The Off Switch Problem

Why enterprises cannot turn off AI, and why the off switch is a governance artifact before it is a security control

CTRL ALT Elite · Fortune 100-caliber infrastructure, operations, and security leadership

By René López, Founder and Principal · ORCID https://orcid.org/0009-0008-4993-0088

Executive Summary

The largest enterprises in the United States spent the week of October 12, 2026 announcing that they are blocking AI agents, and that fact is the clearest evidence yet that most organizations have no working off switch. JPMorgan, Schwab and Amazon can refuse an outside agent at the login page because that is the one place they can still see it. Inside the perimeter, where employees' assistants, developers' coding agents and vendors' automations already run on inherited credentials, the same companies cannot name every agent, cannot say who owns it, and cannot stop one without stopping the human account it borrowed.

Turning off AI fails for six connected reasons: the agent cannot be found, it has no accountable owner, its credential is a person's credential, the only control is all or nothing, the business now depends on it, and nothing proves what it was allowed to do. A sub-50ms kill switch, the fix most vendors are selling this month, solves none of the first five.

> *The off switch is a governance artifact before it is a security control.*

This paper argues that an organization can turn off an agent when, and only when, it keeps a register of agents and keys with an owner, a scope, a budget and a revoke path for each. The recommendations close with a 90-day sequence a CIO or CISO can run with existing tools, and six questions a board should be asking now.

The Blocking Reflex

In one week, three of the most sophisticated technology operators in the country chose the bluntest control available: refuse all outside agents. JPMorgan Chase confirmed on October 13 that it blocks external agents such as Meta's Muse from operating retail accounts, with Jamie Dimon warning of rogue agents in the thousands. Schwab said it stops outside agents tens of times a week and will offer read-only MCP access instead, steering clients to its own Charley assistant. Amazon rewrote its robots.txt to shut out Google's shopping agents, while Shopify requires Shop Pay so a human approves every purchase and Walmart is not blocking at all.

Three things stand out.

1. **The block works only at the one chokepoint these firms control:** the public login or checkout page. An agent arriving with a customer's own credentials through Plaid is already partly inside; JPMorgan allows read-only balance access precisely because it cannot cleanly separate the agent from the customer.
2. **The block is as much commercial as defensive.** Schwab's cash-sweep revenue and Amazon's search placement are both threatened by agents that shop around. An off switch that protects margin will be defended harder than one that protects data, and regulators will notice.
3. **Every one of these firms runs agents internally and is building customer-facing ones.** The policy is not against AI; it is against AI they did not issue and cannot revoke. That is the governance gap this paper is about, and it exists inside every enterprise, not only at the perimeter of the largest ones.

Six Reasons the Off Switch Fails

The failures are sequential: each one is possible because the one before it went unaddressed. Most vendor attention goes to the fourth, the blunt control, because that is where the symptom shows. The cause sits in the first three.

FIGURE 1: WHY THE OFF SWITCH FAILS, SIX LINKED FAILURES

Failure	What happens	Which makes the next one possible
1. Cannot find it	An agent nobody registered runs in production	It has no owner
2. No owner	No named human is accountable for the agent or its key	It runs on whatever credential launched it
3. Borrowed credentials	The agent acts as a person, with everything that person can reach	It cannot be scoped on its own
4. All or nothing	The only safe action is to block everything	Nobody pulls the switch until a team depends on it
5. Dependency	A workflow, a team or revenue now relies on the agent	Stopping it needs proof of what it was doing
6. No evidence	Nothing recorded what it was permitted to do	Nobody will sign the order to stop it

Read it top to bottom. An agent nobody registered has no owner; an agent with no owner runs on whatever credential launched it; a credential shared with a person cannot be scoped, so the only safe action is to block everything; by the time anyone considers that, a team depends on the agent; and because nothing recorded what it was permitted to do, nobody will sign the order to stop it.

Failure 5 is the one executives underestimate. Gartner's finding that only 13% of organizations are confident in their agent governance sits beside Forbes' finding that 79% already run agents in production. The gap is not awareness. It is that the agents were useful before they were governed, and useful things are hard to switch off.

Evidence From the Field

The incidents and surveys published between October 12 and 18, 2026 each illustrate one of the six failures. None of them was stopped by a runtime control; most were discovered after the fact.

FIGURE 2: THIS WEEK'S INCIDENTS, SORTED BY FAILURE

Failure	Incident or finding	What it shows
1. Cannot find it	Employees connecting personal agents to enterprise systems; Zapier: 79% of execs say staff bypass AI policy	Policy exists on paper; the agents exist in production
2. No owner	Forbes: 79% run agents, 2% have identity tooling; Bitsight: stolen OpenAI, Anthropic and Google keys sold on 16 forums	A key with no owner is not noticed when it is stolen; the victim pays the inference bill
3. Borrowed credentials	ASOS: one employee credential, 138,828 customers exposed; CERT-In: agents under test read live production data	Authenticated is not authorized; agents inherit everything the human could reach
4. All or nothing	JPMorgan, Schwab and Amazon block agents outright; Amex builds Intent ID to scope purchases	Without per-agent scope, the only control is refusal; Amex shows the alternative
5. Dependency	Security concerns slowing AI in ERP; Schwab steers clients to its own assistant	Firms cannot stop AI they rely on, so they replace outside agents with their own

Failure	Incident or finding	What it shows
6. No evidence	OpenAI model spent third-party ETH and omitted it from its summary; Just Security on the liability gap	The agent's own report is not a record; nobody can show what was authorized

One more case belongs here because it involved no AI at all. Refrigeration controllers in 300 French supermarkets were told to extend defrost cycles from one hour to eight through an authentication bypass in outdated firmware. Standing access with no verification of who issues a command is the same failure, and the industry has had thirty years to solve it for operational technology. It has not.

What a Working Off Switch Requires

An off switch is the last step of a chain, not a product you buy. Each requirement below answers one of the six failures, and each can be met with tools most enterprises already own.

FIGURE 3: SIX REQUIREMENTS, ONE PER FAILURE

Requirement	Answers	Minimum viable implementation
Inventory	Cannot find it	Route all model traffic through one gateway; reconcile against provider billing, SSO app logs, repo secret scans and endpoint plugin inventories monthly
Ownership	No owner	Every key and agent has a named human, a business purpose and an expiry date; unowned keys are frozen after 30 days
Separate identity	Borrowed credentials	Agents get their own short-lived credentials issued by the gateway, never a person's token; human accounts cannot be used by automation
Scope and budget	All or nothing	Allowed models, tools and data classes per agent; a token budget with a hard cap; exceeding either pauses the agent, not the team
Graceful degradation	Dependency	Every agent-dependent workflow has a documented manual or read-only fallback, tested the way disaster recovery is tested
Evidence	No evidence	Immutable log of what each agent was permitted, what it did, who approved it and who froze it; mapped once to NIST AI RMF and ISO 42001

The revoke action itself is then trivial: freeze the key at the gateway, notify the owner, invoke the fallback. What makes it trivial is the five rows above it. Organizations that buy the kill switch first discover they have nothing to point it at.

Two design principles follow. First, scope beats speed: a control that pauses one agent's access to one system is worth more than one that halts the fleet in fifty milliseconds, because the second will never be used in anger. Second, the register must be the system of record for finance as well as security. When the CFO reads token spend by team from the same ledger the CISO reads scope from, ownership gets maintained, because someone is paying for it.

Two further capabilities turn an off switch into control over the data itself.

Measure the round trip. Count, for every protected value, how often it was sent to a model, returned, and restored to a real value, and to which destination. The ratio of values restored to values protected shows whether an agent needed the real data at all, and a value restored to a new destination is the earliest sign a stolen key is in use. When data must later be destroyed, this record is also the map of every copy that escaped.

Destroy keys at fine grain. If each person, data class or document has its own key, destroying one key erases one guest or one leaked file everywhere it was ever sent, including the model provider's logs, without touching anything else. Keys that expire on their own enforce retention without a cleanup job. The audit trail stays permanent because it never held the values, only the tokens; that is what lets an organization be both erasable and auditable.

Recommendations for CIOs and CISOs

Start with the register, not the switch. The sequence below assumes an existing identity provider, a cloud bill and an engineering team willing to route traffic through one gateway; nothing else needs to be purchased in the first 90 days.

1. **Days 1 to 30: find and name.** Pull every model provider invoice and every API key from secret scans, SSO logs and the gateway. Assign a human owner to each or freeze it. Publish the count to the executive team; it will be three to ten times what anyone expected.
2. **Days 31 to 60: separate and scope.** Issue agent credentials from the gateway and ban human tokens in automation. Set a token budget and an allowed-tool list per agent. Pause anything that exceeds either and see what breaks; that is your dependency map.
3. **Days 61 to 90: prove and rehearse.** Map the register to NIST AI RMF and ISO 42001 once. Run a tabletop exercise in which one agent and then one team is frozen, and time how long the fallback takes. Report the result to the board alongside the quarterly cyber update.

SIX QUESTIONS A BOARD SHOULD ASK NOW

How many AI agents and model keys do we run, and how confident are you in that number?

Who owns each one, and what happens when that person leaves?

Can any agent act with a human employee's credentials? Which ones?

Which agent, if frozen today, would stop revenue or a regulatory filing?

What is our monthly token spend by business unit, and who approved it?

If a regulator asked what an agent was authorized to do on a given date, what would we show them?

> *An organization that can answer all six has an off switch. One that cannot has a block list, and will keep adding to it.*

About the Author

René López is Founder and Principal of CTRL ALT Elite LLC, a fractional CIO, CTO, and CISO advisory practice for Series A and B startups and private-equity-backed operators. He brings 25 years of Fortune 100 infrastructure and operations leadership, most recently as VP of Infrastructure and Operations at Royal Caribbean Group and Managing Director of Enterprise Applications, Infrastructure and Operations at Delta Air Lines. He completed the EC-Council Certified CISO (C|CISO) program in 2026, is a FinOps Certified Practitioner, and co-authored SAN patents at Vicom Systems.

CTRL ALT Elite

reboot@ctrlaltelite.io · 954-516-6909 · ctrlaltelite.io · Pompano Beach, FL

René López, Founder and Principal · ORCID https://orcid.org/0009-0008-4993-0088

Disrupt · Innovate · Elevate

© CTRL ALT Elite. Shareable in full with attribution. Vendor-neutral; no product is endorsed. Incidents and figures current as of October 18, 2026; verify against primary sources before relying on them.

Sources: Incident summaries were drawn from The RuntimeAI Brief, week of October 12 to 16, 2026 (runtimeai.io), a vendor digest, and checked against the primary reports linked in the tables above; vendor framing in that digest was not adopted. Cybernews, JPMorgan blocks outside AI agents; The Next Web, Schwab and JPMorgan block agents moving customer money; American Banker, Schwab points clients to Charley; TipRanks, Amazon blocks Google shopping agents; The Hacker News, Bring Your Own Agent; BankInfoSecurity, Gartner on agent governance; Forbes, 79% run agents, 2% have identity tooling; Help Net Security, stolen AI API keys fund autonomous scans; eSecurity Planet, ASOS breach via employee credentials; The Hacker News, CERT-In advisory on AI agents; TechCrunch, Amex Intent ID; The Hacker News, OpenAI model spent third-party test ETH; Just Security, Can the law stop AI hackers; ERP Today, security concerns and AI in ERP; eSecurity Planet, French supermarket refrigeration attack.