

Governing AI Data-Centre Resources Before Consumption: The ZERO Framework

Maria Kollia

karnw79@gmail.com

Independent Researcher

Perspective

Keywords: artificial intelligence, data centres, circular economy, resource governance, workload scheduling, carbon neutrality, water-aware computing

Posted Date: August 19th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10593841/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.
[Read Full License](#)

Additional Declarations: No competing interests reported.

Abstract

AI data-centre sustainability is usually evaluated after electricity, water and materials have already been committed. This Perspective develops ZERO as an operational governance framework that moves the decision boundary upstream. ZERO combines hard admissibility constraints, workload characterisation, local resource-state assessment, federated comparison, continuation conditions and auditable evidence. Its contribution is distinguished from efficiency metrics, carbon-aware scheduling, water-aware scheduling and circular recovery by requiring a multi-resource decision before execution. An illustrative computational case study shows how the framework can exclude a low-carbon but water-critical facility and select a lower-impact admissible alternative. The case study is deliberately synthetic and demonstrates decision logic rather than measured savings. The paper concludes with empirical evaluation requirements, limitations and a staged research programme.

1 Introduction

Artificial intelligence is often described through models, parameters and computational capability. Its expansion is also a physical infrastructure project. Training and inference depend on processors, electricity, cooling systems, water, buildings, networks and equipment supply chains. Data-centre resource burdens vary with workload, facility design, cooling configuration and geography, making both energy and water boundaries consequential (Masanet et al. 2020; Siddik et al. 2021; Mytton 2021; Lei et al. 2025). Recent energy-system assessments expect data-centre electricity demand to rise substantially during the present decade, although the scale and location of growth remain uncertain (International Energy Agency 2025; Shehabi et al. 2024). The environmental consequence cannot be inferred from electricity alone. The same workload may interact differently with grid carbon intensity, cooling technology, watershed conditions, available heat users, equipment utilisation and local infrastructure constraints.

Efficiency remains necessary, but it does not settle the prior question of admission. Most sustainability controls become visible after a workload has been accepted, a facility has been selected and resources have entered operation. Power usage effectiveness (PUE), water usage effectiveness (WUE) and carbon usage effectiveness (CUE) then describe the performance of an operating system. They do not by themselves determine whether a particular computation should start, whether its scale is justified, whether it should be delayed, or whether a technically available facility is locally admissible. Reporting can make a burden visible without giving that information authority over execution.

This Perspective develops ZERO as a resource-governance framework for decisions before and during AI workload execution. ZERO is the name of the proposed framework, not an acronym and not a claim of literal zero physical consumption. The name expresses its governing premise: avoidable demand and preventable resource loss should be challenged before they are treated as inevitable. The framework integrates workload necessity, resource state, local constraints, recovery pathways, network coordination, evidence freshness and accountable intervention. Earlier conceptual papers introduced

separate elements of zero-waste infrastructure, circular recovery and adaptive resource governance (Kollia 2026a, 2026b). The present paper consolidates those elements into a more explicit operational model, reduces reliance on those earlier texts by grounding the argument in a broader peer-reviewed literature, and provides an illustrative case study of the decision process.

The scientific contribution is therefore not a new efficiency indicator. It is a decision architecture with five testable properties: (1) hard local constraints are applied before optimisation; (2) workload and resource states are represented separately; (3) only admissible options enter network comparison; (4) approval remains conditional on evidence validity and changing conditions; and (5) the resulting decision and physical outcome are recorded against a credible baseline. These properties permit future comparison with conventional local execution, carbon-aware scheduling and water-aware scheduling.

2 Existing approaches and the remaining decision gap

Data-centre sustainability research has developed several strong but partly separate traditions. Facility metrics standardise the measurement of power, water and carbon performance. Cooling research examines thermal management and the recovery of low-grade heat (Ebrahimi et al. 2014; Wahlroos et al. 2018; Yuan et al. 2023). Carbon-aware computing shifts flexible workloads across time or space according to electricity-system conditions (Radovanovic et al. 2023; Hall et al. 2025). Water-aware scheduling exploits geographical and temporal differences in water efficiency (Islam et al. 2018). Multi-energy models coordinate workload flexibility, renewable generation, cooling and heat recovery (Liu et al. 2012; Wang et al. 2020). Recent geo-distributed formulations broaden this optimisation space further: Moore et al. (2025) co-optimize service quality, carbon, water and energy cost for LLM inference; Ruparel (2026) jointly treats cost, carbon, latency and uncertainty; and Wang et al. (2026) coordinates workload migration with electricity and carbon-market conditions. Circularity-oriented work also extends the system boundary to equipment design and value retention (Andrews and Kerwin 2026).

These approaches establish that workload timing and placement can influence environmental performance. They also show why a single objective is insufficient. A lower-carbon location can have higher water use. A facility with efficient cooling can face a locally constrained watershed. A heat-recovery installation can produce little useful substitution if temperature, distance or demand are unsuitable. Spatial movement can add transfer energy, latency, duplicated computation and operational risk. Carbon-aware scheduling itself has limits when forecasts, service constraints and marginal grid effects are not represented adequately (Sukprasert et al. 2024).

ZERO differs in where it places the governing decision. Its novelty is not the simultaneous consideration of carbon, water or other resource variables, which already appears in integrated scheduling and multi-resource optimisation. Instead, ZERO formalises a governance sequence in which binding admissibility conditions, local decision authority, evidence validity, veto rights and revalidation have authority before and during optimisation. This distinction is methodological rather than merely terminological. Conventional constrained optimisation can define a feasible set at solve time; ZERO treats admissibility

as an authority-bearing state that can expire, be locally vetoed, or be reopened while execution is under way when evidence or conditions change. It does not begin by ranking every technically available site. It first removes options that fail legal, security, safety, resilience or local environmental requirements. Optimisation occurs only within the remaining set. It also separates resource capture from productive reuse, requires the validity period of evidence to be explicit, and retains a local veto after network comparison. The framework is therefore best understood as constrained admissibility followed by multi-objective selection embedded within a continuing governance lifecycle, rather than as another unrestricted environmental objective function. The comparison with established and closest integrated approaches is summarised in Table 1 below.

Table 1

Position of ZERO relative to facility metrics, sustainability-aware scheduling and integrated multi-resource optimisation. The comparison separates optimisation objectives from ZERO's additional governance functions: admissibility, evidence validity, local authority, veto and execution-time revalidation.

Approach	Primary decision object	Strength	Boundary not normally resolved	ZERO extension
PUE, WUE and CUE	Facility performance ratio	Comparable operational measurement	Does not decide workload necessity or admission	Uses indicators as inputs to an execution decision
Carbon-aware scheduling	Time or site of execution	Aligns flexible load with lower-carbon electricity	May omit water, cooling reserve, heat use and local veto	Applies carbon only after multi-resource admissibility
Water-aware scheduling	Time or site of execution	Responds to water intensity and spatial variation	May omit carbon, legal constraints, heat recovery and evidence authority	Combines water state with carbon and other binding conditions
Multi-energy optimisation	Joint operation of workload and energy assets	Coordinates renewable supply, cooling, heat and cost	Often optimises within a predefined feasible system	Adds governance tests that determine whether an option is admissible
Circularity and heat recovery	Recovery or lifecycle pathway	Extends value retention beyond direct efficiency	Recovery can be used to excuse avoidable demand	Places avoidance and reduction before recovery
Integrated carbon-water / multi-resource scheduling	Time or site under joint resource objectives	Represents cross-resource trade-offs explicitly	Optimisation may begin from a predefined feasible set without explicit evidence expiry, local veto or execution-time revalidation	Adds governance-controlled admissibility, local authority, evidence validity and revalidation before and during selection
ZERO	Admission, continuation and accountable intervention	Integrates hard constraints, multi-resource state, evidence and authority	Requires local thresholds and empirical calibration	Operational governance sequence illustrated here through a synthetic demonstration

3 The ZERO hierarchy

ZERO moves the decision boundary upstream through an ordered resource hierarchy. The first duty is to avoid computation that has no defensible purpose or duplicates an adequate result. Necessary computation should then be reduced to an appropriate scale through model choice, precision, batching, reuse or workload redesign. Flexible work should be aligned with suitable temporal and geographical conditions. Residual energy, heat, water and material value should be recovered where technically and

environmentally justified. Productive reuse must be distinguished from capture. Regeneration is the final and most demanding claim because it requires evidence that the wider system has been strengthened rather than merely compensated. The hierarchy is shown in Fig. 1 below.

The order matters. Heat reuse cannot justify unnecessary computation. Renewable electricity does not remove the need for efficient software and hardware. Water replenishment in one location cannot conceal pressure in another. Recycling does not justify avoidable equipment turnover. This ordering is consistent with circular-economy principles of value retention and systems thinking, but it adds an execution question: which conditions must be satisfied before a workload is permitted to proceed?

The hierarchy and the later weighted score serve different functions and are deliberately non-compensatory. The hierarchy and hard admissibility rules determine whether an option may enter the candidate set at all. Failure of a binding requirement cannot be offset by a favourable value elsewhere. Weighted scoring is applied only after admissibility has been established and only to compare alternatives that have already passed the non-compensable gate. Thus, optimisation can trade among soft objectives within the admissible set, but it cannot convert an inadmissible option into an admissible one.

4 Operational formulation

4.1 Workload and resource representations

A workload request is represented as $W = \{p, e, d, u, l, q, x\}$, where p is purpose and necessity, e is expected IT energy demand, d is duration, u is urgency and deadline, l contains location and jurisdiction restrictions, q contains service and resilience requirements, and x describes flexibility, including interruption, division, model substitution and migration. This representation prevents the environmental decision from being reduced to a single energy estimate.

For each facility i and decision time t , the resource state is $R(i,t) = \{c, w, k, h, g, m, v\}$. The elements denote electricity carbon intensity c , water condition w , cooling reserve k , productive heat-use opportunity h , grid or local infrastructure condition g , material and equipment condition m , and evidence validity v . Each element includes a measured or forecast value, source, confidence and time-to-live. Local operators translate raw telemetry and policy into operational states such as favourable, normal, constrained, critical, unavailable or unknown. The labels are common, but their numerical thresholds remain local because engineering design, climate, watershed conditions and legal obligations differ.

4.2 Hard admissibility before optimisation

Hard admissibility is represented by a binary gate, but the workflow distinguishes a provisional hold from a passed or failed gate. Let $C_H(i,W,t)$ denote the set of binding legal, security, safety, data-sovereignty, resilience and local environmental constraints applicable to facility i , workload W and decision time t . Define the hard-admissibility indicator $H(i,W,t) = 1$ only when every applicable constraint in $C_H(i,W,t)$ is

satisfied and the evidence required to evaluate those constraints remains valid; otherwise $H(i,W,t) = 0$. The admissible facility set is $A(W,t)=\{i : H(i,W,t) = 1\}$. A failed binding constraint produces EXCLUDED status. If a binding value is unresolved, stale or awaiting confirmation but can reasonably be resolved before execution, the workflow may assign HOLD/CONDITIONAL status; this is not a pass and the facility does not enter A until the missing condition is resolved. This distinction preserves a mathematically binary admissibility gate while allowing the operational process to represent uncertainty without silently treating it as favourable.

This rule is intentionally stricter than weighted optimisation. A sufficiently low carbon score cannot compensate for a critical cooling reserve, a prohibited data location or a binding water restriction. Hard constraints therefore determine membership of A before any composite score is calculated. A HOLD/CONDITIONAL case remains outside A until its binding evidence is resolved; a FAIL case remains excluded unless the relevant state materially changes and is revalidated.

4.3 Selection among admissible options

Only facilities in A are compared. For each admissible action a, including immediate execution, intensity reduction, alternative hardware, division, delay, relocation, suspension or rejection, ZERO calculates an action score $S(a)$. The score combines normalised carbon, water, cooling, productive recovery, transition cost, service risk and evidence confidence. The weights are decision-specific and must be disclosed. They operate only within the already admissible set and therefore cannot override or compensate for a failed hard constraint. A general form is $S(a) = \sum(w_j z_j(a)) + P(a)$, where z_j is a normalised burden or benefit, w_j is its declared weight and P is a penalty for transition, uncertainty or service degradation.

Weights require governance because an arbitrary vector can make a transparent score appear more objective than it is. For empirical deployment, ZERO therefore permits two auditable routes. A stakeholder or policy-derived vector may be elicited using the analytic hierarchy process (AHP), including a consistency check on pairwise judgements; an objective vector may be estimated using an entropy-weight method when a sufficiently large decision matrix exists, so that attributes with greater observed information dispersion receive greater weight. A hybrid vector can then be declared as $w = \lambda w_{\text{AHP}} + (1-\lambda) w_{\text{ENT}}$, $0 \leq \lambda \leq 1$. The method, reference data, λ and final weights must be fixed before inspecting the preferred outcome. Because the present Perspective contains neither an expert-elicitation dataset nor an empirical decision matrix, it would be inappropriate to fabricate AHP or entropy estimates. Section 6.4 therefore uses predeclared synthetic weight profiles only as a sensitivity analysis, not as calibrated operational weights.

A decision record contains the selected action, rejected alternatives, binding constraints, evidence sources, validity period, responsible actor and continuation triggers. This record is part of the operational model, not a later reporting layer. It allows retrospective testing of whether the same information would have produced a different decision under another method.

4.4 Continuation and revalidation

Admission is not a permanent licence. Long-running workloads continue only while their material conditions remain valid. Reassessment is triggered when a resource state crosses a declared threshold, evidence expires, confidence falls, a local restriction changes, or the predicted consequence differs materially from the observed one. The framework does not react to every fluctuation. Trigger severity, persistence and confidence are specified in advance to avoid oscillation and unnecessary migration.

The least disruptive admissible response is considered first. New flexible demand may be restricted before active work is paused. Intensity or hardware may be changed before relocation. Migration is justified only when the avoided burden exceeds checkpointing, transfer, duplicated computation, receiving-site impact and service risk. The receiving facility performs a final local confirmation immediately before execution. The operational sequence is shown in Fig. 2 below.

This continuing step is an additional distinction from a one-time feasible-set construction. A facility that was admissible at t_0 does not retain authority merely because it passed an earlier optimisation. A material threshold crossing, evidence expiry or local veto reopens the decision, recomputes $H(i, W, t)$, and can remove the currently executing option from the admissible set. Conversely, a previously excluded or held option can re-enter only after the relevant condition is resolved and revalidated. Section 6.5 gives a small rolling-horizon illustration of this mechanism.

5 Federated coordination and authority

Electricity, water, cooling conditions and heat users are place-bound, while some computational workloads are movable. This asymmetry creates a legitimate role for federated coordination. Each facility publishes a validated, workload-specific resource state without exposing every sensor, customer record or infrastructure weakness. A network scheduler compares only facilities that have already passed local admissibility and technical eligibility. Jurisdiction, security, latency and transition cost are therefore applied before environmental preference.

Federation does not remove local authority. The facility layer governs physical truth and retains a veto where current conditions make execution unsafe, unlawful or locally irresponsible. The network layer governs comparison among admissible choices. A standard state message should identify the workload class, relevant resource conditions, confidence, evidence age, validity period, protected reserve and binding constraints. Stale or unknown information must remain visible; otherwise, a poorly monitored facility can appear preferable to a facility that reports its limits honestly.

Authority follows the ability to intervene. Facility operators control physical safety, maintenance reserve and local environmental boundaries. Workload owners define genuine deadlines, location restrictions and service requirements. Network schedulers compare eligible options but cannot override a local refusal. Material exceptions require a named decision owner, reason, duration, mitigation and later review. Repeated exceptions are evidence that a threshold, service definition, contract or infrastructure design requires correction.

5.1 Operational overhead and lightweight implementation

ZERO adds control-plane work rather than workload-scale computation. For n candidate facilities, d published state fields and m admissible actions, a direct implementation requires $O(n d)$ state ingestion, $O(n)$ hard-gate evaluations for a fixed rule set, and approximately $O(|A| m)$ scoring after local filtering. The potentially larger costs are operational rather than arithmetic: telemetry acquisition, evidence verification, checkpointing, migration, human exception handling and repeated revalidation. These costs are not measured in the present synthetic study and must be reported explicitly in deployment trials alongside resource outcomes.

A lightweight implementation can limit overhead through local pre-filtering before federation, compact state messages instead of raw telemetry, time-to-live caching for evidence that remains valid, event-driven revalidation rather than continuous global recomputation, hysteresis and minimum dwell times around thresholds, bounded candidate shortlists, incremental re-scoring only for affected facilities, and shadow-mode deployment before any automated control. These design choices preserve local authority while reducing unnecessary network traffic, oscillation and control-loop cost. Future pilots should measure state-message volume, decision latency, CPU and memory overhead, revalidation frequency, migration/checkpoint cost and operator intervention time.

6 Illustrative computational case study

6.1 Purpose and design

The following case study is an illustrative demonstration of the decision logic, not a validation of the framework and not an empirical claim about a specific operator. Its values are synthetic but physically plausible and are reported completely so that the calculation can be reproduced. The purpose is to demonstrate how the stated rules change the feasible decision set relative to local-default execution and carbon-only scheduling when a low-carbon site is subject to a binding water constraint. Because Site B is deliberately defined as water-critical and the ZERO rule excludes new flexible demand under that condition, the scenario should be read as a transparent worked example of the governance sequence, not as evidence of superiority over conventional constrained multi-resource optimisation.

A flexible AI training stage requires 100 MWh of IT energy and must finish within 18 hours. It can be delayed by up to six hours or moved once. The 100 MWh quantity defines the IT-energy accounting boundary of this synthetic example. Facility overhead is not modelled through a site-specific PUE, so the reported electricity-related emissions are not total facility emissions; they represent the stated IT workload plus transition energy multiplied by the receiving-site carbon intensity. For consistency in this simplified example, receiving-site operational water intensity is also applied to the workload plus stated transition energy. A full empirical evaluation would need to include cooling and auxiliary electricity, site-specific PUE or equivalent facility overhead, and the physical location of transfer-related energy use. Three technically capable facilities are available. Site A is the default local facility. Site B has the lowest

electricity carbon intensity but is in a critical water state. Site C has moderate carbon intensity, low water intensity, adequate cooling reserve and an identified heat user. Relocation to Site C adds 3 MWh for checkpointing, transfer and restart. The hard water rule excludes new flexible workloads when the local state is critical. The synthetic case-study inputs are reported in Table 2 below.

Table 2

Synthetic inputs for the base worked example. Values are constructed for reproducibility and are not measurements from an operator; the hard-admissibility result is evaluated before any weighted comparison.

Variable	Site A	Site B	Site C
Electricity carbon intensity (kg CO2e/MWh)	480	140	260
Operational water intensity (m3/MWh IT)	0.30	1.10	0.18
Water state	Normal	Critical	Normal
Cooling reserve	22%	19%	15%
Productive heat-use fraction	0.20	0.10	0.65
Evidence confidence	0.90	0.92	0.85
Transition energy	0 MWh	4 MWh	3 MWh
Hard admissibility result	Pass	Fail: critical water state	Pass

6.2 Baselines and ZERO decision

The local-default baseline executes at Site A. It uses 100 MWh, produces an operational electricity footprint of 48.0 t CO2e and consumes 30.0 m3 of water under the stated intensities. A carbon-only scheduler selects Site B despite 4 MWh of transition energy. Its electricity footprint is 14.56 t CO2e, but its calculated operational water consumption is 114.4 m3 and the action violates the declared critical-water rule. Carbon optimisation therefore identifies the lowest-carbon technical option but not an admissible option.

ZERO removes Site B before scoring. Sites A and C remain. For illustration, the disclosed comparison weights are carbon 0.35, water 0.25, cooling reserve 0.15, productive heat use 0.10, transition burden 0.10 and evidence confidence 0.05. Normalisation is performed only between admissible options. Site C is selected because its lower carbon and water burdens and stronger productive heat opportunity outweigh its transition energy and smaller cooling reserve. The selected action uses 103 MWh, corresponding to 26.78 t CO2e and 18.54 m3 of operational water under the synthetic assumptions. The indicative outcome comparison is reported in Table 3 below.

Table 3

Indicative outcomes for the base worked example under the stated synthetic assumptions. The comparison demonstrates the decision sequence only and does not estimate generalisable savings or establish superiority.

Method	Selected site	Electricity footprint (t CO ₂ e)	Operational water (m ³)	Admissible under stated rules?
Local default	A	48.00	30.00	Yes
Carbon-only scheduling	B	14.56	114.40	No
ZERO	C	26.78	18.54	Yes

6.3 Interpretation and sensitivity

The case study does not validate ZERO or show that it is superior across operating conditions. It illustrates one consequence of the stated governance rule: a hard constraint changes the feasible set and prevents a favourable carbon result from masking a locally unacceptable water condition. Under the synthetic assumptions, the ZERO-selected action has a 44% lower calculated electricity footprint and 38% lower operational water consumption than local-default execution. These percentages are arithmetic properties of this constructed scenario only; they are not measured savings, effect estimates or evidence of generalisable performance.

The result is sensitive to transition energy, evidence confidence and the relative importance assigned to water, carbon, cooling and reliability. That sensitivity is not evidence against the architecture; it is a reason to expose rather than conceal the governing assumptions. The base case therefore should not be read as a unique optimum. Section 6.4 varies both resource conditions and predeclared weight profiles to show which changes affect the soft ranking and which changes bypass ranking because a hard admissibility condition fails.

The case also identifies measurable hypotheses for future work. First, applying hard local constraints before optimisation should reduce environmentally adverse load shifting. Second, explicit evidence expiry should reduce decisions based on stale resource conditions. Third, multi-resource selection should produce different placement decisions from carbon-only scheduling in a non-trivial proportion of cases. Fourth, the benefit should be evaluated after including migration, service and governance overheads.

6.4 Multi-scenario and weighting sensitivity

To provide a limited robustness check without presenting synthetic stress tests as empirical validation, three admissible variants of Sites A and C were rescored using fixed reference normalisation: $z_c = \min(c/500, 1)$, $z_w = \min(w/1.2, 1)$, $z_k = \max((20-k)/20, 0)$, $z_h = 1-h$, $z_t = \min(T/40, 1)$, and $z_v = 1-v$, where lower scores are preferred. These scales are illustrative reference ranges only and are not proposed as universal thresholds. Four predeclared weight profiles were examined in the order (carbon, water, cooling, heat, transition, evidence): balanced (0.30, 0.25, 0.15, 0.10, 0.10, 0.10); drought-priority (0.20,

0.40, 0.15, 0.05, 0.10, 0.10); carbon-priority (0.45, 0.15, 0.10, 0.05, 0.15, 0.10); and reliability-priority (0.05, 0.05, 0.35, 0.05, 0.25, 0.25). The exercise is a deterministic sensitivity analysis, not an AHP or entropy calibration. The resulting sensitivity comparison is shown in Table 4 below.

Table 4

Synthetic weight-sensitivity results among already admissible Sites A and C. Lower score is preferred. In the renewable-rich variant, Site C carbon intensity is reduced from 260 to 120 kg CO₂e/MWh, cooling reserve rises to 18%, and evidence confidence rises to 0.90. In the high-transition variant, Site C transition energy rises from 3 to 30 MWh. The table shows that soft ranking can change with declared priorities; it does not alter hard admissibility.

Weight profile	Weight vector (c,w,k,h,t,v)	Base case	Renewable-rich C	High-transition C
Balanced	(.30,.25,.15,.10,.10,.10)	C: .289 < A: .441	C: .177 < A: .441	C: .356 < A: .441
Drought-priority	(.20,.40,.15,.05,.10,.10)	C: .242 < A: .342	C: .158 < A: .342	C: .309 < A: .342
Carbon-priority	(.45,.15,.10,.05,.15,.10)	C: .325 < A: .520	C: .179 < A: .520	C: .427 < A: .520
Reliability- priority	(.05,.05,.35,.05,.25,.25)	A: .126 < C: .195	C: .116 < A: .126	A: .126 < C: .364

Two additional stress cases demonstrate the non-compensable gate. If Site C enters a declared critical-water state, it is EXCLUDED before scoring and Site A is the remaining admissible choice under every weight profile. If Site C evidence confidence falls to 0.55 under a predeclared minimum binding-confidence requirement of 0.60, Site C is placed on HOLD/CONDITIONAL status and remains outside A until evidence is refreshed. These cases separate sensitivity of the soft ranking from changes that remove an option from the feasible governance state altogether.

6.5 Rolling-horizon revalidation example

A four-step synthetic trace illustrates how authority can change after an initial admission. The same checkpointable workload is assumed to begin at Site C after the base-case decision. A protected cooling-reserve threshold of 10% is declared in advance. Site A remains technically eligible and is rechecked before any relocation. The example isolates revalidation logic; it does not claim that these timings or thresholds are representative of a real operator. The rolling-horizon trace is reported in Table 5 below.

Table 5

Synthetic rolling-horizon revalidation trace. The example demonstrates that an earlier pass is not a permanent licence: a protected-reserve violation removes Site C from the admissible set, while later physical recovery is insufficient until expired evidence is refreshed.

Time	Site C state/evidence	H_C	Trigger	Governed action
t0	water normal; k = 15%; v=.85	1	initial decision	execute at C
t + 4 h	water constrained but within local limit; k = 13%; v=.82	1	state change within admissible range	continue; shorten review interval
t + 7 h	k = 8%, below protected 10% reserve	0	hard threshold crossing / local veto	checkpoint; revalidate A; relocate to A if H_A = 1
t + 10 h	k recovers to 14%, but C evidence TTL has expired	0 / HOLD	evidence expiry	do not return to C until fresh evidence restores H_C = 1

This trace provides the dynamic behaviour that is absent from the static base case. The key output is not the specific relocation itself, but the reopening of authority: H_C changes with current conditions, the local veto is effective during execution, the receiving site is revalidated before use, and recovery of a physical metric does not silently override expired evidence.

7 Evidence, evaluation and claims

ZERO does not replace established indicators. It changes how they enter a decision. Ratios should be reported with absolute quantities, system boundaries and time periods. Electricity demand, water withdrawal, water consumption, thermal generation, captured heat, delivered heat, productive reuse and equipment turnover remain distinct. A composite score can support comparison, but it must not erase the underlying values or a binding constraint.

Governance performance requires its own evidence. Four questions are central. Did relevant resource information reach an actor or system with authority? Did it change admission, timing, location, intensity or recovery when justified? Did the selected action improve the physical outcome relative to a credible alternative? Did the result change later policy, thresholds or infrastructure design? These questions connect decision influence, outcome accuracy and organisational learning. A sophisticated dashboard is not evidence of governance if workload behaviour remains unchanged.

Claims of regeneration require a baseline, additionality and geographical honesty. A heat-recovery project should identify the conventional heat displaced and the energy used for upgrading and delivery. A water claim should distinguish withdrawal from consumption and show whether local pressure changed. A workload move should include transfer overhead and the receiving site's resource state. Independent assurance may strengthen a material claim, but verification cannot substitute for weak physical performance or incomplete boundaries.

8 Validation programme and limitations

The base case, multi-scenario sensitivity exercise and rolling-horizon trace illustrate internal decision logic only. They do not empirically validate numerical thresholds, comparative performance, organisational acceptance, interoperability or measured environmental benefit. Empirical validation should therefore proceed as a staged research workflow. Stage 1 should use retrospective operational traces and numerical simulation to replay admission and scheduling decisions across drought, heat, grid, workload and evidence-quality scenarios, including uncertainty, forecast error, transition cost and threshold sensitivity. Stage 2 should use a shadow-mode or hardware-in-the-loop prototype at a single facility, where ZERO receives realistic telemetry and produces recommendations without operational control. Stage 3 should conduct a controlled real-world pilot for a bounded workload class with human approval retained for interventions, measuring false exclusions, decision latency, service effects and physical resource outcomes. Stage 4 should test cross-federation operation across multiple facilities to examine local vetoes, confidentiality, heterogeneous thresholds, unequal data quality, transfer overhead and the ability of a network scheduler to respect site-level authority. Progression between stages should depend on predeclared safety, reliability and performance criteria rather than on the apparent success of a synthetic scenario.

Across these stages, ZERO should be compared with credible baselines rather than with a system that ignores sustainability. Relevant comparators include local-first execution, carbon-aware scheduling, water-aware scheduling, integrated carbon-water scheduling and unconstrained weighted multi-objective optimisation. Outcomes should include avoided demand, absolute and intensity changes, total facility electricity where available, service performance, transition cost, productive reuse, exception frequency, false exclusions, decision latency, revalidation frequency and learning after error. Distributional effects also matter: a network-level improvement is incomplete if pressure is transferred to a community or watershed with less power to resist it.

Several limitations remain. The present study contains one constructed base scenario, controlled synthetic stress variants and a small rolling-horizon trace; these additions improve internal stress testing but cannot establish robustness, effect size or external validity. Data may be commercially sensitive, unavailable or inconsistent across operators. Carbon and water signals may be measured at different temporal and geographical resolutions, and forecast uncertainty may cause both false exclusions and unsafe admissions. Resource states can conflict, while normalisation and weight selection can influence ranking among admissible options. The sensitivity analysis uses declared illustrative weights rather than calibrated AHP or entropy weights because no expert-elicitation dataset or empirical decision matrix is available. Customers may classify most work as urgent; contracts may prevent movement or interruption; and operators may have incentives to report states or exceptions strategically. Local vetoes can be misused, while central platforms may favour utilisation over local evidence. Cross-facility federation introduces additional questions of data interoperability, trust, jurisdiction, security and unequal institutional capacity. Governance itself also consumes time, computation, network capacity

and administrative effort. The present paper provides analytical complexity and mitigation principles, but does not measure these overheads; they must be quantified in prototype and pilot evaluation.

The framework does not determine universal thresholds. A critical water state, protected cooling reserve or acceptable evidence age must be defined by local engineering, regulation and public-resource conditions. Nor does ZERO resolve the political question of who has authority to limit computational demand. It makes that authority visible and records its use. Human judgement, regulation and infrastructure investment remain necessary where evidence is incomplete or consequences extend beyond operator boundaries.

9 Conclusion

AI data-centre sustainability cannot be reduced to efficient equipment, annual reporting or a single environmental objective. Those approaches provide essential evidence, but they address only part of the decision. ZERO places resource governance before and during execution. It asks whether demand is necessary, whether it can be reduced or aligned, whether a facility is admissible under current local conditions, whether residual value reaches productive use, and whether the evidence remains valid.

The revised framework is operational in the limited scientific sense required at this stage: its inputs, constraints, candidate actions, selection process, evidence requirements and continuation triggers are explicit. The base case demonstrates how admissibility can precede optimisation under a declared rule set, while synthetic sensitivity and rolling-horizon examples show how soft priorities and changing conditions affect the decision without being treated as empirical validation. A low-carbon option can still be rejected when a binding water, cooling or evidence condition is present. The next step is staged testing with retrospective data, numerical simulation, shadow-mode deployment, controlled real-world pilots and cross-federation trials. Until such evidence exists, ZERO should be understood as a testable governance architecture rather than a demonstrated source of measured savings or proven superiority over established optimisation approaches.

Statements and Declarations

Funding: The author received no funding for this work.

Competing interests: The author has no relevant financial or non-financial interests to disclose.

Author contributions: Maria Kollia developed the conceptual framework, conducted the literature and policy synthesis, designed the illustrative case study, prepared the figures, and wrote and reviewed the manuscript.

Data availability: No empirical dataset was generated or analysed. All values used in the illustrative case study are synthetic and are reported in the manuscript.

Ethics approval: Not applicable.

Acknowledgements: OpenAI ChatGPT was used during manuscript preparation for English language drafting and editing. The conceptual framework, scientific argument, source selection, interpretation, conclusions and final verification are the author's own.

Prior dissemination and originality: This Perspective develops a new synthesis from the author's earlier conceptual research deposited in Zenodo. Those works are cited explicitly. The revised operational formulation, comparison tables and illustrative case study were prepared for this submission and are not under consideration elsewhere.

References

1. Andrews D, Kerwin K (2026) Design for circularity: a data centre equipment case study. *Mineral Economics* 39:53-70. <https://doi.org/10.1007/s13563-025-00587-7>
2. Ebrahimi K, Jones GF, Fleischer AS (2014) A review of data center cooling technology, operating conditions and the corresponding low-grade waste heat recovery opportunities. *Renewable and Sustainable Energy Reviews* 31:622-638. <https://doi.org/10.1016/j.rser.2013.12.007>
3. Hall S, Micheli F, Belgioioso G, Radovanovic A, Dörfler F (2025) Carbon-aware computing for data centers with probabilistic performance guarantees. *IEEE Transactions on Power Systems*. <https://doi.org/10.1109/TPWRS.2025.3630499>
4. International Energy Agency (2025) Energy and AI. IEA, Paris
5. Islam MA, Ahmed K, Xu H, Tran N, Quan G, Ren S (2018) Exploiting spatio-temporal diversity for water saving in geo-distributed data centers. *IEEE Transactions on Cloud Computing* 6(3):734-746. <https://doi.org/10.1109/TCC.2016.2535201>
6. Kollia M (2026a) Zero-Waste AI Data Centers: Rethinking Resource Consumption in Artificial Intelligence Infrastructure. Zenodo. <https://doi.org/10.5281/zenodo.20750510>
7. Kollia M (2026b) Circular AI Infrastructure: A Resource Recovery Framework for Future Data Centers. Zenodo. <https://doi.org/10.5281/zenodo.20927844>
8. Lei N, Lu J, Shehabi A, Masanet ER (2025) The water use of data center workloads: a review and assessment of key determinants. *Resources, Conservation and Recycling* 219:108310. <https://doi.org/10.1016/j.resconrec.2025.108310>
9. Liu Z, Chen Y, Bash CE, Wierman A, Gmach D, Wang Z, Marwah M, Hyser C (2012) Renewable and cooling aware workload management for sustainable data centers. *ACM SIGMETRICS Performance Evaluation Review* 40(1):175-186. <https://doi.org/10.1145/2254756.2254779>
10. Masanet E, Shehabi A, Lei N, Smith S, Koomey J (2020) Recalibrating global data center energy-use estimates. *Science* 367:984-986. <https://doi.org/10.1126/science.aba3758>
11. Moore H, Qi S, Hogade N, Milojevic DS, Bash CE, Pasricha S (2025) Sustainable carbon-aware and water-efficient LLM scheduling in geo-distributed cloud datacenters. *Proceedings of the Great Lakes Symposium on VLSI*, pp 929-934. <https://doi.org/10.1145/3716368.3735301>

12. Mytton D (2021) Data centre water consumption. *npj Clean Water* 4:11.
<https://doi.org/10.1038/s41545-021-00101-w>
13. Radovanovic A, Koningstein R, Schneider I et al (2023) Carbon-aware computing for datacenters. *IEEE Transactions on Power Systems*. <https://doi.org/10.1109/TPWRS.2022.3173250>
14. Ruparel H (2026) Carbon-aware scheduling and distributionally robust optimization for cloud systems: forecasting, battery storage, and adaptive ambiguity refinement. *Journal of Cloud Computing* 15:84. <https://doi.org/10.1186/s13677-026-00904-7>
15. Shehabi A, Smith SJ, Hubbard A et al (2024) 2024 United States Data Center Energy Usage Report. Lawrence Berkeley National Laboratory. <https://doi.org/10.71468/P1WC7Q>
16. Siddik MAB, Shehabi A, Marston L (2021) The environmental footprint of data centers in the United States. *Environmental Research Letters* 16:064017. <https://doi.org/10.1088/1748-9326/abfba1>
17. Sukprasert T, Souza A, Bashir N, Irwin D, Shenoy P (2024) On the limitations of carbon-aware temporal and spatial workload shifting in the cloud. *Proceedings of EuroSys 2024*, pp 924-941. <https://doi.org/10.1145/3627703.3650079>
18. Wahlroos M, Pärssinen M, Rinne S, Syri S, Manner J (2018) Future views on waste heat utilization: case of data centers in Northern Europe. *Renewable and Sustainable Energy Reviews* 82:1749-1764. <https://doi.org/10.1016/j.rser.2017.10.058>
19. Wang P, Cao Y, Ding Z (2020) Flexible multi-energy scheduling scheme for data center to facilitate wind power integration. *IEEE Access* 8:88876-88891. <https://doi.org/10.1109/ACCESS.2020.2990454>
20. Wang G, Mi D, He P, Wang J et al (2026) Deep Reinforcement Learning-Based Migration and Scheduling Strategy for Data Centers in Electricity and Carbon Markets. *IEEE Transactions on Industry Applications*, early access. <https://doi.org/10.1109/TIA.2026.3691718>
21. Yuan X, Liang Y, Hu X, Xu Y, Chen Y, Kosonen R (2023) Waste heat recoveries in data centers: a review. *Renewable and Sustainable Energy Reviews* 188:113777. <https://doi.org/10.1016/j.rser.2023.113777>

Figures

ZERO RESOURCE-GOVERNANCE HIERARCHY

Ordered stages; later stages cannot compensate for avoidable failure at an earlier stage

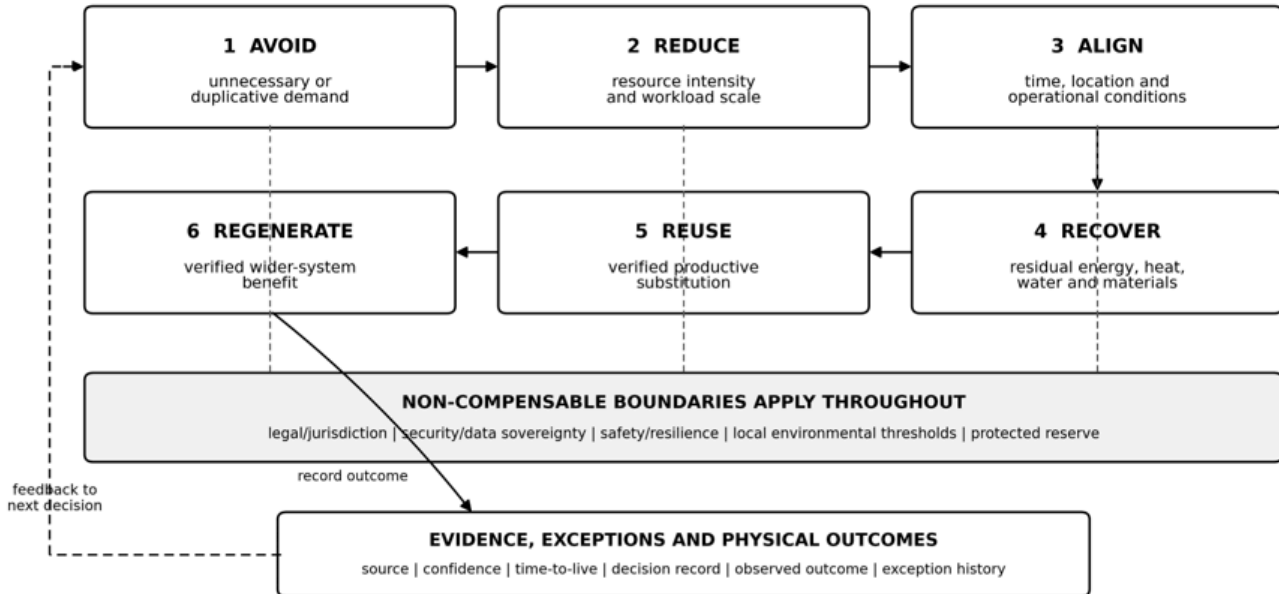


Figure 1

The ZERO resource-governance hierarchy. Avoid, reduce, align, recover, reuse and regenerate are ordered stages. Binding legal, jurisdictional, security, data-sovereignty, safety, resilience and local environmental constraints are non-compensable and apply throughout; evidence, exceptions and observed outcomes feed back into subsequent decisions.

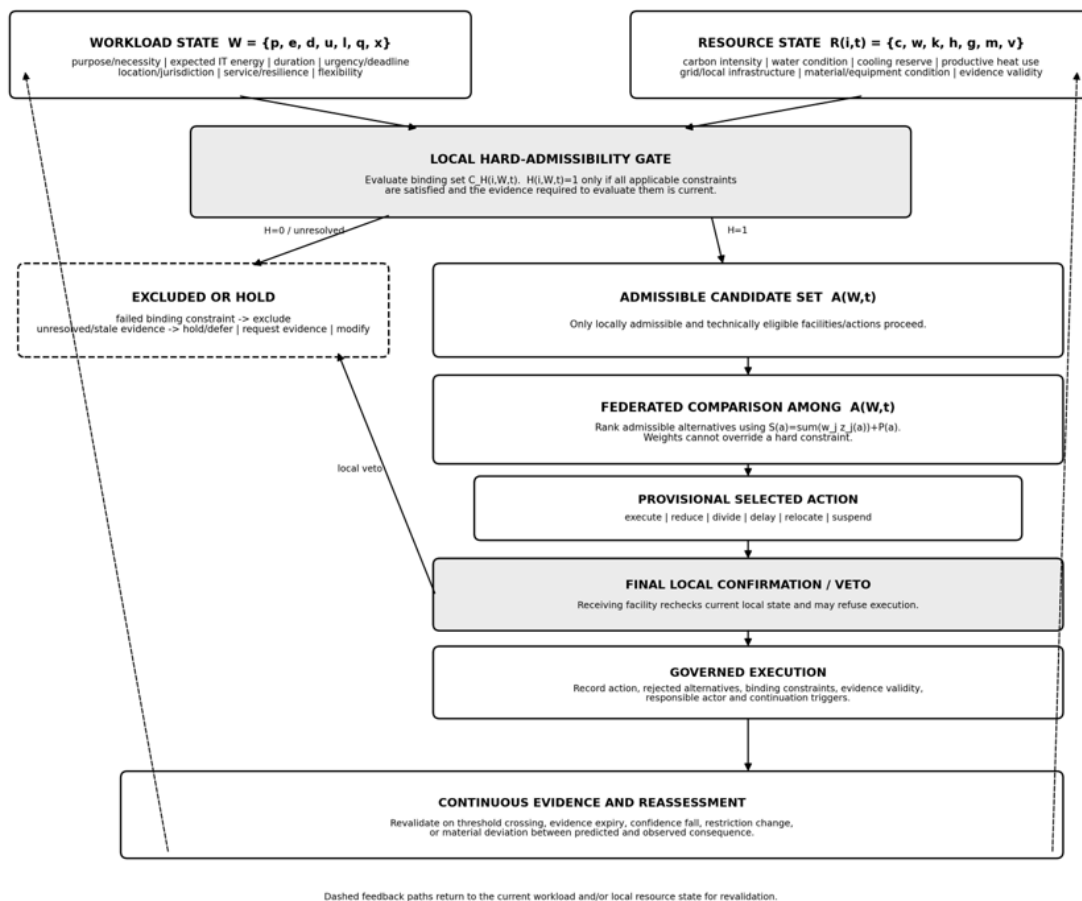


Figure 2

Formal operational sequence for local and federated ZERO governance. The workload state W and local resource state $R(i,t)$ feed a local hard-admissibility gate. Only admissible candidates enter federated comparison. The receiving facility retains a final local veto, while evidence validity and operational monitoring trigger revalidation when material conditions change.