GUEST EDITORIAL

National Science Review
13: nwag689, 2026
https://doi.org/10.1093/nsr/nwag689
Advance access publication 31 October 2026

Special Topic: Embodied Artificial Intelligence

Embodied artificial intelligence: from understanding the world to changing it

Han Ding[1,2,*,†], Wenwu Zhu[3,*,†] and Ye Yuan[2,4,*,†]

Artificial intelligence (AI) is transforming information processing and advancing the ability to understand the world, opening new possibilities for knowledge discovery, problem solving, reasoning and decision-making. Injecting these capabilities into the physical world calls for embodied AI (EAI), which heavily relies on a physical body to enable machines, or even every object, to act on, adapt to and even change their surroundings for creating environments, eventually the world, that support better life, work, technological development and economic growth. Such a paradigm of changing the world through a physical body places physical interaction at the core of EAI, as actions alter the environment and generate new sensory evidence that shapes subsequent understanding and behavior. The central challenge is therefore to make understanding and action reinforce each other in a reliable perception–action loop that adapts to changing physical conditions while remaining safe and responsive to human intent. This special topic on EAI, including reviews, perspectives and research articles, aims to address the challenge through intent-driven and self-evolving frameworks, the grounding of reasoning in measurement and control, adaptive contact-rich manipulation and industrial deployment and physical human–robot interaction. Viewed as a whole, they aim to help readers connect conceptual advances with physical capabilities, assess progress toward reliable deployment and identify open problems in sustaining this loop across bodies, tasks and environments.

At the conceptual level, understanding needs to be organized around intent. Agents trained on correlation alone compromise causal coherence and value alignment [1]. Intent must therefore be explicit, revisable and verifiable, mediating human goals, environmental constraints and behavior. This closed loop must also persist across time. Most embodied systems are built on given memory, tasks, embodiment and environment. Maintaining this closed loop under changing conditions requires components that update memory, switch tasks, predict the environment, adapt the embodiment and evolve the model on their own, together with software–hardware co-evolution [2]. Purposeful understanding and a maintained capacity to change frame the contributions that follow.

Between understanding and changing lies an interface that remains poorly resolved in current systems. An open-ended command, such as inspecting a damaged roof, does not specify a motion. It poses a question about the task's state that only measurement can answer. Understanding is thus acquired through action, extending active perception and informative planning, and world models must predict how knowledge of the task will evolve in addition to how images will change [3]. In the reverse direction, understanding must be verified before it changes anything, because action cannot be revoked once taken. Reasoning is slow where control must be fast, and latency, occlusion and intermittent communication amplify errors in swarms. In the swarm architecture proposed in this special topic, reasoning is therefore confined to planning and every plan is checked before execution [4]. This interface, rather than either side alone, determines reliability.

At its most physical, changing the world happens through contact, which is non-smooth, only partially observable and switches between discrete contact modes. The field has moved from explicit contact models to learned policies. This is best read as a redistribution of contact knowledge across models, learned representations and feedback adaptation, not as a replacement of physics by data [5]. For learning, two consequences follow. Objects of different geometry and friction exhibit different contact dynamics, and a single policy trained on all of them inherits their conflicts. Identifying the latent modes of the hybrid system and composing specialized policies mitigates this conflict and, in the reported twisting tasks, generalizes zero-shot to unseen real objects [6]. Experience collected under past dynamics also misleads once the physics drifts, and correcting for both policy and dynamics shifts together keeps old experience tied to present conditions [7]. Reliable change thus depends on knowing when prior experience applies, and more data is not, by itself, more valid experience.

Whether this closed loop holds is ultimately decided in deployment. Industrial settings quantify what changing the world costs. Precision, real-time performance, reliability and safety must all hold at once and at bearable cost, and flexibility missing any one of them does not survive the production line [8]. The interaction data, simulators and validation protocols involved are indispensable instruments, though instruments rather than the substance of intelligence itself. Interaction also extends from objects to people. In walking

© The Author(s) 2026. Published by Oxford University Press on behalf of China Science Publishing & Media Ltd. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

assistance, large-area tactile coupling at the torso lets intent, balance and load be exchanged continuously, without explicit commands [9]. Contact is at once the physics to be controlled and the channel through which intent is read. The robot understands the human through the same coupling with which it supports them, and the body becomes a shared interface rather than a one-way instrument.

These contributions also point to a problem with how progress is measured. Benchmarks typically report the average success rate on a fixed set of tasks. Deployment asks harder questions. Real conditions rarely hold still. A robot's behavior changes as it learns, and the physical dynamics it acts on drift as well [7]. Industrial users judge a system by its worst case, and a single catastrophic failure can rule it out [8]. A deployed system must also recognize the limits of what it knows. For aerial robots, that means testing whether a drone avoids false certainty and can recover when its first reading of a command turns out to be wrong [3]; in vehicle swarms, unsafe or low-confidence plans should be caught and replaced before they are executed [4]. A useful benchmark would therefore report not only whether a task was completed, but how wide a range of conditions the system can handle and how it fails when it reaches their edge. Nor can safety and value alignment be added afterwards. In the frameworks presented here, they are part of how an agent represents intent [1], how it updates itself [2] and how it checks evidence and plans before acting [3,4].

Taken together, the contributions in this special topic advance our understanding of how intelligence can be translated into physical action through perception, learning, control and interaction. They also reveal persistent challenges in generalizing capabilities across tasks and environments, adapting to changing physical conditions and ensuring reliable and safe operation. Addressing these challenges requires evaluation beyond isolated task success, with greater attention to sustained performance, physical constraints and failure recovery. Meanwhile, EAI extends beyond the models and control methods emphasized in this collection. Although some contributions address physical embodiment and sensing, deeper integration of intelligent models with perception, actuation, hardware design and physical embodiment remains essential. Such a seamless integration could enable more general intelligent capabilities across diverse embodiments while accommodating their distinct physical characteristics and constraints. Achieving this vision calls for collaboration across AI, robotics, control, mechanics, materials science and human–robot interaction, alongside appropriate consideration of ethical and regulatory issues. Ultimately, EAI should not be confined to any particular robotic form, but should enable diverse physical systems to learn, adapt and act through interaction, advancing AI from understanding the world to changing it and thus bringing about a better future for humanity.

Han Ding[1,2,*,†], Wenwu Zhu[3,*,†] and Ye Yuan[2,4,*,†]

[1]School of Mechanical Science and Engineering, Huazhong University of Science and Technology, China; [2]State Key Laboratory of Digital Manufacturing Equipment and Technology, Huazhong University of Science and Technology, China; [3]Department of Computer Science and Technology, Beijing National Research Center for Information Science and Technology, Tsinghua University, China and [4]School of Artificial Intelligence and Automation, Huazhong University of Science and Technology, China

*__Corresponding authors__. E-mails: dinghan@hust.edu.cn; wwzhu@tsinghua.edu.cn; yye@hust.edu.cn

†Guest Editors of the Special Topic.

Conflict of interest statement. None declared.

REFERENCES

1. Zheng N. *Natl Sci Rev* 2026; **13**: nwag476.
2. Feng T, Wang X, Zhu W. *Natl Sci Rev* 2026; **13**: nwag662.
3. Zhou X, Wu Y, Liu Z *et al. Natl Sci Rev* 2026; **13**: nwag404.
4. Zhang HT and Hu BB. *Natl Sci Rev* 2026; **13**: nwag401.
5. Zhu BK, Ding Y, Zhang D *et al. Natl Sci Rev* 2026; **13**: nwag618.
6. Liu X, Dong Y, Wan J *et al. Natl Sci Rev* 2026; **13**: nwag351.
7. Luo Y, Lv L, Sun F *et al. Natl Sci Rev* 2026; **13**: nwag506.
8. Yu H, Song C, Zhang Y *et al. Natl Sci Rev* 2026; **13**: nwag505.
9. Sun Y, Wei J, Chen L *et al. Natl Sci Rev* 2026; **13**: nwag402.

© The Author(s) 2026. Published by Oxford University Press on behalf of China Science Publishing & Media Ltd. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.