AI & PLANNING SYSTEMS

The Board Wants Agentic AI. What Should the Supply Chain Leader Do?

Christian Kroschl · 14 August 2026

Article on kroschl.com · Read on LinkedIn

Why AI pressure may be the best opportunity in years to turn a technology programme back into a business transformation

Boards want higher productivity, lower headcount, less working capital, fewer write-offs, better service and protected revenue. Agentic AI is the current vehicle for that expectation, and supply chain leaders are expected to answer, often while their planning transformation is still running.

What follows is written for the supply chain leader, not the board directly. The mechanics and the evidence are the leader's homework. The board only needs what comes at the end, in *What the board needs to decide* and *The position*.

The predictable response is a workshop. Forecasting agents, inventory agents, scheduling agents, S&OP copilots, autonomous planning. The ideas get ranked by value and feasibility, turned into a roadmap, and presented as progress.

A use-case list creates activity. It does not create value. An agent has no business case. A changed decision has one.

Value appears along a single chain: economic loss, decision failure, changed decision, measurable outcome. Anything that cannot be placed on that chain is a demonstration.

See Figure 1 · page 6

Agents can only act on decisions that are machine readable

Most of ours are not. Making them readable is the work, it is finite, and it produces value with or without an agent.

A decision is machine readable when four things are written down. The trigger that raises it. The constraints that bound it. The rule that resolves it. The owner who signs it.

Where those exist, an agent can prepare, recommend and eventually execute. Where they do not, an agent can only summarise the confusion faster.

That is not a reason to wait. It is the value case.

See Figure 2 · page 7

What agents can infer, and what they cannot

The claim now being sold is that agents read transaction data, infer what is going on, and clean master data becomes optional.

The distinction is narrow but decisive: transaction data can reveal the measured value. It cannot reveal the business intent behind it.

Derivation from transactions is correct for the measured value and worthless for the intent. Actual lead time is a distribution and it comes from movements, as it should. Which point of that distribution becomes the planned lead time comes from

a service decision, and that decision appears in no movement anywhere. A parameter is the intent.

A safety stock of 300 units looks identical in the transaction record whether it was derived from a service target last quarter or carried over at go-live in 2019. A lead time of 14 days looks the same whether it reflects the supplier or reflects nobody having looked since.

Two consequences follow, and both are worth stating to a board.

Raising a planned lead time to recover schedule adherence can make adherence worse. Orders release earlier, queues grow and actual lead times rise. The parameter is not a measurement of the process. It is an input to it.

And cumulative lead time is not the exposure every inventory position faces. The relevant number is the unprotected, decoupled path between the customer and the next effective buffer. That is what a promise can be quoted from.

Master data is not a cleansing project. Those end, and the data drifts again from the day they close. A parameter is a decision with an expiry date, and each field that drives planning needs five things: an owner, the business intent recorded with the value, the measurement it derives from, a permitted range, and a review cycle.

The scope objection answers itself. You are not doing this to every field in the material master. You are doing it to the fields behind the decisions you chose.

A composite case

The pattern below is assembled from several life sciences programmes. No single company is described.

Several ERP instances from more than one vendor. Demand planning, tactical supply planning, detailed scheduling and analytics on separate platforms. A tactical planning solution currently being replaced. Forecast quality limited. Tactical supply planning largely unconstrained. Rough-cut capacity planning either absent or without influence on the released plan. Network planning that is substantially distribution requirements planning.

Planners reconcile by hand. Capacity problems surface late. Inventory and write-offs stay high. The business value of the transformation is not clearly demonstrable.

The board asks where agentic AI can be deployed. The ambition is not the problem. The starting point is.

Three criteria instead of a value and feasibility matrix

Leaving a workshop with forty ideas is the failure mode. Leaving with four is only better if the four were chosen against a criterion.

Is there a named loss already visible in the accounts? An inventory position, a write-off line, a premium freight spend, a service penalty, a batch that expired. If the loss cannot be pointed at, the value case will be built from assumptions and defended the same way.

Is the loss driven by elapsed time between a signal and an action? Total response time has three parts: from the event to the signal being available, from the signal to a decision, from the decision to the action taking effect. AI compresses the first and part of the second. It does not compress the third and it does not supply the policy.

See Figure 3 · page 8

Where the loss is structural rather than temporal, an agent detects it faster and changes nothing. A lane that is permanently uneconomic, a site that cannot make the volume, a portfolio carrying more variants than the changeover capacity supports. Those are network and portfolio decisions.

Does the policy exist, or can it be written? An agent cannot invent customer priority, risk appetite, the service floor or the regulatory boundary. Where the policy exists it can be applied. Where it does not, writing it down is the work.

There is a structural version of the use-case trap worth naming, because it is what most vendor and advisory roadmaps currently look like. An agent behind every process step and every role is an organisation chart rendered in software. It automates the handovers instead of removing them, and the number of agents a landscape appears to

need is a measurement of how much handover its design contains. The best outcome is a decision that disappears, not a decision that gets an agent.

The easier it is to name an AI use case, the more likely it is that the decision underneath it was never designed.

Constrained supply allocation

Demand for a critical product family exceeds capacity for the next eight weeks. Demand sits in the planning application, production orders and inventory across ERP instances, schedules elsewhere, customer priorities and contractual commitments in local files or in a few people's heads.

A workshop produces three agent ideas. There is one decision: how should constrained supply be allocated across customers, markets and products, while protecting contractual commitments, strategic revenue, customer criticality and future supply stability.

An agent can assemble the confirmed supply position, real capacity rather than nominal capacity, priorities, regulatory restrictions, contractual obligations, service floors and feasible production alternatives. It can compute allocation scenarios, show the impact on service, revenue, backlog and inventory, route the recommendation to the accountable decision-maker and document what was approved. It cannot decide whether one strategic customer outranks another. That is a policy.

There is a harder point that allocation discussions avoid. Under scarcity, contractual allocation is weaker than it looks. A supplier who is short diverts material to its largest and most profitable accounts, and that rule is often already coded into its own order promising. A contract that only performs when supply is adequate is not a risk response. The same applies in reverse to your own customers, which is why allocation rules are negotiated and instrumented before scarcity rather than invoked during it.

Value: protected revenue, fewer lost sales, lower premium freight, reduced backlog, better service for priority customers, more stable production.

That decision is worth trying rather than reading about. Sixty seconds, the same scarcity, and everything an agent would have: allocate, then watch two more runs on identical data argue their way to different answers.

Open interactive tool Open in a new tab

Write-off prevention

Write-offs rarely result from one planning error. Forecast bias, shelf-life constraints, campaign sizes, market shifts, delayed release, packaging restrictions and slow reallocation accumulate over time.

An agent can identify materials likely to expire before the risk reaches a financial report. Detection alone is worth nothing. The question is whether there is still time to act. In a regulated environment the options are bounded by registration, batch release, labelling, transport conditions and minimum remaining shelf life, and a recommendation that ignores any of them costs credibility rather than earning it.

Value: earlier detection, more intervention time, feasible reallocation or production changes, lower obsolescence and write-offs.

Planner capacity

The potential is real. Planners spend substantial time collecting data, preparing exceptions, reconciling systems and chasing actions, and much of that can be supported or automated.

Released capacity is not headcount value.

Headcount value appears when roles are redesigned, manual reports retired, coordination layers removed and duplicated planning stopped. A published case makes the point better than a projection. At a semiconductor manufacturer, planners overrode the integrated planning system more than ninety percent of the time. The cause was not model quality and not data: order acknowledgment sat with business unit planners protecting their own delivery metrics, while capacity allocation sat centrally. Consolidating the accountability cut overrides to under fifty percent within six months, moved shipping performance from

seventy five to ninety five percent and added two inventory turns, with no change to the system.[1]

Override rate is often a governance signal rather than a measure of system quality. Concentrated overrides usually point to a specific conflict in accountability, policy, parameters or model design rather than general planner resistance.

Value: less manual data collection, less exception preparation, less coordination effort, and FTE reduction only after roles, reports and organisational layers are redesigned.

What agentic AI should actually do

Four capabilities are sold as one thing, and separating them is most of the argument. Machine learning detects patterns, bias and anomalies in structured data. Language models extract events from supplier communication, regulatory notices and unstructured documents. Simulation evaluates how a disruption or a decision propagates over time. Optimisation calculates and ranks feasible responses. The agent orchestrates these, calls systems, assembles context and manages workflow.

The language model detects. Simulation explains. Optimisation proposes. Governance decides.

See Figure 4 · page 9

That is how the current research builds it. A 2026 paper in Annals of Operations Research specifies a decision support system for tariff induced disruption in exactly this shape: a language model monitors unstructured sources and raises an alert, discrete event simulation predicts the impact, and a mixed integer optimisation computes the reconfiguration.[2]

One consequence belongs in front of a board directly. A detection module without a measured latency, precision and recall is not a control element. It is a subscription.

Everyone talks about self-learning planning. Almost nobody has built the controlled return path from execution measurement to an approved parameter change. That path needs an owner, a cadence, a threshold and an audit trail. A loop that updates every cycle is not automatically better than one that acts on a threshold. Frequent parameter changes can reschedule or destabilise the open order book, even when the new value is statistically more accurate. Observe continuously, change only outside a governed band.[3]

1. Can capacity be calculated from maintained routing standards? 2. Are confirmations posted against the resource that actually performed the work?

Both are answerable in an afternoon and neither needs a consultant.[4]

Re-anchor the transformation rather than adding to it

Many supply chain transformations have quietly become planning system implementations. This is not a technology failure and should not be framed as one. Modern planning systems do what they claim. Finite capacity scheduling, sequence dependent setup handling, multi-site netting, available to promise and rule-based sequencing have worked for decades. The open questions are fit, configuration and ownership, and they sit with the company.[5]

The difference is what the programme is asked to deliver. A technology project delivers configuration, integration, testing, migration, training and go-live.

A business project delivers feasible plans, earlier constraint visibility, lower inventory, lower write-offs, faster decisions, less manual planning effort, and protected service and revenue.

See Figure 5 · page 10

The same workstreams, described the second way: Tactical planning delivers feasible supply decisions and earlier constraint visibility, not a module.

Rough-cut capacity planning means no tactical plan is released without a capacity check on key resources, meaning bottlenecks, capacity that is expensive or slow to change, key materials and key suppliers. That is smaller than most people fear.

Network planning covers capacity, sourcing, inventory and production alternatives, not distribution requirements alone. The data platform creates a

trusted decision state, not access. Scheduling integration creates a controlled handover to executable scheduling, not an order transfer. Agentic AI reduces detection, decision and coordination latency.

What the board needs to decide

Five things, and none of them is a technology purchase. Name an accountable owner for planning parameters, with authority to set and change them inside a stated range. Approve recalibration as a standing procedure with limits and an audit trail, rather than a change request each time.

Fund the measurement, meaning retain the data that makes variance visible: forecast by lag against actual, lead time distributions by supplier, confirmation quality by work centre.

Admit AI initiatives to the portfolio only with a named decision, a named owner and a stated threshold at which action is taken.

Re-anchor one existing workstream on an outcome rather than a go-live date, and report it that way.

What to measure

Not pilots and not agents. Inventory and working capital. Write-offs. Premium freight. Plan adherence. Protected service and revenue. Time from event to signal, from signal to decision, from decision to effect. Share of planning parameters with a named owner and a current derivation. Realised productivity value after organisational redesign.

The position

The pressure is legitimate. A long use-case list is not a sufficient answer. Value comes from changed decisions, and those decisions have to become machine readable. That work brings a running planning transformation back to its business case.

We will use agentic AI on a small number of high-value decisions, chosen because there is a loss we can point at, elapsed time we can compress and a policy we can write. We will make those decisions machine readable and use them to bring the planning transformation back to the business case it started with.

Notes

1. Published case. Cohen, S. and Roussel, J., *Strategic Supply Chain Management: The Five Disciplines for Top Performance*, McGraw-Hill, 2005, chapter 3. The company is named in the source. Figures as reported by the authors. 2. Ivanov, D., A generalized concept of a decision-support system for responses to tariff-induced uncertainty and disruptions, *Annals of Operations Research*, 2026, DOI 10.1007/s10479-026-07303-w. Open access. 3. The return path itself is not new. It is derived independently in several lines of work between 1997 and 2006, none of which cites the others: production operating curves relating work in process, output rate and throughput time in Schuh, G. (ed.), *Produktionsplanung und -steuerung*, Springer 2006, chapter 3.6 by Wiendahl, Nyhuis and colleagues; safety stock from measured deviation against a chosen service level, with a named owner and a stated cadence, in Wild, T., *Best Practice in Inventory Management*, Woodhead 1997; cycle time targets derived from measured queue time in Herrmann, J. (ed.), *Handbook of Production Scheduling*, Springer 2006; and ex post feedback named as a coupling parameter in Stadtler, H. and Kilger, C. (eds.), *Supply Chain Management and Advanced Planning*, 3rd ed., Springer 2005. The threshold rule follows from Wild's caution that each parameter update rewrites the open order book, and from the fixed cost formulation in Sethi, S., Yan, H. and Zhang, H., *Inventory and Supply Chain Management with Forecast Updates*, Springer 2005, where a fixed cost of amending an order makes the optimal response a band rather than a recomputation. 4. Reject rates on shop floor confirmation data of twenty to fifty percent, with incorrect work centre assignment the most frequent cause, are reported from industrial studies cited in Schuh 2006, chapter 3.6. 5. Reported by PRTM's benchmarking practice and cited in Cohen and Roussel 2005, chapter 6: firms implementing sophisticated planning technology without matching attention to process and organisation performed worse than firms using no advanced technology at all. Provenance note: the underlying analysis was presented in a benchmarking webcast rather than a peer reviewed study, so the direction is better supported than the magnitude.

Figure 1

FIGURE 1

Where value has to come from



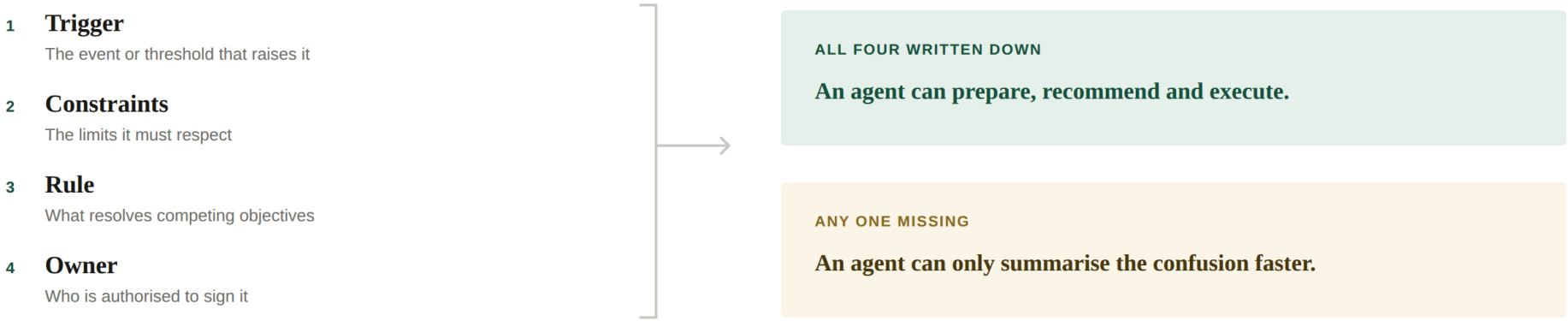
Anything that cannot be placed on this chain is a demonstration, not a business case.

Back to text · page 1

Figure 2

FIGURE 2

When a decision is machine readable



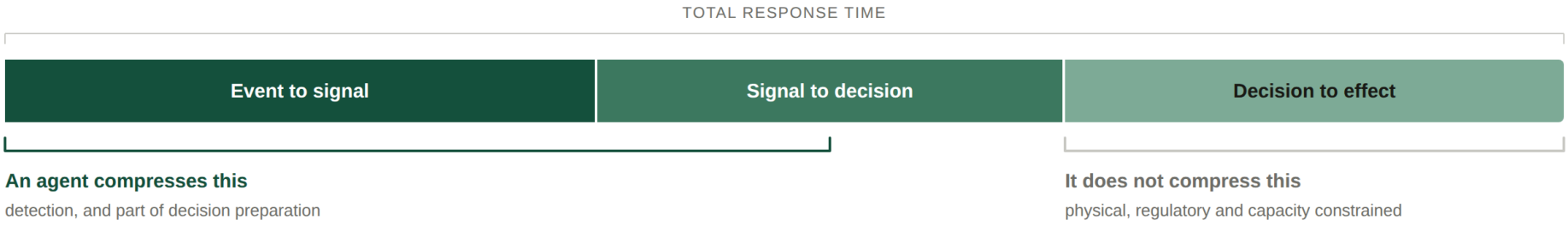
Written down, not known, not understood, not agreed in a meeting once.

Back to text · page 1

Figure 3

FIGURE 3

Total response time, and the part an agent can compress



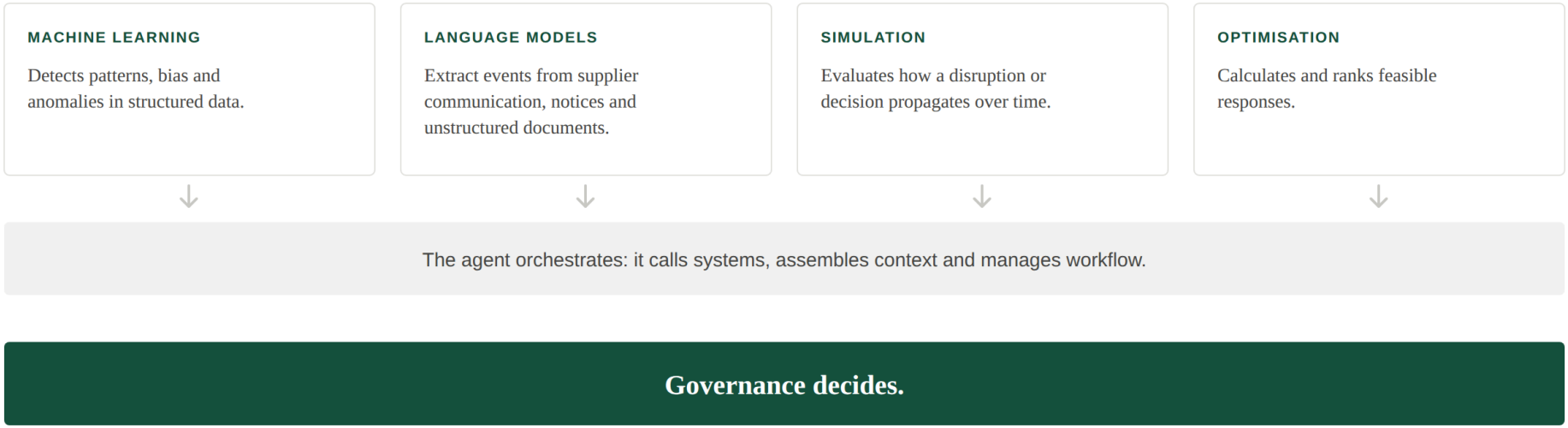
Segment widths are schematic. The proportions differ by decision and have to be measured. The anatomy does not.

Back to text · page 2

Figure 4

FIGURE 4

Four capabilities, sold as one thing



None of the four decides anything. Separating them is most of the argument.

Back to text · page 4

Figure 5

FIGURE 5

The same workstreams, described twice

	A TECHNOLOGY PROJECT DELIVERS	A BUSINESS PROJECT DELIVERS
Tactical planning	Module implementation, model configuration, integration	Feasible supply decisions and earlier constraint visibility
Rough-cut capacity	Resource model and load profiles	No plan released without a capacity check on key resources
Network planning	Distribution requirements and deployment	Decisions across capacity, sourcing, inventory and production alternatives
Data platform	Interfaces, data lake, reporting access	A trusted decision state across planning and execution
Scheduling integration	Order transfer and interface monitoring	A controlled handover to executable scheduling
Agentic AI	Agents, copilots, pilots	Reduced detection, decision and coordination latency

Nothing in the left column is wrong. It is simply not what the board asked for.

Back to text · page 4