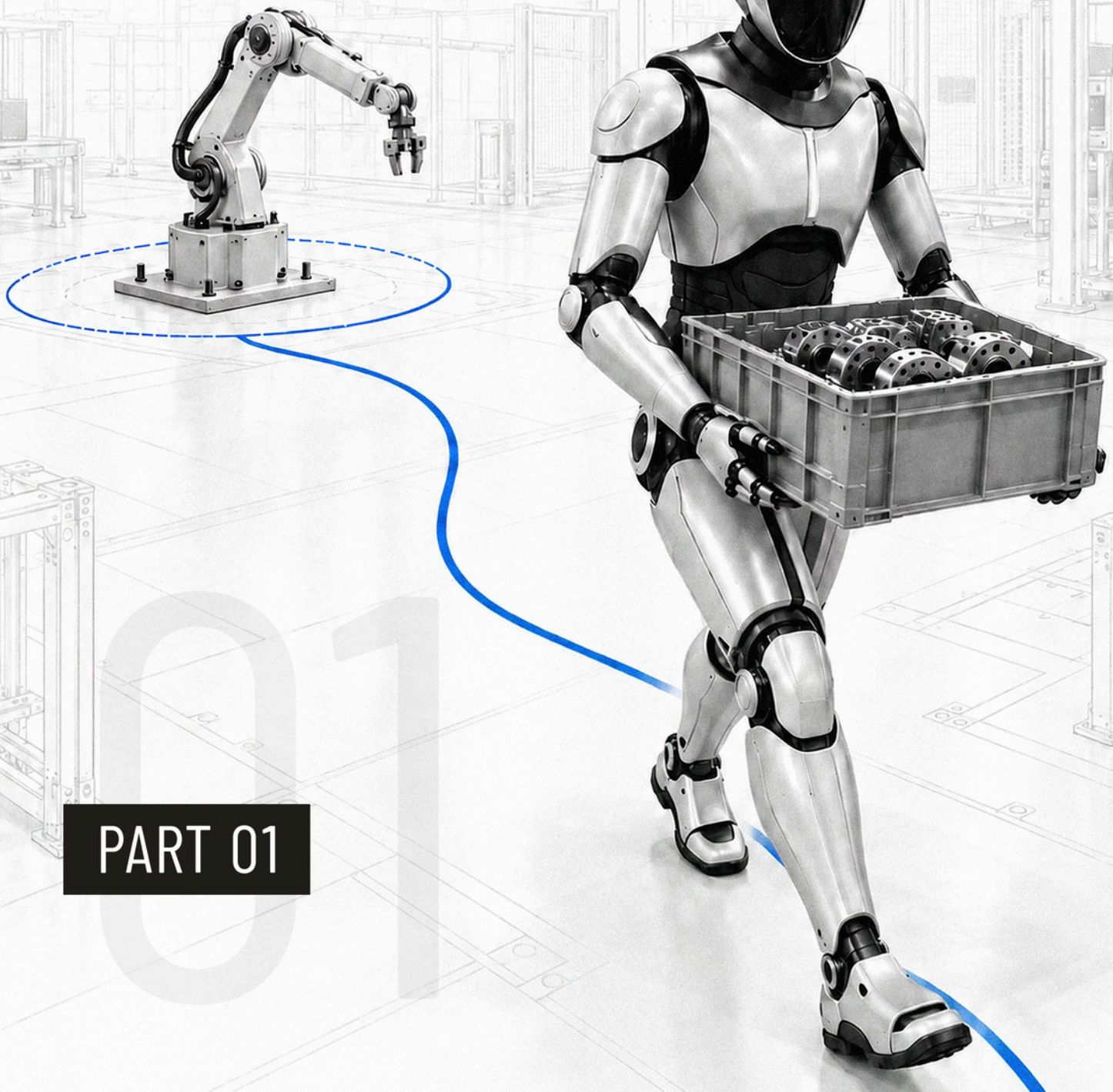
DSCI
PROMOTING DATA PROTECTION

THE HUMANOID HORIZON



PART 01

1. Context

For about sixty years, industrial robots have been excellent at doing the same task the same way, inside a safety cage, with people kept at a distance. That model built the modern car plant and it still earns its place. What is changing is the arrival of machines that are meant to leave the cage. Humanoid robots, which pair a human shaped body with learned models that see, listen, and decide, are moving out of laboratories into early commercial work. Industry trackers put global shipments at roughly 18,000 units in 2025, and the first half of 2026 is estimated to have passed 19,000, with Chinese vendors supplying the large majority. These are indicative figures and the counting methods differ between trackers, but the direction is clear.

Four developments arrived together and made this possible. Multimodal foundation models gave robots a way to be instructed in plain language instead of being programmed line by line. Edge compute became powerful enough to run those models on the robot itself. Brushless motors and precision gearing became lighter, stronger and cheaper, largely because drones and electric vehicles paid for the research. And GPU based simulation made it practical to rehearse millions of hours of practice before a robot touches a real shelf.

THE CORE SHIFT

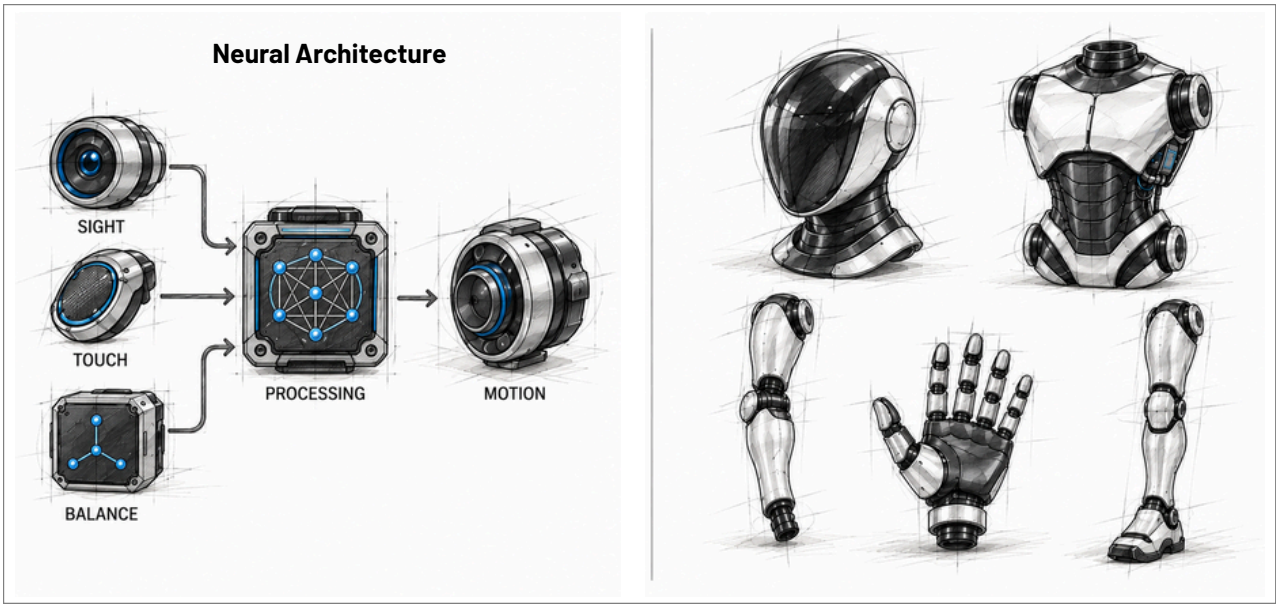
A modern humanoid is best understood as a mobile embodied intelligence platform: general purpose neural networks controlling a high dimensional physical body in real time, in spaces that were designed for people.

This paper is the first in a series. It explains how humanoids are built, how they differ from the robots we already know, who is building them, where early deployments are happening and why the subject matters for India's industry, security and policy. We have tried to keep the engineering honest and the language plain, and where a number is uncertain, we say so.

1.1 Modern Humanoid

A modern humanoid combines two systems that have to work as one. The first is a body: a torso, a head, two arms with hands, and either two legs or a wheeled base in place of legs. The second is a neural architecture that can process sight, touch and balance continuously and turn them into movement. A traditional machine is tuned for a single repetitive job. A humanoid is meant to be versatile. It uses vision language action models to interpret instructions given in natural language,

builds a picture of its surroundings from several cameras and depth sensors, and adjusts its movement to whatever is in front of it.



The human form is a means rather than a goal. Factories, warehouses, hospitals and homes were built around human bodies, with doors, stairs, shelves and hand tools sized for us. A robot that fits that world can be placed in an existing building without redesigning it, which is the economic case often called brownfield reuse. The counter argument deserves equal airtime. A purpose built machine will usually beat a humanoid at any single task. The humanoid wager is that one adaptable machine, helped by better AI and falling hardware costs, will eventually beat many specialized machines wherever the work keeps changing.

Humanoid is not a chatbot with legs: a language model can understand a request, but safe and reliable physical action is a separate and much harder problem. And it is not a remote-controlled teleoperated device. Many viral demonstrations rely on teleoperation, which is a legitimate way to gather training data but should never be presented as autonomy.

2. Global Landscape and the Indian Ecosystem

The race is being shaped by ageing populations, difficulty in staffing physically demanding shifts, and a growing concern about who controls the underlying technology. Three regions anchor it, North America, China and Europe, and India is taking its first steps toward a place in it.

2.1 North America, China and Europe

North America leads on frontier foundation models, end to end learned policies and the sheer volume of venture funding. Tesla, Figure AI, Apptronik, Agility Robotics and Boston Dynamics are the names most often cited, and most of them train on very large GPU clusters. China leads on the speed of hardware iteration, supply chain depth, low cost precision gearing and manufacturing scale, through companies such as Unitree, AgiBot and UBTECH. Europe's strengths lie in collaborative safety practice, tactile sensing and fine motor control, and 1X Technologies, which began in Norway, is among the best known examples.

2.2 India's position and capabilities

India is a credible regional contender in mobile robotics and warehouse automation. Addverb Technologies leads domestic development, building autonomous mobile platforms and articulated manipulators, and it has extended into humanoid platforms. Younger companies such as Perceptyne and iHUB Robotics have first products in the market, are active in dynamic balance algorithms and spatial computer vision. There is, however, seems to be no humanoid specific policy yet, and no verified count of Indian humanoid patents.

The structural weakness is the supply chain. High torque frameless brushless motors, harmonic drive reducers, high efficiency battery cells and high performance edge compute chips are still largely imported from Japan, Germany, China and the United States. India's real leverage today sits elsewhere: in software development, simulation, systems engineering and synthetic data generation, all areas where the country already has depth.

Region	Current State
United States	Frontier AI foundation models and Sim2Real frameworks, high manufacturing cost
China	Rapid prototyping, low cost actuation and mass production, Limited access to advanced chips
Europe	Precise manipulation, safety standards and actuator efficiency Slower deployment of venture capital
India	**Software stack, simulation and systems integration, Heavy import dependence for precision hardware**

Table 1: Comparative account of regional position and capabilities

3. Industrial Automation Giants

The large industrial automation groups have mostly chosen not to build bipedal hardware from scratch. They are concentrating instead on software integration, digital twin platforms and industrial edge infrastructure, which makes sense, because they already own the factory floors that humanoids will have to walk onto.

Siemens

is building software pipelines that connect industrial digital twins to embodied AI platforms. Coupled with GPU simulation, this lets developers train humanoid policies inside a digital copy of a factory before any hardware arrives on site.

Honeywell

is drawing on its warehouse automation business to bring intelligent mobile manipulators and humanoid platforms into supply chain facilities. The emphasis is on linking fleet orchestration engines with warehouse management systems, so that humanoids can work alongside the conveyors already in place.

Bosch

has partnered with the London-based robotics startup Humanoid to manufacture its HMND01 wheeled humanoid robot at an industrial scale in Europe.

Mitsubishi Electric

works on precise manipulation and balance control through its research laboratories (MERL). Its efforts point toward hybrid actuation, combining industrial reliability with the compliance needed around people.

Rockwell Automation

relies on its partner network. By linking programmable logic controllers to edge AI nodes, it aims to let humanoids talk directly to existing plant infrastructure and safety relay systems.

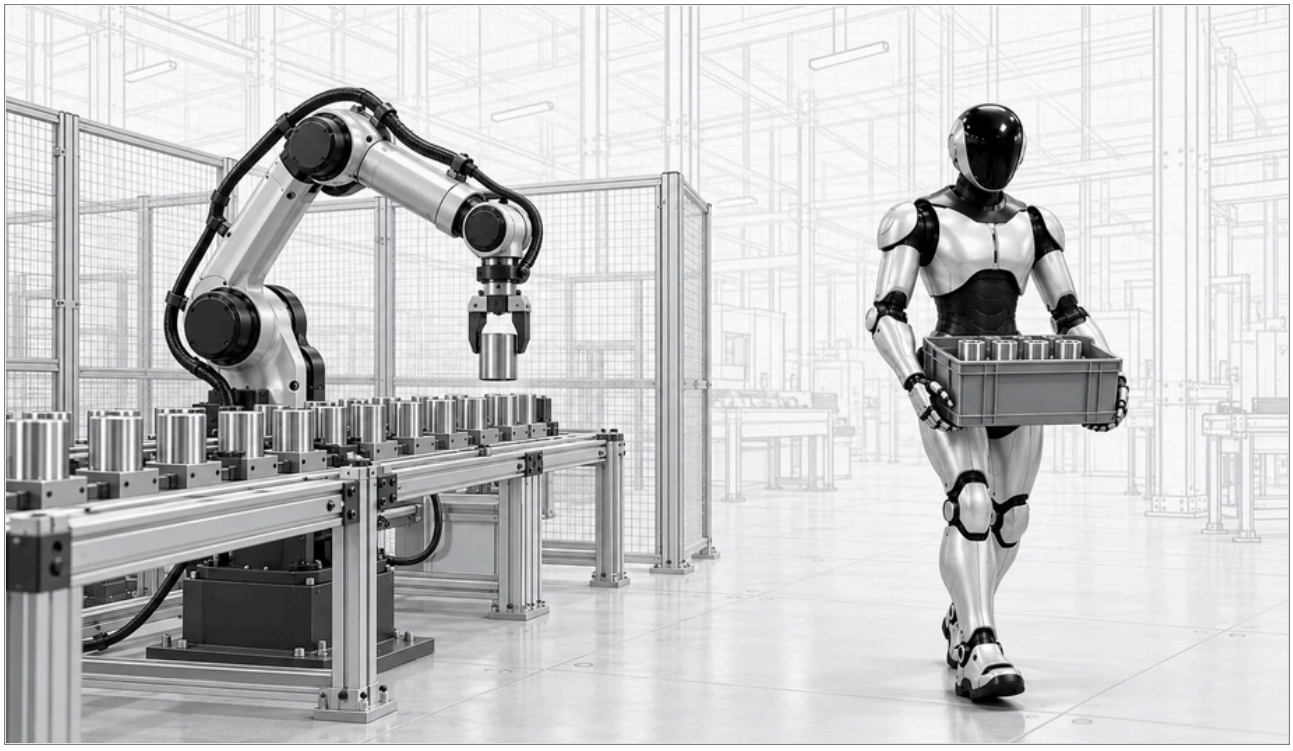
The common thread is integration. A humanoid that cannot speak to the plant's control system, scheduling software and safety circuits is a demonstration, not a worker.

4. Humanoids vs Traditional Industrial Robots

The two categories differ in body, in how they are programmed, in where they can work and in how they are kept safe. Table 2 sets out the main contrasts.

Feature	Traditional industrial robot	Embodied humanoid robot
Morphology	Fixed base with a six or seven degree of freedom arm	Bipedal or wheeled body with 20 to 50 or more degrees of freedom
Programming	Explicit scripts and ladder logic	Learned neural policies, guided by natural language
Environment	Structured, caged and highly predictable	Unstructured, changing and shared with people
Generalization	Optimized for a single task	General purpose, able to switch between tasks
Safety model	Physical isolation and perimeter cutoffs	Active compliance and vision based safety networks

Table 2: Contrast between Humanoids and Traditional Industrial Robots



Traditional arms depend on high stiffness and absolute position accuracy inside an isolated cell. They repeat a path that never changes, and they do it superbly. Humanoids make a different trade. They give up some of that rigidity in return for compliance, richer perception and constant, active balancing. Instead of replaying a fixed path, a humanoid keeps adjusting its movement policy to match the physical conditions of the moment, which is why it can share an aisle with a person.

Form factor matters here too. Full size bipeds attract the headlines, but wheeled humanoids, with a human like upper body on a mobile base, are a growing share of early industrial deployments because they are simpler, more stable and efficient on flat factory floors. Small and medium bipeds are found mostly in research, education and entertainment. Quadrupeds, autonomous mobile robots and conventional cobots are valuable machines, but they are not humanoids.

5. Key Components: Hardware and Software Codesign

Building a humanoid is an exercise in codesign. The mechanical engineer, the embedded software developer, the sensor specialist and the machine learning researcher cannot work one after another, because a decision in one discipline constrains all the others. A heavier hand changes the balance model. A slower camera limits how bold the walking policy can be. The system also has to live with very different clocks, from motor current loops that run in tens of microseconds up to planning that unfolds over several **Architectural** hours.

Many builders describe the top two layers as a pair of brains. A slower reasoning model, often called System 2, interprets the scene and the instruction. A faster visuomotor policy, System 1, turns those decisions into smooth movement. We use that framing throughout this paper. Figure 1 shows the complete logical architecture.

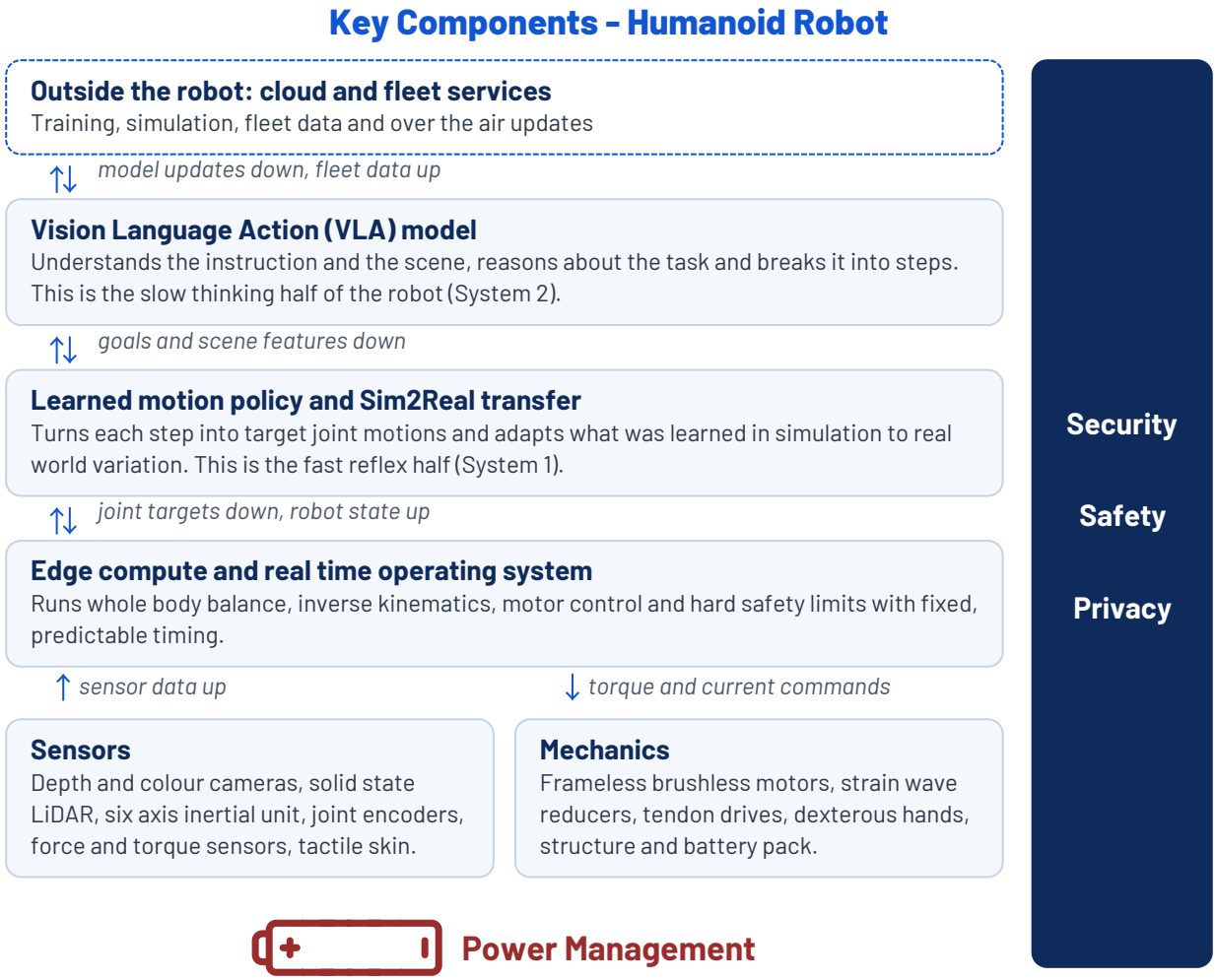


Figure 1: Key components of a humanoid robot

5.1 Vision Language Action model

A **Vision-Language-Action (VLA) model** is a multimodal foundation neural network that unifies visual perception, natural language understanding, and physical motor control into a single end-to-end framework. Instead of breaking a robot's software stack into separate hard-coded blocks for object detection, path planning, and joint actuation, a VLA takes raw camera images and text instructions as input and directly generates executable robot commands.

Core Architecture Components

Vision Encoder: Converts raw pixels from RGB camera feeds into continuous visual feature vectors or tokens using pre-trained vision transformer backbones like SigLIP or DINOv2

Language Backbone: Processes natural language instructions (and sometimes internal robot states) using large language model architectures (such as Gemma or Llama) to interpret high-level semantic intent.

Multimodal Fusion Layer: Combines visual, textual, and proprioceptive (joint position, velocity) tokens through cross-modal attention mechanisms inside a transformer network.

Action Decoder (Action Head): Maps the unified latent representation directly into continuous motor values (joint angles, velocities) or discrete action tokens

Action Representation Strategies

- **Discrete Token Output:** Treats physical movements like words by quantizing continuous motion dimensions into uniform bins (e.g., 256 bins), allowing an autoregressive transformer to generate action strings just like text. Models like OpenVLA utilize this approach.
- **Continuous Output via Diffusion or Flow Matching:** Bypasses vocabulary discretization to output smooth, high-frequency joint trajectories (often up to 50–200 Hz) using denoising or vector-field prediction networks.

Training and Scaling Paradigm

- **Internet-Scale Pretraining:** Models inherit broad physical reasoning, object recognition, and linguistic common sense from massive internet image-text datasets before ever touching a robot.
- **Imitation Learning Fine-Tuning:** The combined network is fine-tuned via behavior cloning on large robotics trajectory datasets (such as the Open X-Embodiment collection) mapping observations to action sequences

5.2 Motion policy and Sim2Real transfer

The layer is a learned neural network policy, and it runs on the humanoid's onboard compute hardware which converts each step into movement. It evaluates a pretrained policy, generates trajectory goals and adapts to the small physical differences between a simulator and a real floor, such as a slippery patch or a package heavier than expected. This is the layer where simulation training pays off, and it is where the quality of the Sim2Real transfer decides whether a robot looks graceful or clumsy.

5.3 Edge compute and the real time operating system

Below the learned layers, behavior becomes deterministic. Embedded compute runs whole body balance models, inverse kinematics and the emergency limits on torque, speed and position, on a real time operating system whose timing can be guaranteed. A neural network may suggest a motion, but this layer decides whether the motion is safe to execute. The industry has largely converged on a clean separation here, so that a failure or a slow response in the AI layers can never starve the control loop.

5.4 Sensors

The sensory layer gives the robot its awareness of itself and its surroundings. Proprioceptive encoders report the angle of each joint, a six axis inertial unit reports body motion, depth cameras and solid state LiDAR map the space, and tactile electronic skin and force sensors tell the hands how firmly they are gripping. Sensor data flows upward through Layer 3 to the learned layers, and a fast subset flows straight into the balance controller.

5.5 Mechanics

The bottom layer is the body itself: motors, reducers, tendon drives, dexterous hands, the structure and the battery pack. Together they produces movement and carries the machine. The mechanical layer demands exceptional power density. Modern humanoids use custom frameless brushless motors paired with strain wave harmonic gears, a combination that delivers high instantaneous torque in a package small enough to fit inside human limb proportions.

5.6 Power Management

Every layer of a humanoid depends on one shared energy source, so power deserves to be treated as an architectural concern and not only as a battery pack in the legs or torso. Most platforms today use lithium-ion packs, and the energy they hold decides how long a robot can work, which is why heavy-duty operation is still limited to roughly two to four hours before recharging or swapping. Around the cells sits a battery management system that monitors voltage, temperature and state of charge, balances the cells and cuts power when something goes wrong. Power electronics then convert and distribute that energy at the voltages each subsystem needs, from high current motor drives to the lower power compute modules and sensors. Because motors and onboard processors both turn much of that energy into heat, thermal management is part of the same design, since a robot that overheats must slow down or stop long before its battery is empty. The power path also matters for safety, because the hardware emergency stop must be able to isolate the motors regardless of what the software is doing, and a battery fault near people is a physical hazard. Many industrial deployments therefore plan around hot swappable or autonomously exchanged packs, and the planning layers increasingly take the remaining charge into account, for example by returning to a swap station before the battery runs low. In short, the battery sets the length of the working shift, and the management around it determines whether the robot can work safely for that whole time.

6. The Frontier AI Stack: Embodied Foundation Models

The intelligence of a modern humanoid comes from vision language action models trained on very large multimodal datasets. A vision language action model takes a stream of pixels together with a natural language command and produces motor actions as its output.

6.1 Multimodal policy architectures

Models treat robot actions as tokens in the same way a language model treats words. Visual and tactile inputs are then mapped to physical coordinates by the same machinery that predicts the next word. Newer designs, including pi zero and NVIDIA GR00T, add components that generate smooth, continuous motion. Open models such as OpenVLA and GR00T give builders something they can inspect, adapt and host locally.

These models can handle objects and instructions they were not explicitly trained on, to a useful degree. Reliability on genuinely novel tasks is still limited, so claims of zero shot generalization should be read with some care and tested on the actual job.

SIMULATION TO REALITY TRANSFER

GPU accelerated physics engines such as NVIDIA Isaac Gym and Isaac Lab, MuJoCo and Genesis let thousands of simulated robots practice at once, accumulating years of experience in hours. Domain randomization, which varies friction, mass, lighting and sensor noise during training, helps the resulting policy survive the differences between simulation and the real world.

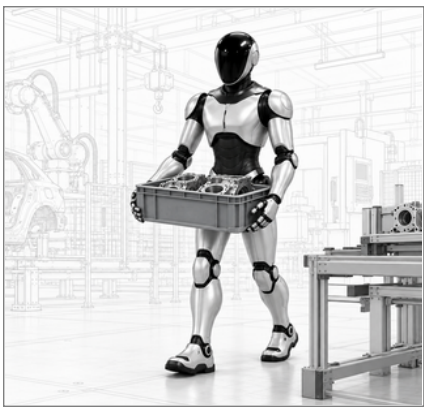
6.2 Fleet learning and data pipelines

A deployed fleet is also a data network. Demonstrations collected by teleoperation, together with the everyday experience of every robot in the field, are processed in central compute centers to fine tune policy networks, and the improved versions are sent back to the machines over the air. This is the flywheel that makes the economics work, because every robot makes the next one better.

7. Key Commercial and Industrial Verticals

Early commercial deployment is concentrated where labor shortages, physical fatigue or hazards create real operational risk, and where the tasks are repetitive enough to be learned but variable enough to defeat conventional automation.





7.1 Automotive and heavy manufacturing

Automotive assembly is the primary target. Humanoids are being tried at tasks such as placing trim panels, installing wiring harnesses and moving material between fixed stations. Pilots have been reported at BMW's Spartanburg plant with Figure and at a Hyundai plant in Georgia with Boston Dynamics' Atlas, and Chinese carmakers are running their own trials. Most of this work is still limited to material handling and parts transfer rather than delicate assembly.

7.2 Logistics and warehousing

Warehouse work involves dynamic sorting, unloading container trucks and picking packages that are never quite the same shape twice. Humanoid platforms can move through tight aisles, lift heavy parcels and build fulfilment pallets directly. Agility Robotics' Digit, working on tote movement at customer sites under a robot as a service model, is one of the clearest early examples.

7.3 Other sectors to watch

Hospital logistics, elder care and hazardous inspection are emerging areas. For India, the most relevant near term openings are probably night shifts that are hard to staff, ergonomically harsh tasks in auto component plants and warehouses, and support roles in healthcare.

8. Defense Applications and Geopolitical Significance

Humanoid platforms have genuine dual use potential. The same machine that moves totes in a warehouse can move crates in a depot, which is exactly why governments are paying attention.

8.1 Tactical logistics and hazardous operations

In military logistics, humanoids could carry supplies over rough ground, handle munitions and carry out base maintenance in contaminated environments. Their value is first of all a human one: they can take people out of harm's way during chemical, biological, radiological or nuclear contamination response.

SOVEREIGN TECHNOLOGY INDEPENDENCE

Building indigenous hardware and software capability in embodied intelligence is essential for long term defense resilience. Export controls on advanced chips and certain components have already shown how quickly access to critical technology can change.

8.2 Strategic labor mitigation

Many industrial nations face shrinking workforces that threaten national productivity, and general purpose humanoids could act as a structural buffer, preserving output without depending on foreign manufacturing. India's situation is different, because it has a young population. For India the argument is about raising productivity, filling roles that are hard to staff and building an export capability, and not about replacing jobs.

9. Technical Bottlenecks and Outlook

Despite rapid progress, broad commercialization still faces some stubborn technical problems.

- **Energy density and thermal limits**
  Current lithium battery technology limits high load operation to roughly two to four hours before the robot needs to recharge or swap its battery. Sustained heavy work also pushes motors and compute modules against their thermal limits.
- **Real time tactile processing**
  Covering a hand with a dense array of electronic skin generates a flood of signals. Handling them with low enough latency to react to a slipping object requires dedicated signal processing, or the system becomes a bottleneck.
- **Long horizon reasoning**
  Keeping a multi-step plan on track when something unexpected interrupts it, such as a dropped tote or a person walking across the path, remains an open AI challenge.
- **Reliability and cost**
  Hours between failures, maintenance effort and price per unit will decide whether pilots turn into purchase orders. These are less glamorous than the AI questions, but they are what a plant manager will ask first.

As foundation models improve and hardware supply chains scale, humanoids will move from experimental pilots toward routine operational infrastructure. We expect that path to be uneven, with a long period of narrow, supervised deployments first. The organisations and countries that use this period to build skills, standards and supply chains will be in the strongest position when it ends.

Looking Ahead

Next papers in the series will move from foundations to evidence and practice. This series will present an original baseline search of Indian patents, an analysis of the component supply chain, and a set of case studies. It will propose safety and cyber security testing for humanoids, assess workforce and skilling needs, and examine the dual use and strategic dimensions in more depth, including precedents set by export controls.

References

1. Brohan, A., et al. (2023). RT 2: Vision Language Action Models Transfer Web Knowledge to Robotic Control. arXiv:2307.15818.
2. Driess, D., et al. (2023). PaLM E: An Embodied Multimodal Language Model. International Conference on Machine Learning (ICML).
3. Ahn, M., et al. (2022). Do As I Can, Not As I Say: Grounding Language in Robotic Affordances. Conference on Robot Learning (CoRL).
4. Black, K., et al. (2024). pi zero: A Vision Language Action Flow Model for General Robot Control. Physical Intelligence. arXiv:2410.24164.
5. Kim, M. J., et al. (2024). OpenVLA: An Open Source Vision Language Action Model. arXiv:2406.09246.
6. NVIDIA. (2025). GR00T N1: An Open Foundation Model for Generalist Humanoid Robots. arXiv:2503.14734.
7. Tobin, J., et al. (2017). Domain Randomization for Transferring Deep Neural Networks from Simulation to the Real World. IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). arXiv:1703.06907.
8. Makoviychuk, V., et al. (2021). Isaac Gym: High Performance GPU Based Physics Simulation for Robot Learning. arXiv:2108.10470.
9. Todorov, E., Erez, T., and Tassa, Y. (2012). MuJoCo: A physics engine for model based control. IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS).
10. International Organization for Standardization. ISO 10218: Robots and robotic devices, safety requirements for industrial robots. ISO 13482: Safety requirements for personal care robots.

Author

Prasad Deore
Vice President, DSCI

ABOUT DSCI

Data Security Council of India (DSCI), set up by Nasscom, is a not-for-profit industry body and think tank on data protection, cyber security, and critical technologies in India. DSCI is committed to making cyberspace safe, secure, and privacy-assured, and to advancing critical technologies vital for national security through policy advocacy, thought leadership, capacity building, and outreach initiatives. DSCI engages with the government, regulators, defence, law enforcement agencies, embassies, industry members, and academia to position India as a leader in the global digital ecosystem

Data Security Council of India
4th Floor, Nasscom Campus, Plot No. 7-10, Sector 126, Noida, UP-201303

data-security-council-of-india
dsci.connect
dsci.connect
DSCI_Connect
dscivideo
Security Chips

All Rights Reserved © DSCI 2026