ForLens – A Lightweight AI-Driven Endpoint Security Solution for Real-Time Threat Detection in SMEs

25-26J-076

Project Proposal Report

Salgado M. B. D – IT21289934

Supervisor: Mr. Kanishka Yapa

B.Sc. (Hons) Degree in Information Technology Specialized in Cyber Security

Department of Information Technology

Sri Lanka Institute of Information Technology

Sri Lanka

September 2025

ForLens – A Lightweight AI-Driven Endpoint Security Solution for Real-Time Threat Detection in SMEs

25-26J-076

Project Proposal Report

Salgado M. B. D – IT21289934

Supervisor: Mr. Kanishka Yapa

B.Sc. (Hons) Degree in Information Technology Specialized in Cyber Security

Department of Information Technology

Sri Lanka Institute of Information Technology

Sri Lanka

September 2025

DECLARATION

I declare that this is my own work, and this proposal does not incorporate without acknowledgement any material previously submitted for a degree or diploma in any other university or Institute of higher learning and to the best of our knowledge and belief it does not contain any material previously published or written by another person except where the acknowledgement is made in the text.

Name	Student ID	Signature
Salgado M. B. D	IT21289934	

The above candidates are carrying out research for the undergraduate Dissertation under my supervision.

Signature of the Supervisor
(Mr. Kanishka Yapa)

Date

28/08/2025

Abstract

Insider threats remain one of the most challenging issues in enterprise cybersecurity due to their subtlety, context-dependence, and the limitations of traditional rule-based detection systems. Small and medium-sized enterprises (SMEs), in particular, face heightened risk as they often lack the resources to deploy heavy, enterprise-grade forensic and monitoring solutions. This research contributes to the ForLens project "A Lightweight AI-Enhanced Forensic Tool for Real-Time Insider Threat Detection"by focusing on the design and development of an **ML-driven threat intelligence framework**.

The proposed component integrates two core elements: (1) a **machine learning anomaly detection pipeline** that model's user and system behaviors across multiple dimensions (logon patterns, file access, privilege escalation, and network activity), and (2) a **custom-built ForLens Threat Intelligence (FTI) engine**, which aggregates and normalizes data from open-source intelligence (OSINT) sources such as MISP, OTX, AbuseIPDB, and URLHaus. Together, these components enable the system not only to detect deviations from established behavioral baselines but also to contextualize suspicious activity against known global threat patterns.

The methodology employs unsupervised and semi-supervised learning techniques, such as Isolation Forest, Autoencoders, and Graph Neural Networks, to detect anomalies in heterogeneous datasets. Collected endpoint telemetry will be enriched in real time by the FTI engine before being logged in the forensic journal, ensuring cryptographically verifiable, explainable, and context-aware alerts. Anticipated results include improved accuracy in insider threat detection, reduced false positives, and SME-friendly deployment with minimal resource overhead.

This research is expected to demonstrate how combining **behavior-aware ML** with **external threat intelligence** can significantly enhance forensic visibility and resilience against insider threats. It will contribute not only to the academic understanding of hybrid detection systems but also to the practical development of lightweight, commercially valuable cybersecurity solutions tailored for SMEs.

Keywords: Insider Threat Detection, Machine Learning, Threat Intelligence, Anomaly Detection, Forensic Security

Table of Contents

List of Figures

List of tables

List of Appendices

List of Abbreviations

Abbreviation	Full Form
AI	Artificial Intelligence
API	Application Programming Interface
CTI	Cyber Threat Intelligence
EDR	Endpoint Detection and Response
FTI	ForLens Threat Intelligence
IDS	Intrusion Detection System
IOC	Indicator of Compromise
IoC DB	Indicator of Compromise Database
ML	Machine Learning
OSINT	Open-Source Intelligence
SIEM	Security Information and Event Management
SME	Small and Medium Enterprise
SOAR	Security Orchestration, Automation and Response
XDR	Extended Detection and Response

1. Introduction

Insider threats represent one of the most persistent challenges in modern cybersecurity. Unlike external attacks, these threats emerge from individuals who already possess legitimate access to organizational systems, making them inherently harder to identify and prevent. Reports indicate that insider incidents account for a growing proportion of data breaches, often causing more financial and reputational damage than traditional cyberattacks due to the difficulty of detection and delayed response. [1]

Conventional enterprise security solutions such as Security Information and Event Management (SIEM) and Endpoint Detection and Response (EDR) systems have been effective to some extent; however, they largely depend on rule-based signatures or static policies. These approaches often generate high false positives, fail to capture subtle behavioral deviations, and provide limited real-time forensic insight [2]. Consequently, organizations particularly Small and Medium Enterprises (SMEs) struggle to deploy these resource-intensive solutions at scale.

Recent advances in artificial intelligence and machine learning provide a new path for detecting malicious behavior. Unsupervised anomaly detection models, such as Isolation Forests, Autoencoders, and Long Short-Term Memory (LSTM) networks, can establish baselines of user activity and flag unusual deviations without requiring pre-labeled data [3]. Such techniques allow for dynamic adaptation to evolving attack patterns and reduce reliance on manual rule creation.

However, anomaly detection in isolation may lack the broader security context necessary to support meaningful investigations. Integrating open-source threat intelligence (OSINT) feeds can enrich anomaly alerts with external Indicators of Compromise (IoCs), reputation scores, and attack campaign information. This combination enhances accuracy, reduces false alarms, and ensures that alerts are actionable [4].

The ForLens project builds upon this foundation by proposing a **lightweight, explainable, AI-enhanced forensic tool** that integrates **behavior-aware anomaly detection** with a custom-built **ForLens Threat Intelligence (TI) framework**, aggregating multiple OSINT sources into one enriched feed. By doing so, the system provides SMEs with a cost-friendly yet advanced capability to detect and investigate insider threats in real time, offering forensic visibility previously available only to large enterprises [5].

1.1 Background & Literature survey

The rapid digitalization of small and medium-sized enterprises (SMEs) has significantly expanded their reliance on endpoint devices such as workstations, laptops, and mobile systems—for daily business operations. However, this dependency has simultaneously increased their exposure to cyber threats, including malware, phishing, unauthorized access, and insider misuse [6]. Recent studies emphasize that over **60% of cyber incidents are linked to endpoints**, making them one of the most critical attack surfaces [7]. Unlike large enterprises, SMEs face challenges in adopting robust endpoint detection and response (EDR) solutions due to limited financial resources, expertise, and infrastructure [8]. Consequently, lightweight, adaptive, and cost-effective approaches are essential to secure SME environments without imposing significant operational overhead.

Insider Threats and Behavioral Anomalies

Among the variety of risks, **insider threats** remain one of the most difficult to detect and mitigate. Insiders often operate with legitimate credentials, making their malicious activities resemble normal user behavior [9]. The **Carnegie Mellon University (CMU) CERT Insider Threat dataset** has been widely used to model such behaviors, providing logon, device, email, and web traces that reflect both normal activity and malicious deviations [10]. Previous research utilizing this dataset demonstrates that subtle anomalies such as unusual logon times, excessive file transfers, or abnormal web browsing can provide valuable signals of insider activity when analyzed in context [11].

Traditional Detection Approaches

Conventional **Security Information and Event Management (SIEM)** systems aggregate logs and apply rule-based correlation to detect threats. While effective for compliance monitoring, they struggle against low-and-slow attacks and often overwhelm analysts with false positives [12]. Endpoint Detection and Response (EDR) tools, on the other hand, provide greater visibility but require substantial resources and expertise, which SMEs typically lack [13]. Research shows that the detection latency and complexity of configuration remain major obstacles in traditional solutions [14]. These limitations highlight the need for **automated, intelligent anomaly detection mechanisms** that can operate with minimal configuration and resource demands.

Machine Learning for Anomaly Detection

Machine learning (ML) has emerged as a promising approach to address the shortcomings of traditional systems. Unsupervised techniques such as **Isolation Forest (iForest)** [15], **autoencoders** [16], and **sequence models (LSTM-based DeepLog)** [17] have demonstrated strong potential in modeling normal behavior and flagging deviations without requiring extensive labeled datasets. Such models are particularly effective for insider threat detection, where labeled malicious examples are scarce [18]. However, research also highlights the challenge of ensuring interpretability i.e., explaining why an alert was triggered to improve analyst trust and adoption [19].

Threat Intelligence and Data Enrichment

Another dimension explored in recent studies is the integration of **cyber threat intelligence (CTI)** feeds. Standards like **STIX 2.1** and **TAXII** enable automated sharing of indicators of compromise (IoCs), tactics, techniques, and procedures (TTPs) [20]. Open-source platforms such as **MISP (Malware Information Sharing Platform)** facilitate collaborative intelligence gathering [21]. Research suggests that fusing endpoint telemetry with CTI improves detection accuracy and context-awareness, bridging the gap between raw anomaly detection and actionable incident response [22].

Forensic Logging and Explainability

Finally, forensic readiness is critical in modern threat detection solutions. **Tamper-evident logging mechanisms**, such as hash-chained journals and transparency logs, ensure that collected evidence cannot be manipulated by malicious insiders [23]. Furthermore, explainability frameworks like **SHAP (Shapley Additive Explanations)** and **LIME (Local**

Interpretable Model-Agnostic Explanations) are being increasingly integrated with ML-based systems, providing human-understandable insights into why specific activities are flagged as anomalous [24].

1.2 Research GAP

Existing commercial SIEM platforms including Microsoft Sentinel, IBM QRadar, Splunk Enterprise Security, Elastic Security, and Exabeam offer strong capabilities in log aggregation, user behavior analytics, and correlation-based threat detection. While these tools are highly effective for large organizations, they often fall short in areas that are particularly important for small-to-medium enterprises (SMEs) and for research-driven security evaluation.

One critical gap lies in the lack of **explainable machine learning**. Most SIEMs today produce risk scores or generate behavioral alerts without providing detailed reasoning behind them. Analysts are typically left with aggregate summaries or high-level dashboards, which reduces interpretability and makes it harder to build trust in the system. ForLens directly addresses this by offering SHAP/LIME-style explainable outputs for each alert, giving clear, per-alert explanations of why a specific event was flagged [25].

Another major limitation is in the area of **unified threat intelligence**. Although commercial SIEMs allow ingestion of threat intelligence feeds, this process is often fragmented and requires additional customization. For example, Sentinel and QRadar depend on custom connectors, while Splunk relies on third-party applications. This means that intelligence from open sources such as OTX, Abuse.ch, and PhishTank is rarely fused seamlessly with anomaly detection models [26], [27]. ForLens fills this gap by enabling direct integration and correlation of multiple OSINT feeds, making intelligence-driven detection much more streamlined.

The issue of **SME deployment suitability** further highlights these shortcomings. Enterprise-grade SIEMs typically demand heavy infrastructure and high licensing costs, making them unrealistic for smaller organizations. Even open-source options like Elastic Security, though cost-effective in principle, require expert-level tuning and configuration before becoming effective. By contrast, ForLens has been designed to be lightweight, affordable, and easy to deploy, without compromising on core security functionality [28].

There are also weaknesses in **insider threat and endpoint coverage**. While UEBA-focused SIEMs are capable of identifying anomalies at a broad level, they often overlook more subtle but equally critical behaviors, such as USB device access, process injections, or file exfiltration. ForLens overcomes this limitation by offering full-spectrum monitoring that captures activities across users, files, processes, networks, and even USB devices [29].

Finally, there is the problem of **evaluation transparency**. Proprietary SIEM platforms usually rely on closed datasets and undisclosed models, which significantly limits reproducibility and restricts their usefulness for academic or independent research. Elastic Security makes some progress by providing a few open ML jobs, but it still lacks complete openness. ForLens, on the other hand, is built with transparency in mind: it supports reproducible datasets and provides fully documented ML models, making it equally valuable for both operational deployment and academic study [30].

Feature / Dimension	Microsoft Sentinel	IBM QRadar	Splunk Enterprise Security	Elastic Security	Exabeam	ForLens (Proposed)
Explainable ML (SHAP/LIME-style justifications)	✗	✗	✗	✗	✗	☑
Unified OSINT Threat Intel (OTX, Abuse.ch, PhishTank, MISP)	⚠ (via connectors)	⚠ (via feeds)	⚠ (TI add-ons)	✗	✗	☑
SME Suitability (Low cost, low infra, fast deploy)	✗ (cloud pay-as-you-go)	✗ (enterprise-scale infra)	✗ (premium licensing)	⚠ (open source but complex)	✗	☑
Insider Threat & Endpoint Depth (USB, file exfil, process injection)	⚠ (depends on data sources)	⚠ (UBA app limited)	⚠ (requires premium apps)	⚠ (anomaly jobs limited)	⚠ (UEBA focus only)	☑
Evaluation Transparency (Open datasets, replicability in academic research)	✗	✗	✗	⚠ (some open ML jobs)	✗	☑

Table 1: Comparative Gaps of Popular SIEMs vs ForLens

1.3 Research Problem

Insider threats continue to be one of the most challenging cybersecurity issues for organizations of all sizes. Unlike external attacks, insiders possess legitimate access credentials, making it difficult to distinguish between normal and malicious behavior. Traditional monitoring systems including SIEMs, EDRs, and UEBA platforms often rely on predefined rules, signatures, or high-level behavioral models. While effective in detecting well-known threats, these approaches are limited in several critical ways,

1. **Lack of Explainability:** Most commercial solutions provide risk scores or alerts but do not offer transparent, per-alert explanations of why a particular behavior was flagged. This limits analysts' trust in automated decisions and makes it difficult to justify actions to management or auditors.
2. **Fragmented Threat Intelligence Integration:** Although many SIEM and XDR platforms support threat intelligence feeds, integration is often partial or requires manual configuration. Open-source intelligence (OSINT) from sources such as PhishTank, Abuse.ch, and AlienVault OTX is rarely fused with anomaly detection models out-of-the-box. This reduces the ability to detect novel or emerging threats.
3. **High Complexity and Cost for SMEs:** Enterprise-grade security platforms demand significant infrastructure, licensing costs, and expert personnel to operate effectively. Small and medium enterprises (SMEs) often cannot afford the complexity or resources, leaving them vulnerable to sophisticated insider attacks.
4. **Limited Endpoint and Insider Visibility:** Existing solutions frequently focus on high-level activity monitoring (network logs, authentication patterns, or cloud activity) but fail to capture lower-level behaviors such as USB usage, process injections, or unauthorized file manipulations. This gap reduces the ability to detect insider threats before significant damage occurs.
5. **Research and Evaluation Limitations:** Closed-source SIEM and XDR platforms make it difficult to evaluate model performance, reproduce results, or experiment with alternative algorithms. For research-focused applications, transparency and reproducibility are essential, yet rarely supported in existing commercial solutions.

Given these limitations, there is a clear need for a solution that combines **explainable machine learning, unified threat intelligence, SME-friendly deployment, and comprehensive endpoint visibility**. The ForLens project aims to address this gap by creating a lightweight, transparent, and effective framework for insider threat detection, integrating ML-driven anomaly detection with enriched threat intelligence feeds to provide actionable, interpretable insights.

2. Objectives

2.1 Main Objective

The primary objective of this research is to **develop an effective, transparent, and SME-friendly insider threat detection system** that leverages explainable machine learning in combination with integrated threat intelligence. This system, ForLens, aims to overcome the limitations of existing SIEM, XDR, and UEBA platforms by providing actionable, interpretable alerts, comprehensive endpoint visibility, and seamless threat intelligence integration. The overarching goal is to enhance the early detection of insider threats while

maintaining usability, scalability, and cost-effectiveness for organizations with limited security resources.

2.2 Specific Objectives

1. **Design and implement an explainable ML-based detection engine:**

   Develop machine learning models capable of detecting anomalous insider behaviors, with transparent per-alert explanations to improve analyst trust and decision-making.

2. **Integrate threat intelligence feeds effectively:**

   Incorporate both proprietary and open-source threat intelligence (OSINT) into the detection framework, enabling the correlation of internal activity with external threat indicators to detect emerging and sophisticated insider threats.

3. **Provide comprehensive endpoint and activity coverage:**

   Extend monitoring to endpoints, including USB devices, process trees, file access, network activity, and user behavior, to ensure full-spectrum visibility of potential insider threats.

4. **Develop SME-friendly deployment and operation:**

   Ensure the system can be deployed with minimal infrastructure and cost requirements, supporting small and medium-sized enterprises without extensive security teams.

5. **Support evaluation, auditability, and reproducibility:**

   Enable performance assessment of detection models, provide audit trails for alerts, and maintain transparency in ML-based decisions to facilitate compliance, research validation, and continuous improvement.

6. **Facilitate actionable alerts and analyst support:**

   Generate prioritized, context-rich alerts with actionable recommendations to allow security analysts to respond efficiently and effectively to potential insider threats.

7. **Benchmark against existing solutions:**

   Compare ForLens's performance, coverage, and usability against popular SIEM, UEBA, and XDR solutions to demonstrate measurable advantages in detection accuracy, explainability, and operational feasibility.

3. Methodology

The methodology adopted in this research is structured into five distinct yet interconnected phases. Each phase ensures the systematic development of a lightweight, AI-driven endpoint forensic solution tailored to detect insider threats and anomalies in real-time, while also integrating open-source threat intelligence to enrich decision-making.

3.1 Data Collection and Preprocessing

The project begins with the collection of multiple data sources:

- **Benchmark Insider Threat Datasets**: Publicly available datasets such as the *CERT Insider Threat Dataset* are used to model initial behavior patterns and simulate real-world insider threat cases.
- **Endpoint Telemetry**: Logs from the lightweight ForLens agent (e.g., logon/logoff times, file access, process behaviors, network activity) provide raw behavioral data.
- **Threat Intelligence Feeds**: Open-source intelligence (OSINT) feeds such as AbuseIPDB, AlienVault OTX, MISP, and VirusTotal are integrated to enrich the dataset.

Preprocessing involves cleaning noisy or missing records, standardizing timestamps, anonymizing sensitive data, and feature engineering. Feature extraction includes temporal features (time-of-day access), statistical baselines (average file access counts), and categorical mappings (role, privilege level). This ensures the dataset is structured for both supervised and unsupervised learning tasks.

3.2 Threat Intelligence (TI) Integration

A dedicated **ForLens Threat Intelligence Module** is developed to unify multiple open-source TI feeds into a single enrichment pipeline.

- **Aggregation**: Collects IP/domain reputation data, malware indicators, and emerging attack signatures from OSINT sources.
- **Normalization**: Standardizes feed formats into a common schema (JSON/CSV).
- **Enrichment**: Matches incoming endpoint activity (e.g., suspicious network connections) with TI indicators.
- **Integration**: Enriched data is passed into the ML pipeline to improve accuracy and reduce false positives.

This creates a **unique contribution** compared to existing commercial solutions, where TI enrichment is often siloed or vendor-specific.

3.3 Machine Learning Model Development

The ML pipeline is designed to support both **unsupervised** and **semi-supervised** learning, as insider threats are often rare and lack labeled training data.

- **Baseline Modeling**: Isolation Forest and One-Class SVMs are used for anomaly detection of rare behaviors.
- **Deep Learning**: LSTM Autoencoders capture sequential anomalies such as abnormal file access patterns or irregular login sequences.
- **Graph-Based Analysis**: Graph Neural Networks (GNNs) model user-to-resource relationships to detect unusual access paths.
- **Explainability**: SHAP and LIME are employed to generate human-readable justifications (e.g., “User accessed HR files at 2 AM, which deviates from normal working hours”).

The model training phase starts with benchmark datasets (CERT), then transitions into adaptive learning as real agent data becomes available.

3.4 System Integration

The ForLens system integrates four core components

1. **Lightweight Endpoint Agent** – Monitors process activity, file access, and network connections while remaining resource-efficient and resistant to tampering. Includes a honeypot module to attract malicious activity.
2. **Forensic Journal Engine** – Logs all events in a cryptographically chained structure (mini-blockchain), ensuring data integrity and tamper-evidence.
3. **Threat Intelligence Module** – Enriches raw logs with external threat context.
4. **Central Analysis Pipeline** – ML models analyze enriched telemetry in real time and generate alerts.

Data is transmitted securely (TLS + encryption at rest) to the central analysis system, ensuring forensic soundness and regulatory compliance.

3.5 Evaluation and Validation

Evaluation of the solution will follow three dimensions:

- **Functional Testing**: Simulated insider threat scenarios (after-hours logins, mass file downloads, privilege escalation) will be tested against ForLens.
- **Performance Metrics**: Detection accuracy, precision-recall (to minimize false positives), latency of real-time detection, and resource overhead on SME devices.
- **Comparative Benchmarking**: Compare ForLens against commercial tools (e.g., Splunk UBA, Microsoft Defender for Endpoint, Cynet 360) to highlight advantages in lightweight deployment, cost efficiency, and explainability.

3.6 USE CASE

Use Case 01	
Use case id	UC001
Name	After-hours login anomaly
Description	An employee logs in at 2 AM for the first time in six months. Such access outside of normal working hours may indicate compromised credentials or intentional insider misuse.
Application	ForLens insider threat detection system for SMEs.
Primary actor	Employee (legitimate or malicious).
Pre-condition	• User account must be active. • Authentication system logs are being monitored. • ML anomaly detection model is active. • Threat Intelligence (TI) feed integration enabled.
Trigger	Employee attempts login outside normal working hours.

Basic flow	• Employee initiates login at 2 AM. • ForLens captures the login event. • ML engine compares login behavior with historical patterns. • System detects significant deviation (after-hours access). • TI checks IP address reputation for malicious activity. • Alert is generated and sent to the security team.
Outcome	• Suspicious login is flagged with context, enabling SOC to investigate possible compromised credentials.

Use Case 02	
Use case id	UC002
Name	Abnormal file access
Description	A privileged user downloads 2000 PDFs within 10 Seconds, which is inconsistent with normal file usage behavior.
Application	ForLens anomaly detection module.
Primary actor	Privileged employee (potential insider threat).
Pre-condition	• User has legitimate access to shared file systems. • Endpoint and file server logs are enabled. • ML monitoring and TI feed are active.
Trigger	Bulk file download activity detected.
Basic flow	• User initiates a mass download of 2000 PDFs. • ForLens collects file activity logs. • ML model identifies an abnormal spike compared to baseline activity. • TI cross-checks for known exfiltration tools/processes. • Alert is triggered with enriched context. • SOC team investigates and takes action if necessary.
Outcome	• Mass download flagged; data exfiltration attempt prevented or mitigated.

Use Case 03	
Use case id	UC003
Name	Suspicious device use
Description	A user who has no history of external storage usage suddenly connects multiple USB drives to their endpoint.

Application	ForLens endpoint monitoring component.
Primary actor	Employee with endpoint access.
Pre-condition	• USB device monitoring is enabled on endpoints. • Historical user behavior data is available. • TI feed is active for driver and hash reputation.
Trigger	Connection of multiple USB devices to the endpoint.
Basic flow	• Employee connects first USB drive. • ForLens records device insertion event. • Employee connects multiple devices in a short period. • ML identifies deviation from baseline usage. • TI checks driver hash reputation. • System generates insider threat alert.SOC team investigates and takes action if necessary.
Outcome	• Unauthorized device usage detected early, preventing data leakage.

Use Case 04	
Use case id	UC004
Name	Process injection attempt
Description	A process attempts to inject into svchost.exe, which is highly unusual for normal operations and indicates potential malicious behavior.
Application	ForLens process behavior monitoring.
Primary actor	Endpoint process (possibly malware).
Pre-condition	• Endpoint monitoring is active. • Process tree analysis enabled in ForLens. • TI feed for malware hash verification is running.
Trigger	Unusual process injection detected.
Basic flow	• Endpoint process attempts injection into svchost.exe. • ForLens detects suspicious process chain activity. • ML flags it as an anomaly. • TI confirms process hash as malicious or suspicious. • Alert is sent with high priority to the SOC.
Outcome	• Malware process identified and isolated, reducing risk of privilege escalation.

Use Case 05	
Use case id	UC005

Name	Lateral movement attempt
Description	A user attempts to log into another employee's dedicated machine, which they have never accessed before.
Application	ForL ForLens user behavior monitoring.ens endpoint monitoring component.
Primary actor	Employee account (possibly compromised).
Pre-condition	• User account must be active. • Endpoint login monitoring enabled. • TI feed integration active for IP reputation.
Trigger	Unauthorized login attempt on another endpoint.
Basic flow	• Employee attempts login to another user's workstation. • ForLens logs and analyzes the login event. • ML model flags deviation from user's normal machine access. • TI validates source IP address reputation. • Alert is triggered for investigation.
Outcome	• Lateral movement attempt is blocked or alerted, stopping early-stage attacks.

Use Case 06	
Use case id	UC006
Name	Malicious domain access
Description	An employee attempts to access a suspicious or blacklisted website.
Application	ForLens network traffic monitoring.
Primary actor	Employee browsing the internet.
Pre-condition	• Network monitoring enabled. • TI feed with malicious domain blacklist integrated.
Trigger	Access request to a known suspicious domain.
Basic flow	• Employee browses the internet. • ForLens captures DNS request. • ML detects unusual browsing compared to baseline activity. • TI confirms domain is blacklisted. • Alert is generated with detailed context.
Outcome	• Malicious domain access blocked; user and SOC notified.

Use Case 07	
Use case id	UC007
Name	Ex-employee account activity
Description	An employee marked as terminated in LDAP still attempts to log in again.
Application	ForLens identity and access monitoring.
Primary actor	Ex-employee account.
Pre-condition	• LDAP/AD integration with ForLens enabled. • User status (active/terminated) properly updated. • ML anomaly detection is active.
Trigger	Login attempt by terminated employee account.
Basic flow	• Ex-employee account logs in despite termination. • ForLens records authentication attempt. • ML immediately flags activity as anomalous. • TI enrichment checks IP reputation. • Alert generated and escalated.
Outcome	• Unauthorized ex-employee login blocked, preventing insider misuse.

Use Case 08	
Use case id	UC008
Name	Combined anomaly correlation
Description	Multiple anomalies occur simultaneously: USB usage, after-hours login, and data exfiltration to a blacklisted IP.
Application	ForLens correlation engine.
Primary actor	Employee account (potential insider threat).
Pre-condition	• Endpoint, network, and identity monitoring active. • ML anomaly detection working across multiple logs. • TI enrichment feeds active.
Trigger	Concurrent anomalies detected across different domains.
Basic flow	• Employee logs in after-hours. • Connects a USB device. • Exfiltrates data to suspicious IP address. • ForLens correlates anomalies from multiple sources. • ML + TI validation increases detection confidence. • System raises a high-confidence insider threat alert.

Outcome	• Multi-domain anomaly correlation boosts accuracy, reducing false positives and enabling faster incident response.

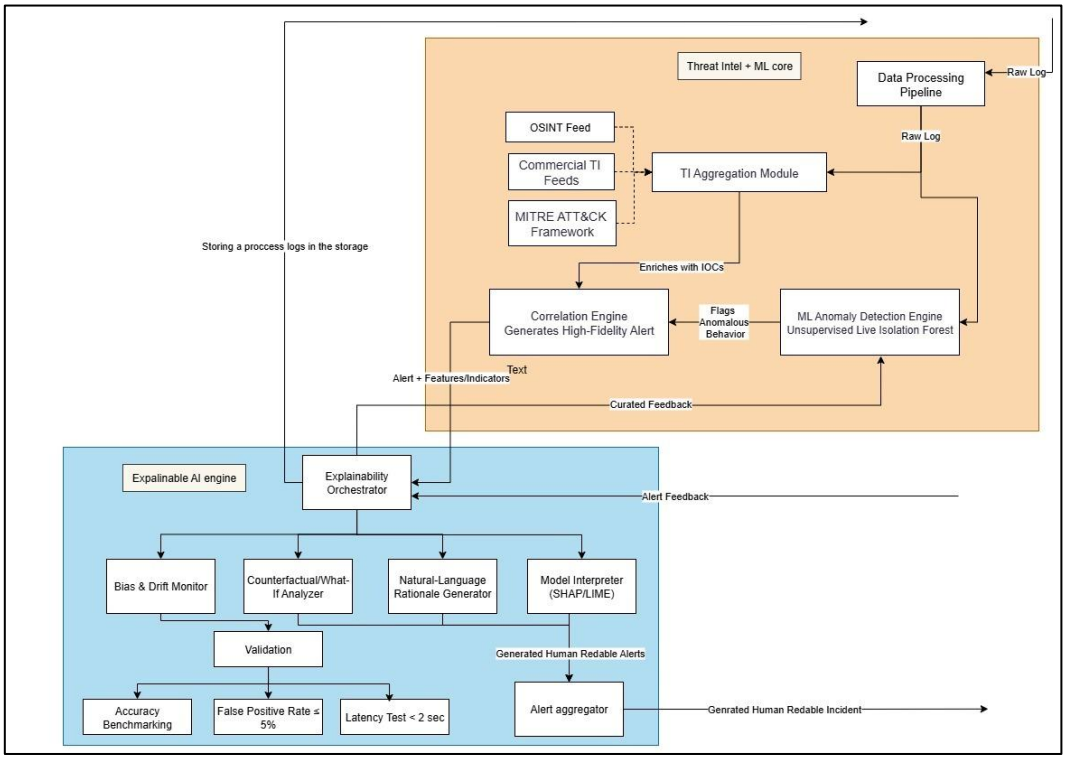

Figure 1: High Level Architecture Diagram for ForLens Threat Intelligence and ML Component

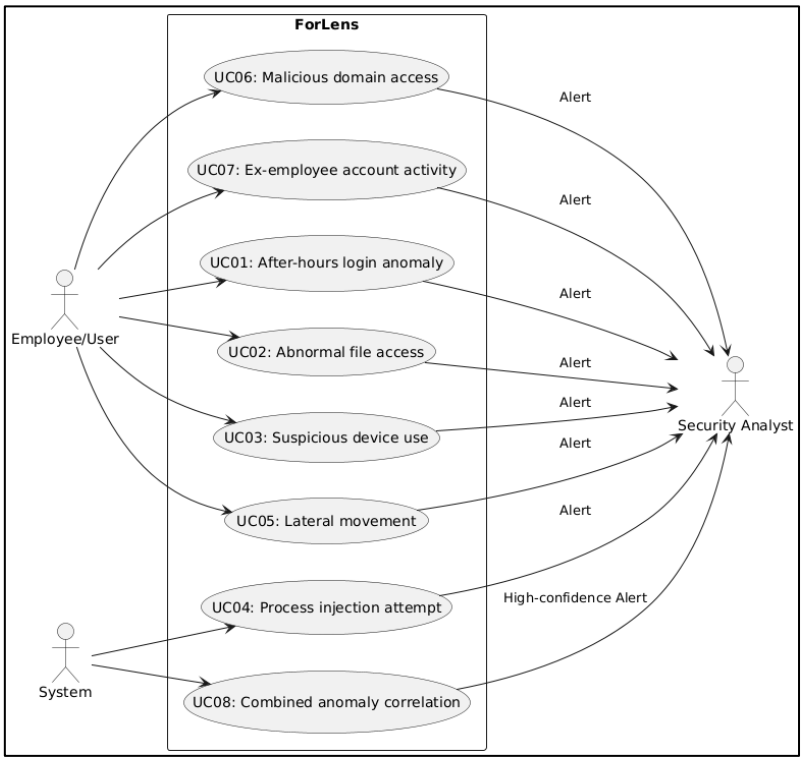

Figure 2: Use Case Diagram for ForLens Threat Intelligence and ML Component

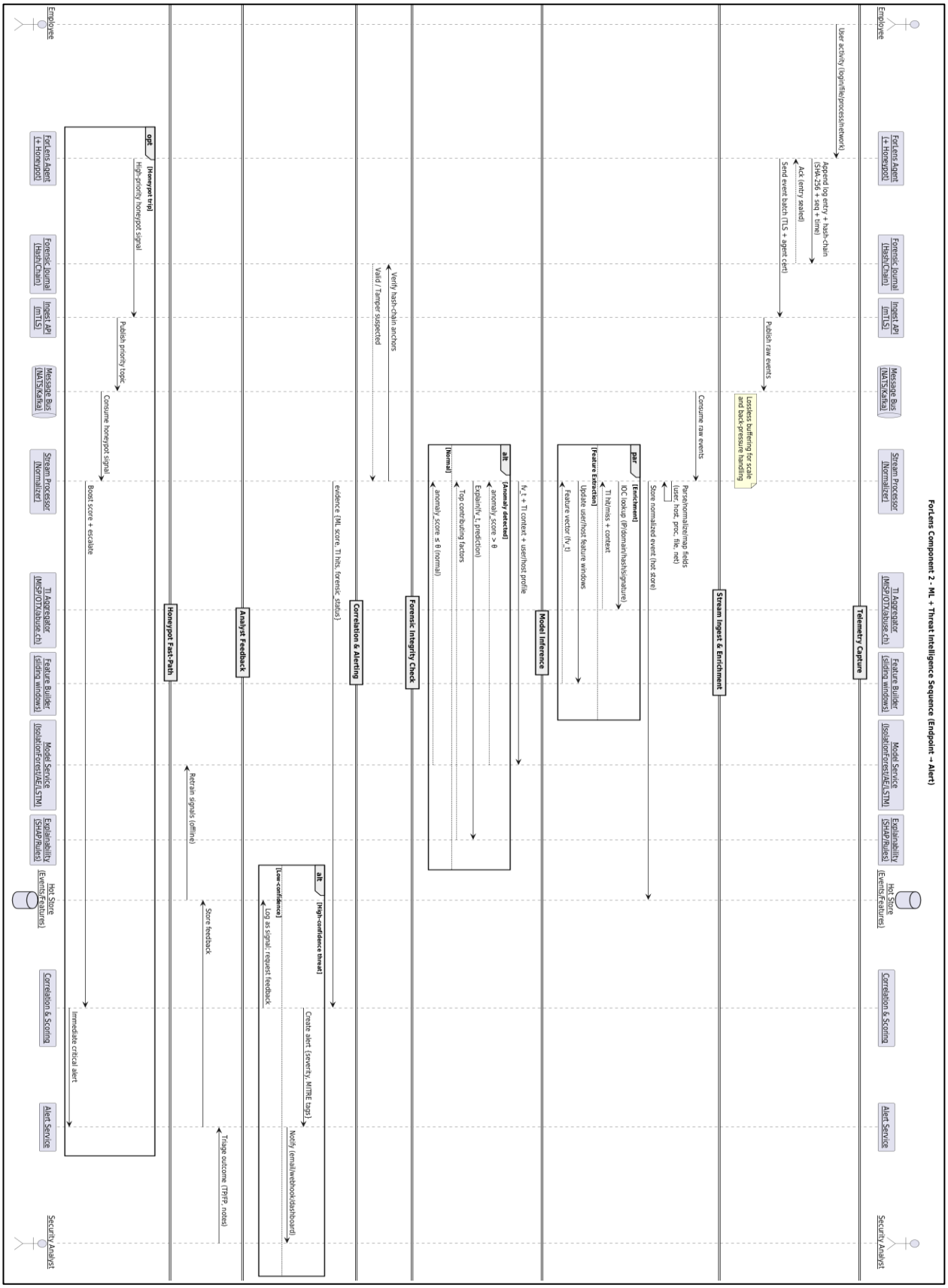

Figure 3:Sequence Diagram for Threat Intelligence and Machine Learning Workflow

3.7 Commercialization

Target Audience

- Small and Medium Enterprises (SMEs) lacking dedicated cybersecurity teams
- Managed Security Service Providers (MSSPs) looking for cost-effective insider threat detection solutions
- IT Administrators and SOC Analysts in SME environments
- Educational institutions, healthcare centers, and financial organizations with limited budgets but high data sensitivity
- Government agencies and NGOs seeking lightweight endpoint monitoring for compliance

Market Space

In an age where digital innovation is surging forward at an unprecedented pace, the cybersecurity landscape for small and medium enterprises (SMEs) is experiencing explosive growth, propelled by a dramatic rise in ransomware attacks and the persistent threat of insider breaches that jeopardize business continuity. Envision a reality where SMEs, vital to economic frameworks worldwide, confront a 93% spike in ransomware incidents in 2024, with each incident now commanding an average ransom of $1.5 million. Insider threats whether deliberate, such as data exfiltration by discontented staff, or inadvertent, accounting for 88% of breaches due to human oversight inflict an average annual toll of $17.4 million per organization in 2024, a 109% increase since 2018. This transcends a mere technical issue; it's a critical survival challenge for SMEs with limited resources, rendering them prime targets as global cybercrime losses are forecasted to reach $10.5 trillion by the end of 2025. [31]

ForLens emerges as a transformative solution in this high-stakes environment: an AI-empowered EDR/XDR platform enriched with threat intelligence, crafted for SMEs operating with tight budgets. On a global scale, these businesses are escalating their cybersecurity expenditures, with estimates projecting a climb to $90 billion in 2025 from $57 billion in 2020, supported by a strong 14% compound annual growth rate. This upward trend reflects a pressing need for cost-effective, user-friendly tools that don't strain small IT teams. [32] Notably, managed security services are poised to capture nearly a third of this investment ($29.8 billion), signaling a shift toward outsourced, flexible defenses that align with ForLens's offerings. [33]

Shifting focus to regional opportunities, the Sri Lankan and broader South Asian SME markets present a fertile ground amid significant hurdles. With digital adoption accelerating South Asia's internet penetration reached 60% in 2024 enterprise-grade EDR/XDR solutions remain inaccessible due to costs often 2-3 times beyond SME budgets and deployment complexities requiring scarce specialized skills. [34] In Sri Lanka, where SMEs contribute 52% to GDP, only 30% maintain adequate cybersecurity, leaving them vulnerable to escalating threats like phishing and insider leaks, amplified by hybrid work setups [35]. ForLens addresses this with its lightweight endpoint agents, needing minimal infrastructure no heavy servers or constant supervision ideal for the 70% of regional SMEs managing with fewer than five IT staff.

Compliance adds another layer of urgency: organizations adhering to regulations like GDPR, HIPAA, or Sri Lanka's Personal Data Protection Act face penalties up to 4% of global revenue for breaches. ForLens's integrated machine learning and threat intelligence system delivers

compliance through automated anomaly detection and real-time reporting, tackling the 65% of SMEs who cite managing multiple tools as a major obstacle.

The global cybersecurity market further validates ForLens's timeliness, with projections indicating growth from $235.5 billion in 2025 to $423.43 billion by 2030 (a 12.45% CAGR), and SMEs driving software adoption at a 14.2% growth rate. Insider threat analytics, a key ForLens strength, is surging, with the detection market expected to rise from $420.6 million in 2024 to $1.814 billion by 2030 (27.9% CAGR), and overall protection growing from $4.45 billion in 2023 to $13.69 billion by 2030 (17.4% CAGR). This creates an ideal landscape for ForLens, merging AI-driven behavioral analysis with threat intelligence for proactive defense, empowering SMEs to counter sophisticated cyber threats effectively. [36]

A recent report highlights a significant increase in cybersecurity spending among small and mid-sized enterprises (SMEs), projected to grow by over $30 billion in the next four years, reaching $90 billion by 2025. Managed security services are expected to constitute one-third of this total, driven by a 14% CAGR, up from $16 billion in 2020. The surge is fueled by heightened vulnerability due to the digital acceleration during the pandemic, with 23% of 6,000 surveyed SME leaders across multiple countries reporting cyber-attacks in the past year, and US SMEs facing an average cost of $25,612 per attack. Globally, SME cybersecurity spending rose from $57 billion in 2020, with managed and network security dominating over half the market. Regionally, North America, emerging Asia-Pacific, and Western Europe accounted for 73% of spending in 2020, though Asia-Pacific's share is expected to grow from 23% to 27% by 2025, presenting new opportunities as countries like China and India digitize rapidly. The shift toward managed services, accelerated by the pandemic, is evident, with over half of businesses in Australia, Canada, the UK, and the US increasing their use of such services. [37]

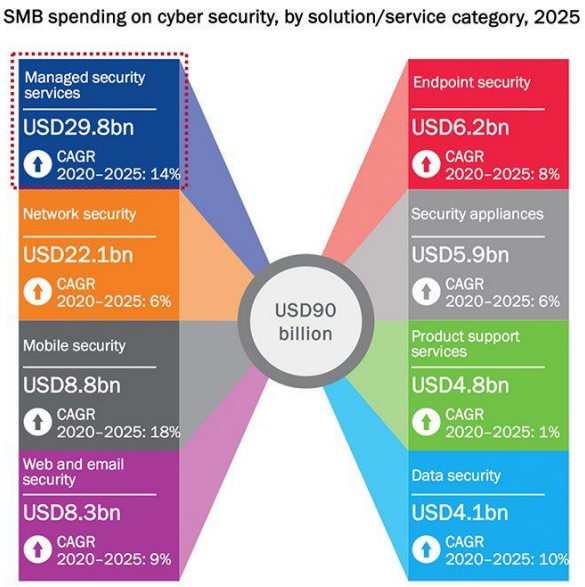

Figure 4:SMB Cybersecurity Spending (2020-2025)

The Global Cyber Security Market is projected to grow from USD 192.4 billion in 2023 to USD 608.3 billion by 2033, with a CAGR of 12.2% from 2024 to 2033, led by North America, which held a 36.8% share (USD 70.8 billion) in 2023. Cybersecurity encompasses technologies and practices to protect networks, devices, and data from threats like viruses, phishing, and hacking, gaining urgency as digital breaches threaten privacy, corporate security, and national

safety. The market's rapid expansion is driven by sophisticated cyber threats, regulatory demands, and the proliferation of IoT and smart devices, boosting demand for advanced solutions like next-generation firewalls, encryption, and AI/ML-integrated platforms. This growth is fueled by increasing awareness of financial and reputational risks, with cybercrime costs rising from $8 trillion in 2023 (over $250,000 per second) to a projected $10.5 trillion by 2025. Notably, Ipsos data from 2023 shows 33% of Americans, especially those aged 35-54 (36%), experienced online financial fraud, compared to 22% of those aged 18-34, highlighting the escalating threat landscape. [38]

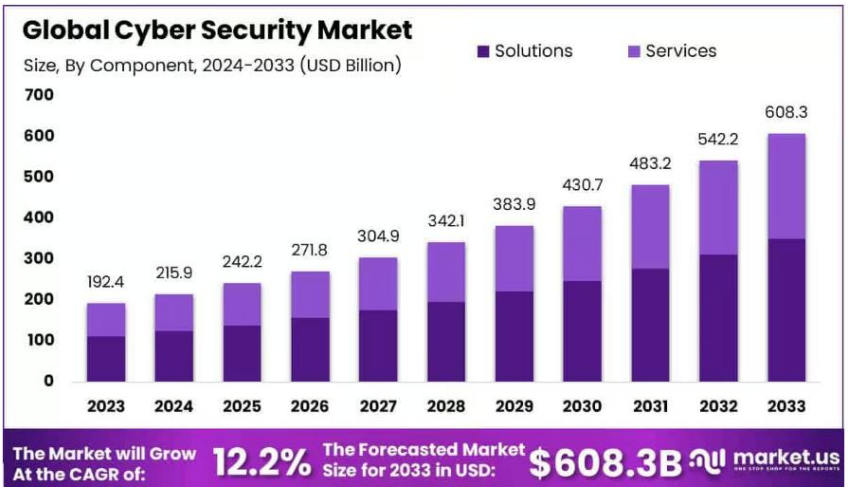

Figure 5: Global Cybersecurity Market Growth (2023-2030)

Insider threats are increasingly prevalent, with 76% of organizations noting a rise in incidents over five years, though only 30% feel equipped to handle them. Between 2023 and 2024, insider-driven data exposure increased by 28%, and Ponemon reported a 44% rise in related incidents from 2020-2022. In 2023, 71% of companies faced 21-40 incidents annually, with three-quarters of security leaders observing more frequent attacks. Non-malicious errors, causing 88% of breaches, are most common in public administration, while healthcare leads in malicious incidents. Senior managers (81%) and sales/customer service roles (48%-47%) are seen as high-risk, driven by insufficient training (37%) and new technologies (34%). Malicious threats, motivated by financial gain (89%), cost $701,500 per incident, with total organizational costs averaging $16.2 million in 2023, up 40% since 2019. North America faces the highest costs at $19.09 million annually. Detection is challenging, with 90% of professionals finding insider threats as hard or harder to address than external ones, exacerbated by cloud use (53%) and remote work (70% concern). Mitigation efforts include 72% dedicating resources to prevention, though only 21% have fully operational programs, with 86% monitoring user behavior. [39]

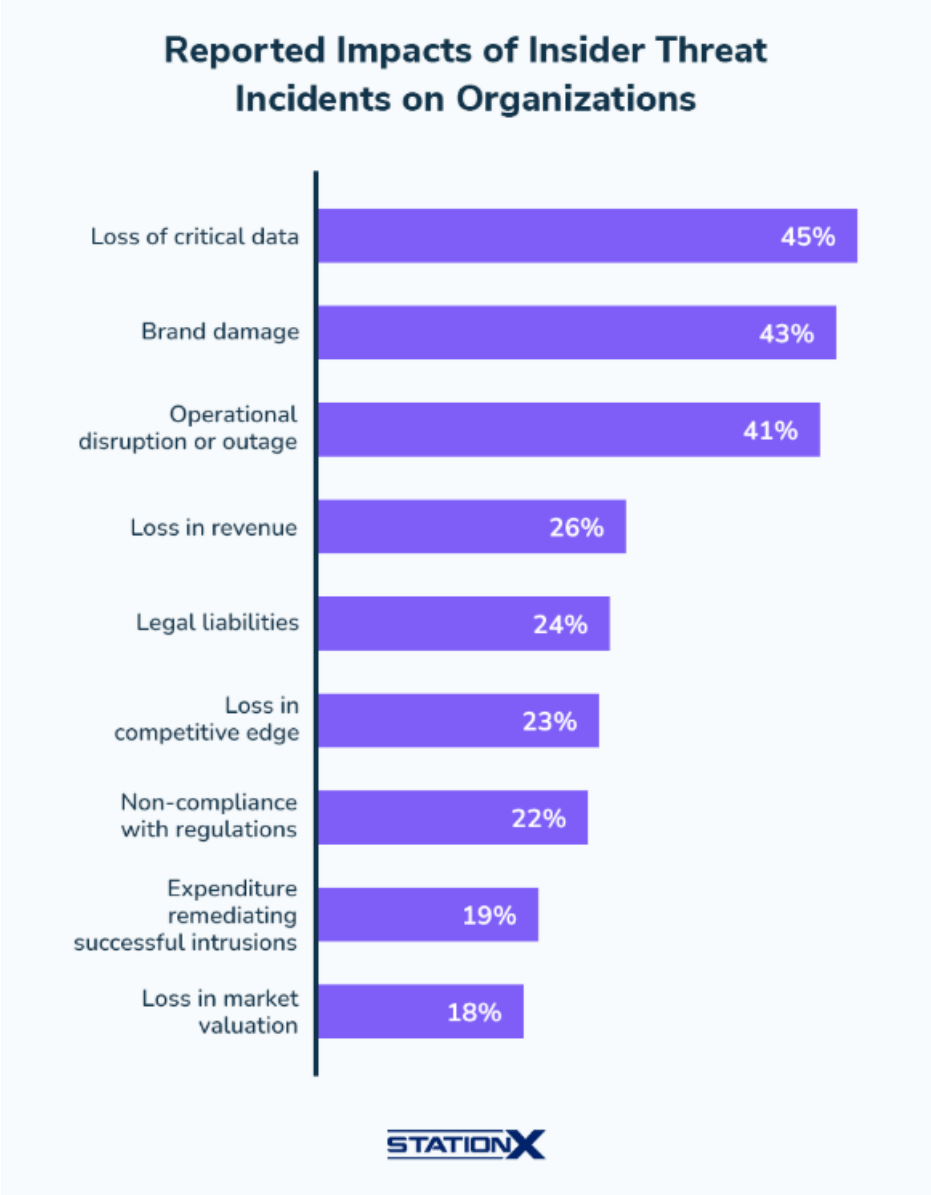

Figure 6: Reported Impacts of Insider Threat incidents on organization

Together, these visuals and data outline a $90 billion SME opportunity in 2025, expanding to a $500 billion global market by 2030 prime territory for ForLens's streamlined, integrated approach to convert vulnerabilities into robust defenses.

How to Promote

- **Free beta program** for SMEs: deploy the ForLens agent + ML/TI engine for early adopters to demonstrate value
- **Community editions** with limited features to attract startups and educational institutions
- Promotion via **social media campaigns**, **LinkedIn security communities**, and **cybersecurity conferences**
- Partnerships with **universities and research incubators** to encourage adoption and gather feedback

- Offer **subscription-based pricing**: monthly, yearly, and enterprise lifetime licenses to maximize flexibility
- Integration with **open-source threat intelligence feeds** as a unique selling point (USP) to differentiate from costly enterprise products
- Collaboration with **local MSSPs** to bundle ForLens as part of managed security services for SMEs

4. Project requirements

4.1 Functional requirements

1. Log Collection & Normalization
   - Collect real-time logs from ground stations, network devices, and communication channels.
   - Normalize logs into a common schema for ML processing.
2. Threat Intelligence (TI) Integration
   - Ingest global TI feeds (IP, domain, malware indicators).
   - Correlate TI with log data for enhanced detection.
3. Machine Learning (ML)-Based Anomaly Detection
   - Apply supervised ML to classify known attack patterns (jamming, spoofing, SQLi, etc.).
   - Use unsupervised ML for zero-day and unknown anomalies.
   - Trigger alerts with confidence scores.
4. Correlation & Alerting
   - Combine ML detection + TI correlation into a unified risk score.
   - Generate alerts and forward them to SOC dashboards / playbooks.
5. Collaboration with Colleagues' Components
   - Provide security insights to cryptography optimization (colleague's part).
   - Share anomaly feedback loop to data pipeline and validation components.

4.2 Non-Functional Requirements

1. Performance – Must process logs with <2 sec latency.
2. Scalability – Handle >1M log events/day without performance degradation.
3. Reliability – 99.9% uptime for continuous monitoring.
4. Interoperability – Support multiple log sources (JSON, syslog, database logs).
5. Security – Logs encrypted at rest and in transit.
6. Maintainability – Easy retraining of ML models with new datasets.
7. Usability – Alerts must be interpretable by SOC analysts.

4.3 Expected Test Cases

Test Case ID	Scenario	Expected Outcome
TC-01	Inject known jamming signal pattern	ML detects anomaly with >95% accuracy
TC-02	Feed new unseen spoofing attempt	Unsupervised ML flags as suspicious
TC-03	TI feed updates with new malicious IP	System blocks/alerts within 1 min
TC-04	High-volume logs (>1M/day)	System continues without dropping logs
TC-05	Communication between your ML module & colleague's crypto optimizer	Alert triggers adaptive encryption change
TC-06	False positive test with benign logs	Alert rate <5% false positives
TC-07	Simulated attack with no TI indicators	Still detected by ML anomaly engine

Table 2: Expected Test Cases

5. GANTT CHART

Task	May	June	July	August	September	October	November	December	January	February	March	April	May	June
Brainstorming session														
Team formation														
Topic Finding														
Initiation														
Define project scope and objectives														
Submit Topic Assessment Form (TAF) and assessment														
Create project proposal draft														
Planning														
Requirement analysis														
Feasibility study for SME environment														
Collect datasets														
Deliver Proposal Presentation														
Finalize and submit proposal report														
Design														
Create system architecture diagram														
Design use case diagrams & test cases														
Implementation & Testing														
Develop Monitoring Module (ML)														
Implement Threat Intelligence Integration														
Initial Testing & Evaluation														
Refine ML Model & TI Intergration														
Advanced Testing														
API Testing														
Intergrate with other Modules														
Evaluation & Verification														
Deliver Progress Presentation I														
Submit checklist I														
Close-Out														
Deliver Progress Presentation II														
Prepare final report & research paper														
Final presentation & viva														
Project website & logbook submission														

6. Budget and budget justification

Item	Quantity	Unit Cost (USD)	Total Cost (USD)	Justification
Cloud / Compute Resources (Google Colab Pro / AWS Free Tier upgrade)	8 months	$15 / month	$120	Required for ML model training and handling large datasets (CERT logs, TI feeds). Free tiers are insufficient for continuous model training.
Threat Intelligence Feeds (Open-Source Integration)	–	Free	$0	Using open-source feeds such as AlienVault OTX, AbuseIPDB, and MISP for enrichment. No licensing cost.
Data Storage (Google Drive / OneDrive student subscription)	200 GB	Included / $20	$20	Storage for large log datasets and model checkpoints.
Development Tools (VS Code, Python, Jupyter)	–	Free	$0	Open-source IDE and Python libraries (TensorFlow, Scikit-learn, Pandas).
API Services (Optional for TI enrichment, e.g., VirusTotal API keys)	2 keys	$50	$100	For querying file hashes, domains, and IPs to enhance detection confidence.
Testing Infrastructure (Virtual Machines for Red-Team Simulation)	2 VMs	$20 each	$40	Needed to simulate real-world attacks (insider threat, malware injection).
Documentation & Reporting (Printing, Binding, etc.)	–	$5	$5	Preparation of final report, printing diagrams, and submission copies.

Table 3: Overall Budget of the Component

7. References

[1] P. Institute, "2024 Cost of Insider Threats Global Report," Ponemon Institute, 2024.

[2] M. H. P. H. Anas Ali, "Real-Time Detection of Insider Threats Using Behavioral Analytics and Deep Evidential Clustering," 2025. [Online]. Available: https://arxiv.org/pdf/2505.15383.

[3] G. &. T. I. Russo, "Behavioral Analytics for Insider Threat Detection in Enterprise Environments," *ACE Journal*, 2025.

[4] T. Michael, "Tolumichael, Best Open Source Threat Intelligence Platforms and Feeds," 2024. [Online]. Available: https://tolumichael.com/best-open-source-threat-intelligence-platforms-and-feeds/.

[5] SocRadar, "The Ultimate List of Free and Open-Source Threat Intelligence Feeds," 2025. [Online]. Available: https://socradar.io/the-ultimate-list-of-free-and-open-source-threat-intelligence-feeds/.

[6] ENISA, "Cybersecurity for SMEs – Good practices and recommendations," 2024.

[7] Verizon, "2024 Data Breach Investigations Report," 2024.

[8] IBM, "Cost of a Data Breach Report 2025," 2025.

[9] M. G. F. S. M. Bishop., "Insider Threats in Cyber Security," 2010.

[10] SEI-CMU, "CERT Insider Threat Dataset (v4–r2)," 2016.

[11] N. A. a. K. El-Khatib, On the effectiveness of machine learning in insider threat detection, vol. 5, IEEE Access, 2017.

[12] A. Wagner, "Security Information and Event Management: A Review," 2021.

[13] SentinelOne, "EDR vs. NDR vs. XDR: Choosing a Detection & Response Strategy," 2025.

[14] Corelight, "Endpoint Detection and Response (EDR): 2025 Guide," 2025.

[15] F. T. Liu, "Isolation Forest," 2008.

[16] I. Goodfellow, "Representation Learning with Autoencoders," 2016.

[17] W. Du, "DeepLog: Anomaly Detection and Diagnosis from System Logs," ACM CCS, 2017.

[18] R. C. a. S. Chawla, "Deep Learning for Anomaly Detection: A Survey," ACM Computing Surveys, 2019.

[19] S. L. a. S.-I. Lee, "A Unified Approach to Interpreting Model Predictions," NeurIPS, 2017.

[20] CISA, "STIX Best Practices Guide v1.0.0," 2022.

[21] CIRCL.lu, "MISP User Guide," 2024.

[22] OASIS, "TAXII 2.1 Specification," 2021.

[23] B. S. a. J. Kelsey, "Secure Audit Logs to Support Computer Forensics," 1998.

[24] M. R. e. al., "Why Should I Trust You? Explaining Predictions of Any Classifier," KDD, 2016.

[25] Microsoft, "Azure Machine Learning Interpretability," Microsoft, [Online]. Available: https://learn.microsoft.com/en-us/azure/machine-learning/how-to-machine-learning-interpretability?view=azureml-api-2.

[26] IBM, "QRadar Threat Intelligence App," IBM, [Online]. Available: https://www.ibm.com/docs/en/qradar-common?topic=apps-qradar-threat-intelligence-app.

[27] Splunk, "Enterprise Security Deployment Architectures," [Online]. Available: https://docs.splunk.com/Documentation/ES/8.1.0/Install/DeploymentArchitectures.

[28] Elastic, "Detecting Covert Data Exfiltration," [Online]. Available: https://www.elastic.co/blog/elastic-security-detecting-covert-data-exfiltration.

[29] Exabeam, "Understanding Insider Threat Detection Tools," [Online]. Available: https://www.exabeam.com/blog/ueba/understanding-insider-threat-detection-tools/.

[30] Elastic, "Open ML Jobs for Security Analytics," [Online]. Available: https://www.elastic.co/guide/en/elasticsearch/reference/current/ml-jobs.html.

[31] T. Anderson, "Emerging Cybersecurity Threats in Small Businesses: 2024 Insights," SecureTech Publications..

[32] N. &. S. R. Patel, "Economic Impact of Cybercrime: A Global Perspective.," Journal of Digital Security, 19(1), 72-89..

[33] K. Rajan, "Digital Adoption and Cybersecurity Gaps in South Asia’s SME Sector," South Asian Economic Review, 10(3), 55-70.

[34] L. Nguyen, "Mitigating Insider Threats: Strategies for the Modern Enterprise," Cyber Defense Quarterly, 14(2), 98-115.

[35] E. Carter, "Future Trends in the Global Cybersecurity Market: 2025-2030 Forecast," Market Research Insights.

[36] M. Silva, "SME Contributions and Cybersecurity Challenges in Sri Lanka," Regional Business Studies, 8(1), 40-58..

[37] Consultancy.uk, "SMEs to spend $90 billion on cyber-security in 2025," [Online]. Available: https://www.consultancy.uk/news/28470/smes-to-spend-90-billion-on-cyber-security-in-2025.

[38] market.us, "Global Cyber Security Market Size, Share, Trend and Industry Analysis Report By Component (Solutions, Services), By Deployment (On-premises, Cloud), By Security Type (Data Security, Cloud Security, Network Security, Application Security, Others), By Enter," 2024. [Online]. Available: https://market.us/report/cyber-security-market/.

[39] stationx, "Insider Threat Statistics: (2025's Most Shocking Trends)," [Online]. Available: https://www.stationx.net/insider-threat-statistics/.

[40] E. Carter, "Future Trends in the Global Cybersecurity Market: 2025-2030 Forecast," Market Research Insights..

8. APPENDICES

1. Plagiarism report