
# The Approval Queue Is Not Governance. It Is a Peace Treaty.

*Autonomy Platform · Public Blog*

PERSPECTIVE

Strategy August 2026

# The Approval Queue Is Not Governance. It Is a Peace Treaty.

Enterprise AI's real blocker is not accountability. It is that most organisations have never written down what they optimise for, and the per-decision approval queue is where that unresolved argument gets settled one order at a time. The fix is to govern the yardstick instead of the decision.

Tim Krug wrote a post in Aug'2026 that I mostly agree with.

He separates two things that get conflated: autonomous execution and human accountability. Humans design the system, define acceptable behaviour, set the constraints, test it, monitor it, intervene, and remain accountable for it. The machine makes the individual decisions. Requiring a human to click "approve" 100,000 times is not governance. It is theatre.

That is right. I want to disagree with his example, and then with what he says the governance problem actually is.

# The example

Tim uses autonomous vehicles. I would avoid it. Not because it is wrong, but because it is the wrong shape, and because a good part of his audience has not accepted it. AVs are a perception problem, in public, with bodily risk and unsettled liability. Enterprise planning is a judgment and allocation problem, in private, with financial risk and contractual liability. Arguing from one to the other invites the objection instead of answering it.

There are better examples, and most people never think of them.

Every five minutes, an optimizer sends binding output instructions to power stations across a continent. Automatic generation control trims the balance on a seconds-level cycle. Protective relays trip transmission lines in milliseconds. Nobody approves an instruction. Humans set the reliability criteria and the market rules. The machine dispatches.

Airline and hotel pricing has run this way since the late 1980s. Millions of price and inventory decisions a day, no human in the loop. Which is worth sitting with for a moment: the same executives who want an approval queue on a replenishment order ceded pricing to an algorithm thirty five years ago, and pricing sits closer to the P&L than replenishment does.

Card payments: billions of accept or decline decisions a day, each in about a hundred milliseconds, decided by a model. Liability sits with the issuer by regulation, independent of what the model decided. Tim’s split is not a proposal there. It is law.

In aviation, TCAS goes further than anything he is arguing for. A resolution advisory is not advice. Pilots are required to follow the machine even when it contradicts a human air traffic controller. That rule was tightened after

Überlingen in 2002, where one crew obeyed the controller, the other obeyed TCAS, and they collided. A safety regulator looked at a fatal accident and concluded that the machine should outrank the human, because the machine's error rate was measurable and lower.

And if you want a transport example without the baggage, use driverless metros. Docklands since 1987, Copenhagen since 2002, Paris Line 1 converted while in service. Millions of passengers a day, for decades, with no controversy at all.

Refineries run model predictive control on distillation columns. Insulin pumps dose a lethal drug continuously with no per-dose approval. None of this is contested.

So why is enterprise planning different?

## The part I think is wrong

Tim frames the blocker as accountability. I do not think it is.

Nobody in a planning organisation seriously believes a planner is personally accountable for each of a hundred thousand decisions. They know they are rubber-stamping. Ask them.

The approval queue survives for a different reason. It is the last remaining forum for an argument that was never settled.

Sales wants service level. Finance wants working capital. Operations wants utilization and a stable schedule. Procurement wants supplier leverage. Every one of those is legitimate, and they conflict. In most companies that

conflict has never been resolved explicitly, because resolving it means somebody loses. So it gets resolved implicitly, one decision at a time, in the approval step, by whoever is in the room and how much political capital they are carrying that quarter.

The approval queue is not a control. It is a peace treaty. It lets an organisation avoid ever writing down what it actually optimises for.

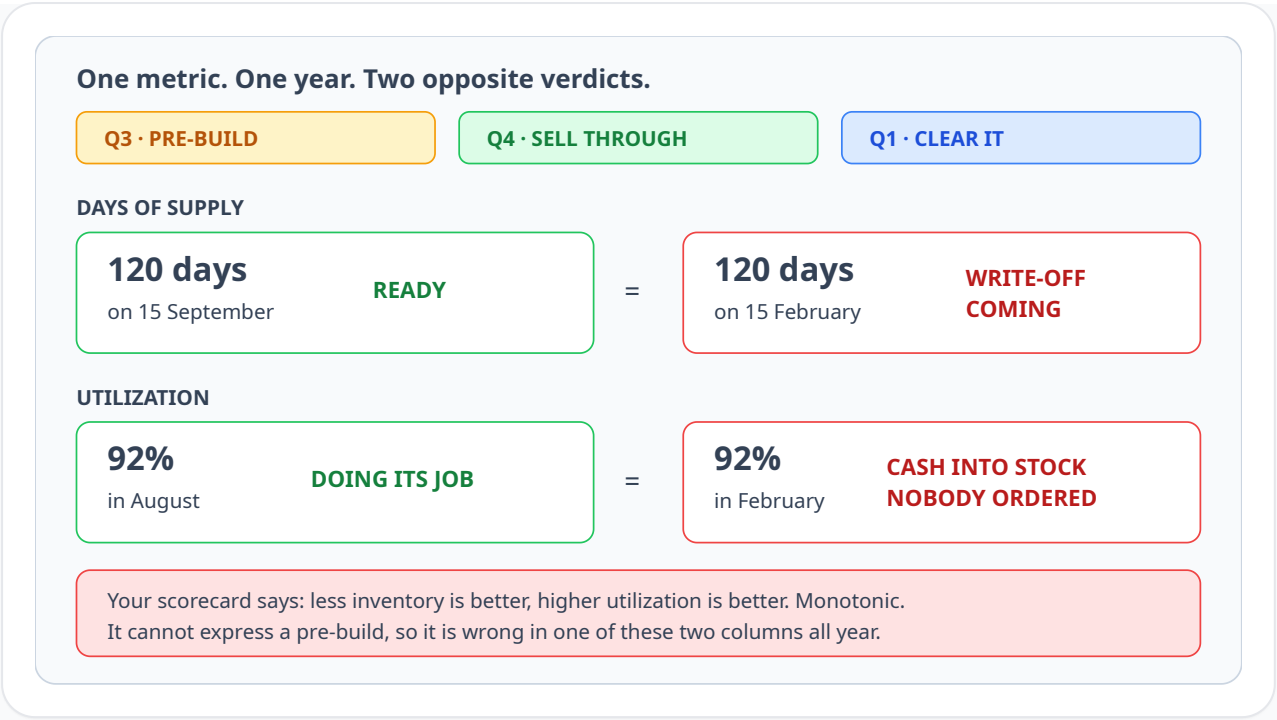
That is why “the AI made a decision I disagree with” is so rarely an argument about the AI. It is an argument about priorities, wearing a technical costume.

## The tell: your scorecard has no time axis

There is a specific and very common version of this that makes the whole problem visible.

Take a company that does 25 percent of its annual sales at Christmas. To serve that, it pre-builds. Utilization runs hot in Q3. Finished goods peak going into Q4 and sell down through it. By Q1, both inventory and utilization should fall sharply, and anything left over is a markdown waiting to happen.

Now look at what that does to the measurements.




120 days of supply on 15 September is a company that is ready. The same 120 days on 15 February is a company with a write-off coming.

92 percent utilization in August is a plant doing its job. 92 percent utilization in February is a plant converting cash into stock nobody has ordered.

The number is identical. The verdict is opposite.

Almost every enterprise scorecard I have seen is monotonic in each metric. Less inventory is better. Higher utilization is better. Higher fill is better. A monotonic scorecard is structurally incapable of expressing a pre-build. It will penalise the correct behaviour in Q3 and reward the wrong behaviour in Q1, every single year, and everybody in the building knows it.

So what happens? The planners correct for it manually. They override. And the place where that correction actually gets applied is the approval step.

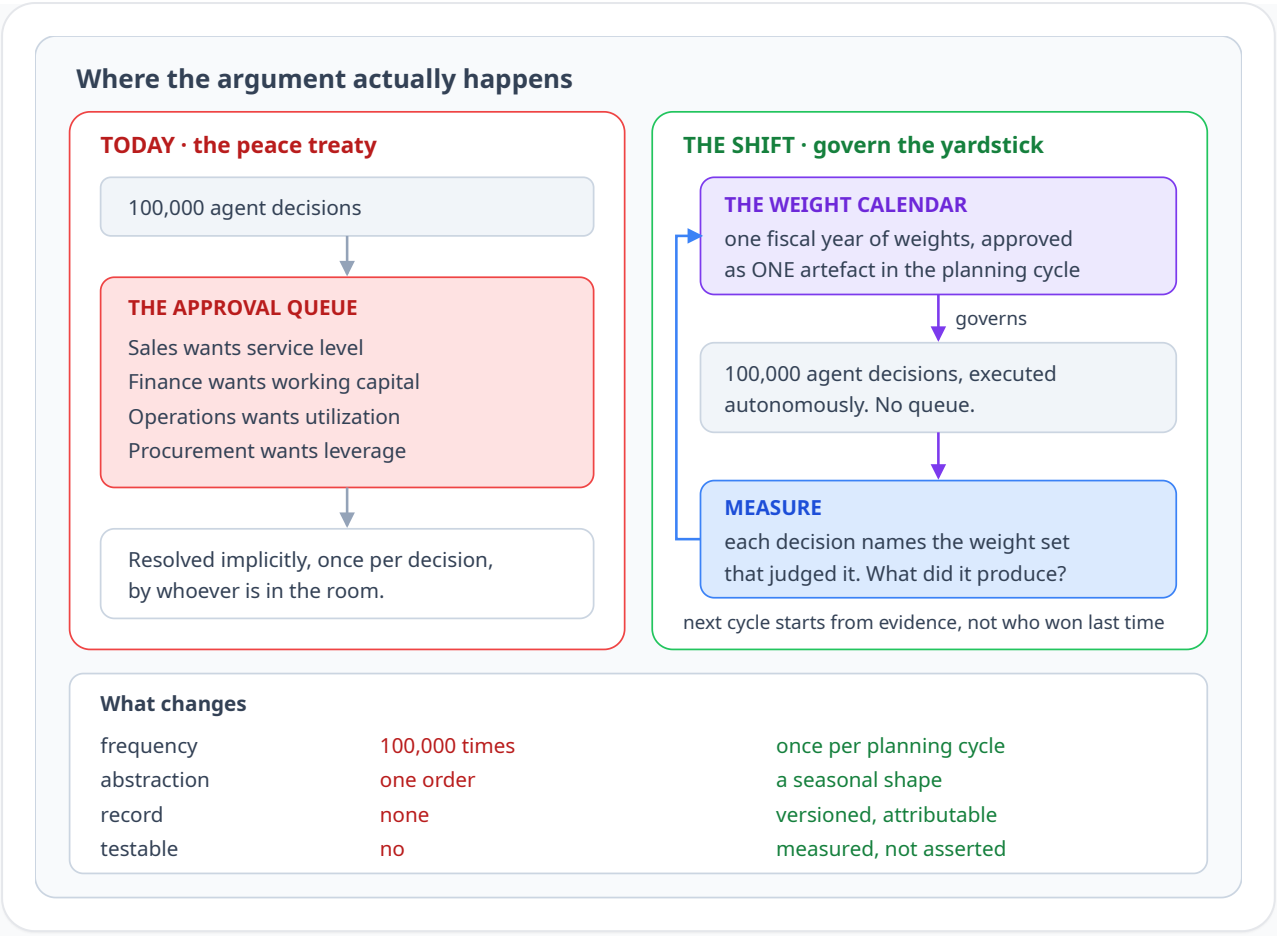
Which means the approval queue is not primarily a control at all. It is a patch for a missing dimension in the measurement system. The organisation has a seasonal objective function and a non-seasonal scorecard, and the gap between them is being closed by hand, one decision at a time, in somebody's head.

Worse, it is being closed inconsistently. The seasonal curve in the sales director's head is not the one in the CFO's head, and neither is written down. That is the entire Q4 argument, every year, in one sentence.

## The remediation: move the argument to how we measure good

The obvious objection to everything I am about to propose is that a declared objective function is itself a political artefact, negotiated by the same people, and just as gameable as an approval queue. That objection is correct, and it is the right place to focus, because the response to it is the whole design.

You do not remove the politics. You relocate it, and you change four things about it: its frequency, its abstraction level, whether it leaves a record, and whether it can be tested.




**Frequency.** The trade-off between service and working capital gets made once per planning cycle, by the people who actually hold the authority to make it, instead of a hundred thousand times, implicitly, by people who were never given that authority and cannot be held to it.

**Abstraction.** This one matters more than it looks. In July, in the abstract, nobody is defending a specific order. It is far easier to agree “we will accept 60 days of supply and 90 percent utilization through Q3 to protect Christmas fill, and we will accept eight points of utilization loss in Q1 to clear it” than to have that same fight four thousand times in October about individual SKUs. Abstraction lowers the political temperature, because no one is losing a particular battle in front of their team.

**Record.** A weight that was agreed, versioned and timestamped can be argued with next year on evidence. A decision that was waved through in a meeting cannot be argued with at all, because there is nothing to point at.

**Testability.** Once the shape is written down and every decision names the shape it was judged under, “was last year’s Q4 right?” becomes a measurement rather than a negotiation.

Concretely, that means an objective function with a time axis in it. In our case that is a balanced scorecard across financial, customer, internal process and learning, with weights the customer sets, resolving on five axes: the tenant, the product hierarchy, the geography, the planning horizon phase, and the calendar event. So the trade-off in December for a promotional line in the Southeast is legitimately allowed to differ from the trade-off in March for a base line nationally.

Attainment is measured against the target for that phase, not against a direction. Building to 120 days in September scores well. Sitting at 120 days in February scores badly. Same metric, same company, opposite verdict, written down in advance rather than reconstructed afterwards by whoever is defending themselves.

The guardrails phase too. The envelope the agent is allowed to act inside is wider on inventory in Q3 and tighter in Q1.

# Three controls, or it is just politics with logging

Here is where I want to be more useful than encouraging, including about our own work. A declared objective function without the following three controls is not governance. It is the same peace treaty with a nicer audit trail. If you are evaluating anyone in this space, including us, these are the questions to ask.

**One. Is it declared forward, never backward?** The seasonal shape has to be agreed before the season, in the cycle that commits to it. If weights can be backdated over decisions already taken, you have not built a governance system, you have built an excuse generator, and “the weights were different that quarter” becomes an unfalsifiable defence for any miss. Ask whether the system will physically refuse a backdated effective date, and whether there is a flag that turns the refusal off. There should not be one.

**Two. Are decisions evaluated against the weights that were in force when they were taken**, rather than the weights in force at review time? This requires two things most systems do not have: the objective must be versioned and immutable, and every decision row must name the version that judged it. Without that binding you cannot separate “the world changed” from “the yardstick changed”, which means you cannot attribute anything. Ask whether weight history can be deleted. If it can, the answer to this question is no, whatever else you are told.

**Three. Is a change to the objective itself a governed decision?** Author, timestamp, stated reason, an approver distinct from the proposer, and a record on the same stream as every other decision. Changing the objective

mid-quarter is legitimate. Changing it quietly is not.

I will be straight about where we are on these. The five-axis, time-varying scorecard is built and running. The three controls above are designed and specified, and are being built now. I am publishing the bar before we have finished clearing it, because the bar is right whether or not any particular vendor has reached it, and because a market that does not ask these questions will be sold audit trails instead of governance.

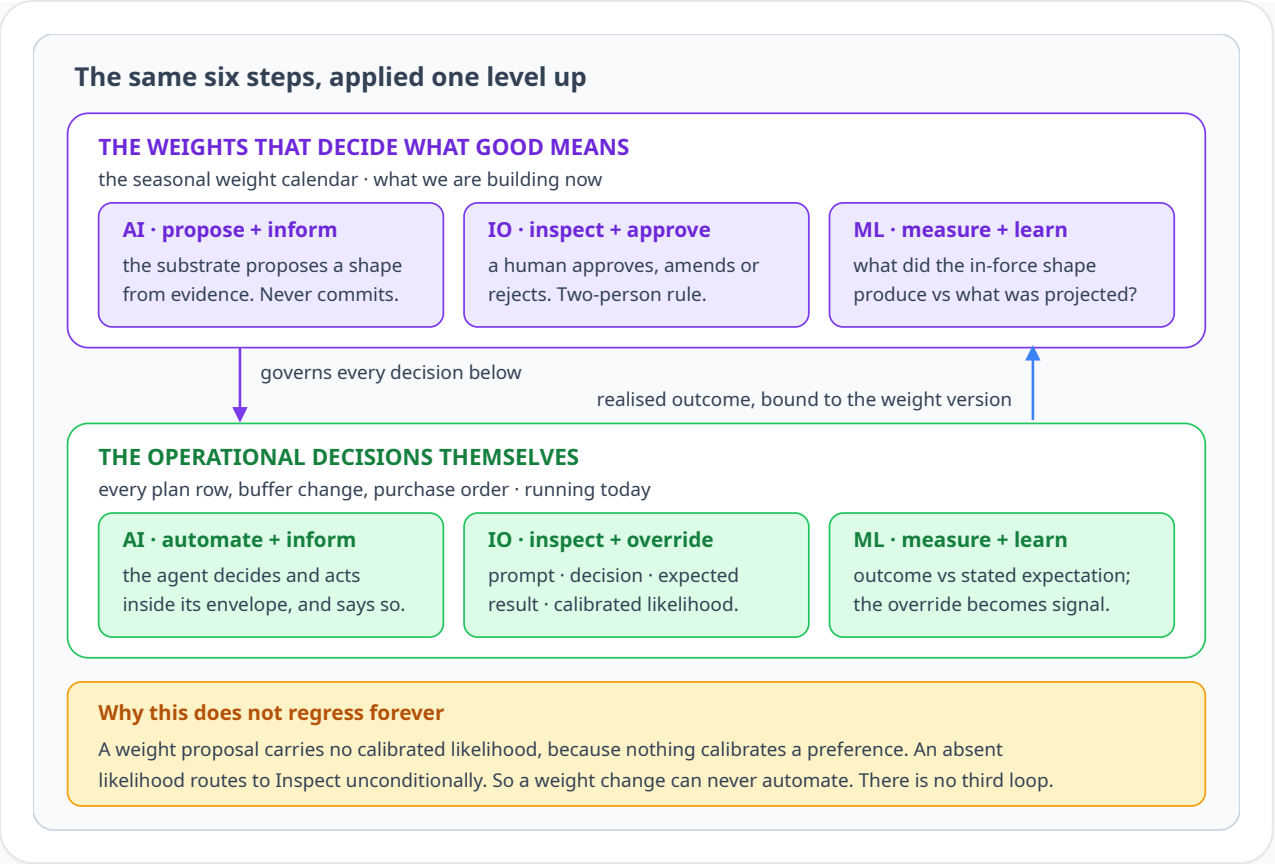
## AI, IO, ML, applied one level up

There is a cleaner way to say all of this, and it is the operating model we build everything else on.

**AI** is the agent's two steps: Automate, then Inform. It decides and acts inside its envelope, and it surfaces the decision when that matters.

**IO** is the human's two steps: Inspect, then Override. A person reads what prompted the decision, what was decided, what result was expected and how likely that result is, and supersedes it when they know something the agent did not.

**ML** is the substrate's two steps: Measure, then Learn. The outcome is scored against the stated expectation, and it feeds back as training signal. Most agent stacks get as far as Measure. The rung that separates a system that measures from a system that learns is the one after it.




We apply those six steps to every operational decision. We have not been applying them to the decision that governs all of them, which is how “good” is defined across the year. That is the actual gap, and stating it as a gap in a model we already have is more honest than presenting it as a new idea.

Applied one level up it reads:

**AI.** The substrate proposes a seasonal weight calendar from the evidence of the last cycle, and surfaces it. It proposes. It never commits, and that is a deliberate architectural line rather than caution: the weights are the preference, so tuning them against the scorecard would be optimising the yardstick against itself.

**IO.** Humans inspect the proposal with the same four fields as any other decision, and approve, amend or reject it. This is the only approval queue worth having, and it has a few dozen entries a year rather than a hundred thousand.

**ML.** At season close, measure what the in-force shape actually produced against what was projected when it was approved, and hand that to next year's proposal.

One question always comes back at this point, and it deserves a direct answer: if a weight change is itself a decision, what governs that decision, and does it regress forever?

No, and the reason is structural rather than a matter of policy. Every decision in our system routes by its calibrated likelihood, and a weight proposal does not have one, because nothing calibrates a preference. An absent likelihood routes to Inspect unconditionally. So a weight change can never automate. The buck stops at a named human by construction. There is no third loop.

## Earned how, exactly

Tim closes by saying autonomy should be earned. Agreed, but I would be far more specific, because "earned" as usually stated is a trust-building exercise with no defined end, and that is a very comfortable place for a buyer to park indefinitely.

In every accepted example above, autonomy was not earned by elapsed time. It was earned the moment somebody built the instrument that made the machine's error rate observable and bounded. TCAS has a certified

false-advisory rate. Card fraud has chargeback accounting. Grid dispatch has N-1 security criteria. The insulin pump has alarm limits and a manual override.

For planning, that is four instruments: a calibrated likelihood on every decision, so confidence is a number with a coverage guarantee rather than a vibe; a declared envelope the agent provably cannot exceed, enforced structurally rather than by convention; an audit record good enough for a regulator; and causal attribution of realised outcomes back to the decisions that caused them, so the envelope widens on evidence instead of on optimism.

Build those and the approval queue becomes visibly redundant. Skip them and no amount of demonstrated performance will ever retire it, because there is nothing to demonstrate performance against.

## The objection this actually answers

Someone will say the approval queue has nothing to do with measurement systems. It survives because a planner who approves a hundred thousand decisions has a hundred thousand opportunities to be seen adding value, and no governance architecture competes with that.

I used to think that was the unfixable residue. It is not. It is precisely what the ML half of the loop is for.

Volume is only a proxy for value while nobody measures value. Every override in our system is measured against what the agent’s decision would have produced: the counterfactual, not the outcome on its own. That result accrues to a posterior held per planner, per decision type. And

the posterior is not a report card that gets filed somewhere. It sets how much weight that planner's overrides carry the next time the model retrains.

That inverts the incentive completely. A planner whose overrides are consistently right does not merely get credit. Their judgment is compiled into the operating policy, by name, and the system defers to them more next cycle. A planner who overrides indiscriminately in order to be visible produces a visible record of low-value overrides, and their weight decays.

The currency changes. It stops being how many decisions you touched and becomes whether touching them helped. That is a far better deal for a good planner than the approval queue ever was, because in a rubber-stamp queue the expert and the box-ticker are indistinguishable. Nobody was ever promoted for the two overrides a year that genuinely mattered, because nothing separated them from the other nine hundred.

It is also how institutional memory stops walking out of the door. A planner who retires after thirty years takes their judgment with them today. Once their overrides have been measured and weighted, the judgment stays.

## What it still does not fix

Two honest limits.

A per-planner record is a performance instrument, and performance instruments change behaviour whether or not you intend them to. If planners come to believe the record is used punitively they will under-override, including in the cases where they genuinely do know more. That is the worse failure, because those are exactly the overrides the system

needs in order to learn anything at all. So the instrument has to measure override value and never override rate. A low override count is not a good score, it is an absent signal. And below the sample size at which the number means anything, it should refuse to show a number rather than show a noisy one.

And the weights themselves are still set by people. A scorecard whose weights are decided by whoever won the last argument is the same politics with better instrumentation, and anyone who tells you a governance layer makes an organisation rational is selling something. What changes is that the argument becomes periodic instead of continuous, abstract instead of transactional, recorded instead of verbal, and testable instead of asserted. Over three or four cycles the evidence starts constraining it, because a position that has been measurably wrong twice is hard to hold a third time. That is a real improvement, and it is a smaller claim than the one usually made.

The uncomfortable version, for anyone still reading: if your organisation cannot state its objective function, including how it changes across the year, the problem was never the AI. The AI just made the omission impossible to keep ignoring, and the approval queue was how you were hiding it.

# See Autonomy in action

Walk through how Autonomy models, executes, monitors, and governs supply chain decisions with autonomous AI agents.

See It Live