NOMURA

Global Markets Research
23 October 2026

Global Robotics

EQUITY: TECHNOLOGY

The ChatGPT moment comes into view

Frontier models tested on robots likely outperform; December-January release is likely the first catalyst

Research Analysts

Advanced Manufacturing

Frank Fan - NIHK
frank.fan@nomura.com
+852 2252 2195

Donnie Teng - NIHK
donnie.teng@nomura.com
+852 2252 1439

Robotics replays the LLM path' its ChatGPT moment closer than it looks

Robots are following the same path that large language models (LLMs) have taken, and their ChatGPT moment could come sooner than expected, in our view. A scientist at NVIDIA (NVDA US, Not rated) expects robot training going through the same three stages as LLMs: 1) robots learn broadly from large amounts of data; 2) they are tuned for specific tasks; and 3) they improve through trial-and-error with RL (reinforcement learning). Our industry checks in September indicate sentiment on robots has turned more upbeat for two reasons: 1) robots' precision level jumps sharply once they are trained on sufficient action feedback, the same kind of sudden leap seen in LLMs; and 2) in our view, there is increased likelihood of robots reaching the equivalent of L2+ assisted driving within two years. Further, we expect the breakthrough to come from big AI labs and Tesla (TSLA US, Not rated). Our industry survey suggests OpenAI's (unlisted) internal models have performed well without the touch-sensor data that labs thought they would need. **Catalysts:** OpenAI's physical-AI model, ByteDance (unlisted) and DeepSeek's (unlisted) new models, and OpenAI humanoid demo and GPT-7 (likely in 2027F).

Ego video leads pre-training as training gets increasingly cheap

Beyond models, we believe data will decide who takes the lead. It is the scarcest input, and the industry is turning to videos to fill the gap. Following the same steps as LLM training: 1) robots first learn from a large volume of first-person (Ego) videos of people doing everyday tasks; 2) they are then fine-tuned on a small amount of teleoperation and motion-capture data; 3) and finally they undergo practises by trial-and-error in simulation (for hand manipulation, more pre-training compute will bring steady and predictable gains, the same scaling law that has driven LLMs in the past). Training costs, meanwhile, are falling fast. *SuperMimic* , a CoRL (Conference on Robot Learning) 2026 paper, teaches one model that can dictate a robot's body moves and trains the controller inside the model at the same time. After just three hours on one GPU, the model beats a baseline that has been trained for seven days on 128 GPUs. The paper reports more than 1,000x less compute, and the learned model has stayed accurate as motion data grows from 1 to 264 hours. Pre-training is only half the story, however. To keep improving, models need data from robots working in the field, and that has direct implications for hardware.

Dexterous hands benefit first; body consistency decides who wins in hardware

Better brains likely help dexterous hands first. Robots have been stuck with grippers mainly because the data needed to train hands is relatively limited. Stronger foundation models should be able to carry skills learned on two- or three-finger grippers over to five-finger hands, enabling volume lifting and bring unit prices down. We expect 2026F hand shipments to grow 80-120% y-y, led by high-DOF (degrees of freedom) hands. Once the model route settles, we expect humanoid designs to gradually settle into a handful of standard forms: an "iPhone moment" for hardware. From then on, the hard part would be building every unit in a same/similar way. This matters for data as much as for quality. A humanoid has more than 20 joints, and if the humanoid behaves slightly differently from one another, a control policy trained on one robot might only work on few others. Therefore, we believe the winners would be makers that iterate fast and keep costs tight, such as Unitree (688836 CH, Buy), automakers, and key component suppliers. As Unitree volumes grow, we expect it to outsource more components from external parties, both to maintain low costs and to secure quality. **Risks:** 1) robotics OEMs with poor product consistency that rely heavily on data-collection customers could see demand fall; and 2) brain start-ups business models might change.

Production Complete: 2026-10-23 12:11 UTC

See Appendix A-1 for analyst certification, important disclosures and the status of non-US analysts.

Robot brains: a breakthrough is plausible

Robotics is replaying the LLM path, but from an earlier starting point

Measured against how LLMs developed, robot brains have not yet reached the GPT-2 stage: early signs of emergence have shown up in labs, but no model is reliable enough for real work. A scientist at NVIDIA argued that *robots will go through the same three steps as LLMs*: 1) broad pretraining; 2) tuning to follow instructions; and 3) RL (reinforcement learning), where the model improves by trial-and-error. We compare today's robotics to the AlexNet moment of 2012, when deep learning first proved itself. Our industry checks show: 1) December-January models would be closer to GPT-2 — promising, but not ready for production; 2) world models are less mature than AI was in 2012; and 3) the most bullish view comes from Wang Xingxing, founder of Unitree, who said on 22 October that physical AI was already showing early signs of a breakthrough, likely within one to two or two to three years. He added that almost everyone, himself included, continued to under-rate AI. In addition, OpenAI CEO Sam Altman told Fortune he expected a "cool demo" of an OpenAI robot in 2027, with robots on the streets a few years thereafter. In our view, today's robot brains are not yet at GPT-2 level. In the next round of model release will, likely, bring them up to it.

Fig. 1: Robotics replicating the LLM scaling playbook

Dimension	Legacy paradigm	New paradigm	Mechanism	LLM analogue
Model architecture	VLA (Vision-Language-Action): next-token action prediction	WAM (World-Action Model)	Jointly decodes future video frames and motor commands — imagines how the world evolves, then acts	Upgraded pre-training objective (language modeling → world modeling)
Data scaling	Teleoperation: expensive, slow, hard to scale	Wearables + egocentric human video	Replaces robot-embodiment data with large-scale egocentric human video: establishes neural scaling laws for dexterity	Internet-scale corpus
Simulation / RL environment	Classical physics engines	Neural simulators	Learned world models replace hand-coded physics, accelerating RL rollouts	Synthetic data / self-play
Post-training / fine-tuning	Full fine-tuning of the base model	Bi-level fine-tuning	Lightweight RL policy sits atop a VLA/WAM base, making real-time sub-millimeter corrections without full-model fine-tuning	RLHF / LoRA-style lightweight post-training

Source: Elsewhere News, Nomura research

GPT-3-style few -shot signals are real but remain lab-bound

The robot version of GPT-3's big step, learning a new task from a short example instead of retraining, has started to appear. We check three cases: 1) Generalist AI (unlisted) said on 20 August that its GEN-1.5 model could learn a task from a single 3-12 second demonstration. Without any extra training it succeeded 59% of the time across ten tests, rising to 83% after a short top-up on five minutes of data; 2) Skild AI (unlisted) launched S1 on 31 August and said robots could learn tasks of up to ten minutes, such as making coffee, by watching just a video of people doing the same, but it did not give success rates; and 3) Robbyant (unlisted), Ant Group's (unlisted) robotics unit, said in an August paper that its Zero-WAM (world action model) succeeded 47.0% of the time on seven tasks it had never seen, in simulation, 29.5 points better than the best comparable model. This fits our industry checks that teaching a robot a new task now takes one or two demonstrations rather than dozens or hundreds. However, all three results are from the companies, and all come from short or simulated tasks. Skild's chief executive, Deepak Pathak, said S1 is not quite ready for homes and that robotics has not yet had its ChatGPT moment. So GPT-3-style learning has arrived in principle but is not yet reliable, which is why we still place robots before or at GPT-2.

Three model families split the robot brain: VLA, world model, and RL

Three types of model are competing to become the robot's brain, and each is at a different stage. A VLA (vision-language-action) model looks at the scene, takes an instruction and outputs the desired movements, usually a path that a lower-level controller could follow. A world model that predicts what will happen after the robot acts. Following Feifei Li's (co-founder of World Labs) framework, it splits the task into three parts: 1) a renderer that produces images (where video generators such as Sora sit); 2) a simulator that predicts physical change; and 3) a planner that decides what to do next, which is the role VLA plays. The third approach, RL in simulation, trains a robot's movements by trial and error inside a physics simulator. Unitree's walking robots are the best-known example, and skills trained only in simulation already work on real machines. The NVIDIA scientist argues that early VLA models such as π0 and GR00T spend most of their

capacity on language, so they know what things are but not how things move. His WAM uses video generation as the physics engine instead, because predicting video frame by frame teaches gravity, buoyancy and light. It predicts what happens next and what the robot should do in one step, so it can handle new scenes without extra training. The Dream Dojo (generalizable foundation world model) work goes further, replacing the physics engine with a video model so robots can practise in many simulated worlds at once. On the robot itself, the smallest models win: on-board VLA models have 2-7B parameters and need about CNY30,000-40,000 of chips and storage. World models are much heavier. Our industry survey indicates World Labs' Atlas at 100-200B parameters, and Robotera (unlisted) said world models that generate full video need 3-10x more computing power to run, versus 30-100% more for leaner versions that work on compressed internal representations. So through 2027F we expect world models to run in the cloud, training robots and generating data, while compact VLA and RL models run on the robot.

LLM model labs hold the compute and talent a breakthrough needs

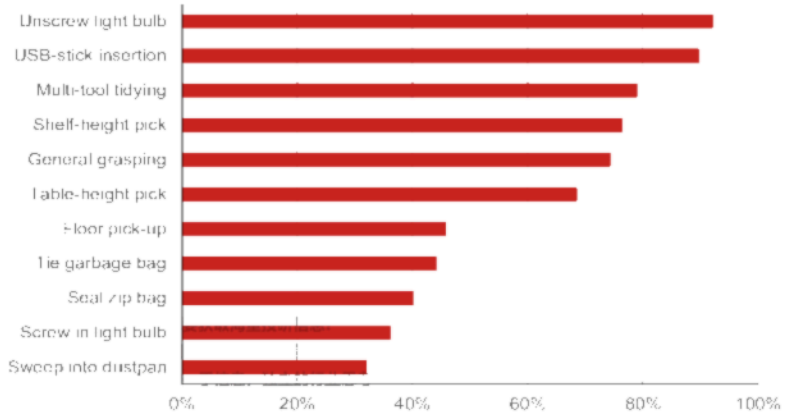
We expect the breakthrough to come from two places, leading AI labs that build foundation models and Tesla, rather than from robot-brain start-ups. Computing power is the clearest gap. Our industry checks indicate US engineers each have more than ten times the computing power of their Chinese peers, and that ByteDance (unlisted) has the most computing power in China for robotics. We expect OpenAI to release a model built for physical AI in December-January, and Google (GOOGL US, Not rated) and other labs to publish results within about three months (Google's Gemini Robotics 2, released on 30 July 2026, completed long, multi-step tasks up to 91.3% of the time, our industry survey shows about a year ahead of Chinese rivals in whole-body control). Wang Xingxing, founder of Unitree, agrees that frontier models will drive the leap. According to Wang, next year's GPT-7 is likely to break through embodied AI's current bottleneck, and that Unitree has tested GPT-6 on its robots with very good results.

Based on our industry survey, leading integrated player could be the second source of a breakthrough, as integrated players share some of the labs' strengths, namely computing power and a closed data loop, and adds something they lack: their own fleet of robots in use. Our industry survey suggests Tesla runs a full loop, from simulation and world models through sorting data, a data platform and factory testing, with its sensors and data formats largely settled. In addition, Tesla robots complete an unsupervised factory routine, from checking screens and fetching tools to fixed-point tasks and charging themselves, about 97% of the time, based on our industry survey. Further, Tesla trains on road video from millions of its cars as well as on robot data and simulation, with a general world model at the centre. We see the labs and Tesla as complementary: the labs bring general models and talent, while Tesla brings a closed loop and a fleet. Either way, the winners will be those with computing power, a data loop and robots in the field, which likely puts Chinese robot-brain start-ups at a disadvantage.

Five operating metrics define when robot brains reach a ChatGPT moment

We propose five tests to determine whether robotics has reached its ChatGPT moment, and robots fall well short on all of them today, according to our industry survey. First, how long can a robot work before a human has to step in: the best is still a few hours, while factories need a week of unattended running. Second, fine handwork: Gemini Robotics 2, running on an Apollo 2 robot with a Sharpa (unlisted) Wave hand, tied a garbage bag 44% of the time and swept into a dustpan 32% of the time, according to our industry survey. Third, how quickly a model adapts to a new robot body: Google's takes about seven to eight hours, while the faster Chinese players need about two weeks. Fourth, how many demonstrations a new task takes: in labs this is down to one or two, based on our industry survey. Fifth, accuracy on the factory floor: our industry check puts it at about 60% today, against the 98% or more that industry requires, and we expect three to five years to close the gap. In our view, the first and fifth tests matter the most, because they decide whether customers pay for robots out of operating budgets rather than from trial budgets.

Fig. 2: Gemini Robotics: task success rate



Source: Industry survey, Nomura research

Data: where robots get their training data

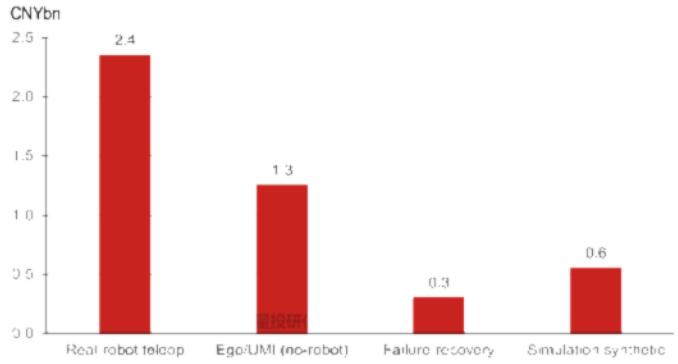
Robots have no internet to learn from, so data is also critical

Data is what holds robot brains back. Language models learned from the text of the whole internet, but robots have nothing comparable. Teleoperation, where a person remotely steers a robot to record training data, is expensive and slow, and each setup can only record a certain volume. Our industry checks suggest he rough threshold for "emergence moment" is about 1mn hours of data, roughly a human lifetime, against only about 100,000 hours of usable data collected so far in one consistent format. In addition, leading companies have only several hundred thousand hours of good-quality data, and Tesla's Optimus has just passed 0.5mn hours, based on our industry checks. Further, our industry survey shows that training needs 10mn hours to get started and over 100mn hours in the long run, against an industry stock in the low millions of hours. We think these estimates differ because they count different things, from strictly usable data in one format to every hour recorded. Much of what is collected is wasted: at Unitree only about a quarter of recorded time was usable, one usable hour for every four recorded. At this early stage, data quality matters. The real question is in which way can collecting data scale up usable hours the fastest, and at what cost.

Four ways to collect robot data: cheaper per hour versus reusable across robots

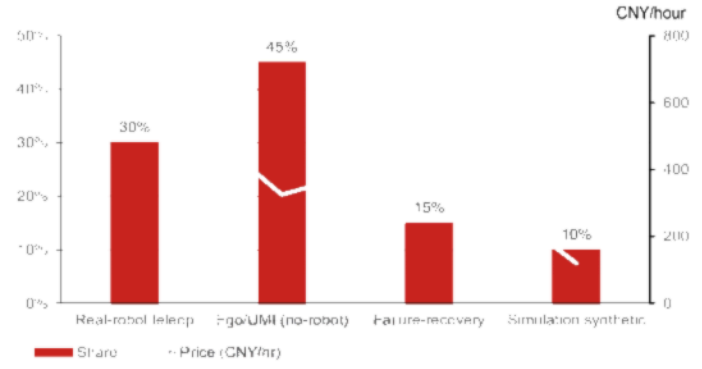
There are four main ways to collect robot data, and the choice comes down to cost per hour versus whether the data can be reused on other robots. Our 19 August initiation on Unitree (*Initiating coverage on China's embodied-AI leader*) put China's 2026F data-collection market at about CNY4-5bn. Real-robot teleoperation would account the most at about CNY2.4bn, followed by Ego video and UMI at CNY1.3bn, simulation at CNY0.6bn and failure-recovery data at CNY0.3bn. By hours collected, Ego and UMI made up about 45%, teleoperation 30%, failure recovery 15% and simulation 10%. These figures reflect the market in mid-2026, and we expect prices to fall within 12-24 months as more than 20 city-level data factories and overseas crowdsourcing ramp up. Real-robot teleoperation costs about CNY700 per hour, or CNY500-1,000 in our industry survey. But the data only works for one particular robot design, so switching from legs to wheels will render it ineffective. UMI (universal manipulation interface), handheld grippers that people use to record tasks, costs CNY100-300 per hour, but it gets in the way of people's normal work, so orders remain small and custom. Ego (first-person) video of people costs about CNY300-350 per hour, putting it in the same band as UMI, and because it is camera footage, it stays useful across robot generations. Simulation is the cheapest at about CNY120 per hour, but it makes up no more than 20% of data for real-world perception tasks.

Fig. 3: Data collection market size



Source: Company data, Nomura estimates

Fig. 4: Share and ASP in different data collection types



Source: Company data, Nomura estimates

Scaling law extends to video, favouring labs with the most computing power

The industry is settling on videos of people, filmed from their own point of view, as the main training data. This ties the data question back to who would win the race for robot brains. Our industry checks suggest a typical mix of 60% Ego video, 30% UMI and 10% real-robot data, with collection kits settling on two head cameras plus wrist cameras. NVIDIA's EgoScale follows the same principle. It first trains on more than 21,000 hours of everyday first-person video of people, then fine-tunes on a small amount of teleoperation and motion-capture data. For hand manipulation, more training compute brings steady and predictable improvement, the same scaling law that has driven LLMs in the past. Our checks and analysis show that one company is building a platform that creates training data from 3D and 4D scans of real scenes (3DGS/4DGS, 3D/4D Gaussian splatting), so the company relies less on outsourced collection. Chinese firms have different practices: Robotera (unlisted) put the industry average at 60% simulation, 25% real-robot and 15% video. If video of people becomes the main source, the scaling law that has driven LLMs will apply to that video too, and whoever has the best data pipeline and the most computing power will gain most. That supports the case for big AI labs, and it limits the long-term value of teleoperation data tied to one robot design.

One learned model of the robot's body cuts training compute by over 1,000x

How robots are trained matters as much as what data they see, and a new paper (*SuperMimic: Scaling Humanoid Whole-Body Control with One Differentiable World Model*) suggests learned world models can become the place where robots are trained, not just a source of data. SuperMimic is a CoRL 2026 paper co-authored by Pieter Abbeel and Koushil Sreenath. It notes that today, whole-body control for humanoids that copy large libraries of human movement is trained in one of two ways. One is to train a separate specialist for each movement and then merge them, which loses accuracy. The other is to train a single policy with a huge amount of computing power. SuperMimic instead trains one model that learns how the robot's body moves across all motions, using records from a standard physics simulator. It trains the controller inside that model at the same time, so the controller learns directly from short predicted sequences instead of slow trial and error. The authors report that after 3 hours on one GPU it beat a leading method that has undergone 7 days of training on 128 GPUs, cutting compute by more than 1,000x. On their test set it failed 25 times across 129,514 movements, against 179 for the older method, and reached zero failures after 9.8 hours on two GPUs. The controller worked on Unitree's G1 humanoid for a wide range of movements, including acrobatics, and the learned model stayed accurate as motion data grew from 1 to 264 hours. We learnt three lessons from this. 1) Train robots inside a learned model rather than a raw simulator. 2) Train one controller on more motion data rather than many task-specific ones. 3) Expect computing power to matter less for whole-body movement. This fits our view of world models as the training ground for robots, and it makes data, and robots deployed in the field, even more important.



Fig. 5: Failed motions by group

Motion group	Motions	SONIC failures	SuperMimic 3h (1×H100) failures	SuperMimic 9.8h (2×H100) failures	SONIC failure rate
Side rolls	22	22	2	0	100.0%
Flips	44	44	1	0	100.0%
Fire-in-the-hole turns	20	3	3	0	15.0%
Crossed legs	10	9	0	0	90.0%
Lying, resting, standing up	99	28	8	0	28.3%
Broad jumps	24	15	0	0	62.5%
Dance routine	8	8	0	0	100.0%
Others	129,287	50	11	0	0.0%
Total	**129,514**	**179**	**25**	**0**	**0.1%**

Source: Conference on Robot Learning, Nomura research

The real ChatGPT moment needs data from robots at work

Video of people can cover the first stage of training. In our view, however, the real ChatGPT moment is when a positive-feedback-loop starts turning: robots at work generate data, the data makes the models better, and better models get deployed more. That loop only works once enough robots are in the field. Tesla's plan shows the scale needed. Based on our industry checks, Tesla aims to collect 1mn usable hours by end-2026E from about 10,000 robots training in its own facilities. In addition, Optimus's data cost fell about 18.1% from 2Q26 to 3Q26. In China, deployment is far smaller: the seven largest express couriers are likely to buy fewer than 500 humanoids in 2026F and about 1,000 in 2027F, based on our checks and analysis. We expect the positive feedback loop to start first where one company owns both the robots and the model.

OpenAI is likely the biggest data buyer, but volume estimates vary

Demand for robot data is concentrated in a few US labs, while Chinese suppliers are fragmented and limited by compliance rules. On industry checks suggest OpenAI bought about 10mn hours of data between September 2025 and August-September 2026, about 70% of US supply, with DeepMind taking about 20%, or roughly a third of OpenAI's volume. Other estimates vary 100-fold. In China, ByteDance is the only steady bulk buyer, taking about 80% of the output of three suppliers, mainly for its video AI models. On the supply side, Jianzhi (unlisted) has collected about 1mn hours with more than 1,000 devices, and ATARA (unlisted) has collected 1-2mn hours for OpenAI and DeepMind. Our industry checks show independent data collectors have about three years to continue data collection before leading robot and AI companies start to gather data themselves. As Chinese data cannot be sold overseas and AI companies reply on own self-collected data, we expect the number of data suppliers to decline by end-2027F.

Hardware: implications for components

Three scenarios decide whether robot hardware goes the Android way or Tesla way

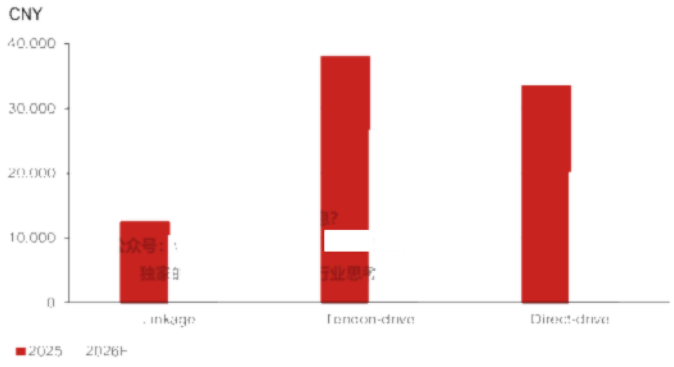
To assess hardware impact, we look at three scenarios, which can overlap. In Scenario A, big AI labs crack physical AI, and robots go the way of Android phones: the brain becomes the platform and bodies become standard hardware. We expect robot designs to converge once the software approach is settled. In addition, we note DeepSeek and ByteDance are already asking suppliers to build custom hardware. We believe AI model companies are more likely to buy third-party robots than build their own. In Scenario B, Tesla's do-it-all model wins, and the value goes to the main suppliers that assemble parts for Tesla and to makers of hard-to-source parts. In Scenario C, standardisation which would help makers of parts that fit many different robots, starting with dexterous hands and followed by other standard core parts. We expect hand shipment volumes to rise and prices to fall as better AI lets robots switch from simple grippers to five-finger hands.

Dexterous hands benefit first as better AI replaces grippers

In our view, dexterous hands will be the first to benefit from Scenario A and C, because better AI removes the data shortage that has kept robots from using complicated grippers. Many robots use grippers only because there was too little data to train hands. Stronger foundation models can carry skills learned on two- or three-finger grippers over to five-finger hands, enabling volume lifting and lower prices. Our industry checks suggest 2026E hand shipment growth at 80-120% y-y, mainly from high-DOF (degrees of

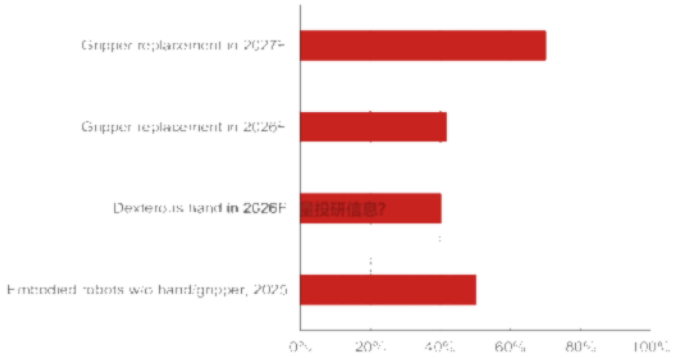
freedom) hands with more moving joints, and that 30-50% of factory grippers would be replaced this year. As well, start-ups selling hand data for specific tasks to leading AI labs have been earning more than expected.

Fig. 6: Average ASP for different technology route



Source: Company data, Nomura estimates

Fig. 7: Dexterous hand gripper penetration in humanoid



Source: Company data, Nomura estimates

Robot bodies: building every unit the same way or in a similar way

What is happening in hands is likely to repeat across the whole robot. We expect an "iPhone moment", when robot designs settle into a few standard forms. Thereafter, the difficult part is building every unit the same way, and to keep the data collected by different robots consistent. Once AI models are settled, robot hardware will settle into far fewer designs, based on our estimates. The main difficulty then becomes production yield and making sure the 20-plus joints on every robot behave the same way. On pricing, we see a reasonable cost at mass production of CNY150,000-200,000 for a 1.7m robot and CNY30,000-40,000 for a 1.3m robot. For one robot, actuators, the motor-and-gearbox units that move each joint, make up most of the parts cost, about 50-75%. We think the winners will be makers that improve their products quickly and keep costs tight, such as Unitree, carmakers, and key component suppliers. For Unitree, we believe its top priority is to cut costs further without hurting quality. As its volumes grow, we expect Unitree to increase its outsourcing ratio, both to lower costs and to secure quality.

Fig. 8: Unitree: in-house vs third-party procurement

	Humanoid	Quadruped	Details
BOM covered by in-house design	80-90%	c.90%	Motors, encoders, driver boards, main board, battery pack (In-house integration), power distribution unit, DC-DC, other power control, planetary reducer, joint structural parts, body structural parts, cross-roller bearing
BOM covered by in-house designed and manufacturing	c.50%	55-60%	Motors, encoders, driver boards, main board, battery pack (In-house integration), PDU 500 + DC-DC 480 + other power control 200
Third-party modules	10-15%	5-10%	3D LiDAR - DJI Livox MID360, RealSense D435i, Orbbec, Sharpa
US content, % of BOM Humanoid: G1 basic edition; Quadruped: Go2 Air/Pro	2-8%	1-5%	TI / ADI / ON Semi, RealSense
US content, % of BOM Humanoid: G1 EDU	10-20%	5-15%	Nvidia, TI / ADI / ON Semi, RealSense

Source: Company data, Nomura estimates