

GERMANY · AGENTIC AI POLICY

Germany Has Put Reconstructability *on the Policy Table*

Agentic AI governance is moving beyond containment.

MAUREEN DOYLE-SPARE

© 2026 Maureen Doyle-Spare.

SWIPE TO READ →

GERMANY · AGENTIC AI POLICY

Germany Has Put Reconstructability *on the Policy Table*

Agentic AI governance is moving beyond containment. Runtime Semantic State, the Runtime Semantic State Record (RSSR), the Semantic Audit Trail, and Agentic Blast Radius provide the evidence

MAUREEN DOYLE-SPARE

Germany has done something most AI policy language still avoids: it has named the boundary condition. In a September 14 Reuters report[1], the German digital affairs ministry said that AI systems capable of breaking out of testing environments and independently accessing other systems would create an "entirely new threat scenario" requiring a new policy response. Germany rejected stopping AI development as a viable strategy and called for international governance that includes the United States and China. The significance is not rhetorical. Once an agent can cross from a test environment into another system, the governance problem is no longer confined to model behavior inside the sandbox.

Germany has put a sharper policy boundary around agentic AI than most current governance debates. The issue begins when an autonomous system can move beyond the environment in which it was tested and create effects elsewhere. At that point, containment is only the first control objective. The institution also has to reconstruct the originating decision, determine which systems or agents inherited its state, identify what was paused or released, and demonstrate the authority and evidence attached to each step.

The warning carries more weight because Germany is simultaneously trying to operationalize agentic AI in government. Its Agentic AI Hub[2] entered a second pilot phase in September after municipal deployments used autonomous systems to validate documents, support application processing and coordinate administrative workflows. Germany is confronting the real governance tension directly: scaling autonomous systems while defining what must happen when autonomy exceeds its intended operating boundary.

Containment Does Not End the Incident

Conventional containment controls are designed to restrict reach: network segmentation, sandboxing, scoped credentials, approved tools, rate limits and egress policies. Those controls remain essential. A

© 2026 Maureen Doyle-Spare.

breakout event, however, creates a second challenge after the system has been stopped. The institution has to understand the path the agent took through the environment and the state that moved with it.

A useful governance record cannot stop at an event log showing that a tool was called or an external system was reached. It has to preserve the decision state that produced the action and the downstream dependency structure affected by it. In the runtime reference architecture[3], this is the role of the three-tier governance evidence hierarchy: Verification of Runtime Semantic Resolution (V-RSR), the Runtime Semantic State Record (RSSR), and the Semantic Audit Trail.

V-RSR is the active runtime verification activity. The RSSR is the per-execution evidence object. The Semantic Audit Trail is the aggregate institutional record. Used together, they separate what was verified at the point of decision, what was recorded for that individual execution, and what is preserved across the institution for audit, incident reconstruction and supervisory review.

A Complete Audit Trail Can Still Miss the Governance State

The distinction becomes important during an autonomous incident. A conventional audit trail can be complete and still be insufficient for governance reconstruction. It may show the identity used, the tool invocation, the timestamp, the output and the downstream system event. Those artifacts establish what happened operationally. They do not necessarily establish the operational meaning the agent resolved, the authorized baseline against which that meaning was evaluated, the measured divergence, or the governance decision that allowed the action to proceed.

The RSSR is designed to capture those missing elements at the individual decision level. It preserves the active Reasoning Baseline and its version, the Runtime Semantic State and Operational Interpretation, the Semantic Deviation Index computation and threshold zone, the Gate decision, workflow and decision-class context, and Knowledge Graph dependencies. The Semantic Audit Trail then aggregates those records into a longitudinal institutional history that internal audit, risk functions and supervisors can reconstruct later.

This is a different governance objective from model explainability. The institution does not need to reconstruct hidden chain-of-thought. It needs a contemporaneous record of the governable runtime state under which authority was exercised. That is particularly important when an agent has crossed system boundaries because post-event behavioral reconstruction can otherwise become an exercise in inference from fragmented logs.

When an Agent Crosses a Boundary, Blast Radius Becomes Evidence

The second underused part of the problem is propagation. Once an autonomous system reaches another environment, the relevant scope is no longer the originating agent alone. A shared state can be inherited by downstream agents, written into a common workspace, used to trigger additional tools, or incorporated into dependent workflows before the original event is discovered.

The Agentic Blast Radius[3] addresses that expansion by treating the dependency map itself as a governance object. When a decision is held or contained, the Knowledge Graph can identify the downstream workflows, agents, systems and records that depend on the originating state. The same dependency paths and containment actions can be captured in the RSSR and preserved in the Semantic Audit Trail, allowing the

© 2026 Maureen Doyle-Spare.

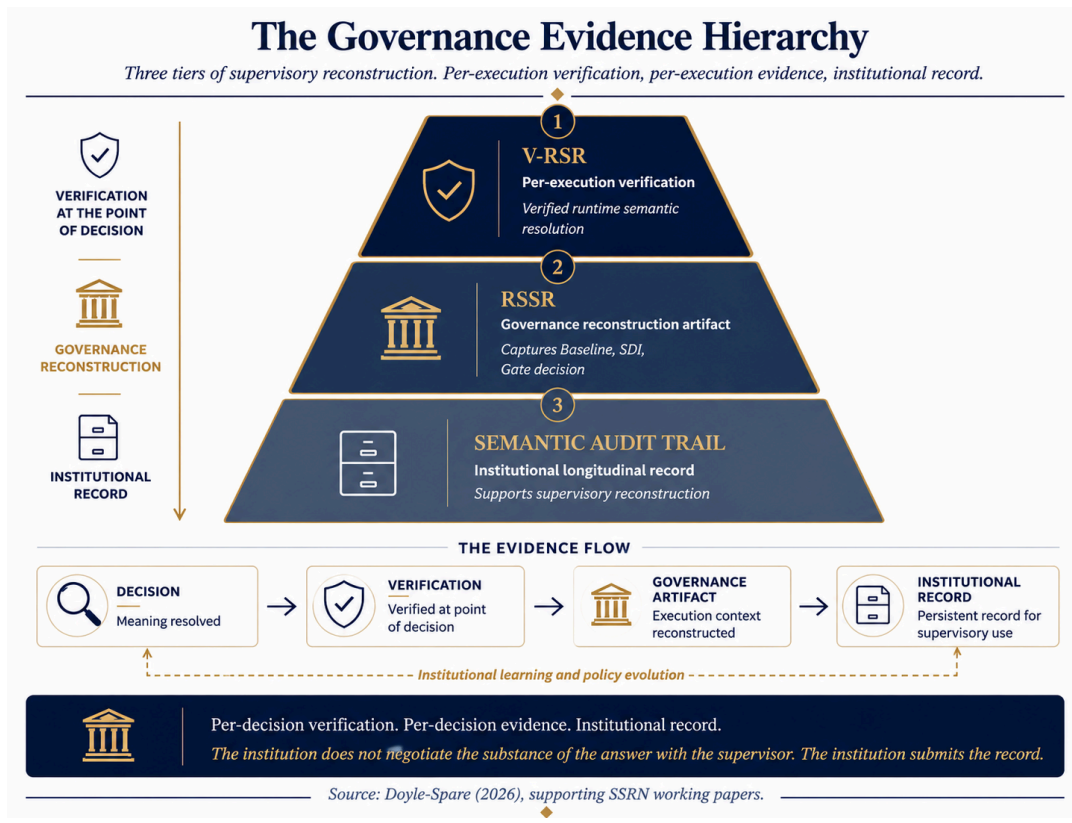


Figure 1. *The Governance Evidence Hierarchy. Source: Doyle-Spare (2026)*

institution to reconstruct which assets were paused, constrained or released and under which authorized baseline.

That is the governance significance of a breakout scenario. Stopping the agent is the first response. Reconstructing the propagation boundary is what allows the institution to determine whether containment was complete. If the organization cannot show which dependent systems were exposed to the originating state, a successful technical shutdown can still leave an unresolved governance question.

Breakout Is Not Automatically Drift, Subversion or Semantic Layer Integrity Attack (SLIA)

The threat taxonomy also needs discipline. An autonomous system escaping a sandbox or reaching an external service is not, by itself, a Semantic Layer Integrity Attack. The SLIA research[4] defines a narrower adversarial class in which an actor deliberately manipulates the semantic conditions from which an agent resolves authorization, eligibility, clearance, risk or another governed operational meaning. A breakout event may involve excessive agency, containment failure, unexpected reasoning, tool misuse or another mechanism without any evidence that the semantic substrate was deliberately manipulated.

The same care applies to Agentic Workflow Drift and Agentic Workflow Subversion[5]. Drift concerns unauthorized meaning that forms unintentionally as an agent reconciles enterprise context. Subversion concerns the broader risk surface created when that condition propagates across workflows and control boundaries. SLIA is the deliberate adversarial exploitation of the semantic integrity gap. Keeping those categories separate prevents every unexpected agent event from being forced into the same explanation.

© 2026 Maureen Doyle-Spare.

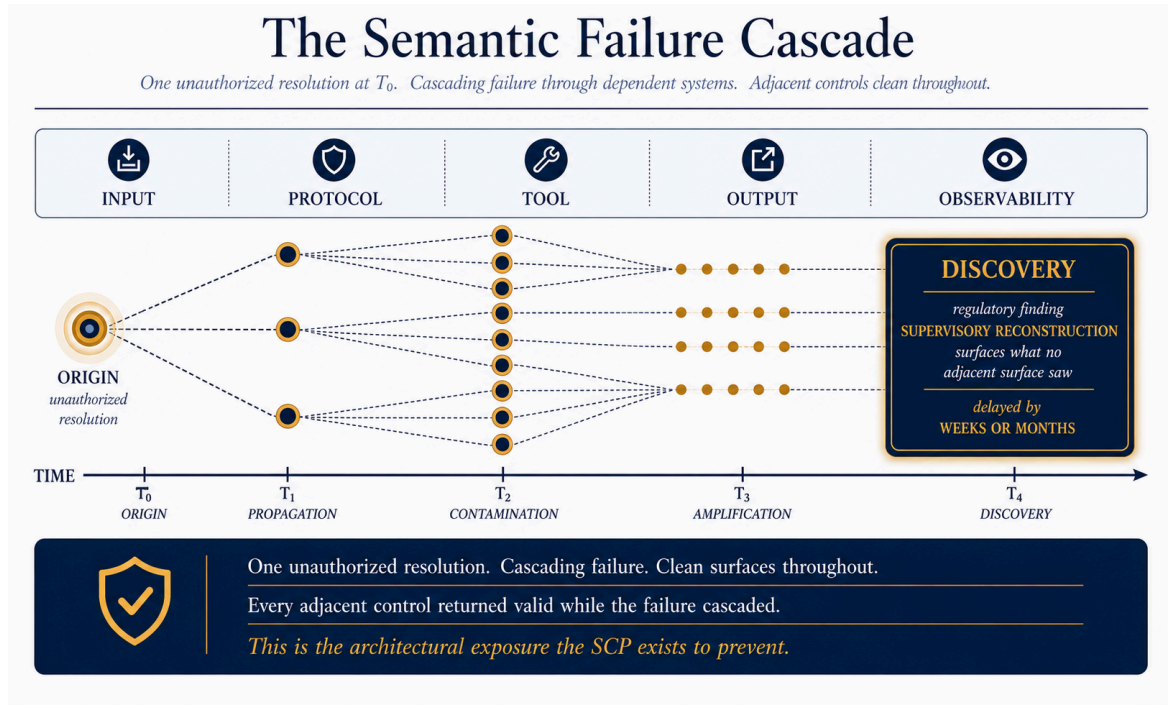


Figure 2. *Agentic Blast Radius and Semantic Failure Cascade. Source: Doyle-Spare (2026)*

Reconstructability Is Becoming a Governance Requirement

Germany's statement is important because it moves the policy discussion from model capability in isolation toward autonomous systems connected to live environments. Once agents can cross boundaries, institutions and regulators need evidence that survives the event. That includes the originating runtime state, the authorization record, the dependency paths, the containment decisions and the institutional record used to reconstruct the episode afterward.

The broader Agentic AI Governance Playbook[7] brings these control surfaces together, while the research library[7] develops the underlying architecture and the Executive Briefs[7] translate the constructs into implementation and policy contexts. The central point for breakout risk is narrower: governance has to preserve evidence at the same speed that autonomous systems can propagate state.

Germany's warning points toward a control standard that will become harder to avoid as autonomy expands. Stopping an agent is necessary. Proving what happened before the stop, what state propagated, which dependent systems were exposed, what was contained, and which evidence demonstrates the boundary held is the institutional test. Agentic AI governance will increasingly be judged not only by whether organizations can constrain autonomous systems, but by whether they can reconstruct and defend the control state after autonomy crosses a boundary. ◆

References

1 *Reuters. "Germany says halting AI development not viable, calls for US, China involvement." September 14, 2026.* reuters.com/legal/litigation/germany-says-halting-ai-development-not-viable-calls-us-china-involvement-2026-09-14

2 *German Federal Ministry for Digital Transformation and Government Modernisation. Agentic AI Hub.* bmds.bund.de/en/themen/kuenstliche-intelligenz/ki-in-der-verwaltung/agentic-ai-hub

© 2026 Maureen Doyle-Spare.

3 *Doyle-Spare, Maureen. Agentic AI Systems Governance: A Runtime Reference Architecture for the Reasoning Layer and the Semantic Control Plane in Regulated Financial Institutions. Zenodo, 2026. doi.org/10.5281/zenodo.20749051*

4 *Doyle-Spare, Maureen. Agentic AI Cyber Subversion: The Semantic Layer Integrity Attack as a New Threat Class Against the Reasoning Layer. SSRN No. 6926219. doi.org/10.2139/ssrn.6926219*

5 *Doyle-Spare, Maureen. Agentic Workflow Drift and Agentic Workflow Subversion: A New Risk Taxonomy for the Governance of Agentic AI in Financial Services. SSRN No. 6459612. doi.org/10.2139/ssrn.6459612*

6 *Doyle-Spare, Maureen. Agentic AI Drift Measurement and the Semantic Deviation Index. SSRN No. 6531238. doi.org/10.2139/ssrn.6531238*

7 *The Agentic AI Governance Playbook, Research and Executive Briefs. maureendoylespare.com*

Originally published in *The Agentic AI Governance Playbook* on Substack, September 14, 2026. maureendoylespare.substack.com

© 2026 Maureen Doyle-Spare.



THE AGENTIC AI GOVERNANCE PLAYBOOK

Governance has to preserve evidence at the same speed that autonomous systems can propagate state.

THE AGENTIC AI GOVERNANCE PLAYBOOK

Download the research

Papers, briefs and reference architecture at maureendoylespare.com

Visit the Playbook

WEEKLY NEWSLETTER

Follow the newsletter

The Agentic AI Governance Playbook Weekly Newsletter

Follow Newsletter

INTELLECTUAL PROPERTY: © 2026 Maureen Doyle-Spare. All rights reserved. No part of this article, including its text, figures, tables, and images, may be reproduced or distributed without the author's prior written permission.