JETSTREAM

The JetStream AI Kill Switch™

Turn off one agent in seconds, without turning off your AI.

A white paper from JetStream Security Inc.

5201 Great America Parkway
Suite 346
Santa Clara, CA 95054
go.jetstream.security | sales@jetstream.security

v1

Contents

At a glance

What it is. The JetStream AI Kill Switch™ stops a single AI agent on demand by terminating the entitlement it needs to act. You turn off one agent, not your whole AI estate.

Why it matters. When an agent is compromised, overspends, or must be halted for compliance, most teams lack this granular control. Not being able to stop one agent is the sharpest edge of the AI trust gap.

How it works. From the agent's Blueprint page, an administrator clicks the Kill Switch to disable it. The agent stops in seconds, every other agentic flow keeps running, and the action is reversible and fully audited.

The problem: you can't stop just one

Enterprises are adopting agentic AI faster than they can govern it, and trust is the real blocker: only 17% of organizations run AI at production scale (UBS Evidence Lab), held back by missing control rather than by the models. Part of that missing control is blunt and specific. The credential an agent runs on is usually a raw, shared provider secret, so cutting it off for one misbehaving agent cuts off every other workflow that shares it. You are left with two bad options: shut down far more than the one agent causing the problem, or leave it running and hope.

The off switch is the sharpest edge of the trust gap. If you cannot stop one agent, you do not really control it. And if you do not control it, you cannot trust it. That is what the AI Kill Switch™ solves: precision control, agent by agent, so a single agent can be stopped in seconds without touching the rest.

You reach for the Kill Switch in three moments:

1. When an agent is compromised or turns unsafe.
2. When it blows its token budget.
3. When a policy or regulation demands you suspend it now.

All three come down to the same move, done in one pinpointed action.

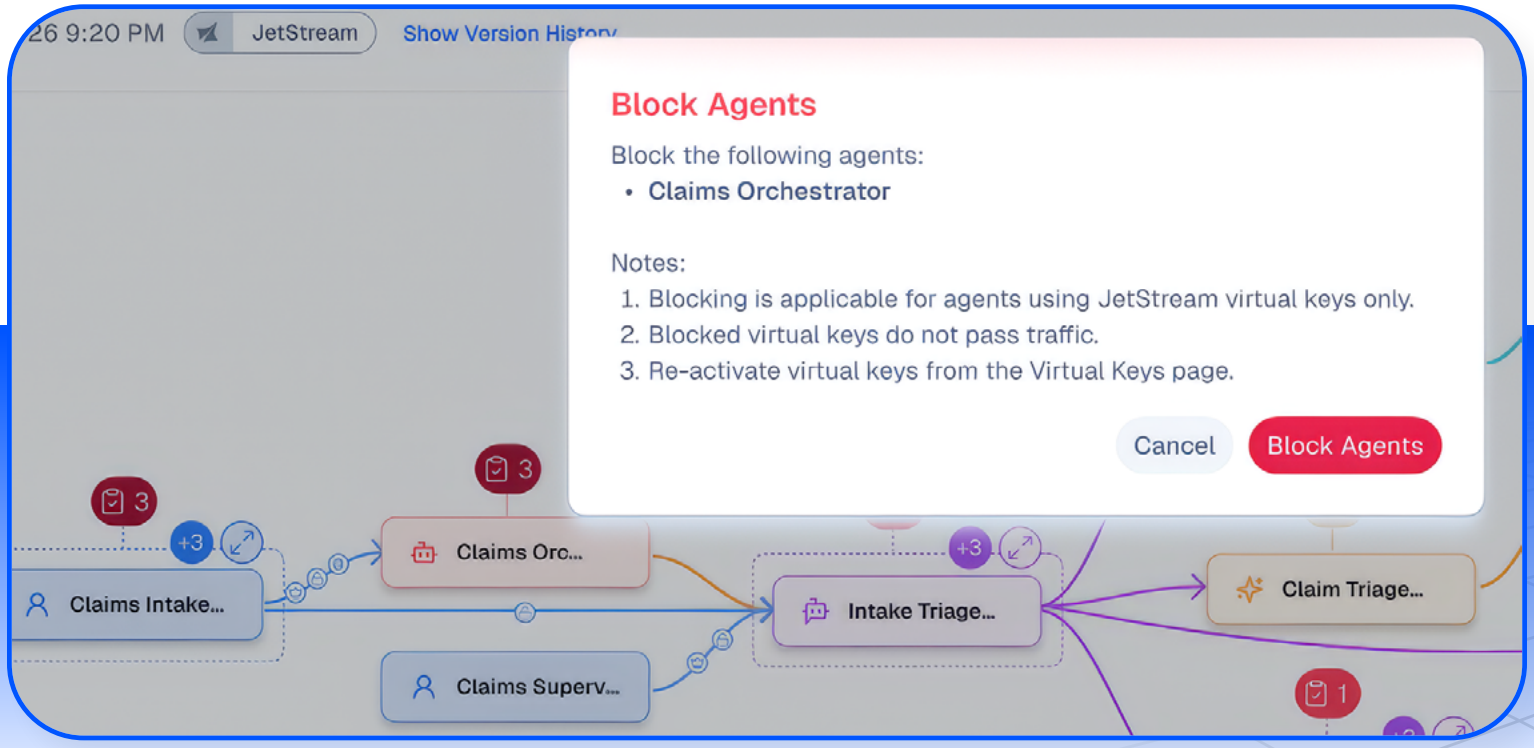
What the JetStream AI Kill Switch™ is

Underneath every agent, JetStream maps identities to their entitlements. Each agent, and the human or agentic identity behind it, is bound to the specific entitlements it uses to reach models, datasets, MCP servers, and tools. That identity-to-entitlement map is defined in JetStream AI Blueprints™, the approved operational contract for the workflow. Because the platform already knows which entitlements belong to which identity, the Kill Switch can revoke exactly the right entitlements and shut off the correct components in one pinpointed action. No entitlement, no action. The agent stops, and the rest of the estate keeps running.

The Kill Switch is selective by design. Whether you have 10 agents or 1,000, every one of them is governed by a discrete Blueprint, and every action runs in a specific identity and entitlement context. What matters is that stopping one no longer means stopping everything. You are not turning off AI. You are turning off a specific agent, on purpose, and leaving everything else running.

Regulations and Standards

Regulators and standards bodies have quietly converged on something most enterprises haven't caught up to: if you can't stop an AI agent, you don't control it. The EU AI Act now requires a way to bring a high-risk system to a safe halt, and both the NIST AI Risk Management Framework and ISO/IEC 42001 expect organizations to be able to disengage or shut down AI that operates outside its approved boundaries. A stop button is no longer a nice-to-have. It's the line between AI you govern and AI that's governing itself.



A live graph of JetStream AI Blueprints with a Jetstream AI Kill Switch™ in action; clicking it disables the agent and its entitlement goes inactive, neighboring Blueprints unaffected.

Who reaches for it, and what worries them

An AI Kill Switch is rarely one person's decision. Several leaders share the AI estate, and each comes to the Kill Switch with a different concern. Here is what keeps each of them up, and what the control gives them.

Who	What keeps them up	What the Kill Switch gives them
CEO, CIO, CAIO and the board	The board's question, "what happens if something goes wrong," and personal accountability for AI decisions made several layers down.	Provable control, and a scoped, reversible failsafe that is rarely needed and always there.
CISO and security leaders	A compromised or runaway agent they cannot stop fast enough, with no way to contain one without a redeploy or taking down more than the problem.	A contained stop in seconds, no redeploy, and a full record of the stop and the restart.
CTO, Head of Platform and AI Engineering	Governance that slows shipping, and blunt controls that take everything down to fix one thing.	Revocation as infrastructure. One agent dies, the rest run, with no ticket queue.
CFO, finance, and procurement	An AI bill that grows without warning and is gigantic before anyone sees it coming.	A hard cap on a runaway agent at the source, tied to an accountable identity, before the overage lands.
Chief Risk and Compliance Officer	Non-compliant processing they cannot halt quickly, and audit evidence reconstructed after the fact.	An immediate, defensible suspension with time-stamped evidence that's compliance-framework- and auditor-ready.
Chief Privacy Officer, DPO, or General Counsel	A data-handling exposure that can kill a deal, and automated decisions that must stop under regulation.	A circuit breaker scoped to the one workflow, with a defensible record of the action.

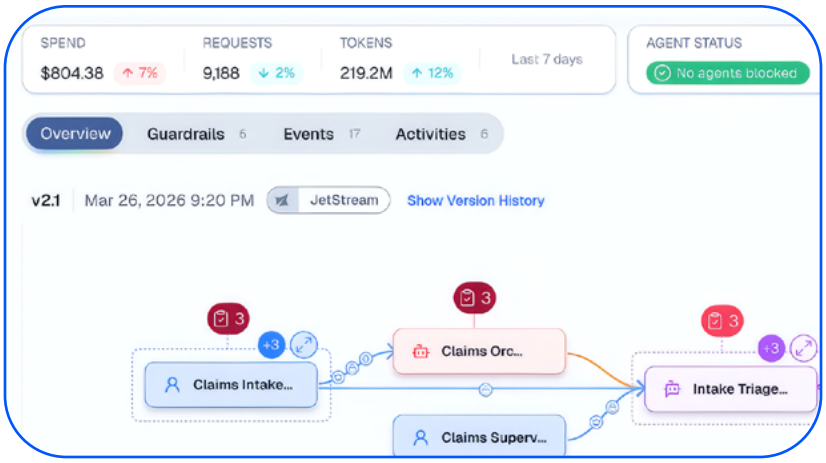
How it works, and why you stay in control

A Kill Switch is only worth trusting if it is precise, and if the people accountable for the agent are the ones who decide when to use it. Both are built in.

The precision comes from identity and its entitlements. Every agentic workflow has three identities that must be governed together: the human who invoked the action, the agentic identity that carries it out, and the human product owner accountable for that agent. JetStream Identity Broker™ binds all three into one accountable chain, so every action traces back to a responsible owner by design. Authority then travels as a scoped, revocable entitlement, not as a standing secret. JetStream Key Broker™ issues virtual, revocable entitlements scoped to the approved Blueprint, in place of raw provider credentials. Because each agent's authority is carried by its own entitlements, acting on them affects that agent alone.

The control stays with your team, and nothing happens on its own. The flow is simple:

1. **Click the Kill Switch.** From the agent's Blueprint page, an administrator selects the Kill Switch to disable the agent, which revokes the entitlement it runs on. The Blueprint shows the blast radius, exactly what that agent can reach, before you act.
2. **Access stops at once.** Because every LLM and MCP call must resolve a valid entitlement, a disabled agent can no longer act anywhere, taking effect on the very next request.
3. **Controlled reactivation.** Re-enable the agent from the same Blueprint page once the issue is resolved. An administrator can also enable or disable the entitlement from the Key Management page.
4. **Fully audited.** Every disable and every re-enable is recorded with the actor and the timestamp. agent is disabled, scoped to the approved Blueprint, while the rest of the estate runs on. During a broader incident or a malicious campaign, the same control halts many agents, or the entire platform, at once.



That audit trail is a defensible account and usable evidence of what was stopped, by whom, and when, ready for post-incident review or an examiner rather than reconstructed after the fact. The Kill Switch is contained, attributable to a named owner, and reversible. That is why it can be trusted rather than feared. You can't trust what you can't *attribute*.

One move, three situations to make it

The Kill Switch answers three control needs: security, cost, and compliance. All three come down to the same move: revoke one agent's entitlement.

1 Security control: cut a compromised or misbehaving agent

When a credential is compromised, or an agent starts acting maliciously or outside its approved Blueprint, the Kill Switch is a single action that stops it. Terminate the entitlement and the agent goes dark. There is no redeployment, and no race to rotate a shared secret across every workflow that uses it. The affected agent is disabled, scoped to the approved Blueprint, while the rest of the estate runs on. During a broader incident or an injection campaign, the same control disables many agents, or the entire platform, at once.

2 FinOps control: stop the spend before the bill lands

Agentic AI spends money continuously, and a single agent can run far beyond plan without a code change. The risk is not hypothetical. Thousands of Google Gemini keys recently leaked through client-side code. One developer watched a single stolen key run up $82,314 in about 48 hours (The Register, March 3, 2026). When an agent's token budget is exceeded and the business wants to stop paying for it, the owner disables that agent from its Blueprint page. Spend is capped at the source, in the moment, rather than reconciled as an overage at month's end. Because the action is scoped to one Blueprint, only that agent stops. A per-Blueprint Kill Switch makes a runaway agent a one-agent event, not an estate-wide scramble.

3 Compliance and policy control: suspend on a regulatory or policy trigger

Some events demand that an AI workflow stop now, not after a review cycle. A policy change, a regulatory obligation, or an automated decision that must be halted under a governing framework.

The Kill Switch gives compliance and legal teams an immediate, defensible suspension: disable the specific agent, contain the exposure, and produce a time-stamped record of the action for auditors. It is a circuit breaker for processing that has become non-compliant, scoped to the one workflow rather than the whole system.

Why JetStream

The Kill Switch is not a standalone function. It is the last move in a single governance loop, and the earlier stages are what make the stop precise and safe.

JetStream discovers the AI estate, turns discovered agents into approved Blueprints, binds every action to an accountable identity through Identity Broker and Key Broker, and watches live behavior against the approved design. JetStream AI Drift Detection™ surfaces the very things that make you reach for the Kill Switch: a model swap, a new tool, a permission change, a cost spike, a compromised component.

Detection tells you when something is wrong. The Kill Switch is how you stop it. Because authority already travels as a revocable, identity-scoped entitlement, the stop is already built in. The JetStream SAIG Platform™ is an AI control plane that makes the Kill Switch a standing capability, not an emergency improvisation.

A governed agent is one you can stop. That is the peace of mind the Kill Switch delivers. You can turn off a single agent, or several, exactly when you need to, and leave the rest of your AI running.

See it run

The fastest way to trust a Kill Switch is to watch one work. We will show you a live workflow, an agent disabled from its Blueprint page with the Kill Switch, and re-enabled once the issue clears, on your own terms. Book a live Kill Switch demo.

About JetStream Security

JetStream makes AI governable at enterprise scale by providing unified visibility and control across all AI activities so that AI becomes a managed asset, not an unmanaged liability.

Our Mission

Accelerate Enterprise AI Adoption

The Result

Governance enables trust. Trust unlocks advantage.
Governance → Trust → Advantage

Contact us at sales@jetstream.security

For more information about JetStream Security

v1
5201 Great America Parkway
Suite 346 Santa Clara, CA 95054
go.jetstream.security | sales@jetstream.security

JETSTREAM