ISSN (Online): 3139-7689
Available online: www.ijsrmes.com
International Journal for Scientific Research in Modern Engineering and Science
Published in Conference Special Issue: Vol. 1, Special Issue 1, pp. 361–368, 2026

ARCHITECTING THE POST-SILICON TRANSITION: A MULTIDIMENSIONAL ANALYSIS OF GREEN AI INFRASTRUCTURE AND THE WATER-INTELLIGENCE NEXUS

Rahul Dewangan[1*], Rishabh Deshmukh[2], Maddu Divyansh[3], Pranav Shekhar[4], Shankar Sharan Tripathi[5]

[1,2,3,4]*CSE(AIML), Shri Shankaracharya Technical Campus.*

[5]*Computer Science and Engineering, Shri Shankaracharya Technical Campus.*

Abstract

Artificial Intelligence is more than just a useful tool now; it has become an enormous power-hungry machine. For those concerned about sustainability, this is an important issue. AI enables us to operate power grids more cleanly, create smarter cities and improve agricultural productivity; however, something has to give. According to projections, U.S. data centres will consume approximately 300 terawatt-hours of electricity and 720 billion gallons of fresh water each year by 2028. That is a staggering amount.

Four distinct solutions to these issues are presented within this document: Circular Thermal Systems, Liquid Immersion Cooling, Neuromorphic Computing and DNA-based Storage. The document will also provide a practical engineering pathway encompassing the requirements of the UN's Sustainability Goals and Global Ethics Stds. The conclusions drawn from this report show that if we take action soon, through some combination of these solutions, we can reduce the environmental impact of AI by at least 50% over the next ten years, and potentially significantly more through innovative bio-inspiration technologies created in the future. Let's explore the details.

Keywords: *Sustainable AI; Data Centers; Liquid Immersion Cooling; Neuromorphic Computing; DNA Storage; EcoServe 4R; Silicon Thirst; SDGs.*

* *Corresponding author*

1. Introduction

It may seem like artificial intelligence is going to impact progress on the United Nations' Sustainable Development Goals, but there's a problem with the technology that is moving us forward. For example, while AI can help produce cleaner energy and create smarter cities and combat climate change, it is also pushing our planet closer to its limits. Currently, on target for meeting less than 35% of the SDG Goals by 2025, energy use demands from today's top supercomputers adds to that target being missed.The United States will likely triplicate energy used for all data centers from 2023-2028. This could equate to about 12% of total U.S. electricity consumption. And it gets worse, as cooling this hardware requires significant amounts of water. Recent projections put water used for cooling U.S. data centers at over 720 billion gallons annually—equal to the amount of water used indoors by approximately 18.5 million households.This paper discusses four ways that we can address these issues: (1) use waste heat generated by servers to heat buildings, (2) convert data center cooling methods to liquid immersion cooling that has negligible water consumption, (3) begin to utilize neuromorphic chips to help reduce greatly power demands of specific processing, and (4) develop DNA molecules for cold archival data storage..

Fig. 1. Growing public awareness of data centers' environmental toll.

2. Silicon Thirst: The Water Dimension

The article "Silicon Thirst" explores the impact of AI on global water resources; in short, it states that as AI continues to expand, it will consume vast amounts of fresh water. To cool large server facilities (especially those using evaporative cooling), a large amount of freshwater is required. For example, it is estimated that training GPT-3 used over 700,000 liters of water alone, which is a small fraction of the total water used by billions of queries each day. Once all of these individual water usages are summed, it begins to look similar to the amount of water required to provide a typical city with water. There is a trade-off regarding using evaporative cooling for data center energy efficiency and the amount of water that can be lost through such methods. Water Usage Effectiveness (WUE) is a metric that illustrates the trade-off between how energy savings in electricity, as an example, can lead to increased water usage.

In addition to the issues at the global level, the issues at the local level are substantial. As aquifer levels fall, cities' water supplies are strained, farms or homes struggle to compete for the remaining resources. Hyperscalers typically seek out rural areas for cheap land and available energy, which exacerbates the issues faced by those who live in the area. The "Giant Soda Straw" metaphors represent these data centers which consume resources, competing with whole cities, and without adequate accountability on their actions. The question is, who are the winners of this AI boom, and, conversely, who is the loser.

3. Climate Change and Energy

A great deal of energy is consumed by artificial intelligence, and if fossil fuels are the source of energy that supplies electricity to the power grid, then the use of the artificial intelligence resource will go a long way to increasing greenhouse gas (GHG) emissions from the generation of electricity needed for these uses. If the annual electricity consumption for just the U.S. artificial intelligence data centers reached 300 TWh/year, this would create a significant portion of the national GHG emissions unless the power source for these centers transitions to renewable sources of energy rapidly. In addition to the day-in and day-out energy consumption account for the use of the system, the large amount of carbon that is emitted producing the processors (CPUs, GPUS and memory chips) adds to the GHG emissions associated with the operation of these systems.

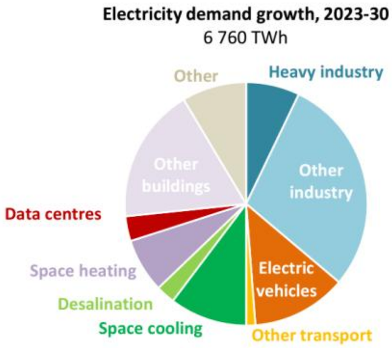

Fig. 2. Electricity demand growth by sector, 2023–30. Data centers highlighted.

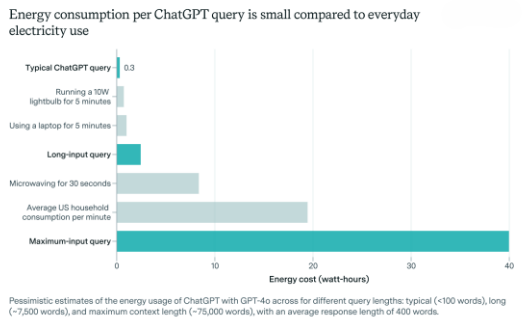

Fig. 3. Energy per ChatGPT query vs. everyday electricity uses (watt-hours).

A. The EcoServe 4R Framework

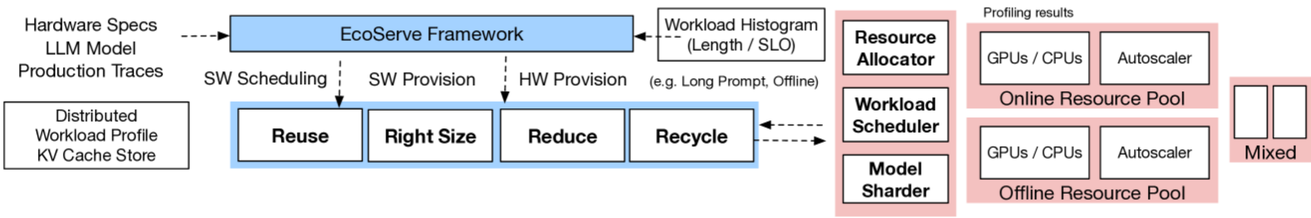

Fig. 4. EcoServe 4R Framework: Reuse, Right-Size, Reduce, Recycle applied to AI workload orchestration.

The EcoServe 4R Framework— The 4R framework is intended to reduce the amount of energy that these systems use as well as the GHG emissions that were generated in the manufacturing of the hardware (CPUs, GPU and memory chips) used in the production of LLM. The 4R framework consists of the following principles: reuse, right-size, reduce.

4. Methodology / Approach

In exploring four options for reducing emissions from medium-sized data centers (defined as data centers with IT loads between 100-500 kW), we modeled three different ten-year scenarios using detailed hourly energy utilization models, lifecycle assessments compliant with ISO 14040-44 protocols, and Monte Carlo sensitivity analyses. We considered a variety of variables, including the carbon intensity of the electric grid regionally, the rate at which workloads are expected to increase over time, and potential costs associated with new technologies.

Dimension	A — Baseline	B — Green Stack	C — Bio-Inspired

Cooling	Air / evaporative	Liquid immersion (LIC)	LIC + neuromorphic
Compute	GPU-centric servers	GPU + EcoServe 4R	GPU + neuromorphic chips
Storage	Disk / tape archives	Optimised disk/tape	DNA molecular archival
Heat Recovery	None	Greenhouse / district heating	Full circular heat reuse
Carbon Awareness	Limited	Renewable-aware scheduling	Fully optimised

Evaluation Framework

The evaluation tracked metric data points such as: net energy consumption (measured in megawatt hours/year), water use, lifecycle greenhouse gas emissions (measured in CO2e), and total cost of ownership. The study was also sensitive to social variables, including: local water stress caused by the operation of the data center; any positive social impacts resulting from energy consumption by the data center. The study included community input into the site selection process as well as required reporting at regular intervals to ensure open and transparent communication regarding pilot project implementation

5. Results and Discussion

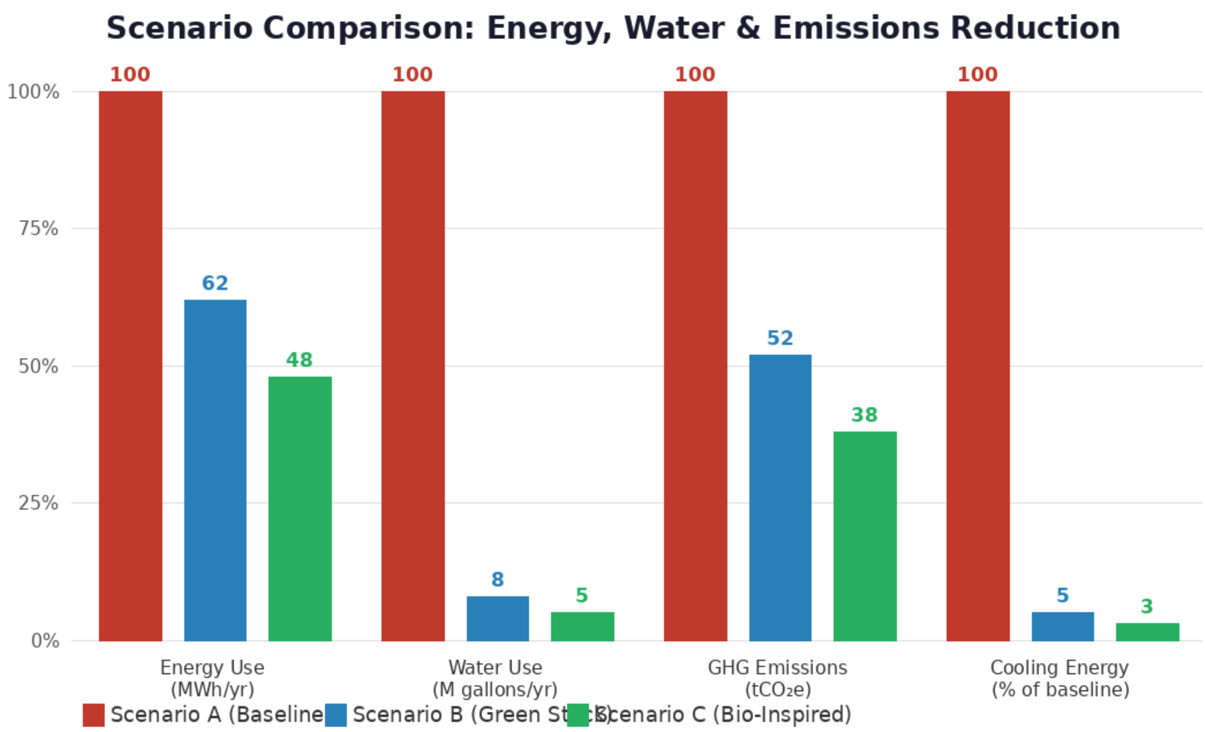

Fig. 5. Scenario comparison: energy, water, GHG emissions, and cooling energy indexed to Scenario A = 100%.

A. Circular Thermal Infrastructure

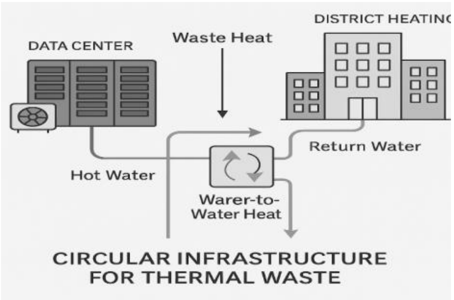

Fig. 6. Data center waste heat coupled to district heating loop.

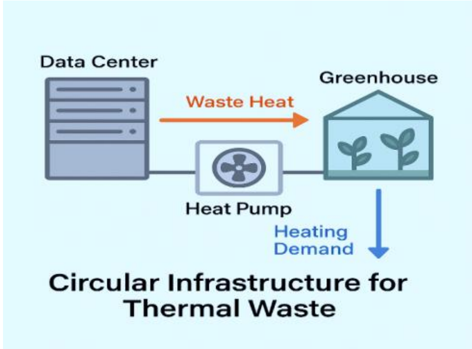

Fig. 7. Waste heat from data center diverted to greenhouse via heat pump.

Data centres are beginning to integrate their residual heat into district heating systems or to be used in greenhouses. As a result, many more users of the energy being produced gain a benefit from it. Farms are becoming more efficient and so are the local communities. The statistics confirm this. With respect to the greenhouses, the amount of money spent on electricity has been reduced by nearly 30%, while the amount spent on heating has increased by an average of 16%. For example, Dutch tomato and strawberry growers have reduced their overall heating costs by 25%. These types of circular systems are currently producing approximately 25% to 30% of local energy. The cooling issues that data centres used to deal with can now be considered a success, with total site energy use reduced to as low as 40% of capacity. Depending on the prevailing weather conditions, this number can even be reduced further.

B. Liquid Immersion Cooling (LIC)

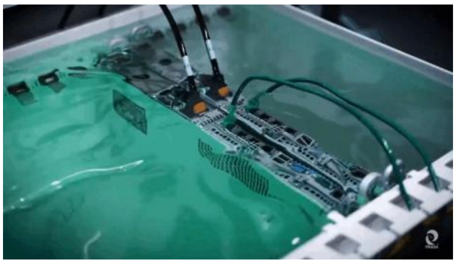

Fig. 8. Hardware submerged in dielectric fluid: LIC in practice.

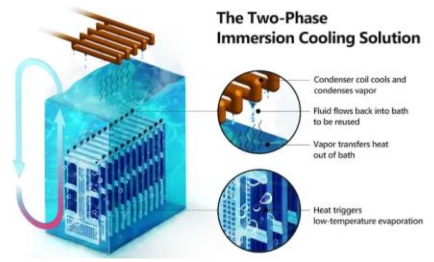

Fig. 9. Two-phase immersion: vapour condenses and recirculates for reuse.

Liquid immersion cooling has changed how we look at efficiency in data centers, reducing cooling-related power consumption by as much as 95% and achieving very low Power Usage Effectiveness (PUE), typically between 1.03 to 1.05. A high-density rack can reliably operate under 40kW to 100kW without ever needing to be concerned about thermal throttling, which is common in conventional air-cooled systems. There are also no evaporative water losses due to the closed-loop design utilizing dry heat rejection, which is especially important given that there is increasing pressure on global supply chains associated with fresh water sources as water stress continues to grow globally. Overall, many organizations will see significant savings on operational costs through this technology, with an average payback period of only 2-4 years on the initial infrastructure investment. Additionally, LIC has the flexibility to provide an infrastructure that supports denser deployment and "futureproofs" the facility from the increasing power and cooling requirements of next-generation AI hardware. By greatly reducing both energy and water footprints, LIC represents the fundamental solution to creating a sustainable and scalable architecture for the digital world.

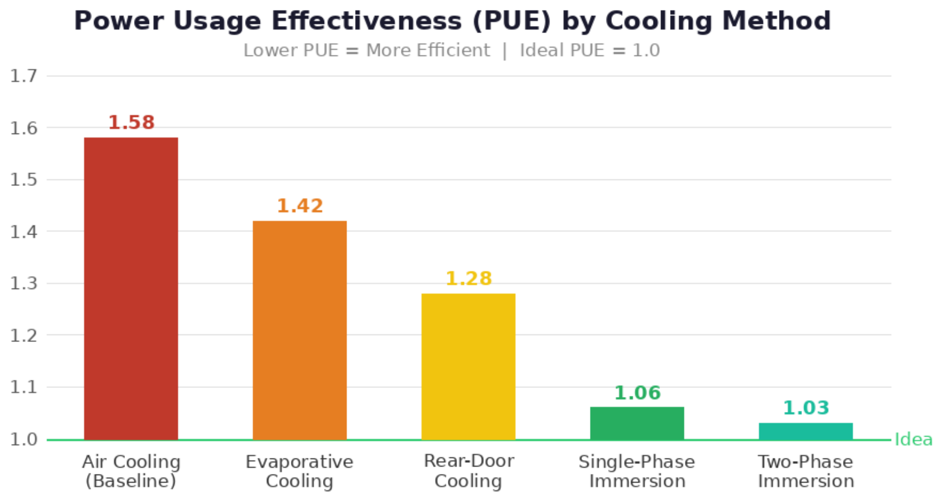

Fig. 10. PUE by cooling technology—two-phase immersion approaches the theoretical ideal of 1.0.

C. Neuromorphic Computing

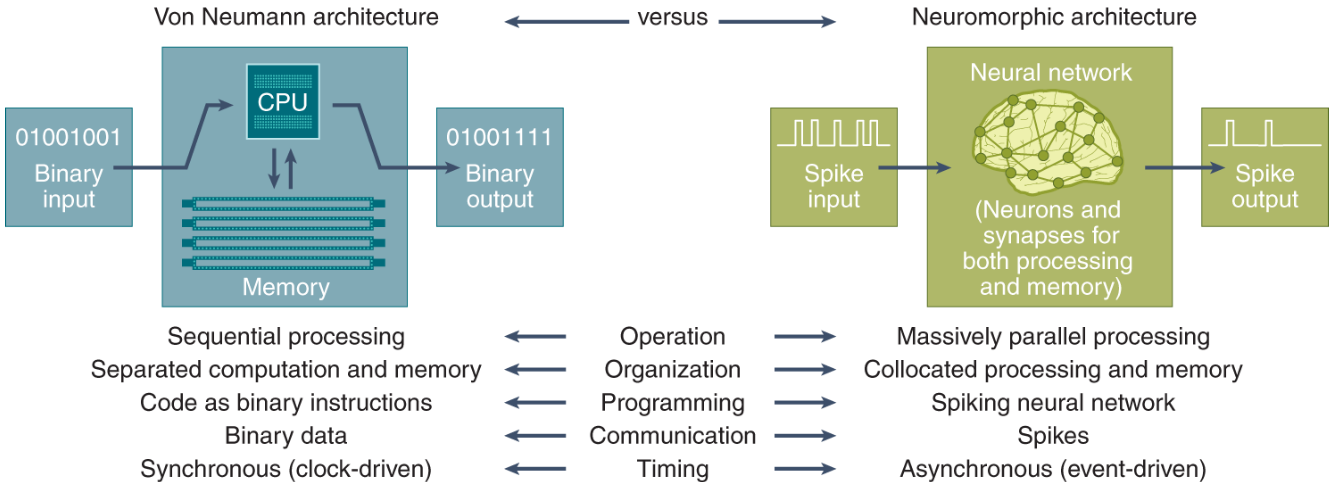

Fig. 11. Von Neumann vs. Neuromorphic: co-located memory and processing eliminates the 90% data-movement energy bottleneck.

Chips utilizing brain-like qualities could achieve as much as 1,000x better energy efficiency than traditional hardware for tasks that are performed only when an event needs to be processed by having both memory and processing capabilities in the same physical location, thus eliminating the energy expense (90%) associated with data movement in traditional hardware. By moving 20% to 30% of the total inference workload to brain-like chips, total compute energy could be reduced by another 10% to 20% compared to Scenario B, and power consumption savings for always-on edge deployments would help reduce the expected growth of upstream data centers.

D. DNA-Based Archival Storage

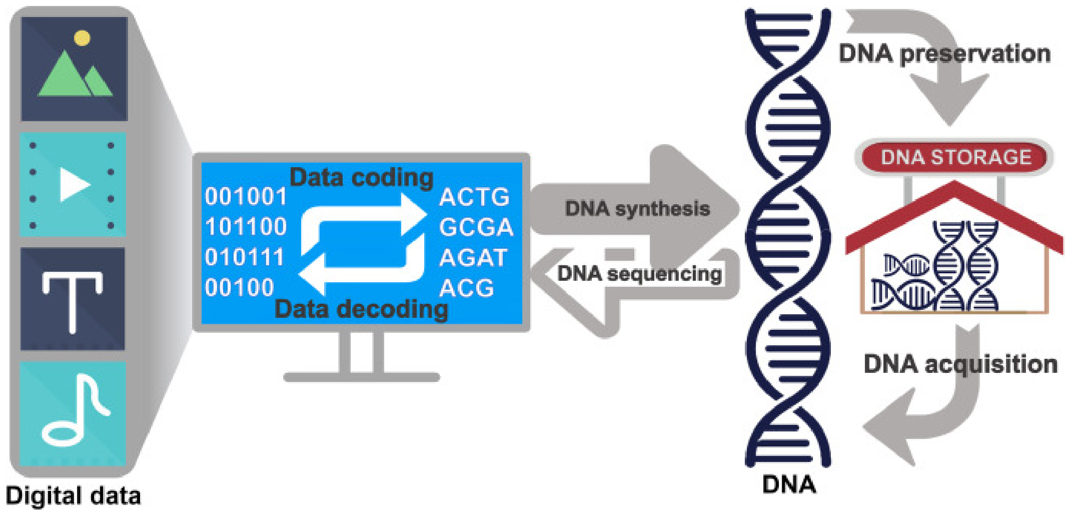

Fig. 12. DNA storage pipeline: encode → synthesise → store → sequence → decode, achieving 215 PB/g with near-zero maintenance energy.

Chips inspired by the brain revolutionize the way tasks are accomplished that are driven by events. These types of chips place the memory for processing as close to the actual processing as possible to eliminate the need for ever 'shuffling' (moving) the data, which typically consumes over 90% of energy consumed in conventional technology. The energy efficiency of chips using this method is increased by as much as 1,000 times. By moving just 20 to 30% of the inference workload from conventional chips to neuromorphic (brain-inspired) chips, the total energy consumption of computing will be reduced by 10 to 20% beyond that obtained in Scenario B.

6. Conclusion

This study investigated if artificial intelligence (AI) is a major contributor of environmentally-sustainable innovations while also constituting a very serious source of environmental impact, and this multi-disciplinary roadmap affirms that this transformation is immediately possible and will occur during this decade. This conclusion is based on a number of criteria, such as:

• Short-Term Solutions Include: EcoServe's 4R + Life Cycle Indicators + circular heat recuperation can offer up to 50% reduction to AI's environmental footprint; practically none of AI's cooling will consume fresh water; the Total Cost of Ownership (TCO) payback will occur in a timeframe of 2 to 4 years.

• Long-Term Solutions Will Be: Neuromorphic Computing and DNA Storage will effectively uncouple the growth of AI from a linear growth trajectory in the use of energy, water and materials via orders of magnitude efficiency improvements.

• Governance & Policy: AI governance must require Water and Energy Use (WUE), Power Usage Efficiency (PUE), and Greenhouse Gas (GHG) disclosure, prevent evaporative cooling in watersheds subject to water scarcity, require community benefit agreements for heat re-use projects, and ensure alignment with UNESCO's Recommendation on the ethics of artificial intelligence as prerequisites to ensuring equitable outcomes.

Ultimately, AI needs to evolve to a new paradigm—the Biological Machine Paradigm—that is efficient, resilient, and becomes integrated with the earth's natural cycles/destructive cycles. This transformation provides AI with the ability to not simply remain a threat to sustainability but to become a regenerative force in support of a new resilient digital future that functions within the limits of this planet.

References

[1] A. Shehabi et al., LBNL-1005775, 2016.
[2] E. Masanet et al., Science, 367(6481), 2020.
[3] S. Deng et al., ACM e-Energy, 2020.
[4] P. Wu et al., Renew. Sustain. Energy Rev., 151, 2021.
[5] S. A. Cook et al., IEEE Trans. Sustain. Comput., 7(3), 2022.
[6] J. H. Koomey, Analytics Press, 2021.
[7] K. G. Brill & K. S. Lange, IEEE ITherm, 2021.
[8] A. T. A. M. de Jong et al., Appl. Energy, 276, 2020.
[9] M. Dayarathna et al., IEEE Commun. Surveys Tuts., 18(1), 2016.
[10] S. Furber, J. Neural Eng., 13(5), 2016.
[11] M. Davies et al., IEEE Micro, 38(1), 2018.
[12] N. Goldman et al., Nature, 494, 2013.
[13] Y. Erlich & D. Zielinski, Science, 355(6328), 2017.
[14] UNESCO, Ethics of AI Recommendation, 2021.
[15] OECD, Digital Economy Papers, No. 317, 2021.
[16] Green Software Foundation, SCI Spec v1.0, 2023.
[17] World Resources Institute, Aqueduct Atlas, 2023.