Agent Behavior Analytics

Detect and investigate threats from your growing AI workforce

Agent Behavior Analytics (ABA) extends New-Scale Analytics to monitor the non-human agents operating in your environment. It introduces dedicated behavioral models that identify unsafe or unexpected agent actions and guardrail violations. By adding this context to investigation workflows, your team can detect misuse, compromised agents, and operational risk introduced by automated systems.

Detect Risky Agent Activity

ABA identifies high-risk agent behavior using new purpose-built models. Instead of profiling individual agents, it focuses on detecting the actions most likely to indicate misuse, compromise, or unauthorized activity. These models provide an early warning when an automated process begins acting outside approved patterns.

Models assess indicators such as:

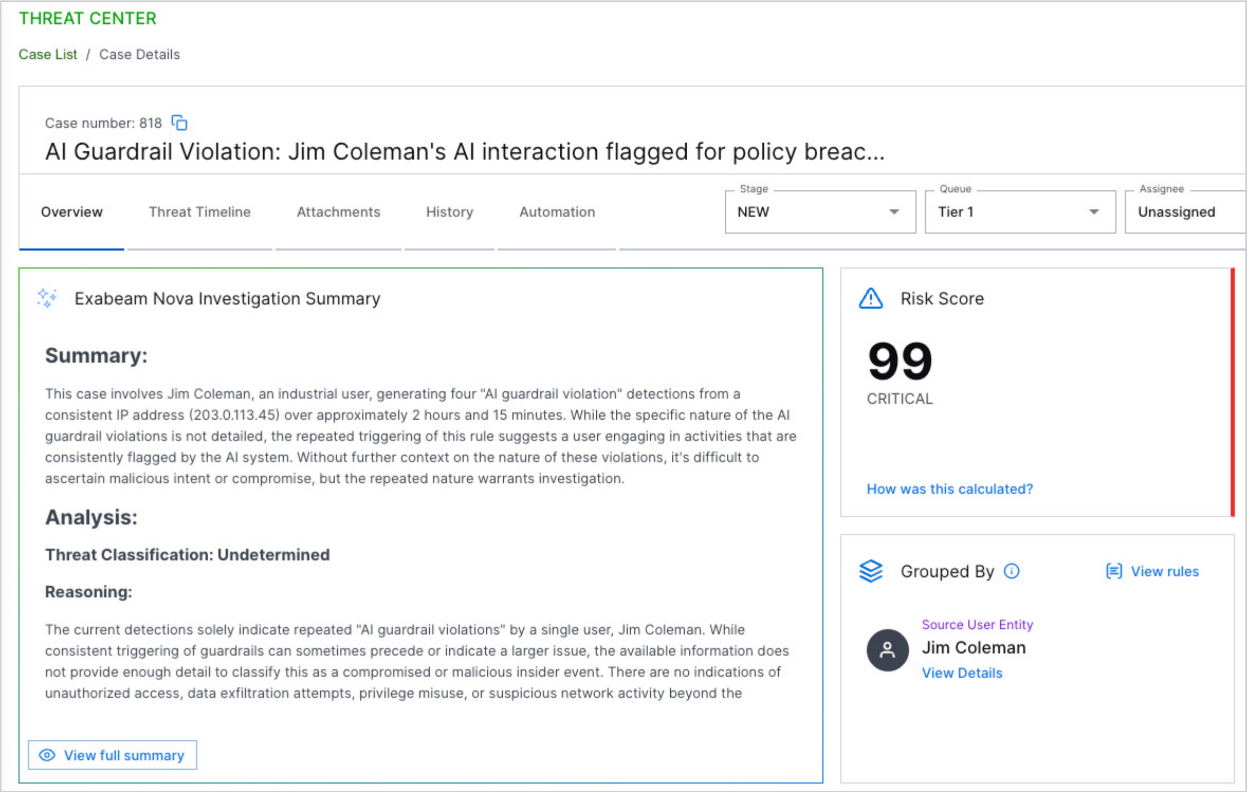
- First-time or unusual agent actions
- Violations of defined security or operational guardrails
- Behavior that suggests malicious prompts or data exfiltration
- Suspicious events observed in agent-specific log sources

Investigations Enriched With Agent Behavior Context

Agent-related threats often overlap with human activity. Exabeam enriches existing user timelines with agent detections rather than creating separate views. Analysts gain a complete picture of incidents involving both users and agents without switching views or reconstructing context.

This integrated view helps you:

- See when a user instructed an agent to take a risky action.
- Understand attacks involving compromised agents.
- Distinguish malicious agent use from normal automation.
- Apply existing investigation workflows to agent-driven threats.



THREAT CENTER

Case List / Case Details

Case number: 818

AI Guardrail Violation: Jim Coleman's AI interaction flagged for policy breac...

Stage: NEW

Queue: Tier 1

Assignee: Unassigned

Exabeam Nova Investigation Summary

Summary:

This case involves Jim Coleman, an industrial user, generating four "AI guardrail violation" detections from a consistent IP address (203.0.113.45) over approximately 2 hours and 15 minutes. While the specific nature of the AI guardrail violations is not detailed, the repeated triggering of this rule suggests a user engaging in activities that are consistently flagged by the AI system. Without further context on the nature of these violations, it's difficult to ascertain malicious intent or compromise, but the repeated nature warrants investigation.

Analysis:

Threat Classification: Undetermined

Reasoning:

The current detections solely indicate repeated "AI guardrail violations" by a single user, Jim Coleman. While consistent triggering of guardrails can sometimes precede or indicate a larger issue, the available information does not provide enough detail to classify this as a compromised or malicious insider event. There are no indications of unauthorized access, data exfiltration attempts, privilege misuse, or suspicious network activity beyond the

Risk Score

99

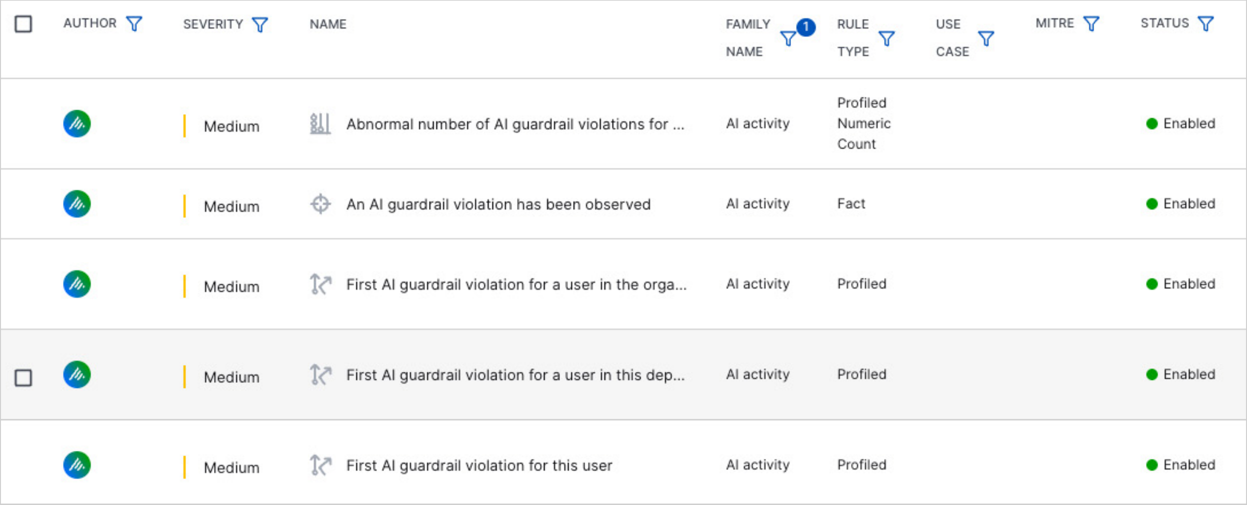
CRITICAL

Grouped By

Source User Entity

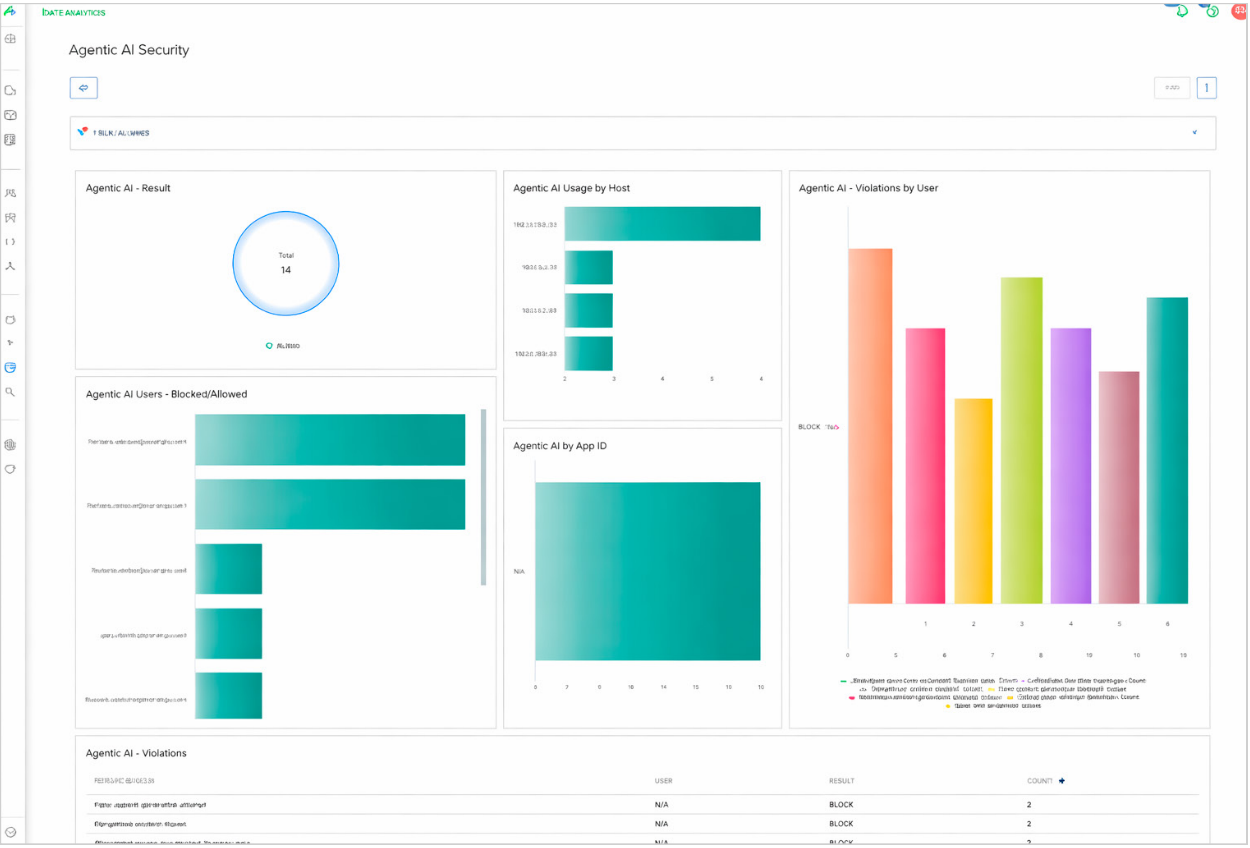
Jim Coleman

Figure 1. Behavioral activity of AI agents is integrated across threat detection, investigation, and response (TDIR) workflows within Threat Center.



☐	AUTHOR	SEVERITY	NAME	FAMILY NAME	RULE TYPE	USE CASE	MITRE	STATUS
		Medium	Abnormal number of AI guardrail violations for ...	AI activity	Profiled Numeric Count			● Enabled
		Medium	An AI guardrail violation has been observed	AI activity	Fact			● Enabled
		Medium	First AI guardrail violation for a user in the orga...	AI activity	Profiled			● Enabled
☐		Medium	First AI guardrail violation for a user in this dep...	AI activity	Profiled			● Enabled
		Medium	First AI guardrail violation for this user	AI activity	Profiled			● Enabled

Figure 2. Behavioral detections for AI Agents highlight abnormal activity that rule engines may miss.



Agentic AI - Violations

[illegible]	USER	RESULT	COUNT
[illegible]	N/A	BLOCK	2
[illegible]	N/A	BLOCK	2
[illegible]	N/A	BLOCK	2

Figure 3. Use natural language prompts to build dashboards and visuals for agent security.

Visualize and Report on AI Security

As organizations adopt more automated systems, demonstrating oversight of agent activity is becoming essential. ABA provides dashboards and reporting that track agent behavior trends and risk indicators, helping teams communicate security posture to leadership, auditors, and other stakeholders.

Visualizations and reports support the ability to:

- Present AI agent monitoring to executives and the board.
- Provide evidence of oversight for compliance.
- Justify investment in securing automated operations.
- Validate safe deployment of new AI tools

Key Capabilities

- Behavioral models: Five models designed to detect high-risk agent activity, including first-time actions and guardrail violations
- Agentic AI security use case: A dedicated Outcomes Navigator use case that helps you track coverage over time, measure readiness, benchmark your program against industry peers, and align with frameworks like MITRE ATT&CK®.
- Enriched Investigation Timelines: Agent detections appear directly in user timelines to provide unified context.
- Centralized visibility: A single view for monitoring agent activity across your environment
- Reporting and compliance: Executive-ready reports that demonstrate oversight of automated systems.

> “As AI adoption accelerates, one of our greatest priorities is understanding and managing agent behavior. The new connected capabilities from Exabeam provide the ability to see when an AI agent deviates from expected patterns, follow its activity through a unified investigation, and continuously improve our defenses with posture insights. This level of connected visibility and governance for AI agent activity is extremely valuable for ourselves and our end customers, and I look forward to seeing Exabeam continue to expand upon these capabilities”
>
> **– Michael Cole, Chief Information Security Officer, First Financial Bank**

A Strategic View for the Entire Security Team

- **For security leaders:** Provides assurance that automated systems are monitored and controlled, enabling clear, data-driven communication on organizational readiness.
- **For security architects and engineers:** Establishes a consistent approach for monitoring new agent capabilities and expanding automated workflows safely.
- **For security analysts:** Reduces investigation time by adding agent activity to timelines, helping analysts instantly determine whether behavior is expected or malicious.

About Exabeam

Exabeam is a leader in intelligence and automation that powers security operations for the world’s smartest companies. As a global cybersecurity innovator, Exabeam provides industry-proven, security-focused, and flexible solutions for faster, more accurate threat detection, investigation, and response (TDIR).

exabeam™

Learn more at www.exabeam.com →