

Earning a Seat on the Staff

By: [Ben Van Roo](#), CEO and Co-Founder

The Staff Has Absorbed New Machines Before

New machines join the staff by meeting its standards, not by lowering them.

The modern staff is itself a technology, invented to solve a coordination problem. By the middle of the nineteenth century, war had outgrown what one commander could hold in his head. Prussia's answer was an institution: trained officers, common methods, and the new machines of the age wired into the process. [Moltke was among the first senior officers to grasp what railroads meant for moving armies](#), and he prized the telegraph because it could [“direct separated parts of the army according to a single will”](#). He ran the opening moves of two wars from an office in Berlin. Every army in the world copied the model, which is why every army in the world has a staff.

The staff has absorbed new machines ever since, each on the same terms. The newest arrived faster than any predecessor. The Department of War's GenAI.mil grew from [roughly 80,000 users at its December 2025 launch](#) to [nearly 1.7 million by July 2026](#), with more than 100,000 custom agents built along the way and [five of the six services designating it their enterprise platform](#). That chapter is a success story, and it puts the next question on the table sooner than anyone planned: what does it take for these systems to join the work staffs do together? Generative AI is the first machine in the lineage that can participate in the work, not just accelerate one person inside it.

The question has a boundary. In sensor processing, autonomy, and electronic warfare, model performance is often the constraint and better models are the answer. Staff work is different: a team sport played through artifacts, orders, estimates, overlays, and matrices, built by many hands and checked against each other at the end, when contradictions are most expensive to fix. An assistant in every seat makes each officer faster, and the drafts still meet the way they always have. The failures that cost a staff the most do not happen inside anyone's session. They happen between sessions. Getting from an assistant in every seat to a staff that works with agents is an architecture question, and architecture questions have answers: five requirements, three commitments that make a shared surface safe, and a way for leaders to evaluate any system that claims a seat.

The Private Session Is a Structural Ceiling

A private AI chat reaches one step of a process built from twelve.

Follow one appendix to an operations order through the process. A plans officer drafts it, the chief of plans marks it up, the fires and sustainment cells coordinate on it because it makes assumptions

about both, and it goes into the orders crosswalk to be checked against the base order and every other annex. It gets briefed, changed, and approved, and at that moment the document acquires legal and moral weight. It directs people to do dangerous things, and a specific human being is accountable for its contents.

A private assistant reaches one step of that. The plans officer gets a better first draft, faster, and the moment the draft leaves the session the assistant is out of the loop. It cannot see the fires cell's objection or the Tuesday change to the sustainment assumption, and it cannot check the revised appendix against the base order, because the base order lives in another system and the other annexes live in eleven other private sessions. Provenance evaporates at the boundary, and when the crosswalk finds a contradiction, no one can say where the bad assumption entered.

Four characteristics of military work break that model. The unit of work is the shared artifact, not the conversation. The work carries attribution, because an anonymous paragraph pasted out of a private chat is a liability in a system built on accountability. The work outlives any individual's attention, because battle rhythm runs in shifts and personnel rotate on cycles that bleed institutional knowledge. And the context is governed, because who may see which source, at which classification, with which caveat, is load-bearing. None of these is a technology preference; they are what command authority looks like implemented as an organization, and a private assistant cannot meet them.

Five Requirements Before an Agent Joins the Work

Each is the price of admission to a process that ends in an authorized decision and a consequence.

The four characteristics above produce five architectural requirements.

Shared, persistent sessions. Humans and agents operate in the same workspace, with roles, and the session survives its participants. A planner who joins on day four inherits what the agents did, which sources they used, and which changes reviewers accepted. Plans decay, and what decays first is the shared understanding the planning built. [RAND reports that a typical Army unit loses 20 to 30 percent of its personnel each year](#). At that rate, a persistent session is how the next team inherits the assumptions and the decision record rather than just the PDF.

Artifacts as the unit of work. An agent's output is a versioned draft in the shared workspace, subject to the same review as a human draft, and an approved version does not stay stranded there. It leaves as a published order, a ticket, a record update, or a call into a connected system, without severing the record of its approval and attribution. This is the requirement most collaboration tooling never reaches: a system that produces a better document has improved drafting, not staff work.

Attribution of every action. When an agent modifies an artifact, retrieves a source, or carries out a workflow step, the action is logged, attributed, and inspectable. Attribution is what makes an agent's

contribution admissible in a chain of review. Without it, the annex contains an anonymous paragraph that will direct action.

Governed context. The material an agent may use is scoped by permission, provenance travels with content, and the organization sets the steps where the agent proposes and a person decides. Classification is the hardest case: an agent may propose a marking on content it writes and may not set one, because classification is an authorized human determination.

Persistence across the seams. This requirement comes not from the nature of staff work but from where the work happens. Sessions, context, and artifacts have to survive shift changes, team turnover, and the loss of the network, because the seams in national security work include the network itself. [The case for designing disconnected-first has been made in this series](#), and it applies with full force here. A capability that requires a cloud connection is a garrison capability.

The capability side of this argument gets a longer treatment in [The Mission is Multiplayer](#). This paper stays on the requirements, because requirements are what a leader can act on.

The Coordination Layer Is Where the Gains Live

The estimates go stale, and the crosswalk catches the contradictions at the end, after they have propagated.

A planning effort is structurally a coordination machine: specialists each maintaining a running estimate of their own domain, collectively producing one integrated product under time pressure, with a base order and a dozen or more annexes that all have to agree. That last clause is where planning efforts go to die. Doctrine requires every staff section to maintain a running estimate and an orders crosswalk before publication. In practice the estimates go stale within days, because the officers who own them are busy producing the artifacts in front of them, and the crosswalk happens at the end, after the contradictions have propagated. Army researchers observing brigades at the combat training centers in 1998 found staffs frequently deficient in synchronizing the battle. The Army's 2023 planning handbook opens by saying it exists to reverse trends observed at those same centers. Twenty-five years separate the two documents. They describe the same problem.

The pattern repeats on the watch floor, where the most dangerous hour is the hour after shift change, and through sustainment, where every echelon has a human reconciling the reports from below, so the picture a theater commander sees describes a force that existed days ago. GAO documented the Army and Marine Corps [losing visibility over equipment during the Desert Storm buildup](#), and [the same problems a decade later in Iraq](#). The optimization models for these problems have existed for decades. Some of them are older than the officers running the process. They can only operate on the state they are given, and the shared state does not exist as shared state. It exists as thousands of documents describing the force at different times, held by different people, reconciled by hand.

Agents that satisfy the five requirements attack that layer directly, and the mechanism is worth stating plainly, because “AI makes staffs faster” is a claim that has been made and broken before. The staff process itself does not change, and doctrine survives contact largely intact. What changes is which parts of it consume human hours. One agent maintains the running estimates as the artifacts change, because it does not have to choose between maintaining the estimate and producing the annex. Another checks every artifact against every other on every revision, so the contradiction between the fires appendix and the airspace annex surfaces while the work is still forming. Humans keep the decisions the organization designates as consequential: intent, course-of-action selection, risk acceptance, classification, and release.

There is a precedent for what fixing a coordination layer is worth. In 1995 the Army and RAND launched Velocity Management against chronic delay in ordering and shipping repair parts, a process averaging just over 22 days. It bought no trucks and built no depots. It defined the process end to end and re-engineered it, and [By December 1997, order-and-ship times in the continental United States had fallen roughly 50 percent at every percentile](#). That campaign ran on periodic measurement. Agents holding shared state could run the same discipline continuously, on every requisition.

A Shared Surface Makes a Bad Assumption Travel Faster

The mechanism that surfaces a contradiction on day three propagates one on day three.

The strongest objection to everything above is that a private session traps value and also contains damage. An agent that pulls a stale supply figure into one officer's draft has corrupted one draft. The same agent working in the shared order has corrupted the order, and eleven people are now planning against it. Any staff officer who has watched a bad number survive four briefings will raise this objection before the second slide, and they are right to.

Attribution does not answer it. Attribution tells you who, not whether. Knowing which agent introduced the figure is a post-mortem, and the orders brief already happened. Nothing in the five requirements establishes that the figures an agent pulled are correct, and a system that governs authorship perfectly while governing correctness not at all is a faster way to be wrong.

Three architectural commitments make the trade worth taking, and a system without them should not be trusted with a shared artifact. First, proposed state and accepted state are different states wherever policy designates a change as consequential. Routine changes commit within an agent's permissions and delegated authority; a gated change stands only when an authorized reviewer accepts it, and the history preserves both states and the transition. This is the same gate that turns an approved artifact into governed action, and it carries load in both directions. Second, an agent draws from systems of record rather than from recall. The figure in the annex resolves to the system it came from and the time it was read. A model generating a plausible number from memory was a nuisance in a private chat and is a hazard on a shared surface, and that is the sharpest line between a system that belongs in staff work and one that does not. Third, blast radius is a design parameter. An agent with write access to

twelve annexes is a different risk object than an agent with write access to one, and the difference should be a deliberate choice.

That discipline has a cost, and the cost belongs in the open. Some of the recovered time goes straight back into verification, and the trade is worth taking only because checking a figure against a linked source and a retrieval time is cheaper than deriving it by hand. A system that cannot make verification cheaper than production has moved the labor rather than reduced it. That is a test to run in an exercise before it is a claim to make in a brief. And none of it makes an agent's judgment about a source trustworthy, because it is not supposed to. Whether a source holds up is the part a person keeps, which is why the gate exists.

The Benefit Arrives as Recovered Time and as Error Economics

A staff that saves ten hours of transcription and spends them polishing slides has gained nothing.

Recovered time is easy to count and easy to squander, because recovered hours only convert into output if commanders restructure the battle rhythm around them. Error economics is harder to count and worth more. A contradiction caught on day three costs a revision. The same contradiction caught on day nine costs a night, a ripple through four annexes, and sometimes the credibility of the plan. Moving discovery earlier is the compounding benefit, and it is invisible to any metric that only counts speed.

No public study yet measures agents working inside a military staff's shared artifacts. Other high-consequence domains have paid to measure what context gaps cost. The I-PASS study, a nine-hospital trial covering 10,740 patient admissions, found that [standardizing the shift handoff reduced medical errors by 23 percent and preventable adverse events by 30 percent](#). That does not establish an effect size for a military staff. It does show that changing how context crosses a boundary reduces errors without changing the underlying work. The 1988 Piper Alpha explosion, which killed 167 people, began with a safety valve whose removal was recorded on paper but never conveyed at shift handover. The record existed. It did not move with the work.

Governance Is the Precondition, Not the Feature Tier

The gap between a helpful summary in a chat channel and a product a battle captain will sign is the entire governance stack.

Commercial software treats scoped access, action logging, and approval workflow as enterprise features, priced as a tier. In national security they are the condition of entry: an organization that cannot audit an AI's participation in its work will not, and should not, let it participate. The requirement binds hardest in operations centers, because operations centers act, and an unattributed AI contribution in a chain that moves forces or feeds targeting is unacceptable. Everyone who has worked in one of those buildings knows it. [The architectural case for governed execution has been](#)

[made in this series](#), and it holds here with one addition. The approval gate is not there to slow the agent down. It exists because judgment is the part of the work an organization cannot delegate, and a gate is how an organization says so in software.

This is the premise Legion built on, not a feature added for a release: agents work under human command, every action is attributed, and consequential steps wait for a person. [Legion Multiplayer](#) carries that same governance from the single operator into the work a staff does together. It is not a separate product to sign up for. It is how the platform operates. At a U.S. Department of Energy national laboratory, [a recurring report is drafted today as one shared artifact](#), each agent contribution linked to its session, with subject matter expert review built into the approval step. That is evidence the architecture is buildable and running. It is not yet a measurement of the error-economics gain, and the difference matters. The edge record is narrower still: at Scarlet Dragon 26-01, Legion ran on a tactical SIPR network, and [Centurion by Legion Intelligence](#) went from field install to operational capability in a day. That demonstrates forward deployment and install time, not a session surviving a network partition.

Command Authority Is the Question This Forces

An agent working in the artifact enters the chain of accountability the moment it contributes.

Command authority is the question the private session let everyone defer. A private assistant is easy to govern because its output carries no weight until a human adopts it. An agent that drafts the annex, maintains the estimate, and flags the contradiction is a participant in staff work, and participants sit inside a chain of accountability. When an agent's contribution shapes a plan that goes wrong, the accountability question will have an answer, and the answer cannot be the vendor's terms of service. Attribution and approval gates are the technical foundation for that answer, not the whole of it.

That has a direct consequence for where the money goes. If generation is not the binding constraint on staff work, model selection is not the center of the investment there. Record systems store what happened. The durable layer is the system of work: where shared state, authority, provenance, and institutional practice live, and where the work itself happens under governance. Models should be replaceable above that layer without rebuilding it, though every substitution still demands evaluation, testing, and in most environments renewed authorization. The model is the fastest-moving layer of the stack, and an organization that has bet its architecture on one model has bet on the part most likely to be obsolete before the accreditation clears.

What to Demand Before an Agent Joins Your Staff

The argument above is testable and should be tested in exercises before it is believed in a brief.

Ask to see the five requirements in the artifact history. The five are inspectable properties, not roadmap promises. If a vendor cannot show you who changed a passage, from which source, at what time, on a disconnected network, the requirement is not met.

Ask who sets the gates. The organization designates which steps are consequential, not the vendor. Proposed state and accepted state must be different states, with the transition preserved. If that gate does not exist, the accountability question has no answer.

Run the verification test. Give a real staff product to a real staff section in an exercise and measure whether checking an agent's figure against its linked source is cheaper than deriving the figure by hand. A system that fails this test has moved the labor, not reduced it.

Measure on three axes. Recovered time: staff hours per product, and time from tasking to reviewable draft. Error economics: the age of a contradiction when it is discovered, and rework hours per conflict. Correctness: extraction accuracy against known sources, and reviewer overturn rates. [The evaluation discipline for agentic systems already exists](#); apply it here. Provenance is a separate check, not a proxy for correctness.

Write the conventions now. Who may task an agent, what an agent's draft is worth in review, how a flag is adjudicated when it contradicts a section chief, and when an agent must stop and wait for a person. Doctrine will lag, as it always does. The formations that write these conventions first, in exercises, with real staffs producing real products, will hold an advantage that has nothing to do with whose model sits underneath.

The mission has always been multiplayer, and now the software is too.

References

1. Headquarters, Department of the Army. FM 5-0, Planning and Orders Production, 2022.
2. Wampler, Centric, and Salter, The Military Decision-Making Process: A Prototype Training Product, U.S. Army Research Institute, 1998.
3. Center for Army Lessons Learned, Military Decision-Making Process: Organizing and Conducting Planning Handbook, 2023.
4. Cullen, The Public Inquiry into the Piper Alpha Disaster, HMSO, 1990.

Continue the series

Follow Legion Intelligence on LinkedIn: [linkedin.com/company/legion-intel](https://www.linkedin.com/company/legion-intel)

Explore all Command Papers: legionintel.com/command-papers

Ready to talk: <https://www.legionintel.com/contact>