BERNSTEIN
SOCIETE GENERALE GROUP

Humanoid Robotics: The emperor's new brain and the AlphaGo moment - four predictions shaping the next three years

Jay Huang, Ph.D. +852 2123 2631 jay.huang@bernsteinsg.com
Elsa Fu +852 2123 2617 elsa.fu@bernsteinsg.com
Dien Wang, Ph.D. +852 2123 2622 dien.wang@bernsteinsg.com
Weibin Liang, Ph.D. +852 2123 2666 weibin.liang@bernsteinsg.com

To study early-stage technologies, such as humanoid robotics, the best approach is a Bayesian approach — one updates the probability of a hypothesis as new evidence becomes available, combining prior knowledge with new observations. This path has led us a long way from our initial views three years ago to the latest predictions in this note about technology frontier (brain models, data and training technique), products, and industry structure. These predictions shape our forecasts of industry size (what humanoid robots are capable of, at what ROI) and future winners (what drives value and differentiation).

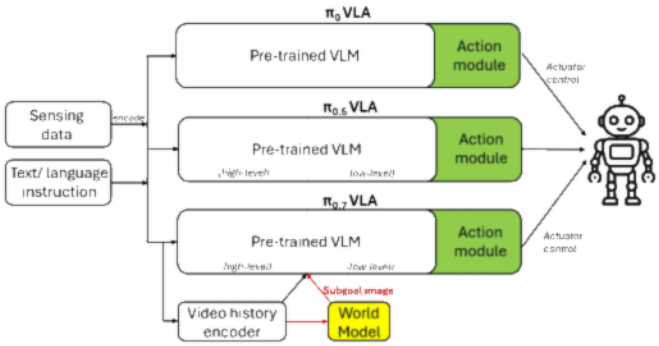
MODEL: THE EMPEROR'S NEW BRAIN

A brain that imagines

The notable breakthrough in robotic brain models today is the newly acquired ability to imagine. Instead of a linear translation from instructions to actions using a traditional VLA (vision-language-action) model, the World Model paradigm imagines the outcome, as actions are planned and determined (see here). We firmly believe this is the right path forward. However, dramatic statements such as "VLAs are dead, long live the World Models" are misleading. The true path is likely a fusion of the two, such as in this example of π0.7 (Exhibit 1). The essence is the model's ability to imagine outcomes, rather than the name tag.

We believe that in the next three years, World Model — especially the subcategory that imagines both the actions and their effects on the environment, and outputs both at the same time, the so-called World Action Model (WAM) — will be widely adopted in the quest for long-horizon autonomy and cross-task, cross-embodiment generalization. However, one important research conclusion in the last two years was that visual realism does not equal physical correctness. The best World Models in 2026 only score 22% when comprehensive "physical understanding" is evaluated. This may still remain a challenge three years from now.

EXHIBIT 1: **The latest Physical Intelligence model integrates VLA and World Model**



Physical Intelligence is private.
Source: Physical Intelligence's website and papers, Bernstein analysis

A brain that starts to remember

Call it a wish — how wonderful it would be if a robot, failing a task, immediately adjusts its actions based on the specific factors behind that failure, so its actions quickly converge toward success; or a robot keeps track of progress in a 30-minute task and adjusts its plan accordingly. Robots are not capable of doing that today, because the brain has minimal memory function. Given the same instructions in the same environment, a robot will take the same action, without considering the contexts of what already happened. Injecting memory into a robot brain is already a new frontier of research. We expect it to be the next great breakthrough.

Multi-modality goes into the brain

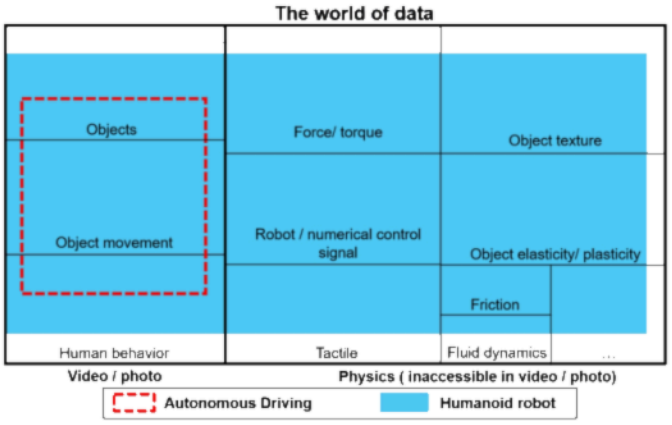
Eighteen months ago, in the note "The hierarchy of infinities", we predicted that non-video, physical data will be a unique challenge differentiating robotics from other AI applications (Exhibit 2). Today, this challenge has materialized as a sudden boom of startups making sensors and UMIs (Universal Manipulation Interface). While it is a great start, it is not sufficient. On the one hand, the variety of physical data remains far too limited. For example, data regarding the physics of the outside world, an integral part of a true World

See the Disclosure Appendix of this report for required disclosures, analyst certifications and other important information. Alternatively, visit our Global Research Disclosure Website.

First Published: 03 Sep 2026 20:30 UTC Completion Date: 03 Sep 2026 09:55 UTC

Model, remains missing. On the other hand, models, as they stand now, already struggle to effectively use the slightly extended data modality. The ideal of true multi-modality is as much a hand-and-eye problem as it is a brain problem. In three years, we predict that robotic brains will become much more capable in fusing third-party / egocentric / palm video, human and robot action, force and tactile sensing data in both model training and real-time action policy generation.

EXHIBIT 2: **Physical data is a challenge unique to robotics, among all AI applications**



Source: Bernstein analysis

DATA: THE "ALPHAGO MOMENT"

Why is manipulation more difficult than locomotion

I still remember the days when I can comfortably beat computer programs in the game Go. That was until 2015, when AlphaGo suddenly beat a professional Go player then a world champion the next year. The great leap forward was enabled by Reinforcement Learning (RL). Before AlphaGo, algorithms learned their moves from numerous human games. AlphaGo introduced a "value network" (i.e., a reward function) to assess whether certain game moves are toward success or failure, hence closing the loop for RL.

The training of robotic models today is the pre-AlphaGo analogy. A model is fed with numerous data to learn tasks, and the goal is to match human behavior. It gets to a basic level of performance, then is stuck. People have long thought of using Reinforcement Learning to greatly improve performance. In fact, the technique worked beautifully to render robot locomotion a solved problem — progressing from barely able to walk to breaking human records in fully autonomous sports games (link) in just a few years. However, there has only been limited success using the same technique to train robots for manipulation tasks. The critical difference is the difficulty in developing effective reward functions for complex tasks. While such reward values (more likely to succeed / more likely to fail) are straightforward for Go (since 2015) and for robotic locomotion (since around 2023), it proves elusive for manipulation without a great deal of manual input. The added complexity is to use RL in a simulator to improve training efficiency. For locomotion, one only needs to simulate the robot, but for manipulation, one also needs precise simulation of interactions and the evolution of the world, which is rudimentary at best.

Therefore, forget the "ChatGPT moment" for robotics — depending on the perspective, one may say that it already took place, because robotic brain models are already built on the transformer architecture, the breakthrough behind ChatGPT, or one may say it will never come, because Physical AI adoption is more of an evolution and will not have a drastic inflection point. Instead, embrace the "AlphaGo moment" in robotics, when people solve the Reinforcement Learning problem for manipulation to push the success rate of most tasks efficiently from 60% to 99%. My best guess is that we are a few years away from important breakthroughs, and robotic manipulation training today is like Go-playing algorithms in 2012.

Data is overrated

Data is critical — that is the "politically correct" statement, but data is also overrated.

In pre-training, advancement in robotic models have greatly expanded the scope of usable data. Slow and costly teleoperation is supplemented, if not completely replaced, by internet videos, ego- and palm/wrist-cameras, and UMIs. "Oversupply of pre-training data" is a slight exaggeration of the status quo, but we are probably not too far from that. Post-training / deployment data is indeed scarce, but the popular notion of an automatic "data flywheel", i.e. "more deployment, more data, better performance", is completely missing the point. The primary challenge is not the amount of post-training data, but the broken link between data and task performance, which, as we've seen, requires breakthroughs in Reinforcement Learning. "Pretrained to imagine, fine-tuned to act" is the tested approach, but cutting edge World Action Models use millions of hours of easily accessible internet video data for pretraining, and only tens of hours of proprietary action data for post-training.

We can now complete the AlphaGo analogy for robotics. Before AlphaGo, data was king, because model training required as many human games as possible. AlphaGo started with a much smaller set of human games for initial training, followed by performance improvement through Reinforcement Learning. Just one year after AlphaGo beat world champion Lee Sedol, AlphaGo Zero beat AlphaGo 100:0, and it was trained exclusively by playing games against itself without any expert data. This approach was further generalized in AlphaZero that learned, with zero data, to play many types of games in addition to Go. This was a clear trajectory of improved algorithms and training techniques drastically reducing the needs of data.

Certain types of data, such as tactile sensing, will likely see much higher demand and continue to be the scarcity, but overall, the data segment is just like everything else in the humanoid robotics industry — low barrier to entry, high ceiling to differentiation (see

here). At the industry level, we believe that **the best place to build a moat about data is in the brain**. So far we've discussed how innovations in models and training techniques are required to accommodate expanded data modality, reduce data burden, and close the missing link in the data flywheel. Another issue is that different models are all selective on data in different ways — one man's meat is another man's poison. It is often said that only 10% of collected data is usable. The most effective way around this challenge, in my view, is to tailor data collection to the specifics of a model. We predict that model-data integrated development, such as the approach taken by Physical Intelligence (Exhibit 3), will become a sustainable moat.

EXHIBIT 3: **Physical Intelligence's model-data integrated approach**

Model	Release time	Training data scope	Progress
π0	10/31/2024	The largest **multi-task and multi-robot** interaction dataset to the release date. incl. - **Internet-scale** vision-language **pretraining** (3Bn parameter) - **Open-source** robot **manipulation** datasets (Open X-Embodiment dataset) - **Own datasets** consisting of **10k hours** of demonstrations from **eight distinct robots**	**Enable the generalist robot policy**
π0.5	4/22/2025	Co-training on **heterogeneous data** - Various robotic data sources from many **different robot types** incl **400 hours mobile manipulation data** collected with the same robot that is use in PI experiments - **Multimodal data** incl high-level subtask commands, instructions, and data from the web	- **Meaningful generalization** to entirely **new environments** (e.g., new homes) - Can accept language commands at various levels of granularity
π0.5 + EGO	12/16/2025	Use **ego-centric data** from humans to provide a 2x improvement on tasks where robot data is limited	Emergence of **human to robot transfer**
π0.6	11/17/2025	Beyond pre-training data, **include post-training data**, such as corrective teleoperation data collected when the robot makes a mistake and robots' autonomous experience	**Reinforcement learning**
π0.6 with MEM	3/3/2026	**Multimodal memory** that enable the model to keep track of tasks up to 15 minutes and even correct mistakes	**Multi-scale embodied memory (MEM)**
π0.7	4/16/2026	**Context and visual subgoals provided** - Broad and diverse data, incl. **data from many different robots, humand data, and autonomous epsodes** - **Diverse context** to the prompt that specify not only what to do but how to do it - **Visual subgoals** produced by a **world model**	- **Perform tasks that were never seen** in its training data - **Cross-embodiment transfer**

Physical Intelligence is private.
Source: Physical Intelligence's website and papers, Bernstein analysis

PRODUCTS: CROSSING THE RUBICON

Approaching the 2.5-year payback threshold

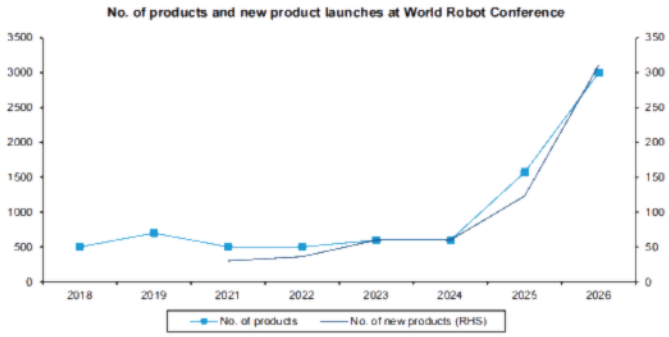
What characterizes Humanoid Robotics in 2026 is the quick shift of focus from locomotion demonstration to commercial adoption. In our recent note (link), we analyzed four key areas of adoption in the next five years — warehouse pick-and-place, factory material handling and inspection, urban and industrial patrol, and last-leg delivery to door — and showed that their payback periods can soon fall below 2.5 years. Combined, the four applications account for about half of our projected 1-million annual shipment in 2031. In our analysis, the shortening payback is primarily driven by reduced robot prices, which in turn are a function of increased volume. Empirical evidence from the EV and battery industries shows that costs typically decline by15-18% when volume doubles, with a much steeper curve at very small volumes.

Based on our experience with industrial robots, for industrial and commercial businesses, crossing the 2.5-year payback is like crossing the Rubicon River — after that there is no return, and adoption decisively accelerates. 2.5-year is not a magic number but an average threshold. Established large enterprises tend to tolerate longer payback periods up to 5 years, while some smaller companies in dynamic industries want it to be as short as 1 year before broad adoption starts.

Platform hardware, killer SKUs

The humanoid robotics industry is currently experiencing explosive expansion of SKUs (Exhibit 4). We believe the trend will continue to cover the expanding scope of applications. At the same time, we predict the rise of a few platform robot products. The industry is unlikely to see consolidation of players in the next three years, but it is likely to have concentration of volume to some killer SKUs. These robot SKUs will have designs optimized for material handling in factories and warehouses, with10-15kg sustainable payload each arm.

EXHIBIT 4: **There have seen explosive growth in the number of total products and new products at WRC in 2025 and 2026**



Source: industry news, Bernstein analysis

Rising to the platform hardware status will have great benefits. Economies of scale in procurement and production is only one of those. Model developers may also follow those platform hardware in model design and training, hence creating a virtuous circle. The critical factors to achieve that status are volume (through commercial success) and consistency / reliability (with hardware design and manufacturing know-how). This layer of potential differentiation among robot OEMs is generally underappreciated. Compared to mature technologies such as industrial robots, product consistency and reliability is a pain point not only for individual players today, but also for the humanoid robotics industry as a whole.

INDUSTRY: "ROSEBUD … A PIECE IN A JIGSAW PUZZLE – A MISSING PIECE"

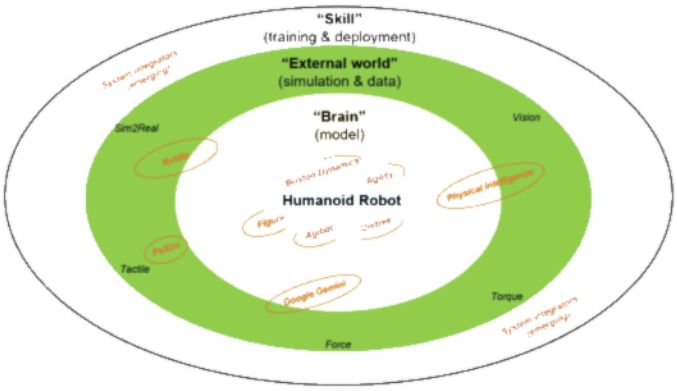
Human brain, eye, hand, joint … all have counterparts in a humanoid robot, but what is the counterpart of "skills"? Too often, robot skills (e.g. "sorting parcels", "loading a dishwasher") are confused with the brain model, and humanoid robot companies are assessed on what

skills their robots have demonstrated. But the distinction is clear in the human analogy — a healthy human child has a capable brain, but she wouldn't be able to fold shirts or wash dishes automatically — trust me, I observed it firsthand.

Most skills are not part of the brain model. They are what a model is capable of learning. The confusion partly arises from the reality that at such early stage of the industry, humanoid robot makers and model developers routinely assume the role of the skill enablers — they do the hard work teaching their robot children to fold shirts and wash dishes.

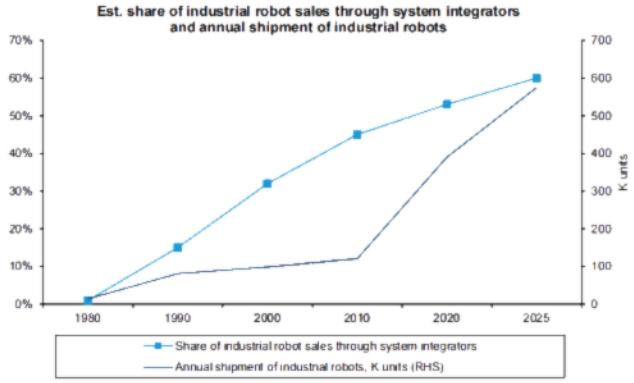
For the industry to scale up, the skills layer needs to develop separately (Exhibit 5), because eventually, there will be several brain models, tens of robot makers, hundreds of robot SKUs, but many thousands of applications and use scenarios. In industrial robotics, the skill layer consists primarily of system integrators, complemented also by OEMs, such as FANUC, and independent technology enablers, such as Mech-Mind. They collaborate to enable new applications and drive deployment to new corners. In humanoid robotics, "system integrator" is the last missing piece of the jigsaw puzzle, and we expect it to emerge in the next few years, partly freeing robot makers and model developers the burden of deployment, data collection, and post-training. Once starting, it will be a decades-long journey (Exhibit 6).

EXHIBIT 5: **The ecosystem of humanoid robotics and player examples– for the industry to scale up, the skills layer needs to develop separately**



Google is covered by Bernstein U.S. Internet team. Nvidia is covered by Bernstein U.S. Semiconductors team. Boston Dynamics' major shareholder is Hyundai (covered by Bernstein Korea Autos & Auto Parts team). Unitree and Mech-Mind are public (not covered). Agility agreed to go public through a merger with the SPAC Churchill Capital Corp XI (not covered). Other companies mentioned in the picture are private.
Source: Bernstein analysis

EXHIBIT 6: **Over four decades, the share of industrial robots sold through system integrators has steadily increased**



Source: industry news, IFR, Bernstein analysis

PICKING WINNERS

What do all of these mean for investors looking for future winners in the industry?

On the next generation brain models, our three predicted directions all point to *Physical Intelligence*. This preference is further strengthened by our conviction in model-data integration ("the moat of data is in the brain"). Other notable "brains" are from *Figure AI* (task generalization and fusion of data modality), *Nividia* (World Models as planner and simulator), and *Google*. Two early-stage startups, *Dyna Robotics* and *Rhoda AI*, have caught our attention recently by demonstrating the first "scaling law" in robotics and using World Models to greatly shift the data burden from the hard-to-acquire robotics data to internet-scale human videos. In China, *Agibot, Galbot, and Robotera* are the rising stars.

Data scarcity coexists with data oversupply. The latest boom of UMIs to generate pretraining data may not be sustainable. In our view, tactile sensing is the more certain opportunity. There are several technology paths for tactile sensing, and it's too early to bet on a single path. *PaXini* and *Tashan Technology* are two players we watch closely.

On the robot itself, the race to be the industry's platform product has no clear front-runners yet. Hardware design and consistency is the only area where *Tesla's Optimus* might claim an advantage, thanks to its supply chain and manufacturing capability, but then it appears so lagging behind in other important technologies including the model and data / training technique. The long list of promising contenders include *Boston Dynamics, Figure AI, Agility, Unitree, Agibot, LimX* … We don't have the final answer, but we know the most important new observation to insert in our Bayesian formula in order to update our prediction — the early evidence of system integrators gathering around certain robot makers.

Note: Nvidia is covered by Bernstein U.S. Semiconductors team. Google is covered by Bernstein U.S. Internet team. Boston Dynamics' major shareholder is Hyundai (covered by Bernstein Korea Autos & Auto Parts team). Unitree and Mech-Mind are public (not covered). Agility agreed to go public through a merger with the SPAC (not covered). Other companies (Physical Intelligence, Figure AI, Dyna Robotics, Rhoda AI, Agibot, Galbot, Robotera, LimX, PaXini, Tashan Technology) mentioned in the picking winners section are private.

人在草木间
更多资料请加V

BERNSTEIN TICKER TABLE

Ticker	Rating	Cur	2 Sep 2026 Closing Price	Price Target	TTM Rel. Perf.	Cur	Reported EPS 2025A	Reported EPS 2026E	Reported EPS 2027E	Reported P/E (x) 2025A	Reported P/E (x) 2026E	Reported P/E (x) 2027E
CGNX (Cognex)	O	USD	59.94	85.00	16.8%	USD	0.67	1.57	1.80	89.5	38.1	33.4
6324.JP (HDSI)	O	JPY	5,430.00	7,800.00	90.2%	JPY	16.99	57.37	79.51	319.6	94.6	68.3
300124.CH (Inovance)	O	CNY	60.15	82.00	(47.4)%	CNY	1.87	2.19	2.65	32.2	27.4	22.7
002747.CH (Estun)	M	CNY	30.81	31.60	(0.0)%	CNY	0.05	0.30	0.30	616.2	101.2	102.8
2715.HK (Estun)	M	HKD	16.50	17.12	NA	CNY	0.05	0.30	0.30	330.0	54.2	55.0
6954.JP (Fanuc)	O	JPY	5,761.00	7,200.00	3.7%	JPY	178.47	221.78	207.13	32.3	26.0	27.8
6861.JP (Keyence)	O	JPY	77,980	95,000	(1.2)%	JPY	1,835.62	2,420.47	2,675.94	42.5	32.2	29.1
HSAI (Hesai)	O	USD	16.59	29.00	(65.5)%	CNY	2.91	3.18	5.26	38.3	35.0	21.2
2525.HK (Hesai)	O	HKD	17.12	29.00	NA	CNY	0.36	0.40	0.66	40.3	36.9	22.3
688017.CH (Leader Drive)	U	CNY	290.41	115.00	58.3%	CNY	0.69	0.76	0.83	419.4	379.6	350.4
601689.CH (Tuopu)	O	CNY	47.83	75.00	(56.1)%	CNY	1.61	1.81	2.36	29.7	26.4	20.3
002050.CH (Sanhua)	M	CNY	36.30	39.00	(13.5)%	CNY	1.03	0.95	1.08	35.2	38.0	33.7
2050.HK (Sanhua)	M	HKD	24.72	27.00	(47.4)%	CNY	1.03	0.95	1.08	20.6	22.2	19.7
002472.CH (Shuanghuan)	O	CNY	35.42	60.00	(35.2)%	CNY	1.50	1.59	1.97	23.6	22.3	18.0
SPX			7,666.60									
JPL			2,650.95									
ASIAX			1,949.13									

O - Outperform, M - Market-Perform, U - Underperform, NR - Not Rated, CS - Coverage Suspended

6954.JP, 6861.JP base year is 2026;

Source: Bloomberg, Bernstein estimates and analysis.