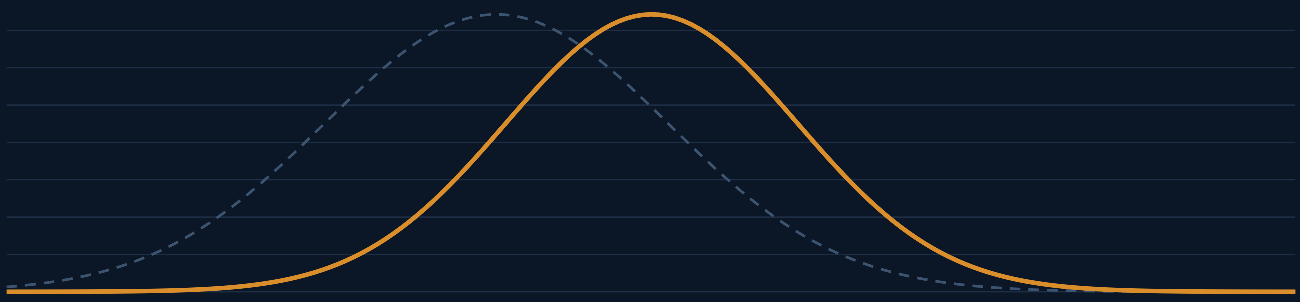
PHYSICAL AI FRONTIERS
ROBOTICS TRACK · EDITION R02 · JULY 2026

The Last Inch



Kinesthetic sensing, dense clutter, and the quantity a camera cannot measure.

Lise DELOBEL
GenAI project manager & consultant — Physical AI, digital twins, and applied ML for industrial and extreme-environment operations. MIT IDSS Data Science & Machine Learning certificate.

LISEDELOBEL.FR · LISE.DELOBEL@ME.COM

01 / EXECUTIVE SUMMARY

The quantity you cannot see

Classical industrial automation handles planned contact well, and has for decades — welding, polishing and force-controlled insertion all depend on it. What it handles badly is unexpected contact in an environment whose geometry cannot be specified in advance. Dense warehouse storage is exactly that environment, which is why it resisted automation long after the arm itself stopped being the obstacle.

The interesting finding from the first production deployment at scale is not that robots learned to touch things. It is what they had to touch things in order to find out. A camera can measure the geometry of a packed bin, even through the elastic bands across its face. It cannot measure how far a bagged garment will compress. That quantity determines whether an item fits, and the only way to obtain it is to push and feel the result.

FINDING 01

The old failure mode was binary

Amazon's applied science lead describes the prior state of the art plainly: on unexpected contact, industrial robots either emergency-stop or smash through, and often do not register the contact at all because they cannot sense it. Neither branch survives a crowded shelf.

FINDING 02

Perception is biased, and touch corrects it

Vision-only estimates of the space a sweep could create underestimated it by 36 mm on average. Estimates informed by kinesthetic feedback from earlier sweeps in the same bin were effectively unbiased, at 0.15 mm. Compressibility is invisible; it is only measurable by contact.

FINDING 03

The deployed system is a hybrid, not a learned policy

The published architecture uses learned perception and learned risk prediction, with engineered behaviours and task planning executing the motions under hybrid force/position control. The authors describe learning the behaviours themselves as the logical next step, not as what shipped.

FINDING 04

Difficulty concentrates in the contact-rich subset

Across 100,000 stow attempts, insertions into existing free space succeeded 90.7% of the time. Sweeps, which must create space by contact, succeeded 66.7% and were unproductive four times as often. The average conceals two very different tasks.

Two definitions used throughout

End-of-arm tooling (EoAT) is the task-specific hardware at the wrist — jaws, plank, sensors — as distinct from the arm; in contact-rich work it carries most of the engineering difficulty. Hybrid force/position control regulates force along the axes where contact matters while holding position on the rest, rather than treating force as a threshold that triggers a stop.

Four things made contact tractable

Force control in robotics is decades old. What changed is that the surrounding stack — occlusion-tolerant perception, high-rate control and learned risk estimation — became good enough to make deliberate contact worth attempting at production rates.

85.9%	224 vs 243	0.24%
STOW SUCCESS	UNITS PER HOUR	DAMAGE RATE
Across 100,000 annotated production attempts, with 9.3% unproductive and 4.8% defects	Robot against human stowers on the same floor, March 2025	Items damaged per stow attempt — the constraint that governs commercial viability

01 Perception learned to see through the occlusion

Fabric pods are covered by translucent elastic bands that defeat off-the-shelf 3D perception. Learned stereo depth and segmentation models, trained partly on synthetic imagery to treat the bands as invisible, cut depth error from 2.14 cm to 1.16 cm. Without that, nothing downstream is possible.

02 Force sensing became production tooling

A six-axis force-torque sensor sits at the base of the tool's extendable plank, a single-axis load cell regulates jaw force, and the control loop runs at 500 Hz with the planner reading feedback at 100 Hz. This is engineered hybrid force/position control, not a learned controller.

03 Learning entered where it pays, not everywhere

Learned models handle bin state, free-space estimation and — in a tabular gradient-boosted risk model — the probability that a given behaviour succeeds on a given item and bin. That risk model alone raised stow rate by about 7% over a frequentist baseline. Execution stayed engineered.

04 The task is enormous and physically punishing

A single warehouse holds more than a million unique items, and stowing is performed manually more than 14 billion times a year. Top rows sit high enough to need a step ladder; the paper estimates that handing only the top rows to robots would raise human stow rates by 4.5% and remove the ladders.

Source: Hudson et al., “Stow: Robotic Packing of Items into Fabric Pods,” Amazon Robotics, arXiv:2505.04572, 2025 · Amazon Science, May 2025.

Four kinds of touch

Once contact stops being a fault condition it has to be classified, because each kind carries different information and justifies a different amount of risk.

TYPE 01

Contact avoided entirely

Where sufficient free space already exists, the best behaviour is still not to touch anything. Direct insertion is both the fastest cycle and the most reliable, and a well-designed system prefers it whenever the geometry allows.

TYPE 02

Exploratory contact

Deliberate touching to find out what cannot be seen. Compressibility, item rigidity and whether a neighbour will yield are all invisible to a camera and recoverable only by pushing and reading the response.

TYPE 03

Functional contact

Contact that performs the task: sweeping existing stock aside to open a gap. Here force is not a disturbance to reject but the actuation itself, and the control problem is to apply enough without applying too much.

TYPE 04

Contact with a person

A different governing logic entirely. Limits are set by biomechanics and by standards rather than by what the inventory tolerates, and they are far lower than anything in the bin.

THE LADDER OF RESPONSES

01

Free-space planning

Plan a path through empty space and assume nothing is in it. Any contact is a fault.

02

Collision avoidance

Sense obstacles in advance and route around them. Still treats touch as something to prevent.

03

Passive compliance

Add a compliant element so incidental contact is survivable without being understood.

04

Force in the loop

Measure contact continuously and use it both to steer and to learn what could not be seen.

The strategic read

Most operators sit on rung two and believe they are close to rung four. The gap is larger than it looks, because rungs three and four differ in kind: compliance makes contact harmless, while force in the loop makes contact informative. Only the second lets the machine discover properties of the world that no sensor can observe at a distance — and only the second requires you to rebuild your definition of a successful cycle.

A system built around one invisible variable

Contact-rich stowing into fabric pods — Amazon Robotics

The deployment worth studying is a stow system placing items into fabric storage pods. It is described in a published engineering paper rather than only in press material, which makes it unusually checkable — and the paper is candid about what does not work.

THE PREMISE

The task as its engineers state it

Items are stored heterogeneously to maximise density, in bins fronted by translucent elastic retention bands, with more than a million unique items in a single warehouse and no reliable prior information about new products. The design target is to stow 80% of items at 300 units per hour, running more than 20 hours a day, seven days a week.

THE TOPOLOGY

Specialised machines, not one general robot

The architecture deliberately favours specialised manipulators. A six-axis arm on a two-axis gantry carries the stow tool; a separate three-axis gantry robot opens the elastic bands with a hooked tool and its own six-axis force-torque sensor; a SCARA robot reorients items beforehand; one human inductor quality-checks items and feeds three robotic workcells.

THE TOOLING

The tool is the argument

The stow tool has parallel jaws with conveyor belts built into their faces, so an item is fed in by conveyor transfer and ejected the same way — the jaws never enter the bin, which would cost storage density. A thin retractable aluminium plank, instrumented with a six-axis force-torque sensor, does the in-bin manipulation. Because the plank rather than the item creates the space, the item is shielded from the contact.

So what

The transferable lesson is that morphology can dissolve a control problem. A human stower does this bimanually, sliding the held item against its neighbours; the robot instead separates the jobs across dedicated hardware so the inserted item never participates in the contact. That decomposition is why the damage rate is low enough to be commercially viable — a hardware decision doing work that no amount of control sophistication would have done as cheaply.

Source: Hudson et al., arXiv:2505.04572, 2025 · Amazon Science, May 2025. Vulcan Pick, which uses suction tooling and a depth camera, is a separate system from the stow system described here.

Why residual coverage may carry disproportionate cost

An economic reading of the published cycle data — offered as hypotheses to test, not as established findings, since facility-level cost data is not public.

Picking and stowing automation is usually sold on the share of items a system can handle. The published cycle data suggests that framing is the wrong one, because the cost of a cycle varies sharply with which behaviour the situation forces.

THE SPLIT

The cheap cycle and the expensive one

A direct insertion into existing space averages 10.9 seconds and succeeds 90.7% of the time. A sweep that must create space averages 13.7 seconds, succeeds 66.7%, and is unproductive 23.7% of the time against 5.7% for direct insertion — and an unproductive cycle costs a full cycle of time and returns nothing.

THE MECHANISM

Rate is a portfolio problem, not a speed problem

Because a failed pod costs roughly 6 seconds to swap and a pod yielding nothing costs at least 12, the planner is really choosing a risk portfolio: fast-but-likely-to-fail against slow-but-reliable. Replacing a fixed heuristic with a learned model of success probability and cycle time raised the stow rate by about 7%, which is a scheduling gain rather than a mechanical one.

THE RESIDUE

The residue is structural

What resists automation is not random. Deformable packaging, thin items under about 2 mm, and compressible goods in bags share the property that they do not transmit force predictably, which is exactly what makes kinesthetic monitoring unreliable for them. Books are singled out in the paper as a recurring damage case during insertion.

So what — stated as a hypothesis

The proposition worth testing in your own operation is that value comes from absorbing complete categories rather than from raising a coverage percentage, because a category fully removed eliminates a process while a thin improvement spread across the catalogue leaves every station and shift pattern in place. This is a procurement hypothesis, not a published finding, and it should be measured against your own exception path before it is believed.

Source: Hudson et al., arXiv:2505.04572, 2025, Tables II-III. Comparative vendor pricing such as Brightpick's advertised US robotics-as-a-service rate from $1,990 per robot per month is specific to that vendor, carries a three-year minimum and a ten-robot smallest deployment, and is not comparable to the system described here.

Designing what the robot will not do

Task allocation, escalation and the limits of the ergonomics argument

The most consequential decision in a contact-rich deployment is the division of labour, and it is usually made as a residual — defined by whatever the machine failed to do — rather than specified in advance.

THE PRINCIPLE

Allocate by human physical cost

Human stow rates vary far more than robotic ones: people are fast with small items and slow when crouching for low bins or using a step ladder for high ones. The paper estimates that assigning robots only the top rows would raise overall human stow rates by 4.5% and eliminate ladder use — an argument that improves both safety and throughput without depending on the robot being faster.

THE HANDOFF

Design the failure that costs least

The system distinguishes an unproductive cycle, where the item stays in the tool and returns to the buffer, from a defect, where an item is dropped or damaged. Kinesthetic checks compare the space actually created against the item width and abort before insertion. That distinction is what allows the planner to attempt risky stows at all.

THE TENSION

Two claims that should not be merged

Amazon has announced 25,000 new European fulfilment roles within a wider investment programme. That is a forward-looking corporate statement covering facility expansion alongside robotics, and it is not evidence that robotics caused net employment growth. Reporting has separately noted falling headcount per facility. A credible programme states its own position rather than borrowing someone else's.

A boundary worth stating explicitly

A robot capable of controlled contact with inventory is not thereby certified for physical contact with people. Those are different regimes with far lower limits, governed by ISO 10218-1 and ISO 10218-2:2025 for industrial robots and cells, and by ISO/TS 15066:2016 where collaborative operation is in scope. Any deployment that puts people within reach inherits that obligation separately, and it is not satisfied by the contact capability itself.

Sources: Hudson et al., arXiv:2505.04572, 2025 · About Amazon, 2025-2026 · ISO 10218-1 and ISO 10218-2:2025 · ISO/TS 15066:2016, published and under revision.

The variable no camera can measure

Creating space in a densely packed bin — contact-rich end-of-arm tooling

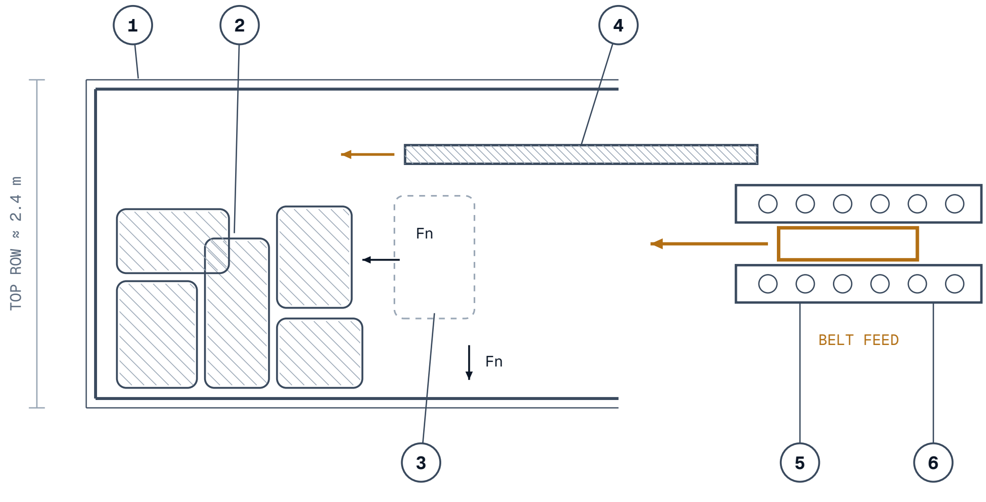
THE CHALLENGE

To place an item into a full bin, something already inside has to move. The planner must therefore answer one question before committing a cycle: how much space could be created here? Get it wrong optimistically and the cycle is wasted or an item is damaged; get it wrong pessimistically and the bin is passed over and density falls.

The difficulty is that the answer depends on compressibility. A bin of bagged apparel yields several centimetres under a modest sweep; the same bin geometry filled with boxed goods yields nothing. The two are close to indistinguishable to a camera — and the camera is also looking through translucent elastic bands stretched across the bin face.

> **Why the obvious solutions fail**
>
> Better perception does not resolve it, because the required quantity is not an optical property at all. Vision-only space estimates were found to underestimate the achievable space by 36 mm on average, with a standard deviation of 40 mm — a systematic pessimism that costs storage density on every bin. Nor does simply pushing harder: as Amazon's applied science lead notes, conventional industrial robots meeting unexpected contact either emergency-stop or smash through, and frequently cannot tell which has happened.

FIG. 01 – FABRIC POD BIN WITH STOW TOOLING, SIDE ELEVATION (SCHEMATIC)



1 Fabric pod bin, 1-in lip, elastic bands
2 Existing inventory — densely packed
3 Space created by controlled sweep
4 Extendable plank — 6-axis force/torque
5 Parallel jaws — load-cell grasp force
6 Jaw conveyors insert; jaws stay outside bin

Measure it by pushing on it

The design move, the trade-off accepted to ship, and the result in production

THE RESOLUTION

The turning point was to treat the sweep as a measurement as well as an action. As the plank consolidates items under a force-based admittance controller, the kinesthetic trace records how far the contents actually moved under a known force. That is a direct observation of compressibility, and it is written back into the bin's stored state so the next decision about the same bin is made with evidence rather than inference.

The effect on the estimate is the whole story. Vision-only prediction is systematically pessimistic by 36 mm. Prediction informed by a previous kinesthetic measurement in the same bin is effectively unbiased at 0.15 mm, with slightly tighter spread. Roughly a third of sweeps were selected using the kinesthetically informed estimate.

FIG. 02 — SPACE-ESTIMATE ERROR

SPECIFICATION — REPRESENTATIVE JOINT UNIT

Plank sensing	6-axis force/torque
Jaw force	Load cell, 80 N fixed
Control loop	500 Hz
Planner feedback	100 Hz
Perception	Stereo, learned depth
Depth error	1.16 cm

Published characteristics of the deployed stow system. The fixed 80 N clamp force is identified in the paper as a cause of damage to lightweight boxes — an acknowledged limitation, not a tuned optimum.

The trade-off accepted to ship it

Contact-rich sweeps are worse cycles by every measure: slower at 13.7 seconds against 10.9, less reliable at 66.7% success against 90.7%, and unproductive four times as often. The system accepts them anyway, because they reach inventory that direct insertion cannot. The engineering judgement was not that contact is better — it is that a worse cycle which sometimes works beats a better cycle that cannot be attempted, and the planner's job is to know which situation it is in.

THE OUTCOME — 100,000 ANNOTATED PRODUCTION ATTEMPTS

85.9%	90.7 / 66.7	500k+
OVERALL SUCCESS	DIRECT VS SWEEP	STOWS COMPLETED
9.3% unproductive, 3.8% amnesty, 0.8% other, 0.24% damage	Success rate for insertion into existing space against space created by contact	In one large fulfilment centre, at stow rates comparable to human stowers

Source: Hudson et al., "Stow: Robotic Packing of Items into Fabric Pods," Amazon Robotics, arXiv:2505.04572, 2025 · Figure 10 and Tables II-III. Drawing is schematic and generic to the tooling class, not a manufacturer drawing.

Buying a capability that touches things

Durations are indicative only and vary substantially with item mix, facility layout, safety classification and data availability.

PHASE 01 · 0–3 MONTHS

Catalogue what your process cannot handle

Before any vendor conversation, document the exception pile — the items your current process routes to a human. Photograph, weigh and classify them by packaging and compressibility. That pile is the specification, and in most operations nobody has ever written it down.

PHASE 02 · 3–9 MONTHS

Pilot on the residue, and instrument damage separately

Vendor demonstrations use tractable items. Insist on a trial against your documented exceptions, at your density, and measure damage and escalation accuracy independently of the robot's own success reporting. A system that fails honestly beats a system that fails quietly.

PHASE 03 · 9–18 MONTHS

Redesign the process, not just the station

Contact-rich automation changes what a successful cycle means and where exceptions surface. Rebuild task allocation, escalation routing and quality metrics around it. Programmes that install the machine into an unchanged process capture a fraction of the value.

PHASE 04 · 18 MONTHS +

Scale with a stated position

Fleet rollout, retraining as the catalogue shifts seasonally, and an explicit account of the employment consequence. Where people share the workspace, machinery-safety and functional-safety obligations apply separately and are not satisfied by the contact capability itself.

> **Organisational readiness**
>
> Contact-rich automation fails on the boundary between operations and engineering. Operations owns the exception pile and rarely documents it; engineering owns the tooling and rarely sees the pile. The most useful preparatory act — cheap, unglamorous, almost never done — is a month of systematically recording what the current process cannot handle.

What goes wrong, and what it costs

Ranked by how often it is missed rather than by severity.

RISK 01

Demo-set bias

Vendor trials run on items chosen for tractability. Coverage measured on a curated set says nothing about performance on the residue that actually drives your staffing.

RISK 02

Averages that hide two tasks

A single success rate conceals the split between insertion into free space and contact-rich space creation. Ask for the breakdown by behaviour, because the second is where the difficulty and the cost live.

RISK 03

Silent damage

A system that touches things can degrade stock without registering failure. Damage must be instrumented independently of the robot's own reporting, or it will not be seen until it is a trend.

RISK 04

Force limits set once

Published systems have shipped with a fixed clamp force that is acknowledged to damage lightweight packaging. Anything constant across a heterogeneous catalogue is a defect source waiting for the wrong item.

RISK 05

Deformable items break the sensing

Compressible and very thin items do not transmit force predictably, so kinesthetic monitoring cannot confirm what happened. These are the same items that resist automation in the first place.

RISK 06

Learned models drift

Risk and space models are trained on the behaviour distribution the current planner produces. Change the behaviours, the packaging or the season and the training data no longer describes the system.

RISK 07

Process left unchanged

Installing contact-capable tooling into a workflow designed around contact avoidance captures a fraction of the benefit, and is the most common failure in practice.

RISK 08

Human-proximity obligations

Capability for controlled contact with inventory confers nothing about contact with people. ISO 10218-1 and 10218-2:2025, and ISO/TS 15066 where collaborative operation applies, are separate and far more restrictive.

08 / WHERE TO START

Three questions worth asking this quarter

01 Has anyone catalogued what our current process cannot handle?

02 Which quantity do our systems infer that they could instead measure?

03 Would our system know when it had failed — and say so?

The second question generalises well beyond warehouses. Most automation carries an estimate it could replace with a measurement, and usually nobody has asked what that estimate costs.

NEXT IN THIS TRACK

Edition R03 — Tighter Than the Robot

Precision assembly, where the clearance between the parts is smaller than the robot's own positional uncertainty — and the two-point jam that makes the obvious insertion strategy fail.

SOURCES & METHOD

Primary: Hudson, Hooks, Warrier, Salisbury, Hartley et al., "Stow: Robotic Packing of Items into Fabric Pods," Amazon Robotics, arXiv:2505.04572, 2025. All success rates, cycle times, space-estimate errors, sensing specifications and stow-rate comparisons are drawn from that paper, principally Figure 10 and Tables II and III.

Secondary: Amazon Science and About Amazon, 2025-2026 · The Robot Report, May 2025. Company-reported deployment and employment figures are identified as such in the text and are not independently audited.

Standards referenced: ISO 10218-1 and ISO 10218-2:2025 · ISO/TS 15066:2016.

Method note: the 500,000 figure counts stow operations in a single fulfilment centre, not customer orders. Coverage percentages reported for different Amazon robotic systems are not like-for-like and are not combined here. Economic propositions in section 04 are labelled as hypotheses because facility-level cost data is not public. The section drawing is schematic.

Lise DELOBEL

GenAI project manager & consultant — Physical AI, digital twins, and applied ML for industrial and extreme-environment operations. MIT IDSS Data Science & Machine Learning certificate.

LISEDELOBEL.FR · LISE.DELOBEL@ME.COM

PHYSICAL AI FRONTIERS · ROBOTICS TRACK · EDITION R02 · JULY 2026