

No Common Crawl for Robots

The Data Problem
in Physical AI

October 2026

AUTHORS:



Steve Sarracino

Founder & General Partner



Ross Nayler

Analyst | Research

✉ ross@activantcapital.com

About Activant

Activant is a research-led global investment firm that partners with high-growth companies. Since 2015, we have invested in category-defining businesses during their most critical phases of growth, partnering with founders who have won their initial battles and are ready for the next challenge.

Our approach pairs deep, proprietary research with patient, flexible capital and hands-on operational partnership. We work alongside founders and leadership teams to refine strategy, strengthen operations, and accelerate sustainable growth.

Activant Research is dedicated to uncovering the most exciting emerging technologies, sectors, and companies we believe will shape the future. Our research-driven perspective informs everything we do, helping us invest at meaningful inflection points and support founders in building enduring, category-leading businesses.

You can find out more about Activant and our research at <https://activantcapital.com/>.

Physical AI needs data that doesn't exist. The internet gifted language models a near-complete record of human text for free, but it holds almost no record of what it feels like to grip a wrench or catch a falling glass.

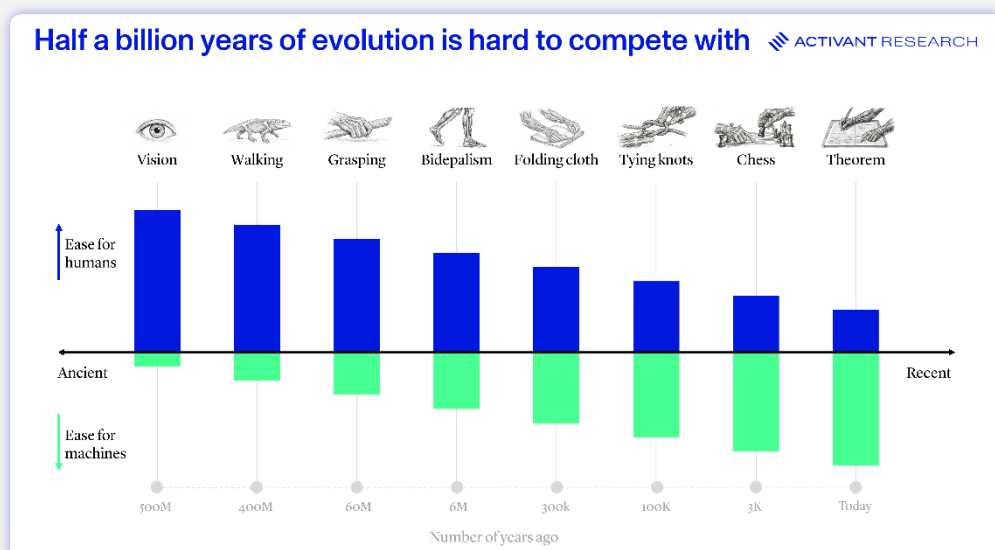
In [a previous robotics piece](#), we discussed how the same scaling laws that powered the LLM revolution appear to hold in physical AI. But the scaling story in the physical world is an inversion of the one in language. For LLMs, data was free and abundant while compute was scarce, so the industry spent nearly a trillion dollars building data centers.¹ Physical AI has the compute, but it does not have the volumes of data required to ride the scaling curve the same way LLMs did: there is no [Common Crawl](#) of robot experience and every token needs to be manufactured, on the right hardware, one demonstration at a time.

This changes who gets paid. When a scarce resource is discovered, you back whoever owns the deposit; but when it is manufactured, owning a deposit means very little. Supply is a function of how cheaply anyone can produce the next unit, and in physical AI, the cost of the next unit is falling from four directions at once. ***We think value accrues to whoever owns the process that keeps manufacturing data: collect, train, deploy, capture, retrain; not to whoever sells it.***

In this piece, we explore what it actually costs to manufacture that data, why the obvious picks-and-shovels trade is the wrong end of the trade, and where the durable returns land. But the place to start is upstream of all three: what a robotics lab inherits from the internet, and what it must build itself.

Building The Brain

In the 1980s, Hans Moravec observed a paradox: the things humans find intuitive – **motor control, perception, and dexterity** – are the exact things that machines find the hardest.² As it turns out, half a billion years of natural selection refined the sensorimotor mechanisms that let us pour our morning coffee without spilling it; abstract reasoning arrived much later. AI inherits this asymmetry in reverse. Reasoning came first, because the internet contains petabytes of it in text; sensorimotor skills have to be built from scratch, because the internet contains almost no record of how we physically interact with the world, even in something as simple as pouring our morning coffee.



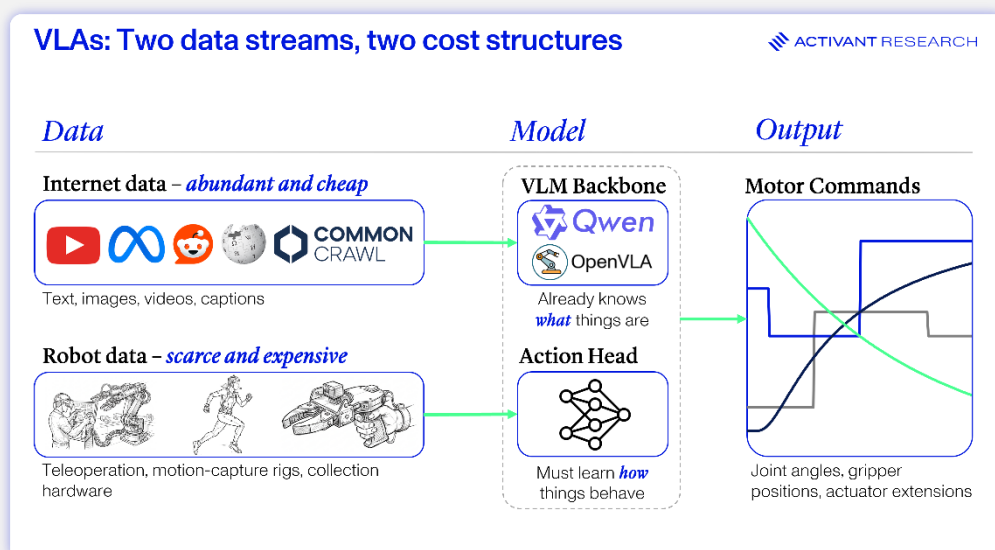
Ex-Meta Chief AI Scientist and founder of [AMI Labs](#), Yann LeCun frames the data gap as follows: a four-year-old child has absorbed more data through vision alone than the entire text corpus used to train a frontier LLM.³ Yet almost all of that physical experience remains unrecorded.

The recent evolution of robotic foundation model (RFM) architectures is really the story of how frontier labs are attacking the harder half of Moravec's paradox. It is easiest to define this evolution with a question: [how much of the model do you inherit from the internet, and how much must you manufacture yourself?](#)

Inheriting Semantics: Vision-Language-Action Models

From 2023 to 2025, the industry standard RFM architecture was the [VLA model](#). It imported a pretrained vision-language model (VLM) that already understood the world semantically, such as what an air fryer looks like or what "load the dishwasher" means, and bolted on an action head trained on robot demonstrations. [Google's](#) RT-2 and PaLM-E were early examples; [Physical Intelligence's](#) π -range carries the pattern forward, as do [Nvidia's](#) GR00T, Google DeepMind's Gemini Robotics, [Figure's](#) Helix, and [Genesis AI's](#) GENE-26.5.^{4,5,6,7,8,9}

The appeal of the VLA is that [scale can be imported for free](#). The VLM half is grounded in internet-scale text, images, videos, and captions: abundant, cheap, and already paid for by someone else. The cost falls on the fine-tuning, where the action head has to learn from real robot data collected via teleoperation or humans in motion-capture gloves and VR rigs. VLAs need two data streams, one essentially free and one essentially priceless.

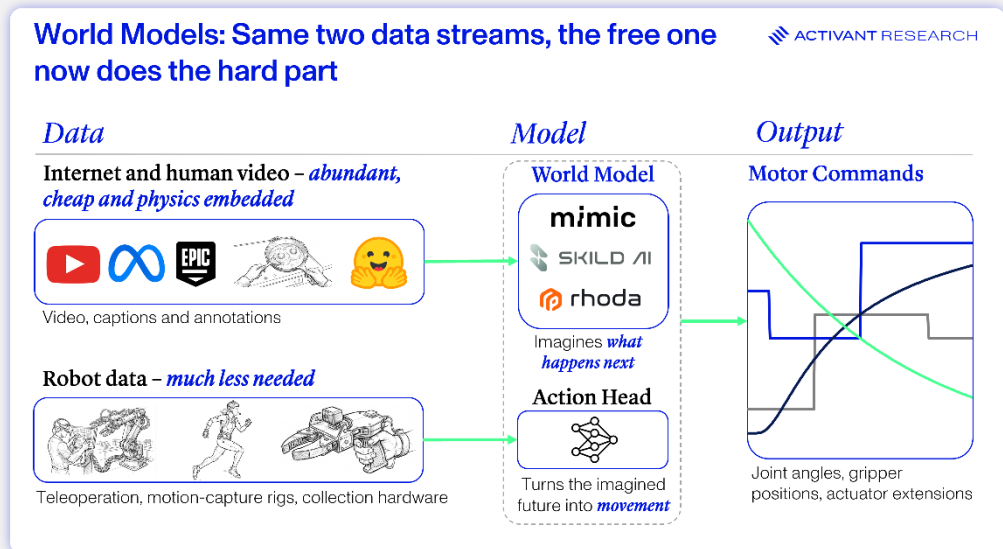


As Nvidia's Director of Robotics Dr. Jim Fan puts it, the problem with this division is that the free half is the [wrong shape](#). VLMs do see physical events in their training data, but their training objective is not to model physics. They are trained to hill-climb static benchmarks like visual question answering and captioning, which have nothing to do with how plastics deform, how a t-shirt folds, or how friction behaves on a stainless-steel countertop. The semantic knowledge is powerful but encyclopedic: models can describe a scene accurately but still struggle to act in it. As a result, much of the physical understanding still has to be learned from robot demonstrations in post-training, which means scaling teleoperation into the hundreds of millions of hours range. This is the data acquisition bottleneck, and the reason the recipe started to change in 2025.¹⁰

Inheriting Dynamics: World Models and Beyond

The first step away from the VLA is to swap the VLM backbone for a generative video model, a [world model](#), trained on a similar scale of internet video. The rationale is that [internet video is cheap, abundant, and implicitly encodes physical dynamics](#): how objects fall, deform, occlude, react to force, get picked up and put down. Instead of mapping observations and instructions straight to actions, the model learns to imagine the future and ground those imaginings in motor commands.

The major distinction isn't necessarily the diet; both world models and VLMs consume video data. [The difference is the training objective](#): a VLM is trained to describe what is going on in a frame, a world model is trained to predict the next one. Captioning a glass on a countertop requires no understanding of the laws of gravity, friction, or momentum; but predicting where it lands after a knock requires a deep understanding of all three. Nobody explicitly annotates physics into a world model; some of that physical understanding emerges through prediction, and because every frame is its own label, [the free half of the data stream starts doing the work the expensive half used to pay for](#).



[Mimic's](#) mimic-video, [Rhoda's](#) Direct Video-Action (DVA), [1X's](#) 1XWM and [Skild's](#) Skild Brain are all building here with the goal of learning physical dynamics from human video while being less dependent on robot data.^{11,12,13,14}

Some labs push this logic even further. [Generalist](#) refuses both the VLA and world model labels; its GEN-1 model is almost entirely trained from scratch on the world's largest proprietary physical interaction dataset.¹⁵ Its bet is that the right pre-training backbone is the one you deliberately manufacture rather than inherit.

The architecture a lab picks is really a bet on how much expensive robot data it can avoid. **That bet is the whole investment question**, because it decides whether the scarce input stays scarce.

Different Models, Different Diets

The architecture bet heavily influences the types of data labs use to train their models. Robot data remains the most sought-after ingredient, but no lab uses it on its own:

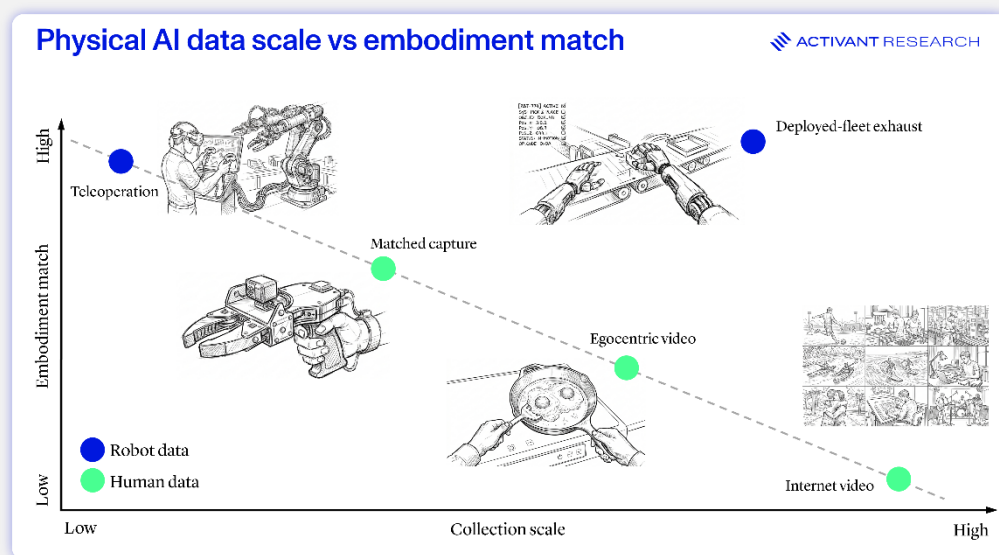
Physical AI Data Modalities ACTIVANT RESEARCH

Source	What it gives you	What it costs	Who leans on it
Deployed-fleet exhaust	Embodiment-native, self-renewing , generated by the target hardware doing the target job	Effectively free once a fleet exists, but requires a fleet	The few companies already operating in the wild
Teleoperation	Fully labeled motor commands on the exact robot	Expensive : scales linearly with headcount	Manipulation-first labs
Matched capture	1:1 mapping from a gripper or glove onto the robot's end-effector	Hardware plus logistics per collector	Labs with human-shaped hands
Egocentric video	Task and scene coverage at human scale	Cheap and scalable, but no motor commands	Humanoid labs
Internet video	Semantic and dynamic priors	Effectively free and infinite	Everyone , as a backbone

Click [here](#) to view live examples on our website

A lab's data mix depends on what it is building. Manipulation-first labs like Physical Intelligence, Generalist, and Genesis lean harder on teleoperation and matched capture. Humanoid labs like Figure and 1X teaching robots to move through the home need large volumes of egocentric and internet video for semantic and task coverage alongside teleoperation and matched capture for manipulation, balance, and locomotion. A newer twist crowdsources teleoperation itself: [Enigma](#) put 100+ of its own robots online for anyone in the world to control, turning public curiosity into native action data.¹⁶

Every lab hits the same data wall. **The data that scales is human data, and humans are not robots.** Our bodies execute tasks differently and our nervous systems are not instrumented, so there are no motor commands attached to pixels. This is the central tradeoff underpinning the whole data layer: **the easier it is to collect, the less it resembles what the robot actually needs.**



There is really one problem underpinning all of this: the **embodiment gap**. It shows up in three places: between a human and a robot, between one robot form factor and another, and between simulation and reality.¹⁷ Fleet exhaust data can sidestep this gap, but it is locked away in companies already running fleets of robots. Everyone else has to pay to close it.

Closing The Embodiment Gap

Four overlapping routes have emerged in pursuit of closing the embodiment gap, and we treat the mix a lab selects as a forecast of its cost curve. Each route tries to answer the same question from a different angle: *what does the marginal unit of usable robot data cost, and does that cost fall over time?* Some convert compute into data and get cheaper as compute does. Others convert human headcount and hardware into data and get cheaper only as fast as you can hire people and ship devices. **Routes that scale with compute compound; routes that scale with people don't.** The race is to close the gap at the lowest cost, because

that is what decides whether robot data stays a durable moat or becomes a commodity racing to zero.

Almost no lab picks a single route; most pick two or three and use them in unison. Some labs stack routes into one cohesive strategy, others run them in parallel and only need one to land. [We read the former as a thesis and the latter as insurance.](#) Each of the four routes buys something different, at a different price.

1. The Hardware Route

Match the collection device to the robot and the gap disappears at the point of capture. Genesis and [Proception AI](#) build the most complete versions: robotic hands with a 1:1 match to the human hand, paired with a glove sharing the same interface.^{18,19} [Sunday Robotics](#) ships a gripper-shaped capture glove to a paid team of collectors at \$60/hour.²⁰ [The humanoid form factor itself is the same bet at the scale of the whole body:](#) build the robot as close to human as hardware allows, and what a human-video-trained model imagines, a human-shaped robot should be able to reproduce.²¹

Unit economics are the main constraint. A pair of [Manus](#) motion-capture gloves costs ~\$5,000, compared to the few dollars it takes to strap a cheap camera to someone's head. The moment you ship hardware to a collector, cheap labor collapses into the logistics of shipping, monitoring, and recovering devices.²² Hardware design even dictates buy-versus-build: Figure and 1X use off-the-shelf gloves because their hands are essentially human hands with a few degrees of freedom removed, while Sunday must build in-house for its mitten-shaped end-effector.*

The hardware route produces the best fidelity per hour and the worst cost curve *today*. We don't read that as disqualifying. Nothing cheaper produces contact-rich data at this fidelity, which is exactly why it commands a premium, and \$5,000 gloves reflect current volumes rather than a floor. [We expect the fidelity advantage to outlast the cost disadvantage.](#)

2. The Model Route

If hardware closes the gap with the body, the model closes it with the brain. The prize is [cross-embodiment learning](#): one policy trained across many robot bodies, so data collected on one transfers to another. Physical Intelligence treats this as the central unlock.²³ This route exploits an asymmetry: [semantic generalization](#) (knowing an apple is still an apple in a stranger's kitchen) is largely solved and inherited from the internet for free, while [skill generalization](#) (twisting a cap or recovering when an object slips) is the expensive half, because it lives in action data the internet does not contain.

* **Degrees of freedom** refer to the number of independent, single-axis movements a robot can make

Figure's Project Go-Big was early proof: trained only on egocentric video gathered through a 100,000-home partnership with [Brookfield](#), Helix learned to navigate homes from natural-language commands with zero robot data.²⁴ It worked for locomotion but has not been shown to solve dexterous manipulation.

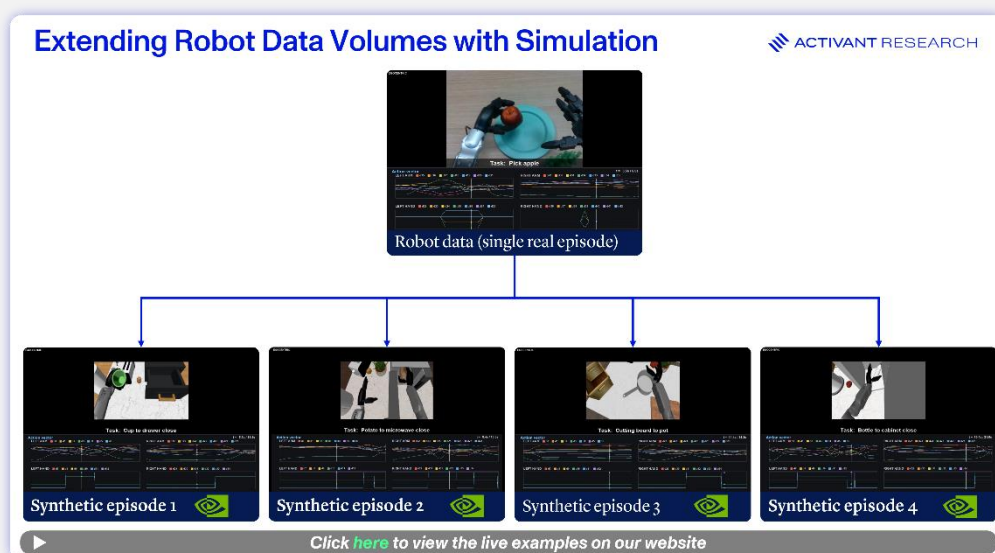
In August, Figure brought [Index](#) out of stealth: a consumer app paying people from 108 countries to film their own chores.²⁵ Index is the Go-Big thesis scaled a hundredfold; the platform has already attracted over 16 million video uploads and pulls in roughly thirty minutes of footage every second.²⁶ If Go-Big provided Helix with sufficient data to solve locomotion with no robot data, then it is very possible that Index will provide Helix with the fuel to solve the next frontier: manipulation.

We think this route will work. If it lands, the marginal cost of usable robot data collapses and the premium on embodiment-native collection goes with it. If it doesn't, a great deal of unlabeled video sits on a very expensive balance sheet.

3. The Simulation Route

This route generates robot data computationally instead. Physics engines like Nvidia's [Isaac Sim](#) and Google DeepMind's [MuJoCo](#) multiply a handful of real demonstrations into large trajectory sets; world models like [Odyssey's](#) Odyssey-2, Nvidia's [Cosmos](#), and Activant portfolio company [General Intuition's](#) MIRA synthesize experience and serve as the environment the model trains in. The appeal is that simulation converts the one input physical AI has in surplus, **compute**, into data, which is why [Skill](#) and [Flexion](#) lean on it so heavily.

The limit is the **sim-to-real gap**: a simulator is only as good as its embedded physics. A hand-authored engine is capped by the physics its designers programmed; a world model is capped by its training data, and it learns the messy dynamics no one can write down by hand. The gap narrows as world models improve, but until it closes, **synthetic data multiplies real data volumes rather than replacing it**, inheriting a slice of the native-data premium rather than escaping it entirely.

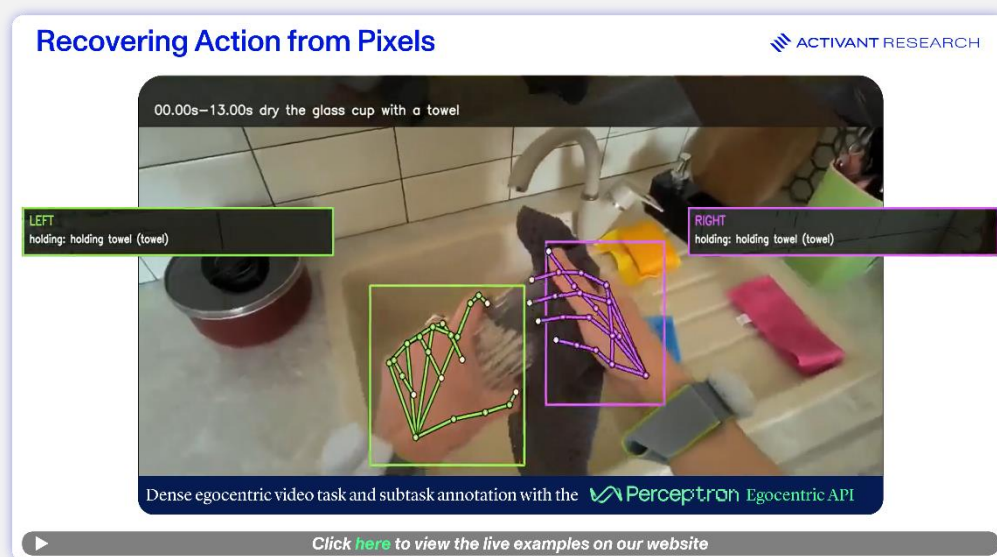


This makes simulation the strongest complement in the stack and the weakest thing to own outright [if the sim-to-real gap stays open](#). This also makes Nvidia the most consequential player in the field without needing to be right about any of this: it sells the multiplier to everyone and carries the gap risk of no one.

4. The Pipeline Route

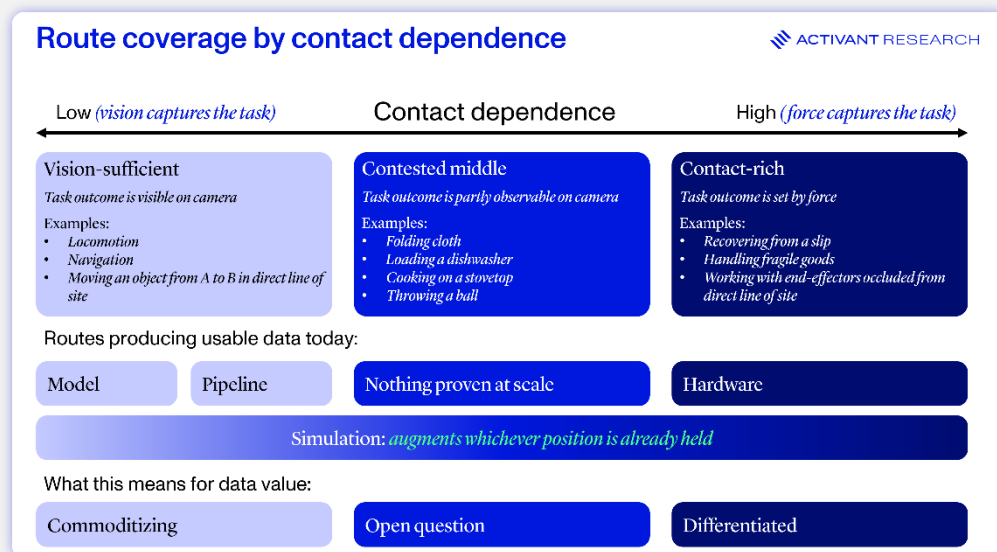
The pipeline route ingests ordinary egocentric video and post-processes it until it resembles robot data. A deployed robot logs the exact command sent to every joint; human video records none of it, so the challenge is inferring those commands precisely enough for a robot to repeat the motion. This is the hardest route, but it is the most leveraged, because it applies a software margin structure to a data problem: [the cheapest input converted into the most expensive output](#).

Pipeline components are starting to productize: [Scale AI](#) is extending its annotation engine into physical AI; [Mecka AI](#) is building EgoVerse; [Perceptron's](#) Egocentric API focuses solely on hand and sub-task annotation for manipulation; and [Neural Motion](#) is training a foundation model purpose-built for cross-embodiment data transfer. [Tesla](#) is the clear vertically integrated example: in mid-2025, it brought its entire data acquisition stack in-house, dropped the motion-capture and teleoperation rigs the rest of the industry depends on, and pivoted to camera-first collection, betting that action can be recovered from vision at scale, straight out of the Autopilot playbook.²⁷ Figure's Index is the newest: crowdsource collection through an app, convert in-house, and treat distribution as the acquisition channel.



[We think the pipeline route is worth underwriting](#), but with a caveat: nobody has published the number that decides if it works. Cost per collected hour is disclosed and falling. Cost *per usable hour* is not. Three of the five components in Figure's pipeline exist to fight the incentives its own payment model creates, which tells us that there is a chasm between collected data and usable data volumes.²⁸

Set the four routes against what is possible today and they stop looking like competitors. They divide the work between them, sorted by how much the task depends on contact.



Our view is that the gap closes from the vision-end inward and slowly: the vision-legible half commoditizes while the hardware route holds the half that pays. The contested middle is where this gets settled. Folding clothes and loading a dishwasher are the tasks every domestic demo shows, and the tasks nothing cheap has yet produced usable data for. Whoever converts that band first sets the price for everything behind it.

Figure, 1X and Tesla agree with most of this: they build their own pipelines rather than buying hours and made their hands near human so video would transfer. The disagreement is about rate. Figure's commitment of over \$1B to data over the next 12 months is a bet that the embodiment gap melts faster and deeper into manipulation than we expect. A dexterous manipulation result from Index data inside the next twelve months would mean we are wrong about the pace.



Teleoperation can't capture true expert behavior. The people operating the robots aren't the professionals who actually do the work, and cameras alone miss the physics of manipulation, including contact force and tactile feedback. To capture real human dexterity, you need to record professionals doing their actual jobs, using video for what they see and tactile gloves for what their hands feel and do."



Junghee Ryu
Founder & CEO, RLWRLD

Touch is the single hardest modality to capture reliably.²⁹ Haptic sensors are improving and the first tactile foundation models are appearing, but the modality

remains unstandardized, which caps the near-term opportunity in sensing itself to component sales into a fragmented buyer base.^{30,31} **We expect embodiment-native tactile data to hold its premium longest.**

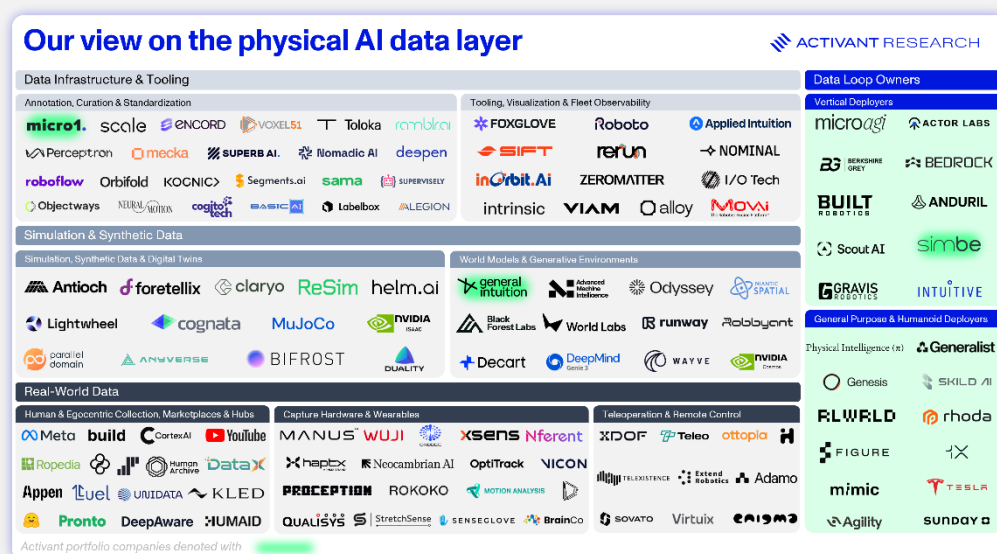
These four routes tell us how the gap gets closed, not who gets paid for closing it.

Where The Value Accrues

Mapping the data layer this way suggests an intuitive trade: if physical AI is bottlenecked by data, back the companies selling data. Picks and shovels have rewarded investors in every previous gold rush, and a whole cohort of physical-AI data companies now exist for exactly that reason.

We think the instinct is right and the target is wrong. In a gold rush you sell picks and shovels. You do not sell ore. Ore is worth whatever the cheapest supplier will take for it, and robot data works much the same way, with one problem on top: it goes stale. Revise the hardware and a good deal of what you already collected no longer matches the robot you are building.

So, the line that matters is not between the data and the tools. **It is between selling the input and owning the loop that keeps producing it:** *collect, train, deploy, capture, retrain*. A sale happens once. A loop pays out every time it turns.



Not The Sellers of Hours

The first physical AI data pure play sells human effort by the hour: paid collectors doing chores in head-mounted cameras; warehouses of rentable robot arms with teleop crews in the Philippines; workers in Shenzhen wearing capture gloves.³² It is a brutally competitive market. Hours go for \$15–\$40 and vendors "spawn like weeds" to the point that the constraint on the labs buying the data isn't finding a vendor but having enough engineers to vet the stream of them.³³

Competition has pushed the hourly rate down, but not far enough for the recipe those hours are meant to fund. Scaling a VLA-style model to general-purpose capability on demonstrations alone plausibly exceeds 100 million teleop hours. At current market prices, 100 million hours would mean \$1.5–\$4 billion in pure data spend before a single robot ships. And that data may only be partly transferable once the robot hardware is revised. **No lab without significant funding is buying its way up the scaling curve at those prices.**

This points to the deeper problem with this market: every one of its customers is funding R&D whose explicit purpose is to need less of the product. All four routes we explored above are attempts to buy *fewer hours*, not more. Barriers are also low enough that the better-capitalized labs simply do it themselves.^{34,35} Figure is explicit about why: it tried buying data first, and reports that "*vendors couldn't hit the throughput, diversity, or quality bar Helix requires.*"³⁶

The price floor is set by forces the vendors don't control either. Open commons like [Open X-Embodiment](#), [Ego4D](#), and [X-DOF's ABC-130K](#), plus state-backed Chinese collection facilities operating at industrial scale, all push the value of *undifferentiated* hours toward zero.^{37,38,39}

Here, the LLM analogy *is* tempting. Data licensing became recurring revenue for [Reddit](#) and the news archives.^{40,41,42} We think it points elsewhere in physical AI. Two things make Reddit's corpus worth a subscription rather than a single cheque: it replenishes daily as the byproduct of a business that earns its money elsewhere, and nobody can recreate twenty years of it at any price. Teleop hours fail both tests. They replenish only while someone is paying, and anyone willing to pay can recreate them. Selling hours solves a real cold-start problem for a lab that has none, but it is a one-time sale of an undifferentiated input into a market racing to the bottom, and as we flagged in [our first robotics piece](#): **it does not compound**. The corpus that does replenish, and that nobody can recreate, is fleet exhaust, but its owners are not selling.

The Sellers of Rails

Adjacent to the data itself sits the unglamorous plumbing: systems that ingest multimodal robot logs, store them, and let engineers search, visualize, and mine them for the rare, useful episode in a flood of repetitive footage. Visualization and annotation tooling has frontrunners in [Foxglove](#), [Sift](#), [Rerun](#), [Encord](#), and [Voxel51](#), but the deeper infrastructure layer has largely resisted productization. The reason is the chasm in the buyer base: research teams are content with off-the-shelf tools, while the handful of well-funded labs that would pay real money build everything in-house.



Data is a commodity, but the cost at scale is still huge, so investing in the right data is paramount. At Voxel51, our aim is to provide the world's best workbench for identifying the right data to feed into your flywheel."



Brian Moore

Co-founder & CEO, Voxel51

We think the category is investable, with a condition. A rails business is the mirror image of selling hours. An hour is a transaction with no integration and no stickiness; a vendor whose tooling carries a customer's data loop compounds switching costs with every pass and **gets paid whichever way the embodiment gap goes**. Ingestion, curation, visualization, and fleet observability are needed identically whether the trajectories arrive from a teleop rig, a capture glove, a simulator, or a video pipeline.

The condition is the chasm: a rails business has to stay horizontal enough to serve the wider market *and* become useful enough to displace the in-house tooling its best-funded customers would otherwise build. That is a narrow path, and the graveyard of ML tooling companies squeezed between open source and platform teams is the reason to underwrite company by company rather than thematically.

The Loop Owner

Building the brain is a manufacturing problem, and the loop is the factory: collect data, train, deploy, capture what the deployment throws off, retrain. Every position discussed so far sells one input into someone else's loop. Only one owns the loop itself.

The obvious objection is circular: capturing fleet exhaust data requires a deployed fleet, running a deployed fleet requires robot capability, and robot capability requires data the fleet has not yet produced. The loop must be started from outside itself, and this is where most of the industry's capital is currently exposed. **What separates a loop owner from a robotics company with a shiny pitch deck is whether the first turn is funded by revenue rather than by capability that does not exist yet.**

This points to a specific shape of business: a wedge task that is narrow enough to be solvable with today's models, valuable enough that a customer pays for the output rather than the novelty, and *just* repetitive enough that serving it generates the data the next task needs. Manufacturing, warehousing, and food production all qualify; a general-purpose domestic robot does not, and won't for some time. This wedge is not the whole market; it is the ignition.

None of this is priced sympathetically though. Loop ownership is the most capital-intensive, longest-payback position in the space, but it is also the most legible: the

marks in humanoids and frontier RFM labs already assume the loop closes. Our view from [our last article](#) has not changed; direction is far easier to underwrite than timing, and at current entry prices you are mostly paying for timing. Two things make the loop bet work anyway. The first is the wedge test above, which converts a binary capability bet into a compounding operating business with revenue today: a different risk shape, and eventually a different multiple. The second is that loop ownership is the only position that does not require us to be right about the embodiment gap.

[If the gap holds](#), embodiment-specific data stays king and the advantage belongs to whoever already has robots in the field. [If it melts](#), cross-embodiment learning and direct human-to-robot transfer make cheap egocentric and internet video usable, the data stockpiles deflate, and value moves to whoever owns the conversion technology. The loop owner is long both states: it holds the embodiment-native asset in the first, and it is the highest-value customer for the conversion technology in the second. Everyone else has to be right about which state arrives, and when. Our read is that it arrives in stages, melting first where contact is simple and vision does the work, last where manipulation is dexterous. The stockpile deflates from the easy end, and what is hardest to collect is what holds its value.

[Microagi](#) is a live example of the profile: it collects its own human data through its [Shift](#) division, deploys robots onto the factory floor, and fine-tunes on what those deployments generate, owning the loop end to end while renting the commoditizing hardware and models around it. Microagi raised the largest seed round in German history, and have reportedly grown their ARR from \$0 to \$100M in less than six months using this recipe.^{43,44} The ex-Waymo team behind [Bedrock Robotics](#) retrofits autonomy onto customers' existing excavators rather than selling new machines, and is running supervised deployments on live jobsites while it works toward full autonomy, so the paid operation *today* doubles as the training set for *tomorrow*.⁴⁵ [Actor Labs](#) skips fleet ownership entirely, paying equipment owners to host its collector nodes and training on operator data from machines already earning.⁴⁶ Activant portfolio company [Simbe](#) has been running this loop since 2016: thousands of its Tally robots scan shelves in stores across ten countries, several times a day, and retailers pay for what the scans tell them rather than for the robot itself.⁴⁷



You don't build great physical AI in a lab. You build it by operating in the real world, every day. After a decade in stores and more than 3,000 robots under contract, every deployment gives us a deeper understanding of how stores actually operate, which makes our technology better and the intelligence more valuable to retailers. That compounding loop is incredibly difficult to recreate."



Brad Bogolea
Co-founder & CEO, Simbe

Our thesis rests on a schedule, so it is worth discussing what would move it.

Signals That Would Change Our Mind ACTIVANT RESEARCH	
Embodiment gap <i>melts</i> faster than expected	Embodiment gap <i>holds</i>
• Cheap video produces reliable dexterous manipulation on hardware the model has never seen • Synthetic trajectories replace real robot data rather than multiplying it • Teleoperation hour prices collapse faster than volumes grow; showing falling demand, not competing supply	• Labs that bet on the model route quietly return to matched capture for contact-rich tasks • Tactile sensing remains unstandardized, with no tactile foundation model reaching production • Fleet operators decline to license their data exhaust at any price. The cleanest revealed preference that the asset is a moat, not inventory

If the embodiment gap is going to melt, we would expect these signals in the next 12–24 months. Several of the ones pointing the other way have already shown up.

The Window

Data is the robot brain's scarcest input today, and the only ingredient in physical AI that nobody inherits. That makes it tempting to treat the input as the asset. **But scarcity has never been a moat; it's a window.** A stockpile of hours depreciates as the premium erodes. A fleet that keeps manufacturing them does not.

We've highlighted that proprietary data generation at scale is a key trait of any durable robotics business. The loop is what it looks like in practice, and the test is not how much data a company holds but whether its own deployments produce the next tranche at a falling cost. **Own the loop while the window is open, and you own the business that remains when it closes.**

If you're building in this space, reach out - we'd love to hear from you. We always learn the most from the people on the ground doing the work.

¹ Activant analysis of company filings and quarterly earnings releases. Figures as reported through Q2 2026.

² Hans Moravec, Mind Children: The Future of Robot and Human Intelligence, 1988

³ Yann LeCun, World Models: Enabling the next AI revolution, 2026

⁴ Google DeepMind, RT-2: Vision-Language-Action Models, 2023

⁵ Google DeepMind, PaLM-E: An Embodied Multimodal Language Model, 2022

⁶ Physical Intelligence, $\pi 0$: A Vision-Language-Action Flow Model for General Robot Control, 2024

⁷ Nvidia, GR00T N1: An Open Foundation Model for Generalist Humanoid Robots, 2025

-
- ⁸ Figure, [Helix: A Vision-Language-Action Model for Generalist Humanoid Control](#), 2025
- ⁹ Genesis, [GENE-26.5: Advancing Robotic Manipulation to Human Level](#), 2026
- ¹⁰ Activant Research, [Robots: Valuations and The Scaling Question](#), 2026
- ¹¹ Mimic, [mimic-video: Video-Action Models for Generalizable Robot Control Beyond VLAs](#), 2025
- ¹² Rhoda, [Causal Video Models Are Data-Efficient Robot Policy Learners](#), 2026
- ¹³ 1X, [1X World Model | From Video to Action: A New Way Robots Learn](#), 2026
- ¹⁴ Skild AI, [Building the general-purpose robotic brain](#), 2025
- ¹⁵ Generalist, [Going Beyond World Models & VLAs](#), 2026
- ¹⁶ TechCrunch, [Enigma raises \\$70M to make controlling a robot as easy as adjusting the volume](#), 2026
- ¹⁷ Anonymous Authors, [The Embodiment Gap in Robot Foundation Models](#), 2026
- ¹⁸ Genesis, [GENE-26.5: Advancing Robotic Manipulation to Human Level](#), 2026
- ¹⁹ Proception AI, [Shipping ProHand: The Human-Like Robotic Hand Built for Dexterous Manipulation](#), 2026
- ²⁰ Sunday, [ACT-1: A Robot Foundation Model Trained on Zero Robot Data](#), 2025
- ²¹ 1X, [1X World Model | From Video to Action: A New Way Robots Learn](#), 2026
- ²² Activant Expert Network
- ²³ Ibid
- ²⁴ Figure, [Project Go-Big: Internet-Scale Humanoid Pretraining and Direct Human-to-Robot Transfer](#), 2025
- ²⁵ Figure, [Introducing Index: Building The World's Largest and Most Diverse Physical Dataset](#), 2026
- ²⁶ Ibid
- ²⁷ Markets.com, [Tesla Shifts Focus to Pure Vision for Optimus Robot Training](#), 2025
- ²⁸ Figure, [Introducing Index: Building The World's Largest and Most Diverse Physical Dataset](#), 2026
- ²⁹ Activant Expert Network
- ³⁰ RLWRLD, [RLDX-1: A Dexterity-First Foundation Model for Robot Hands](#), 2026
- ³¹ Zhou et al., [TouchWorld: A Predictive and Reactive Tactile Foundation Model for Dexterous Manipulation](#), 2026
- ³² Rest of World, [China is training a robot future — one folded shirt at a time](#), 2026
- ³³ Activant Expert Network
- ³⁴ Generalist, [GEN-0 / Embodied Foundation Models That Scale with Physical Interaction](#), 2025
- ³⁵ Figure, [Project Go-Big: Internet-Scale Humanoid Pretraining and Direct Human-to-Robot Transfer](#), 2025

-
- ³⁶ Figure, Introducing Index: Building The World's Largest and Most Diverse Physical Dataset, 2026
- ³⁷ Open X-Embodiment Collaboration, Open X-Embodiment: Robotic Learning Datasets and RT-X Models, 2023
- ³⁸ Facebook AI Research et al., Ego4D: Around the World in 3,000 Hours of Egocentric Video, 2022
- ³⁹ XDOF, ABC-130K: The largest open source teleoperation dataset, 2026
- ⁴⁰ AP News, ChatGPT-maker OpenAI signs deal with AP to license news stories, 2023
- ⁴¹ TechCrunch, OpenAI inks deal with Axel Springer on licensing news for model training, 2023
- ⁴² TechCrunch, Reddit says it's made \$203M so far licensing its data, 2024
- ⁴³ Microagi, microagi raises \$55 million seed round to power Europe's robotic reindustrialization, 2026
- ⁴⁴ Dealroom, MicroAGI claims \$100M+ annual run rate six months after Shift launch, 2026
- ⁴⁵ Bedrock Robotics, We Shouldn't Have to Choose What Gets Built, 2026
- ⁴⁶ Actor Labs, A VLA Policy for an Excavator, 2026
- ⁴⁷ Simbe, Meet Tally, 2026

Disclaimer: The information contained herein is provided for informational purposes only and should not be construed as investment advice. The opinions, views, forecasts, performance, estimates, etc. expressed herein are subject to change without notice. Certain statements contained herein reflect the subjective views and opinions of Activant. Past performance is not indicative of future results. No representation is made that any investment will or is likely to achieve its objectives. All investments involve risk and may result in loss. This newsletter does not constitute an offer to sell or a solicitation of an offer to buy any security. Activant does not provide tax or legal advice and you are encouraged to seek the advice of a tax or legal professional regarding your individual circumstances.

This content may not under any circumstances be relied upon when making a decision to invest in any fund or investment, including those managed by Activant. Certain information contained in here has been obtained from third-party sources, including from portfolio companies of funds managed by Activant. While taken from sources believed to be reliable, Activant has not independently verified such information and makes no representations about the current or enduring accuracy of the information or its appropriateness for a given situation.

Activant does not solicit or make its services available to the public. The content provided herein may include information regarding past and/or present portfolio companies or investments managed by Activant, its affiliates and/or personnel. References to specific companies are for illustrative purposes only and do not necessarily reflect Activant investments. It should not be assumed that investments made in the future will have similar characteristics. Please see "full list of investments" at activantcapital.com/companies/ for a full list of investments. Any portfolio companies discussed herein should not be assumed to have been profitable. Certain information herein constitutes "forward-looking statements." All forward-looking statements represent only the intent and belief of Activant as of the date such statements were made. None of Activant or any of its affiliates (i) assumes any responsibility for the accuracy and completeness of any forward-looking statements or (ii) undertakes any obligation to disseminate any updates or revisions to any forward-looking statement contained herein to reflect any change in their expectation with regard thereto or any change in events, conditions or circumstances on which any such statement is based. Due to various risks and uncertainties, actual events or results may differ materially from those reflected or contemplated in such forward-looking statements.