

009

Restoring aircraft availability with sovereign AI agents

Agentic MRO for defense maintenance, repair and overhaul,
deployed over the systems you already run.

AUTHORS

Mohammed Hashmi · Bahaulddin Shammary · Yassine Say · Ali Akram

/ EXECUTIVE SUMMARY

THE PROBLEM

Defense MRO rarely underperforms because of skill or budget. It underperforms because the operating picture is fragmented across systems that were never designed to work together. Maintenance execution runs in the CMMS (computerized maintenance management system), supply runs in the ERP (enterprise resource planning system), the flying program runs in operations software, and technical records sit in a fourth system. Availability decisions are made on data that is hours or days old, then corrected by hand in a morning meeting.

WHAT WE FOUND

Across platforms and air forces, the same five failures appear. The fleet picture is stale by the time leadership sees it. Planning follows the calendar and reacts to breakdowns, with no forecast of each tail's needs. Supply and maintenance each hold their own version of what is in stock. Reliability data is filed and rarely analyzed. And where a contractor does the maintenance, that same contractor also grades its own performance. Twenty years of predictive-maintenance policy have not closed these gaps.¹ The reason, this paper argues, is that the data already exists and the systems of record already work. What is missing is the layer where the four pictures meet and the daily decisions about tails, parts, and bays are made.

WHAT WE PROPOSE

The proposal starts with what we call the Agent Twin, a live model of the operation that is built by observing the existing systems in shadow mode. Task agents then work on the twin and route every output to a human through Command Center, so that together they form an agentic decision layer while every system of record stays in place. The entire stack, including the AI models, runs inside the operator's perimeter with no data leaving it, and an air-gapped option is available where required. The engagement begins read-only, and the decision to deploy is made on the operator's own verified numbers.

KEY FACTS

- In fiscal year 2024 the US Department of Defense missed its mission capable rate goals for 42 of 45 aircraft types; 27 were more than ten points below goal.²
- The US F-35 fleet's mission capable rate fell from 67 percent to 44 percent between fiscal years 2021 and 2025; the UK fleet's rate in 2024 was roughly half of the Ministry of Defence's target.^{3, 4}
- Cannibalizations consume at least twice the labor of a normal repair. The last fleet-wide count, for fiscal years 1996 to 2000, recorded about 850,000 of them and 5.3 million additional maintenance hours across the US Air Force and Navy.⁵ No service has published a comparable count since.
- In 2018, F-35 squadrons cannibalized at more than six times the services' objective,¹⁴ and a 2025 audit found Navy maintainers still stripping grounded aircraft for parts the supply system could not deliver.²⁷
- Avionics no-fault-found (NFF) rates have been placed between 20 and 50 percent for decades, and avionics account for roughly three quarters of NFF occurrences.^{6, 7}
- The share of US Air Force depot aircraft delivered late against original targets rose from 31 percent in fiscal year 2019 to 73 percent in fiscal year 2024, with targets often revised after the fact.⁸
- Operating and support costs are about 70 percent of a weapon system's total life-cycle cost; the US alone spends nearly \$90 billion a year on weapon system maintenance.^{9, 1}
- Saudi Arabia budgets SAR 240 billion for the military sector in 2026. Since 2019 the United States has notified Congress of at least \$6.6 billion in spares, repair-and-return, and contractor logistics support for Royal Saudi Air Force aircraft alone.^{28, 29, 30, 31}

/ 1 THE STRUCTURE OF TRADITIONAL MRO

The starting point is the daily reality of a maintenance operation, which will be familiar to anyone who has run one. Every morning, the fleet status brief is assembled from CMMS extracts, supply system queries, and phone calls to the flight line, and in most operations it begins to age the moment it is printed.

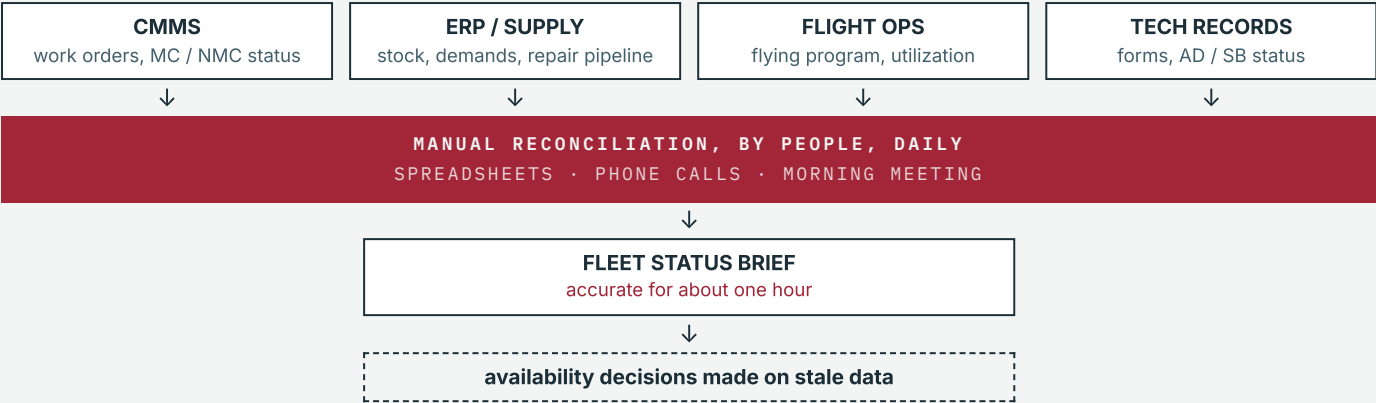


FIG. 1 INFORMATION FLOW IN TRADITIONAL MRO

Figure 1 shows why this happens. The only thing joining the four source systems is people, and the brief they produce is correct for roughly an hour, while every availability decision downstream is made against it.

The same structure appears at fleet scale in the public audit record. Audits of the largest logistics systems in defense aviation describe the same pattern. One is a central F-35 logistics system whose inaccurate or missing data has at times grounded flight-ready aircraft.¹⁰ Another is a US Army enterprise system that does not function where network connectivity is poor.¹¹ A third is a program office that reconciles availability data from three separate databases with three versions of the same spreadsheet.³

An aircraft is reported not mission capable, supply (NMCS), while the required part sits serviceable in a warehouse forty kilometers away, invisible because two inventory systems carry different stock records. A phase inspection is inducted before the parts kit is confirmed complete, which means the bay is occupied and the turnaround clock is running while the aircraft waits on supply. Cannibalization (CANN) keeps priority tails flying at the cost of a doubled maintenance burden, because every CANN action generates one removal and one installation, and it pushes the donor aircraft progressively deeper into the hangar.

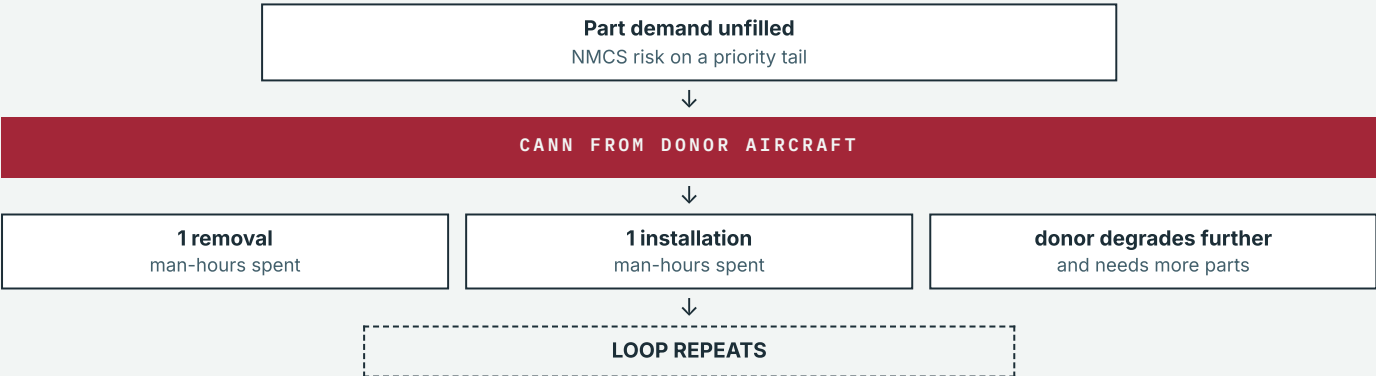


FIG. 2 THE CANNIBALIZATION LOOP

Figure 2 traces the mechanism. A normal repair is one removal and one installation, whereas a cannibalization is two of each, so it consumes at least twice the labor,⁵ and it replaces a broken part with a used one while adding a hole to the donor. The loop persists because it is a rational response to stock that nobody can see, and it tends to stop once that stock becomes visible across the systems involved.

The practice has not gone away. In fiscal years 1996 to 2000, the US Air Force and Navy reported about 850,000 cannibalizations and 5.3 million additional maintenance hours, the equivalent of more than 500 people working full time for five years,⁵ and no service has published a comparable count since. In 2018, F-35 squadrons cannibalized at more than six times the services' objective,¹⁴ and a 2025 audit found Navy maintainers stripping grounded aircraft for parts the supply system could not deliver.²⁷

At depot level the same structural problem appears at larger scale. Turnaround times are quoted at induction and confirmed only at delivery, because work scope grows after teardown and the operator learns about the growth when the schedule slips. In a 2026 review, the US Air Force's own auditors found that the additional time to complete unplanned work is not reflected in depot targets.⁸ They also found that targets for roughly half of delayed aircraft were revised after the maintenance was complete, to match the actual result.⁸ Over the five years to fiscal year 2024, the share of aircraft delivered late against original targets rose from 31 percent to 73 percent.⁸ The US

Navy's aviation depots, for their part, completed between 45 and 63 percent of aircraft on time over a six-year period.¹²

None of this is a competence problem. Organizational (O-level), intermediate (I-level), and depot (D-level) organizations are staffed with people who understand their aircraft in depth. The problem is rather that the maintenance enterprise generates large volumes of data across these systems and very little of it returns to the operation as timely decisions. The money involved is large. Operating and support costs run to about 70 percent of a weapon system's total life-cycle cost, and the US alone spends nearly \$90 billion a year sustaining them.^{9,1}

/ 2 FIVE RECURRING FAILURE MODES

The structure described in Section 1 fails in the same five places, regardless of platform or air force. This section takes each of the five in turn, with the public evidence for it. The pressures are the same where sustainment is being localized. Saudi Arabia, the largest military spender in the Middle East and the eighth largest in the world in 2025,²⁴ has set a target of localizing more than half of its military spending by 2030 and reached 24.89 percent by the end of 2024,²⁵ with maintenance, repair and overhaul capabilities among the priorities for that localization.²⁶

2.1 The operating picture is stale. Fleet status, work in progress, and parts posture live in separate systems that reconcile through people. Leadership therefore sees numbers that describe yesterday, and where the reporting is manual it is also unreliable. An inspection of one air arm's readiness reporting found that nine of ten sampled squadron commanders had reported inaccurately.¹³ The result shows up in the readiness numbers themselves. In fiscal year 2024 the United States missed its mission capable rate goals for 42 of 45 aircraft types.² The F-35 fleet's mission capable rate fell from 67 percent to 44 percent between fiscal years 2021 and 2025.³ In the United Kingdom, the F-35 fleet stood at roughly half of the Ministry of Defence's target in 2024.⁴

2.2 Planning is reactive. Scheduled maintenance follows the calendar and the flying program, and unscheduled maintenance follows failure. The utilization and removal history needed to forecast demand per tail number exists in the records, and in most operations it is never worked into the plan. Condition-based maintenance policy has required this forecasting since 2002, and yet the services still generally lack metrics to show whether it is happening.¹

2.3 Supply and maintenance hold different truths. NMCS hours accumulate because demand signals, stock records, and the repair pipeline are never resolved in one place. Rotable pools are sized on fleet averages while actual consumption varies sharply by tail, by season, and by mission profile. The cost is visible at fleet scale. In one seven-month period, for example, the F-35 fleet was unable to fly nearly 30 percent of the time because of spare parts shortages, with some 4,300 parts awaiting repair.¹⁴

2.4 Reliability data is collected and then archived. Removal histories contain the rogue units and the repeat offenders. Industry studies have placed avionics no-fault-found rates between twenty and fifty percent for decades,⁶ and each NFF event carries the full cost of a removal, a shipment, and a repair-shop induction that ends with a bench check serviceable finding.¹⁵ Avionics account for roughly three quarters of all NFF occurrences.⁷ The engineering signal exists in the data, but in most operations it is rarely extracted.

2.5 The vendor grades its own performance. Where the MRO provider also produces the availability reporting, the baseline belongs to the provider. Contract reviews become disagreements about whose numbers are correct, and the operator negotiates without an independent measurement of the service it is buying. Auditors have recommended verifying contractor performance data on performance-based logistics contracts for two decades.¹⁶ On the F-35, more than \$114 million of \$269 million in available incentive fees was paid over a period in which full mission capable and supply rates stagnated or worsened.³ Nearly half of the reconciled rates moved the contractor into a higher fee threshold.³

/ 3 WHY PREVIOUS FIXES HAVE NOT HELD

Three kinds of remedy have been applied to these failure modes, and each has a structural limit.

Policy without instrumentation. Condition-based maintenance directives date to 2002 and 2007. Twenty years on, the US Government Accountability Office found the services generally lacked metrics, goals, and a single owner for predictive maintenance, and made sixteen recommendations to establish them.¹ The policy, in other words, told the enterprise what to do but gave it no way to see whether it was being done.

Replacement systems. The most expensive failures in defense logistics IT have been attempts to replace the systems of record. The F-35's central logistics system, for example, accumulated years of audit findings on data accuracy and connectivity before it was redesigned,¹⁰ and the replacement inherited the same underlying data. Replacing a system does not reconcile the four pictures, since the new system simply becomes a fifth picture that still has to be reconciled with the others.

Point-in-time studies. Consulting reviews produce a baseline, a set of recommendations, and a report. The baseline is correct on the day it is delivered and ages at the same rate as the morning brief, so the improvement cycle restarts with the next study.

Each of these remedies addressed a system or a document. The decision layer, where the four pictures meet and where the daily choices about tails, parts, and bays are actually made, stayed as it was. The rest of this paper is concerned with that layer.

/ 4 WHAT TO LOOK FOR IN AN AGENTIC DECISION LAYER

Before describing a particular approach, it is worth stating what any agentic decision layer for defense MRO must satisfy. These are the conditions under which a security directorate, a maintenance commander, and an engineering authority can each say yes. They are written so that they can be used as requirements, for example in a tender or a statement of work.

TABLE 1 REQUIREMENTS FOR AN AGENTIC DECISION LAYER

#	REQUIREMENT	WHY IT MATTERS	WHAT TO ASK
1	Read-only first	The first phase must observe every system and write to none, so the operator can evaluate the layer with no risk to systems of record.	Can you deliver a verified baseline before we grant any write access?
2	Systems of record unchanged	The CMMS, ERP, flight ops, and records systems stay authoritative. The layer reconciles; it does not replace.	What do you write to, and when?
3	Sovereign by construction	Models, data, and compute inside the operator's perimeter: on-prem, in-country cloud, or air-gapped. Zero egress verifiable at the firewall.	Where does inference run, and what leaves the network?
4	A human boundary that is explicit	No agent signs off airworthiness, releases an aircraft, or overrides an engineering decision. Sensitive steps wait for a named person.	Show us the list of actions an agent can never take.
5	Full audit trail	Every agent action logged and attributable, so the security directorate can authorize and the operator can reconstruct.	Can we replay any decision from the log?
6	Stop and reverse	Rollback for any change; a kill switch for any agent or all of them.	How long does a full stop take, and who can trigger it?
7	Inventory of agents	A registry of what exists, what each agent touches, and what it may do.	Give us the registry export.
8	Evaluated on our history before go-live	Agents scored against the operator's own past cases, with a threshold agreed in advance.	What was the score, on what cases, and who set the threshold?
9	Independent baseline	Availability, NMCS, CANN, and turnaround measured independently from the operator's own data.	Whose numbers are we measured against?
10	Exit by design	The model of the operation, the datalake, and the evaluation sets remain the operator's property when the engagement ends.	What do we keep if we stop?

Agentic MRO leaves the CMMS, the ERP, and the engineering organization in place and adds a layer that observes them continuously. The architecture has four levels, shown in Figure 3.

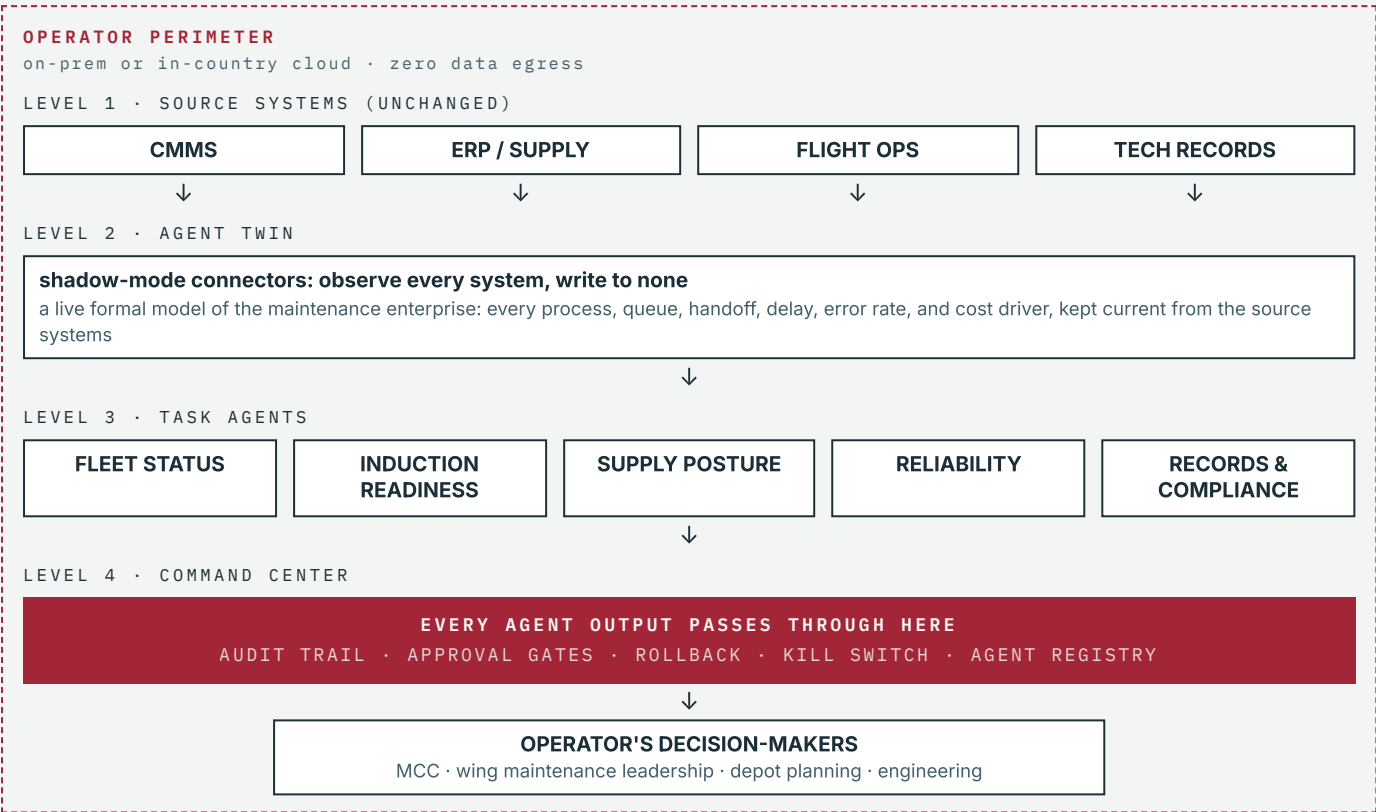


FIG. 3 AGENT TWIN ARCHITECTURE FOR MRO

Compared with Figure 1, the band of manual reconciliation has been replaced by the twin, and the people have moved from doing the reconciling to making the decisions.

Level 1 is the operator's existing systems, which continue to operate exactly as they do today.

Level 2 is the Agent Twin. The name borrows from engineering, where a digital twin is a virtual replica of a physical asset that is kept in step with it by a stream of data.¹⁷ An Agent Twin applies the same idea to the whole maintenance organization. It is a live, machine-readable model of how the maintenance enterprise actually runs, covering every process, queue, handoff, delay, error rate, and cost driver, and it is kept current from the source systems. It is built in shadow mode, meaning the connectors read from every system and write to none. The twin is where agents are designed, tested against the operator's own history, and measured, before any of them act on a live case.

Level 3 is the task agents that work on top of the twin. **Level 4** is Command Center, through which every agent output reaches a human decision-maker.

WHAT THE AGENTS DO NOT DO
No agent signs off airworthiness, releases an aircraft, or overrides an engineering decision. No agent writes to the CMMS, the ERP, or the technical records. No agent executes a step classified as sensitive without a named person clearing it at an approval gate. A qualified individual signs for every release today, and that does not change. The agents prepare the analysis and route it to the person who holds the authority.

Table 2 lists the task agents in an MRO deployment, each with an example of what the responsible person receives.

TABLE 2 THE FIVE TASK AGENTS

AGENT	WHAT IT DOES	EXAMPLE OUTPUT (NOTIONAL)
Fleet status	Reconciles CMMS status codes, supply records, and the flying program into a single live operating picture, and replaces the manually assembled morning brief with a continuously current one.	"Tail 207 is coded NMCS in the CMMS, but the demanded part was receipted at 04:10. Status can be updated; estimated return to MC 11:00."
Induction readiness	Before an aircraft or component is inducted into a bay, verifies that the parts kit, tooling, technical data, and qualified manpower are available, and flags any induction that would begin in a wait state.	"Phase inspection for tail 214 is scheduled Monday. The kit is three items short with earliest arrival Wednesday. Recommend inducing tail 219 first."
Supply posture	Forecasts parts demand per tail number from actual utilization and removal history, identifies NMCS risk before it grounds an aircraft, and records every CANN action with its man-hour cost and its effect on the donor.	"Four tails will exceed the hydraulic pump removal interval within 30 days; stock covers two. Order now to avoid NMCS."
Reliability	Mines removal and repair history for rogue units, repeat failures by ATA chapter, and NFF patterns, and converts them into specific engineering investigation requests.	"Serial 8842 has been removed five times in nine months and bench-checked serviceable each time. Recommend wiring inspection on the tails it was fitted to."
Records and compliance	Audits technical records for completeness and tracks airworthiness directive (AD) and service bulletin (SB) incorporation status across the fleet on a continuous basis.	"AD 2026-04 is incorporated on 31 of 34 tails. The three outstanding are due within 40 flight hours."

/ 6 A WORKED EXAMPLE: ONE TAIL, ONE PART, 72 HOURS

Table 3 below traces a single NMCS event through a traditional operation and through the agentic layer. The base, tail numbers, part, and times are notional, while the sequence of events on the traditional side is the one described in Section 1.

TABLE 3 NOTIONAL 72-HOUR TRACE OF A SINGLE NMCS EVENT

Data are notional. The traditional-path labor figure assumes 2.5 hours per pump removal or installation.

TIME	TRADITIONAL PATH	AGENTIC PATH
Mon 02:40	Tail 207 lands with a hydraulic pump write-up. Line technician confirms failure, opens a work order, raises a demand. Base stock: zero.	Same. The twin registers the work order and the demand within minutes of entry.
Mon 04:10	Nothing. The demand sits in the supply system awaiting the next replenishment cycle.	Supply posture agent matches the demand against every stock record it can see, including the regional depot store on a second inventory system, and finds a serviceable pump 40 km away. It drafts a lateral transfer request and routes it to the MCC (maintenance control center) via Command Center. The step is logged.
Mon 06:30	Morning brief reports 207 NMCS, estimated four days. Tail 207 is a priority tail for Thursday's tasking.	Fleet status agent's brief shows 207 NMCS with a transfer request pending approval and an estimated return to MC of 14:00 if approved. The MCC officer approves at the gate at 06:15.
Tue 08:00	With Thursday approaching and no pump, the MCC authorizes a CANN from tail 231, which is already down for an unrelated avionics fault. Two technicians remove the pump from 231 (2.5 hours) and install it on 207 (2.5 hours).	Pump arrived Monday 10:30. Installed by 13:20. Tail 207 returned to MC at 14:00 Monday. No cannibalization. Tail 231's status is unchanged.
Wed 11:00	The original demand is filled from the regional depot store, where the pump had been serviceable since Sunday. It is now installed on 231 (2.5 hours), which still has its avionics fault.	Reliability agent notes that the pump removed from 207 is its third removal in eleven months, twice bench-checked serviceable, and drafts an engineering investigation request for the pump's electrical interconnect. Routed to engineering. Logged.

Thu 06:00	Tail 207 available. Tail 231 still NMC, now with a 48-hour gap in its records for the pump swap. Total labor spent on the pump: 7.5 hours across three installations and one removal. CANN recorded, or not.	Tail 207 available since Monday. Tail 231 untouched. Total labor: 2.5 hours. One investigation request open on a suspected rogue unit. Every step attributable in the audit trail.
-----------	--	--

Nothing in the agentic path required a new system, a sensor, or an agent taking an action. The difference was that the second inventory system was visible, the transfer was proposed before the morning brief, and a person approved it at the gate. The reliability finding came from data that already existed in the removal history.

/ 7 COMMAND CENTER

Every agent operates under Command Center, which is where the operator's team watches agent activity, approves sensitive actions, and stops anything that needs stopping, as traced in Figure 4.

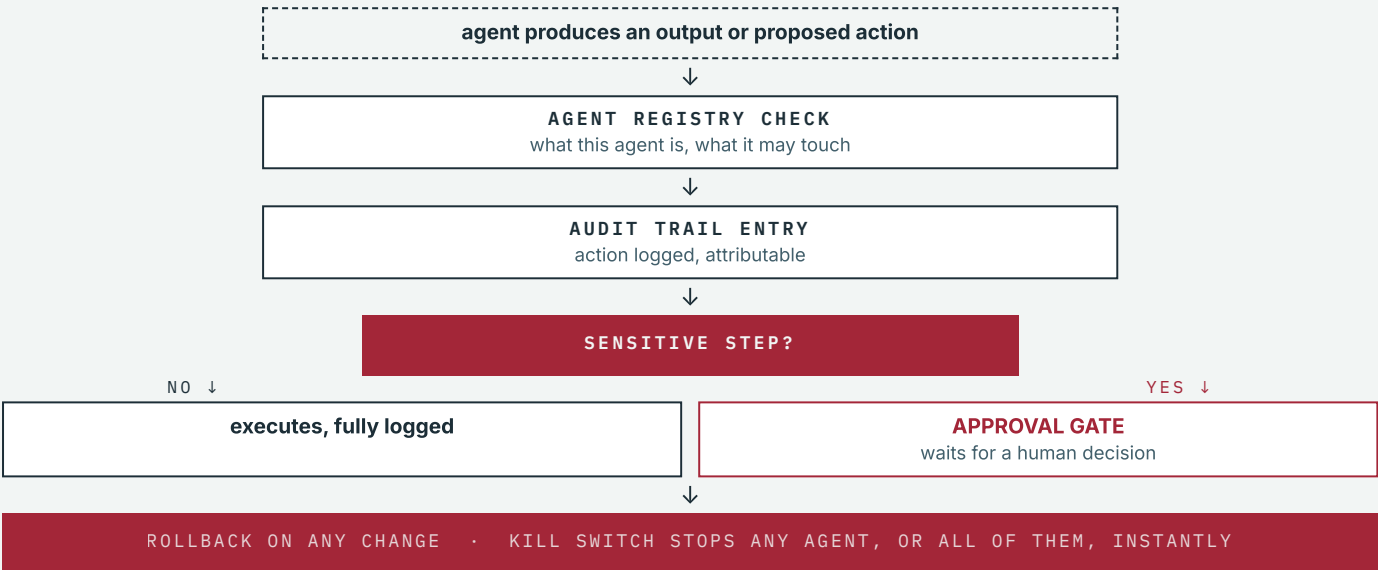


FIG. 4 COMMAND CENTER GOVERNANCE PATH

Every output is checked against the registry and written to the audit trail before anything happens. Routine steps then execute, fully logged, while sensitive steps stop at the approval gate for a named person, and rollback and the kill switch sit underneath the whole path.

Command Center rests on five controls. An audit trail logs every agent action and makes it attributable. Approval gates hold sensitive steps until a human decides. Rollback reverses any change an agent makes. A kill switch stops any agent, or all of them, instantly. An agent registry records what exists, what each agent touches, and what it is permitted to do. In a defense environment these controls are what allow the security directorate to authorize the system in the first place.

They are also, in practice, what the public governance frameworks a defense operator is measured against ask for. Table 4 maps each control to the framework language a reviewer will look for.

TABLE 4 COMMAND CENTER CONTROLS MAPPED TO GOVERNANCE FRAMEWORKS

CONTROL	NIST AI RMF 1.0 ¹⁸	NATO PRINCIPLES OF RESPONSIBLE USE ¹⁹	UK MOD JSP 936 ²⁰	ISO/IEC 42001 ²¹
Audit trail	Govern; Measure (traceability of decisions)	Explainability and traceability	Understanding; Responsibility	Event logging and AI system documentation
Approval gates	Manage (human oversight of consequential outputs)	Responsibility and accountability; Governability	Human-centricity; Responsibility	Human oversight in operation of AI systems
Rollback	Manage (response and recovery)	Governability; Reliability	Reliability	Incident response and change control
Kill switch	Manage (ability to disengage)	Governability (ability to disengage or deactivate)	Reliability; Human-centricity	Operational controls for AI system use
Agent registry	Map (inventory of AI systems and contexts)	Responsibility and accountability	Responsibility	AI system inventory, roles, and resources

/ 8 TRADITIONAL AND AGENTIC MRO COMPARED

TABLE 5 TRADITIONAL AND AGENTIC MRO COMPARED

DIMENSION	TRADITIONAL MRO	AGENTIC MRO
Operating picture	Assembled manually each morning, ages immediately	Reconciled continuously across all source systems
Maintenance planning	Driven by calendar and failure events	Forecast per tail from utilization and removal history
Supply posture	Rotable pools sized on fleet averages	Demand signals tied to the maintenance forecast
Reliability engineering	Data archived, reviewed episodically	Removal and NFF patterns worked continuously
Depot turnaround	Quoted at induction, confirmed at delivery, targets revised after the fact	Induction readiness verified before the bay is occupied; scope growth visible as it happens
Performance baseline	Produced by the MRO vendor	Measured independently from the operator's own data
Improvement cycle	Periodic consulting studies	Continuous measurement against a verified baseline
Validation	Pilot on live cases	Evaluated on the operator's historical cases before go-live
Governance	Distributed across contractors and systems	Command Center, inside the operator's perimeter
Exit	Custom build stops working when the vendor leaves	Model, datalake, and evaluation sets remain the operator's property

/ 9 LIMITATIONS: WHERE THIS DOES NOT HELP

An agentic decision layer is not a remedy for every sustainment problem, and a reader should know its edges before its benefits.

- **It needs a digital trail.** The twin is built from records. Where work is done and not recorded, or recorded on paper that is never transcribed, the twin cannot see it. The Recon phase, the read-only mapping phase described in Section 11, measures the size of this gap, but it cannot close it.
- **Source data is rarely clean, and the twin treats that as a finding.** Status codes are entered late, removal reasons are free text, and inventory records disagree with the shelf. Recon measures where the records are wrong and by how much, and that measurement becomes part of the baseline the operator receives. Agents are therefore designed for the data as it actually is.
- **It does not predict rare events.** Analysis of AI for predictive maintenance in defense finds that it improves failure analysis case by case but cannot forecast events with little history behind them.²² The supply posture agent

forecasts recurring demand, and it does not foresee the first failure of a new mode.

- **It is not sensor-based prognostics.** The agents work on maintenance records and do not use health-monitoring data from the aircraft. Records are the right starting point for fleets where onboard health data is sparse or proprietary, and the approach complements condition-based maintenance programs that use such data.
- **Integration takes the time the operator's security process takes.** Even read-only access to a CMMS on an accredited network requires approvals. Recon starts that clock on its first day, but no vendor can shorten it unilaterally.
- **The operation changes, and the twin must follow it.** A new supply process, a new variant, or a reorganized depot changes the model. Telemetry from production flows back into the twin so that it is re-fitted continuously, and when the engagement ends, the model, the datalake, and the evaluation sets stay with the operator.
- **Adoption is a leadership decision, not a technical one.** The morning brief is also where a wing aligns. Replacing the reconciling with a live picture changes how the meeting works, and the people who assembled the old brief need to own the new one.

/ 10 THE AGENT TWIN

The Agent Twin is built for a type of institution, one whose systems of record must run for decades, whose workflows are its own, and where a wrong decision has real consequences. Plug-and-play software has consistently failed in such places. Governments and heavy-operations industries are where these institutions are found, and defense MRO is one instance among several. The twin is what lets one method serve all of them. Every deployment starts by mapping the institution's own processes, rules, and systems, so the agents that follow apply that institution's policy and nothing else.

Underneath the twin, Agentic MRO is a hybrid architecture developed across successive sovereign deployments, and each method in it does the work it is suited to. Language-model agents do the coordinating and the reading of unstructured records, which is what they are good at.²³ Where a decision can be computed exactly, such as a schedule, a stock level, or the simulated effect of a change, classical optimization, scheduling, and simulation methods compute it, because those methods give guarantees a language model cannot. A deterministic policy layer, which returns the same verdict for the same input every time, governs what each agent may observe, propose, and execute. The language model proposes, and the policy layer decides what it may do.

Every deployment carries a formal model of the institution it serves, and every agent is validated against the operator's historical case archive before go-live, using an evaluation methodology built for environments where an incorrect output has operational consequences. Each agent is scored on the decisions it would have made against the operator's own past cases. An agent that does not clear the threshold agreed with the operator for its task does not deploy. Sustaining this stack requires engineering depth across model serving inside air-gapped networks, integration with systems that have been in service for decades, formal process modelling, and evaluation science, which is why the capability has not emerged from commercial plug-and-play products.

The approach is in production today at a state-owned national registry, where it processes four thousand cases per day on an architecture built for forty thousand. The discipline there is the one proposed here: map the operation, verify the baseline, deploy under Command Center, and measure results against the operator's own data.

/ 11 HOW AN ENGAGEMENT STARTS

The engagement runs in three phases: Recon, Deploy, and Expand. It begins with a shadow-mode mapping phase that gives a maintenance leadership team its own verified numbers before any deployment decision is required, and the deployment decision is then made on those numbers.

TABLE 6 THE THREE ENGAGEMENT PHASES

PHASE	WHAT HAPPENS	WHAT THE OPERATOR RECEIVES
1 Recon	Shadow-mode connection. The twin is built and the operation mapped: flows, queues, delays, error rates, cost drivers. Agents are designed, built, and evaluated inside the twin against historical cases. Read-only throughout.	An independent baseline (availability, NMCS, CANN, turnaround, NFF) from its own data; a map of where to deploy, how, and what it returns, in plain numbers.
2 Deploy	The same agents move from the twin into production under Command Center. Telemetry measures every action against the Recon baseline.	Agents in production; results measured against the baseline it already verified.
3 Expand	Additional agents and sites on a datalake the operator owns.	A twin it owns, with exit by design.

/ 12 CONCLUSION

Defense MRO underperforms at the decision layer, where four accurate but separate pictures are reconciled by people, once a day, into a brief that is stale within the hour. The remedies tried so far have replaced systems, issued policy, or produced reports, and none has touched that layer. An agentic decision layer observes the existing systems, holds a live model of the operation, and routes every output to a named person under audit and approval. It runs inside the operator's perimeter, and no agent ever holds authority over an aircraft. The place to begin, this paper has argued, is with the operator's own numbers.

/ ACRONYMS

AD	Airworthiness directive	MC / FMC	Mission capable / fully mission capable
ATA	Air Transport Association (chapter numbering)	MCC	Maintenance control center
CANN	Cannibalization	MRO	Maintenance, repair, and overhaul
CMMS	Computerized maintenance management system	NFF	No fault found
ERP	Enterprise resource planning	NMC / NMCS	Not mission capable / not mission capable, supply
O / I / D-level	Organizational / intermediate / depot level maintenance	SB	Service bulletin

/ REFERENCES

1. US Government Accountability Office, GAO-23-105556, *Military Readiness: Actions Needed to Further Implement Predictive Maintenance on Weapon Systems*, December 2022. <https://www.gao.gov/products/gao-23-105556>
2. US Government Accountability Office, GAO-25-108104, *Military Readiness: Implementing GAO's Recommendations Can Help DOD Address Persistent Challenges Across Air, Sea, Ground, and Space Domains*, March 2025. <https://www.gao.gov/products/gao-25-108104>
3. US Government Accountability Office, GAO-26-108113, *F-35 Sustainment: Actions Needed to Ensure Updated Strategy Improves Persistent Readiness Challenges*, June 2026. <https://www.gao.gov/products/gao-26-108113>
4. UK National Audit Office, *The UK's F-35 capability*, HC 989, July 2025. <https://www.nao.org.uk/wp-content/uploads/2025/07/the-uks-f-35-capability.pdf>
5. US Government Accountability Office, GAO-01-693T, *Military Aircraft: Cannibalizations Adversely Affect Personnel and Maintenance*, May 2001. <https://www.gao.gov/assets/gao-01-693t.pdf>
6. Han, Park and Lee, "Intermittent failure in electrical interconnection of avionics system," *Reliability Engineering and System Safety* 185 (2019). <https://doi.org/10.1016/j.ress.2018.12.016>
7. Steadman, Berghout, Olsen and Sorensen, "Intermittent Fault Detection and Isolation System," IEEE AUTOTESTCON 2008, pp. 37-40. <https://doi.org/10.1109/AUTEST.2008.4662580>
8. US Government Accountability Office, GAO-26-107890, *Air Force Readiness: Actions Needed to Address Depot Maintenance Delays and Staffing Challenges*, May 2026. <https://www.gao.gov/products/gao-26-107890>
9. US Government Accountability Office, GAO-26-108140, *Weapon System Sustainment: DOD Identified Critical Cost Growth, and the Army Should Take Action to Yield Cost Savings*, April 2026. <https://www.gao.gov/products/gao-26-108140>
10. US Government Accountability Office, GAO-20-316, *Weapon System Sustainment: DOD Needs a Strategy for Re-Designing the F-35's Central Logistics System*, March 2020. <https://www.gao.gov/products/gao-20-316>
11. US Government Accountability Office, GAO-21-313, *Defense Logistics: Army Should Ensure New System Operates in All Situations and Soldiers Complete Training*, April 2021. <https://www.gao.gov/products/gao-21-313>
12. US Government Accountability Office, GAO-20-390, *Military Depots: The Navy Needs Improved Planning to Address Persistent Aircraft Maintenance Delays*, June 2020. <https://www.gao.gov/products/gao-20-390>
13. US Department of Defense Office of Inspector General, DODIG-2018-141, *United States Marine Corps Aviation Squadron Aircraft Readiness Reporting*, August 2018. <https://www.dodig.mil/reports.html/Article/1598583/>
14. US Government Accountability Office, GAO-19-321, *F-35 Aircraft Sustainment: DOD Needs to Address Substantial Supply Chain Challenges*, April 2019. <https://www.gao.gov/products/gao-19-321>
15. Khan, Phillips, Jennions and Hockley, "No Fault Found events in maintenance engineering Part 1: Current trends, implications and organizational practices," *Reliability Engineering and System Safety* 123 (2014). <https://doi.org/10.1016/j.ress.2013.11.003>
16. US Government Accountability Office, GAO-05-966, *Defense Management: DOD Needs to Demonstrate That Performance-Based Logistics Contracts Are Achieving Expected Benefits*, September 2005. <https://www.gao.gov/products/gao-05-966>
17. Grieves and Vickers, "Digital Twin: Mitigating Unpredictable, Undesirable Emergent Behavior in Complex Systems," in *Transdisciplinary Perspectives on Complex Systems*, Springer, 2017. https://doi.org/10.1007/978-3-319-38756-7_4
18. National Institute of Standards and Technology, *Artificial Intelligence Risk Management Framework (AI RMF 1.0)*, NIST AI 100-1, January 2023. <https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-1.pdf>
19. NATO, *Summary of NATO's revised Artificial Intelligence Strategy*, July 2024. https://www.nato.int/cps/en/natohq/official_texts_227237.htm
20. UK Ministry of Defence, *JSP 936: Dependable Artificial Intelligence in Defence, Part 1: Directive*, November 2024. <https://www.gov.uk/government/publications/jsp-936-dependable-artificial-intelligence-ai-in-defence-part-1-directive>
21. ISO/IEC 42001:2023, *Information technology: Artificial intelligence: Management system*. <https://www.iso.org/standard/42001>
22. RAND Corporation, *Understanding the Limits of Artificial Intelligence for Warfighters, Vol. 3: Predictive Maintenance*, RRA1722-3, 2024. https://www.rand.org/pubs/research_reports/RRA1722-3.html
23. Wang, Ma, Feng et al., "A survey on large language model based autonomous agents," *Frontiers of Computer Science* 18, 186345 (2024). <https://doi.org/10.1007/s11704-024-40231-1>
24. Stockholm International Peace Research Institute, *Trends in World Military Expenditure, 2025*, SIPRI Fact Sheet, April 2026. https://www.sipri.org/sites/default/files/2026-04/2604_milex_2025.pdf
25. General Authority for Military Industries (Saudi Arabia), "Localizing the sector." <https://www.gami.gov.sa/en/achievements/localizing-sector>
26. US International Trade Administration, *Saudi Arabia Country Commercial Guide: Defense and Security*, May 2026. <https://www.trade.gov/country-commercial-guides/saudi-arabia-defense-security>
27. US Government Accountability Office, GAO-25-107468, *Weapon System Sustainment: DOD Can Improve Planning and Management of Data Rights*, September 2025. <https://www.gao.gov/products/gao-25-107468>
28. Ministry of Finance (Saudi Arabia), *Budget Statement, Fiscal Year 2026*, December 2025, pp. 40, 85, 89. https://www.mof.gov.sa/en/budget/2026/BudgetStatementDocs/Eng_2026.pdf
29. US Defense Security Cooperation Agency, "Kingdom of Saudi Arabia: F-15 Sustainment," Transmittal No. 25-103, February 3, 2026. <https://www.dsca.mil/Press-Media/Major-Arms-Sales/Article-Display/Article/4396586/kingdom-of-saudi-arabia-f-15-sustainment>
30. US Defense Security Cooperation Agency, "Kingdom of Saudi Arabia: System Logistics and Sustainment Support," Transmittal No. 24-43, July 23, 2024. <https://www.dsca.mil/press-media/major-arms-sales/kingdom-saudi-arabia-system-logistics-and-sustainment-support>
31. US Defense Security Cooperation Agency, "Kingdom of Saudi Arabia: Aircraft Follow-on Support and Services," Transmittal No. 18-21, May 24, 2019. <https://www.dsca.mil/press-media/major-arms-sales/kingdom-saudi-arabia-aircraft-follow-support-and-services>

AGENT 01

Mohammed Hashmi

CHIEF TECHNOLOGY OFFICER

Mohammed is the Chief Technology Officer of 009. Before 009 he was Chief Product Officer at Watad, where he was on the founding team and led the build of AridOS, an agentic operating system for the construction industry in the Gulf. He spent a decade as an engineer at some of the world's most demanding technology companies. At Apple he worked in the Special Projects Group and then as an iPhone systems engineer in Cupertino. He went on to systems engineering at Uber, R&D at Videri, and technical program management at Lyft. He is also a technology partner at Flow Ventures and an angel investor in Riyadh. He holds a degree in Electrical and Computer Engineering from the University of Waterloo, where he won a President's Scholarship.

AGENT 02

Bahaulddin Shammary

FOUNDING ENGINEER

Bahaulddin is a founding engineer at 009, where he builds the verification and governance behind the Agent Twin. Before 009 he was a Lead AI Engineer at Watad, building AridOS, a multi-agent system for construction procurement and credit. Before that he led the production hardening of a multi-agent platform for network operations at OpenText and was a Senior AI Research Scientist in Dell's Office of the CTO, where he filed multiple patents on AI deployment and security. He worked with the Open Source Security Foundation, part of the Linux Foundation, to publish its guide to securing machine learning pipelines. He also co-authored research published by the Association for Computational Linguistics on Arabic dialect intent detection, raising accuracy on Tunisian Arabic from 43 to 94 percent. He holds an MSc in Artificial Intelligence from Munster Technological University, Ireland.

AGENT 03

Yassine Say

FOUNDING ENGINEER

Yassine is a founding engineer at 009, where he leads the work on graph neural networks for process discovery. Before 009 he was a Lead AI Engineer at Watad, where he built the layer that registers, permissions, and orchestrates AI agents across AridOS. Before that he led the team behind Rooceptionist at Vendoroo, a voice agent that handled half a million calls for more than 330 property management companies and resolved about 70 percent of them with no human involved. He spent six years at Dell Technologies, rising from intern to Data Science and Machine Learning Engineer Advisor. He holds a master's in artificial intelligence from Mohammed V University in Rabat.

AGENT 04

Ali Akram

FOUNDING ENGINEER

Ali is a founding engineer at 009, where he leads process discovery and built the Petri net layer of the Agent Twin, the part that turns an institution's event logs into a working model of how its operation actually runs. He has spent seven years building machine learning and AI systems. Before 009 he was founding AI engineer at Focal, a San Francisco startup building AI tools for financial advisors, and before that he built LLM systems for stock analysis at TransCrypts. At Cambridge Cognition he built speech models that detect signs of cognitive impairment, and co-authored a study of automated speaker verification in Alzheimer's disease clinical trials, published by the Association for Computational Linguistics in 2023. He holds a degree in Systems Design Engineering from the University of Waterloo.

/ ABOUT 009 LABS AND THE AGENTS

009 Labs builds sovereign AI agents deployed in the institutions that run nations: governments, ministries, state institutions, and heavy-operations industries across defense, energy, utilities, ports, rail, registries, and customs. The company's method is the Agent Twin, a working model of an institution built by reading its systems from the inside, and Command Center, which keeps the institution's own team in control of every agent. Everything runs on-premises or in the client's in-country cloud, with no data leaving the perimeter and exit by design.

Start with your own numbers.

To see your fleet's availability, NMCS, cannibalization, and turnaround numbers measured independently from your existing systems, request a Recon briefing. Recon is read-only, fixed fee, and the numbers are yours whether or not you proceed.

[REQUEST A RECON BRIEFING →](#)

[009labs.ai](#) · [009labs.ai/contact](#) · © 2026 009 Labs