Ecological Challenges of Modern AI-Driven Robotics

Georgi Angelov
Institute of Robotics
Bulgarian Academy of Sciences
Sofia, Bulgaria
george@robotics.bg

Yasen Paunski
Institute of Robotics
Bulgarian Academy of Sciences
Sofia, Bulgaria
yasen@robotics.bg

***Abstract*: Robotics is increasingly driven by artificial intelligence (AI) through large language models (LLMs), vision–language models (VLMs), and large action models (LAMs), enabling unprecedented autonomy and adaptability. These capabilities, however, rely on hyperscale datacenters which rapidly increasing power density and cooling demands impose significant environmental costs. Rack-level power densities have risen from 2–4 kW in early datacenters to more than 100–140 kW in modern AI facilities, intensifying electricity and water consumption. This article reviews the ecological implications of AI-driven robotics, examines representative gigawatt-scale datacenter projects, and surveys emerging technological pathways aimed at mitigating energy and resource impacts.**

Keywords: AI-driven robotics, hyperscale datacenters, energy consumption, water usage, sustainability, foundation models, robotic autonomy, environmental impact

I. Introduction

Robotics has historically been rooted in deterministic, rule-based control systems. Early industrial robots were programmed using explicit algorithms that defined motion trajectories, sensor feedback loops, and decision rules with limited capacity for adaptation. While these systems proved reliable in structured environments such as automotive assembly lines, they were inherently brittle in the face of uncertainty and environmental variability. Over the past decade, the integration of AI - particularly deep learning and foundation model, has transformed robotics into a data-driven discipline capable of learning, generalization, and multimodal reasoning.

Large language models enable robots to interpret and generate natural language instructions, vision–language models provide semantic understanding of visual scenes, and large action models learn mappings from perception to sequences of physical actions. Together, these systems allow robots to operate in domains such as healthcare, logistics, agriculture, and disaster response, where classical control approaches are insufficient. Yet the intelligence enabling these capabilities is rarely embedded entirely within robotic platforms. Instead, it resides predominantly in remote datacenters that perform both training and inference at scale.

This architectural decoupling between physical robots and computational intelligence has fundamental ecological implications. Hyperscale datacenters consume rapidly growing amounts of electricity and water, contributing to greenhouse gas emissions, grid instability, and resource competition. As robotics becomes increasingly dependent on cloud-based AI, its environmental footprint becomes inseparable from that of global datacenter infrastructure.

II. From Algorithmic Robotics to Foundation Models

The earliest robotic systems relied on control theory, finite state machines, and kinematic modeling to execute predefined tasks. Techniques that used algorithmic control ensured precise motions but required carefully engineered environments. These systems lacked learning capabilities and were unable to generalize beyond their programmed conditions – Fig.1.

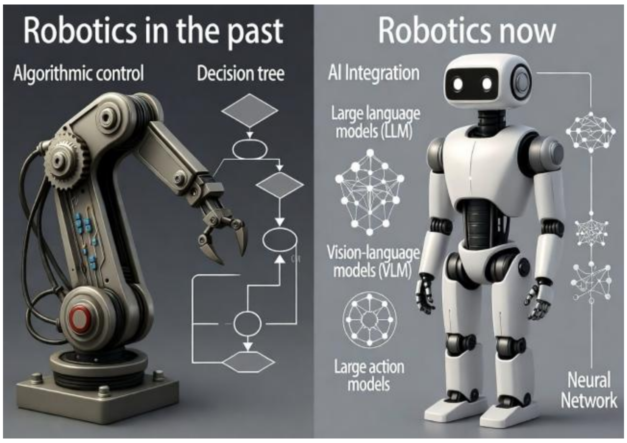

Fig. 1. Algorithmic vs AI enabled robotics control

The introduction of machine learning, and later deep learning, marked a turning point. Reinforcement learning enabled robots to optimize behavior through interaction with environments, while supervised learning improved perception tasks such as object recognition. More recently, foundation models have unified perception, reasoning, and action. LLMs allow robots to parse abstract instructions, VLMs integrate vision and language for contextual understanding, and LAMs generate coherent action sequences directly from sensory input. Systems such as RT-2 demonstrate how textual and visual representations can be mapped directly to robotic

ACKNOWLEDGE THE FINANCIAL SUPPORT OF THE PROJECT WITH FINANCING AGREEMENT NO. PVU-44 OF 05.12.2024 UNDER PROJECT NO. BG-RRP-2.017-0011 "ECOLOGICAL COLLABORATIVE ROBOTS POWERED BY GREEN HYDROGEN" UNDER THE RECOVERY AND RESILIENCE MECHANISM FOR THE IMPLEMENTATION OF AN INVESTMENT UNDER C2I2 "INCREASING THE INNOVATION CAPACITY OF THE BULGARIAN ACADEMY OF SCIENCES (BAS) IN THE FIELD OF GREEN AND DIGITAL TECHNOLOGIES" FROM THE RECOVERY AND RESILIENCE PLAN, BULGARIA.

actions, reducing the need for hand-engineered pipelines [16,17].

These advances, however, require models with billions of parameters trained on massive datasets. The computational cost of training and serving such models has shifted robotics toward centralized computing facilities, fundamentally altering its environmental impact profile.

III. Datacenter-Centric AI and Rising Power Density

Modern AI models are trained and deployed in hyperscale datacenters operated by major technology firms. Unlike earlier enterprise datacenters designed for general-purpose workloads, AI datacenters are optimized for massively parallel computation using GPUs and specialized accelerators. This shift is reflected most clearly in rack-level power density.

In the early 2000s, typical datacenters operated at approximately 2–4 kW per rack [2]. By the late 2010s, cloud computing and early GPU adoption raised this figure to 10–15 kW. In the current AI-driven era, power densities of 40–140 kW per rack are increasingly common – Fig. 2, driven by dense GPU and TPU clusters optimized for neural network training and inference [1,3].

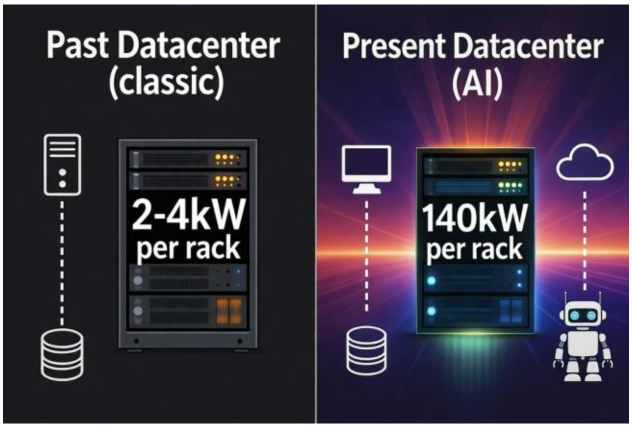

Fig. 2. Typical datacenter rack consumption comparison

The consequences of this escalation are multifold. Higher power density increases heat generation, necessitating advanced cooling solutions and raising operational costs. At a macro scale, datacenters already account for 1–2% of global electricity consumption, with projections suggesting this could rise to 8–10% by 2030 under continued AI growth [8]. Because many electrical grids remain partially fossil-fuel based, AI-driven robotics indirectly contributes to increased carbon emissions.

IV. Case Studies: Gigawatt-Scale AI Infrastructure

A. xAI Colossus

xAI's Colossus supercomputer, located in Memphis, Tennessee, exemplifies the scale of contemporary AI infrastructure. Operational since 2024, Colossus integrates more than 100,000 GPUs and reportedly consumes between 0.3 and 0.5 GW of power, with expansion plans approaching gigawatt levels [12]. Due to local grid constraints, the facility has relied in part on gas turbines, raising concerns about emissions and regulatory oversight [24]. Although Colossus enables rapid training of large action models relevant to AI and robotics, its environmental footprint highlights the tension between speed of deployment and sustainability.

B. OpenAI Stargate

OpenAI's Stargate initiative represents an even more ambitious vision. Planned as a multi-phase network of datacenters, Stargate is projected to require up to 7 GW of power and investments on the order of hundreds of billions of dollars [14]. Facilities such as the Abilene, Texas site, with a planned capacity of approximately 1.2 GW, underscore the emergence of AI infrastructure comparable in scale to major urban power consumers [15]. While these facilities are intended to support advances in LLMs, VLMs, and robotics, they also magnify ecological pressures on energy and water systems.

V. Cooling, Water Consumption, and Local Impacts

Energy consumption represents only one dimension of the environmental footprint of modern AI-oriented datacenters. Equally critical, yet often less visible, is the issue of thermal management and its associated water demand. The extreme power densities characteristic of contemporary AI hardware that frequently exceeds 100 kW per rack results in substantial heat generation that must be continuously dissipated to maintain system reliability and performance. Current cooling strategies rely predominantly on water-intensive methods, making water consumption a central sustainability concern in the operation of AI infrastructure supporting modern robotics.

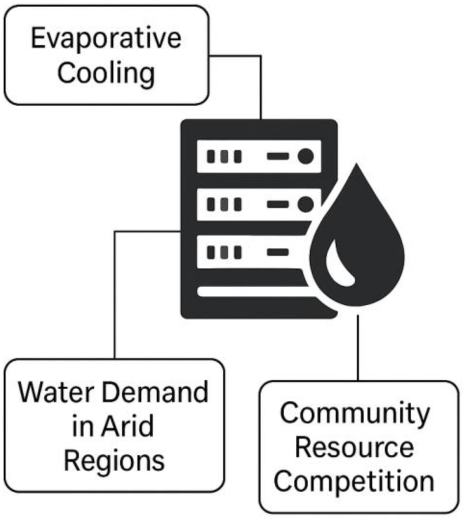

Fig. 3. AI datacenters water challenges

The most widely deployed cooling approach in hyperscale datacenters is evaporative cooling, which exploits the latent heat of vaporization of water to remove thermal energy from server environments - Fig 3. While highly effective in reducing energy overhead compared to purely air-based cooling, evaporative systems consume large volumes of freshwater. Estimates indicate that a single medium-sized datacenter may require on the order of 400 million liters of water annually, depending on climate, cooling design, and operational load [4]. As AI workloads continue to scale, both in model size and inference frequency, this demand is expected to increase correspondingly.

The implications of such water usage are particularly severe in arid and semi-arid regions, where datacenters are often located due to favorable land availability, tax incentives, and proximity to power infrastructure. In these contexts, datacenter water consumption directly competes with municipal, agricultural, and ecological needs. Reports from communities adjacent to large AI facilities describe declining groundwater levels, degraded water quality, and heightened vulnerability during drought periods [5]. These localized impacts illustrate how global AI infrastructure decisions can produce disproportionate environmental burdens at the community level.

Projections suggest that, if current trends persist, AI-driven datacenters could collectively consume up to 1.7 trillion gallons of water annually by 2030 [5]. Such figures raise concerns not only about sustainability but also about long-term resilience in the face of climate change. Increasing temperatures and more frequent droughts are likely to intensify cooling demands precisely when water availability becomes more constrained, creating feedback loops that further stress both technological and natural systems.

Beyond physical resource depletion, water consumption by datacenters introduces ethical and governance challenges. In several documented cases, communities subject to water-use restrictions for residential or agricultural purposes coexist with datacenters that operate under separate regulatory frameworks or receive preferential access to water resources. This asymmetry raises questions regarding environmental justice, transparency, and the prioritization of essential human needs over commercial computational activities. For robotics, whose societal value is often framed in terms of public benefit: such as healthcare, safety, and sustainability - these contradictions warrant particular scrutiny.

Efforts to mitigate water-related impacts include the adoption of closed-loop cooling systems, increased use of reclaimed or non-potable water, and the relocation of datacenters to regions with more abundant water resources. However, such measures introduce trade-offs, including higher capital costs, increased energy consumption, or the displacement of environmental pressures to other regions. As with energy consumption, water use in AI datacenters must therefore be understood as a systemic issue rather than a purely technical one.

In summary, cooling and water consumption constitute a critical but underappreciated component of the ecological footprint of AI-driven robotics. As robotic intelligence becomes increasingly dependent on centralized, high-density datacenter infrastructure, addressing water sustainability will be essential to ensuring that technological progress does not exacerbate existing environmental and social vulnerabilities.

VI. Sustainability-Oriented Technological Pathways

Although the primary emphasis of this article is the identification and characterization of the ecological challenges associated with AI-driven robotics, a comprehensive assessment must also consider the technological pathways currently proposed as potential mitigations. These pathways should not be interpreted as finalized or universally applicable solutions. Rather, they represent active research directions aimed at reducing the environmental footprint of large-scale AI infrastructure that increasingly underpins robotic intelligence.

A frequently discussed pathway is the transition toward low-carbon baseload energy sources for datacenters. Among these, nuclear energy, particularly in the form of small modular reactors (SMRs) has gained renewed attention due to its capacity to deliver continuous, high-density power with minimal direct greenhouse gas emissions. Unlike intermittent renewable sources, nuclear power can provide the stable electricity supply required by AI datacenters operating at gigawatt scale. Several technology companies, in collaboration with government agencies, have announced exploratory investments in SMRs explicitly intended to support datacenter infrastructure, with pilot deployments anticipated later in the decade [6], [7]. Despite these advantages, nuclear integration remains controversial. Challenges related to regulatory approval, long-term waste management, safety assurance, and public acceptance persist, particularly when reactors are colocated with population centers or critical infrastructure. Nevertheless, from a purely energetic perspective, nuclear power remains one of the few viable options capable of supporting sustained AI growth without proportional increases in carbon emissions.

A second major research direction focuses on hardware specialization as a means of improving energy efficiency. Conventional AI datacenters rely heavily on general-purpose GPUs, which, while flexible, are not optimized for all classes of AI workloads. Application-specific integrated circuits (ASICs), neuromorphic processors, optical accelerators, and other non-von-Neumann architectures seek to reduce energy consumption by tailoring hardware design to the computational structure of neural networks. By minimizing data movement and exploiting parallelism at the architectural level, such systems can significantly reduce power draw per operation. Industry and research reports suggest that task-specific accelerators can achieve order-of-magnitude improvements in energy efficiency compared to general-purpose GPUs for selected workloads [9]. For robotics, where perception, planning, and action inference are frequently repeated at scale, these gains could translate into substantial reductions in the upstream energy cost of intelligence.

Bio-inspired and neuromorphic computing paradigms represent a complementary approach grounded in the principles of biological information processing. Spiking neural networks (SNNs), which encode information through sparse, event-driven spikes rather than continuous-valued activations, more closely resemble the operation of biological neurons. This sparse and asynchronous mode of computation enables significantly lower energy consumption, particularly when implemented on neuromorphic hardware. Experimental studies indicate that SNN-based controllers are well suited for robotic tasks such as navigation, sensorimotor integration, and adaptive behavior under strict energy constraints [10]. While current neuromorphic systems remain limited in scale and programmability, they suggest alternative computational models in which intelligent behavior does not require extreme power density or centralized infrastructure.

Quantum computing is often cited as a longer-term possibility for alleviating the computational bottlenecks associated with training and optimizing large AI models. In principle, quantum algorithms could offer exponential or polynomial speedups for specific classes of optimization and learning problems, potentially reducing the energy required for training models relevant to robotics. However, contemporary quantum hardware remains highly experimental, characterized by limited qubit counts, short coherence times, and substantial overhead for error correction. As a result, the practical impact of quantum computing on near-term robotic AI workloads remains speculative, and its role should be regarded as exploratory rather than imminent [11].

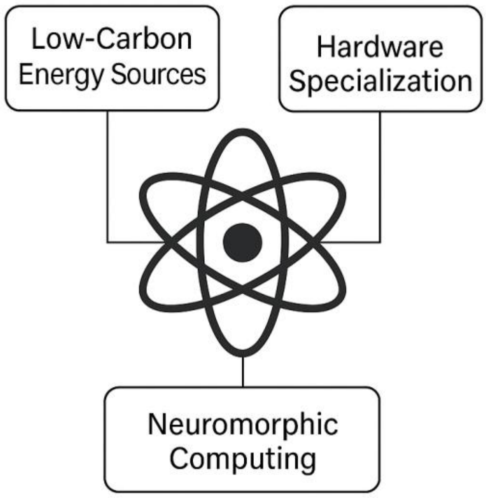

Fig. 4. Solutions of the data center energy problems

Finally, distributed and edge computing approaches aim to relocate portions of AI computation closer to robotic platforms, reducing latency and dependence on centralized datacenters. From an architectural standpoint, this decentralization can improve responsiveness and resilience in robotic systems operating in dynamic environments. However, distribution alone does not inherently reduce total energy consumption. In some cases, shifting computation to edge devices powered by less efficient or more carbon-intensive energy sources may increase overall emissions. Consequently, distributed computing should be viewed primarily as a performance and architectural optimization rather than a standalone sustainability strategy.

In aggregate, these technological pathways highlight the complexity of addressing sustainability in AI-driven robotics. Meaningful reductions in environmental impact are unlikely to result from a single intervention. Instead, progress will depend on the coordinated evolution of energy systems, hardware architectures, and computational paradigms, alongside policy and regulatory frameworks capable of accounting for the externalities of large-scale AI infrastructure.

VII. Discussion and Conclusion

The integration of foundation models into robotics represents a fundamental paradigm shift, enabling unprecedented levels of autonomy and adaptability. However, this progress is underwritten by a rapidly expanding computational infrastructure whose environmental footprint is increasingly difficult to ignore. Rising power densities, escalating water consumption, and gigawatt-scale datacenter projects pose systemic sustainability challenges.

Addressing these challenges requires coordinated advances in energy generation, hardware efficiency, and computational paradigms, as well as policy frameworks that account for the externalities of AI infrastructure. Without such interventions, the ecological costs of intelligent robotics may undermine the societal benefits they promise. Sustainable robotics will therefore depend not only on smarter machines, but also on more responsible ways of powering and cooling the intelligence that drives them.

Acknowledgment

The author acknowledges the financial support of the Project "Green hydrogen-powered eco-collaborative robots" from The Recovery and Sustainability Mechanism for the implementationof investment: PVU-44/2024, BG-RRP-2.017-0011-C01.

References

[1] Ramboll Group, “100+ kW per rack in data centers: The evolution and revolution of power density,” Ramboll Insights, 2024, Online: https://www.ramboll.com/en-us/insights/decarbonise-for-net-zero/100-kw-per-rack-data-centers-evolution-power-density

[2] “The path to power,” Data Center Dynamics, 2025, Online: https://www.datacenterdynamics.com/en/marketwatch/the-path-to-power/

[3] B. Srivathsan et al., “AI power: Expanding data center capacity to meet growing demand”, McKinsey & Company, 2024, Online: https://www.mckinsey.com/industries/technology-media-and-telecommunications/our-insights/ai-power-expanding-data-center-capacity-to-meet-growing-demand

[4] Environmental and Energy Study Institute, “Data centers and water consumption,” EESI Issue Brief, 2025, Online: https://www.eesi.org/articles/view/data-centers-and-water-consumption

[5] M. Fleury, N. Jimenez, “‘I can’t drink the water’: Life next to a US data centre,” BBC News, 2025, Online: https://www.bbc.com/news/articles/cy8gy7lv448o

[6] U.S. Department of Energy, “Advantages and challenges of nuclear-powered data centers,” DOE Energy News, 2025, Online: https://www.energy.gov/ne/articles/advantages-and-challenges-nuclear-powered-data-centers

[7] Deloitte Insights, “Nuclear energy’s role in powering data center growth”, 2025, Online: https://www.deloitte.com/us/en/insights/industry/power-and-utilities/nuclear-energy-powering-data-centers.html

[8] A. Zewe, “Explained: Generative AI’s environmental impact”, MIT News, 2025, Online: https://news.mit.edu/2025/explained-generative-ai-environmental-impact-0117

[9] F. Ferdaus at al, “Evaluating Energy Efficiency of AI Accelerators Using Two MLPerf Benchmarks”, IEEE 25th International Symposium on Cluster, Cloud and Internet Computing (CCGrid) 2025.

DOI: 10.1109/CCGRID64434.2025.00035

[10] R. Guerrero-Criollo et al., “Bio-inspired neural networks for decision-making mechanisms and neuromorphic computing in robotics,” Frontiers in Neurorobotics, vol. 17, 2023, doi: 10.3389/fnbot.2023.1078074.

[11] WisdomTree, “Quantum computing and robotics on the brink of revolution”, WisdomTree Blog, 2025, Online: https://www.wisdomtree.com/investments/blog/2025/01/16/titans-of-tomorrow-quantum-computing-and-robotics-on-the-brink-of-revolution

[12] Wikipedia contributors, “Colossus (supercomputer)”, Wikipedia, accessed 2025, Online: https://en.wikipedia.org/wiki/Colossus_%28supercomputer%29

[14] M. Adams, “OpenAI’s Stargate project will require energy to power a whole city”, Business Insider, 2025. Online: https://www.businessinsider.com/openai-stargate-project-data-center-power-gigawatt-chatgpt-ai-2025-10

[15] G. Buttler, “OpenAI and Oracle to deploy 450,000 GB200 GPUs at Stargate data center,” Data Center Dynamics, 2025, Online: https://www.datacenterdynamics.com/en/news/openai-and-oracle-to-deploy-450000-gb200-gpus-at-stargate-abilene-data-center/

[16] T, Xiao, “Robotics in the era of foundation models,” Agentic Systems, Substack, 2023, Online: https://agentic.substack.com/p/robotics-in-the-era-of-foundation

[17] H. Jeong et al., “A survey of robot intelligence with large language models,” Applied Sciences, vol. 14, no. 19, Art. no. 8868, 2024, doi: 10.3390/app14198868.