AIMONGER
WHITE PAPER

AI Discovery

Talk to One Agent First, Why Crews Fail Without a Loop

Boards are being shown crews, swarms, and agent roadmaps before one production loop has proved useful. The sensible sequence is narrower: one agent, one workflow, allowlisted tools, evaluation, cost envelope, and a staffed human gate. Global agent fashion starts with swarms. Operating capacity starts with one loop that works.

David Saliba . 2026-07-14 . aimonger.com/whitepapers/single-agent-before-crew/

The problem in one sentence

Agent roadmaps are easy to draw; one production loop with a moving metric is hard, and that sequence is reversed in most programmes.

Boards are shown crews, swarms, and orchestration diagrams before anyone can name a workflow metric that moved last month. The sensible milestone is narrower: one agent, one process, allowlisted tools, evaluation, cost ceiling, staffed human gate.

McKinsey's State of AI 2025 survey shows AI use is widespread while enterprise EBIT impact remains concentrated among high performers attributing more than 5 per cent of EBIT. Those firms redesign workflows before they multiply agent topology.

Single-agent loop, plainly: one agent plans and acts with approved tools under policy, with logging, evaluation, and human approval on consequential steps. **In practice:** invoice exceptions: extract, compare to ERP, queue mismatches; one loop, one owner, one metric.

Crew, in short: multiple agents with distinct roles passing work to each other under an orchestrator. **Worked case:** three agents chatting about each document multiplies tokens before a human sees a decision.

Human gate means: a named role with authority to approve, edit, or block outputs before they affect customers, regulators, or systems of record. **On Monday:** no customer email sends without a service manager on the rota who can be audited.

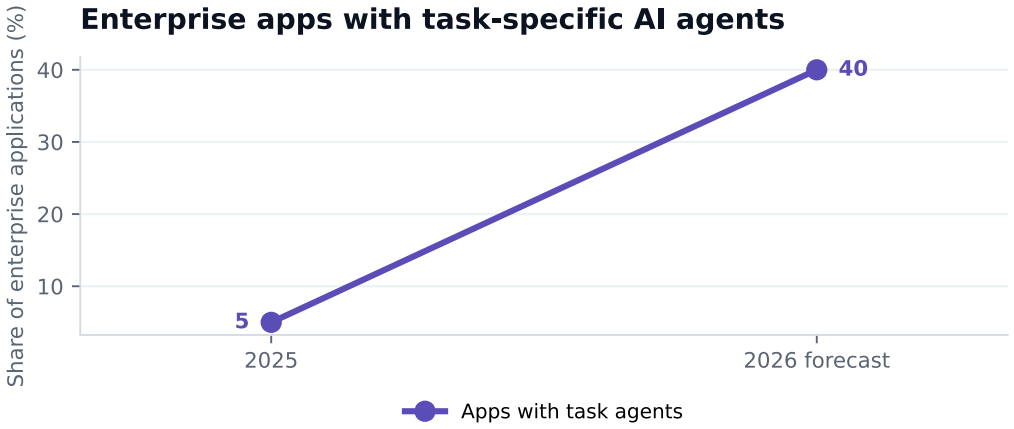
The sequencing error

Enterprise AI programmes often skip from chatbot pilots to multi-agent architectures in one budget cycle. In a CIO office, that skip usually shows up as cost, ownership confusion, and a board pack that cannot name the workflow metric.

What you see globally right now is agent language moving faster than operating proof. US and Asian platforms are packaging agents into suites, startups are selling crews for every workflow, and internal teams feel pressure to show an "agentic" roadmap. Gartner predicts that 40 per cent of enterprise applications will feature task-specific AI agents by the end of 2026. Separately, Gartner predicts that more than 40 per cent of agentic AI projects will be canceled by the end of 2027 due to escalating costs,

unclear business value, or inadequate risk controls. Reuters reported the cancellation forecast on the same day.

The practical reading for boards: agents are entering the product surface of enterprise software, but internal programmes that jump to crews without a working loop will join the cancelled set. McKinsey's State of AI 2025 survey shows why this matters: AI use is widespread, while enterprise EBIT impact remains concentrated among a small high-performer cohort, organisations attributing more than 5 per cent of EBIT to AI and reporting significant value from workflow redesign.



Gartner press release, Aug 2025 (forecast). AIMonger redraw.

Figure 1. Gartner forecast: share of enterprise applications with task-specific AI agents, 2025 vs end-2026. Plain read: agents arrive in vendor suites whether you build them or not. **Operating case:** Salesforce or Microsoft adding an agent is not proof your internal crew works. Source: Gartner press release, 26 August 2025 (https://www.gartner.com/en/newsroom/press-releases/2025-08-26-gartner-predicts-40-percent-of-enterprise-apps-will-feature-task-specific-ai-agents-by-2026-up-from-less-than-5-percent-in-2025). AIMonger redraw.

What "one agent loop" means

A production single-agent loop is not a chat window with a persona.

It includes:

1. **A named workflow with baseline metric.** The loop serves one process with a measured before-state, cycle time, exception rate, or cost per case, so success is operational, not narrative. A baseline number is what lets the team say "before the loop, exceptions took six hours; now they take ninety minutes." Without it, the loop's value is a story the project team tells and the board cannot check. The named workflow is the scope boundary that stops the pilot from becoming an everything-tool.

2. **Allowlisted tools only.** The agent may call only pre-approved systems; anything else requires a change request, preventing silent privilege sprawl during the first production month. An allowlist is the least-privilege control that stops the agent from escalating its own access one justified request at a time. The change-request path is what keeps the boundary governed rather than porous.

3. **Autonomy policy in writing.** Document what the agent may execute versus propose, so operators and auditors know where human approval is mandatory before customer or ledger impact. The policy is the artefact an auditor reads, and it is the line the operator enforces without having to phone the build team. "It can draft, it cannot send" is the kind of sentence the policy needs on page one.

4. **Reconstructable provenance.** Every retrieval step and tool call logs to an export auditors can replay, not only a chat transcript that disappears when the session closes. Provenance is what makes a wrong answer diagnosable rather than mysterious, because the team can replay the exact steps that produced it. An audit export is the difference between an agent the firm can trust and an agent the firm can only hope.

5. **Evaluation with promotion holdout.** Proposition tests and a frozen holdout set gate promotion decisions, so tuning cannot overfit the demo questions that sold the pilot internally. The holdout is the honesty layer that keeps the accuracy claim general rather than memorised. Without it, the team reports results on the questions they built the agent to answer, which proves nothing about the next real question.

6. **Hard cost envelope per run and day.** Token, tool, and wall-clock caps stop runaway loops before finance discovers them on the monthly cloud invoice. A cap is the financial control that makes autonomy affordable, because it turns an open-ended spend into a bounded one. The hard stop is what makes the cap real; a cap that only warns is a cap that does not exist at 2 a.m.

7. **Staffed human gate for consequential actions.** A named role on the rota approves, edits, or blocks outputs before they hit customers, regulators, or systems of record. The gate is the control that makes the loop production-safe, and "staffed" means a person is on the rota with coverage hours, not a title on a slide. An unstaffed gate is autonomy without a brake.

8. **Operator training and override path.** People who run the workflow daily know how to refuse, escalate, and override without paging the build team for every exception. Training is what makes the loop operable beyond the project team, because the operators are the ones who will catch the edge cases the demo never showed. An override runbook is the document that makes the human check fast rather than ceremonial.

If any of these are missing, you do not have an agent programme. You have a demo.

Tool allowlist, put simply: the explicit list of systems an agent may call; nothing else, ever, without a change request. **Operating case:** read ERP and create draft tickets yes; delete records or email customers no until policy says yes.

Cost envelope (plain English): hard caps on tokens, tool calls, and wall-clock time per run and per day. **On the P&L:** if one run exceeds EUR 12, it stops and queues for human, with no silent budget burn.

Eurostat's 20.0 per cent EU enterprise AI adoption figure for 2025 is a useful local baseline. Many mid-market firms are still early enough to build discipline before agent sprawl hardens into architecture debt.

Why crews feel productive and fail

Crews create visible activity: agents assigning tasks to agents, long traces, impressive diagrams. Activity is not throughput.

Coordination costs show up as:

- **Token multiplication from orchestrator chatter.** Each handoff between agents adds planning and summarisation tokens that often exceed the work of the underlying task itself. The orchestrator has to describe the task, the receiver has to confirm understanding, and the result has to be summarised back, and each step is a model call. Three agents on one document can spend more tokens agreeing than doing.

- **Latency that breaks the workflow SLA.** Multi-agent coordination pushes P95 response time past what operators or customers tolerate, even when each individual agent looks fast in isolation. The coordination overhead is serial, not parallel, so the handoffs add up at the tail. A workflow that needs a response in two seconds cannot afford four sequential agent calls.

- **Conflicting writes to shared state.** Parallel agents can update the same ticket, record, or draft without a schema contract, leaving operators to reconcile contradictory outputs manually. Two agents writing to the same field is a race condition with no database lock to save it. The reconciliation labour is a cost the crew diagram never shows.
- **Debugging that requires graph specialists.** Incidents need someone who understands the orchestration topology, which smaller mid-market teams rarely have on call at 2 a.m. A single-loop incident is a stack trace; a crew incident is a topology the on-call engineer has to reconstruct before they can diagnose it. The specialist dependency is a continuity risk the crew introduces by design.
- **No clear owner when the final artefact is wrong.** Crews diffuse accountability so steering cannot ask one name why a customer email, posting, or case closure was incorrect. When every agent contributed and none owned, the post-incident review assigns blame to the architecture rather than to a person. Diffuse accountability is how a programme stops learning from its own mistakes.

A single loop makes ownership obvious. A crew without contracts makes ownership theatrical.

Decision rule for topology

Situation	Default
New domain, weak evaluation habits	Single loop only
Linear workflow, few tools	Single loop
Need a critic/reviewer for high-stakes drafts	Single loop + critic (still simple)
True parallel specialised work with clear contracts	Specialist pair, then crew
Vendor pitch leads with “multi-agent” and no workflow metric	Refuse

Ninety-day path

Days 1-30: Ship one assisted or single-agent path on one workflow. Measure cycle time and exception rate.

Days 31-60: Harden isolation, evaluation, and operator training. Freeze topology.

Days 61-90: Only if metrics move, consider a second specialist role with a one-page handoff contract. If metrics do not move, do not add agents; fix the workflow choice.

Board questions

1. **Which single loop is trusted in production today?** Name the one workflow path with logging, gates, and owners, not a roadmap slide showing six planned agents. The question forces the room to point at a live artefact rather than a future state. A roadmap is not a capability, and a named loop is.
2. **What metric moved because of it?** Report cycle time, exception rate, or cost per case change over at least four weeks, not demo completion or token volume alone. Four weeks is long enough to see absorption rather than a one-off demo glow, and it is the window that separates a working loop from a launch. Token volume is a platform metric; cycle time is an operating metric.
3. **Why is a crew required beyond that loop, with evidence?** Show holdout proof that the single loop failed a quality or latency bar a second role uniquely fixes, not vendor architecture fashion.

The evidence is the holdout score the single loop failed, not the slide the vendor brought. Without it, the crew is a preference dressed as a requirement.

4. **What is cost per successful case at P95?** Finance needs the slow-case unit cost, because agent loops that look cheap at median often break budgets at tail latency. The median is the number the demo reports and the P95 is the number the invoice reflects. A loop that is cheap at median and ruinous at P95 is a loop that fails finance.

5. **Who is the human gate, and are they staffed?** Name the role on the rota with coverage hours; a gate on paper that nobody staffs is autonomy cosplay, not control. Coverage hours are the difference between a control and a poster. An unstaffed gate is how a consequential mistake ships at 2 a.m. with no one to catch it.

6. **What is the kill date if the metric stalls?** State the calendar date and threshold that stops spend, so the programme cannot drift in perpetual pilot mode. A kill date is what keeps a pilot from becoming a permanent line item that nobody ever closes. The threshold is the metric value below which the programme stops, written before the programme starts.

What "good" looks like on a single loop

A board can recognise readiness without reading traces.

Green signals

- **Metric moved four consecutive weeks.** The chosen workflow measure improved or held inside target for a month, showing absorption rather than one-off demo success. Four weeks is the window that distinguishes a real change from a launch artefact, because the demo glow fades by week two. A metric that holds inside target is a loop the operators can live with, not just a loop that ran.
- **Cost per successful case inside ceiling.** Unit economics stay within the pre-approved envelope at P95, so finance is not surprised when volume scales beyond the pilot cohort. P95 is the percentile where the tail lives, and the tail is where agent loops become expensive. Staying inside the ceiling at P95 is the proof the loop is affordable at scale, not just at pilot volume.
- **Operators override without paging engineering.** Front-line staff handle refuses and edits using runbooks, proving the loop is operable beyond the project team that built it. An override the operator can run without the build team is an override that happens in minutes, not hours. This is the signal that the loop has left the project and entered operations.
- **Promotion holdout still untouched during tuning.** Developers did not train against the frozen promotion set, so reported accuracy reflects honest generalisation rather than overfit. The untouched holdout is the integrity check that keeps the accuracy claim trustworthy. A team that tunes against the holdout reports a number that will not hold on the next real question.
- **Incident runbook rehearsed once.** The team walked through a kill-switch or rollback drill, so production failure does not become the first time anyone tests stop conditions. A rehearsed runbook is the difference between a controlled stop and a 2 a.m. scramble. One drill is the minimum bar for calling the loop production-ready.

Red signals

- **Success equals demo script completion.** Steering hears "the agent finished the flow" without a workflow metric tied to customer or finance outcomes. A finished flow is a demo artefact, and without a metric it tells the board nothing about whether the workflow improved. This is the signal that the pilot is measuring theatre rather than absorption.
- **Tool permissions widen weekly without review.** Allowlists expand informally, multiplying OWASP-style privilege abuse surface before anyone updates the threat model. Each informal widening is a privilege the agent granted itself, and the accumulation is how a controlled loop becomes an uncontrolled one. Weekly review is the gate that keeps the allowlist honest.

- **No human gate owner on the rota.** Consequential actions queue to nobody on evenings or weekends, which is when costly agent mistakes typically reach customers. An empty rota is an unstaffed control, and an unstaffed control is a control in name only. The gap is where the consequential mistake ships uncaught.
- **Cost unexplained in finance language.** Engineering reports tokens or GPU hours while the CFO cannot map spend to cost per successful case by workflow. A cost the CFO cannot map is a cost the CFO cannot defend or cut. The translation from tokens to euros per case is the conversation that has to happen for the programme to be governable.
- **Next ask is five more agents before green signals.** Topology expands while the first loop still lacks metric movement, evaluation discipline, or staffed gates. Expansion before green signals is how a small unproven loop becomes a large unproven crew. The ask is the warning that the programme is scaling ambition rather than capacity.

Anti-patterns specific to early agent programmes

1. **Persona theatre.** Multiple chat personalities share one tool and no workflow metric, producing impressive demos that never touch a system of record. The personas are the cast and the demo is the play, but no real workflow ever runs. This is the pattern that looks agentic on a slide and does nothing in production.
2. **Framework shopping.** The team rebuilds the same loop each month in a new library, resetting evaluation and operator training instead of hardening one production path. Each rebuild is a reset of the evaluation baseline and the operator muscle memory, and the net is a year of development with no production gain. The framework is the variable; the loop is the constant.
3. **Autonomy cosplay.** Auto-send or auto-write is enabled before provenance and gates exist, so the first production mistake ships to a customer without reconstructable trace. Autonomy without provenance is a mistake with no audit trail, and the first incident is the one that ends the programme. The gate and the provenance are the prerequisites for autonomy, not the polish.
4. **Eval after launch.** Measurement and holdouts are invented during the first incident, guaranteeing steering debates opinion against opinion instead of against frozen tests. An eval invented post-incident is an eval designed to justify the existing system rather than to test it. The frozen holdout has to exist before launch, or every debate is a rerun.
5. **Crew as compensation.** Extra agents are added to mask a bad workflow choice rather than to fix a documented failure mode the single loop cannot clear on holdout. A crew cannot fix a workflow that was never mapped to a baseline, because the problem is upstream of the agent count. Adding agents to an unmapped workflow is adding cost to an unsolved problem.

Worked example (synthetic)

A mid-market operations team wants faster exception handling on inbound documents.

Wrong start: three agents (classifier, researcher, writer) debating each document.

Right start: one loop that extracts fields, compares to the system of record, and queues disagreements for a human. Metric: time-to-clear exception. After the metric moves, consider a critic pass on edge cases, still not a swarm.

Worked example: IT ticket triage (synthetic)

A mid-market firm wants agents to classify and route internal IT tickets.

Wrong start: classifier agent, researcher agent, resolver agent, notifier agent; four roles, one ticket, P95 latency beyond SLA.

Right start: one loop with allowlisted read access to the ticket system and knowledge base; auto-label and suggest assignee; human approves before priority or customer-facing status changes. Metric: mean time to first correct assignment.

On Monday: prove assignment accuracy for four weeks before any "auto-close" autonomy.

Integration with the rest of AI Discovery

Read with:

- **Absorption capacity.** Decide which workflow deserves the loop first, because agent topology cannot create EBIT if the process was never mapped to a baseline metric. The loop is the absorption mechanism, and the workflow is what is being absorbed. Topology without a baseline is a crew looking for a job.
- **Knowledge trust.** Define how answers and tool outputs are measured, logged, and promoted so the loop earns trust before autonomy expands. Trust is the output of evaluation and provenance, and it is the prerequisite for autonomy. A loop that has not earned trust should not have been given autonomy.
- **Multi-agent board trust.** Use the crew decision criteria when, and only when, a single loop with tools cannot meet quality or latency on documented evidence. The "when and only when" is the gate that keeps crews from becoming the default. The evidence is the holdout the single loop failed, not the deck the vendor brought.
- **Human capacity.** Staff and train the operators who override, refuse, and gate outputs; without them, the loop is a developer demo, not production capacity. The operators are the production capacity the loop runs on, and training is what converts a project team into an operating team. Without them, the loop is a demo that never reaches operations. The difference between success and failure here is single, namely whether the people on the rota can run the loop on a Sunday without paging the engineer who built it.

Appendix: single-loop charter

Complete one charter per workflow before production access. Leave the Entry column blank until the steering group fills it.

Field	Entry
Workflow	
Owner	
Tools allowlisted	
Autonomy boundary	
Human gate role	
Baseline and target metric	
Cost ceiling	
Eval owner	
Kill date	
Crew expansion criterion (if any)	

Plain read: empty cells are intentional. A filled charter is the go-live gate, not a decorative appendix.

Economic test before crew funding

Compute:

```
Coordination tax = (crew cost per success − single-loop cost per success) / quality gain
```

If quality gain is unmeasured, the tax is infinite. If quality gain is marginal and tax is large, refuse. Deloitte's gap between expected agent adoption and mature operating models is the market warning; your local numbers decide the investment.

Coordination tax, briefly: the extra cost and latency of multiple agents versus one loop, divided by measurable quality gain. **Finance reading:** crew costs 3x per successful case but accuracy rises 2%; refuse unless that 2% is worth EUR X in fraud or rework avoided.

Production definition of done for loop v1

- **Metric moved four weeks.** The workflow baseline improved or held inside target for a month on the agreed measure, not on anecdote from the project team. Four weeks is the window that separates absorption from launch glow, and the measure is the one the charter pre-registered. A metric that moves for a week and stalls is a launch artefact, not a production result.
- **Audit export verified.** Legal or compliance successfully reconstructed a sample case from logs without engineering intervention, proving provenance is production-grade. The reconstruction is the test that proves the audit trail works for someone outside the build team. If legal cannot replay a case, the provenance is a developer claim, not an operational capability.
- **Cost ceiling enforced with real stop.** A run that exceeds the cap actually halts and queues, not merely logs a warning that finance discovers later on the invoice. A real stop is the difference between a cap that controls spend and a cap that reports it. The queue is what makes the ceiling a brake rather than a gauge.
- **Operators certified.** Front-line staff completed training and passed override drills, so the loop survives the first week without the build team on call. Certification is the gate that says the operators can run the loop, not just watch it. A loop the operators cannot run is a loop that reverts to the project team the first weekend.
- **Incident rehearsal completed.** The team executed rollback or kill-switch once in a tabletop or staged drill, so production is not the first test of stop conditions. A rehearsal is the cheapest way to find out the kill-switch does not work, and finding out in a drill is infinitely cheaper than finding out in production. One rehearsal is the minimum bar for production readiness.
- **Kill criterion still on calendar.** The graduate-or-stop date and threshold remain visible to steering, preventing silent extension of a pilot that is not moving the metric. A kill date that stays on the calendar is a commitment the programme cannot quietly drop. The visibility is what keeps the pilot from becoming permanent.

Only then open the topology discussion.

Evidence base: agent ambition exceeds operating maturity

Source	Finding	Implication for sequencing
Gartner (Aug 2025)	40% of enterprise apps expected to include task-specific agents by end-2026, up from under 5%	Agent features become table stakes in software suites
Gartner / Reuters (Jun 2025)	Forecast: more than 40% of agentic AI projects cancelled by end-2027 due to cost, unclear value, or weak risk controls	Topology fashion without value and controls is capital at risk
Deloitte State of AI 2026	23% of surveyed leaders report at least moderate agentic use; 74% expect at least moderate use within two years; only 21% report a mature operating model for autonomous agents	Deployment intent is ahead of operating maturity in an engaged AI-leader sample
Stanford AI Index 2026	Organisational AI adoption high in compiled survey evidence; agent deployment still single-digit across nearly all functions	Space remains to build one reliable loop before crew sprawl
OWASP Agentic Top 10 2026	Goal hijack, tool misuse, privilege abuse, cascading failures, and insecure inter-agent communication are named critical risks	Isolation and least agency are prerequisites, not polish

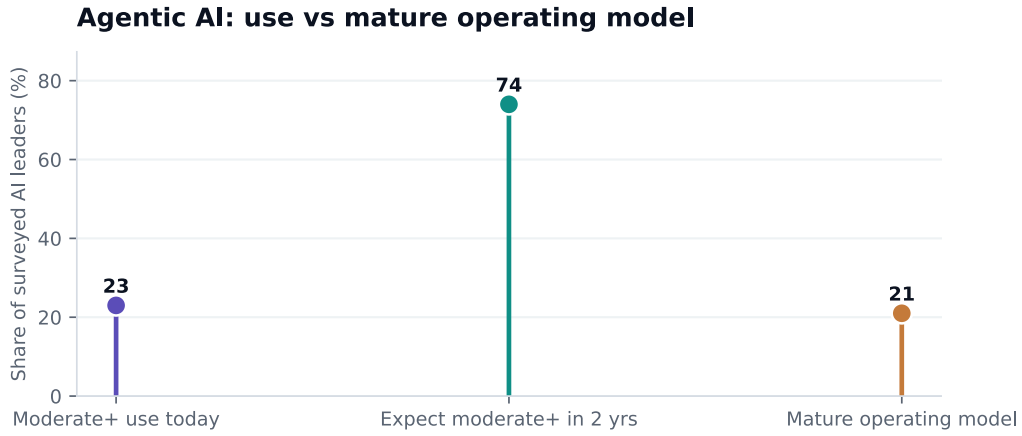


Figure 2. Deloitte 2026: moderate+ agentic use today vs mature operating model. In short: many leaders expect agents; few report mature operating models. **Board reading:** your roadmap slide is probably in the 74% "expect use" bucket; ask which workflow metric moved instead. Source: Deloitte State of AI in the Enterprise 2026 (3,235 leaders). AIMonger redraw. Plain-English read: the market expects agents faster than operating maturity. Starting with one loop is how you avoid joining Gartner's cancelled project forecast.

Methodology note

Deloitte surveyed 3,235 leaders directly involved in AI initiatives across 24 countries (Aug-Sep 2025). Rates therefore describe AI-active organisations, not the full enterprise population. Gartner cancellation figures are forecasts. OWASP is peer-reviewed practitioner consensus, not incidence frequency.

Counter-position: starting with one loop can feel too slow

Competitors may announce multi-agent platforms first. Announcements are not operating capacity. A single loop with a moving metric, audit trail, and kill criterion beats a crew that cannot explain cost or ownership. If a specialist pair is truly required, prove the single loop fails a quality or latency bar first.

Board approval gate for any second agent

Approve a second agent only when the first loop has four weeks of metric movement, a documented failure mode the second role uniquely fixes, a one-page handoff contract, and a cost model that still clears after coordination overhead.

Research addendum: start simple is evidence-backed, not taste

Google's reported sequential-planning penalties for multi-agent topologies, equal-budget comparisons favouring single agents on some multi-hop tasks, and Anthropic's own caveat that dependency-heavy work is often a poor multi-agent fit all support the same operating sequence: master one evaluated loop before funding a crew.

The countercase is real: Anthropic's internal research evaluation showed large gains for multi-agent breadth-first research, at roughly **15x** token cost versus chat. That is a reason to use crews for parallel research with budgets, not a reason to start every workflow with a swarm.

OWASP agentic risks on a single loop first

Before adding agents, close the basics OWASP names for agentic applications: least-privilege tools, input validation on tool arguments, human approval on consequential writes, and logging sufficient to reconstruct goal drift. A crew multiplies these surfaces before you have one reliable loop.

Monthly operating rhythm (CIO)

1. **Metric movement versus baseline.** Report whether the workflow measure improved against the pre-registered baseline, because token usage alone is not an operating outcome. The baseline is the number the charter committed to, and the report is the accountability moment for it. Token usage is a platform metric that does not answer the question the board actually asked.
2. **Cost per successful case at P95.** Review tail unit economics monthly; agent loops that look affordable at median often fail finance at slow-case percentiles. The tail is where agent loops spend money invisibly, because the slow cases are the ones that loop, retry, and escalate. P95 is the percentile that turns a median claim into an honest one.
3. **Override and incident count.** Track how often humans corrected or stopped the agent, since rising overrides signal quality or policy decay before customers complain. Overrides are the ground truth the eval set cannot see, because they capture the cases where the operator's judgement diverged from the system's output. A rising count is the early warning that the loop is drifting in a way the scores miss.
4. **Tool permission change log.** Audit every allowlist expansion or scope change, because privilege creep is a leading indicator of OWASP-style tool misuse incidents. Each expansion is a new attack surface the agent can reach, and the change log is what lets the team review whether the expansion was justified. Unaudited creep is how a least-privilege loop becomes a broad-privilege one.
5. **Evaluation promotion set integrity.** Confirm nobody tuned against the frozen holdout, keeping promotion decisions honest as models and prompts evolve. The holdout's integrity is what makes the accuracy claim trustworthy month after month, and a single tuning incident can invalidate every later result. The check is the control that keeps the eval honest.
6. **Human gate staffing coverage.** Verify the gate role is staffed on rota including evenings if needed; failing this item alone is enough to pause autonomy expansion. An unstaffed gate is a

control that does not exist in the hours when consequential mistakes ship. Coverage is the difference between a control on the rota and a control in operation.

Only item 6 failing is enough to pause autonomy expansion.

Talk to one agent first. Ship with a crew only when the loop is boring.

That discipline is how mid-market firms avoid paying the Gartner cancellation tax while still building real agentic capacity.

References

1. Gartner, "Gartner Predicts 40% of Enterprise Apps Will Feature Task-Specific AI Agents by 2026, Up from Less Than 5% in 2025," press release, 26 August 2025. https://www.gartner.com/en/newsroom/press-releases/2025-08-26-gartner-predicts-40-percent-of-enterprise-apps-will-feature-task-specific-ai-agents-by-2026-up-from-less-than-5-percent-in-2025
2. Gartner, "Gartner Predicts Over 40% of Agentic AI Projects Will Be Canceled by End of 2027," press release, 25 June 2025. https://www.gartner.com/en/newsroom/press-releases/2025-06-25-gartner-predicts-over-40-percent-of-agentic-ai-projects-will-be-canceled-by-end-2027
3. Reuters, "Over 40% of agentic AI projects will be scrapped by 2027, Gartner says," 25 June 2025. https://www.reuters.com/business/over-40-agentic-ai-projects-will-be-scrapped-by-2027-gartner-says-2025-06-25/
4. Deloitte AI Institute, "The State of AI in the Enterprise: The Untapped Edge" (2026 edition); survey of 3,235 leaders, Aug-Sep 2025. https://www.deloitte.com/us/en/what-we-do/capabilities/applied-artificial-intelligence/content/state-of-ai-in-the-enterprise.html
5. Stanford Institute for Human-Centered Artificial Intelligence, "The 2026 AI Index Report," Economy chapter. https://hai.stanford.edu/ai-index/2026-ai-index-report/economy
6. OWASP GenAI Security Project, "OWASP Top 10 for Agentic Applications for 2026," 9 December 2025. https://genai.owasp.org/resource/owasp-top-10-for-agentic-applications-for-2026/
7. McKinsey & Company / QuantumBlack, "The State of AI: Global Survey 2025." https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai
8. Regulation (EU) 2024/1689 of the European Parliament and of the Council (EU AI Act). https://eur-lex.europa.eu/eli/reg/2024/1689/oj
9. Anthropic, "How we built our multi-agent research system." https://www.anthropic.com/engineering/multi-agent-research-system

Published by AIMonger . https://aimonger.com . Malta . Europe