PANTHEON CONTROL / SECURITY RESOURCES

Pantheon Control Buyer Overview

A proposed approach for governing AI agent access and actions

Prepared for security leaders and technical evaluators · October 2026

Pantheon Control is a proposed security control plane for helping teams connect an AI agent's identity and action route to the systems and data it can reach. The concept is to make those paths understandable, apply policy at a supported control point, and coordinate authorized containment and remediation through the organization's existing tools.

The product is not a replacement for identity, endpoint, cloud, application, data, or security operations platforms. Its value depends on integrations that provide trustworthy context and, for prevention, an enforcement point that can inspect an action before the destination executes it.

The problem the product is designed to address

An agent can use more than one route to act. It may call an MCP tool, an API function, a SaaS connector, a browser session, a shell or code runtime, an automated workflow, or another agent. Security teams need to understand which identity authorizes each operation, what target it affects, and whether a control can intervene before execution.

Existing controls remain essential. The product concept focuses on connecting evidence across those controls and preserving the difference between an observed event, a policy decision, an enforced block, a containment action, and a verified fix.

Who should evaluate it

- Security and risk leaders planning to allow agents to reach business systems, sensitive data, or production environments.
- Security operations teams that need to connect agent activity with identity, workload, endpoint, cloud, SaaS, and data context.
- IAM, platform, application, and AI teams responsible for agent owners, delegated access, tool configuration, and workflow permissions.
- Organizations that can name a bounded agent workflow and the control point where its actions could be observed or evaluated.

What the proposed workflow does

Stage	What the team needs to see	Product role proposed
Discover	Agents, owners, runtimes, identities, tools, permissions, and reachable resources.	Build an inventory with source, timestamp, confidence, and known coverage gaps.
Trace	The requested action, actual principal, route, target, and downstream result.	Normalize and correlate source events into an explainable action record.
Evaluate	Policy inputs such as tool, scope, data class, destination, environment, and approval.	Apply customer-defined deterministic rules; leave missing facts explicit.

Stage	What the team needs to see	Product role proposed
Enforce	Whether a compatible control intercepted the request before execution.	Return a deny or approval hold only through a tested, authorized on-path control.
Respond and verify	What was contained or changed and whether source systems confirm it.	Stage approved actions through existing tools and re-read the resulting state.

Coverage depends on the action route

Route	Possible pre-action control	Coverage limit to test
Remote MCP	Managed gateway, MCP server, or target authorization.	Calls that bypass the managed route are outside that gateway's view.
Local MCP over stdio	Agent host, endpoint, or managed runtime.	A network proxy may not see local standard-input/output messages.
API or function call	Application logic, API gateway, tool runtime, or target authorization.	A model request is not proof that the application executed the operation.
SaaS connector or browser	Connector scopes, identity/session policy, or application-side check.	Inherited human access and user-interface activity need separate review.
Shell, CI, workflow, or delegated agent	Endpoint/container, CI permissions, orchestrator, queue, or delegated identity.	Child processes, delayed steps, and agent handoffs can obscure lineage.

How it fits with existing security tools

Existing capability	Relationship to Pantheon Control
Identity and access management	Provides principals, grants, and revocation controls. Pantheon's proposed role is to relate the effective identity to a specific agent action and coordinate a scoped change.
EDR and workload security	Provides host and runtime telemetry and may isolate a workload. Pantheon would use those signals and stage authorized response.
SIEM and SOAR	Provides log analytics, incident workflows, and automation. Pantheon would focus on the agent-action context and the evidence needed to choose or verify a response.
Cloud, application, data, and vulnerability tools	Provide exposure, permission, data, and patch facts plus the controls that change them. Pantheon would connect those facts to the agent path.

Illustrative example

A research assistant is allowed to read approved support cases and draft a summary in a collaboration workspace. It proposes exporting a large set of customer records to an unapproved destination. The proposed control loop resolves the agent, sponsor, delegated identity, tool route, data class, volume, and destination; checks the customer's policy; and asks an on-path gateway, host, application, or destination control to deny or hold the request before execution. If the export

already occurred or another route remains open, authorized responders can narrow the identity grant, disable the connector, or block egress, then verify the source system and logs.

This example is illustrative. Prompt-injection detection is not guaranteed. A policy decision can prevent the export only when the request passes through a compatible control point before the destination acts.

What JADEPUFFER can and cannot establish

Public reporting describes an attacker-operated campaign that entered through an internet-facing vulnerable Langflow service, then used infrastructure access and credential discovery to reach other systems. It was not an organization's approved agent taking an unsafe action. Agent governance or an MCP gateway alone would not have stopped the reported initial exploit.

A broader product concept could help connect exposure, runtime, identity, secrets, network, data, and recovery evidence, then stage authorized response. That is a hypothesis to test through integrations and customer validation, not a demonstrated Pantheon outcome.

How to evaluate a first pilot

1. Choose one named agent, one owner, one business workflow, and one action route. Document adjacent routes that are not covered.
2. Start read-only. Compare the discovered agent, actual principal, tools, permissions, and targets with owner-validated source systems.
3. Run policy in shadow mode. Measure attribution quality, missing fields, analyst agreement, false positives, and decision latency.
4. Test one pre-execution denial and one approval hold in a controlled environment. Capture a source-control receipt and test bypass and outage behavior.
5. Test one approved containment or remediation action, then re-read the source system and confirm both the change and intended workflow.

Questions to ask before buying

- Which agent routes and versions are covered today, and which are not?
- What source proves the actual principal and downstream operation?
- Where is the pre-action enforcement point, and how is a deny acknowledged?
- What happens if telemetry, policy, or the enforcement point is unavailable?
- Which response permissions are required, and can they be separated from read-only discovery?
- How are action data, identifiers, logs, retention, and deletion handled?
- What evidence demonstrates that a fix changed the source system and closed the path?

Status and product claim boundary

Pantheon Control is currently a product concept with a simulated public demo. These documents do not represent live integrations, validated enforcement, customer deployments, or a prevention outcome. Any future product claim should name the tested route, enforcement point, required permissions, bypass conditions, and evidence of the result.

References

- NIST, Artificial Intelligence Risk Management Framework 1.0, NIST AI 100-1, https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-1.pdf. Voluntary AI risk-management framework; it is not a product certification.
- Model Context Protocol, Specification, https://modelcontextprotocol.io/specification/2026-07-28. Protocol documentation; MCP is one agent tool path, not a universal enforcement layer.
- OWASP GenAI Security Project, Top 10 for Agentic Applications, https://genai.owasp.org/2025/12/09/owasp-top-10-for-agentic-applications-the-benchmark-for-agentic-security-in-the-age-of-autonomous-ai/.