

From Solar Panels to AI Decisions: Intelligent Server Utilization for Sustainable Computing [†]

Nikolaos Fragkos , Stylianos Katsoulis , Evangelos Nannos , Fotios Zantalis ,
Ioannis Chrysovalantis Panagou , Panagiotis Tsiakas  and Grigorios Koulouras * 

TelSiP Research Laboratory, Department of Electrical and Electronic Engineering, School of Engineering, University of West Attica, Ancient Olive Grove Campus, 250 Thivon Str., GR-12241 Athens, Greece; eee20387277@uniwa.gr (N.F.); skatsoulis@uniwa.gr (S.K.); enannos@uniwa.gr (E.N.); fzantalis@uniwa.gr (F.Z.); ipanagou@uniwa.gr (I.C.P.); ptsiakas@uniwa.gr (P.T.)

* Correspondence: gregkoul@uniwa.gr; Tel.: +30-210-53851647

[†] Presented at the 1st International Online Conference on Designs, 9–10 February 2026; Available online: <https://sciforum.net/event/Designs2026>.

Abstract

Renewable integration is increasingly important for sustainable off-grid computing. The inherent variability of solar output frequently produces unusable midday surpluses. Leveraging recent Artificial Intelligence (AI) advances and established literature, we evaluate an AI-driven demand-response framework for scaling Large Language Models (LLMs) training servers using real-time solar energy data, Solcast forecasts, and battery storage records collected from Battery Management Systems (BMS), Maximum Power Point Tracking (MPPT) units, and smart inverters. An n8n AI Agent using the Ollama chat model gpt-oss:20b assesses surplus solar energy, activating selected servers to utilize otherwise wasted capacity. Workloads consistently align with solar availability, demonstrating 99% operational reliability, sub-second responsiveness, and accurate surplus-energy detection. This research demonstrates how Artificial Intelligence can repurpose surplus solar output into usable computational capacity, thereby contributing to a broader transition toward renewable-powered infrastructures.

Keywords: sustainable computing; AI-powered demand response; solar power; renewable energy data monitoring; artificial intelligence; AI; surplus energy utilization; energy optimization; large language models; LLMs; server scaling; n8n automation

1. Introduction

The continuous evolution of technology and drastic digitalization of everyday processes have led to a rapid increase in computational workloads, especially when considering high amounts of Artificial Intelligence usage. Data-intensive tasks such as LLM training have significantly increased the energy demands of modern digital infrastructures. From data centers and edge computing systems to distributed home-lab environments, all contribute to escalating electricity consumption, raising concerns about both the environment and operational costs. The first of these already accounted for around five percent of the U.S. electricity usage share of 2023 [1], which is projected to multiply worldwide carbon emissions by 8% in the coming years [2,3]. While this may seem devastating for the sustainable future of the planet, advancements in renewable energy technologies such as solar systems have at the same time created opportunities for contributing to non-polluting ways



Academic Editor: Tommaso Bragatto

Published: 25 June 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

of meeting these high energy requirements. However, Photovoltaic (PV) technology introduces some substantial challenges when it comes to power management. A persistent issue is the occurrence of energy surplus, especially at midday, which originates when power generation and battery storage often exceed demand during these hours due to sunny weather; this can range from minor low-impact problems such as the lack of energy exportation and accelerated battery aging in off-grid or standalone systems installations to technical and economic challenges of paramount importance in enterprise-level grid-connected structures. These burdens may result in higher maintenance costs or even total failure of the system, potentially causing power blackouts. In this case, operators face the risk of enforcing curtailments [4], as seen in California's grid operator, which reduced 3.4 million MWh of renewable power in 2024 (a 29% increase from 2023), with solar energy making up 93% of the reduced output [5,6]. These limitations have motivated researchers to explore efficient solutions from simple load-shifting techniques to AI-enhanced resource allocation. Recent work has highlighted the ability to absorb renewable energy supply fluctuations through flexible data center workloads, predicting short-term Photovoltaic output [7] and performing power usage adjustments by dynamically modulating IT processes without violating service quality [8]. Big Tech corporations such as Google also seem to follow this model, implementing the timing of certain computing tasks to coincide with periods of abundant renewable generation on the grid [9], leading at the same time to the reduction of their carbon emissions. Recent developments in artificial intelligence and workflow automation present new opportunities for creating AI-enhanced decision systems to support demand-response strategies, offering real-time adaptability and predictive reasoning based on heterogeneous data sources in response to renewable energy conditions. Therefore, the purpose of this study is to design an AI-enabled demand-response system that dynamically scales computational workloads based on real-time solar energy availability, examining the feasibility of solar power balancing in off-grid small-scale data centers with the help of components that inherently produce rich telemetry data ideal for AI interpretation. We introduce a practical implementation using a Tools AI Agent Node and the Ollama gpt-oss:20b model on the n8n automation platform. Solcast real-time photovoltaic energy data along with battery metrics from BMS and MPPT units are processed by the agent to determine the current operational state and available renewable energy capacity. When surplus conditions are detected, the system autonomously activates or shuts down custom RAM-only provisioned servers optimized for LLM execution, effectively repurposing excess solar output into computational work. The proposed framework is also applicable to grid-connected research laboratories and home-based infrastructures seeking to maximize renewable use.

2. Methods

2.1. System Orchestration

At the core of the system lies an automation pipeline built using n8n (v2.2.4), an open-source low-code workflow platform configured locally as a Docker Engine (v29.1.3) container. The pipeline executes on a scheduled basis every 30 min and includes sequential operations for data ingestion, solar forecast querying, AI-based reasoning, and control execution. The workflow is modular, allowing for node-specific debugging and reconfiguration without disrupting the entire pipeline. Real-time system state is maintained within n8n's execution context, enabling adaptive logic flow.

The target server is provisioned with a custom RAM-only operating system that is offered via network through a Preboot Execution Environment (PXE). As a stateless implementation, all the needed configuration is provided by means of Ansible playbooks, reforming the system's mandatory setup. This infrastructure retains the ability to handle

heavy computational workloads such as those required for LLM training while operating with high reliability, superior performance, and substantially increased speed.

The overall system supports both local and distributed server orchestration and is designed to run continuously without manual intervention. Its architecture also allows for seamless scalability, enabling the integration of additional workflows or services as operational requirements evolve.

2.2. Energy Infrastructure and Solar Forecasting

The system is organized into three functional layers which collect data from multiple interconnected energy infrastructure components, as illustrated in Figure 1. Solar generation is provided by two photovoltaic arrays, each with a capacity of 3660 Wp, connected to separate Victron SmartSolar 250/100 MPPT (Victron Energy, Almere, The Netherlands) units that regulate the DC output. The energy storage subsystem features a 51.2 V/628 Ah LiFePO₄ battery bank monitored by a Victron Smart Shunt (Victron Energy, Almere, The Netherlands), with a Victron Lynx Power In (Victron Energy, Almere, The Netherlands) unit installed to act as a centralized DC busbar for future scaling. Power distribution is handled by a Victron Lynx Distributor (Victron Energy, Almere, The Netherlands) that serves as the central DC interconnection point of the system, directly linking the outputs of the MPPT charge controllers and the battery subsystem to the DC input of a Victron MultiPlus II 48/8000 inverter (Victron Energy, Almere, The Netherlands). The latter supplies AC power to the computing infrastructure and is rated for a continuous output power of 6400 W at 25 °C, 5500 W at 40 °C, and 4000 W at 65 °C. A Victron Cerbo GX (Victron Energy, Almere, The Netherlands) unit is explicitly integrated as the local monitoring and communication gateway, aggregating real-time measurements from the BMS, MPPT controllers, inverter, and Lynx distribution modules. These are made them available through the Victron Remote Management (VRM) platform API, enabling consistent and reliable data acquisition for the AI decision engine.

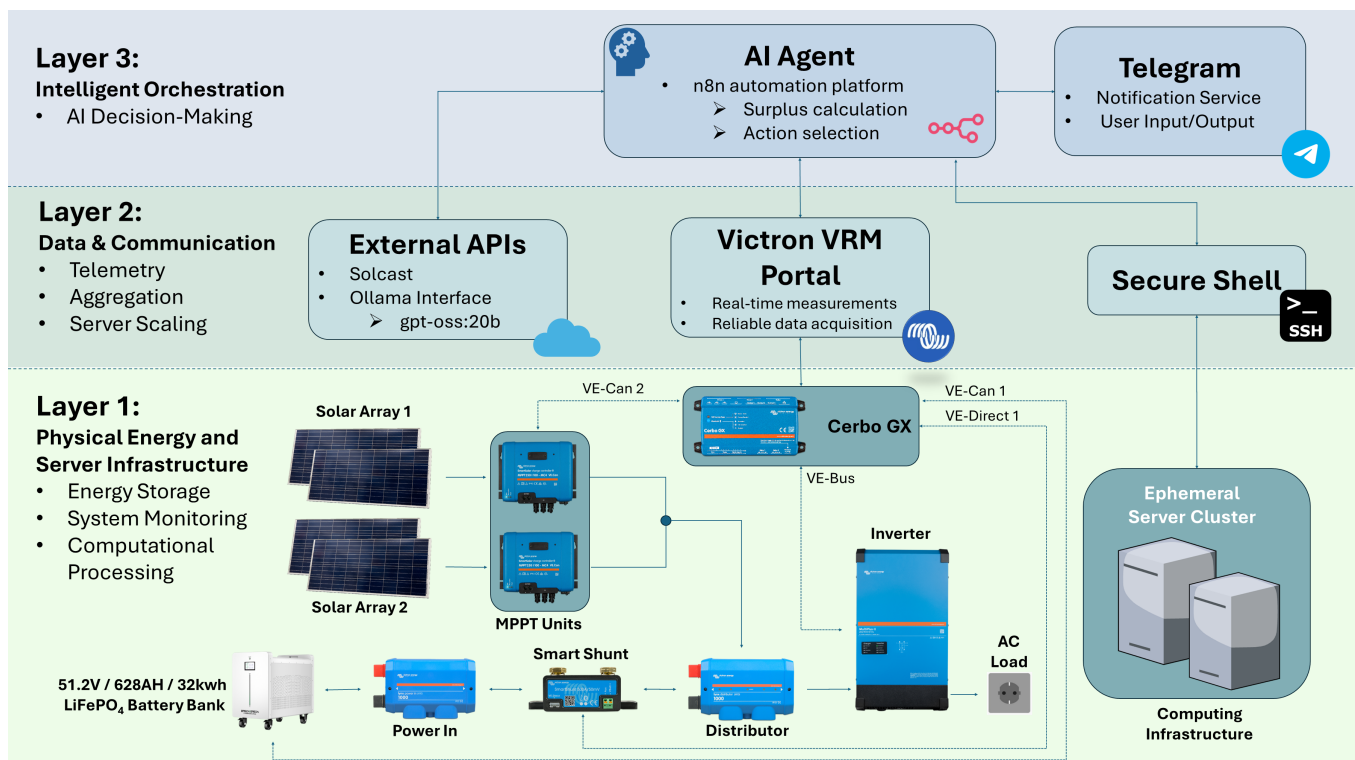


Figure 1. Layered architecture of the proposed framework, illustrating the flow from energy monitoring and BMS telemetry to AI orchestration.

In parallel, high-resolution short-term solar forecasts are retrieved via the Solcast API. Leveraging the “Home User” SOLAR FORECAST plan, we set our system’s location, adjusting the necessary parameters as follows:

- Capacity AC in kW: 5 kW
- Capacity DC in kW: 8 kW
- Azimuth: 0 degrees
- Tilt (horizontal): 30 degrees
- Efficiency factor: 90%

As a result, forecast parameters include real-time total PV power output estimates, shown as 30-min averages, along with forecasted data extending from the current day up to three days ahead. This external forecast data can be accessed in JSON format, which can be merged with live energy readings to create a comprehensive basis for predictive decision-making.

2.3. AI Decision Engine

The core of the decision-making mechanism is implemented as a Tools AI Agent Node configured within the n8n workflow platform. Its interface is provided by selecting a local instance of the gpt-oss:20b language model executed through the Ollama runtime engine (v0.13.5), which provides the LLM API layer for the agent. Instructions are included inside the prompt and system message, guiding the agent to access the Solcast API. The data are filtered to today’s date, summarizing recent energy conditions and forecasting trends. The final results are presented as two JSON objects, “surplus” and “explanation”, using the Output Parser’s extended function. This AI decision engine calculates the possibility of exceeding the upper limit of 80% amount of the 32 kWh battery storage, deciding the appearance of surplus energy. In this case, the identically named object is assigned the value “1”. If there is no adequate solar production due to bad weather conditions or if the energy generation is likely to decline, thereby risking the battery percentage falling below 20%, the surplus accumulation cannot be considered feasible, and a value of “0” is assigned. It is worth mentioning that the implementation also takes into consideration the load consumption of the 300 W server. In any case, an additional summary of the decision is provided within the corresponding “explanation” object. The complete output is then parsed by downstream nodes and routed to the appropriate control module. The AI agent is also integrated with Telegram, which acts as a lightweight input/output interface and notification service. It reports surplus energy decisions and server state actions while supporting simple interaction with the orchestration workflow.

2.4. Server Orchestration and Control Logic

The most important component of the given JSON output is the value of the “surplus” variable. Taking this into account, an algorithm is developed in a JavaScript Code node on the n8n environment, determining the execution path of each outcome. Once the AI agent issues power abundance, returning “1” as output, a Wake-on-LAN (WoL) magic packet is sent to the selected MAC address, activating the ephemeral server. Since WoL is not provided as a pre-installed software package, it was manually installed on the n8n host machine. Conversely, a shutdown command is performed to add the essential private key credentials when solar energy and stored battery power are regarded as insufficient. Both operations are performed via Secure Shell (SSH) connections, established from the n8n container to the host machine and to the activated server, respectively, after passing through a simple “if” function unit for each case, driven by JavaScript-based forwarding logic, as shown in Figure 2.

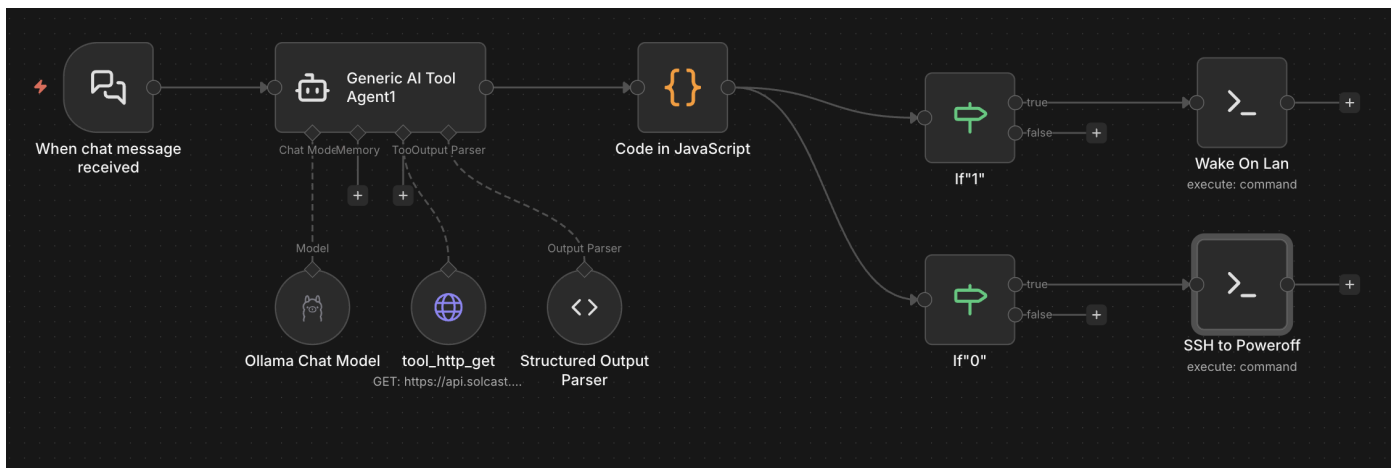


Figure 2. Implementation logic and communication workflow of the AI Agent within the n8n automation platform.

3. Results and Discussion

From data input to command execution, the end-to-end decision-making process exhibited modest latency, primarily attributed to hardware constraints impacting AI inference performance. Once the agent (including the output parser unit) is executed, the remaining nodes occur instantly. The compact prompting format and the lightweight nature of the “Control Logic” unit contributed to improving determinism to the extent that was feasible within our setup, with only minor delays that are negligible in the context of a non-commercial data center. The activated compute nodes reliably executed workloads during surplus intervals and powered down when surplus conditions diminished. WoL and SSH connections were established successfully in 99% of cases, demonstrating that the system achieved its primary objective. Failure cases were limited to external factors, namely, insufficient power supply or loss of network connectivity.

In addition, the ephemeral nature of the server cluster introduces a provisioning latency before the activated compute nodes become fully operational. Specifically, the PXE-based boot process required approximately one minute for the operating system to be loaded over the network. After boot completion, software provisioning and configuration through Ansible required an additional two to three minutes, depending on the assigned server role and the corresponding playbook complexity. Consequently, the total activation time from a positive surplus decision to a fully provisioned compute node was typically in the range of three to four minutes. This delay is inherent to the stateless design of the infrastructure, since each activated server is intentionally reconstructed at runtime rather than maintained as a permanently configured instance.

Regarding the AI agent within the previously described end-to-end process, the observed response time typically ranged from one to two minutes, covering telemetry retrieval, forecast data processing, and local model reasoning. However, this duration does not constitute a strict limitation for the examined use case, since solar surplus detection and battery-state forecasting operate over gradual energy-variation timescales rather than requiring instantaneous actuation. The workflow is executed at 30-min intervals, while photovoltaic generation trends and battery state-of-charge variations evolve over substantially longer time scales than the observed decision latency. Accordingly, the observed response interval with an approximate duration of one to two minutes can be considered acceptable for renewable-aware workload scheduling, particularly in the context of surplus solar utilization and battery protection.

To assess decision quality, predictions from the AI agent were compared against actual solar generation curves and system load patterns leveraged from the Victron Software,

VRM App (v3.0.7). The model correctly identified imminent surplus conditions in the vast majority of cases, with accurate “power-on” decisions occurring whenever the battery’s state of charge could reach its maximum threshold due to excessive PV output. Likewise, a comparable pattern was observed for the power-off condition, where no surplus energy production was detected. The results are presented in Tables 1 and 2, which summarize ten activation events for each server-scaling scenario.

Table 1. Ten consecutive runs on a sunny day.

Execution Runs	Surplus_Tag	Server-State Command
1	“1”	WoL
2	“1”	WoL
3	“1”	WoL
4	“1”	WoL
5	“1”	WoL
6	“1”	WoL
7	“1”	WoL
8	“0”	Poweroff
9	“1”	WoL
10	“1”	WoL

Table 2. Ten consecutive runs on a cloudy day.

Execution Runs	Surplus_Tag	Server-State Command
1	“0”	Poweroff
2	“0”	Poweroff
3	“0”	Poweroff
4	“0”	Poweroff
5	“1”	WoL
6	“0”	Poweroff
7	“0”	Poweroff
8	“0”	Poweroff
9	“0”	Poweroff
10	“0”	Poweroff

It should also be noted that the presented runs are intended to provide an initial functional evaluation of the proposed workflow rather than a complete statistical characterization of all possible operating conditions. Although the number of runs is limited, the results offer a useful indication of the system’s behavior under representative operating cases. A more extensive evaluation over longer seasonal periods would be required in order to quantify the impact of intermediate weather conditions, forecast uncertainty, and variable computational load on the long-term decision accuracy of the system.

4. Conclusions

The presented approach offers a low-cost modular solution adaptable to both off-grid micro-data centers and grid-connected facilities, aiming to maximize clean-energy utilization. Across all evaluated scenarios, the proposed architecture consistently aligned workload activation with surplus solar availability. By integrating real-time solar telemetry, short-term photovoltaic forecasts, and intelligent decision-making through a local LLM agent, the system successfully repurposed midday surplus energy into usable compute capacity. Importantly, the framework operated fully autonomously, requiring no manual intervention while maintaining stable and consistent behavior throughout extended testing.

Beyond its technical performance, this research highlights the ability to intelligently synchronize intensive tasks with periods of solar energy abundance. With the contribution of Artificial Intelligence, the development of renewable-aware infrastructures constitutes not only a dynamic load stabilization tool but also a beneficial application for computing tasks. Future work may introduce more improved system mechanisms by using advanced learning methods, helping the agents to make better decisions over time. Additionally, the current study could be expanded to use a multi-server workflow with dynamic selection of which instances to activate, prioritizing those that have not been used previously.

Author Contributions: Conceptualization, N.F., S.K., E.N., F.Z., I.C.P., P.T., and G.K.; methodology, N.F., S.K., E.N., F.Z., I.C.P., P.T., and G.K.; software, N.F., S.K., E.N., F.Z., I.C.P., P.T., and G.K.; validation, N.F., S.K., E.N., F.Z., I.C.P., P.T., and G.K.; formal analysis, N.F., S.K., E.N., F.Z., I.C.P., P.T., and G.K.; investigation, N.F., S.K., E.N., F.Z., I.C.P., P.T., and G.K.; resources, N.F., S.K., E.N., F.Z., I.C.P., P.T., and G.K.; data curation, N.F., S.K., E.N., F.Z., I.C.P., P.T., and G.K.; writing original draft preparation, N.F., S.K., E.N., F.Z., I.C.P., P.T., and G.K.; writing—review and editing, N.F., S.K., E.N., F.Z., I.C.P., P.T., and G.K.; visualization, N.F., S.K., E.N., F.Z., I.C.P., P.T., and G.K.; supervision, G.K.; project administration, G.K. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: Data are contained within the article.

Acknowledgments: During manuscript preparation, the authors used the Gemini 3.1 assistant solely for basic editorial support. The tool was used to check grammar, sentence structure, spelling, and formatting consistency. The authors have reviewed and edited the output and take full responsibility for the content of this publication.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

AC	Alternating Current
BMS	Battery Management System
DC	Direct Current
LLM	Large Language Model
MPPT	Maximum Power Point Tracking
PV	Photovoltaic
PXE	Preboot Execution Environment
RAM	Random Access Memory
SSH	Secure Shell
VRM	Victron Remote Management
WoL	Wake-on-LAN

References

1. Shehabi, A.; Newkirk, A.; Smith, S.; Hubbard, A.; Lei, N.; Siddik, M.A.; Holecek, B.; Koomey, J.; Masanet, E.; Sartor, D. 2024 *United States Data Center Energy Usage Report*; Technical Report LBNL-2001637; Lawrence Berkeley National Laboratory: Berkeley, CA, USA, 2024. [\[CrossRef\]](#)
2. Andrae, A.S.; Edler, T. On global electricity usage of communication technology: Trends to 2030. *Challenges* **2015**, *6*, 117–157. [\[CrossRef\]](#)
3. Gupta, U.; Elgamal, M.; Hills, G.; Wei, G.Y.; Lee, H.H.S.; Brooks, D.; Wu, C.J. ACT: Designing sustainable computer systems with an architectural carbon modeling tool. In *Proceedings of the 49th Annual International Symposium on Computer Architecture*; ACM: New York, NY, USA, 2022; pp. 784–799. [\[CrossRef\]](#)
4. Rad, M.A.V.; Kasaeian, A.; Niu, X.; Zhang, K.; Mahian, O. Excess electricity problem in off-grid hybrid renewable energy systems: A comprehensive review from challenges to prevalent solutions. *Renew. Energy* **2023**, *212*, 538–560. [\[CrossRef\]](#)

5. U.S. Energy Information Administration. Solar and Wind Power Curtailments Are Increasing in California. *Today in Energy*, 2025. Available online: <https://www.eia.gov/todayinenergy/detail.php?id=65364> (accessed on 15 June 2026).
6. Di Leo, P.; Malgaroli, G.; Spertino, F.; Ciocia, A. Renewable Energy Integration in Emerging Electricity Grids: Technologies, Challenges and System-Level Perspectives. *Appl. Sci.* **2026**, *16*, 5124. [[CrossRef](#)]
7. Li, S.; Tong, L.; Mount, T.D. Energy management for renewable-colocated artificial intelligence data centers. In *Proceedings of the 2026 IEEE PES International Meeting (PES IM)*; IEEE: Piscataway, NJ, USA, 2026; pp. 1–5. [[CrossRef](#)]
8. Zhang, Y.; Wilson, D.C.; Paschalidis, I.C.; Coskun, A.K. HPC Data Center Participation in Demand Response: An Adaptive Policy With QoS Assurance. *IEEE Trans. Sustain. Comput.* **2022**, *7*, 157–171. [[CrossRef](#)] [[PubMed](#)]
9. Radovanović, A.; Koningstein, R.; Schneider, I.; Chen, B.; Duarte, A.; Roy, B.; Xiao, D.; Haridasan, M.; Hung, P.; Care, N.; et al. Carbon-Aware Computing for Datacenters. *IEEE Trans. Power Syst.* **2023**, *38*, 1270–1280. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.