APPLIED FUTURES RESEARCH
REALITY REGISTER

REA·005

The Pilot Trap

Why Automation Deployments Stall Between Proof and Scale

Prepared by Kevin Nussbaum / Applied Futures Research

July 2026

Abstract. Roughly four in five enterprises now run an automation pilot; barely one in seven operate automation at production scale. The gap is not a technology problem, and it is not bad luck. A pilot and a production deployment are structurally different businesses, and the pilot systematically flatters itself: it runs on three hidden subsidies — concentrated attention, human absorption of exceptions, and charitable measurement — that are silently withdrawn the moment the system is asked to carry real volume. When the subsidies lapse, the true cost surfaces all at once. Integration labor, uninvited infrastructure, exception handling, and the compounding carrying cost of a slipped timeline routinely push realized production cost to three-to-five times the pilot business case. This paper dissects the mechanism, prices each hidden line, treats integrator durability as the counterparty risk it has become, and closes with a staged-commitment framework built for buyers rather than vendors: commit capital only as ignorance is retired, with explicit kill criteria at every gate. It is the deployment-floor companion to REA·004, *The Humanoid as Capital Asset*. The asset thesis told you what to buy. This tells you how not to lose the money after you have bought it.

Epistemic tags used throughout

[GROUNDED] verifiable fact or reported datum. **[ANALYTIC]** the author's reasoning from evidence.
[PROJECTION] a modelled or forward estimate. **[SPECULATIVE]** plausible but unproven.

1. The Trap, in One Paragraph

Every operator who has run an automation program knows the shape of the story before it is told. A pilot is built up in a fenced-off corner of the facility. It works. The demo goes beautifully; the numbers are clean; the vendor is attentive; the champion is vindicated. Capital is approved to scale. And then the deployment that looked finished at the pilot turns out to have barely started — costs balloon, timelines slip, the throughput that was effortless in the corner becomes a daily fight across the floor, and eighteen months later the program is quietly declared “on hold.” This is not a rare misfortune. It is the ***modal outcome*** of enterprise automation, and it has a mechanism that can be named, priced, and designed against.

78%	14%	~70%	3–5×
of enterprises run an automation pilot	operate at production scale	of manufacturers stuck in “pilot mode”	realized cost vs. pilot business case

[GROUNDED] The headline numbers are stark and consistent across independent 2026 surveys of enterprise technology leaders: about 78% of organizations report at least one active automation or AI-agent pilot, but only roughly 14% have reached production scale, with manufacturing clustered near 13–16%. Of those that attempt to scale, a majority hit a blocking issue, and most of those remain stalled for more than six months with no clear resolution path. In manufacturing specifically, an estimated 70% of companies describe themselves as trapped in “pilot mode.”

[ANALYTIC] Read those figures as a single sentence and the thesis of this paper appears on its own: the hard part of automation is not building something that works once, in ideal conditions, watched by the people who built it. The hard part is building something that keeps working at volume, in real conditions, watched by no one in particular. The pilot proves the first. It says almost nothing about the second — and it is frequently engineered, without anyone intending fraud, to obscure the difference.

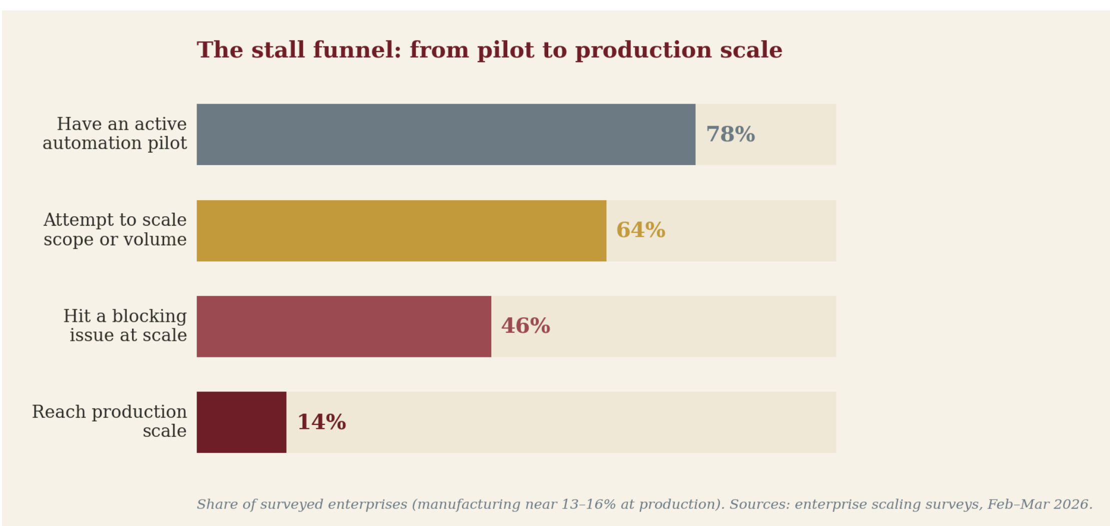

Figure. *The stall funnel. The attrition is not concentrated at the technology-proof step — pilots mostly succeed. It is concentrated at the scale step, where the economics change.*

2. Two Different Businesses

The central error — committed by capable, numerate organizations — is to treat the pilot as a small version of production. It is not a small version of production. It is a ***different business*** that happens to use the same equipment. The two are governed by different cost structures, different failure modes, different people, and different definitions of success. A pilot optimizes for a demonstration; production optimizes for a duty cycle. Confusing the two is the source of nearly every stalled deployment in the record.

Dimension	The pilot business	The production business
Objective	Prove the task is technically possible	Sustain throughput at target cost, unattended
Time horizon	Weeks; a bounded demonstration	Years; a duty cycle across shifts and seasons
Inputs	Curated, clean, representative-of-the-easy-case	The full distribution, including the ugly tail
Labor	Best people, vendor A-team, champion present	Whoever is on shift; vendor at arm's length
Exceptions	Absorbed quietly by humans nearby	Must be handled by the system or they compound
Measurement	Chosen metric, favorable window	Total cost of ownership, every hour, audited
Failure	An interesting data point	A line-down event with cascading cost

Table 1. The pilot and the production deployment are not two points on one line. They are two businesses with opposed incentives.

> *A pilot optimizes for a demonstration. Production optimizes for a duty cycle. Every subsidy in Section 3 is a way the first quietly borrows from the second.*

3. The Three Subsidies Every Pilot Enjoys

[ANALYTIC] A pilot performs better than the eventual production system will, and it does so for reasons that have nothing to do with the technology and everything to do with the conditions of the test. Those conditions constitute three hidden subsidies. None of them appears as a line item. All of them are withdrawn at scale. Naming them is the single most useful diagnostic an automation buyer can carry into a vendor meeting.

3.1 The attention subsidy

During a pilot, the deployment receives a concentration of human attention it will never see again. The vendor fields its strongest engineers because a reference customer is on the line. The internal champion is personally present, unblocking problems in real time. The facility's best operators are assigned to the cell because the pilot matters. Every one of these is a temporary input, and every one of them is quietly performing work the production system is assumed to do for itself. When the vendor's A-team rotates to the next reference account, when the champion returns to their day job, when the cell is staffed by whoever is on shift, the attention subsidy lapses — and performance that was implicitly propped up by people falls to the level the system can sustain alone.

3.2 The exception-absorption subsidy

This is the most expensive subsidy and the least visible. In a pilot, when the system encounters something it cannot handle — a mislabeled tote, an unexpected orientation, a jam, an ambiguous input — a human standing nearby simply fixes it. The intervention is instantaneous, unlogged, and free, and it never enters the success metric. The pilot therefore reports the performance of a human-plus-machine system while being credited as a machine. At production volume the arithmetic turns hostile.

The rare, malformed, and ambiguous cases that make up perhaps one to five percent of real input are negligible in a low-volume pilot and decidedly not negligible at scale.

[GROUNDED] The mechanism is quantifiable. If three percent of inputs defeat the system and the facility processes ten thousand tasks a day, that is three hundred failures every day — each one requiring a human to notice, interrupt, diagnose, and resolve. In the pilot, at a few hundred tasks a day, the same three percent was a handful of quiet saves that no one bothered to record. The exception rate did not change. The volume did, and with it the entire economics of who handles the exceptions.

3.3 Measurement charity

The third subsidy is epistemic. Pilots are measured charitably, and usually without anyone deciding to cheat. The task chosen is the one most amenable to automation, not the representative one. The data is cleaner than production data because someone prepared it. The window is favorable. The metric is throughput or accuracy on the selected task rather than fully loaded cost per unit across the real mix. A pilot scored this way answers the question "can the machine do the good case well?" — which is almost always yes — while leaving unasked the only question that matters for a capital decision: "what does it cost to do the whole job, every hour, for years, including the bad cases and the downtime?"

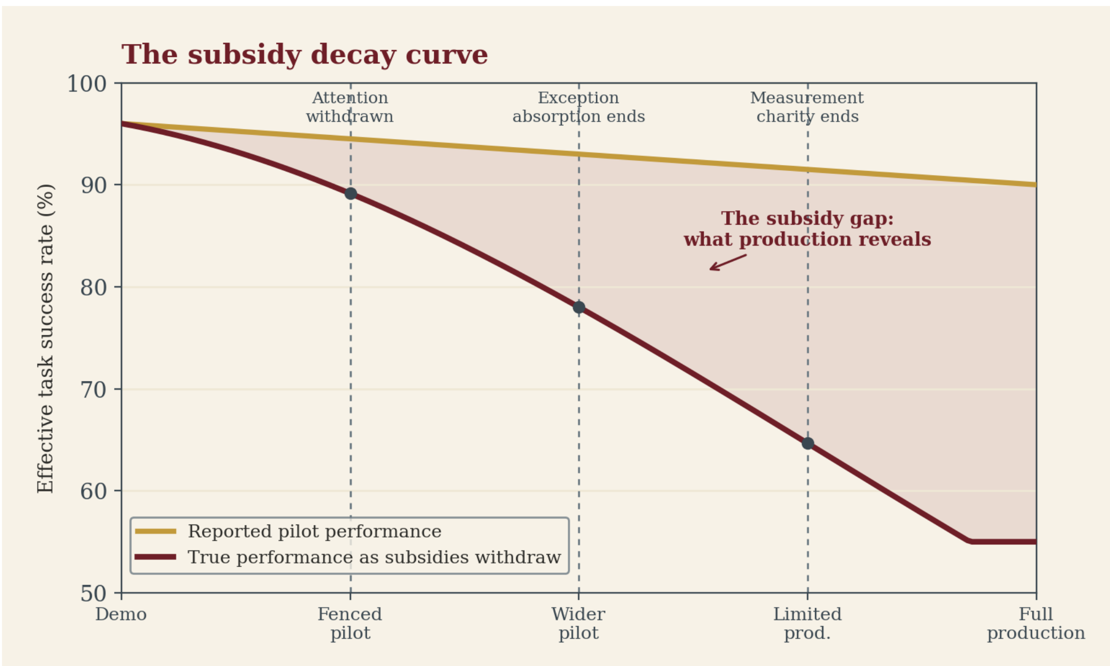

Figure. *The subsidy decay curve. Reported pilot performance (gold) barely moves. True performance (oxblood) falls as each subsidy is withdrawn. The widening band between them is the unpleasant surprise that arrives with scale — and it arrives as cost, headcount, and downtime, not as a chart.*

> *A pilot reports the performance of a human-plus-machine system while being credited as a machine. Production removes the human and keeps the credit — until it can't.*

4. The Scaling Cliff: Where the Money Actually Goes

[ANALYTIC] When the subsidies lapse, their cost does not disappear — it converts into money and appears on the production ledger all at once. This is why the pilot-to-production cost multiple is so reliably brutal. Across warehouse and manufacturing automation, realized production cost commonly lands at three-to-five times the pilot business case, and the overrun is dominated not by hardware but by the work of making the hardware function inside a real operation.

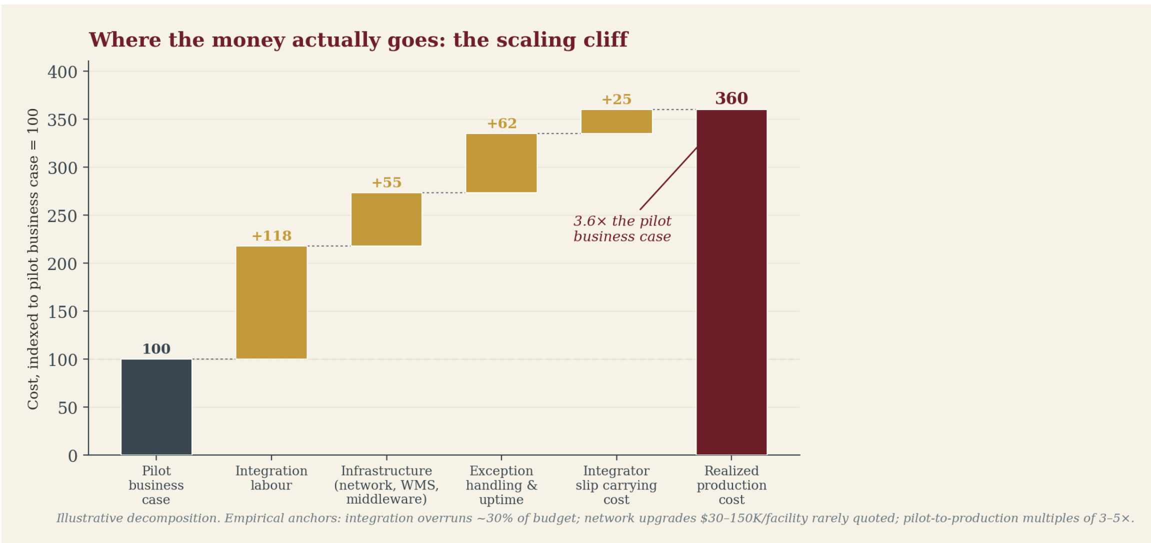

Figure. *From business case to reality. Hardware is rarely the story. Integration labor, uninvited infrastructure, the now-priced exception load, and the carrying cost of slipped time stack into a multiple of the original case. Values indexed to the pilot business case = 100.*

4.1 Integration labor is the dominant hidden line

The largest single addition is almost never on the vendor's quote. It is the labor of integration: connecting the system to the warehouse or manufacturing execution layer, reconciling it with legacy software written before modern interfaces existed, building the middleware bridges, cleaning the data the system depends on, and running the parallel period during which the old process and the new one operate side by side. Legacy execution systems frequently lack usable interfaces, forcing custom bridges that add latency and become single points of failure. Independent market analysis finds integration projects routinely overrunning budgets by around thirty percent and timelines by as much as twelve months.

4.2 The infrastructure that was never in the quote

[GROUNDED] Physical automation has prerequisites that vendor proposals systematically omit because they are the buyer's responsibility on paper. Network infrastructure is the canonical example: facility upgrades to support a robot fleet commonly run thirty to one hundred fifty thousand dollars and are almost never included in automation quotes. Execution software sized for pilot volume stalls under real load — a system comfortable at a thousand orders an hour can cascade into failure at twenty-five hundred during peak. And the connectivity failures that dashboards under-report are real: fewer than ten percent of floor-level device issues are ever reported to IT, meaning the true failure rate can be an order of magnitude worse than the metrics suggest.

4.3 The exception load, now priced

The exception-absorption subsidy from Section 3 reappears here as a permanent operating cost. Someone must handle the three hundred daily failures, and that someone is a headcount, a rota, and a training burden that the pilot business case never carried. Worse, the exceptions that survive to production are the hard ones — the easy tasks were exactly the ones the pilot automated first. The residual work left for humans is more difficult, not less, which inverts the comfortable assumption that automation leaves people with the simple jobs. It leaves them with the ones the machine could not do.

Hidden cost line	Why the pilot hides it	Empirical anchor
Integration labor	Vendor quotes hardware, not the work of wiring it into the operation	Overruns ~30% of budget; the dominant line
Network & facility	Buyer's responsibility; omitted from proposals	$30–150K per facility, rarely quoted
Middleware / legacy bridges	Pilot ran on a clean, isolated feed	Custom bridges add latency; single points of failure
Exception handling	Humans absorbed exceptions silently in the pilot	1–5% of volume; hundreds of events/day at scale
Software scaling	Sized for pilot throughput	Cascading failure above rated load
Timeline slip carry	Pilot had no committed capital at risk	Up to 12 months of carrying + foregone benefit

Table 2. The hidden-cost inventory. A buyer who asks a vendor to price each of these lines explicitly has already avoided most of the trap.

5. The Compounding Cost of Time

[ANALYTIC] Timeline slip is treated inside most organizations as a scheduling inconvenience — a project that will land two quarters late. Framed in capital terms it is nothing of the kind. Once capital is committed to an automation program, a slipped timeline carries a real, compounding cost: the carrying cost of the committed capital itself, plus the operating benefit the asset was supposed to be generating and is not. On a committed program of a few million dollars, a twelve-month slip is measured in hundreds of thousands of dollars of destroyed value before the system produces a single unit at target cost.

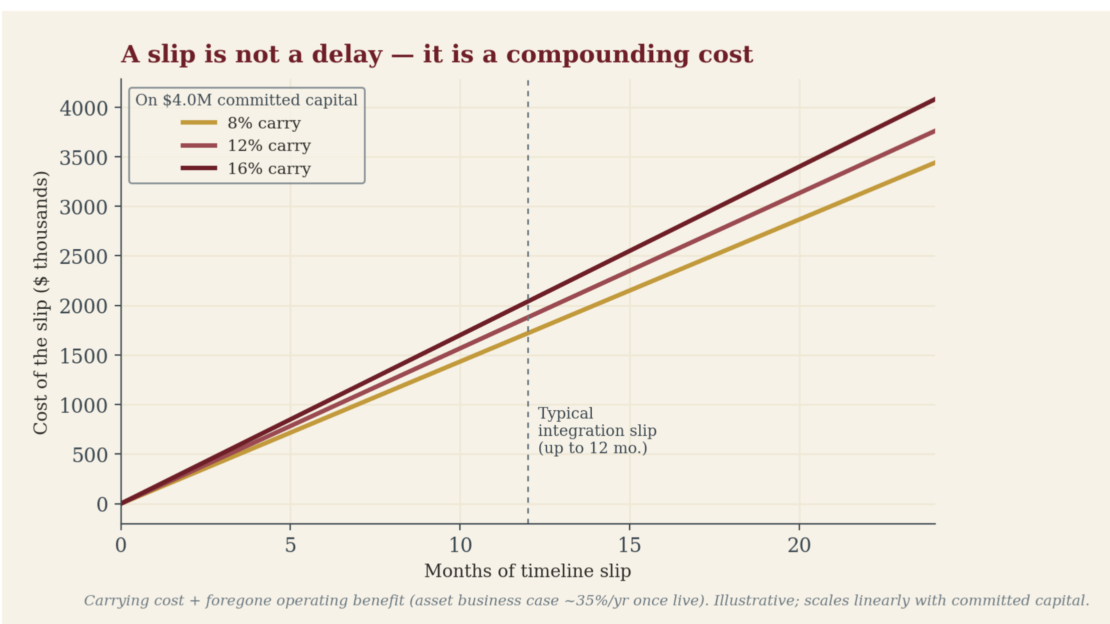

Figure. *A slip is not a delay. On $4.0M of committed capital, a twelve-month integration slip — well within the observed range — destroys value through carrying cost and foregone operating benefit whether or not the vendor ever misses another milestone. The line does not flatten; time is the enemy of committed, non-producing capital.*

5.1 Counterparty durability is now a deployment risk

There is a second, harder-edged form of timeline risk that the automation-buyer literature almost entirely ignores: the vendor may not survive the deployment. The mobile-robotics sector has entered a visible consolidation and distress phase, and an asset with a five-to-ten-year duty cycle purchased from an integrator that exits the business in year two is a stranded asset regardless of how well the pilot performed.

[GROUNDED] The evidence is recent and unambiguous. In December 2025 Zebra Technologies began winding down the autonomous-mobile-robot division it had built around its $290 million acquisition of Fetch Robotics in 2021. iRobot, one of the few household names in consumer robotics, entered a restructuring in which a Chinese creditor took control of its debt. Independent market trackers characterized 2025 as a turbulent year for warehouse automation in which several vendors filed for bankruptcy, divested, or closed outright. Consolidation among the strong (Symbotic's $200M purchase of Walmart's robotics unit; ABB's and Toyota's moves on navigation and materials-handling specialists) is the same signal read from the other side.

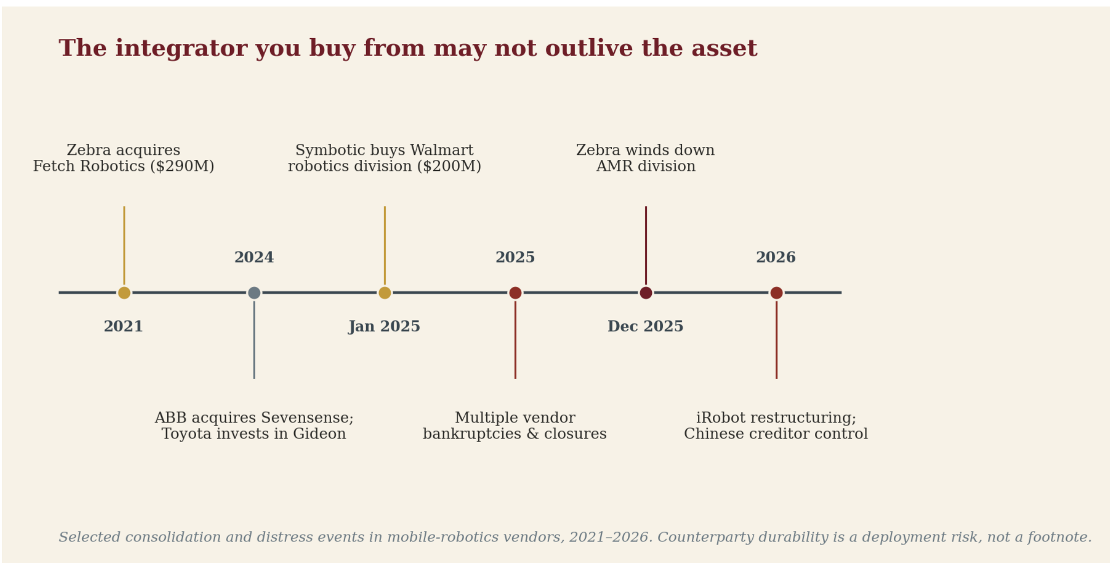

Figure. *A selective timeline of consolidation and distress among mobile-robotics vendors, 2021–2026. For a buyer, counterparty durability belongs in the underwriting, not the appendix.*

Integrator due-diligence question	What a weak answer tells you
Show me three deployments past 24 months at my volume	Your pilot is their R&D, not their product
What is your installed base and its trend?	A shrinking base precedes an exit
Who owns the integration IP if you disappear?	Vendor lock-in becomes a stranded asset
What is your funding runway to the unit I need?	Runway shorter than my duty cycle is my problem
What is the second-source path for parts and service?	No second source is a single point of failure

Table 3. Integrator underwriting. These are equipment-buyer questions, not procurement checkboxes. They price the risk that the counterparty does not outlive the asset.

6. The Organizational Immune Response

[ANALYTIC] There is a final reason pilots stall that no cost model captures, and it is the one operators recognize fastest. An organization tolerates a pilot precisely because it is a pilot — small, contained, and threatening no one. The immune response activates at the exact moment the pilot stops being a curiosity and starts being a change to how people work, are measured, and are staffed. This is not irrational resistance to be lectured away. It is the predictable behavior of a system defending its own equilibrium, and it is most dangerous precisely when the program is succeeding, because success is what makes the threat real.

The consulting literature has converged on a useful decomposition, sometimes rendered as a ten-twenty-seventy split: automation success is roughly ten percent algorithms, twenty percent data and technology, and seventy percent people, process, and organizational change. **[GROUNDED]** Gartner's finding that a large share of AI projects fail on data quality, and its projection that a substantial fraction will be abandoned after the proof-of-concept phase, points at the same reality from the technical side. **The pilot proves the ten percent. The stall happens in the seventy.**

A related fragility deserves explicit mention because it masquerades as success during the pilot. Many deployments run on the expertise of a single person — the engineer who tuned the system, the operator who knows its quirks. An operation with ten such applications does not have an automation capability; it has ten single points of failure. When that person leaves, change's role, or is stretched across too many programs at once, the deployment silently degrades. The workforce dimension of automation ROI — who stays, who leaves, and what tacit knowledge leaves with them — is large enough to be its own subject and is the subject of the next paper in this series.

7. Sequencing Discipline: A Framework Built for Buyers

Everything above argues that the pilot answers the wrong question and hides the cost of the right one. The remedy is not to abandon pilots — it is to *stop asking pilots to prove things they cannot prove*, and to stagger capital commitment so that money is put at risk only in proportion to the ignorance that has actually been retired. The discipline is borrowed directly from how experienced capital allocators fund anything with an uncertain payoff: small stakes to buy information, larger stakes only once the information justifies them, and a pre-committed willingness to stop.

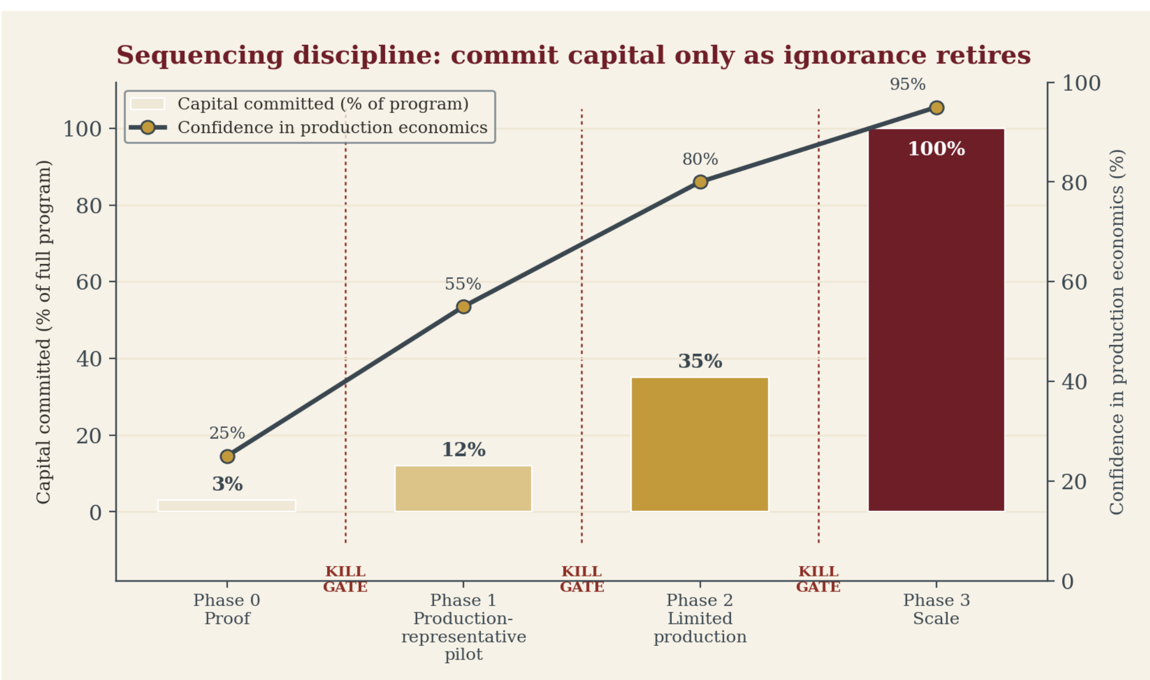

Figure. *The staged-commitment curve. Capital committed (bars) should track confidence in the production economics (line), never lead it. Between every phase sits a kill gate with pre-agreed criteria — the discipline is deciding the exit conditions before the sunk-cost instinct sets in.*

7.1 The phase-gate model

The framework replaces the single make-or-break pilot with four phases separated by explicit economic gates. Each phase has one job: to retire a specific, named uncertainty at the lowest cost that will settle the question. Critically, the pilot phase is redefined. It is no longer a demonstration; it is a deliberately unsubsidized test designed to withdraw the three subsidies on purpose — arm's-length vendor, no champion hovering, real staff, real data, and real exception load — so that the subsidy gap is measured before capital is committed rather than discovered after.

Phase	Question it retires	Exit / kill criteria
0 · Proof	Can the technology do the task at all?	Exit: task demonstrably feasible. Kill: fundamental capability gap.
1 · Production-representative pilot	What is the true unsubsidized performance and cost?	Exit: unit economics hold with subsidies removed. Kill: subsidy gap breaks the case.
2 · Limited production	Does it hold across shifts, seasons, and real staff?	Exit: sustained duty cycle at target cost. Kill: uptime or exception load off-model.
3 · Scale	Can the operation and the vendor sustain it?	Exit: replicate. Kill: integrator or org constraints cap it.

Table 4. The phase-gate model. Each kill criterion is written before the phase begins. A gate with no pre-agreed kill criterion is not a gate; it is a hope.

> *A staged commitment converts an all-or-nothing bet into a sequence of small, cheap, reversible questions. The most valuable thing a phase gate produces is not a green light. It is permission to stop.*

[ANALYTIC] The framework's deepest benefit is cultural. The single-pilot model creates enormous pressure to declare success, because the alternative is to admit a visible failure after a visible investment. The staged model makes stopping cheap and normal: a program killed at Phase 1 has spent roughly a tenth of the full commitment and has bought a genuinely valuable answer. Reframing a kill as a successful, low-cost information purchase — rather than a failure — is what breaks the sunk-cost dynamic that keeps doomed deployments alive long past the point anyone still believes in them.

8. The Capital Allocator's Verdict

The humanoid and warehouse-automation wave is real, and the asset thesis of REA·004 stands: durable value in embodied automation migrates toward whoever owns the model, the data, and the deployment competence, not the hardware. But the deployment floor is where most of the capital will actually be won or lost, and the record is clear that most of it is lost. The three scenarios below frame the decision for a mid-market buyer weighing a first serious automation commitment.

Scenario	What it looks like	Verdict
Run the classic pilot	One make-or-break pilot, vendor-led, measured charitably; scale on a green demo	Avoid. This is the trap. You will discover the 3–5× after committing.
Stage the commitment	Four phases, subsidies withdrawn in Phase 1, kill criteria pre-agreed, integrator underwritten	Proceed. Ignorance is retired before capital. Most of the loss is designed out.
Wait for maturity	Defer until reference deployments and residual-value data exist	Defensible, with a cost. Real risk of ceding a durable operating-capability lead to peers.

Table 5. The verdict table. The middle row is the entire argument of this paper compressed to one line.

Bottom line. The pilot that works is not evidence the deployment will. It is, most often, evidence that the deployment has not yet been tested. Withdraw the subsidies deliberately, price the hidden lines before you commit, underwrite the integrator like the counterparty it is, and stage the capital so that stopping is always cheap. Do that, and the pilot trap is not a risk to be feared but a mechanism to be exploited — because most of your competitors will walk straight into it.

This paper is REA·005 in the Reality register of Applied Futures Research. It follows REA·004, The Humanoid as Capital Asset, and precedes REA·006, Own, Lease, or Rent the Robot, and REA·007, The Crew and the Machine.

References & Notes

Figures are illustrative models built on the empirical anchors cited below unless labelled [GROUNDED]. Cost multiples, overrun ranges, and infrastructure figures are drawn from independent market analysis and enterprise surveys published between late 2025 and mid-2026; deployment and vendor events are matters of public record. Sources are listed by the claim they support; where a figure is a modelled projection it is tagged [PROJECTION] in the text.

1. Enterprise AI/automation pilot-to-production surveys, Feb–Mar 2026: ~78% with pilots vs. ~14% at production scale; 64% hit blocking issues on scaling, 72% of those stalled >6 months; manufacturing at 13–16%. Digital Applied; agility-at-scale.com.
2. "Pilot purgatory" in manufacturing (~70% trapped in pilot mode) and the 12–24-month benefit-realization horizon. Alpha Technical Solutions, 2025.
3. Gartner: share of AI projects abandoned after proof-of-concept; ~85% failing on data quality; the BCG 10–20–70 principle. As reported by Opsio and Astrafy, 2025–2026.
4. The tail-of-distribution mechanism (1–5% of inputs; hundreds of failures/day at 10,000 tasks). Digital Applied, Mar 2026.
5. Integration overruns (~30% of budget; up to 12-month slips), legacy-WMS interface gaps and middleware fragility. Mordor Intelligence, warehouse automation market analysis, 2026.
6. Network/facility upgrades of $30,000–$150,000 per facility rarely included in vendor quotes; sub-10% of floor device issues reported to IT; execution-software scaling failure above rated load. The Network Installers; warehouse automation statistics, 2026.
7. AMRs are "not plug-and-play"; the differentiator is orchestration, not the robot. Logistics Viewpoints, Jan 2026.
8. Vendor consolidation and distress: Zebra winding down its AMR division (built on the $290M Fetch Robotics acquisition, 2021), Dec 2025; iRobot restructuring under Chinese creditor control; Symbotic–Walmart ($200M) and ABB/Toyota moves. The Robot Report, Jan 2026; MCF Corporate Finance, 2025; Interact Analysis via Supply & Demand Chain Executive, 2026.
9. Early humanoid deployment reality: 2025 as the pilot year; Western deployments in the dozens; Agility Digit (~$250K/unit, ~100 units) in GXO and Amazon facilities; "measured in years, not months." Contrary Research; 1023 Jack Q2-2026 status; industry deployment reviews, 2025–2026.
10. REA·004, "The Humanoid as Capital Asset," Applied Futures Research, June 2026.