RESILIENT SUPPLY CHAIN

EXECUTIVE BRIEF

Why supply-chain AI pilots fail to deliver lasting value

What turns a promising model into a dependable operating capability

Based on a structured review of the Resilient Supply Chain interview archive

RESILIENT SUPPLY CHAIN

Contents

Introduction 3

Executive brief 4

The main finding 4

What senior leaders should decide 5

Finding 1 , Production data is an operating capability 6

Finding 2 , Insight must be connected to the place where action occurs 7

Finding 3 , Authority, accountability and trust must be designed 8

Finding 4 , Scaling requires redesigning work, not installing a tool 9

Finding 5 , Value discipline matters, but the archive does not establish it as a universal primary cause .. 10

What successful scaling cases add 11

What leaders should do differently 12

Where the evidence is less conclusive 13

About this research 14

Source interviews 15

About Tom Raftery 16

Introduction

Supply-chain AI pilots often prove that a model can generate a plausible forecast, recommendation or alert. That is not the same as proving that a business can use the output reliably, repeatedly and economically in live operations. Production value appears only when current data reach the model, the output enters a real decision, people and systems act on it, authority is bounded, exceptions are handled and the resulting improvement survives beyond the project team.

This briefing retests an earlier archive analysis against the current *Resilient Supply Chain* research corpus. It does not assume the earlier five-barrier conclusion is correct. Four fresh semantic searches examined the central question, successful scaling, counterexamples and alternative causes, and possible role or temporal differences. Combined with the preserved all-qualifying ledger, they cover 43 interviews and 154 distinct transcript sections.

The archive is a purposive collection of practitioner and provider interviews, not a representative survey. Support categories describe the quality and recurrence of evidence in the material examined; they do not estimate how common a barrier is across industry. Vendor-reported customer outcomes are treated as useful but unverified observations.

The conclusion is more precise than the original briefing: pilots often fail to last when the surrounding operating system is missing. Four mechanisms are strongly supported, production data, connection to execution, designed authority and redesigned work. A fifth, value discipline, is suggestive and important but less directly evidenced as a cause of pilot failure. These are not a universal or ranked causal model. Successful cases, focused modelling and bounded automation show that organisations can create value before every dependency is perfected.

RESILIENT SUPPLY CHAIN

Executive brief

The main finding

A supply-chain AI pilot is rarely scaled by improving the model alone. The unit of scale is the operating capability around it: maintained data and sensing, integration with the system of execution, explicit decision rights, work design, exception handling, accountable ownership and economic measurement.

Leadership exposure	What the archive supports	Evidence grade
Prepared pilot data cannot be sustained in production	Data quality, fragmentation, sensing and partner inputs are continuing operating responsibilities	Strongly supported
The output is separated from action	Value disappears at functional, system, site, equipment or partner hand-offs	Strongly supported
Decision authority is undefined	Trust cannot mature when recommendation, execution, override and accountability are not designed	Strongly supported
The job and process remain unchanged	Adoption fails when the tool adds friction, ignores local constraints or removes the learning path needed for judgement	Strongly supported
The programme starts with AI rather than a material decision	Problem definition, baseline and scale-or-stop economics are necessary disciplines	Suggestive as a distinct failure cause

What senior leaders should decide

Do not approve a pilot as an isolated technology experiment. Approve a bounded production hypothesis: a named operational decision, a measurable baseline, a complete data and action path, an accountable owner, explicit authority limits, a realistic workflow, and thresholds for stopping, revising or scaling. A model demonstration can answer whether an output is possible. An investment gate must answer whether the organisation can operate it.

Finding 1 - Production data is an operating capability

Evidence grade: strongly supported.

A pilot can be assembled around a curated dataset. A production service depends on information that continues to arrive from ERP and planning systems, spreadsheets, emails, equipment, suppliers and logistics partners. The recurring exposure is therefore not merely dirty data. It is the absence of a managed production data chain with owners, quality rules, maintenance, sensing and recovery when inputs degrade.

The maintenance burden is visible in contradictory and unmaintained master data across systems. Andy Kohm, founder and CEO of SCIP, warns that AI accelerates bad decisions when its inputs are wrong. Don Mahoney, Chief Customer Officer at SNP Group, adds that important enterprise history is fragmented across legacy and unstructured sources; selecting and enriching the right information is part of transformation, not preliminary housekeeping. Their evidence supports an enduring operating obligation rather than a one-off cleansing exercise. [1], [2]

The same mechanism extends into the physical world and beyond the enterprise. Doron Hazan, Director of AI at Wiliot, argues that models cannot act on events the business cannot sense. A 3E interview describes supplier information arriving in inconsistent files, on paper or not at all. Data readiness can therefore depend on sensors, connectivity and external participants that the pilot sponsor does not directly control. [3], [4]

The newly published Asim Akram interview strengthens the operational interpretation. Akram, CEO of MultiSensor AI, describes predictive-maintenance models that need continuous multi-modal sensing, site commissioning, customer-specific thresholds and time to accumulate operating history. He says some models may need a year of data and updating before confidence matures. This is provider testimony rather than independently verified performance, but it demonstrates why buying a model is not the same as establishing a dependable signal chain. [5]

> **Executive decision:** Require a production-data plan before funding. Name each critical source, owner, maintenance rule, latency requirement, external dependency, failure mode and fallback. Treat sensing and data stewardship as recurring operating cost.

Finding 2 - Insight must be connected to the place where action occurs

Evidence grade: strongly supported.

An accurate recommendation has no economic value until it changes a live decision. Across the interviews, the break occurs at several hand-offs: between functions, between an analytical tool and the system of execution, between a standard design and site reality, between software and physical equipment, or between organisations.

The functional hand-off fails when insight is not connected to core processes, auditable context and governance, as Stephen Jamieson, CMO Sustainability, describes. Spencer Penn, co-founder and CEO of Lightsource.ai, observes that sourcing decisions may be completed through email and spreadsheets before ERP is updated; intelligence attached only to the system of record arrives after the consequential choice. [6]; [7]

Execution also varies by site. Keith La Londe, Vice President of Systems at PathGuide Technologies, and Scott DeGroot, Executive Director of the Global Supply Chain Institute at the University of Tennessee, warn that conference-room designs can miss transport, labour and floor conditions. In yard operations, Chad Fox, Manager and Delivery Lead at Miebach Consulting, recommends a reusable core template while explicitly planning for site-level exceptions. Repeatable deployment therefore requires a standard core plus local configuration, not an assumption that the pilot workflow can simply be copied. [8], 10 August 2026; [9]

The strongest success counterexample reinforces this finding. Nishith Rastogi, founder and CEO of Locus, describes decision software connected to live last-mile allocation and common operational metrics. He reports sustained fill-rate improvement and lower miles per package. Those figures are vendor-reported and not independently audited, so they cannot establish expected ROI. They do show the architecture of a scaling case: a computationally suitable decision, operational integration and measures that align executives and drivers. [10]

> **Executive decision:** Define the action path before selecting technology. Specify where the output appears, which system executes it, who receives it, how quickly action must follow, what local variation is allowed, which partners and assets are required, and how the outcome returns as feedback.

Finding 3 , Authority, accountability and trust must be designed

Evidence grade: strongly supported.

Trust is not a soft attitude that appears after training. It is an operating state built from bounded authority, visible evidence, reliable performance and retained accountability. A pilot can defer this design because a project team watches every output. Production cannot.

The first boundary is responsibility for consequences. Simon Bezrukov, Chief AI Officer at Bristlecone, separates automating the paperwork around a decision from owning its consequences. He also warns that machine-learning forecasts can become confidently wrong when discontinuities make the past a poor guide to the future. The implication is not that autonomy is impossible; it is that authority must reflect consequence, confidence, reversibility and the evidence available to explain an action. [11]

Authority can then expand with evidence. A Logility interview describes AI moving from information and recommendations towards execution aligned with explicit organisational goals. A SourceReady interview distinguishes rapid AI-enabled supplier analysis from residual risk and commercial judgement that remain human responsibilities. Wayne Scott, Global Governance, Risk and Compliance Lead at Escode, adds continuity risks from erroneous outputs and dependency on an external provider. Together, these accounts support graduated authority, auditability, override and fallback rather than a blanket human-in-the-loop rule. [12]; [13]; [14]

Akram's newly published evidence supplies a useful exception to any claim that every AI action requires permanent human approval. He describes customers who accept recommendations directly, while safety-sensitive applications retain visual verification and human checks. His approximate 60/40 split is a single provider's observation and should not be generalised. The more defensible conclusion is that human involvement should vary by application risk and earned confidence.

> **Executive decision:** Approve an authority matrix for every use case: what the system may observe, recommend or execute; confidence thresholds; evidence retention; escalation and override; the owner of each consequence; and the safe mode when the model, data or provider fails.

Finding 4 , Scaling requires redesigning work, not installing a tool

Evidence grade: strongly supported.

Calling weak adoption resistance to change hides the design problem. The archive repeatedly describes a mismatch between technology and the roles, interfaces, incentives, skills and physical constraints of the work.

Confidence develops when users understand where a system works and where judgement remains, according to John Dony, CEO and co-founder of the What Works Institute. Rastogi notes that the software buyer is not necessarily the operator who must use it and argues for interfaces that let functional experts express intent. Mor Peretz of CaPow emphasises retrofit disruption; Keith Moore of AutoScheduler highlights the difficulty of repeating deployments across facilities; and Gonzalo Benedit of Aera Technology describes organisations accumulating alerts without becoming able to act. [15]; [16].

A future-dated Friddy Hoegener transcript was not included as publication evidence. It nonetheless identifies a question for later testing: automating entry-level tasks may create capacity while weakening the route through which future leaders acquire judgement. That possible second-order effect should be tested after publication rather than folded into the current finding.

> **Executive decision:** Make operating-model redesign a deployment gate. Involve users and supervisors before configuration, rehearse exceptions under realistic pressure, define how each role changes, preserve learning routes for judgement-intensive work and measure sustained use and decision quality, not training attendance.

Finding 5 , Value discipline matters, but the archive does not establish it as a universal primary cause

Evidence grade: suggestive.

The earlier brief presented weak business linkage as one of five co-equal barriers. The retest supports a narrower claim. Several contributors argue that initiatives should begin with a valuable business or customer problem and a measurable baseline. The archive does not directly show enough cancelled AI pilots to establish that missing ROI discipline is as recurrent a causal mechanism as data, integration, authority or work design.

Budget and prioritisation discipline form the common thread. JP Wiggins, CEO of 1Logtech, says AI tools without hard ROI will struggle for budget and places technology inside business strategy. Cynthia Lai urges leaders to ask which business or customer challenge generative AI will solve rather than funding a hot topic. John Beath, founder and CEO of JBE, argues that weak baselines direct spending toward what is visible rather than valuable. These are consistent executive disciplines, but much of the evidence is advisory rather than retrospective evidence about failed pilots. [17]; [18]; [19].

A proof of concept can also create value by resolving uncertainty even when it does not become the final solution. Yeelen Knegtering recommends testing, finding the value and then implementing, with incentives tied to outcomes. This contradicts any simple equation of unscaled pilot with failure. [20]

> **Executive decision:** Require a decision-level baseline, full production cost and explicit learning objective. A pilot may succeed by disproving a hypothesis; it fails when it neither improves a material outcome nor resolves a decision worth funding.

What successful scaling cases add

The fresh challenge search found multiple operational outcomes that were not prominent in the earlier briefing. They are important counterweights to a failure-only narrative, although the claims are generally made by people associated with the solution and are not independently audited.

In sourcing, Penn reports a natural within-programme comparison, not a controlled experiment, in which categories handled through the platform completed sourcing 25% faster and experienced 37% less post-award cost creep than categories handled outside it. These are vendor-reported results from an unnamed automotive customer. [7], 34:10–35:01. In reverse logistics, Terry Boyle, CEO of Trove, gives prospective estimates of 25–40% labour savings and improved return-to-stock performance after workflow standardisation. Because the passage predicts expected benefits rather than documents a completed deployment, it is not treated here as a successful-scaling case. [21]

Physical automation supplies two further provider-reported examples. Peretz describes one robot-fleet implementation in which charging queues fell from about 60 seconds to 15 seconds, alongside a stated 30–40% downtime effect; the interview does not provide the customer, baseline design or independent verification. [16], 22:18–23:01. Hiten Sonpal, CEO of RISE Robotics, describes embedded diagnostics and prognostics that identify belt degradation before breakage and enable scheduled intervention. His broader statement that the system eliminates downtime is a provider assertion without a quantified deployment record in the selected passage. [22]

The contradiction search also found two boundary conditions. First, imperfect information need not stop progress: targeted modelling, simulation and an 80/20 focus can create useful decisions without cleansing the entire data estate. Second, autonomy is not binary. Opening a ticket, retrieving missing data, proposing a replan or selecting from predefined playbooks can be appropriate autonomous or semi-autonomous action when objectives and state are clear. These exceptions weaken any instruction to perfect all data first or retain a human approval step for every action. They strengthen the case for selecting a bounded decision and expanding scope only as evidence accumulates. [11], [18]

What leaders should do differently

1. **Qualify the decision.** Name the consequential decision or workflow, accountable owner and uncertainty the pilot will resolve.
2. **Establish the baseline.** Record current service, cost, speed, risk or sustainability performance and define a minimum worthwhile improvement.
3. **Test the complete production chain.** Include live data maintenance, sensing, integration, equipment, partner participation and site variation.
4. **Design authority.** State what AI may recommend or execute, when a person intervenes, what evidence is retained and who owns the consequence.
5. **Redesign and rehearse the work.** Test normal operation, exceptions, degraded data, provider failure and frontline pressure.
6. **Measure total economics.** Include integration, governance, infrastructure, support, change and vendor-continuity costs, not only model or licence cost.
7. **Scale against evidence.** Expand deployment and authority only when reliability, adoption and value thresholds are met. Stop or redesign when they are not.

Where the evidence is less conclusive

1. The material examined is a qualitative interview archive, not a representative survey. It cannot estimate prevalence or prove causality.
2. Four new semantic searches selected 87, 94, 59 and 65 sections respectively. After deduplication and combination with the preserved broad ledger, the analysis covers 43 interviews and 154 distinct transcript sections.
3. The current RSC transcript, chunk and embedding inventories match at 489 records. One eligible guest episode did not appear in the combined semantic evidence set; coverage is broad but not proof of saturation.
4. Most contributors in the relevant set are solution providers, advisers or association leaders. The evidence is inadequate for a defensible vendor-versus-operator comparison.
5. The examined publication window is short and reflects available corpus coverage and editorial focus. It does not support a robust longitudinal claim.
6. Reported outcomes are usually contributor or vendor claims. They indicate mechanisms and possible results, not independently audited benchmarks.

About this research

This briefing analyses the *Resilient Supply Chain* interview archive available on 31 August 2026. Four fresh semantic searches and the preserved broad ledger cover **43 eligible interviews** and **154 relevant transcript sections** after deduplication. The resulting evidence spans 3 November 2025 to 31 August 2026; that is a property of the available material, not a designed study period. Selection was purposive and mechanism-led, using the repository's current embedding index and separate challenge queries for success, contradiction, role differences and time. The interviews are not a representative survey, and support counts and grades apply only to the archive examined. The prose uses checked paraphrases and short transcript pointers; it contains no publication-verified direct quotations.

Source interviews

The briefing cites these interviews directly; the evidence appendix records their contribution and passage ranges.

1. Andy Kohm, *Bad Data Is Slowing Supply Chain Decisions*.
2. Don Mahoney, *Supply Chain Transformation Isn't Failing Because of Technology*.
3. Doron Hazan, *Supply Chain AI Needs Better Sensing, Not Smarter Models*.
4. Lily Hogan, *Why Supplier Data Is Breaking Supply Chain Resilience*.
5. Asim Akram, *The Hidden Resilience Risk Inside Your Automated Warehouse*.
6. Stephen Jamieson, *No Carbon Data, No Market Access*.
7. Spencer Penn, *AI in Procurement: When ERP Is Too Late*.
8. Keith La Londe and Scott DeGroot, *Dashboard Theatre*.
9. Chad Fox, Kurt Neutgens and Matt Yearling, *Why an Autonomous Truck Won't Automate Your Yard*.
10. Nishith Rastogi, *Miss the Delivery Slot, Lose the Customer*.
11. Simon Bezrukov, *AI in Supply Chain: Automation Is Not Autonomy*.
12. Jonathan Doller, *Finding the One Thing That Can Break Your Supply Chain*.
13. Ricky Ho, *AI Can Find Suppliers. Humans Still Own the Risk*.
14. Wayne Scott, *When Critical Software Becomes a Supply Chain Risk*.
15. Mike Swain and John Dony, *Industrial Safety Metrics Are Improving. Serious Harm Isn't*.
16. Mor Peretz, Keith Moore and Gonzalo Benedit, *Why More Robots Don't Always Fix Warehouse Performance*.
17. JP Wiggins, *The Real Supply Chain Bottleneck Isn't AI. It's Integration*.
18. Cynthia Lai, *Poor Supply Chain Emissions Data Could Cost You More*.
19. John Beath, *Measure First*.
20. Yeelen Knegtering, *Paperwork Errors Still Break Supply Chain Resilience*.
21. Terry Boyle, *The 5-Point EBITDA Opportunity in Reverse Logistics*.
22. Hiten Sonpal, *Hydraulics Waste 75% of Energy , Inflating Heavy Equipment Electrification Costs*.

RESILIENT SUPPLY CHAIN

About Tom Raftery

Tom Raftery is a sustainability and technology executive, analyst and broadcaster whose work focuses on climate, energy, digital transformation and resilient supply chains. He hosts the **Resilient Supply Chain** and **Climate Confident** podcasts, where he interviews business leaders, policymakers and innovators about the practical changes required to decarbonise industry and build more adaptable operating systems.

Previously a Global Vice President at SAP, Tom worked on strategy at the intersection of sustainability, energy and emerging technology. He has also spent more than a decade as an independent industry analyst and futurist, including seven years leading sustainability research at RedMonk and work with Gerd Leonhard's Futures Agency.

Earlier in his career, Tom co-founded an Irish software-development company, a social-media consultancy and Cork Internet eXchange, an energy-efficient data centre. That combination of technical, entrepreneurial and advisory experience shapes his approach: he connects new technology to the commercial, organisational and environmental realities that determine whether it creates value.

Tom is an international keynote speaker, a guest lecturer at Instituto Internacional San Telmo and an advisor to startups working in climate and supply-chain technology. Through Podcast Archive AI, he is also developing an evidence-led publishing practice that turns years of expert interviews into traceable, decision-focused research for senior business leaders.