Co-funded by the European Union

NEUROKIT2E Markets Newsletter #2

Turning Edge AI into Real-World Robotics

Content

Robotics has long relied on learning to promise autonomy, yet real-world deployment has remained limited. The true bottleneck is no longer algorithms, but the ability to execute intelligence reliably under physical constraints. This newsletter explores how Edge AI has become a structural requirement for robotics – enabling learning-based systems to move from demonstrations to real-world deployment. Here, Edge AI refers to intelligence executed directly onboard robots, under constraints of latency, energy and safety.

Edge Deployment as the New Bottleneck in Robotics

Recent discussions in robotics highlight a shift from improving AI models to deploying intelligence reliably on real robots. Edge AI brings computation closer to where robots act. In this context, AIDGE provides the industrial deployment framework, while Neurokit2E advances compression and transfer learning to make robotic intelligence scalable and adaptable.

Edge Deployment as the New Bottleneck in Robotics

Intelligent Robots in 2026: Are We There Yet ? Nikita Rudin

Recent discussions in robotics, including those led by Nikita Rudin (ETH Zurich robotics researcher specializing in reinforcement learning for real-world robots), converge on a clear observation: the challenge has shifted from AI capability to the ability to deploy intelligence reliably on real robots operating in the physical world. This is where AIDGE plays a strategic role by providing an open, sovereign and industrial-grade framework for edge deployment.

More broadly, Edge AI reflects the need to bring intelligence closer to where robots act, rather than relying on distant cloud infrastructures. In this context, AIDGE provides the practical means to turn advanced AI into usable robotic intelligence, while Neurokit2E complements this effort by investigating innovative neural network compression and transfer learning methods that make such edge deployment efficient, adaptable and scalable. Together, these initiatives contribute to making long-term visions of modular, learning-based and platform-driven robotics achievable beyond demonstrations and into real-world systems.

Cloud Robotics vs Edge Robotics

Written in 2015, this paper was strikingly prescient, identifying the limits of cloud-centric robotics long before embedded AI technologies existed to make its proposed shift toward distributed, edge-based intelligence a practical reality.

Deep Edge AI: Enabling Intelligence Where Robots Act

Long before Edge AI became a central topic in robotics, this review clearly identified the structural shift from cloud-based intelligence to onboard execution. It explains why autonomy in the physical world requires intelligence to run where robots act – under constraints of latency, energy, privacy and reliability – and not in distant data centers.

Cloud Robotics vs Edge Robotics

Goldberg, K., Kehoe, B., et al. (2015). Cloud Robotics and Automation: A Survey. IEEE.

Published in 2015, this survey on Cloud Robotics and Automation was remarkably forward-looking. At a time when embedded AI hardware, efficient inference frameworks, and edge deployment tools largely did not exist, the authors already identified the core limitations of cloud-centric robotics: latency, network dependency, safety risks, and poor suitability for real-time interaction with the physical world. The paper clearly anticipated that scalable robotics would ultimately require intelligence to move closer to the robot itself, through distributed and hybrid architectures. Nearly a decade later, this analysis appears strikingly prescient. What was a conceptual limitation in 2015 has now become a technological opportunity, as advances in edge computing and embedded AI make it possible to realize the architectures the paper could only outline – precisely the gap that modern Edge AI frameworks such as AIDGE are designed to address.

Deep Edge AI: Enabling Intelligence Where Robots Act

Chen, J., & Ran, X. (2019). Deep Learning With Edge Computing: A Review. Proc. IEEE

This review anticipated the structural shift from cloud-based AI to intelligence executed directly at the edge. It explains how advances in model compression, hardware–software co-design and task-aware optimization make it possible to run complex neural networks under strict constraints of latency, energy and reliability. Beyond its relevance to robotics, this perspective is fully aligned with the objectives of the Neurokit2E project, which explores innovative approaches to neural network compression and transfer learning in order to enable efficient, adaptable intelligence on embedded robotic platforms. Together, these efforts reinforce the idea that real-world autonomy depends not only on better models, but on making them deployable where robots actually act.

Learning Was the Promise — Edge AI Makes It Real

The challenge in robotic learning is no longer how to learn, but how to run learned intelligence reliably where robots act in the physical world

When Robots Had to Walk, AI Had to Be Embedded

Robotics has already moved AI onto robots – but turning embedded intelligence into a robust, reusable and industrial capability remains the real challenge

Making Vision-Language-Action Work at the Edge

Among the first works to show Vision-Language-Action models running onboard real robots rather than in the cloud.

Learning Was the Promise — Edge AI Makes It Real

Argall, B. D., Chernova, S., Veloso, M., Browning, B. (2009). A Survey of Robot Learning from Demonstration. Robotics and Autonomous Systems.

Written at a time when large-scale learning and embedded AI were still limited, this article lays foundational concepts that continue to influence modern robot learning approaches. It provides a comprehensive overview of Learning from Demonstration (LfD), a family of techniques in which robots learn control policies directly from human demonstrations rather than from explicit models or reward functions. The paper systematically categorizes LfD approaches according to how demonstrations are collected, how policies are derived, and how performance can be improved beyond the demonstrated examples. It highlights the strengths of LfD as an intuitive and accessible way to transfer skills from humans to robots, especially in complex real-world tasks, while also clearly identifying its limitations, notably poor generalization, sensitivity to data quality, and dependence on the states encountered during demonstrations.

When Robots Had to Walk, AI Had to Be Embedded

Lee, J., Hwangbo, J., Wellhausen, L., Koltun, V., Hutter, M. (2020). Learning Quadrupedal Locomotion over Challenging Terrain. Science Robotics.

This article illustrates how legged robots have played a pivotal role in the early realization of embedded AI in robotics. Because walking robots must react continuously to contacts, terrain variations and balance disturbances, they imposed real-time constraints that made cloud-based control impractical, forcing intelligence to move onboard. The work demonstrates concrete achievements, including learned locomotion policies running directly on the robot, capable of adapting online to complex and unpredictable environments. However, it also exposes key limitations: the resulting systems are often tightly coupled to specific hardware, require careful tuning, and remain difficult to generalize or deploy across platforms. While legged robots proved that embedded AI could work in the physical world, the article implicitly highlights the remaining challenge – turning these successful but specialized implementations into scalable, robust and reusable edge AI deployments.

Making Vision-Language-Action Work at the Edge

Williams, J., Datta Gupta, K., George, R., Sarkar, M. (2024). LiteVLA: Efficient Vision-Language-Action Control on CPU-Bound Edge Robots. arXiv preprint

Recent advances now show that multimodal robot intelligence can run directly on real robots under strict edge constraints. LiteVLA demonstrates how vision-language-action models can be executed entirely on-board, enabling a mobile robot to perceive its environment, interpret high-level instructions and generate actions without relying on cloud connectivity. By combining lightweight model architectures, aggressive quantization and efficient fine-tuning techniques, the system achieves real-time control on CPU-only embedded hardware. At the same time, the work clearly exposes current limits in performance, thermal constraints and hardware coupling, highlighting that while embedded AI for robotics is now technically feasible, turning these successes into robust, scalable and reusable systems remains the next major challenge.

Conclusion

Taken together, these works show that robotics is no longer waiting for better AI models, but for the means to deploy them reliably, efficiently and sustainably in the physical world. The challenge has shifted from raw model performance to adaptability, compression, transferability and real-world robustness. In this context, Neurokit2E addresses a foundational layer of future robotic systems by investigating advanced neural network compression and transfer learning methods, enabling scalable and portable intelligence across heterogeneous platforms. Moving from isolated demonstrations to reusable and deployable robotic intelligence is no longer merely a technological question – it is a systemic transformation that Neurokit2E actively helps shape

NEUROKIT2E is co-Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or Chips Joint Undertaking. Neither the European Union nor the granting authority can be held responsible for them.