LockThreat

Is Your Deployed AI Actually Being Monitored?

Runtime AI Assurance Self-Assessment for Deployed AI Systems and Agents

An AI agent does not wait to be asked. It monitors inputs, makes decisions, and takes actions. It does so continuously, at machine speed, without a human approving each step. Such agents are already in production today across financial services, healthcare, retail and e-commerce, government, and more.

Agents break the core assumptions of traditional GRC, which assumes that decisions are made by people. Agents act autonomously. GRC assumes periodic audits are sufficient, yet agents act continuously.

At least five governance gaps emerge in agentic systems: agent identity, authorization scope, behavioral drift, decision audit trails, and multi-agent trust. The last gap means one agent instructs another with no human in the chain. A sixth gap sits one layer down, in the AI applications underneath the agents: model drift, where a compliant system quietly becomes non-compliant as real-world data diverges from training data.

Most organizations are struggling to govern the use of AI tools by employees (layer 1), and are only partially aware of AI applications (layer 2). Almost no organization has extended governance to AI agents (layer 3). The self-assessment in this document tells you where the gaps are.

Seven Questions to Ask About Every Deployed AI System and Agent

01

Identity & Ownership

Does every production agent have a defined identity and a named, accountable owner?

02

Authorization Scope

Does every agent have a defined authorization scope, approved tools, and credential boundaries, with transaction limits and step-up approval for high-impact actions?

03

Behavioral Monitoring

Is there real-time behavioral monitoring that detects when an agent deviates from its authorized scope?

04

Drift Monitoring

Is there continuous drift monitoring on your AI systems, not just a point-in-time pre-deployment review, so you'd know if a compliant model quietly became non-compliant?

05

Decision Audit Trails

Do you retain audit trails capturing the inputs, actions, tool calls, approvals, and outputs for every agent decision, enough to reconstruct exactly what happened?

06

Revocation & Override

If an agent behaves unexpectedly, is there an immediate, defined way for a human to revoke its authority?

07

Multi-Agent Trust

In any environment where agents interact, are there defined trust rules, rather than one agent's instruction automatically becoming another agent's authority?



If you answered "no" or "not sure" to any of these, you have an ungoverned AI system in production. For most organizations deploying agentic AI today, that's the current state, not the exception.

AI System Risk Classification Worksheet

Use this worksheet to classify any AI system by risk level and determine the appropriate governance tier. Complete one worksheet per AI system. Each completed worksheet becomes part of your documented AI risk inventory.

> **Score each dimension 1 (Low), 2 (Medium), or 3 (High) for any AI system.**
> Add the scores. Match the total to the governance tier below.

DIMENSION	WHAT TO ASK	SCORE (1-3)
Data Sensitivity	Does this system process personal, confidential, or regulated data?	
Decision Autonomy	Does this system make or significantly influence decisions without human review?	
Regulatory Exposure	Does this system fall under the EU AI Act high-risk category, DORA, or sector-specific regulation?	
Deployment Breadth	How many people, customers, or processes does this system affect? (1 = Few; 2 = Moderate number; 3 = Large number or organization-wide)	
Human Oversight Availability	How easy is it for a human to understand, monitor, and intervene in this system's decisions? (1 = Easy; 2 = Requires effort; 3 = Difficult to impossible)	
Human Override Capability	If this system behaves unexpectedly, how quickly and completely can it be stopped or overridden? (1 = Immediately and fully; 2 = Possible but requires effort; 3 = Difficult or slow to stop)	

System: _____ **Date:** _____ **Total score:** _____

Important: The score is a starting point, not a substitute for regulatory classification. Legal or policy requirements override the numeric score. High-risk uses, including employment, credit, healthcare, biometrics, critical infrastructure, or autonomous decisions, may require mandatory escalation regardless of score.

TOTAL SCORE	6 to 9	10 to 14	15 to 18
GOVERNANCE TIER	**Tier 1:** Standard	**Tier 2:** Enhanced	**Tier 3:** Critical

LockThreat

Understanding the Tiers

Tier 1 systems

STANDARD

Have limited impact, autonomy, data sensitivity, and regulatory exposure. Governance may be lighter, but core requirements such as ownership, approved use, inventory, change control, incident escalation, and evidence retention still apply.

Tier 2 systems

ENHANCED

Have meaningful impact or regulatory exposure, so governance must be more active and more rigorous.

Tier 3 systems

CRITICAL

Have high impact and high regulatory exposure, significant autonomy, or access to sensitive data or critical operations. Their governance must be continuous, automated, and defensible to senior management, auditors, customers, and regulators.

The same core requirements apply across all three tiers: a named owner, pre-deployment checks, drift and behavioral monitoring, human oversight, human override capability, explainability for decisions affecting individuals where required, and audit-ready evidence. What changes between tiers is the standard to which each requirement is implemented.



This worksheet is a classification tool, not a complete governance requirements guide. For implementation guidance specific to your organization, consult qualified legal and GRC counsel.

About LockThreat

LockThreat is the enterprise GRC platform for the era of AI, three layers within one platform: Enterprise GRC, Cyber Compliance, and AI Governance and Security. This is GRC 3.0, a platform that combines a System of Record with a System of Defense for AI. Full governance, risk, and compliance across your entire organization, not just IT and cyber. Deployed as SaaS, private VPC, or on-premises to meet data residency and sovereignty requirements. Serving enterprises in North America and the Middle East, including regulated industries and government.