ISSN: P-2455-0515
E- 2394-8450

Educreator Research Journal

OPEN ACCESS

Volume–XIII, Special Issues–I(b) Jan – Feb 2026

Original Research Article

MITIGATING THE AI WATER CRISIS THROUGH ECO-ADAPTIVE SPATIO-TEMPORAL WORKLOAD SCHEDULING

** Pareek A., **Gayke A., ***Shingole A. & ***Chandani B.*

Students, B. K. Birla College (Empowered Autonomous Status), Kalyan.

Abstract:

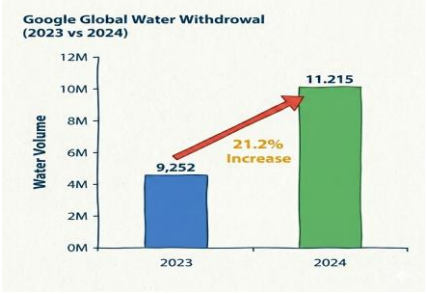
The quick rise of Generative AI is sparking a rapid rise in hyperscale data centers. The extreme carbon impact associated with these facilities is being studied in great detail, however, very little attention has been given to the extreme amount of freshwater used in evaporative cooling.

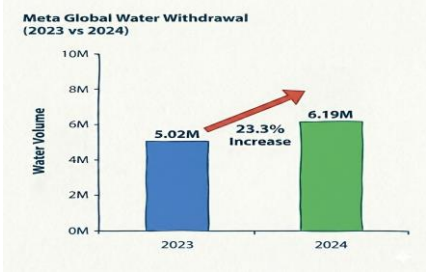
Current hardware-focused solutions like mechanical liquid cooling can be very expensive and shift a large portion of the carbon footprint toward also generating a large amount of electricity.

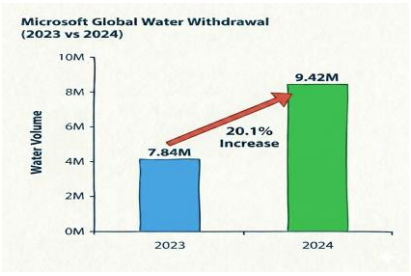
The Eco-Adaptive Water Brain Framework being discussed in this paper is an innovative software based architecture that will drastically reduce water use by designing solutions at the OS level and at the routing level. This framework will utilize spatio-temporal workload scheduling to continuously evaluate the local server's water stressed condition. The framework will route active urgent queries to energy efficient models Low Water Mode while heavy non urgent training workloads will be queued to run later at night when evaporative cooling requirements are naturally reduced with cooler ambient temperatures.The multi layered algorithmic process used in this new framework provides a hardware agnostic and sustainable method that will greatly reduce the water footprint of generative AI

Introduction:

Through the overwhelming use of Generative Artificial Intelligence across all industries, the technological environment of 2024 - 2026 is being impacted significantly, however, in order to train and operate the Large Language Models, the volumes of compute power needed by these high-volume hyper scale servers rely extremely on massive data centre based infrastructure. The thermal load on these hyper scale servers therefore means that there is a high volume of evaporative cooling towers that require millions of litres of clean potable water every day. Due to the growing rate of growth among Cloud Service Providers, the hidden environmentalconsequences of these providers utilisation of Artificial Intelligence capabilities, the hidden environmental cost is becoming a large and measurable Sustainability Crisis.







ISSN: P-2455-0515
E- 2394-8450

Educreator Research Journal

OPEN ACCESS

Volume–XIII, Special Issues–I(b) Jan – Feb 2026

Original Research Article

Statement of Problem:

The rapidly increasing need for Generative Artificial Intelligence and Large Language Models has created a staggering demand for ultra-large data center hyperscale infrastructure to house them. While the carbon emissions of these data centers have been well documented, the tremendous amount of water that is required to cool those data centers has largely been ignored and represents a major, and often unaddressed, crisis. Data centers are increasingly relying on evaporative cooling towers as a method of dissipating the tremendous amounts of heat produced by AI processors graphics processing nodes and preserving hardware from overheating. Millions of liters of clean drinking potable water can be used in the training of one very large AI model. This has a major effect on local water sources and contributes to regional droughts. Today, the industry is responding to this challenge primarily by retrofitting data centers with upgraded capital intensive hardware such as mechanical closed-loop liquid cooling systems. Implementing these types of hardware solutions on existing data centers is very expensive as well as time-consuming. Additionally, the elimination of evaporative water cooling will eliminate one source of water, but it will create new sources of costly electricity from mechanical chillers.Thus, it is critical that we develop software that will automatically and dynamically reduce the amount of water used in data centers based on real time events associated with server loading and user demand without requiring large dollar, or capital intensive, hardware upgrades.

Significance of the Study:

This research represents a landmark study as it looks to expand the definition of Green IT from being left to the rhetorical discourse around the immediate, localized threats to water scarcity that have been generated by hyper-scale computing.

The Eco-Adaptive Water Brain Framework established in this study is the first software-only, and hardware agnostic solution to the world DBA's water crisis caused by AI. As demonstrated through the implementation of spatio-temporal workload scheduling and dynamic model routing, cloud providers can conserve billions of liters of freshwater worldwide by implementing intelligent, operating system level traffic management.

This means that current data centers do not have to spend billions on physical cooling infrastructure upgrades to be environmentally sustainable therefore, if this framework is adopted by large hyperscale cloud providers, it should enable all providers globally to help ensure that the future growth of Generative AI will be both computationally capable and ecologically sustainable.

Limitations of the Study:

The Eco-Adaptive Water Brain Framework, although utilized to minimize evaporative cooling water consumption in hyperscale data centers such as those operated by Microsoft, Google, & Meta cannot be practically validated because of the limited availability of proprietary hardware utilized by these companies. Consequently, the Eco-Adaptive Water Brain Framework's research methodology relies mainly on theoretical architectural designs and on public secondary telemetry data found in the ESG reports published by these hyperscale cloud companies. Additionally, the Eco Adaptive Water Brain Framework's research has a singular goal of reducing evaporative cooling water consumption, while the effects of hardware degradation and or

ISSN: P-2455-0515
E- 2394-8450

Educreator Research Journal

OPEN ACCESS

Volume–XIII, Special Issues–I(b) Jan – Feb 2026

Original Research Article

network latency from rapidly routing AI-related workloads from larger servers to smaller servers in a Low-Water operating mode were not factored into this study.

Objectives of the study:

In this study, we will show how much water is going to be used on Generative AI workloads and show how technologists can remedy this growing effect through a practical software solution.

Ultimately,

- We will analyze ESG disclosures from leading hyperscale computing companies to find current trends associated with water usage.
- We will develop a hardware independent multi-layered scheduling framework (Eco-Adaptive Framework) to direct AI jobs to local servers based on their real-time water stress.
- We will demonstrate that spatio-temporal workload migration, specifically delaying non-critical AI training jobs until night, presents a sustainable alternative to upgrading commercial mechanical liquid cooling systems

Hypothesis of the study:

The foundational hypothesis behind this research is that using a spatio-temporal workload shift and a dynamic model route for the operation of an eco adaptive scheduling framework at the operating system level will reduce the overall evaporative cooling water use for hyperscale AI data centers when compared to how they're operated today.

Review of Literature:

Authors	Research Finding	Research gap
Jun Zhang , Ruiyong Mao, Chao Li , Jiang Lan , Xiaoyan Yi and Zujing Zhang (2023)	Optimization air-conditioning system and thermal management of data center via fan-wall free cooling technology	The study says that modern data centers use air-side and water-side free cooling to reduce electricity use, saving up to 50–68% energy. However, a 1–1.5 MW facility can consume 7–11 million liters of water per year, costing around ₹5–8 lakhs annually in India excluding maintenance.This creates a major water–energy trade-off, especially due to water scarcity, high operational costs, and reduced efficiency in hot climates

ISSN: P-2455-0515
E- 2394-8450

Educreator Research Journal

OPEN ACCESS

Volume–XIII, Special Issues–I(b) Jan – Feb 2026

Original Research Article

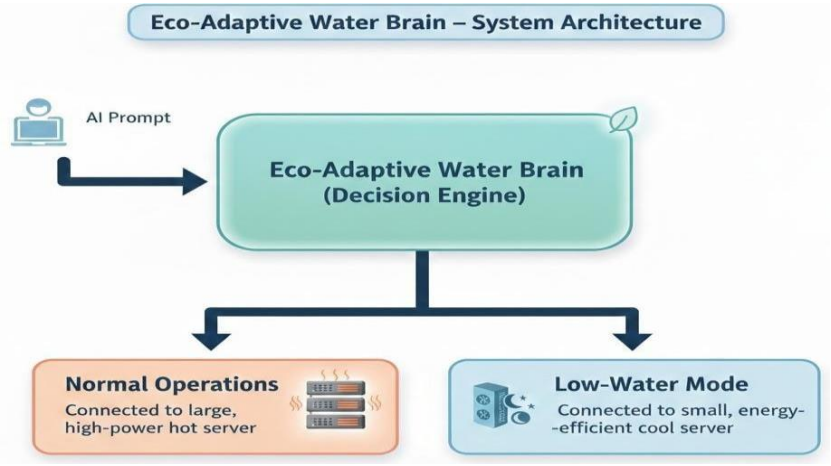
Ranran Mi ,Xuelian Bai , Xin Xu and Fei Ren (2023)	Energy performance evaluation in a data center with water-side free cooling	The study of Energy and Buildings reports that water-side free cooling uses about 1.5–3 liters of water per kWh due to cooling tower evaporation, leading to significant annual water consumption. While it reduces chiller electricity use, it increases costs for makeup water, treatment, and maintenance. Actual energy savings were lower than expected because of cooling tower limitations and climate dependency, showing a clear water–energy trade-off.

Research Methodology:

This project proposes an Eco-Adaptive Water AI system that manages computing tasks based on local water availability. The main objective of this system is to reduce water usage in data centers while maintaining system performance. The methodology followed in this project includes system design, decision logic development, task routing mechanism, and evaluation strategy.

First, the overall architecture of the system was designed. The system consists of an AI-based decision engine called the "Eco-Adaptive Water Brain." This component acts as the central controller. It receives incoming AI tasks and checks environmental conditions, especially the local water level status. The system assumes that large servers use more water for cooling, while smaller servers consume less water. Based on this concept, the architecture connects to two types of servers: a high-power large server for normal conditions and a small energy-efficient server for low-water conditions.

Second, a logical decision-making flow was developed. Whenever a new AI task arrives, the system first checks whether the local water level is critical or not. If the water level is normal, the system runs the task in normal



ISSN: P-2455-0515
E- 2394-8450

Educreator Research Journal

OPEN ACCESS

Volume–XIII, Special Issues–I(b)

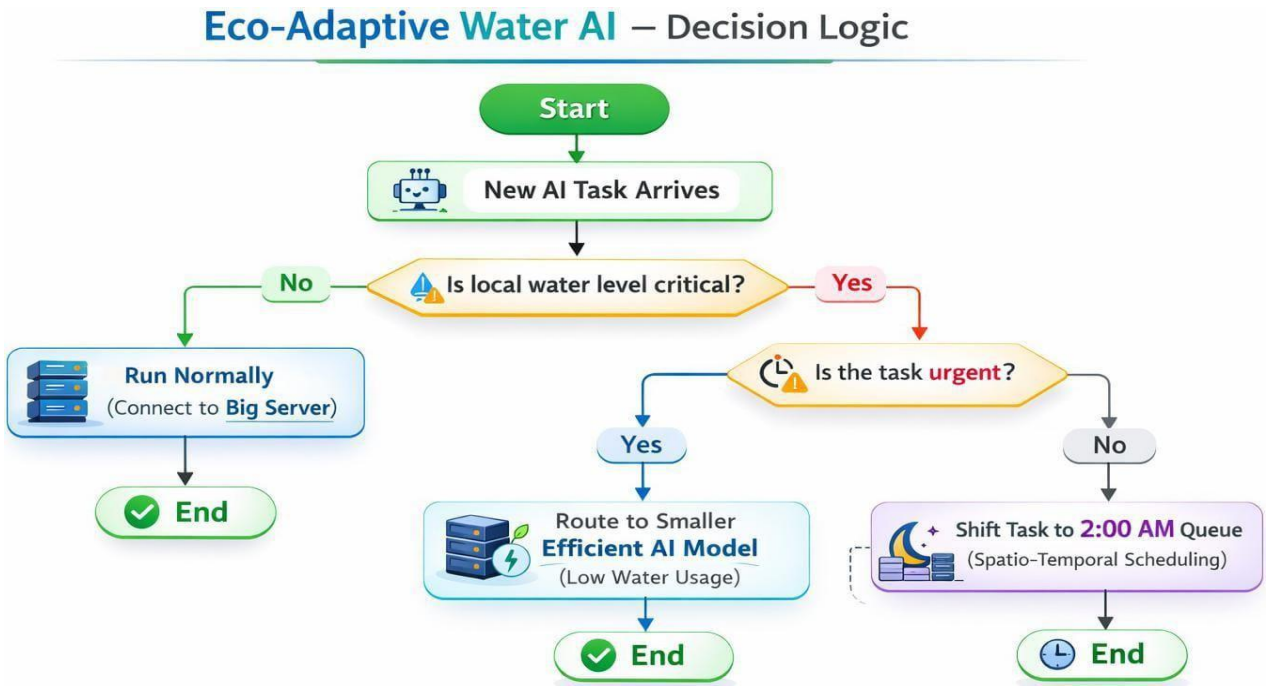
Jan – Feb 2026

Original Research Article

mode by connecting to the large server. This ensures high performance and faster task completion. However, if the water level is critical, the system activates a water-saving strategy.

In water-critical situations, the system further checks whether the incoming task is urgent. If the task is urgent, it is routed to a smaller, energy-efficient AI model that consumes less water for cooling. Although this model may have slightly lower performance, it helps reduce environmental impact. If the task is not urgent, it is shifted to a delayed processing queue. In this project, the delayed tasks are scheduled during off-peak hours, such as 2:00 AM. This is based on the idea of spatio-temporal scheduling, where workloads are moved to times when environmental impact is lower.

Third, the routing mechanism was conceptually implemented using conditional logic. The system works step-by-step: task arrival, water-level check, urgency check, and final execution path selection. This structured decision flow ensures that every task is processed efficiently while also considering environmental sustainability.



Finally, the proposed model was evaluated conceptually by comparing three scenarios: normal operation, urgent task under water-critical condition, and non-urgent task under water-critical condition. The comparison focused on performance efficiency and water conservation strategy. The results suggest that adaptive routing can balance system performance and environmental responsibility.

In conclusion, the methodology combines system architecture design and logical decision-making to create an environmentally aware AI computing framework. The approach demonstrates how artificial intelligence systems can adapt dynamically to environmental constraints while still maintaining operational efficiency.

ISSN: P-2455-0515
E- 2394-8450

Educreator Research Journal

OPEN ACCESS

Volume–XIII, Special Issues–I(b) Jan – Feb 2026

Original Research Article

Data Analysis & Interpretation:

This section evaluates the upward trajectory of water usage across the operations of Microsoft, Google, and Meta. While these companies have implemented sophisticated mitigation strategies, the data from their 2024 and 2025 environmental reports reveals that the rapid expansion of Cloud and AI infrastructure is driving significant increases in absolute water withdrawal and consumption.

1. **Microsoft: The Scale of AI-Driven Demand**

   Microsoft's recent data highlights the challenge of maintaining sustainability goals amidst unprecedented infrastructure growth.

- Rising Consumption: In its 2024 reporting, Microsoft identified that its goal to be "Water Positive" is currently "not yet on track" relative to its 2030 baseline.
- Operational Footprint: The company noted that the "first pillar" of their strategy is to address the rising water and carbon intensity of their operations as they build and operate more datacenters to meet AI demand.
- Regional Stress: Microsoft's data emphasizes that global water stress is increasing, with projections that one in three people may live in high water-stress areas by 2050, putting further pressure on the data centers located in these basins.

2. **Google: Variability and Infrastructure Expansion**

   Google's data demonstrates how geographic expansion and the heat-intensive nature of high-performance computing contribute to rising water totals.

- High-Volume Consumption: Data from 2023 shows substantial water consumption at specific sites to maintain cooling requirements. For example, Google's Berkeley County, SC facility consumed 763.4 million gallons, while its Council Bluffs, IA site reached 1,069.9 million gallons in a single year.
- The PUE-Water Trade-off: While Google has achieved low PUE (Power Usage Effectiveness) ratings of 1.08 to 1.10, the reports acknowledge that "climate-conscious cooling" often involves a choice between water and energy. In many high-heat environments, water-intensive evaporative cooling is still used to keep energy consumption (PUE) low.
- Future Uncertainty: Google explicitly states that while AI can provide climate solutions, the environmental impact of the AI models themselves is a growing factor in their total resource footprint.

3. **Meta: Resource Management for AI Workloads**

   Meta's reporting focuses on the methodology of tracking these increasing totals and the specific demands of AI-ready hardware.

- Withdrawal Volumes: Meta reports its water withdrawal based on local utility data. For its data centers, consumption is calculated by the difference between withdrawal and discharge, highlighting that a significant portion of water is "lost" to evaporation during the cooling process.
- AI Power Profiles: To manage the heat generated by AI, Meta has had to implement custom power profiles for servers. This data suggests that as AI workloads increase, the resulting thermal load necessitates more intensive cooling cycles.

ISSN: P-2455-0515
E- 2394-8450

Educreator Research Journal

OPEN ACCESS

Volume–XIII, Special Issues–I(b) Jan – Feb 2026

Original Research Article

- Construction Impact: Beyond daily operations, the 2024 report shows that the construction of new data centers to meet demand added an additional 1,780,000 cubic meters of water withdrawal.

Challenges:

Although the new Eco-Adaptive Water Brain Framework has the potential to produce many ecological benefits, there are some operational issues concerning the hypothetical application of the framework in hyperscale data center applications. The biggest issue is likely the imposition of network latency onto the application and the negative impact this would have on the end-user experience. When generative AI jobs are delayed and performed at night time due to the use of spatio-temporal scheduling, this would break the immediacy of response expected in modern cloud computing customers. Second, the determination of AI workload urgency, either urgent or non-urgent, would involve complex algorithmic decision making without the involvement of human decision making. Should a business-critical query be incorrectly assessed as non-urgent, resulting in the processing of the query at a later time, this would breach the Service Level Agreements (SLAs) of the corporation. Lastly, the imposition of the new routing layer onto the legacy operating systems of the data center would potentially result in an increased energy consumption of the routing servers themselves

Remedies:

In order to mitigate these latency and classification issues, a user-centric opt-in mechanism called "Eco-Mode" is implemented within the framework, which is similar to the energy-saving options available on mobile devices. By giving end-users the option to mark their prompts as "delayable" in exchange for a documented reduction in their water footprint, these issues are resolved without any algorithmic guesswork. In order to mitigate any potential Service Level Agreement (SLA) violations, a dynamic timeout protocol is implemented within the spatio-temporal scheduler; when a task is near a critical deadline within a queue, it is automatically promoted to the "Low Water Mode" for immediate processing on energy-efficient servers. In order to mitigate potential computational overheads, the decision engine is implemented as a lightweight microservice to ensure that the energy consumption required to power the "Water Brain" is negligible when compared to the massive volume of water saved through evaporation..

Conclusion:

The rapid scaling of Generative AI has inadvertently triggered an ecological crisis, with hyperscale data centers consuming unsustainable volumes of freshwater for evaporative cooling. As evidenced by recent corporate telemetry data, relying solely on capital-intensive hardware upgrades—such as mechanical liquid cooling—is an insufficient and economically restrictive strategy for legacy infrastructure. The Eco-Adaptive Water Brain Framework proposed in this study demonstrates that the AI water crisis can be effectively mitigated at the software level. By integrating spatio-temporal workload scheduling and dynamic model routing, cloud providers can drastically reduce peak-hour water evaporation natively. Ultimately, optimizing how AI data centers route traffic, rather than just how they cool physical servers, provides a scalable, hardware-agnostic pathway toward a truly sustainable future for artificial intelligence.

ISSN: P-2455-0515
E- 2394-8450

Educreator Research Journal

OPEN ACCESS

Volume–XIII, Special Issues–I(b) Jan – Feb 2026

Original Research Article

References:

Research Papers

1. *Jun Zhang , Ruiyong Mao , Chao Li , Jiang Lan , Xiaoyan Yi and Zujing Zhang (2023) Optimization air-conditioning system and thermal management of data center via fan-wall free cooling technology https://doi.org/10.1016/j.applthermaleng.2023.121245*
2. *Ranran Mi ,Xuelian Bai , Xin Xu and Fei Ren (2023) Energy performance evaluation in a data center with water-side free cooling https://doi.org/10.1016/j.enbuild.2023.113278*
3. *Microsoft 2024 Environmental Sustainability Report https://share.google/DbghRIJsBYhCJQcJv*
4. *Microsoft 2025 Environmental Sustainability Report https://www.microsoft.com/en-us/corporate-responsibility/sustainability/report/ 2025 microsoft ESG report*
5. *Google 2024 Environmental Sustainability Report 2024 Environmental Report https://share.google/CGc2Ksfm7rTMXqwgG 2024 google ESG report*
6. *Google 2025 Environmental Sustainability Report 2025 Environmental Report https://share.google/dEKretnazNBcyydh7 2025 google ESG report*
7. *Meta 2024 Environmental Sustainability Report https://share.google/ogzaZ5X7mSdKYNlZm 2024 meta ESG report*
8. *Meta 2025 Environmental Sustainability Report 2025 Sustainability Report - Meta Sustainability https://share.google/SE73v371gP9TjWLyJ 2025 meta ESG report*

Cite This Article:

Pareek A, Gayke A, Shingole A. & Chandani B . (2026). *Mitigating the AI Water Crisis Through Eco-Adaptive Spatio-Temporal Workload Scheduling.* **In Educreator Research Journal: Vol. XIII (Issue I),** pp. 56–63. **Doi: https://doi.org/10.5281/zenodo.19916273**