DOI: 10.62897/COS2025.3-1.6

10-12, November 2025, Győr

SUSTAINABLE INTELLIGENCE AT THE EDGE: USING SUSTAINABILITY AS A METRIC FOR AI CLOUD-EDGE DESIGN

Mustafa Abdulkadhim[*,a,b], Sandor R. Repas[a]
[a]Széchenyi István University, Department of Telecommunications, Gyor, Hungary
[b]Al-Nahrain University, College of Information Engineering, Baghdad, Iraq
abdulkadhim.mustafa@sze.hu

The swift growth of Cloud-Edge infrastructure to support IoT and artificial intelligence workloads with low latency is exceeding the capacity of our management to control the environmental impact. Whereas there are energy-efficient hardware and software, the measures of sustainability, such as lifecycle assessment (LCA), carbon intensity, and circularity, are inconsistently defined and seldom reported. This piece of work brings sustainability to first-class design and operations measures of Cloud-Edge systems, inquiring how it can be directly incorporated into benchmarking and decision-making. The literature review and cross-venue trend analysis (2018-2025) have been conducted in a structured way to map existing practices and gaps. Based on these results, we suggest the Sustainability-conscious Hybrid Edge Assessment Blueprint (SHEAB): a workflow that integrates automated lifecycle inventory, power/CO2 estimation at workload, and circular economy into what is today taken as performance benchmarking. The review is a revelation of ongoing LCA scope, boundary setting, and reporting gaps; in its turn, SHEAB unifies a small, similar metrics portfolio, e.g., Performance-per-Watt (PPW), CO2e per inference/transaction, and embodied carbon per device, and offers a sustainable evaluation pipeline. Using the blueprint to calculate representative edge workloads and device classes shows end-to-end feasibility and uncovers viable trade-offs between throughput and carbon intensity among deployment options. When sustainability is the key metric, the choices of the system change: hardware is selected, its location is different, and the timeframe is varied when CO2e and circularity can be placed next to latency and price. We end up with reporting guidelines and lessons learned to be put into action to fast-track the truly sustainable intelligent edge systems as the AI demand and its carbon footprint continue to increase.

1. Introduction

The need to process data immediately close to data sources, which mostly arises in the IoT, autonomous systems, and real-time analytics, has helped the growth of edge computing. Though edge computing is vastly better in terms of latency, bandwidth, and responsiveness, its ecological impact, especially in the context of energy usage and carbon footprint, is to be considered thoroughly. Conventional Cloud-based solutions do not take sustainability metrics into consideration, which inevitably causes possibly severe ecological effects as edge devices become massive. This research aims to systematically explore sustainability metrics as crucial factors in designing cloud-edge infrastructures. We assert that adopting sustainability metrics as core design principles can profoundly affect energy use, operational efficiency, and environmental responsibility, thus addressing global ecological goals. This research aims to focus on sustainable computing by explicitly addressing the sustainability gaps in edge and cloud computing. This paper aims to provide a pivotal foundation for future research and practical application in sustainable intelligent edge computing.

2. Detailed literature analysis

Kim et al. (2023) describe the development of an open-source, microservice-based MLOps platform. This platform aims to facilitate the efficient development and deployment of AI services within an edge-cloud environment. Although the paper introduced a novel platform, the research was obviously lacking in the consideration of sustainability and CO2 footprint considerations for ML workload.

Abdulkadhim and Repas (2025), in their paper titled "SHEAB: A Novel Automated Benchmarking Framework for Edge AI", introduced a novel benchmarking framework for automated testing for edge AI-capable hardware. Later, they introduced a novel Large Language model assessment tool for the edge (Abdulkadhim and Repas, 2026).

Nezami et al. (2025), in their paper titled "Descriptor: Benchmark Dataset for Generative AI on Edge Devices (BeDGED)", introduced a new benchmarking framework for edge devices. They have implemented it on a Raspberry Pi 5. Zhang et al. (2024) research focuses on end-to-end workflows and tooling to enable reproducible builds, scalable testing, and automated rollouts across heterogeneous clusters and robots.

Welagedara et al. (2021) in their research designed an edge-intelligence collaborative learning system for resource-constrained IoT devices, selecting partitioned model training after reviewing existing approaches.

Seisa et al. (2022) compare two edge setups for remote Model Predictive Control (MPC) of a UAV: a Docker-only design and a Kubernetes-orchestrated design, both using ROS for closed-loop control. evaluation centers on latency/round-trip time and CPU utilization; there's no power/energy measurement, CO_2 accounting. Nkenyereye et al. (2023) in their research propose a deep-reinforcement-learning framework to manage containerized edge-intelligence inference requests across heterogeneous IoT edge nodes.

Uses an actor–critic agent (two critics + a policy network) to adaptively throttle/route inference requests so the edge-computing device isn't overloaded while maintaining service success. Nevertheless, their research centers on QoS and overload avoidance; it does not report energy or carbon metrics or any sustainability considerations.

Heckmann and Ravindran (2023), in their research titled "Evaluating Kubernetes at the Edge for Fault Tolerant Multi-Camera Computer Vision Applications" Evaluates K3s (lightweight Kubernetes) on a Raspberry Pi–based edge cluster running a multi-camera computer-vision gateway (VEI + NATS).

Jhingran et al. (2024) propose a decentralized deployment of generative-AI models using a microservices architecture that runs across cloud/edge, coordinated via API gateways and container platforms like

Kubernetes/Docker to boost scalability and fault tolerance. Kum et al. (2024) propose an edge-AI management framework for smart farms that uses cloud–edge orchestration to deploy AI services across geographically dispersed fields. Motivates vision-based tasks (growth monitoring, pest/disease ID) that need GPU-class compute, and targets rural constraints (limited connectivity/resources).

Karam and Abdulkadhim (2022) proposed an IoRT "Internet of robotic things" implementation for IoT in robotics that is fuzzy-based. Abdulkadhim and Hassan (2021) suggested an implementation of network security at the network edge; their research contributes to the top security research. Al-Kadhimi and Abdulkadhim (2021) proposed a queuing-based model for the analysis of customers in shopping malls. This kind of industrial edge computing contributes to the fact that edge computing might be applied to different types of businesses. The previous research focuses on the framework and deployment with no consideration of the sustainability or CO2-related emissions. A brief comparison and highlight of the research gap were added in Table 1.

Table 1: Literature review that introduces sustainability as a metric

Research	Architectural Focus	Energy Efficiency	Hardware Optimization	AI Integration	Sustainability Metrics
Kim et al., 2023	High	Moderate	Moderate	High	Low
Nezami et al., 2025	Moderate	High	Moderate	High	Moderate
Zhang et al., 2024	High	Moderate	High	High	Moderate
Welagedara et al., 2021	Moderate	High	Moderate	High	Moderate
Szántó et al., 2022	Moderate	High	High	Moderate	Low
Seisa et al., 2022	High	Moderate	Moderate	Moderate	Moderate
Nkenyereye et al., 2023	High	Moderate	High	High	Low
Heckmann and Ravindran, 2023	High	Moderate	High	Moderate	Low
Jhingran et al., 2024	Moderate	Moderate	High	High	Low
Kum et al., 2024	Moderate	Moderate	Moderate	High	Moderate

3. Contribution

This paper makes four contributions. First, we develop an automated lifecycle-assessment framework covering manufacturing through end-of-life for edge devices. Second, we present a carbon-footprint tracking system that visualizes real-time emissions of edge deployments. Third, we integrate circular-economy principles by outlining strategies for reuse, refurbishment, and recycling. Fourth, we introduce standardized, reproducible sustainability benchmarks and metrics for realistic edge-computing scenarios. We detail these in Sections 4 - 6. By incorporating these metrics into the core design processes, cloud-edge computing can be optimized not only for performance but also for environmental sustainability. Below are the details for each suggested framework.

4. Lifecycle Assessment Framework

The proposed Lifecycle Assessment (LCA) Framework for edge computing devices aims to systematically evaluate and minimize the environmental impacts associated with all stages of their life cycle, including manufacturing, deployment, operational usage, and end-of-life management. Key Elements of the Lifecycle Assessment Framework, and the framework is better explained in Figure 1.

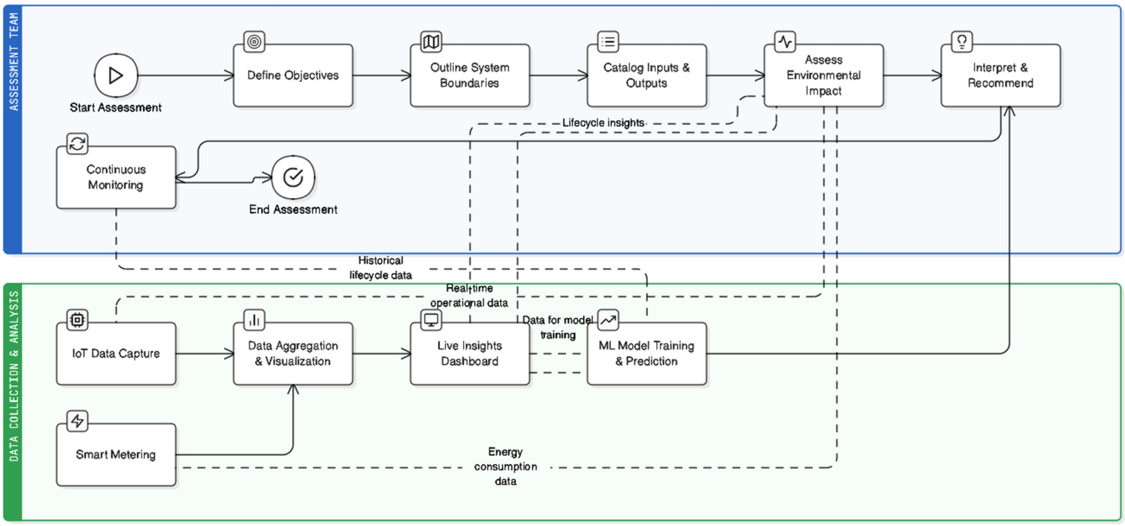

Figure 1: The proposed Lifecycle Assessment framework

4.1 Scope and Goal Definition

Identify and clearly define the objectives of the assessment.
Outline system boundaries, such as specific edge hardware, geographic locations, and intended applications.

4.2 Inventory Analysis

List every input and output on the lifecycle of the product, including, but not limited to, raw-material extraction, manufacturing processes, transport logistics, energy used in operation, maintenance and upgrades, and end-of-life processes, such as disposal or recycling.

4.3 Impact Assessment

Quantify the environmental impact in terms of carbon footprint (CO_2 emissions), energy use (joules, kWh), resource depletion (raw materials and rare metals), toxicity and chemical hazards (hazardous substances), and waste generation (electronic and hazardous waste).

4.4 Automated Data Collection and Analysis

Take advantage of IoT sensors and smart metering to record real-time data when the device is used, use automated software tools to keep on aggregation, visualization, and analysis of data, and make live lifecycle insights available through interactive dashboards.

4.5 Interpretation and Recommendations

Generate actionable insights that pinpoint opportunities to improve sustainability, and recommend alternative practices—such as hardware reuse, optimized maintenance scheduling, and integrating renewable energy—to achieve measurable gains.

4.6 Continuous Monitoring and Improvement

Regularly reassess the lifecycle effects and compare the progress with the past baselines and peers, and use the history of the past to train machine-learned models that forecast hotspots and prescribe interventions, allowing for optimization of future sustainability performance in advance.

5. Key Methodologies for Carbon Footprint Tracking

The proposed system leverages several key methodologies to accurately measure and dynamically manage the carbon footprint:

5.1 Real-Time Data Collection

Adopt sensor-enabled energy monitoring by installing smart IoT sensors and energy meters on edge devices to monitor real-time power usage, both computational (CPU, GPU, FPGA) and non-computational (cooling and network infrastructure) components; and monitor network communications by monitoring data flows and measuring the energy used by data transfers between edge nodes, gateways, and cloud data centers.

5.2 Dynamic Emission Factor Integration

Incorporate locality-aware emissions by integrating regional grid emission factors—from utilities or open databases—to reflect the actual energy mix (coal, gas, renewables) and complement this with renewable-mix accounting that measures, in real time where possible, the share of onsite renewables (e.g., solar, wind), enabling more precise and defensible emissions calculations.

5.3 Automated Carbon Accounting Engine

Implement continuous, low-latency emission accounting by applying automated algorithms that combine real-time energy-use data with dynamically updated grid-emission factors to produce ongoing carbon estimates; in parallel, use predictive analytics and machine-learning models to forecast future emissions from historical patterns, enabling proactive energy-management strategies.

5.4 Visualization and Dashboard

Deliver real-time carbon footprints, trend lines, cross-deployment or cluster comparisons with interactive, real-time dashboards and match with actionable analytics and alerts: alert the stakeholders when the carbon footprint goes beyond predefined limits, providing a notification, and also specific recommendations such as moving workloads, adjusting resource allocation, or optimization of hardware utilization so they can take immediate corrective action. Figure 2 presents one of the available CO2 footprint tracking.

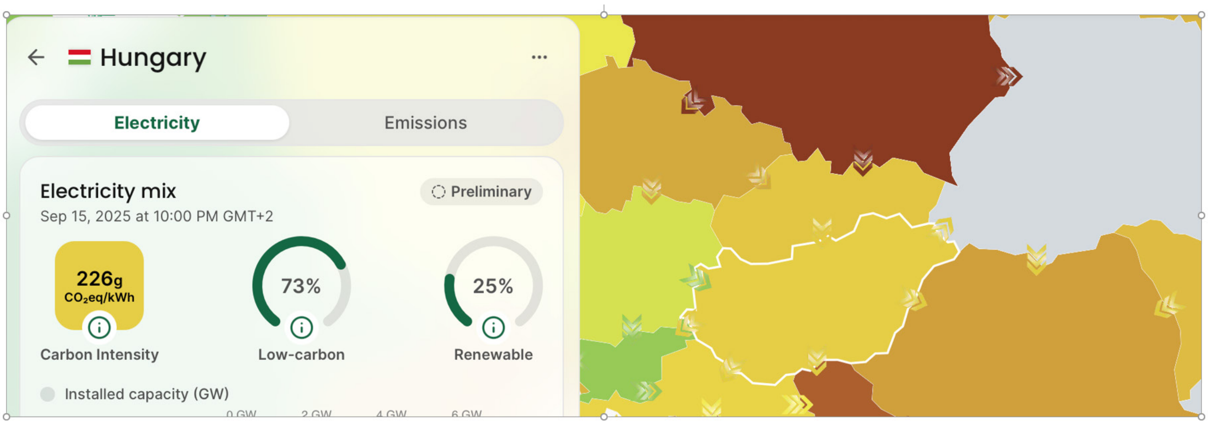

Figure 2: Carbon footprint map of Hungary

5.5 Integration with Energy Management Systems

Enable automated control by integrating the carbon-tracking system with energy-management platforms so carbon metrics can directly trigger interventions—such as workload balancing, device throttling, or staged activation/deactivation—and complement this with carbon budgeting and reporting that automatically generates standardized, auditable summaries of historical emissions, compliance with environmental standards, and progress toward sustainability goals.

6. Benchmark Development for Sustainability

The establishment of elaborate benchmarks with a sustainability focus is a crucial activity in relation to gauging and comparing the impact of different edge infrastructures on the environment objectively. The sustainability benchmarks would quantify and equalize the performance of various ecological parameters so as to offer useful information to focus on enhancement and to be able to comply with the environmental standards.
Introduced in Figure 3 was an example benchmarking of edge devices following the execution of Large Language Models and their energy consumption.

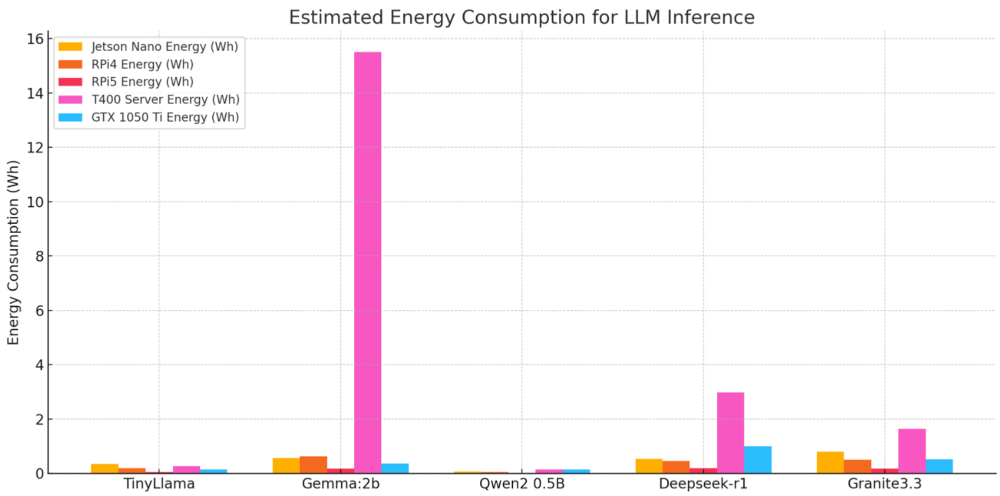

Figure 3: Estimated energy consumption for LLM Inference

The Key Components of the Sustainability Benchmark Framework are as follows:

6.1 Energy Efficiency Metrics

Evaluate energy consumption across idle, peak, and typical utilization levels, and quantify performance per watt (PPW) to measure how efficiently each device converts power into computational work.

6.2 Carbon Footprint Metrics

Measure CO2 emissions per operation of typical computational operations and include the carbon intensity of the underlying source of energy, including regional grid mixes (renewable vs. non-renewable), to obtain the real impact on the environment.

6.3 Lifecycle Sustainability Metrics

Assess the environmental impacts of sourcing raw materials, and account for embodied carbon—the total greenhouse-gas emissions embedded in manufacturing, transportation, and installation—to capture the full upstream footprint of the system.

6.4 Operational Sustainability Metrics

Measure utilization rates to determine how effectively edge devices are employed—reducing idle capacity and resource waste—while simultaneously quantifying the volume and composition of operational waste (electronic, packaging, and other streams) to guide reduction, reuse, and responsible disposal.

6.5 End-of-Life (EoL) Sustainability Metrics

Calculate a recyclability index by estimating the share of components and materials that can be recovered or recycled, and, in parallel, evaluate hazardous-materials management practices to verify safe handling, storage, and end-of-life disposal.

7. Conclusion and Future Work

Measures of sustainability have become pertinent yet under-investigated factors in the development of intelligent edge computing-based infrastructures. Automated lifecycle assessment, intensive carbon footprint monitoring, and circular economy models are inevitable components of real sustainability in edge computing. In the analysis, it is evident that although performance optimization is still at the forefront of the research agenda, approaches that focus on sustainability are much needed. The avenues of future research should be aimed at incorporating these sustainability measures into the framework of the standardized design to maintain a comprehensive balance between technological development and environmental sustainability.

The constant evolution and testing of these frameworks will offer effective instruments to the researchers and industry practitioners to assess and improve the sustainability of cloud-edge deployments, which will eventually lead to sustainable intelligence at the edge.

References

- Abdulkadhim M., Hassan S., 2021, Boosting the Network Performance using Two Security Measure Scenarios for Service Provider Network, https://doi.org/10.24996/ijs.2021.SI.1.24.
- Abdulkadhim M., Repas S.R., 2025, SHEAB: A Novel Automated Benchmarking Framework for Edge AI. Technologies, 13(11), 515, https://doi.org/10.3390/technologies13110515.
- Abdulkadhim M., Repas S.R., 2026, Introducing LEAF: LLM Edge Assessment Framework for Generative AI on the Edge. Machine Learning and Knowledge Extraction, 8(2), 48, https://doi.org/10.3390/make8020048.
- Al-Kadhimi A., Abdulkadhim M., 2021, Queueing model analysis of shopping malls in COVID-19 pandemic era: a case study. International Journal of Advanced Technology and Engineering Exploration, 8(85),1545-1556, https://doi.org/10.19101/IJATEE.2021.874672.
- Heckmann O., Ravindran A., 2023, Evaluating Kubernetes at the Edge for Fault Tolerant Multi-Camera Computer Vision Applications. 2023 IEEE/ACM 23rd International Symposium on Cluster, Cloud and Internet Computing Workshops (CCGridW), 269–271, https://doi.org/10.1109/CCGridW59191.2023.00054.
- Jhingran S., Bansal N., Chaturvedi R., Singh A., Arora Y., 2024, Decentralized Generative AI Model Deployment Using Microservices. 2024 International Conference on Artificial Intelligence and Quantum Computation-Based Sensor Application (ICAIQSA), https://doi.org/10.1109/ICAIQSA64000.2024.10882424.
- Karam Z.A., Abdulkadhim M., 2022, Design, Implementation, Interfacing and Control of Internet of Robot Things for Assisting Robot. International Journal of Computing and Digital Systems, 11, 388–401, https://doi.org/10.12785/ijcds/110132
- Kim C., Kim G.-Y., Kim S., 2023, A Microservice-based MLOps Platform for Efficient Development of AI Services in an Edge-Cloud Environment. 2023 14th International Conference on Information and Communication Technology Convergence (ICTC), 1507–1509, https://doi.org/10.1109/ICTC58733.2023.10392296.
- Kum S., Oh S., Moon J., 2024. Edge AI Framework for Large Scale Smart Agriculture, in: 2024 27th Conference on Innovation in Clouds, Internet and Networks (ICIN). Presented at the 2024 27th Conference on Innovation in Clouds, Internet and Networks (ICIN), IEEE, Paris, France, 143–147, https://doi.org/10.1109/ICIN60470.2024.10494451
- Nezami Z., Hafeez M., Djemame K., Zaidi S.A.R., Xu J., 2025, Descriptor: Benchmark Dataset for Generative AI on Edge Devices (BeDGED). IEEE Data Descriptions, 2, 50–55, https://doi.org/10.1109/IEEEDATA.2025.3552083.
- Nkenyereye L., Baeg K.-J., Chung W.-Y., 2023, Deep Reinforcement Learning for Containerized Edge Intelligence Inference Request Processing in IoT Edge Computing. IEEE Transac5rt 2aqrde3tions on Services Computing, 16(6), 4328–4344, https://doi/org/10.1109/TSC.2023.3320752.
- Seisa A.S., Satpute S.G., Nikolakopoulos G., 2022, Comparison between Docker and Kubernetes-based Edge Architectures for Enabling Remote Model Predictive Control for Aerial Robots. IECON 2022 – 48th Annual Conference of the IEEE Industrial Electronics Society, 1–6, https://doi.org/10.1109/IECON49645.2022.9968933.
- Szántó P., Kiss T., Sipos K.J., 2022, Energy-efficient AI at the Edge. 2022 11th Mediterranean Conference on Embedded Computing (MECO), 650–655, https://doi.org/10.1109/MECO55406.2022.9797178.
- Welagedara L., Harischandra J., Jayawardene N., 2021, Edge Intelligence Based Collaborative Learning System for IoT Edge. 2021 IEEE 12th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON), 667–672, https://doi.org/10.1109/IEMCON53756.2021.9623215.
- Zhang Y., Pasch F., Mirus F., Scholl K.-U., Wurll C., Hein B., 2024, Conquering the Robotic Software Development Cycle at Scale: Using KubeROS from Simulation to Real-world Deployment. 2024 IEEE 20th International Conference on Automation Science and Engineering (CASE), 1503–1509, https://doi.org/10.1109/CASE59546.2024.10711591.