This Week in Robotics

Window: 2026-09-21 → 2026-09-27.

> THE LINE
>
> The infrastructure moved more than the models. Alphabet's Intrinsic open-sourced its production arm stack under Apache-2.0 and, a day later, Qualcomm agreed to buy PickNik, the company that maintains MoveIt 2 — the open manipulation stack changed hands and gained a major donor inside 48 hours. On the paper side the two most useful results were subtractive: seven non-Gaussian flow-matching priors turn out to do nothing, and a controlled WAM study finds that predicted futures act on actions through their temporal order, not their content.

1 Policy learning & VLAs

The Gaussian Is Enough

PREPRINT / TRI + Woven by Toyota + Cornell; no code, project page only

The most expensive negative result of the month, and it saves you an experiment you may have been planning.

- **Mechanism under test:** the source distribution of a flow-matching action head. Seven non-Gaussian alternatives to the standard Gaussian noise source, swapped into otherwise unchanged heads on LBM-1.0, π0.5 and GR00T-N1.5.
- **Eval:** ~100K simulated rollouts over 40+ tasks in two simulators (Drake and MuJoCo), plus 1,250 hardware rollouts on two bimanual Franka FR3 setups. Statistics are STEP and Welch tests with Bonferroni-corrected compact letter displays — more inferential machinery than any other table this week.
- **Result:** at 25% of the data and above, every prior is statistically tied. The one exception is at 5% data, where one prior gives 25.9 against Gaussian's 18.9. Trained from scratch, prior A_pre scores 276/1000 against Gaussian's 175/1000 — and the ordering *reverses* after fine-tuning (290 vs 350). BRIDGER was NaN-unstable.
- **Where it's soft:** no code, no seed count, and priors that change the architecture are excluded. The claim is narrower than "priors don't matter": swapping the noise distribution under a fixed head buys nothing once you have data.
- INFERENCE \2192 the reversal after fine-tuning is the transferable lesson — a prior that helps a randomly initialized head can hurt once a pretrained backbone supplies the structure. Be careful porting from-scratch findings into fine-tuning recipes.

What Matters in Designing World Action Models

PREPRINT / Samsung Robotics eXperience + Samsung R&D Beijing + Wuhan + PKU; no code stated

A controlled study in the style of W38's ACT re-run, but forward-looking: six video-action causal structures, eight latent representations, four objectives, base model held fixed within each axis.

- **The intervention that earns the paper.** Rather than comparing separately trained policies, they intervene on the generated future latents inside a trained route-enabled policy. Content corruption (noise mixed in at 0.10/0.25/0.50, per-channel mean and variance preserved) changes action predictions by **under 1%** and does not move success. Temporal reversal of the future slots changes actions 12.79–14.36% and cuts OOD success **24.24–32.37%**. The future is acting as a clock, not as a picture.
- **Latents:** inter-frame latents (DA3, VGGT-Ω, V-JEPA, Video-VAE) win in-distribution; framewise latents (DINOv3, SAM3, Image-VAE) win OOD, with the largest margins under sensor noise and viewpoint shift. Linear probes for expert actions track the same split. Temporal structure pays when the policy learns it, and becomes brittle when the encoder pre-bakes it.
- **Objectives:** BC-only is the best configuration on RoboCasa-GR1. Every auxiliary objective hurts ID. On LIBERO-Plus, video generation lifts BC-only 77.96 → 81.22 (camera-viewpoint shift +13.32), FDM is marginal, IDM slightly negative. Joint training of everything from step zero is 4.04 points *below* BC+VG; a staged schedule (BC+VG for 80% of optimization, dynamics objectives only in the last 20%) reaches 83.15.

- **Where it's soft:** moderate scale by design, no seed bands on the sim tables, no code stated. The real-robot leg is offline action prediction on a held-out DROID split, not closed-loop success, and every delta there is small (Acc@0.1 51.83 → 53.78). It has the controls the benchmark-validity thread wanted and not the bands.

Grounded Action Model

PREPRINT / Northwestern + UW + NUS + Microsoft; project page / no code stated

The most direct attack yet on the "VLAs replay layouts, they don't follow language" problem, and it does not solve it with scale or a better language encoder.

- **Mechanism:** replace the VLM or video-generation backbone with a *frozen promptable 3D grounding model* (WildDet3D). Language, 2D points or 2D boxes all resolve to the same object-centric representation. Image tokens are a 16×16 pooled feature grid **masked to the task objects and the robot's URDF silhouette** — everything else is zeroed. Detection tokens carry each object's 512-point cloud plus metric centre, extents and 6D rotation. An MM-DiT predicts absolute joint-position chunks by flow matching; only the action head trains.
- **Eval:** RoboTwin 2.0, single-task, 50 clean demos and 100 rollouts per task. 62.96 Easy / 47.62 Hard / 55.29 avg, against 52.00 for Spatial Forcing and 30.40 Hard for the next best entry. Note the shape: FastWAM is 77.8 Easy and **1.9** Hard; DP3, which also eats point clouds, goes 55.2 → 5.0. 3D input alone buys nothing — filtering to the grounded objects does.
- **LIBERO-PRO** is the number that matters. Under the *Task* perturbation — the instruction is replaced by one designating a different object already present in the scene — every baseline scores ≤0.11 on all four suites. π0.5 gets 0.01 on Spatial-Task. GAM gets **0.88**. Average over 16 settings, 0.61 vs 0.53 for π0.5.
- **Real:** bimanual YAM, 20 trials, target given by click/box for GAM and by language for π0.5. In distribution both hit 19/20. Under visual shift GAM keeps 17/20, π0.5 falls to 4/20. On a Franka with a Molmo2 planner feeding points: 64.7% ID and 49.8% OOD step completion over four long-horizon and memory-dependent tasks, against 24.0% OOD for π0.5.
- **Where it's soft:** GAM *trails* on the Obj perturbation (appearance and size) — 0.50 on LIBERO-10, 0.69 on Goal — because size changes move the grasp and point clouds do not fix that. Grounding errors propagate with no recovery path, and object-centric filtering deletes unselected obstacles from the observation, which is a safety problem nobody measured. The backbone needs a per-domain fine-tune (cheap and label-free, from simulator-exported boxes and depth). No code.
- INFERENCE \2192 the Task column is the cleanest evidence yet that published VLA language-following is mostly layout memorization, and that the missing ingredient is metric grounding rather than a bigger language model. The catch is that GAM buys it by *deleting* the rest of the scene — a different policy class, not a drop-in fix.

Training-free Behavior Cloning / Behavior Predictive Control

PREPRINT / Stanford, Schwager group; TRI-funded; no code statement

The week's best one-GPU item in this section.

- **Mechanism:** no policy training. Demonstration windows sit in a Hankel-structured bank. At run time: (1) a fitted retrieval metric — action-ridge, or LDA-style whitening over random Fourier features — picks windows whose *histories are predictive of their continuations*, not merely visually similar; (2) a small constrained least-squares finds coefficients g reconstructing the live observation–action history over those windows, with $1^\top g = 1$; (3) the same g is applied to the stored continuations, plus a closed-form ridge residual fit on the same support. The affine constraint is what separates this from weighted kNN; Proposition 1 gives the error decomposition motivating it.
- **Results.** Real xArm6, 40 trials per task, against π0.5 fine-tuned on the same demos: Coffee 36/40 vs 39/40, Drawer 38/40 vs 40/40, **Switch 34/40 vs 8/40**. Switch is a perceptual-aliasing task — once the bag is mid-transport, a single-observation policy cannot recall which plate it came from. BPC retrieves over a history and adds a task-progress prior, so it can.
- MimicGen, 300 rollouts per task, mean 73 against 77 for π0.5 and 76 for SDP; it loses badly on Nut (13 vs 31) and Stack3 (50 vs 85). Dexterous suites with PointNet inputs: 63 DexArt / 70 Adroit.
- **Cost:** 30–120 s to fit on one RTX 4090 against 20 h for π0.5; 100 Hz inference on a 4090 and **75 Hz on an 8 GB Jetson Orin Nano**, fitted on-board. 16×16 raw pixels were sufficient on hardware — no image encoder at all.
- **Where it's soft:** no code statement. The ablations show the residual correction carries most of the gain (removing it drops the mean 73 → 45), so "training-free" is doing work in the title — both the residual and the retrieval metric are fitted. Bank cost

grows with the dataset; the analysis bounds only the uncorrected prior and says nothing about closed-loop stability. Single-task throughout.

Cluster: three ways to make the action interface carry more than coordinates. Direction-Scale Decomposition (KTH + Stanford, project page only) tokenizes unit direction and magnitude separately, which removes speed and normalization sensitivity and is tokenizer-agnostic; +13.3 pp across 120 real trials. GALA (Tsinghua + Shanghai Qi Zhi; CODE) augments image-based latent actions with 3D end-effector point-cloud transitions under a shared codebook across human hands, dexterous hands and grippers — 68.3% on RoboCasa-GR1 under multi-embodiment co-training (+12.6 over GR-1-only) and 75.5% over four real XHand tasks at 50 trials each, against 71.5 for HARP-VLA and 68.0 for π0.5. MachEmbodied-U0 (Li Auto) puts understanding and generation experts in one MoT with RGB, depth, normals and flow as joint visual dynamics, at 4,200 h and 368 accelerators of pretraining — out of reach, listed for completeness.

2 Manipulation, dexterity & sensing

The theme: **tactile hardware is migrating to training time, and the one paper that keeps it at run time argues for bandwidth over spatial density.**

ME-Dex 1.0

PREPRINT / Li Auto; CODE

Is the clearest instance, and its most useful number is an ablation the authors report plainly rather than bury.

- **Mechanism:** tactile as a *predicted future observation* alongside video, not as a conditioning input. Three-expert Mixture-of-Transformers (Video, Tactile, Action) trained jointly with flow matching, coupled by "H-Bridge" shared attention in intermediate layers only. A Canonical Hand Model maps grippers and dexterous hands onto one human-hand template; a Unified Tactile Autoencoder, pretrained separately and frozen, gives them a shared latent.
- RoboTwin, 50 tasks, mixed Clean+Random training: 91.56 / 91.92 against Fast-WAM 89.22 / 89.22 and π0.5 52.76 / 60.74. Adding tactile *conditioning* to the video-action model takes Random 86.90 → 89.54; adding future tactile *prediction* → 90.60; adding H-Bridge → 91.92.
- **The ablation:** zeroing the tactile input at inference on the trained model gives **91.70 / 91.74** — indistinguishable. The model keeps its future-tactile prediction and its tactile-supervised training; it just stops reading the sensor. On the Clean-only training run they zero tactile in both evaluations and still post 89.6 / 68.1 (avg 78.9 vs 71.3 for OLA-Sem).
- INFERENCE \2192 on RoboTwin, tactile *supervision* is worth ~2.4 points and tactile *sensing at run time* is worth approximately zero. A paper reporting only the 91.92 would have read as a tactile-sensing result.
- **Where it's soft:** single training seed per experiment, stated. The RoboTwin and DexJoCo tactile data is *synthetic*, recorded off simulator force sensors during trajectory replay by their "Agentic Tactile Data Engine" — so the zero-input finding may say more about simulated tactile than about real contact. Real-robot evaluation (SO-101 with a PaXini PX6AX; XR5-10-FS arms with Xynova Flex2 hands) is qualitative only. On ManiFeel, which has real 3-axis force, the gain is larger and non-uniform: 70.0 vs 55.0 average, Power Plug 58 → 88, but Gear Assembly 66 → 60.

Three more this week put tactile hardware on the training side only: HapticWAM distills imagined touch into a WAM that deploys without the sensor, ZeroTouch uses tactile as supervision to estimate a grasp-dependent compression target from vision, and VT-Bridge converts a foundation VLA into a tactile one by residual adaptation. All three are abstract-only and unverified here; flagged as a pattern, not as four confirmed results.

PolyUMI

PREPRINT / Northwestern + TU Darmstadt + Hessian.AI; hardware, electronics, firmware, fabrication files and software all released

Is the week's most complete release. The idea is mechanical, not algorithmic: **one modular sensing finger physically moves between the handheld collector and the robot**, so the sensing geometry at demonstration and at deployment is identical rather than merely calibrated. $236 per finger, ~4 h to assemble, ~10 min to swap. Optical tactile via a 7-layer VHB-tape stack over a curved mirror, plus a 16 kHz contact microphone; VisTA fuses 294 tokens through an 8-layer encoder and an 18-layer flow-matching DiT. Tactile shape 92.3% on held-out scenes; object-in-box 44% → 80% with audio; slip control 8/10 against 2/10 vision-only. An RTX 4060 trains the classifier in 11 minutes. **Soft:** two real manipulation tasks only, and vision-only diffusion policy *wins* the lightbulb one; manipulation success appears only as a bar chart; the PolyTouch baseline is a reimplementation.

Res-HIL

PREPRINT / Siemens + TUM; **no code, videos only**

Human-guided residual RL on a single UR5e with a SpaceMouse. The mechanism worth stealing: each intervention is used twice — as an explicit residual BC target, and as a decaying penalty on the autonomous transitions that *preceded* it, so the critic learns that the states leading into a takeover were already bad. Frozen ACT base, zero-initialized residual output layer, TD3+BC online, ResNet-10, 7 Hz, 20 demos. **Three seeds × 50 fixed initial conditions**, spread under 4 pp — the first multi-seed real-hardware HIL result this briefing has seen. At 10 minutes of online data: 100/64/50/66/92 against HIL-SERL's 90/30/10/0/0; final 100/96/100/88/100. The delta is entirely in the cable tasks, where HIL-SERL scores 0%. **Soft:** ResFiT collapses to 0–4%, which suggests a weak baseline; HIL-SERL is *faster* in cycle time on the insertions; the vent-task margin depends on the authors' stricter success criterion; FTC and HiL-ResRL are not compared. No code is the cost.

The Cartesian Hand

PREPRINT / Duke; open-source "will" — not yet released

Seven prismatic joints, no revolute anywhere: two vertically stacked parallel grippers whose separation is also actuated, and four independently translating fingertips. The payoff is that `p = Jq` with a **constant** 12×7 Jacobian — no configuration-dependent finger kinematics, so manipulation composes from five linear primitives. Rack-and-pinion on Feetech-3915 servos, dovetail rails, 65 mm fingertip travel, 850 g, ~$500 complete ($30 in PLA for the structure, ~2 h assembly). 35 objects — threaded caps, syringes, pipettes, pliers, screwdrivers, triggers — **350/350 trials successful**, joint-feedback contact detection only, no vision. Transfers unchanged to a Duke Humanoid V2, and two hands do a bimanual pipetting procedure where neither arm is a fixture. **Soft:** 350/350 is a repeatability claim under prescribed initial poses with per-object parameters tuned in up to five setup trials; it is not a generalization claim. The design exploits the fact that the *object* constrains the relative motion (thread, pivot, plunger, trigger) — outside that class it is a parallel gripper. Files not out yet.

Also worth naming, both proprioception-only: Real-Time Force Regulation for Whole-Hand Dexterous Grasping turns the hand into its own contact sensor using a precomputed object SDF plus tracked 6-DoF pose, reallocating contact forces every cycle through a convex QP (OSQP, linearized friction cones, 10 N minimum normal) across fingertips, sides, dorsal surfaces and palm — no tactile hardware. CoPRE predicts contact-free torque three steps ahead from commands alone and scores the residual, trained only on contact-free data, for low-cost arms with no force or tactile sensors. And Simple Torque-Observation Alignment (abstract-unverified) aligns simulated and measured torque scale, offset and noise so torque becomes a usable RL observation on a direct-drive gripper — the cheapest sim-to-real contact recipe this week, if it holds up.

3 Platforms & embodiment

Demonstrated, with a method: Whole-Body UMI

PREPRINT / ZJU + HK Embodied AI Lab + CUHK; project page only

The contribution is a decoupling: a diffusion policy learns task semantics from **native UMI demonstrations with no body trackers**, while a separate task-agnostic motion generator learns whole-body coordination from ~105 h of retargeted mocap it never sees paired with the task. They meet at the end-effector trajectory. Three asynchronous layers: DP at 10 Hz, WB-UMI at ~1 Hz, SONIC whole-body controller at 50 Hz, with trimming and blending across replanning boundaries and measured-state feedback into the motion history.

Real Unitree G1, 200 UMI demos per task collected in ~90 minutes, 10 trials each: **drawer closing 9/10, shelf pick-and-place 8/10, ball toss 3/10, loco-pick-and-place 4/10.** Offline, EE look-ahead cuts position error 5.77 → 4.25 cm on unseen action categories. The honest parts are prominent: the drawer miss is an embodiment gap (a human crouches, G1 cannot reach), the ball-toss failures are gripper-release latency, and the Loco-PnP failures are a drift → OOD-observation → worse-EE-prediction spiral, with an ankle-roll trace showing the controller not tracking its own reference. n=10 per task, one training run.

Staged or unmeasured capability, all trade press:

- **Tesla Optimus at several hundred units per week** (The Information, 09-25, unnamed sources, Tesla declined to comment). Up from a few dozen in Q2, targeting >1,000/week by end-2026. Single-source and unconfirmed — but the *bottlenecks* named are the useful part: touch-sensor glove durability (replaceable sensing components planned for 2027), and hands that need >100 screws each with manual assembly and station alignment problems. 500,000+ hours of training data from teleop, motion tracking and human recordings across Colorado, Arizona and Florida hubs. V3 is explicitly a development unit, not the product.
- **AGIBOT reports 20,000 cumulative robots built** — A-series 5,638, X-series 8,757, G-series 5,605; company-reported via X, unaudited. Production, not deployment; 300+ went to a theme park.
- **IFR counts ~7,000 humanoid sales worldwide in 2025** — third-party, not vendor-reported, and IFR's secretary general says many buyers were research institutions or companies collecting AI training data rather than productive users; automotive pilots ran to single or low-double-digit counts. Set against AGIBOT's 20,000 built and Tesla's hundreds per week, this is the sharpest number of the week: **units produced, units sold, and units working are three different quantities and only the first is being reported by vendors.** Sources conflict here in kind, not just in magnitude, and IFR itself warns the definitions don't match.
- **Unitree Dex5-S hand, from $6,500** (39,900 yuan). 22 DoF, human-hand size, all joints backdrivable with impact-torque protection. Weight, grip force, tactile configuration, interfaces and delivery are all unverified — and 22 DoF does not establish 22 independent actuators. Still the cheapest human-scale multi-finger hand a small lab can quote this year.
- **Agility confirms it is evaluating wheeled form factors**, following the wheeled concept in the Digit 5 launch material (W38). That is the fourth wheeled-bimanual signal in five weeks after Nori, Nucleus and the Digit 5 concept.
- **Robros adds full-body teleoperation to IGRIS-C** — 18 sales of 36 units produced, which is at least a disclosed ratio.

Video-only, no method: Figure 04 founder claims, Lumos NexCore "skills in days", Spirit AI's 2027 prediction, XPENG IRON showroom memory demo. Seen and discarded.

---

4 Open source & tooling

Ordered by friction removed this month.

1. **Intrinsic Core** (Apache-2.0, announced at ROSCon Toronto 09-22; coverage). Alphabet's Intrinsic released the production layer under its industrial arm business: a hardware-agnostic real-time control framework, motion planning with collision avoidance, grasp planning, pose estimation built on NVIDIA FoundationPose, simulation and calibration services, and Intrinsic-ROS drivers. Demonstrated on Universal Robots and FANUC hardware via their Open Machine Tending Solution. This is the largest Apache-2.0 drop of vendor-grade manipulation infrastructure since MoveIt itself. Caveats: industrial-arm-shaped — real-time control and planning, not learned policy. And the repo URL 404s as of 09-27, so the contents and the `20260922.0` tag are from the announcement, not read off the tree. Check before you plan around it.
2. **MuJoCo 3.14.0** (2026-09-22). Two changes matter. An experimental `ipc` flag on the discrete integrator gives **penetration-free flex contact** via an incremental-potential formulation with continuous collision detection — the first time MuJoCo has offered a non-penetrating deformable contact path. And procedural model editing is no longer quadratic: building a spec with 8,000 geoms is roughly 10× faster, because name scanning and signature computation became incremental. If you generate scenes programmatically, that second one changes your data-generation wall clock today. Also: archive resource providers, `<frame>` elements now round-trip with position and quaternion intact, a breaking `mj_readCtrl` change for multi-input actuators, and weld-constraint torque fixes.

3. **SpectRobot** (Apache-2.0 code; data in LeRobotDataset v3.0). See §6. If you have an SO-101 and LeRobot already, this adds a tactile modality for roughly €100 of parts and no architecture change.
4. **REBOOT** (PREPRINT, Calgary + Mila; data on HF, CC-BY-4.0, LeRobot v3). 2,160 real bimanual WidowX AI episodes over 18 NIST Task Board #1 connector install/remove tasks, **half recovery-from-failure**. What makes it useful: every episode carries five annotated phase boundaries (Align-pick → Engage-pick → Transport → Align-place → Engage-place), and recovery episodes additionally carry the originating failure phase and a categorical mode (misalignment, slip, premature release, jamming, off-axis collision). The recovery episodes are not hypothesized — they roll out ACT, Diffusion Policy and π0-FAST, log where each actually fails, and allocate recovery demos in proportion. Per-connector clearances are published, from RCA interference fit at 0 mm to HAN 10E at 2.73 mm. **Caveats the authors state:** one annotator, right-arm bias (83 of 2,160 episodes are genuinely bimanual), fixed-duration episodes with trailing static segments, depth recorded but unused in every baseline. 24 h per task on one A100.
5. **Isaac ROS 5.0** (NVIDIA, ROSCon). The one item here that reduces friction rather than adding a feature: a FoundationStereo fine-tuning skill that adapts stereo depth to *your* cameras and scene. FoundationPose inference claimed 5.5× faster; ROS Lyrical and Ubuntu 24.04; Orin Nano through Thor; pick-and-place as a standalone agent-ready skill. Vendor claim, no independent benchmark, no GitHub tag visible yet.
6. **EmbodiedSWE** (PREPRINT, ByteDance Seed; code released). A coding agent writes solution programs against full privileged simulator state for 28 Isaac Lab tasks (contact-rich, deformables, liquids, Franka/bimanual/G1), then diversifies them into VLA training data. Agent solve rate spans 82% → 11% by difficulty; PPO scores 0.00–0.06 on the same tasks; a VLA goes 18% at 10 demos to 69% at 400. One 4090 for ~4 h per run. **Soft:** authors' own benchmark, privileged state, and a reported **43% hack rate** on two models — the agent satisfies the scored predicate rather than the task nearly half the time.
7. **isaac_asimov** (BSD-3). An Isaac Lab PPO+AMP humanoid locomotion recipe that actually discloses per-joint torque, stiffness, damping, friction and actuator delay, plus the full reward and penalty config and reference motions — the parameters most locomotion papers omit. Tested on 3090 and 4090, single and multi-GPU. The claimed checkpoint is absent from the repo and there are no eval numbers.
8. Smaller: Robotiq published official SimReady USD packages for Isaac Sim under CC-BY-4.0 — a gripper vendor shipping its own sim assets is a first in this ledger. ABC-VLA on Jetson Thor ships prebuilt TensorRT engines with an HTTP `/act` endpoint and a no-robot smoke test (25 GB, FP16/BF16 only, gemma license, self-reported benchmarks). Isaac ROS cuMotion on the ROBOTIS AI Worker is a worked dual-arm-plus-lift MoveIt 2 integration.

No release in window from LeRobot, Isaac Lab (v3.0.0-EA, GA targeted end of October), ManiSkill, RoboCasa or Genesis. ROS 2 had no package sync; the 09-18 Lyrical sync is still the latest.

5 Money & moves

Qualcomm to acquire PickNik Robotics (trade press, announced 09-23, terms undisclosed, customary closing conditions). PickNik maintains MoveIt 2 and sells MoveIt Pro; Qualcomm commits to keeping MoveIt under existing licensing and community roadmaps, and plans to integrate it with Dragonwing and with Arduino's VENTUNO Q boards. **INFERENCE** \2192 a silicon vendor buying the motion-planning layer is the mirror image of NVIDIA buying Hugging Face (W36). Both bets are that the defensible position is owning the software layer developers already use and pulling it toward your silicon. As with LeRobot, an open-source *commitment* was stated and **no governance structure was named** — no foundation, no independent steering body. That is the same gap flagged in W36 and it is now load-bearing for two of the three pieces of the open manipulation stack.

Intrinsic open-sources Intrinsic Core (company announcement via trade press). **INFERENCE** \2192 Alphabet is giving away the plumbing one day before a competitor buys the incumbent plumbing vendor. Read as strategy, this prices control software at zero and moves the contested layer up to policies and down to silicon. Read as a gift, it is the single largest reduction in build friction for industrial manipulation this year. Both readings are available; the timing relative to Qualcomm/PickNik is not obviously coincidental.

Tesla Optimus ramp (The Information, single source, unconfirmed). **INFERENCE** \2192 the binding constraints reported are the *hand* — >100 screws, manual assembly, sensor-glove durability — not the policy. That is the same message as Apptronik's

CEO warning that US actuator and component supply is the limiter, and it is where a mechanically literate EE is scarce.

Feather launches a $30,000 wheeled humanoid for developers with $7.6M; **Vesoma exits stealth** with a learning-first humanoid, no method or figure; **O-ID raises $1.2M** for modular swappable humanoids. **INFERENCE** \2192 the venture side was near-silent — no significant round closed on three of the week's five business days. What money there was went to developer-priced hardware and to form factors, not to policy teams. After four weeks of $80M–$400M rounds (W37–W38), a single quiet week is noise, not a trend; note it and watch.

South Korea plans a 2027 robot-AI push starting with logistics — robot foundation models, world models and a national physical-data library, with postal and logistics trials. **INFERENCE** \2192 the deliverable named is a *data library*. Three governments and a dozen vendors have now converged on the view that the scarce asset is embodied data, not model architecture. That is also the bet behind W38's Shutu and Noetive rounds.

Analysts see a PEEK boom in humanoids; Kingfa's robotics lead calls it hype. **INFERENCE** \2192 a materials supplier publicly contradicting a Barclays demand thesis is a rare useful disconfirmation — the supply chain is being priced on projected humanoid volumes that the people selling into it do not believe. Set beside IFR's 7,000 (§3).

6 Deep read of the week

"Learning tactile perception from high-bandwidth single-point sensing" (SpectRobot)
— arXiv:2609.24621v2 (2026-09-23), preprint, Joseph Rigal and Emmanuel Virot (Wormsensing) with Caroline Pascal (Hugging Face). Code, Apache-2.0; data, LeRobotDataset v3.0. **Declared conflict of interest: Wormsensing manufactures the Dragonfly sensors evaluated and employs two of the three authors.**

Why this one over the TRI negative result or GAM: it is the only paper this week that a reader with one arm and one GPU can execute end to end for about €100 of new parts, and its central claim contradicts the direction the tactile field has been moving in for three years.

Setup.

- Task: a four-class sorting task under deliberate visual occlusion. An opaque box contains one of empty / 1 plastic spacer / 7 plastic spacers / 7 metallic nuts. The robot picks it up, shakes it vertically, and must place it in the matching bin. The classes are chosen so the policy must separate *count* within a material and *material* at fixed count.
- Embodiment: a single SO-101 (WowRobo, pre-calibrated), leader-follower teleoperation for collection, ACT via LeRobot 0.5.2, ResNet-18 trained from scratch alongside the transformer.
- Data: 40 demonstration episodes per dataset, constant artificial lighting.
- Held out: nothing in the train/test sense — evaluation is 80 autonomous rollouts per condition (20 per box class) on the physical robot. Grasp failures are counted in the denominator.
- Compute: one consumer workstation, i5 + 32 GB + **RTX 5070**, 100,000 optimization steps, ~3 hours per run.
- Sensing: seven technologies across three physical modalities — MEMS accelerometer (ADXL356CZ), IEPE accelerometer (PCB TLD352A56), IEPE load cell (PCB 208C02), metallic strain gauge in full bridge (HBK), IEPE and passive Dragonfly strain sensors, and a bare 15 mm PZT disk. All mounted on the *upper* gripper, mechanically coupled to the contact points but away from abrasion.

Method.

- Inherited: ACT, LeRobot, the whole imitation pipeline. Nothing about the policy changes.
- New: the representation. A high-rate scalar channel (up to 200 kS/s, 24-bit, ~6,600 tactile samples per camera frame) is turned into a 224×224 **grayscale spectrogram** — power spectral density, Tukey window, 50% overlap, nFFT 64–4096 — regenerated every 33 ms and handed to the vision encoder in the same slot a camera would occupy. Raw time-domain signal is never retained.
- Two design choices are argued rather than inherited. **Grayscale, not RGB**, because a scalar signal has no reason to be mapped to three channels. **Linear frequency axis, not Mel**, because Mel warping encodes human auditory perception and

mechanical harmonics have no such prior.

- The consequence that matters architecturally: acquisition bandwidth and policy rate are decoupled. You can raise the sample rate arbitrarily without changing the size or rate of the policy input, so training and inference cost are unchanged.

Results.

- **Vision-only scores 23%** against a 25% chance floor. Every tactile condition beats it: MEMS accelerometer **82%**, IEPE Dragonfly 77/71/80 across three gripper designs, PZT disk 73%, IEPE accelerometer 71%, passive Dragonfly 67%, metallic strain gauge 58%, IEPE load cell 48%.
- The number that matters is not any of those. It is the **bandwidth-versus-context comparison**: at 0.36 s of temporal history, 0–10 kHz gives ~31%; at 0.29 s, 0–100 kHz gives ~32%. Both are at chance. Extend to 2.9 s and 2.3 s respectively and they go to **~86% and ~92%**. Ten times the bandwidth at matched duration is not resolvable; eight times the history is worth 55 points.
- The low-cost chain is the second result. Swap the Dewesoft IOLITE for a ZONRI IEPE converter, an ADS8688 16-bit SAR ADC and a Teensy 4.1 — **~€100 total** — and RMS noise rises about an order of magnitude, but the MEMS accelerometer, IEPE Dragonfly and PZT disk all land at **~90%** with overlapping Wilson intervals on a separately collected comparable dataset.
- A blank-spectrogram substitution at inference did not merely degrade classification — it **prevented the policy from completing the grasp**, which says the tactile channel is entangled with motor control, not an additive classifier input.

Where it's soft.

- **No comparison against a spatial tactile array.** The paper's framing claim — single-point bandwidth as an alternative or complement to spatial density — is argued from first principles and from seven single-point sensors beating vision-only. No GelSight, AnySkin, e-skin or pressure pad is run on this task. The central claim is therefore unmeasured.
- **Conflict of interest, disclosed and structural.** The Dragonfly is the common reference channel in all three grippers. The mitigating evidence is real and the authors surface it: the cheapest device in the table (a <€100 MEMS accelerometer) scores highest, and they state the top rankings are not statistically resolved at n=80.
- **Bandwidth and duration are confounded by construction**, and the paper says so: nFFT selection ties them, so "short vs long window" compares configurations, not isolated factors.
- **Dataset-level variance is roughly the size of the effects being ranked.** The same Dragonfly sensor gives 77, 71 and 80 across three independently recorded datasets — a ~9-point spread comparable to most between-sensor gaps, and the confusion matrices differ even where the totals agree.
- One task, one embodiment, four classes, no seed count, one trained policy per condition.
- The entanglement observation comes from a **single ablated model** and is explicitly called preliminary.
- No cherry-picking signal: Wilson intervals throughout, the failure case (load cell at 48%) is reported at the same prominence as the wins, and the ablation that undercuts the tidy story is in the main text.

Steal list.

1. **Spectrogram-as-image for any scalar high-rate channel you already have.** Motor current, joint torque, a $5 contact microphone. Fixed-size input regardless of sample rate, drops into an existing image slot, zero added inference cost, no architecture change. Use a linear frequency axis and grayscale.
2. **Budget temporal context before bandwidth.** 0.36 s → 2.9 s was worth 55 points; 10 kHz → 100 kHz at matched duration was worth nothing resolvable. If you are choosing nFFT, choose it for history length.
3. **Mount the sensor off the contact surface.** Mechanically coupled, not abraded. This is the maintenance argument for vibration sensing over skins, and it costs nothing to adopt.
4. **Use the blank-input substitution as a diagnostic, not just an ablation.** If substituting a blank tactile frame breaks the *grasp* rather than the *decision*, your channel is entangled with control and a classification-only ablation will mislead you. Compare with ME-Dex's zero-tactile result in §2, where zeroing the input changed nothing — same probe, opposite answer, and the difference is informative.

5. **Wilson score intervals on n=80**, and report them even when they overlap and spoil your ranking.

Cost to reproduce a scaled-down version.

- Hardware, assuming you already have an SO-101 and a workstation: a PZT disk or a MEMS accelerometer (both under €100 in the paper's cost tiers), plus the ADS8688 + Teensy 4.1 acquisition chain at ~€100 total. The ADS8688's 15 kHz anti-alias cutoff caps you at the 0–10 kHz configuration, which is the one the paper recommends anyway. Skip the IEPE converter unless you buy an IEPE sensor.
- Data: 40 teleop episodes per condition. Two conditions (vision-only, one tactile sensor) is ~80 demos, call it 2–3 hours of operator time. Nothing to download.
- Training: ~3 h per 100k-step ACT run on an RTX 5070. Two conditions × 3 seeds = 6 runs ≈ **18 GPU-hours**. The paper ran one seed; running three is the cheapest improvement available and would let you report a real band.
- Evaluation: 80 physical rollouts per condition. This is the real cost — wall-clock robot time and an operator to reset the boxes, not GPU time. Budget a day per condition.
- Total for a credible minimal replication: ~€100 of parts, ~1 day of teleop, ~18 GPU-hours, ~2–3 days of rollouts.
- **The one-arm extension actually worth doing** is not a replication. Run the same seven-sensor protocol on a task where the information is a millisecond-scale transient — a peg insertion or a slip arrest — rather than a 2.9-second active shake. See §10.

7 Relevant to what you're building

`CONTEXT.md`'s `## Active build` section is still an unfilled placeholder, so there are no workstreams to map this week's items onto. Section skipped.

8 Market signal

Visibly hiring

primary job boards, checked 2026-09-27

- **Figure**: 97 open roles, 13 on the AI–Helix team, all San Jose, five days in office. Titles: Modeling, Pretraining, Video Pretraining, Reinforcement Learning, Robot Learning, Perception, Generative AI, Data Infrastructure, Training Performance, Localization and Mapping, Android, Backend, XR. Down from 104 total a week ago; the Helix count is unchanged.
- **OpenAI**: 23 robotics roles, against 26 read on 2026-09-20. Composition unchanged and still hardware-weighted: actuator design, actuator gear design, motor manufacturing, dynamometer and actuator test infrastructure, thermal simulation, PCB layout, firmware, prototyping lab technician, robotics data-acquisition program management. **Zero titles containing policy, VLA, imitation or robot learning.**
- **Physical Intelligence**: 13 roles — manufacturing, NPI program management, robot operator, robot build technician, forward-deployed robotics engineer, build & release, controls, mechanical design, ML infra for data systems. Also no policy-research titles.

Treat the counts as soft: W38 recorded 21, 26 and 27 for OpenAI from three sources on the same day. A ±3 swing is rendering noise, not a hiring signal.

Comp datapoints, primary postings, base plus equity:

- OpenAI, ML Engineer, Distributed Data Systems – Robotics (SF, hybrid 3 days): **$380K–$500K + equity**, unchanged week over week.
- Figure, Helix AI Engineer, Modeling (San Jose, 5 days in office): **$200,000–$400,000 base**.
- Physical Intelligence publishes one rate: Robot Operator at **$25/hour + benefits**. Every engineering role has compensation withheld.

INFERENCE \2192 the only lab hiring model people in volume is also the only one publishing a band for them, and that band tops out **$100K below** OpenAI's band for the person who builds the data pipeline. Across all three, the roles with published numbers and the roles being hired in bulk are the systems *around* the policy — actuators, test rigs, data acquisition, distributed storage, forward deployment. The $25/hour robot operator and the $380–500K data-systems engineer are the same bet from two ends: the scarce input is episodes, and the premium goes to whoever can move them at scale.

Skills that recurred in work that got attention this week:

- **Making the training-time/run-time split explicit.** ME-Dex's zero-tactile-input result, HapticWAM, ZeroTouch, VT-Bridge, and What Matters' staged-objective schedule are all the same move: buy the benefit during training, pay nothing at deployment. A candidate who can say which of their design choices is free at inference is answering a question interviewers are now asking.
- **Grounding and object-centric filtering over bigger backbones.** GAM beats π0.5 by 10 points on RoboTwin single-task by *deleting* most of the observation, and its ablation shows the filtering is worth more than the backbone (46.8 → 10.3 without it). The failure mode it fixes — policies replaying a memorized layout instead of following the instruction — is now documented well enough to discuss concretely.
- **Statistical literacy as a differentiator, not a formality.** Gaussian Is Enough used Welch and STEP tests with Bonferroni-corrected compact letter displays across 100K rollouts. SpectRobot reported Wilson intervals that spoiled its own ranking. Res-HIL ran three seeds × 50 fixed initial conditions on real hardware. These are still the exception.
- **Contact estimation without contact hardware.** Object SDF plus proprioception, motor-current residuals, torque-observation alignment, high-bandwidth vibration off the contact surface. This is the EE-shaped skill and it is now appearing in VLA and WAM papers rather than in controls papers.

Table stakes, updated. Released code at submission — this week the majority of the top items actually cleared it (SpectRobot, PolyUMI, REBOOT, EmbodiedSWE, ME-Dex, Intrinsic Core, GALA). Latency beside success. An ID/OOD split. And, since W38, paired evaluation with more than one training seed.

Still differentiating:

Multi-seed real-hardware evaluation (Res-HIL); inferential statistics over a rollout population (Gaussian Is Enough); releasing fabrication files and firmware, not just code and weights (PolyUMI); and a controlled intervention that tests *whether the mechanism you claim is the mechanism operating* (What Matters' temporal-reversal probe, ME-Dex's zeroed input, SpectRobot's blank spectrogram, GAM's observation ablations). Four papers this week did the last one. That is the fastest-moving norm in the field right now.

9 Threads

THREAD	STATUS	THIS WEEK	NEXT REAL UPDATE
Does VLA scale buy language-conditioned control?	**advancing** (first move in 5 weeks)	GAM: under LIBERO-PRO's *Task* perturbation every baseline scores ≤0.11 on all four suites (π0.5: 0.01 on Spatial-Task); GAM gets 0.88. The fix was frozen metric 3D grounding plus object-centric masking, not a language encoder or more scale	Independent reproduction of that Task column; and still, a sub-10M real-hardware multi-task policy against a published baseline (five weeks unchanged)

THREAD	STATUS	THIS WEEK	NEXT REAL UPDATE
Benchmark validity in physical AI	**ADVANCING**	REBOOT scores per-phase completion with failure rates measured from real ACT/DP/π0-FAST rollouts (Align-place + Engage-place = 54% of install failures vs 8% for Transport). RoboRecover: UnifoLM 98.8% initial → 48.0% recovery. Gaussian Is Enough runs Welch + STEP with Bonferroni CLD over ~100K rollouts. What Matters is controlled by design but reports no seed bands	A major-lab table with seed bands *and* paired evaluation — What Matters has the controls, not the bands. Or verifier exploitability scored under search
Contact sensing without dedicated tactile hardware	**ADVANCING**	Strongest week yet. ME-Dex zeroed at **INFERENCE** `\2192` 91.70/91.74 vs 91.56/91.92 — but its tactile is simulator-generated. SpectRobot: 82–92% vs 23% vision-only on a €100 chain. Whole-Hand Force Regulation: object SDF + proprioception. CoPRE: motor current only. Against: PolyUMI's entire case is that the physical finger must travel with the data	An estimated- or predicted-force policy matching a real-sensor baseline on a task whose labels were **not** collected with that sensor. Unmet
Paying at training time, not run time	**ADVANCING**	ME-Dex is the cleanest instance yet: keep tactile supervision and future-tactile prediction, discard the sensor reading, score unchanged. HapticWAM, ZeroTouch, VT-Bridge push the same way (abstract-only). What Matters adds a scheduling version: BC+VG for 80%, dynamics last 20% → 83.15 vs 79.11 naive joint	Whether ME-Dex's zero-tactile result holds with real sensor data. Still open: a matched one-step-head comparison on one backbone
Online correction as the last mile	**ADVANCING**	Res-HIL: 3 seeds × 50 fixed inits on a real UR5e, spread under 4 pp — the first multi-seed real HIL result here. At 10 min, 100/64/50/66/92 vs HIL-SERL 90/30/10/0/0, with the whole delta in two cable tasks where HIL-SERL scores 0%. No code	A third-party rerun of EXPO-FT (code two weeks out, nobody has), or Res-HIL code
World-action models as unified policy + simulator	**ADVANCING**	The mechanism question got an answer: corrupting future-latent content moves actions <1%; reversing their temporal order moves actions 12.8–14.4% and cuts OOD success 24–32%. Video generation is the only auxiliary objective that helps OOD (77.96 → 81.22); all of them hurt ID. ME-Dex adds tactile to the target set	A released world model used for policy improvement by someone who did not build it. OpenWAM repos unchanged since 09-09
Who owns the open robot-learning stack	**advancing** (was contradicted)	Two events in 48 hours. Intrinsic Core released under Apache-2.0 (real-time control, motion and grasp planning, calibration, FoundationPose, ROS 2 drivers). Qualcomm agreed to buy MoveIt's maintainer, pledging existing licensing and community roadmaps with **no governance structure named** — same gap as NVIDIA/Hugging Face. LeRobot still has not shipped anything since the NVIDIA deal closed, four weeks on	Either commitment written into a foundation or steering body rather than a press quote; the first post-deal LeRobot release

THREAD	STATUS	THIS WEEK	NEXT REAL UPDATE
Simulation fidelity for transmission and contact	**advancing** (was stalled)	MuJoCo 3.14.0: experimental `ipc` flag on the discrete integrator gives penetration-free flex contact via incremental potential with CCD — the first non-penetrating deformable path in MuJoCo. Procedural editing ~10× faster at 8,000 geoms. Isaac Lab unchanged; MuJoCable still has no repo	A sim-to-real result crediting the IPC flex contact or the discrete integrator; a MuJoCable repo
Low-cost open hardware as training substrate	**ADVANCING**	SpectRobot: €100 chain on an SO-101, Apache-2.0 code and LeRobot v3.0 data. PolyUMI: $236 transferable sensing finger, fully released down to fabrication and firmware. Cartesian Hand: 7-DoF all-prismatic in-hand manipulator at ~$500, 350/350 trials — files not out. Unitree Dex5-S, 22 DoF from $6,500. Still no third-party Aero Hand, BRIDGE or Peg-in-Bench build	An independent group reporting numbers on released low-cost hardware. Cartesian Hand files landing would give this thread an in-hand manipulator at hobby cost
Claimed open releases vs actual artifacts	**narrowing** (was contested/worse)	Genuinely better. Out: SpectRobot, PolyUMI (hardware, electronics, firmware, fabrication, software), REBOOT, EmbodiedSWE, Intrinsic Core, ME-Dex, GALA. Missing: Gaussian Is Enough, GAM, What Matters, Whole-Body UMI (no code statement); Res-HIL (videos only); Cartesian Hand ("will"); isaac_asimov (checkpoint claimed, absent). **StreamPI official weights now five weeks pending**	Whether the narrowing holds a second week; StreamPI weights
Humanoid capital vs demonstrated capability	**ADVANCING**	The sharpest production-vs-deployment contradiction yet: AGIBOT 20,000 cumulative built (self-reported), Tesla several hundred/week (one outlet, unnamed sources) — against IFR's third-party count of ~7,000 humanoids *sold worldwide in 2025*, many to research institutions and data collectors. Tesla's named bottlenecks are hands and touch sensors, not policy. Whole-Body UMI is the only real-number humanoid manipulation result this week, and it is academic	A humanoid company first-authoring a method with released code; an independent count of humanoids in productive operation
Manipulation acquired rather than built	**ADVANCING** (adjacent)	Qualcomm/PickNik: a silicon vendor buying the manipulation *software* layer. Third consecutive week where the acquirer is a capital or platform holder, not a logistics incumbent (SoftBank/RAI W38, NVIDIA/Hugging Face W36)	A logistics incumbent buying an end-effector or policy team. The W35 in-quarter prediction has ~4 weeks left; kill on lapse

10 Open question

Does high-bandwidth single-point tactile sensing survive on transient contact, or only on active interrogation?

SpectRobot's task is an active information-gathering motion: the robot shakes a box for 2.9 seconds and reads the spectrum. That is close to the best possible case for a spectrogram — a long, energetic, repeating excitation, where averaging over 224 time columns is a feature. The result that carries the paper is precisely this: temporal context is worth 55 points and bandwidth above 10 kHz is worth nothing resolvable.

But the contact events that motivate tactile sensing in manipulation are the opposite shape. A slip onset, a peg catching a chamfer, a connector seating — these are single millisecond-scale transients inside an otherwise quiet window. At 2.9 s over 224 columns

you get ~13 ms per column, so the event occupies one column and is averaged against nothing. The same paper's short-window configuration, which is the one you would need for a transient, scored at chance (31–32%). It is entirely possible that the bandwidth-versus-context tradeoff *inverts* for transients — that the reason 100 kHz bought nothing here is that this task has no high-frequency information worth having, and that a task which does would show the opposite ordering. The authors say as much in one sentence and do not test it.

What would settle it:
The same seven-sensor benchmark, same rig, same €100 acquisition chain, on a peg-in-hole insertion or a slip-arrest task — one where success depends on detecting an event within tens of milliseconds. Run both the short-window (0.29–0.36 s) and long-window configurations at 10 kHz and 100 kHz, with the same Wilson intervals. And run a spatial tactile array on the same task, which is the comparison the paper never makes and which its framing claim requires. That is about a week of one arm, one GPU and one operator, and it would either generalize the result or bound it to interrogation tasks. Given that the code, the data format and the sensor list are all public, this is the rare open question a single reader could close.

Source types are labelled inline. All arXiv items are unreviewed preprints. Every paper result above is single-lab and unreproduced unless stated; most real-robot numbers are one training run at n=10–80 per cell. Release status was checked against repos, model cards and PyPI rather than paper claims; the one exception is Intrinsic Core, whose repo 404s and whose contents are taken from the announcement. Trade-press items rest on company statements unless marked otherwise; the Tesla Optimus figures rest on one outlet's unnamed sources.

ALSO THIS WEEK

Scored high, didn't earn a section:

- Representation World Model (Tsinghua, project page only) — latent interpolation plus inverse dynamics replaces the forward model and search entirely; 0.03B parameters, LIBERO-Goal 93.0%, trained on 8×RTX 5090.
- RoboRecover (RUC, release unverified) — replays action prefixes to reconstruct off-nominal states across 2,000 RoboTwin/LIBERO scenarios; UnifoLM's 98.8% initial success becomes 48.0% recovery. Sim-only.
- AD-WM (Nanjing, code on project page) — an action-recovery regularizer forces latent dynamics to discriminate between candidate actions for counterfactual MPC; hard-start tasks 3.7% → 52.0%, and the auxiliary heads are discarded at deployment. Built on V-JEPA2 with zero-shot transfer to DROID.
- MachEmbodied-U0 (Li Auto) — mentioned in §1; LIBERO 99.0, LIBERO-Plus 82.5, 4,200 h and 368 accelerators of pretraining.