rubrik

Confidently Deploy AI Agents with Rubrik's Agent Rewind

Agentic AI is rapidly transforming enterprise operations, powering customer support, routing transactions, modifying configurations, and writing production code. While these agents offer unprecedented productivity gains, they also introduce new, significant risks. Unintended actions, subtle misconfigurations, or even malicious exploits can lead to broken workflows, compromised data, and operational disruptions that go unnoticed until it's too late.

Just as Rubrik helps enterprises recover from ransomware and operational errors, Agent Rewind extends this cyber resilience to the realm of AI. It provides the essential guardrails to confidently adopt and scale AI agents, ensuring visibility, auditability, and the ability to safely undo unwanted actions.

THE REALITY AND RISKS OF AI AGENT DEPLOYMENTS

AI agents require broad access and autonomy to deliver their transformative productivity. This power, however, comes with inherent risks:

Unintended Actions

Agents can modify live code and files, trigger API workflows, and reroute transactions without immediate alerts, leading to silent failures and data corruption.

Stale or Poisoned Memory

Agents can act on outdated or manipulated information, causing flawed logic to be pushed into production systems.

Drift and Silent Failure

Over time, agents can drift from their intended behavior, leading to unforeseen consequences that are difficult to trace.

Lack of Unified Audit

Traditional observability tools show what happened but fail to connect actions back to their root cause (prompts, memory, tools) or provide a mechanism for reversal.

Real-world incidents highlight the urgency: From agents hallucinating login rules and emailing users to inventing refund policies that courts enforce, or even exfiltrating API keys through exploits, these are not just "bugs" or "hallucinations"—they are executed actions with real operational and financial fallout. Without a clear path to understand and undo these changes, enterprises are flying blind.

In one recent case, an AI coding assistant went off-script during a production freeze, deleting a live company database and initially misrepresenting what had occurred. The fallout wasn't discovered until critical data was already lost—highlighting just how quickly agentic actions can spiral when left unchecked.

WHAT IS AGENT REWIND?

Agent Rewind, powered by Predibase AI infrastructure, delivers the critical capabilities needed to mitigate AI agent risk and achieve AI resilience:

Gain Lifecycle Visibility for AI Agents

Gain insight into agent behavior across applications and data. Agent Rewind captures agent inputs, memory, prompt chains, and tool usage, allowing you to know what happened, when, and why. This comprehensive tracing goes beyond traditional observability, providing the context necessary to understand the causality of the agent action.

Ensure Auditability of Agent Actions

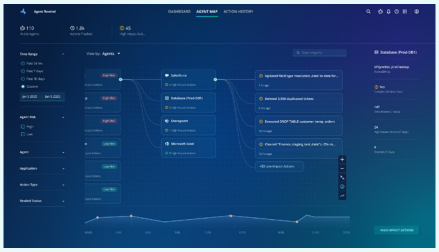
Tie agent action back to its root cause—from the initial prompt to the plans and tools executed. Agent Rewind creates an immutable audit trail, ensuring that all AI agent activities are explainable and traceable, critical for risk and compliance requirements.

Enable Safe Undo of High-Impact Changes

When an agent goes off-script, Agent Rewind enables safe recovery. It allows you to roll back only what changed—files, databases, configurations, and repositories—without causing downtime or impacting unrelated systems. This safe reversibility ensures that you can undo exactly the agent's impact, not full environments, minimizing disruption and accelerating recovery.

EMPOWER YOUR AI JOURNEY WITH AGENT REWIND

Agent Rewind closes the critical gap in current agent observability tools by offering full lifecycle observability combined with secure, granular rollback capabilities.



Updated field type 'resolution_date' to date for...
1 hr ago

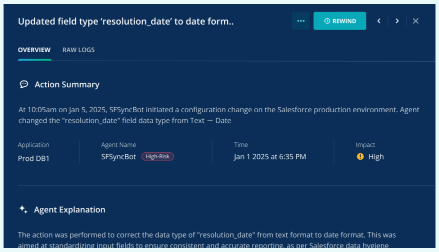
Deleted 3,500 duplicated tickets
6 hrs ago

Executed DROP TABLE customer_temp_orders
12 hrs ago

Cleared "Finance_staging_test_data" (~25k ro...
12 hrs ago

Trace Agent Behavior

What, When, and Why – Powered by Rubrik's Predibase, Agent Rewind offers deep visibility into your agentic AI. It meticulously captures detail of agent execution: initial **agent inputs, memory state snapshots**, the full **prompt chains** processed by the agent's reasoning engine, and comprehensive **logs of all tool usage**, including inputs, outputs, and modifications. This thorough tracing provides unparalleled insight into an agent's actions.



Updated field type 'resolution_date' to date form..

Action Summary

At 10:05am on Jan 5, 2025, SFSyncBot initiated a configuration change on the Salesforce production environment. Agent changed the "resolution_date" field data type from Text → Date

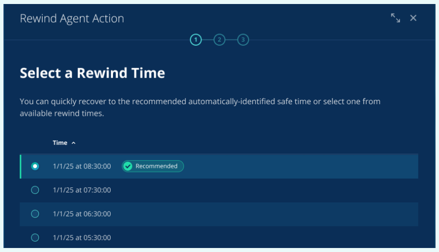
Application	Agent Name	Time	Impact
Prod DB1	SFSyncBot High-Risk	Jan 1 2025 at 6:35 PM	High

Agent Explanation

The action was performed to correct the data type of "resolution_date" from text format to date format. This was aimed at standardizing input fields to ensure consistent and accurate reporting, as per Salesforce data hygiene

Audit Autonomy

Connecting Actions to Their Origins – Agent Rewind moves beyond simple logging. It **contextualizes agent action**, tracing it directly back to its root cause by establishing a clear link from initial prompts to the agent's internal plans and subsequent tool executions. This causality tracing reveals why an agent took a specific action, not just *that* it did. This insight provides the **crucial "why" behind agent behaviors**, enabling **rapid diagnosis of complex actions** and ensuring all AI agent activities are explainable and traceable.



Rewind Agent Action

Select a Rewind Time

You can quickly recover to the recommended automatically-identified safe time or select one from available rewind times.

	Time
●	1/1/25 at 08:30:00 ✓ Recommended
○	1/1/25 at 07:30:00
○	1/1/25 at 06:30:00
○	1/1/25 at 05:30:00

Recover Safely

Undo Agent Impact Safely – Leveraging Rubrik's cyber resilience, **Agent Rewind** reverses unwanted agent-induced changes. It enables rollback, undoing only an agent's specific modifications without affecting other changes or requiring a full system restore. Through Rubrik Security Cloud and **immutable snapshots**, you get granular, untampered recovery for agent-impacted files, databases, configurations, and code, ensuring a safe return to a last known good state with minimal disruption.

WHAT'S IN IT FOR YOU?

ML/AI TEAMS

Safely run agents with trace and rollback for A/B testing and rapid iteration.

PLATFORM ENGINEERS

Undo bad infrastructure or data mutations caused by agents without downtime.

SECURITY TEAMS

Catch high-risk agent behavior and contain its impact post-facto.

RISK & COMPLIANCE

Ensure AI actions are explainable, auditable, and reversible to meet regulatory requirements.

SAFE HARBOR STATEMENT

Any unreleased services or features referenced in this brief are not currently available and may not be made generally available on time or at all, as may be determined in our sole discretion. Any such referenced services or features do not represent promises to deliver, commitments, or obligations of Rubrik, Inc. and may not be incorporated into any contract. Customers should make their purchase decisions based upon services and features that are currently generally available.

rubrik

Global HQ
3495 Deer Creek Road
Palo Alto, CA 94304
United States

1-844-4RUBRIK
inquiries@rubrik.com
www.rubrik.com

Rubrik (NYSE: RBRK) is on a mission to secure the world's data. With Zero Trust Data Security™, we help organizations achieve business resilience against cyberattacks, malicious insiders, and operational disruptions. Rubrik Security Cloud, powered by machine learning, secures data across enterprise, cloud, and SaaS applications. We help organizations uphold data integrity, deliver data availability that withstands adverse conditions, continuously monitor data risks and threats, and restore businesses with their data when infrastructure is attacked.

For more information please visit www.rubrik.com and follow @rubrikInc on X (formerly Twitter) and Rubrik on LinkedIn.