https://doi.org/10.1038/s42256-026-01239-3

From embodied intelligence to physical AI

Several frameworks from different disciplines are converging on the scientific question of what it takes for a system to not just predict, simulate or reason about the world, but to act physically and intelligently within it.

Contemporary artificial intelligence (AI) has transformed science and technology, with advances in machine learning, particularly in deep neural networks, enabling a wide range of applications across many domains. Yet despite progress in language, vision and foundation models that can model the world, the development of systems that can robustly perceive, act and adapt effectively within it remains an open challenge.

Nearly all of AI is digital, virtual or otherwise removed from direct engagement with the physical world. The result is an asymmetry between expectation and capability. Although there is much talk of 'artificial general intelligence', commercial robotic systems continue to struggle to perform relatively mundane tasks such as opening ordinary doors in all their real-world variation. Moravec's paradox still holds: cognitive processes such as abstract reasoning, which humans regard as demanding, are often easier for machines than the sensorimotor skills that humans can perform effortlessly.

The physical world changes the problem of intelligence for several reasons. Vision data such as images and video, for example, are comparatively abundant and readily accessible, particularly for narrowly defined tasks such as object classification or facial recognition. But unlike image analysis tasks with relatively fixed labels, physical interaction is fundamentally uncertain. Identical real-world situations can lead to different responses, even from the same individual, and there is often no single 'correct' action to learn from, especially in complex or dynamic settings. As a result, data that capture physical interactions are much harder and more time consuming to obtain, making robust real-world behaviour more difficult for machines to learn.

To argue that we lack the right data is, in part, to acknowledge that we do not yet understand what must be measured when bodies, environments and actions interact. Several overlapping frameworks have begun to converge on this challenge from different directions, including world models, embodied intelligence, ecological psychology and neuroscience, as well as physical AI and robotics. All approaches address questions of agency, embodiment and the relation between perception and action, even if they do not use the same language.

The idea of a world model can be traced to Kenneth Craik, a Scottish psychologist and philosopher, who proposed in 1943 that an organism could act more effectively by maintaining a "small-scale model" of the world alongside representations of its own possible actions[1]. That intuition underlies many current uses of the term in AI, in which world models are typically understood as internal representations that enable prediction, planning and the evaluation of action. Current approaches vary in emphasis. Some emphasize physical world models that capture aspects of the environment such as objects and their dynamics, whereas others focus on mental world models that represent aspects of human context, including goals and social interactions[2].

Embodied intelligence places the emphasis elsewhere. In this view, intelligence is not computation that can be abstracted from the body. Instead, the body helps to determine what an agent can detect, learn and do[3]. Recent arguments in NeuroAI have sharpened this point by suggesting that embodiment is not peripheral to intelligence and brain function but a part of its structure[4]. As Bing Brunton, a professor of neuroscience at the University of Washington, recently emphasized at the 2026 COSYNE conference, the brain neither senses nor acts on the world except through the body. Any model of intelligent behaviour must therefore include not only the nervous system and the external environment, but also the body that mediates between them.

Ecological psychology and ecological neuroscience shift the focus from how organisms perceive the world to how they perceive what the world affords in action. In the account of James J. Gibson, an American psychologist from the mid-twentieth century, affordances are opportunities for action that emerge from the relationship between body, environment and task[5]. A door handle, for example, affords grasping because it is directly perceived in terms of possible action, not because its 'meaning' must first be inferred. A 2019 News Feature in *Nature Machine Intelligence* suggested that this perspective could shift robotics away from detailed world reconstruction and towards the actionable structure of everyday environments[6]. More broadly, it offers a way to link perception, behaviour and environment without reducing intelligence either to internal representation or to mere reaction.

Recent discussions of 'physical AI' give this shift towards embodiment and affordances an engineering expression. The emphasis is not only on embedding AI within machines, but on recognizing that sensing, morphology, materials and control all contribute to adaptive behaviour[7]. In this view, intelligence is not confined to software alone but is partly realized in materials, morphology and mechanics.

Together, these overlapping approaches are converging on the scientific problem of what it takes for a system to act intelligently in the world. They differ in what they identify as the key ingredients. Some emphasize internal predictive structure; others focus on embodiment, sensorimotor coupling or affordances; still others locate aspects of intelligence in materials, morphology and mechanics. These differences are potentially productive because they sharpen scientific questions and help to guide progress.

Robotics is where systems acting in the physical world encounter consequences in real time, making it a testing ground for competing claims about world models, embodiment, affordances and physical AI. It also supports proposals such as an embodied Turing test, in which the question is not whether a machine can convincingly imitate human conversation, but whether it can acquire and deploy skilled action in real-world settings[8]. The next phase of AI may be defined less by better descriptions or predictions of the world, and more by increasingly capable action within it.

Published online: 24 April 2026

References

1. Craik, K. *The Nature of Explanation* (Cambridge Univ. Press, 1943).
2. Fung, P. et al. Preprint at https://doi.org/10.48550/arXiv.2506.22355 (2025).

3. Brooks, R. A. *Artif. Intell.* **47**, 139–159 (1991).
4. Brunton, B. & Tuthill, J. *The Transmitter.* https://doi.org/10.53053/CHML6687 (2025).
5. Gibson, J. J. *The Ecological Approach to Visual Perception* (Lawrence Erlbaum Associates, 1986).
6. Hsu, J. *Nat. Mach. Intell.* **1**, 124–127 (2019).
7. Miriyev, A. & Kovač, M. *Nat. Mach. Intell.* **2**, 658–660 (2020).
8. Zador, A. et al. *Nat. Commun.* **14**, 1597 (2023).