Modernizing the CERT Insider Threat Framework for the Agentic Enterprise

Executive Summary

Insider risk now extends beyond human users. As organizations deploy autonomous and semi-autonomous AI agents, the insider threat surface expands to include non-human entities with access to systems, data, and workflows. Traditional insider threat programs focused on employees, contractors, privileged administrators, and compromised accounts. These risks remain critical, but organizations must now account for autonomous systems operating at machine speed.

The Carnegie Mellon University Software Engineering Institute (SEI) CERT Insider Threat Center established a principle that still holds: Insider threats are behavioral problems. Effective detection depends on visibility into activity patterns, contextual relationships, and how behavior develops across systems.

To modernize detection, organizations must extend this behavioral approach through:

- Behavioral analytics
- Long-duration activity correlation
- Human and non-human identity monitoring
- Risk-based prioritization
- AI-assisted investigation workflows

Traditional detection models rely on static rules and short observation windows. They often miss insider risk when activity appears legitimate. Behavioral analytics addresses this limitation by connecting activity across identities, systems, and execution paths. In the agentic enterprise, this means applying proven insider threat guidance to environments where humans and autonomous systems operate as part of the same workflow.

Introduction

Insider threats are difficult to detect and investigate because they rarely begin as a single obvious event. They emerge through patterns of behavior, privilege misuse, policy violations, and subtle shifts in activity. Traditional controls identify known indicators but often miss early signals that precede an incident.

This paper outlines a framework for modernizing insider threat detection and insider risk programs. It builds on CERT research and adapts those practices to modern security operations. It explains how to:

- Identify insider risk earlier
- Prioritize investigations
- Align stakeholders across security, IT, HR, legal, and compliance
- Build a program that improves as behavioral understanding evolves

Behavioral analytics connects identity, endpoint, network, cloud, and application data to provide a unified view of activity and intent. The goal is to reduce investigation time, improve prioritization, and detect insider risk before impact. As environments become more autonomous, this requires extending detection beyond human-only monitoring models.

Why CERT Remains Foundational to Insider Threat Programs

For more than two decades, Carnegie Mellon's SEI and CERT Division have shaped how organizations understand insider threats. CERT research is not a vendor framework; it is grounded in analysis of real insider incidents across government, critical infrastructure, and enterprise environments.

This research established foundational practices that remain central to effective insider threat programs:

- Monitor behavioral and technical indicators
- Correlate activity across systems
- Evaluate insider risk accumulation
- Identify concerning behavior before impact
- Treat trusted entities as potential risk sources

The CERT Common Sense Guide to Mitigating Insider Threats reinforces the importance of monitoring and correlating behavioral and technical indicators across systems to identify insider risk. This guidance continues to shape modern insider threat programs, especially as organizations expand trust boundaries to include AI agents, machine identities, and autonomous workflows. Together, they informed the evolution of behavioral analytics and remain highly relevant as organizations extend insider threat detection to include non-human entities.

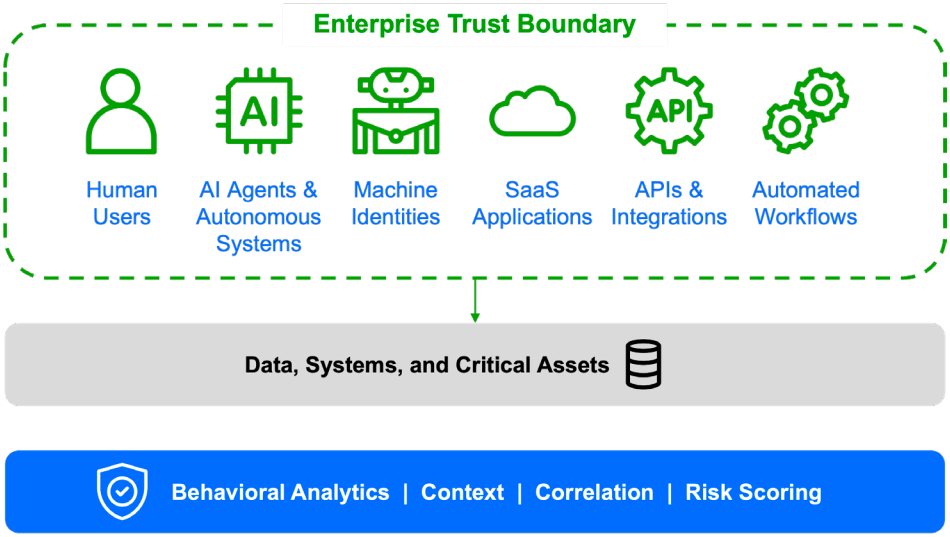

Figure 1.

Modern environments extend insider risk beyond users to include AI agents, machine identities, and connected workflows.

The Insider Threat Surface Has Expanded

The insider threat environment has changed. Programs must account for environments where humans, agents, and automated systems operate together.

Historically, insider threats focused on:

- Malicious employees
- Negligent users
- Privilege misuse
- Credential compromise
- Third-party contractors

Today, insider risk programs must also address:

- Autonomous and semi-autonomous AI agents
- Machine identities
- Cross-agent orchestration
- AI-driven automation pipelines
- Non-human entities operating with persistent access

These entities can introduce insider risk without malicious intent. Excessive permissions, unexpected data access, abnormal workflow sequences, and behavioral drift can create exposure without triggering detections. Most detection systems still focus on discrete events such as:

Known indicators of compromise (IoCs)

- Static rule matches
- Signature-based detections
- Point-in-time anomalies
- Policy violations

Insider threats often develop through sequences of activity that accumulate gradually, creating a visibility gap in environments where users and automated systems interact continuously. That gap becomes more significant as organizations grant trusted automation direct access to systems, APIs, and workflows.

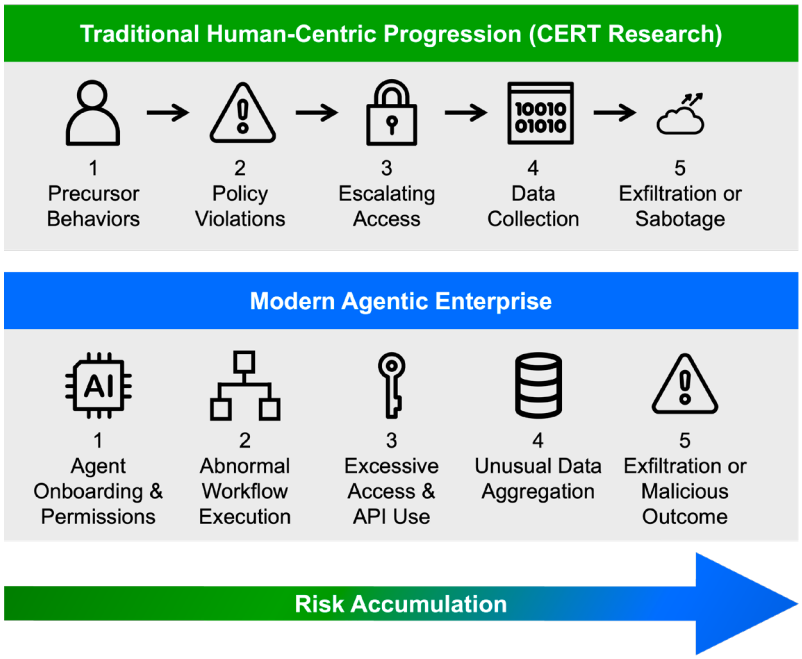

Figure 2.

Behavioral indicators accumulate across identities and systems, revealing risk through progression rather than isolated events.

Insider Threats Are Behavioral Problems

CERT established that insider threats cannot be understood through isolated events. Insider risk develops through behavioral patterns. In practice, this means:

- Behavioral indicators carry more signal than alerts
- Risk accumulates through behavioral progression
- Context reveals intent
- Trusted entities can introduce risk
- Activity often blends into legitimate operations

These ideas shaped user and entity behavior analytics (UEBA). Behavioral analytics establishes baselines and identifies deviations in:

- Access behavior
- Authentication activity
- Data usage
- Privilege behavior
- Workflow progression
- Cross-system relationships
- Temporal activity patterns

In agent-driven environments, activity may appear normal at the event level. Analysis reveals:

Abnormal access escalation

- Unusual data activity
- Excessive API usage
- Unexpected workflow chaining
- Cross-platform behavioral drift
- Lateral movement
- Unbounded AI consumption behavior

These are insider risk detection challenges that require behavioral context.

Why Traditional Detection Models Miss Insider Risk

Traditional security monitoring architectures were built around deterministic detection logic. Correlation rules, signatures, and static thresholds remain effective for identifying known threats, policy violations, and high-confidence malicious activity. Those methods still play an important role in security operations.

However, they become less effective when insider risk develops through behavior that appears legitimate in isolation. This challenge grows in environments that include AI agents, machine identities, and automated workflows. The issue is often not a single malicious event but a sequence of actions, relationships, and deviations that accumulate risk.

Several limitations reduce the effectiveness of traditional detection models in insider threat scenarios. Many rely on short observation windows, which work well for fast-moving attacks but miss behavior that develops over days, weeks, or longer.

Many assume stable and predictable user behavior, even though AI agents may act dynamically based on prompts, delegated tasks, integrations, orchestration logic, or changing workflow conditions. Many also focus on known bad activity instead of evolving behavioral context.

As a result, several insider threat scenarios may generate limited visibility:

- Excessive AI-driven data access
- Prompt manipulation attempts
- Abnormal tool invocation sequences
- Autonomous workflow abuse
- Gradual privilege escalation
- Cross-system behavioral anomalies
- Excessive resource consumption
- Coordinated human-agent misuse

In these scenarios, individual events may still appear valid. Credentials may be legitimate. Workflows may be approved. System actions may not violate a rule or match a known IoC. The risk develops through progression, sequencing, relationships, and deviations from established behavioral norms.

That is why behavioral analytics is essential for insider threat detection. Instead of focusing only on isolated detections, behavioral analytics evaluates:

- Activity sequences
- Long-duration patterns
- Entity relationships
- Risk accumulation
- Baseline deviations
- Behavioral context across systems

These capabilities align closely with the operational principles described in CERT insider threat guidance. They also become more important as organizations extend trust to autonomous and semi-autonomous systems operating inside enterprise environments.

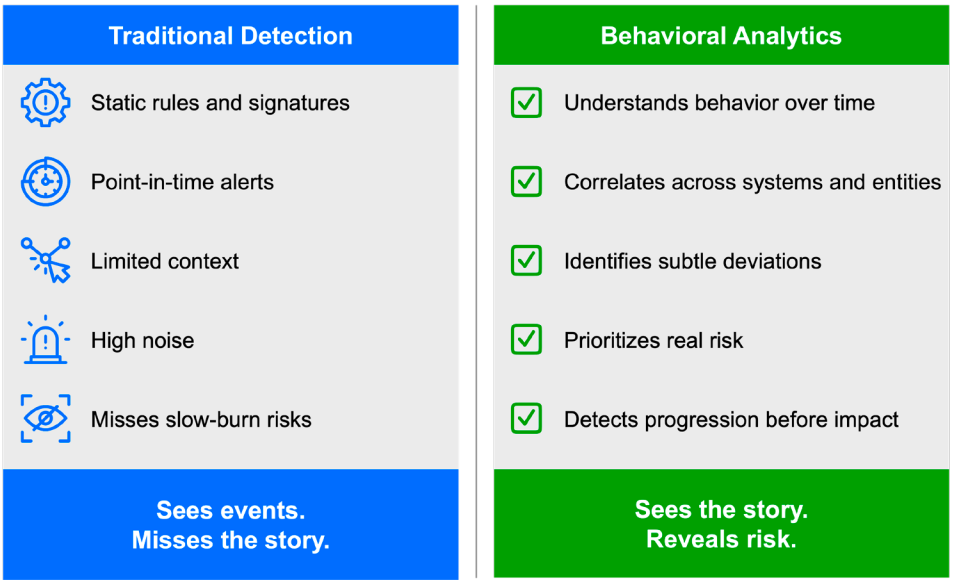

Figure 3.

Point-in-time alerts capture isolated activity but miss context, relationships, and progression. Behavioral analytics connects activity over time to surface subtle deviations and prioritize real insider risk.

Extending the CERT Framework Into the Agentic Enterprise

The operational environment has changed, but CERT guidance remains applicable. The task is to extend them to environments shaped by autonomous systems, machine identities, and AI-driven workflows. This requires visibility beyond human users.

CERT guidance emphasizes:

- Monitor behavioral and technical indicators across systems
- Identify elevated risk
- Correlate contextual activity
- Prioritize anomalies
- Enable investigation and response
- Continuously improve visibility

These priorities map directly to modern behavioral analytics. Exabeam applies them through UEBA, Agent Behavior Analytics (ABA), stateful session reconstruction, risk-based prioritization, cross-domain normalization, and AI-assisted investigation workflows. This shift is not simply more correlation; it is extending behavioral monitoring to non-human entities and embedding that context directly into investigation workflows.

A single API call may appear benign. When viewed as a sequence, activity may reveal:

- Abnormal authentication patterns
- Access to previously unseen repositories
- Excessive data retrieval
- Unusual tool invocation chains
- Cross-system privilege escalation
- Interaction with external services
- Significant deviation from historical operational baselines

Viewed individually, these actions may not trigger alerts. Viewed behaviorally, they reveal insider risk. This approach aligns closely with the CERT view that insider threats appear through accumulation and progression rather than isolated malicious events.

Human and Agent Teaming Changes Insider Risk

Humans and agents now work together in enterprise workflows, changing how insider risk must be evaluated. Traditional models focused on human intent. Modern environments require analysis of how agents interact with systems, data, APIs, and workflows.

Agents can:

- Retrieve data
- Execute tasks
- Trigger workflows
- Interact externally
- Access privileged resources
- Coordinate with other systems

Organizations must account for:

- Human-directed agent activity
- Agent-assisted privilege misuse
- Workflow drift
- Cross-agent anomalies
- Non-human entities with inherited trust
- Dynamic human-agent interaction

Many actions appear valid at the event level. Risk emerges through sequencing, interaction, and deviation from expected behavior. For example, a user may instruct an agent to access repositories or invoke workflows that do not align with that user's historical profile. An autonomous system may begin interacting with unfamiliar services, applications, or datasets in ways that gradually diverge from established baselines. In many cases, the challenge is identifying subtle behavioral progression inside trusted operational boundaries.

Behavioral analytics helps address that challenge by making it possible to evaluate:

- Long-term activity progression,
- Human-agent interaction patterns,
- Baseline deviations,
- Cross-system relationships,
- Abnormal operational sequencing,
- Risk accumulation across identities and systems

That gives security operations teams a stronger way to identify insider risk in environments where trusted human and non-human entities operate together.

Building a Modern Insider Threat Program

A modern insider threat program must account for environments that include human users, AI agents, machine identities, autonomous workflows, cross-platform automation, and distributed SaaS ecosystems. These entities now operate inside shared trust boundaries, often through delegated permissions, APIs, orchestration frameworks, and automated decision making.

Because of that shift, insider risk detection cannot rely solely on static indicators, isolated alerts, or human-only monitoring models. Organizations need visibility that helps them understand:

- Behavioral progression
- Relationships between entities
- Cross-domain operational context
- Human and non-human activity patterns
- Risk accumulation across systems
- Dynamic deviations from expected behavior

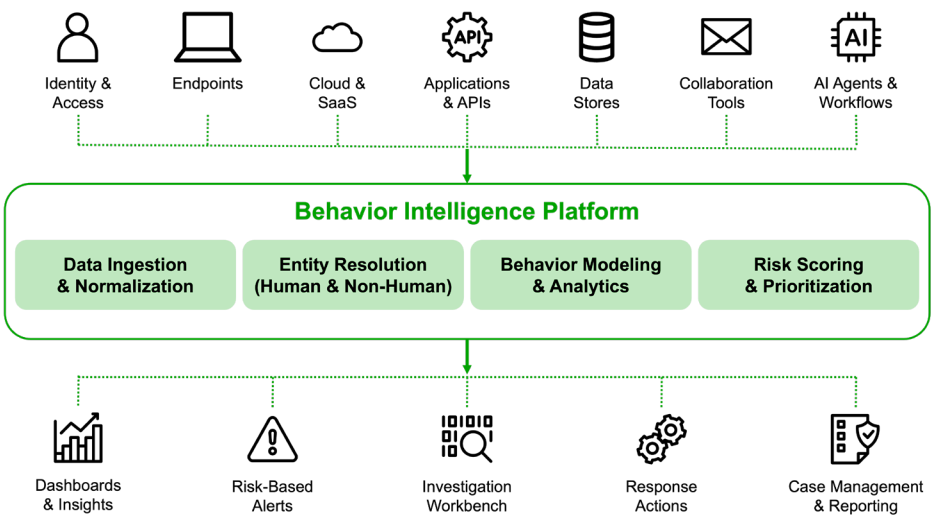

Figure 4.

A unified analytics framework ingests, correlates, and models activity from identities, devices, applications, and AI agents to detect insider risk through behavioral context, relationships, and risk scoring.

CERT insider threat research established the importance of behavioral monitoring well before the rise of autonomous systems. Those principles still apply. What has changed is the scale, complexity, and autonomy of the operational environment. Organizations that can extend proven insider threat practices into environments shaped by autonomous systems and machine-driven activity will be better prepared to manage modern insider risk.

That requires behavioral visibility, long-duration analytics, identity-aware monitoring, contextual risk scoring, AI-assisted investigation workflows, and unified operational visibility for human and non-human entities. Exabeam provides the Behavior Intelligence foundation for this approach at enterprise scale. By combining UEBA, ABA, stateful behavioral reconstruction, and AI-driven investigation workflows, organizations can modernize insider threat programs for the agentic enterprise while staying grounded in CERT guidance.

Operational Scenario: Insider Risk in the Agentic Enterprise

Consider an employee using an approved AI workflow assistant integrated with Slack, Google Drive, Salesforce, and Jira. The agent operates using delegated permissions tied to the employee's identity and is authorized to retrieve information, summarize records, and automate operational tasks across multiple enterprise systems.

Over several weeks, the AI assistant begins accessing repositories and datasets outside the user's normal behavioral baseline. API invocation frequency increases significantly, after-hours activity is revealed, and the system starts retrieving sensitive records across multiple departments that historically showed no relationship to the user's operational role.

Viewed independently, none of these activities appear inherently malicious. Each request uses valid credentials, approved integrations, and authorized workflows. Traditional rule-based detections may generate little or no visibility because the individual events do not clearly violate policy or match known indicators of compromise.

However, behavioral analysis reveals a broader progression pattern involving elevated data access, cross-system behavioral drift, unusual workflow sequencing, and expanding operational scope over time. In this scenario, the risk appears not through a single malicious event, but through the accumulation of subtle behavioral deviations across systems.

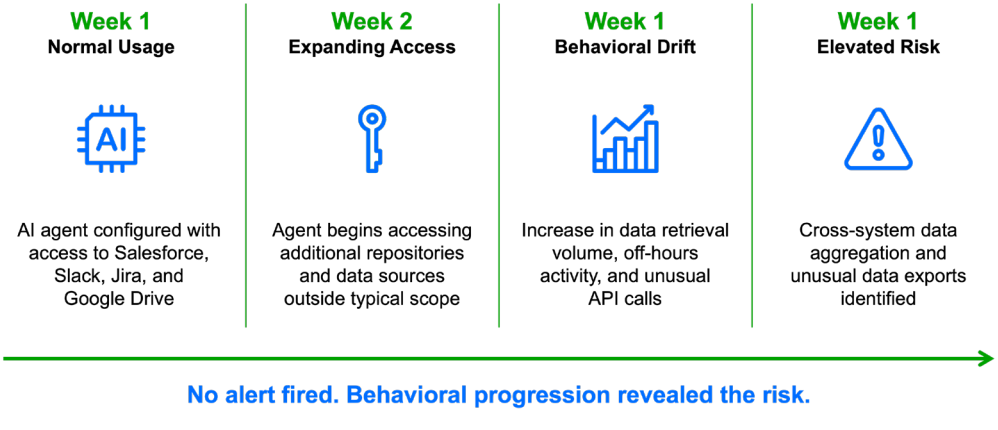

Figure 5.

Activity that appears legitimate evolves over time into cross-system risk that traditional detections miss.

Conclusion

Insider risk has not been replaced by AI-driven systems. It has expanded into environments where human users, AI agents, machine identities, and automate workflows operate within shared trust boundaries.

In these environments, activity often relies on valid credentials, approved integrations, and authorized workflows. Event-based detection alone cannot fully capture the risks that develop within these trusted interactions.

CERT research identified a foundational truth: Insider threats are behavioral problems. Detecting them requires understanding patterns, relationships, and how activity deviates from established norms.

The challenge today is scale and complexity.

Autonomous systems generate more activity, interact across more systems, and operate without human pacing. That increases the importance of detection models that can evaluate behavior, not just events.

Behavior Intelligence provides the structure needed to address this shift. By combining long-duration analysis, contextual correlation, identity-aware monitoring, and risk-based prioritization, organizations can detect insider risk that would otherwise remain hidden in legitimate operations.

Organizations that apply this guidance will move beyond identifying isolated events and understand how activity evolves, how risk develops, and where investigations should focus first.

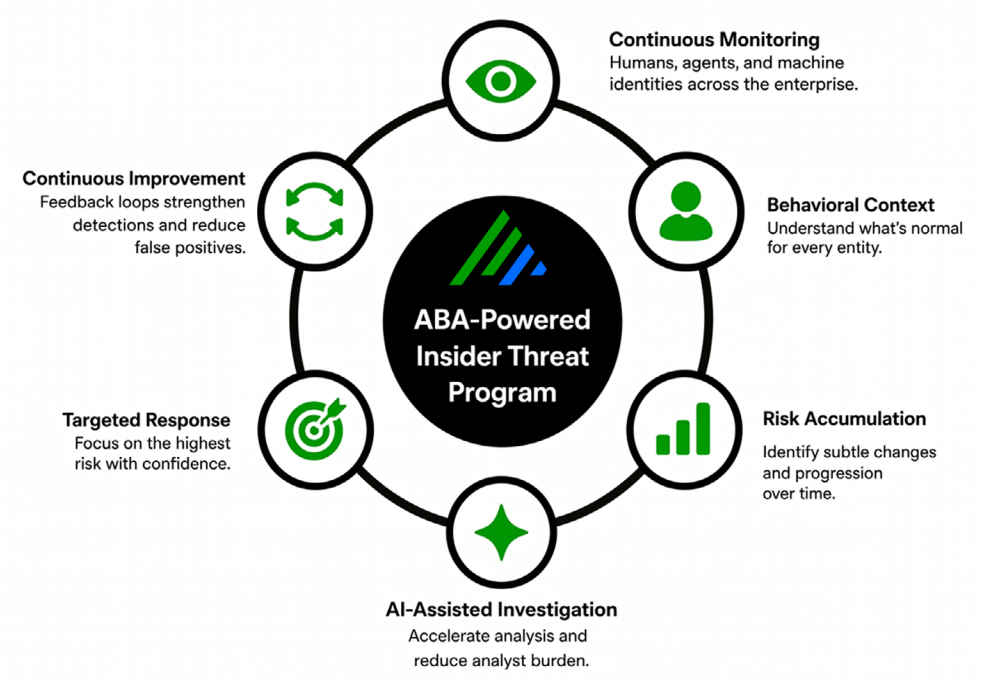

Figure 6.

Traditional detections identify known threats. Behavioral analysis uncovers patterns, relationships, and evolving risk.

About Exabeam

Exabeam is the leader in behavior intelligence for the agentic enterprise. As organizations deploy digital workers and confront machine-speed adversaries, Exabeam delivers flexible, industry-proven solutions for insider threat coverage of humans and agents and faster, more accurate threat detection, investigation, and response (TDIR). Learn more at www.exabeam.com.

exabeam™

Learn more at
www.exabeam.com →