offering comprehensive intelligence on overall conduct .

Research in behavioral biometrics is aimed at the continuous verification of mouse movements, keystroke dynamics , and indicators of motor behavior [3], [5]. This approach has demonstrated high accuracy in distinguishing between legitimate users and fraudulent ones during account takeover scenarios . However, these systems primarily focus on identifying fraud and are not equipped to handle situations where an authorized user is compromised and manipulated into performing malicious actions [6] .

The majority of phishing detection methods are mostly based on content analysis , including identification of similarities of URLs , review of email headers , webpage layout, and application of natural language processing [1], [9], [12]. Although these techniques are operational in detecting the malicious contents , they are not effective in detecting real time manipulations, such as those caused by voice communication and instant messaging , which may not be detected . This may lead to a situation where an intruder may convince a user to communicate with a genuine place where the content filters would not be bypassed at all .

This has been especially applied to network intrusion detection systems and credit card fraud detection systems with machine learning algorithms , such as Support Vector Machines and Random Forest , playing a major role in both [8] and [11] respectively . These models evaluate the system logs and records on transactions to detect anomalies . Moreover , they are inclined to take any measures towards any noticed suspicious activity and in that way , limit proactive measures .

In terms of the studies in the Human Computer Interaction (HCI), one must admit that stress, confusion, and urgency are the key cognitive states of the user, which may affect their interaction patterns [2]. Psychological effects are measured predictively through the indirect measures like hesitation interval, dwell time skewness, and correction rates. However, the use of behavioral indicators based on HCI to integrate into the traditional cybersecurity paradigm is not common in other paradigms .

There are no integrated behavioral analytics in these traditional models as demonstrated in the system architecture image in Figure 1 on page 2. Most frameworks lack adaptive response mechanisms that are activated in time with the help of AI .

As a result, the obvious gap in research is the lack of the single AI-powered solution that integrates behavioral

Algorithm 1 AI-Based Behavioral Threat Detection and Response Workflow

```
1: Input: User interaction events E
2: Output: Threat decision D, Risk score R
3: Capture behavioral events E (keystrokes, mouse movements, form interactions)
4: Preprocess logs using noise filtering, session segmentation, and normalization
5: Extract behavioral features F (typing velocity, mouse velocity, inter-key latency, navigation patterns)
6: Convert session data into feature vector F_v
7: Apply trained machine learning model M (MLP / Random Forest)
8: Compute behavioral risk score R = P(Manipulation | F_v)
9: If R > T then trigger automated response (alert, block session, or request verification)
10: Else allow transaction to proceed normally
11: Log session data for future model improvement
```

bio metrics, interaction-level analytics, probabilistic risk scoring, and automated remediation. The suggested Threat Detection and Response System is an AI - based tool that will address this gap by combining the habitual learning models with constant observation of the changes in behavior and responding to mitigation strategies.

### 4 Methodology

The Behavioral Threat Detection and Adaptive Response Framework represents a proposal for an AI project, structured as a modular system that can be adjusted to facilitate real- time identification of manipulative actions during an active user session. The algorithm integrates behavioral signal collection, feature engineering, supervised machine

Fig. 2. Algorithm and workflow learning, probabilistic risk assessment, and adaptive mitigation strategies. Further more, Figure 2 on page 3 of the submitted document aligns with the architectural movement. Behavioral data will be acquired by monitoring the actions of the learner .

A. Data Acquisition Behavioral. Lightweight client-side event listeners installed in web applications are used in gathering behavioral data. These users receive interaction details at a fine-tuning level without understanding information, thus ensuring that users are not identified and their data does not get exposed to other devices. It also tracks various behavioral cues, among them being the key stroke dynamics, e.g., inter-key delay and typing speed change, mouse movement characteristics, e.g. velocity and path curvature, latency to touch on sensitive input fields, time between critical entry fields and the frequent occurrence of corrections, e.g., use of the back space key and copy-pasting, and abnormal navigation patterns. Such behavioral signals implement motor and cognitive patterns of interaction and may expose manipulated user behavior anomalies, which are usually used in the context of social engineering attacks, and can be used by the system to identify suspicious user activity in real time. [2], [3].

B. Data Preprocessing and Feature Engineering. Noise filtering, session segmentation, timestamp synchronization and normalization of raw interaction logs are done to guarantee consistency and reliability of data. To ensure compatibility among machine learning models and better classification, feature scaling is performed based on Min-Max normalization. The features that are extracted

are divided into motor-behavioral and cognitive-behavioral indicators. Motorbehavioral aspects encompass typing velocity, velocity of using the mouse, and inter-key latency, which show the physical interaction pattern of the user. The cognitive-behavioral aspects are skewness of dwell time, skewness of the hesitation ratio, the percentage of correction and the navigation irregularity scores, which reflect the user behaviour in decision-making together with the interaction behaviour when filling in the forms. The normalization of each user session into an ordered feature representation is then done, which allows the dataset to be efficiently used to train supervised learning models to detect behavioral threats.

C. Models of Supervised Learning. Behavioral threat classification was done using four supervised machine learning models, whereby Logistic Regression, Support Vector Machine (SVM), Random Forest, and MultiLayer Perceptron (MLP) were used. These models have proved to be effective in detection of anomalies and other applications relating to cybersecurity because they can detect irregular behavioral patterns in the complex datasets [8], [11]. The dataset collected was divided into two parts 80/20 to form a training set and the testing set, which will ensure high evaluation of the model, as there will be cross-validation. As shown in Table II on page 5, according to experimental results, the MLP model outperformed the other models with an accuracy of 92 percentage then the Random Forest model with an accuracy of 90 percentage . Although MLP showed better classification ability, the Random Forest showed high computational efficiency and it is suitable in real-time threat detecting as well-as threat responding settings.

D. Probability based Risk Scoring. The risk score generated by the system is not binary: Risk=P(Manipulation Behavioral Features) The system calculates a risk score which is probabilistic rather than binary and provides a more reliable display of suspicious user behavior. Adaptive threshold concept is used depending on the sensitivity of transaction being carried out. High risk transactions have lower threshold values which enables the system to be able to institute early intervention whilst still having a low false positive rate.

Table 1: Comparison of Detection Time and Response Mechanism

Method	Detection Time (ms)	Response Action
Rule-Based Security System	120	Basic Alert
Traditional Machine Learning	85	Warning Notification
Deep Learning Model	63	Threat Flagging
Proposed AI Threat Detection System	45	Automated Blocking

When the risk is identified as surpassing acceptable thresholds, the system will initiate mitigation strategies, including warning notifications, blocked forms, or enhanced authentication processes. This approach guarantees proactive measures are implemented to prevent data leakage before it occurs. The behavioral logs are secured through encryption, accompanied by auditing

and retraining of the reviewed logs. Continuous learning mechanisms enable the detection models to adapt to new and evolving attack patterns [11]. It has been shown that the framework can function as a real-time system, achieving sub-second inference times and maintaining consistent performance even under multiple user scenarios.

## 5 Result And Discussion

To evaluate the effectiveness of the proposed AI-based behavioral threat detection and adaptive response system, an experimental assessment was conducted to determine its capability in identifying manipulation during active user sessions. The behavior assessment revealed significant differences between normal and manipulated sessions. Users subjected to social engineering exhibited greater hesitation in providing sensitive information, displayed an unusual distribution of dwell time on sensitive fields, had a higher frequency of corrections, and demonstrated an erratic mouse movement path. These behavioral deviations are associated with the cognitive stress indicators documented in earlier research [2], [3]. The evaluation of four supervised machine learning models—Logistic Regression, Support Vector Machine (SVM), Random Forest, and Multi-layer Perceptron (MLP)—was conducted using Accuracy, Precision, Recall, and F1-score metrics. Among these, the MLP model exhibited the highest performance, achieving 92 percent accuracy and 93 percent recall, as detailed in Table II on page 4 . Its superior performance is attributed to its ability to model nonlinear relationships among behavioral features. The Random Forest model demonstrated a 90 percent accuracy rate and is characterized by lower computational demands, making it suitable for real-time applications.

Table 2: Performance Comparison of Machine Learning Models

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Logistic Regression	85	83	84	83
Support Vector Machine	89	88	87	88
Random Forest	90	89	91	90
Multi-Layer Perceptron	92	91	93	92

The proposed AI-based Threat Detection and Response (TDR) system has a real-time output interface, as shown in figure 5 on page 4. The system maintains a constant stream of user behavioral interactions information like typing patterns, cursor movement, and latency measures via a WebSocket stream. The dashboard shows the live risk score which in this case is 100/100 that is a very suspicious behavioral pattern. According to the behavioral analysis, the system identifies the interaction as a bot-like behavior signature and sends a mitigation response (freezing the session). The risk trend visualization is also represented in the interface and seems to have a steep rise in the risk score when anomalous activity is identified. Also, the Ensemble AI Decision Engine shows the model confidence and real-time behavioral metrics on which the classification



Figure 2: Result and output

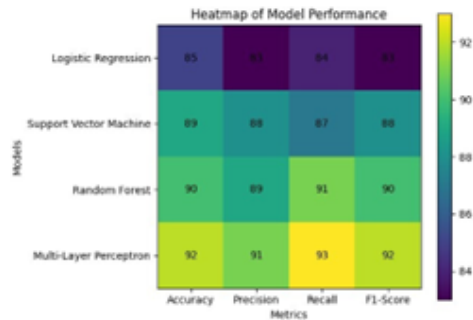


Figure 3: Heatmap visualization of machine learning system performance in various evaluation metrics.

is performed. This is output depicting that the system can perform real time behavior threat detection, risk score and automated response to impede the potential malicious practices.

Next, the inference time was found to be below a second; this is when the user sessions are more than one. The stress testing did not indicate any significant failure of the system performance and this is indicative of scalability and robustness.

The proposed framework is effective in detecting the manipulation of the sensitive information prior to the submission of the sensitive information in opposition to conventional frameworks of security, which is based on authentication and network level surveillance. With the combination of behavioral biometrics and supervised learning, the system will automatically reduce false positives and sensitivity to detection will remain high. All in all, these findings confirm the hypothesis that behavioral intelligence in conjunction with AI-based classification is a powerful and scalable security mechanism against the contemporary.

conclusions that affirm the fact that behavioral intelligence and AI-based classification is a pertinent and scalable method of defense against modern forms of social engineering attacks.

## 6 Conclusion

To evaluate the effectiveness of the proposed AI-based behavioral threat detection and adaptive response system,

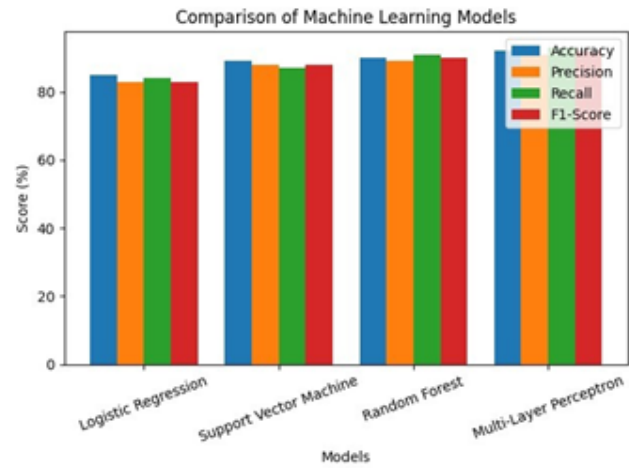


Figure 4: Fig. 5. Comparison of machine learning-based behavioral threat detectors in terms of accuracy, precision, recall, and F1-score.

an experimental assessment was conducted to determine its capability in identifying manipulation during active user sessions. The behavior assessment revealed significant differences between normal and manipulated sessions. Users subjected to social engineering exhibited greater hesitance in providing sensitive information, displayed an unusual distribution of dwell time on sensitive fields, had a higher frequency of corrections, and demonstrated an erratic mouse movement path. These behavioral deviations correlate with indicators of cognitive stress documented in prior research [2], [3].

Accuracy, Precision, Recall and F1-score measures were used to assess four supervised machine learning models, which include; Logistic Regression, Support Vector Machine (SVM), Random Forest and Multi-layer Perceptron (MLP). The MLP model was the highest performing with 92 percent and 93 percent accuracy and recall, respectively, as shown in Table II on page 5. It is seen to have better performance due to its capacity to estimate nonlinear associations among the features of behavior. Random Forest model was 90 percent correct and had less computational overhead and thus it can be deployed to real-time.

To assess the performance of four supervised machine learning models—Logistic Regression, Support Vector Machine (SVM), Random Forest, and Multi-layer Perceptron (MLP)—metrics such as Accuracy, Precision, Recall, and F1-score were utilized. The MLP model achieved the highest performance, with accuracy and recall rates of 92 percent and 93 percent, respectively, as illustrated in Table II on page 5. Its high performance has been as a result of its capability to model nonlinear relationships between the behavioral aspects. Random Forest model had 90 percent accuracy and had less computational overhead thus could be deployed in real time.

The comparative performance chart shown in Figure 5 on page 5 indicates the unwavering effectiveness of the AI

based framework as compared to the traditional rule based framework in all the evaluation metrics. The probabilistic risk scoring mechanism offers the commonplace of adjusting to the dissimilar levels of transaction perception, in contrast with rodent threshold mechanism.

Latency measurements showed the duration of inference to be less than one second, even when there are many user sessions. The stress test did not indicate any serious loss of system performance, which shows the scalability as well as the robustness.

The proposed framework is good at detecting the manipulation prior to the transfer of sensitive data as opposed to the traditional security framework, which is based on authentication and surveillance on the network level. Behavioral biometrics with the supervised learning can minimize false positives whilst exhibiting high detection sensitivity.

Comprehensively, the findings confirm the concept that behavioural intelligence is an effective and scalable defence strategy against the contemporary social engineering-based attacks when it is combined with the AI-driven classification.

## References

- [1] A. Jain and B. Gupta, "Phishing detection: Analysis of visual similarity based approaches," *Journal of Information Security and Applications*, vol. 36, pp. 68–81, 2017.
- [2] Y. Alsharnouby, F. Alaca, and S. Chiasson, "Why phishing still works: User strategies for combating phishing attacks," *International Journal of Human-Computer Studies*, vol. 82, pp. 69–82, 2015.
- [3] D. Bojinov, E. Bursztein, D. Boneh, and P. Bursztein, "Using behavioral biometrics for continuous authentication in real-world settings," in *Proc. IEEE Security and Privacy Workshops*, 2014, pp. 187–193.
- [4] J. Acien, A. Morales, R. Vera-Rodriguez, J. Fierrez, and R. Tolosana, "BeCAPTCHA: Behavioral bot detection via mouse dynamics," *arXiv preprint arXiv:2005.13655*, 2020.
- [5] S. Kumar, A. Arora, and R. Sharma, "Hybrid anomaly detection using mouse and keystroke dynamics for phishing prevention," *International Journal of Cybersecurity and Digital Forensics*, vol. 11, no. 2, pp. 77–87, 2022.
- [6] L. F. Cranor, "Security warning fatigue: A case study," in *Proc. 26th Annual Computer Security Applications Conference (ACSAC)*, 2016, pp. 121–130.
- [7] Bhad, "Behavioral biometrics in preventing identity fraud," Master's thesis, Dept. Comput. Sci., Univ. Oxford, Oxford, U.K., 2025.
- [8] S. Chiasson, P. C. van Oorschot, and R. Biddle, "A usability study and critique of two phishing user studies," in *Proc. 15th USENIX Security Symposium*, 2006, pp. 1–14.
- [9] C. Whittaker, B. Ryner, and M. Nazif, "Large-scale automatic classification of phishing pages," in *Proc. Network and Distributed System Security Symposium (NDSS)*, 2010.
- [10] K. Snow and F. Calabrese, "Behavioral fingerprinting for intrusion detection," in *Proc. IEEE Conference on Communications and Network Security (CNS)*, 2014, pp. 493–498.
- [11] A. Moghimi, A. B. Nassif, and R. M. A. Abdullah, "Machine learning for detecting social engineering attacks: A review," *IEEE Access*, vol. 10, pp. 12098–12115, 2022.
- [12] A. Aleroud and L. Zhou, "Phishing environments, techniques, and countermeasures: A survey," *Computers & Security*, vol. 68, pp. 160–196, 2017.
- [13] M. Jakobsson and S. Myers, *Phishing and Countermeasures: Understanding the Increasing Problem of Electronic Identity Theft*. Hoboken, NJ, USA: Wiley-Interscience, 2006.