



## The Embodiment of Intelligence: The Dawn of Physical AI

Olmer García<sup>1</sup>

Hector Florez<sup>2</sup>

During the last decade, the world witnessed a digital cognitive revolution. In particular, the numerous large language models (LLMs) that have been recently developed now dominate text processing and image generation. Nevertheless, generative artificial intelligence (AI) cannot navigate physical systems with the same ease. For instance, generating an essay seems to be trivial for a modern model, but asking it to manipulate a delicate object still poses a serious challenge, illustrating Moravec's paradox: performing complex operations is easy for computers, but replicating human skills is very difficult.

In this context, an alternative called *physical AI* has emerged, which is able to learn how to control robotic systems. Traditionally, the field of robotics has operated under rigid control paradigms, dependent on finite-state machines and explicit programming. Some popular robotics languages are based on the Robot Operating System (ROS). However, vision-language-action (VLA) models have sparked a radical paradigm change, shifting from isolated control policies to generalized robotics. This entails the creation and expansion of several architectures that are able to transform vision language models (VLMs) from passive AI generators into active decision-making agents.

Consequently, new taxonomies have emerged: (i) autoregressive approaches (e.g., RT-2 or OpenVLA) that discretize reality by treating actions as sequential tokens and (ii) diffusion-based models (e.g., physical Intelligence or diffusion policies) that model continuous actions to achieve fluid and multimodal movement. In addition, there are hybrid and reinforcement learning strategies that enable high-level semantic reasoning and dynamic execution in unstructured environments.

Training these models used to be the privilege of a few laboratories with significant computing and infrastructure resources. However, initiatives like Hugging Face's LeRobot are standardizing data formats and offering accessible pre-trained models (e.g., SmallVLA). LeRobot is allowing independent researchers and developers to fine-tune cutting-edge models with their own datasets.

This paradigm shift underscores two points of view. On the one hand, startups like SkildAI and Physical Intelligence are building an 'omni-bodied brain', suggesting that, based on a robust foundational model, hardware might become secondary. On the other hand, Nvidia, with models like GR00T, and Google DeepMind, with its Gemini Robotics initiative, are demonstrating how the immense multimodal reasoning capacity of its models can be adapted to navigate and manipulate the physical world. This demonstrates that the embodiment of AI has been addressed by both new startups and major computing companies.

1. Ekumen Inc, United States. [olmerg@ekumenlabs.com](mailto:olmerg@ekumenlabs.com)

2. Universidad Distrital Francisco José de Caldas, Colombia. [haflorezf@udistrital.edu.co](mailto:haflorezf@udistrital.edu.co)

However, it is still necessary to take these models from mere demonstrations to robust industrial infrastructure, wherein open-source community plays a fundamental role. Recent initiatives, such as the Special Interest Group on Physical AI of the Open-Source Robotics Alliance (OSRA), serve as a reminder that AI does not replace systems engineering, but that it must be integrated with it. This group has identified critical areas that require standardization in order to construct unified interfaces and messaging to ensure interoperability, create robust data collection pipelines, and establish reference platforms.

For the scientific community, and specifically for the readers of *Revista Científica*, this poses both a challenge and an opportunity. On the one hand, classical control engineering must now engage with data science and deep learning. On the other hand, the simulation-to-real gap is closing, not only with better simulators, but with models that understand the uncertainty of the real world.

All in all, the world in which AI only lived on servers is coming to an end; soon, AI systems will be able to walk and manipulate objects alongside us. Physical AI is not just a technological evolution; it is the ultimate validation that intelligence must be able to interact with the physical world.