ARTIFICIAL INTELLIGENCE

How Physical AI Is Reshaping Robotics Today—and What Comes Next

By Tilman Buchner, Martin Kleinhans, Daniel Küpper, Jonathan Brown, Max Ludwig, Simon Rees, and Francesco Palmegiano

ARTICLE APRIL 14, 2026 15 MIN READ

The world is buzzing about humanoid robots—and the stakes are measured in billions. Analyst projections for the humanoid robotics market by 2030 range from under 1 million annual units to more than 6 million. This means that tens of billions of dollars in potential annual revenue and substantial capital are already flowing into the space. But there is great uncertainty. If these forecasts materialize, humanoids could fundamentally redefine how we live and work. If they do not, the sector risks becoming one of the largest misallocations of industrial capital in recent years.

The excitement—and the uncertainty—are understandable. Manufacturers face persistent labor shortages, rising costs, and increasing variability in products and processes. Traditional automation has delivered enormous value, but it struggles with changeovers, complex handling tasks, and unstructured environments. Physical AI promises to address these limitations by enabling robots to perceive, adapt, and manipulate beyond rigid, preprogrammed routines.

> Improvements in perception, dexterity, planning, and reasoning are unfolding at different speeds, and highly visible demonstrations can obscure which capabilities are mature and which remain experimental.

The challenge is not whether progress is occurring—it is how to interpret it. Improvements in perception, dexterity, planning, and reasoning are unfolding at different speeds, and highly visible demonstrations can obscure which capabilities are mature and which remain experimental. Without a structured way to distinguish deployable performance from aspirational demos, leaders risk either overcommitting capital too early or waiting too long to capture near-term gains.

To cut through this uncertainty, BCG has developed a five-level framework for physical AI that clarifies what robotic systems can reliably do today, what is nearing industrial readiness, and what remains a longer-term bet. The objective is not to predict the future of humanoids, or physical intelligence in general, but to help decision makers sequence investments with discipline—capturing value where it exists now while positioning for what comes next.

Why Physical AI Is Hard to Interpret

Physical AI refers to the next generation of robotic systems that can perceive and act within the physical world. These systems operate in unstructured or dynamic environments, execute dexterous manipulation tasks comparable to those performed by human hands, and are able to reason about the physical consequences of their actions. They can infer human intent, adapt to unfamiliar situations, and autonomously plan and execute workflows to achieve defined objectives. Crucially, physical AI is hardware-agnostic: the same capabilities can be deployed across a wide range of robotic embodiments—from humanoid platforms and drones to industrial automation systems.

Progress in physical AI is inherently nonlinear. Different capabilities mature at different speeds, shaped by distinct technical constraints. Perception has advanced rapidly with modern computer vision and simulation techniques, while dexterous manipulation and reasoning about physical cause and effect have proven far more difficult to scale. This asymmetry reflects a principle in AI known as Moravec's paradox: tasks that are easy for humans—such as grasping irregular objects or navigating cluttered spaces—are often extremely hard for machines, but tasks that seem cognitively complex can be comparatively easier to automate. As a result, improvements in one capability can create the impression of broader advancement, even when other constraints remain.

This paradox makes the current state of robotics particularly difficult to interpret. Highly visible demonstrations—especially humanoid robots performing coordinated movements—can signal rapid progress, even when foundational capabilities such as dexterity or causal reasoning remain constrained. Humanoid form factors further blur this distinction by conflating appearance with

intelligence. A human-like body suggests general-purpose capability, even when the underlying perception, manipulation, or reasoning systems remain narrow. For decision makers, this dynamic can distort expectations about what current systems can reliably deliver in real operating environments.

> Focus on capabilities rather than embodiment. What matters is not a robot's form factor, but which physical AI capabilities it reliably possesses.

A more useful perspective is to focus on capabilities rather than embodiment. What matters is not a robot's form factor, but which physical AI capabilities it reliably possesses—and how those perform under real-world variability. This capability-based view provides the foundation for assessing maturity, value, and risk across the robotics landscape.

A New Decision Framework to Cut Through the Robotics Hype

To operationalize the capability-first view for decision making, it is helpful to think of physical AI as developing through five distinct levels of maturity. Each level represents a step change in what robots can reliably do, the environments they can operate in, and the types of use cases that become economically viable (see Exhibit 1):

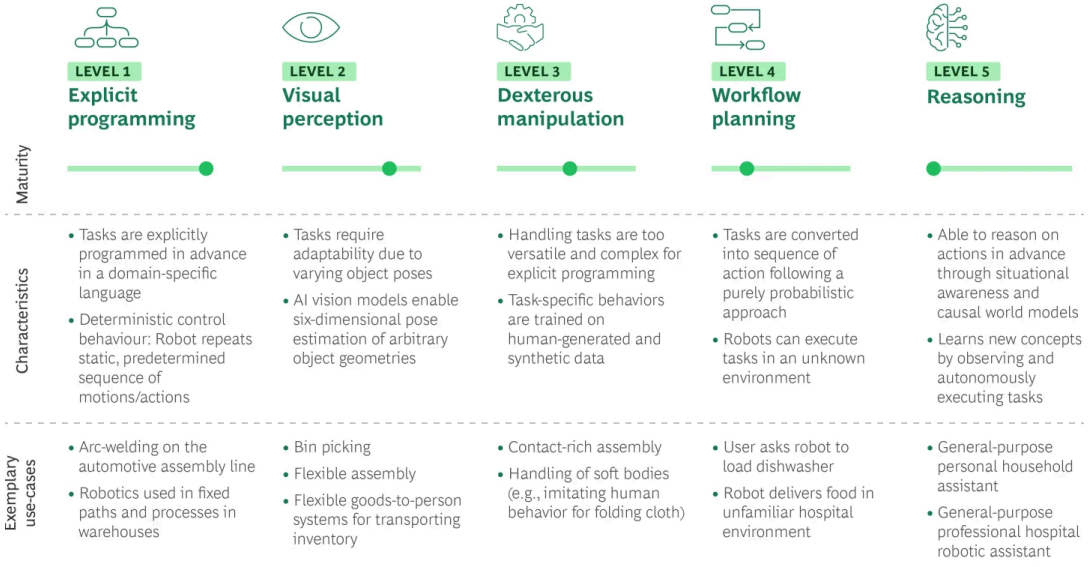
- **Level 1: Explicit Programming**. Robots execute predefined motion sequences with high precision in stable, tightly controlled environments. These systems are mature and widely deployed, but they cannot adapt to variability without manual reprogramming.
- **Level 2: Visual Perception.** Real-time three-dimensional perception allows robots to recognize objects and estimate their position and orientation, enabling flexible handling and faster changeovers. This level expands automation into semistructured environments where exact positioning cannot be guaranteed.
- **Level 3: Dexterous Manipulation.** Robots begin to handle contact-rich tasks and variable or deformable objects by coordinating perception, motion, and force. Although capabilities

improve significantly, performance often remains tightly coupled to specific embodiments and training setups.

- **Level 4: Workflow Planning.** Systems can interpret high-level goals and autonomously sequence tasks, shifting from explicit programming to intent-driven execution. Behavior remains largely probabilistic and reactive, however, without a causal understanding of the physical world.
- **Level 5: Reasoning.** Robots maintain causal world models that allow them to predict outcomes, reason under uncertainty, and pursue complex goals over time. This level enables true general-purpose autonomy—but remains largely aspirational today.

EXHIBIT 1

The Evolution of Physical AI Within Robotics



Sources: Company information; BCG analysis.

With this framework in place, attention shifts to understanding where value can be captured today and how quickly that scope can expand as capabilities mature. The sections that follow focus on Levels 2 and 3, where tangible economic impact is already being realized, before turning to the frontier represented by Levels 4 and 5. Because Level 1 automation is already broadly adopted across industries, it does not materially shape today's strategic choices around physical AI.

Where Robotics Is Delivering Value Today

Although Levels 2 and 3 capabilities do not deliver general-purpose autonomy, they allow companies to automate classes of tasks that were previously uneconomical or impossible to scale. Their impact is felt less in headline-grabbing demos and more in fundamental changes to the cost structure, flexibility, and scalability of automation.

Level 2: Visual Perception. For decades, industrial automation was constrained not by motion control but by perception. Robots could move with extreme precision, yet they required parts to be presented in exactly the right position and orientation. As a result, automation projects front-loaded costs into fixtures, feeders, and painstaking manual tuning, making them viable only for stable, high-volume production.

At Level 2, this constraint shifts. Once robots can reliably perceive objects and estimate their full six-dimensional position—their exact location and orientation in space—perception ceases to be the primary barrier to automation. The critical constraint becomes engineering speed: how quickly systems can be configured, validated, and adapted as products and processes change. This marks a fundamental inflection point in automation economics.

This advance is enabled by a fundamental change in how perception is implemented. Traditional vision systems relied on handcrafted, rule-based pipelines tuned for specific parts and layouts. Modern approaches instead leverage AI-based vision models trained on large volumes of image data to deliver robust perception across variable conditions. Rather than encoding explicit rules for every scenario, these models learn to generalize from examples.

> “As perception becomes model-based, the surrounding engineering workflow can also be restructured—from bespoke, manually tuned pipelines to data- and simulation-driven processes.

The deeper transformation, however, lies beyond the model itself. As perception becomes model-based, the surrounding engineering workflow can also be restructured—from bespoke, manually tuned pipelines to data- and simulation-driven processes. Companies can increasingly automate key steps such as ingesting engineering data, generating synthetic training data, refining models, validating performance in simulation, and deploying updates with minimal manual intervention.

Perception thus evolves into a continuously improving software capability rather than a one-off solution.

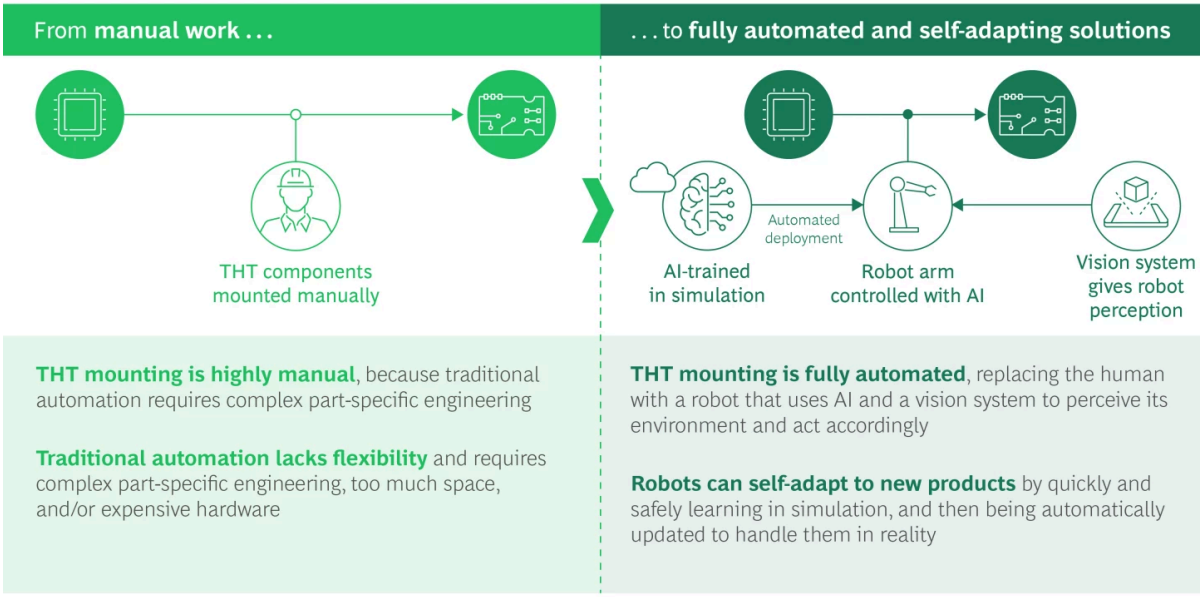
This evolution is already visible in industrial deployments across sectors such as sheet-metal processing and electronics assembly, where perception-driven automation enables flexible handling without custom fixtures or manual retuning. By automating elements of the perception-engineering workflow, these systems can seemingly self-adapt as products and geometries change. (See “Perception in Practice.”)

— Perception in Practice

Two companies illustrate how Level 2 systems replace rule-based vision pipelines with data- and simulation-driven perception engineering:

- In sheet-metal processing, Intrinsic, an AI robotics group at Google, has collaborated with TRUMPF to demonstrate how advanced perception enables robots to reliably detect and handle reflective parts with arbitrary geometries. These tasks previously required custom fixtures and extensive manual tuning.
- In electronics manufacturing, Foxconn has shown how Level 2 perception allows robots to grasp components directly from unsorted bins in through-hole technology assembly. This technique eliminates rigid part-feeding systems and enables automation to adapt as product designs change. (See the exhibit.)

Level 2 Perception Enables the Automation of Handling Operations Too Versatile for Traditional Robotics



Sources: Foxconn; BCG analysis.
Note: THT = through-hole technology.

In both cases, the underlying value does not come from a single vision breakthrough, but from the ability to reconfigure systems quickly and repeatedly without costly and time-consuming reengineering.

The cost implications are significant. In traditional robotics deployments, roughly 75% of the total cost of ownership (TCO) is tied to initial setup and reengineering—configuring workflows, adapting systems to new products, and integrating them into existing operations. Our analysis shows that software-defined approaches can reduce setup and reengineering costs by up to 50%, making roughly 37.5% of TCO addressable. As a result, engineering effort shifts from scarce human experts to scalable computing. Initial setup costs fall, reengineering and changeover costs drop dramatically, and the marginal cost of adapting a system to a new product variant approaches zero. In effect, automation transitions from a series of bespoke projects to a software-defined capability.

The operational consequence is that robots can now be deployed in semistructured and variable environments, not just in tightly controlled production lines. Hardware becomes a platform for continuous learning, and robotic systems evolve into self-adapting assets. This is why Level 2 capabilities are already enabling automation beyond high-volume, stable production and unlocking value in industries long considered too variable to automate.

Level 3: Dexterous Manipulation. Traditional industrial robots handle rigid objects well, but they struggle with tasks that humans perform effortlessly—manipulating soft materials, managing contact forces, and adapting mid-action when conditions change. As a result, many handling and assembly tasks have remained stubbornly manual despite decades of automation investment.

> By integrating perception, semantic understanding, and action generation into a single learning framework, VLA models allow robots to perform more complex, contact-rich tasks.

Recent advances in learning-based robotic control have begun to close this gap, with vision-language-action (VLA) models emerging as the latest, most promising approach to human-level dexterous performance. By integrating perception, semantic understanding, and action generation into a single learning framework, VLA models allow robots to perform more complex, contact-rich tasks. They can adapt grasps based on context, reorient objects dynamically, and execute behaviors that were previously infeasible with hand-engineered control logic. In controlled environments, this progress is already delivering measurable results. (See “Scaling Dexterous Manipulation in Warehouses.”)

— Scaling Dexterous Manipulation in Warehouses

Vulcan, Amazon’s most advanced robotic manipulation system to date, offers a glimpse of what Level 3 systems make possible. Unlike traditional warehouse robots that rely on rigid, hand-engineered routines, Vulcan is powered by perception and action models that let it adapt to real-world variability.

Amazon reports that Vulcan can already reliably handle around 75% of the more than 1 million unique items in its fulfillment catalog—a level of generalization previously unimaginable for robotic picking systems. It can grasp unfamiliar objects, extract tightly packed items from cluttered bins, manipulate deformable goods, and choose context-dependent grasp points without human intervention.

But the limitations of Level 3 are as important as its promise. Today's dexterous manipulation systems remain tightly coupled to their physical embodiment. (See Exhibit 2.) Skills learned by one robot do not readily transfer to another with different kinematics, sensors, or end-effectors. Achieving reliable performance requires training and refining the system using data generated by the specific robot performing the task under realistic conditions. As a result, scaling dexterity across diverse environments remains costly and complex.

In practice, this means that Level 3 systems can deliver impressive results within well-defined domains, but they are not general-purpose. They excel at specific tasks under controlled conditions with carefully engineered training pipelines, yet performance can degrade quickly outside those boundaries. Collecting training data and refining models also require substantial effort. Along with the new software-defined capabilities needed to deploy and operate these systems, this keeps entry costs high.

Ultimately, the next breakthrough in dexterous manipulation will depend less on sheer data volume and more on improving embodiment transfer and physical grounding—the ability to reliably apply learned skills across different robotic forms.

The Critical Boundary Between Planning and Reasoning

Levels 4 and 5 define the frontier between impressive automation and general-purpose autonomy. These higher levels are particularly consequential for humanoid robots, whose economic case depends less on isolated task performance and more on the ability to operate flexibly across many contexts.

Although Levels 4 and 5 are often discussed together, they represent fundamentally different capabilities—planning versus reasoning—and failing to distinguish them is a major source of inflated expectations.

Level 4: Workflow Planning. At Level 4, robots move beyond explicit programming and begin to operate from high-level goals. Advances in large language models allow systems to interpret human intent, decompose tasks into steps, and generate action sequences dynamically. Instead of being told how to perform each motion, robots can be instructed what to achieve.

> Level 4 systems can already coordinate complex workflows, adapt task order when conditions change, and integrate multiple tools or subsystems without manual reprogramming.

This shift is significant. It enables faster setup, greater flexibility, and more natural human–robot interaction. In semistructured environments, Level 4 systems can already coordinate complex workflows, adapt task order when conditions change, and integrate multiple tools or subsystems without manual reprogramming.

It is critical to understand, however, what Level 4 does not provide. These systems rely on generative, decoder-based models that produce probabilistic action sequences. They do not reason causally about the physical world; they predict plausible next steps based on learned patterns. As a result, Level 4 intelligence remains fundamentally reactive rather than deliberative. It can generate convincing plans, but it cannot reliably evaluate whether those plans will succeed under novel physical conditions.

Increasingly, leading researchers are recognizing that limitation. They highlight that current language-model-based systems lack robust causal reasoning and hierarchical planning capabilities—qualities essential for reliable physical autonomy. For robotics, this distinction matters: plausible plans are not the same as executable ones.

Level 5: Reasoning. Level 5 represents a qualitatively different capability. At this level, robots maintain an internal world model that allows them to reason about state, causality, and consequence. They can infer hidden conditions, anticipate how the environment will change as a result of their actions, and choose among alternatives based on expected outcomes rather than statistical likelihood alone.

> Putting a plate into a dishwasher is a Level 2 or 3 task. Cleaning the kitchen, by contrast, requires reasoning: recognizing dependencies, sequencing actions appropriately, handling uncertainty, and adapting when unexpected obstacles arise.

The difference becomes clear when considering the deployment of a humanoid robot in a kitchen. (See Exhibit 3.) Putting a plate into a dishwasher is a Level 2 or 3 task. Cleaning the kitchen, by

contrast, requires reasoning: recognizing dependencies, sequencing actions appropriately, handling uncertainty, and adapting when unexpected obstacles arise. This is the level of intelligence that humanoid robots implicitly promise, but it remains largely out of reach.

Achieving Level 5 requires more than advances in language models. It depends on robust internal world models that represent state, causality, and long-horizon consequences. Yet there is no settled approach to building such models for robotics. Prominent AI researchers, including pioneers in deep learning, have argued that new, nongenerative architectures may be required to achieve reliable reasoning, even as major industry players continue to invest heavily in advancing generative models. (See "Competing Approaches to World Modeling.") What is clear, however, is that Level 5 reasoning is the gating constraint for truly general-purpose robotics.

— Competing Approaches to World Modeling

Two architectural approaches dominate current research on world modeling:

- **Generative (decoder-based) models** attempt to simulate the future in rich detail. These systems compress observations into an internal representation and then generate what the next scene might look like, effectively imagining future video frames. Their appeal lies in visual richness and interpretability. But the tradeoffs are significant: high computational cost, sensitivity to irrelevant visual details, and the risk of producing scenes that look plausible without being physically correct.
- **Joint-embedding (decoder-free) models** take the opposite approach. Rather than reconstructing imagery, they learn compact representations of the underlying state of the world and how it changes over time. This makes them more efficient and better suited to control and planning. The downside is reduced transparency: without imagined visuals, validation becomes harder, and certain forms of visual reasoning are constrained.

Both paradigms are actively evolving, and hybrid approaches are possible. The key takeaway for decision makers is that robust causal reasoning remains an open engineering challenge rather than a mature, deployable capability.

How to Sequence Robotics Investments with Discipline

The five levels of physical AI—and the risks associated with misjudging them—lead to a clear conclusion: durable automation strategies are built on mature capabilities and upgrade paths, not singular bets on form factors or near-term breakthroughs. For companies shaping automation roadmaps, three imperatives stand out.

Let capabilities, not form factors, drive investment decisions. Roadmaps anchored to specific robot types risk locking in assumptions about intelligence that the technology has not yet earned. This is especially true for humanoid robots, which may ultimately play a meaningful role, particularly in environments designed for human workers. But their economic viability depends on progress at the highest levels of physical AI—especially Level 5 reasoning—which remains uncertain. A capability-driven approach explicitly maps use cases to the levels of perception, manipulation, planning, and reasoning required. This enables realistic sequencing of investments, clearer risk management, and flexibility as technical boundaries shift.

Prioritize mature capabilities. The strongest near-term returns come from capabilities that expand flexibility while reducing engineering effort. Level 2 perception allows automation to move beyond fixed, high-volume processes, while Level 3 dexterous manipulation unlocks targeted handling and assembly tasks that were previously manual. For most organizations, these levels offer the best balance between technical readiness and economic impact. Investments should focus not only on individual use cases, but also on reusable platforms, workflows, and data assets that can scale across products and sites.

Architect for continuous learning and upgradeability. As automation becomes increasingly software-defined, long-term value depends less on initial deployment and more on how systems evolve. Architectures should support continuous data collection, model retraining, simulation-based validation, and modular integration of new components. This allows organizations to incorporate advances in perception, manipulation, and planning as they mature—without wholesale redesign. In practice, this requires building software, data engineering, and simulation capabilities alongside traditional automation expertise.

Turning Robotics Progress into Durable Advantage

Companies across industries are already reaping the value of robotics. They're deploying systems that deliver measurable gains in flexibility, productivity, and cost by automating work once considered too variable or complex to scale. The strategic question is not whether value exists, but how far—and how quickly—it can expand as physical AI advances.

For boards and executives, this is ultimately a capital allocation problem. The largest risks lie not in investing in robotics, but in misjudging capability maturity—either overcommitting to technologies that haven't yet crossed critical thresholds or underinvesting in capabilities that are already economically viable. As perception and dexterity mature, significant value can be captured today, even as reasoning remains an open frontier.

The companies that turn robotics progress into a durable advantage will not be those chasing spectacle or waiting for a singular breakthrough. Rather, they'll be sequencing investments with discipline, building architectures for continuous learning, and anchoring their roadmaps in proven capabilities, not aspirations. In a market where capital is moving quickly and expectations are rising, disciplined execution—not hype—will separate leaders from laggards.

Authors

Tilman Buchner

Partner & Director
Munich

Martin Kleinhans

Project Leader
Munich

Daniel Küpper

Managing Director & Senior Partner
Cologne

Jonathan Brown

Managing Director & Partner
Hamburg

Max Ludwig

Managing Director & Partner
Munich

Simon Rees

Managing Director & Partner
Boston

Francesco Palmegiano

Consultant
Zurich

ABOUT BOSTON CONSULTING GROUP

Boston Consulting Group partners with leaders in business and society to tackle their most important challenges and capture their greatest opportunities. BCG was the pioneer in business strategy when it was founded in 1963. Today, we work closely with clients to embrace a transformational approach aimed at benefiting all stakeholders—empowering organizations to grow, build sustainable competitive advantage, and drive positive societal impact.

Our diverse, global teams bring deep industry and functional expertise and a range of perspectives that question the status quo and spark change. BCG delivers solutions through leading-edge management consulting, technology and design, and corporate and digital ventures. We work in a uniquely collaborative model across the firm and throughout all levels of the client organization, fueled by the goal of helping our clients thrive and enabling them to make the world a better place.